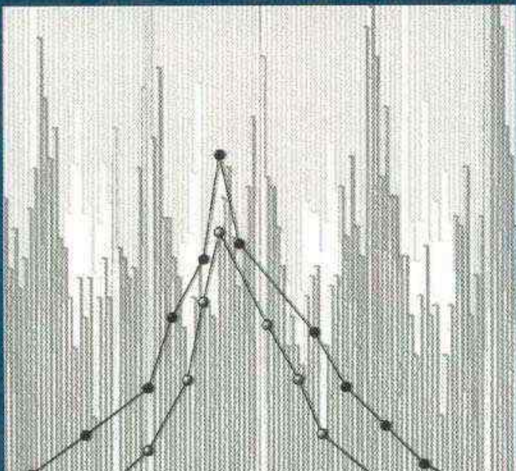
Advanced Series on Statistical Science & Applied Probability

Vol. 3

ESSENTIALS OF STOCHASTIC FINANCE

Facts, Models, Theory

Albert N. Shiryaev



World Scientific

ESSENTIALS OF STOCHASTIC FINANCE

Facts, Models, Theory

ADVANCED SERIES ON STATISTICAL SCIENCE & APPLIED PROBABILITY

Editor: Ole E. Barndorff-Nielsen

Published

Vol. 1: Random Walks of Infinitely Many Particles
by P. Revesz

Vol. 3: Essentials of Stochastic Finance: Facts, Models, Theory
by Albert N. Shiryaev

Vol. 4: Principles of Statistical Inference from a Neo-Fisherian Perspective
by L. Pace and A. Salvan

Vol. 5: Local Stereology
by Eva B. Vedel Jensen

Vol. 6: Elementary Stochastic Calculus — With Finance in View
by T. Mikosch

Vol. 7: Stochastic Methods in Hydrology: Rain, Landforms and Floods
eds. O. E. Barndorff-Nielsen et al.

Forthcoming

Vol. 2: Ruin Probability
by S. Asmussen

Advanced Series on
Statistical Science &
Applied Probability
Vol. 3

ESSENTIALS OF STOCHASTIC FINANCE

Facts, Models, Theory

Albert N. Shiryaev
Steklov Mathematical Institute and Moscow State University

Translated from the Russian by
N. Kruzhilin

World Scientific
Singapore • New Jersey • London • Hong Kong

Published by

World Scientific Publishing Co. Pte. Ltd.

P O Box 128, Farrer Road, Singapore 912805

USA office: Suite 1B, 1060 Main Street, River Edge, NJ 07661

UK office: 57 Shelton Street, Covent Garden, London WC2H 9HE

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

ESSENTIALS OF STOCHASTIC FINANCE: FACTS, MODELS, THEORY

ISBN 981-02-3605-0

Printed in Singapore.

Contents

Foreword

The author's intention was:

- *to select and expose subjects that can be necessary or useful to those interested in stochastic calculus and pricing in models of financial markets operating under uncertainty*;
- *to introduce the reader to the main concepts, notions, and results of stochastic financial mathematics*;
- *to develop applications of these results to various kinds of calculations required in financial engineering.*

The author considered it also a major priority to answer the requests of teachers of financial mathematics and engineering by making a bias towards probabilistic and statistical ideas and the methods of stochastic calculus in the analysis of *market risks*.

The subtitle "Facts, Models, Theory" appears to be an adequate reflection of the text structure and the author's style, which is in large measure a result of the 'feedback' with students attending his lectures (in Moscow, Zürich, Aarhus, ...).

For instance, an audience of mathematicians displayed always an interest not only in the mathematical issues of the 'Theory', but also in the 'Facts', the particularities of real financial markets, and the ways in which they operate. This has induced the author to devote the *first chapter* to the description of the key objects and structures present on these markets, to explain there the goals of financial theory and engineering, and to discuss some issues pertaining to the history of probabilistic and statistical ideas in the analysis of financial markets.

On the other hand, an audience acquainted with, say, securities markets and securities trading showed considerable interest in various classes of stochastic processes used (or considered as prospective) for the construction of models of the dynamics of financial indicators (prices, indexes, exchange rates, ...) and important for calculations (of risks, hedging strategies, rational option prices, etc.).

This is what we describe in the *second* and the *third chapters*, devoted to stochastic 'Models' both for discrete and continuous time.

The author believes that the discussion of stochastic processes in these chapters will be useful to a broad range of readers, not only to the ones interested in financial mathematics.

We emphasize here that in the discrete-time case, we usually start in our description of the evolution of stochastic sequences from their *Doob decomposition* into *predictable* and *martingale* components. One often calls this the 'martingale approach'. Regarded from this standpoint, it is only natural that martingale theory can provide financial mathematics and engineering with useful tools.

The concepts of 'predictability' and 'martingality' permeating our entire exposition are incidentally very natural from economic standpoint. For instance, such economic concepts as *investment portfolio* and *hedging* get simple mathematical definitions in terms of 'predictability', while the concepts of *efficiency* and *absence of arbitrage* on a financial market can be expressed in the mathematical language, by making references to martingales and martingale measures (the *First fundamental theorem* of asset pricing theory; Chapter V, § 2b).

Our approach to the description of stochastic sequences on the basis of the Doob decomposition suggests that in the *continuous-time* case one could turn to the (fairly broad) class of *semimartingales* (Chapter III, § 5a). Representable as they are by sums of processes of bounded variation ('slowly changing' components) and local martingales (which can often be 'fast changing', as is a Brownian motion, for example), semimartingales have a remarkable property: one can define stochastic integrals with respect to these processes, which, in turn, opens up new vistas for the application of stochastic calculus to the construction of models in which financial indexes are simulated by such processes.

The *fourth* ('statistical') *chapter* must give the reader a notion of the statistical 'raw material' that one encounters in the empirical analysis of financial data.

Based mostly on currency cross rates (which are established on a global, probably the largest, financial market with daily turnover of several hundred billion dollars) we show that the 'returns' (see (3) in Chapter II, § 1a) have distribution densities with 'heavy tails' and strong 'leptokurtosis' around the mean value. As regards their behavior in time, these values are featured by the 'cluster property' and 'strong aftereffect' (we can say that 'prices keep memory of their past'). We demonstrate the fractal structure of several characteristic of the volatility of the 'returns'.

Of course, one must take all this into account if one undertakes a construction of a model describing the actual dynamics of financial indexes; this is extremely important if one is trying to *foresee* their development in the future.

'Theory' in general and, in particular, *arbitrage theory* are placed in the *fifth chapter* (discrete time) and the *seventh chapter* (continuous time).

Central points there are the *First* and the *Second fundamental asset pricing theorems.*

The *First theorem* states (more or less) that a financial market is *arbitrage-free* if and only if there exists a so-called *martingale* (*risk-neutral*) *probability measure* such that the (discounted) prices make up a martingale with respect to it. The *Second theorem* describes arbitrage-free markets with property of *completeness*, which ensures that one can build an investment portfolio of value replicating *faithfully* any given pay-off.

Both theorems deserve the name *fundamental* for they assign a precise mathematical meaning to the economic notion of an 'arbitrage-free' market on the basis of (well-developed) *martingale theory.*

In the *sixth* and the *eighth chapters* we discuss pricing based on the First and the Second fundamental theorems. Here we follow the tradition in that we pay much attention to the calculation of rational prices and hedging strategies for various kinds of (European or American) *options*, which are *derivative* financial instruments with best developed pricing theory. Options provide a perfect basis for the understanding of the general principles and methods of pricing on arbitrage-free markets.

Of course, the author faced the problem of the choice of 'authoritative' data and the mode of presentation.

The above description of the contents of the eight chapters can give one a measure for gauging the spectrum of selected material. However, for all its bulkiness, our book leaves aside many aspects of financial theory and its applications (e.g., the classical theories of von Neumann–Morgenstern and Arrow–Debreu and their updated versions considering investors' behavior delivering the maximum of the 'utility function', and also computational issues that are important for applications).

As the reader will see, the author often takes a lecturer's stance by making comments of the 'what-where-when' kind. For discrete time we provide the proofs of essentially all main results. On the other hand, in the continuous-time case we often content ourselves with the statements of results (of martingale theory, stochastic calculus, etc.) and refer to a suitable source where the reader can find the proofs.

The suggestion that the author could write a book on financial mathematics for *World Scientific* was put forward by Prof. Ole Barndorff-Nielsen at the beginning of 1995. Although having accepted it, it was not before summer that the author could start drafting the text. At first, he had in mind to discuss only the *discrete-time* case. However, as the work was moving on, the author was gradually coming to the belief that he could not give the reader a full picture of financial mathematics and engineering without touching upon the *continuous-time* case. As a result, we discuss both cases, discrete and continuous.

This book consists of two parts. The first ('Facts. Models') contains Chapters I–IV. The second ('Theory') includes Chapters V–VIII.

The writing process took around two years. Several months went into typesetting, editing, and preparing a camera-ready copy. This job was done by I. L. Legostaeva, T. B. Tolozova, and A. D. Izaak on the basis of the Information and Publishing Sector of the Department of Mathematics of the Russian Academy of Sciences. The author is particularly indebted to them all for their expertise and selfless support as well as for the patience and tolerance they demonstrated each time the author came to them with yet another 'final' version, making changes in the already typeset and edited text.

The author acknowledges the help of his friends and colleagues, in Russia and abroad; he is also grateful to the Actuarial and Financial Center in Moscow, VW-Stiftung in Germany, the Mathematical Research Center and the Center for Analytic Finance in Aarhus (Denmark), INTAS, and the A. Lyapunov Institute in Paris and Moscow for their support and hospitality.

Moscow
1995 – 1997

A. Shiryaev
Steklov Mathematical Institute
Russian Academy of Sciences
and
Moscow State University

Part 1

FACTS

MODELS

Chapter I. Main Concepts, Structures, and Instruments. Aims and Problems of Financial Theory and Financial Engineering

1. Financial Structures and Instruments

In modern view (see, e.g., [79], [342], and [345]) *financial theory* and *engineering* must analyze the properties of financial structures and find most sensible ways to operate financial resources using various financial instruments and strategies with due account paid to such factors as *time*, *risks*, and (usually, random) *environment*.

Time, *dynamics*, *uncertainty*, *stochastics*: it is thanks to these elements that probabilistic and statistical theories, such as, e.g.,

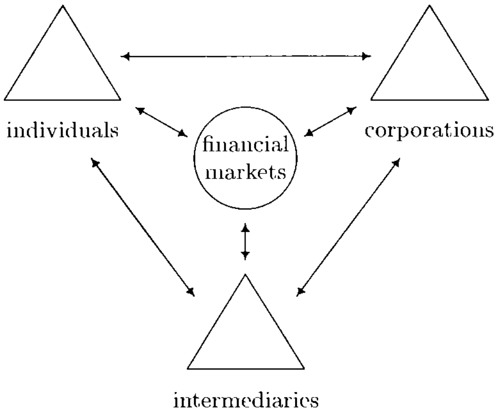
- the theory of stochastic processes,
- stochastic calculus,
- statistics of stochastic processes,
- stochastic optimization

make up the machinery used in this book and adequate to the needs of financial theory and engineering.

§ 1a. Key Objects and Structures

1. We can distinguish the following *basic objects* and *structures* of financial theory that define and explain the specific nature of financial problems, the aims, and the tools of financial mathematics and engineering:

- *individuals*,
- *corporations*,
- *intermediaries*,
- *financial markets*.



As shown in the chart, the theory and practice of finance assign the central role among the above four structures to *financial markets*; they are the structures of the primary concern for the mathematical theory of finance in what follows.

2. Individuals. Their financial activities can be described in terms of the dilemma '*consumption–investment*'. The ambivalence of their behavior as both consumers ('consume more now') and investors ('invest now to get more in the future') brings one to optimization problems formulated in mathematical economics as *consumption—saving* and *portfolio decision making*. In the framework of *utility theory* the first problem is treated on the basis of the (von Neumann–Morgenstern) postulates of the *rational* behavior of individuals under uncertainty. These postulates determine the approaches and methods used to determine *preferable* strategies by means of a quantitative analysis, e.g., of the mean values of the *utility functions*. The problem of 'portfolio decision making' confronting individuals can roughly be described as the problem of the best allocation (investment) of funds (with due attention to possible risks) among, say, property, gold, securities (bonds, stock, options, futures, etc.), and the like. The idea of diversification (see § 2b) in building a portfolio is reflected by such well-known adages as "Don't put all your eggs in one basket" or "Nothing ventured, nothing gained". In what follows we describe various opportunities (depending on the starting capital) opening for an individual on a securities market.

Corporations (companies, firms), who own such 'perceptible' valuables as 'land', 'factories', 'machines', but are also proprietors of 'organization structures', 'patents', etc., organize businesses, maintain business relations, and manage manufacturing. To raise funds for the development of manufacturing, corporations occasionally issue stock or bonds (which governments also do). Corporate management must be directed towards meeting the interests of shareholders and bondholders.

Intermediaries (intermediate financial structures). These are banks, investment companies (of mutual funds kind), pension funds, insurance companies, etc. One can put here also stock exchanges, option and futures exchanges, and so on.

Among the world's most renowned exchanges (as of 1997) one must list the USA-based NYSE (the New York Stock Exchange), AMEX (the American Stock Exchange), NASDAQ (the NASDAQ Stock Market), NYFE (the New York Futures Exchange), CBOT (the Chicago Board of Trade), etc.[a]

3. Financial markets include money and Forex markets, markets of precious metals, and markets of financial instruments (including securities).

In the market of financial instruments one usually distinguishes

- **underlying** (primary) instruments and
- **derivative** (secondary) instruments;

the latter are hybrids constructed on the basis of underlying (more elementary) instruments.

The *underlying* financial instruments include the following securities:

- bank accounts,
- bonds,
- stock.

The *derivative* financial instruments include

- options,
- futures contracts,
- warrants,
- swaps,
- combinations,
- spreads,
- ...

We note that *financial engineering* is often understood precisely as manipulations with *derivative* securities (in order to *raise capital* and *reduce risks* caused by the uncertain character of the market situation in the future).

We now describe several main ingredients of financial markets.

---

[a]Throughout the book we turn fairly often to American financial structures and to financial activities taking place in the USA. The main reason is that the American financial markets have deep-rooted traditions ('the Wall Street'!) and, at the same time, these are the markets where many financial innovations are put to test. Moreover, there exists an enormous literature describing these markets: periodicals as well as monographs, primers, handbooks, investor's guides, and so on. The reader can easily check through the bibliography at the end of this book.

§ 1b. Financial Markets

1. Money dates back to those ages when people learned to trade 'things they had' for 'things they wanted to get'. This mechanism works also these days—we give money in exchange for goods (and services), and the salesmen, in turn, use it to buy other things.

Modern technology means a revolution in the ways of money circulation. Only 8% of all dollars that are now in circulation in the USA are banknotes and coins. The main bulk of payments in retailing and services are carried over by checks and plastic cards (and proceed by wires).

Besides its function of a 'circulation medium', money plays an important role of a 'measure of value' and a 'saving means' [108].

2. Foreign currency, i.e., the *currency of other nations* (its reserves, cross rates, and so on) is an important indicator of the nation's well-being and development and is often a *means of payment* in foreign trade.

Economic globalization brings into being *monetary unions* of several nations agreeing to harmonize their monetary and credit policies and regulating exchange rates between their currencies.

One example is the well-known Bretton-Woods credit and currency system. In 1944, Bretton-Woods (New Hampshire, USA) hosted a conference of the major participants in the international trade, who agreed to maintain a currency system (the 'Bretton-Woods system') in which the exchange rates could deviate from their officially declared levels only in very narrow ranges: 1% in either side. (These parity cross rates were fixed against the USA dollar.) To launch this system and oversee it, the nations concerned established the International Monetary Foundation (IMF).

However, with the financial and currency crisis of 1973 affecting the major currencies (the USA dollar, the German mark, the Japanese yen), the Bretton-Woods system was acknowledged to have exhausted its potentialities, and it was replaced by *floating* exchange rates.

In March 1979, the majority of the European Union (EU) nations created the European Monetary System. It is stipulated in this system that the variations of the cross rates of the currencies involved should lie, in general, in the band of $\pm 2.25\%$ around the official parity rates. If the cross rates of some currencies are deemed under the threat of leaving this band, then the corresponding central banks must *intervene* to prevent this course of events and ensure the *stability* of exchange rates. (This explains why one often calls systems of this kind 'systems of *adjusted* floating rates'.)

Other examples of monetary unions can be found between some Caribbean, Central American, and South American, countries, which peg exchange rates against some powerful 'leader currency'. (For greater detail, see [108; pp. 459–468].)

3. Precious metals, i.e., gold, silver, platinum, and some other (namely, the metals in the platinum group: iridium, osmium, palladium, rhodium, ruthenium) played throughout the history (in particular, in the 19th and the beginning of the 20th century) and still play today an important role in the international credit and currency system.

According to [108] the age of gold standard began in 1821, when Britain proclaimed pound sterling convertible into gold. The United States did the same soon afterwards. The gold standard came into full blossom in 1880–1914, but it could never recover its status after the World War I. Its traces evaporated completely in 1971, when the US Treasury abandoned formally the practice of buying and selling gold at a fixed price.

Of course, gold keeps an important position in the international currency system. For example, governments often use gold to pay foreign debts.

It follows from the above that one can clearly distinguish three phases in the development of the international monetary system: 'gold standard', 'the Bretton-Woods system', and the 'system of adjusted floating exchange rates'.

4. Bank account. We can regard it as a security of bond kind (see subsection 5 below) that reduces in effect to the bank's obligation to pay certain interest on the sum put into one's account. We shall often consider *bank accounts* in what follows, primarily because it is a convenient 'unit of measurement' for the prices of various securities.

One usually distinguishes two ways to pay interest:

- m times a year (*simple interest*),
- continuously (*compound interest*).

If you open a bank account paying interest *m times a year* with *interest rate* $r(m)$, then on having put an initial capital B_0, in N years you obtain the amount

$$B_N(m) = B_0\left(1 + \frac{r(m)}{m}\right)^{mN}, \tag{1}$$

while in $N + k/m$ years' time (a fractional value, $0 \leqslant k \leqslant m$), your capital will be

$$B_{N+k/m}(m) = B_0\left(1 + \frac{r(m)}{m}\right)^{m(N+k/m)}.$$

In the case of compound interest with *interest rate* $r(\infty)$ the starting capital B_0 grows in N years into

$$B_N(\infty) = B_0 e^{r(\infty)N}. \tag{2}$$

Clearly,

$$B_N(m) \to B_N(\infty)$$

as $r(m) \to r(\infty)$ and $m \to \infty$.

If the compound interest $r(\infty)$ is equal to r, then the adequate 'rate of interest payable m times a year' $r(m)$ can be found by the formula

$$r(m) = m(e^{r/m} - 1), \tag{3}$$

while the compound rate $r = r(\infty)$ corresponding to fixed $r(m)$ can be found by the formula

$$r = m \ln\left(1 + \frac{r(m)}{m}\right). \tag{4}$$

In the particular case of $m = 1$, setting $\widehat{r} = r(1)$ and $r = r(\infty)$ we obtain the following conversion formulas for these rates:

$$\widehat{r} = e^r - 1, \qquad r = \ln(1 + \widehat{r}). \tag{5}$$

Besides the 'annual interest rate $\widehat{r}$', the bank can announce also the value of the 'annual discount rate $\widehat{q}$', which means that one must put $B_0 = B_1(1 - \widehat{q})$ into a bank account to obtain an amount $B_1 = B_1(1)$ a year later. The relation between $\widehat{r}$ and $\widehat{q}$ is straightforward:

$$(1 - \widehat{q})(1 + \widehat{r}) = 1,$$

therefore

$$\widehat{q} = \frac{\widehat{r}}{1 + \widehat{r}}.$$

t \ m	1	2	3	4	6	12	∞
0	10000	10000	10000	10000	10000	10000	10000
1	11000	11025	11034	11038	11043	11047	11052
2	12100	12155	12174	12184	12194	12204	12214
3	13310	13401	13433	13449	13465	13482	13499
4	14641	14775	14821	14845	14869	14894	14918
5	16105	16289	16353	16386	16419	16453	16487
6	17716	17959	18044	18087	18131	18176	18221
7	19487	19799	19909	19965	20022	20079	20138
8	21436	21829	21967	22038	22109	22182	22255
9	23579	24066	24238	24325	24414	24504	24596
10	25937	26533	26743	26851	26960	27070	27183

One can also ask about the time necessary (under the assumption of continuously payable interest $r = a/100$) to raise our capital twofold. Clearly, we can determine N from the relation

$$2 = e^{aN/100},$$

i.e.,

$$N = \frac{\ln 2 \cdot 100}{a} \approx \frac{70}{a}.$$

In practice (in the case of interest payable, say, twice a year) one often uses the so-called *rule* 72: if the interest rate is $a/100$, then capital doubles in $72/a$ years.

To help the reader make an idea of the growth of an investment for various modes of interest payment ($m = 1, 2, 3, 4, 6, 12, \infty$) we present a table (see above) of the values of $B_t(m)$ (here $B_0 = 10000$) corresponding to $t = 0, 1, \dots, 10$ and $r(m) = 0.1$ for all m.

5. Bonds are *promissory* notes issued by a government or a bank, a corporation, a joint stock company, any other financial establishment to raise capital.

Bonds are fairly popular in many countries, and the total funds invested in bonds are larger than the funds invested in stock or other securities. Their main attraction (especially for a conservative investor) is as follows: the interest on bonds is fixed and payable on a regular basis, and the repayment of the entire loan at a specified time is guaranteed. Of course, one cannot affirm beyond all doubts that government or corporate bonds are *risk-free* financial instruments. A certain degree of risk is always here: for instance, the corporation may go bust and default on interest payments. For that reason government bonds are less risky than corporate ones, but the coupon yield on corporate bonds is larger.

An investor in bonds is naturally eager to know the *riskiness* of the corporations he considers for the purpose of bond purchase. Ratings of various issuers of bonds can be found in several publications ("Standard & Poor's Bond Guide" for one). Corporations considered less risky (i.e., ones with higher ratings) pay lower interest. Accordingly, corporations with low ratings must issue bonds with higher interest rate to attract investors.

We can characterize a bond issued at time $t = 0$ by several numerical indexes ((i)–(vii)):

(i) the *face value* (par value) $\mathsf{P}(T,T)$, i.e., the sum payable to the holder of the bond at

(ii) the *maturity date* (the year the bond matures) T; (the time to maturity is usually a year or shorter for *short-term* bonds, 2 to 10 years for *middle-term* ones, and $T \geqslant 30$ years for *long-term* bonds);

(iii) the *bond's interest rate* (coupon yield) r_c defining the *dividends*, the amount payable to its holder by the issuer, by the formula $r_c \times$(face value);

(iv) the *original price* $\mathsf{P}(0,T)$ of the bond of T-year maturity issued at time $t = 0$.

If the bond issued at time $t = 0$ has, say, 10-year maturity, the face value \$ 1000, and the coupon yield $r = 6/100$ ($= 6\%$), then having purchased it at the original (purchase) price of \$ 1000 we shall in ten years receive the profit of \$ 600, which is the sum of the following terms:

face value		\$ 1000
+		
interest for 10 years	$1000 \times 6\% \times 10 =$	\$ 600
−		
purchase price		\$ 1000
=		\$ 600

Of course, the holder of a bond of T-year maturity who bought it at time $t = 0$ can keep it all these T years for himself collecting the interest and the face value (at time T). On the other hand he may consider it unprofitable to keep this bond till maturity (e.g., if the *inflation* rate r_{inf} is greater than r_c). In that case the bondholder can use his right to sell the bond (with maturity date T) at its

(v) *market value* $\mathsf{P}(t,T)$

at time t (in general, t can be an arbitrary instant between 0 and T).

By definition, $\mathsf{P}(0,T)$ is the original price of the bond at the moment of flotation, while $\mathsf{P}(T,T)$ is clearly its face value. (Both are equal to \$ 1000 in the above example.) Although it is in principle possible that the market value $\mathsf{P}(t,T)$ is larger than the face value $\mathsf{P}(T,T)$, one typically has the inequality $\mathsf{P}(t,T) \leqslant \mathsf{P}(T,T)$.

Assume now that the bondholder decides to sell the bond two years after the flotation, and the market value $\mathsf{P}(2,10)$ (here $t = 2$ and $T = 10$) of the bond is \$ 800. Then his profit from having purchased the bond and having held it for two years is as follows:

market value		\$ 800
+		
interest	$1000 \times 6\% \times 2 =$	\$ 120
−		
purchase price		\$ 1000
=		−\$ 80

Thus, his profits are in fact *negative*: his losses amount to \$ 80. On the other hand an investor who buys this bond at its market value $\mathsf{P}(2,10) =$ \$ 800 will pocket the following return in 8 years:

face value		\$ 1000
+		
interest	$1000 \times 6\% \times 8 =$	\$ 480
−		
purchase price		\$ 800
=		\$ 680

We must emphasize that the interest r_c on the bond, its coupon yield, is *fixed* over its life time, whereas its market value $\mathsf{P}(t,T)$ *wavers*. This is the result of the influence of multiple economic factors: demand and supply, interest payable on other securities, speculators' activities, etc. Regarding $\{\mathsf{P}(t,T)\}$, $0 \leqslant t \leqslant T$, as a *random process* evolving in time we see that this is a *conditional* process: the value of $\mathsf{P}(T,T)$ is fixed (and equal to the face value of the bond).

In Fig. 1 we depict possible fluctuations on the time interval $0 \leqslant t \leqslant T$ of the market value $\mathsf{P}(t,T)$ that takes the prescribed face value $\mathsf{P}(T,T)$ at the maturity date T.

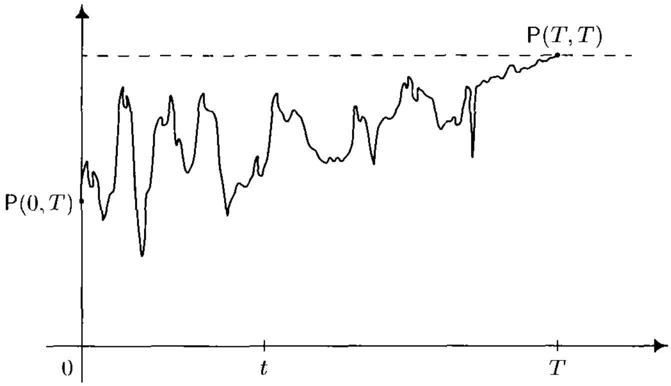


FIGURE 1. Evolution of the market value $\mathsf{P}(t,T)$ of a bond

A frequently used characteristic of a bond is

(vi) the *current yield*

$$r_c(t,T) = \frac{r_c \cdot \mathsf{P}(T,T)}{\mathsf{P}(t,T)}, \qquad 0 < t < T,$$

the ratio of the yearly interest and the (current) market value,

which is important for the *comparison* of different bonds. (In the above example $r_c(0,10) = r_c = 6\,\%$ and $r_c(2,10) = 6\,\% \cdot 1000/800 = 7.5\,\%$.)

Another, probably the most important, characteristic of a bond, which enables one to estimate the *returns* from both the final repayment and the interest payments (due for the remaining time to maturity) and offers thus yet another opportunity for the comparison of different bonds, is

(vii) the *yield to maturity* (on a percentage basis) or the *profitability*

$$\rho\,(T-t,T)$$

(here $T-t$ is the remaining life time of the bond); its value must ensure that the sum of the discounted values of the interest payable on the interval

$(t, T]$ and the face value is the current market value of the bond. In other words, $\rho = \rho(T - t, T)$ is the root of the equation

$$\mathsf{P}(t,T) = \sum_{k=1}^{T-t} \frac{r_c\,\mathsf{P}(T,T)}{(1+\rho)^k} + \frac{\mathsf{P}(T,T)}{(1+\rho)^{T-t}}\,. \tag{6}$$

(Here we measure time in years, $t = 1, 2, \dots, T$.)

In the case when the market value $\mathsf{P}(t,T)$ is the same as the face value $\mathsf{P}(T,T)$, we see from this definition that $\rho\,(T - t, T) \equiv r_c$.

We plot schematically in Fig. 2 a typical '*yield curve*', the graph of $\rho(s,T)$ as a function of the *remaining* time $s = T - t$.

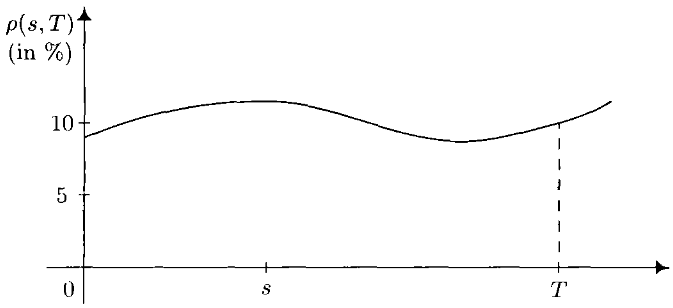


FIGURE 2. Profitablity $\rho = \rho(s,T)$ as a function of $s = T - t$

In the above description we did not touch upon the *structure* of the market prices $\{\mathsf{P}(t,T)\}$ (regarded from the probabilistic point of view, say). We consider this, fairly complicated, question, further on, in Chapter III.

Here we only note that in discussions of the structure and the dynamics of the prices $\{\mathsf{P}(t,T)\}$ from the probabilistic standpoint one usually takes one of the following two approaches:

a) *direct* specification of the evolution of the prices $\mathsf{P}(t,T)$ (the price-based approach);
b) *indirect* specification, when in place of the prices $\mathsf{P}(t,T)$, one is given the time structure of the yield $\{\rho(T - t, T)\}$ or a similar characteristic, e.g., of 'interest rate' kind (the term structure approach).

We note also that formula (6), which is a link between the price of a bond and its yield, is constructed in accordance with the 'simple interest' pattern (cf. (1) for $m = 1$). It is also easy to find a corresponding formula in the case of *continuously* payable interest ('compound interest'; cf. (2)).

We have already noted that bonds are floated by various establishments and for various purposes.

Corporate bonds are issued to accumulate capital necessary for further development or modernization, to cover operational costs, etc.

Money obtained from issuing *government bonds* issued by national goverments or *municipal bonds*, issued by state governments, city councils, etc., is used to carry out various govermental programms or projects (construction of roads, schools, bridges, ...) and cover budget deficits. In America government bonds include Treasury bonds, notes, and bills, which can be purchased through the Federal Reserve Banks or brokers.

Information on corporate bonds and their properties is available in many publications (e.g., in "New York Stock Exchange", section "New York Exchange Bonds", or in "American Stock Exchange").

This infomation is organized in the form of a list of quotations of the following form:

$$\text{IBM-JJ15-7\%} \quad \text{of} \quad '01,$$

which means that the bonds in question are issued by IBM, the interest is payable[b] on January, 15, and July, 15, with rate $r_c = 7\%$ and maturity year 2001.

Of course, one can also find in these publications data on the face value of the bonds and their current market price. (For details, see, e.g., [469].)

6. Shares (stock) are also issued by companies and corporations to raise funds. They mainly belong to one of the two types: *ordinary shares* (equities; common stock) and *preference shares* (preferred stock), which differ both in riskiness and the conditions of *dividend* payments.

An owner of common stock obtains as dividends his share of the profits of the firm, and their amount depends on its financial successes. If the company goes bancrupt, then the shareholder can lose his investment. On the other hand, preferred stock means that the investor's risk to lose everything is smaller and his dividents are guaranteed, although their amount, in general, does not increase with the firm's profits.

There also are other kinds of shares characterized by different degrees of involvement in the management of the corporation or some peculiarities of divident payments, and so on.

When buying stock (or bonds) many investors are attracted not by dividends (or interest—in the case of bonds), but by an opportunity to make money from fluctuations of stockprices, buying cheap ahead of anyone else and selling high (also before the others).

According to some estimates there are now more than 50m shareholders in the USA. (For instance, 2 418 447 investors had shares in AT&T by the end of 1992, with the total number of shares 1 339 916 615; see [357].)

---

[b] Interest is commonly payable once a year in Europe, but twice a year, every 6 months in the USA.

To buy or sell shares one should address a brokerage house, an investment company that is a member of a stock exchange. It should be noted that, although the number of individual stockholders increases with time, the share of individuals holding shares directly decreases: individuals are usually not themselves active on the market, but participate through *institutional investors* (mutual or pension funds, insurance companies, banks, and others, 'betting' on securities markets and, in particular, stockmarkets).

Many countries have stock exchanges where shares are traded. Apparently, one of the first was the Amsterdam Stock Exchange (1602), which traded the shares of joint-stock companies. Traditionally, banks exert strong influence on European (stock) exchanges, while the American stock exchanges has been separated from the bank system since the 1930s.

The two largest American exchanges (as of 1997) are the NYSE (New York Stock Exchange—the name under which it is known since 1817; it had 1366 seats in 1987) and the AMEX (American Stock Exchange, organized on the basis of the New York Curb Exchange founded in 1842). To make its shares eligible for trade on an exchange a company must satisfy certain requirements (concerning its size, profits, and so on). For example, the requirements of the NYSE are as follows: the earinings before taxes must be at least \$2.5 million and the number of shares floated must be at least 1.1 million, of market value at least \$18 million. Hence only the stock of well-known firms may be traded on this exchange (2089 listed corporations by 1993). Trade on the AMEX goes mostly in stock of medium-sized companies; the number of listed firms there is 841.

The NASDAQ (National Association of Securities Dealers Automatic Quotations) Stock Market is another major American share-trading establishment; the trade there proceeds by electronic networks. 4700 firms—large, medium-sized, and small but rapidly developing—are registered here.

An impressive number of firms (around 40 000, [357]) are participants of the OTC (Over-the-counter) securities market. This market has no premises or even central office. Deals are made by wire, through dealers who buy and sell shares on their accounts. This trade stretches over most diversified securities: ordinary and preferrence shares, corporate bonds, the US Government securities, municipal bonds, options, warrants, foreign securities.

The main reason why firms that are small, newly established, or issuing small numbers of shares use OTC dealers is that they meet there virtually no (or minimal) restrictions on the size of assets and the like.

On the other hand companies whose shares are eligible for trade on other exchanges often go to the OTC market if the manner of bargaining and deal-making accepted there seems more convenient than the routine of 'properly organized' exchanges. There can be other reasons for dealing through the OTC system, e.g., a firm may be reluctant to disclose its financial state, as required by big exchanges.

Of course, it is important for investors, 'big' or 'small' alike, to have *information* on the health of firms issuing stock, stock quotations, and the dynamics of prices.

Information about the global state of the economy and the markets, as expressed by several 'composite', 'generalized' indexes, is also of importance.

As regards indicators of the 'global' state of the economy, the most well-known among them are the Dow Jones *Averages* and *Indexes.* There are four *Averages*:

- the DJIA (*Dow Jones Industrial Average*, taken over 30 industrial companies);
- the *Dow Jones Transportation Average* (over 20 air-carriers, railroad and transportation companies);
- the *Dow Jones Utility Average* (over 15 gas, electric, and power companies);
- the *Dow Jones* 65 *Composite Average* (over all the 65 firms included in the above three averages).

We note that, say, the DJIA (the *Dow*), which is an indicator of the state of the 'industrial' part of the economy and is calculated on the basis of the data on 30 large 'blue-chip' companies, is not a mere arithmetic mean. The stock of corporations of higher market value has greater weight in the composite index, so that large changes in the prices of few companies can considerably change the index as a whole [310].

(*On the backgrounds of the Dow.* In 1883, Charles H. Dow began to draw up tables of the average closing rates of nine railroad and two manufacturing companies. These lists gained strong popularity and paved way to "The Wall Street Journal" founded (1889) by Dow and his partner Edward H. Jones. See, e.g., [310].)

Alongside the Dow Jones indexes, the following indicators are also widely used:

- Standard & Poor's 500 (S&P500) Index,
- the NYSE Composite Index,
- the NASDAQ Composite Index,
- the AMEX Market Value Index,
- Value Line Index,
- Russel 2000 Stock Index,
- Wilshire 5000 Equity Index,

. . .

Standard & Poor's 500 Index is (by conrast with the DJ) calculated on the basis of data about many companies (500 in all= 400 industrial companies + 20 transportation companies + 40 utilities + 40 financial companies); the NYSE Index comprises the stock of all firms listed at the New York Stock Exchange, and so on.

(*On the backgrounds of Standard & Poor's.* Henry Varnum Poor started publishing yearly issues of "Poor's Manual of Corporate Statistics" in 1860, more than 20 years before the Dow Jones & Company's first publication of the daily averages of closing rates. In 1941, *Poor's* Finance Services merged with *Standard* Statistic Company, another leader in the collection and publishing of financial information. The result of this union, *Standard & Poor's* Corporation became a major information service and publisher of financial statistics (see, for instance, [310]).)

Besides various pulications one can obtain information on instantaneous 'bid' and 'ask' prices of stock from the NASDAQ electronic system (covering shares of about 5 000 firms), Reuters, Bloomberg, Knight Ridder, Telerate. Brokers can at any time learn about prices through dealers and get in direct touch with the ones whose prices seem more attractive.

In view of economic globalization, it is important that one knows not only about the positions of national companies, but also about foreign ones. Such data are also available in corresponding publications. (Note that in the standard American nomenclature the attributes 'World', 'Worldwide', or 'Global' relate to all markets, including the USA, while 'International' means only foreign markets, outside the USA).

One can learn about the activities at 16 major international exchanges from "The Wall Street Journal" daily, which publishes in its "Stock Market Indexes" section the closing composite indexes of these exchanges and their realtive and absolute changes from the previous day.

As everybody knows from numerous publications, even in mainstream newspapers, stock prices and the values of various financial indexes are permanently changing in a tricky, chaotic way.

We depict the changes in the DJIA as an example:

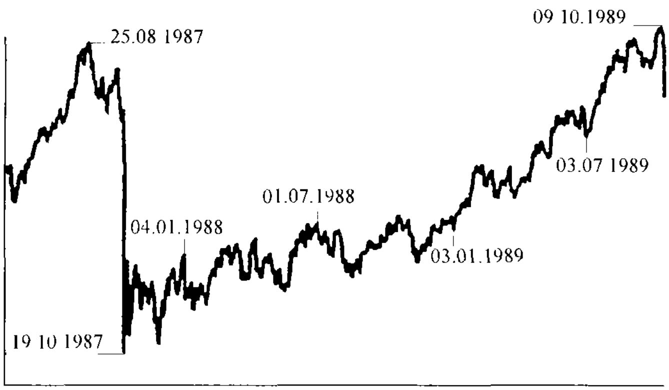


FIGURE 3. Dynamics of the DJIA (Dow Jones Industrial Average). On October 19, 1987, the day of crash, the DJIA fell by 508 points

In the next chart, Fig. 4, we plot the *daily* changes $S = (S_n)$ of the S&P500 Index during 1982–88. It will be clear from what follows that, with an eye to the analysis of the 'stochastic' components of indexes, it is more convenient to consider the quantities, $h_n = \ln \dfrac{S_n}{S_{n-1}}$, rather than the S_n themselves. We can interprete these quantities as 'returns' or 'logarithmic returns' (see Chapter II, § 1a). Their behavior is more 'uniform' than that of $S = (S_n)$. We plot the corresponding graph of the values of $h = (h_n)$ in Fig. 5.

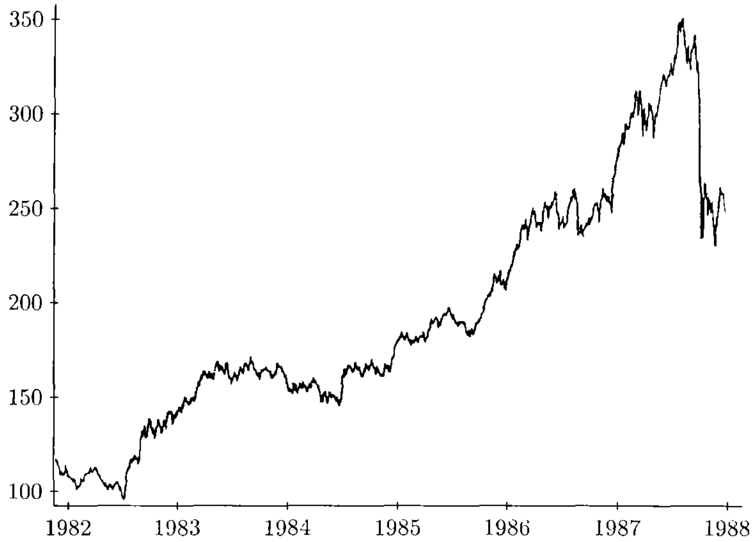


FIGURE 4. The S&P500 Index in 1982–88

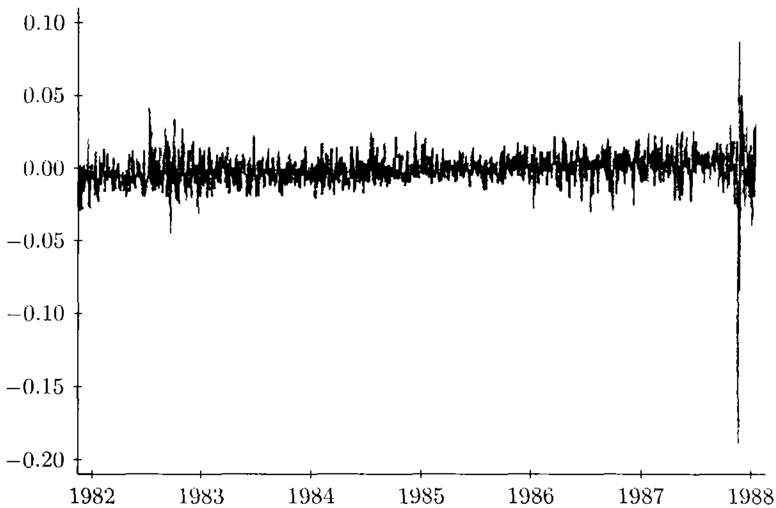


FIGURE 5. Daily values of $h_n = \ln \dfrac{S_n}{S_{n-1}}$ for the S&P500 Index; based on the data in Fig. 4

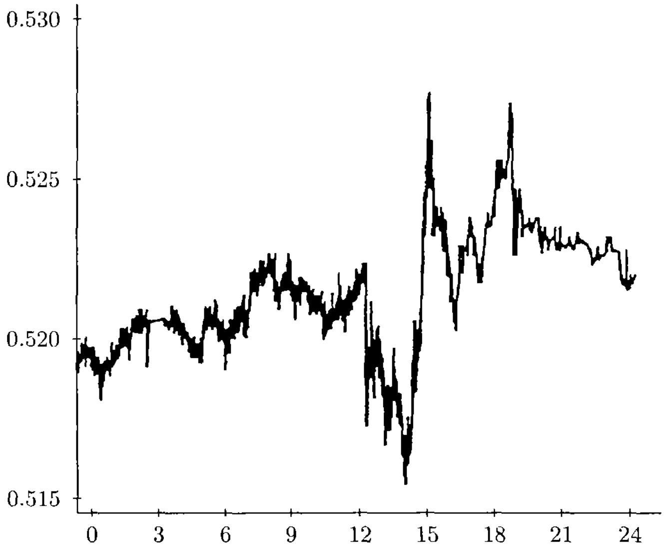


FIGURE 6. Round-the-clock dynamics of the averaged (between the 'bid' and 'ask' prices) DEM/USD cross rate, August 19, 1993 (The label 0 corresponds to 0:00 GMT)

The dip at the end of 1987, clearly seen in Figs. 3 and 4 is related to the famous *October crash*, when stock prices fell abruptly and the investors, afraid of losing everything, rushed to sell. Mass sales of shares provoked an ever growing emotional and psychological mayhem and resulted in an avalanche of selling bids. For example, about 300 m shares changed hands at the NYSE during entire January, 1987, while there were 604 m shares on offer on the day of crash, October 19, and this number increased to 608 m on October 20.

The opening price of an AT&T share on the day of crash was \$ 30 and the closing price was \$ $23\frac{5}{8}$, so that the corporation lost 21.2% of its market value.

On the whole, the DJIA was 22.6% lower on October 19, 1987, than on December 31, 1986, which means \$ 500 bn in absolute figures.

During another well-known October crash, the one of 1929, the turnover of the NYSE was 12.9 m shares on October 24 (before the crash) and 16.4 m on the day of crash. Accordingly, DJIA on October 29 was 12.8% lower than on December 31, 1928, which comprised \$ 14 bn in absolute figures.

We supplement Figs. 3–5, in which one clearly sees the vacillations of the DJIA and the S&P500 Index over a period of *several years*, by Fig. 6 depicting the behavior of the DEM/USD cross rate during *one business day* (namely, Thursday, October 19, 1993).

The first attempt towards a mathematical description of the evolution of stock prices $S = (S_t)_{t \geqslant 0}$ (on the Paris market) on the basis of *probabilistic concepts* was made by Louis Bachelier (11.03.1870–28.04.1946) in his thesis "Théorie de la spéculation" [12] published in Annales Scientifiques de l'École Normale Supérieure, vol. 17, 1900, pp. 21–86, where he proposed to regard $S = (S_t)_{t \geqslant 0}$ as a random (stochastic) process.

Analyzing the 'experimental data' on the prices $S_t^{(\Delta)}$, $t = 0, \Delta, 2\Delta, \dots$ (registered at time intervals Δ) he observed that the differences $S_t^{(\Delta)} - S_{t-\Delta}^{(\Delta)}$ had averages zero (in the statistical sense) and fluctuations $|S_t^{(\Delta)} - S_{t-\Delta}^{(\Delta)}|$ of order $\sqrt{\Delta}$.

The same properties has, e.g., a random walk $S_t^{(\Delta)}$, $t = 0, \Delta, 2\Delta, \dots$, with

$$S_t^{(\Delta)} = S_0 + \sum_{k \leqslant [\frac{t}{\Delta}]} \xi_k^{(\Delta)},$$

where the $\xi_k^{(\Delta)}$ are identically distributed random variables taking two values, $\pm\sigma\sqrt{\Delta}$, with probability $\frac{1}{2}$.

Passing to the limit as $\Delta \to 0$ (in the corresponding probabilistic sense) we arrive at the random process

$$S_t = S_0 + \sigma W_t, \qquad t \geqslant 0,$$

where $W = (W_t)_{t \geqslant 0}$ is just a *Brownian motion* introduced by Bachelier (or a *Wiener process*, as it is called after N. Wiener who developed [476] in 1923 a rigorous mathematical theory of this motion; see also Chapter III, § 3a).

Starting from a Brownian motion, Bachelier derived the formula $\mathbb{C}_T = \mathsf{E} f_T$ for the expectation (here $f_T = (S_T - K)^+$), which from the modern viewpoint gives one (under the assumption that the bank interest rate r is equal to zero and $B_0 = 1$) the value of the *reasonable* (*fair*) *price* (premium) that a buyer of a standard call option must pay to its writer who undertakes to sell stock at the maturity date T at the strike (exercise) price K (see § 1c below). (If $S_T > K$, then the option buyer gets the profit equal to $(S_T - K) - \mathbb{C}_T$ because he can buy stock at the price K and sell it promptly at the higher price S_T, while if $S_T < K$, then the buyer merely does not show the option for exercise and his losses are equal to the paid premium $\mathbb{C}_T$.)

Bachelier's formula (see Chapter VIII, § 1a)

$$\mathbb{C}_T = (S_0 - K)\,\Phi\Big(\frac{S_0 - K}{\sigma\sqrt{T}}\Big) + \sigma\sqrt{T}\,\varphi\Big(\frac{S_0 - K}{\sigma\sqrt{T}}\Big), \tag{7}$$

where

$$\varphi(x) = \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}, \qquad \Phi(x) = \int_{-\infty}^{x} \varphi(y)\, dy,$$

was in effect a forerunner of the famous *Black–Scholes formula* for the rational price of a *standard call option* in the case where $S = (S_t)$ can be described by a *geometric* (or *economic*) Brownian motion

$$S_t = S_0 e^{\sigma W_t + (\mu - \sigma^2/2)t}, \tag{8}$$

where $W = (W_t)_{t \geqslant 0}$ is a usual Brownian motion.

Under the assumption that the bank interest rate r is equal to zero and $B_0 = 1$, the Black–Scholes formula gives one the following formula for the rational price $\mathbb{C}_T = \mathsf{E}(S_T - K)^+$ of a standard call option:

$$\mathbb{C}_T = S_0 \Phi(d_+) - K\Phi(d_-), \tag{9}$$

where

$$d_\pm = \left[\ln \frac{S_0}{K} \pm \frac{\sigma^2 T}{2}\right] \bigg/ \sigma\sqrt{T}. \tag{10}$$

(As regards the general Black–Scholes formula for $r \neq 0$ and $B_0 > 0$, see Chapter VIII, § 1b).

It is worth noting that by (7), if $K = S_0$, then

$$\mathbb{C}_T = \sigma \sqrt{\frac{T}{2\pi}},$$

which gives one an idea of the increase of the rational option price with maturity time T.

The problem of an adequate description of the dynamics of various financial indicators $S = (S_t)_{t \geqslant 0}$ of stock price kind, is far from being exhausted, and it is the subject of many studies in probability theory and statistics (we concentrate on these issues in Chapters II, III, and IV). We have already explained (see subsection 4) that a similar (but maybe even more complicated) problem arises in connection with the description of the stochastic evolution of prices in bond markets $P = \{\mathsf{P}(t,T)\}_{0 \leqslant t \leqslant T}$, which are regarded as random processes with fixed conditions at the 'right-hand end-point' (i.e., for $t = T$). (We discuss these questions below, in Chapter III.)

§ 1c. Market of Derivatives. Financial Instruments

1. A sharp rise in the interest to securities markets throughout the world goes back to the early 1970s. It would be appropriate to try to understand the course of events that gave impetus to shifts in economy and, in particular, securities markets.

In the 60s, the financial markets— both the (capital) markets of long-term securities and the (money) markets of short-term securities—were featured by extremely low *volatility*, the interest rates were very stable, and the exchange rates fixed. From

1934 to 1971, the USA were adhering to the policy of buying and selling gold at a fixed price of \$ 35 per ounce (= 28.25 grams). The USA dollar was considered to be a gold equivalent, 'as good as gold'. Thus, the price of gold was imposed from outside, not determined by market forces.

This market situation restricted investors' initiative and hindered the development of new technology in finance.

On the other hand, several events in the 1960-70s induced considerable structural changes and the growth of volatility on financial markets. We indicate here the most important of these events (see also § 1b.2).

1) The transition from the policy of fixed cross rates (pursued by several groups of countries) to rates freely 'floating' (within some 'band'), spurred by the acute financial and currency crisis of 1973, which affected the American dollar, the German mark, and the Japanese yen. This transition has set, in particular, an important and interesting problem of the optimal timing and the magnitude of *interventions* by a central bank.

2) The devaluation of the dollar (against gold): in 1971, Nixon's administration gave up the policy of fixing the price of gold at \$ 35 per ounce, and gold shot up: its price was \$ 570 per ounce in 1980, fell to \$ 308 in 1984, and is fluctuating mainly in the interval \$ 300–\$ 400 since then.

3) The global oil crisis provoked by the policy of the OPEC, which came forward as a major price-maker in the oil market.

4) A decline in stock trade. (The decline in the USA at this time was larger in real terms than during the Great Depression of the 1930s.)

At that point the old 'rule of thumb' and simple regression models became absolutely inadequate to the state of economy and finances.

And indeed, the markets responded promptly to new opportunities opened before investors, who gained much more space for speculations. Option and bond futures exchanges sprang up in many places. The first specialist exchange for trade in standard option contracts, the Chicago Board Option Exchange (CBOE), was opened in 1973. This is how the investors responded to the new, more promising opportunities: 911 contracts on call options (on stock of 16 firms) were sold the opening day, April 26. A year later, the daily turnover reached more than 20 000 contracts, it grew to 100 000 three years later, and to 700 000 in 1987. Bearing in mind that one contract means a package of 100 shares, we see that the turnover in 1987 was 70 m shares each day; see [35].

The same year, 1987, the daily turnover at the NYSE (New York Stock Exchange) was 190 m shares.

1973 was special not only because the first proper exchange for standard option contracts was opened. Two papers published that year, *The pricing of options and corporate liabilities* by F. Black and M. Scholes ([44]) and *The theory of rational option pricing* by R. C. Merton ([357]), brought a genuine revolution in pricing methods. It would be difficult to name other theoretical works in the finances literature that could match these two in the speed with which they found applications in the practice and became a source of inspiration for multiple studies of more complex options and other types of derivatives.

2. The most prominent among the derivatives considered in financial engineering are *option* and *futures contracts*. It is common knowledge that both are highly risky; nevertheless they (and their various combinations) are used successfully not merely to draw (speculative) profits but also as a *protection* against drastic changes in prices.

An **Option** is a security (contract) issued by a firm, a bank, another financial company and giving its buyer the *right* to buy or sell something of value (a share, a bond, currency, etc.) on specified terms at a fixed instant or during a certain period of time in the future.

By contrast with option contracts giving a *right* to buy or sell,

a **Futures contract** is a *commitment* to buy or sell something of value (e.g., gold, cereals, foreign currency) at a preset instant of time in the future at a (futures) price fixed at the moment of signing the deal. (One can find an indispensable source of statistical data on all securities, as well as options and futures, in an American financial weekly "Barron's".)

α) *Futures* are of practical interest both to sellers and buyers of various goods.

For example, a clever farmer worried about selling the future crop at a 'good' price and afraid of a drastic downturn in prices would prefer to make a 'favorable' agreement with a miller (baker) on the delivery of (not yet grown) grain instead of waiting the grain to ripe and selling it at the (who knows what!) market price of the day.

Accordingly, the miller (baker) is also interested in the purchase of grain at a 'suitable' price and is seeking to forestall large rises in grain prices possible in the future.

In the end, both share the same objective of *minimizing the risks* due to the uncertainty of the future market prices.Thus, a futures contract is a form of an agreement that can in a way be convenient for both sides.

Before a substantial discussion of futures contracts it seems appropriate to consider a related form of agreement: so-called forward contracts.

β) As in the case of futures contracts,

a **Forward contract** is also an agreement to deliver (or buy) something in the future at a specified (forward) price.

The difference between futures and forward contracts is as follows: while forwards are usually sold without intermediaries, futures are traded at a *specialized exchange*, where the seller and the buyer do not necessarily know each other and where a special-purpose settlement system makes a subsequent renouncement of the contract uneconomical.

Usually, the person willing to buy is said to 'hold a *long* position', while the one undertaking the delivery in question is holding a '*short* position'.

A cardinal question of forward and futures contracts is that of the preset 'forward' ('futures') price, which, in effect, can turn out distinct from the market price at the time of delivery.

Roughly, transactions involving forwards run as follows:

$$\begin{tikzcd} \boxed{\text{seller}} \arrow[r, shift right, "\text{goods}"'] & \boxed{\text{buyer}} \arrow[l, shift right, "\text{money}"'] \end{tikzcd}$$

Here we understand 'goods' in a broad sense. For example, this can be currency. If you are interested in trading, say, US dollars for Swiss franks, then the quotations that you find may look as follows:

$$\begin{aligned} \text{Current price } 1 \text{ USD} &= 1.20 \text{ CHF}, \\ (\text{i.e., we can buy } 1.20 &\text{ CHF for } \$\,1); \\ \text{30-day forward } 1 \text{ USD} &= 1.19 \text{ CHF}; \\ \text{90-day forward } 1 \text{ USD} &= 1.18 \text{ CHF}; \\ \text{180-day forward } 1 \text{ USD} &= 1.17 \text{ CHF}. \end{aligned}$$

This picture is typical in that you tend to get less for $ 1 with the increase of delivery time, and therefore if you need CHF 10 000 in 6 months' time, then you must pay

$$\frac{10\,000}{1.17} = \$\,8\,547.$$

However, if you need CHF 10 000 now, then it costs you

$$\frac{10\,000}{1.20} = \$\,8\,333.$$

Clearly, the *actual*, *market* price of CHF 10 000 can in 6 months be different from $ 8 547. It can be less or more, depending on the CHF/USD cross rate 6 months later.

(*On the background of futures.* According to N. Apostolu's "Keys to investing in options and futures" (Barron's Educational Series, 1991), "the first organized commodity exchange in the United States was the Chicago Board of Trade (CBOT), founded in Chicago in 1848. This exchange was originally intended as a central market for the conduct of cash grain business, and it was not until 1865 that the first futures transaction was performed there. Today, the Chicago Board of Trade, with nearly half of all contracts traded in the United States, is the largest futures exchange in the world.

Financial futures were not introduced until the 1970s. In 1976, the International Monetary Market (IMM), a subsidiary of the Chicago Mercantile Exchange (CME), began the 90-day Treasury bill futures contract. The following year, the

CBOT initiated the Treasury bond futures contract. In 1981, the IMM created the Eurodollar futures contract.

A major financial futures milestone was reached in 1982, when the Kansas City Board of Trade introduced a stock index futures contract based upon the Value Line Stock Index. This offering was followed in short order by the introduction of the S&P500 Index futures contract on the Chicago Mercantile Exchange and the New York Stock Exchange Composite Index traded on the New York Futures Exchange. All of these contracts provide for cash delivery rather than delivery of securities.

Trading volume on the futures exchanges has surged in the last three decades—from 3.9 million contracts in 1960 to 267.4 million contracts in 1989. In their short history, financial futures have become the dominant factor in the futures markets. In 1989, 60 percent of all futures contracts traded were financial futures. By far the most actively traded futures contract is the Treasury bond contract traded on the CBOT.")

γ) A forward contract, as already mentioned, is an agreement between two sides and, in principle, there exists a risk of its potential violation. More than that, it is often difficult to find a supplier of the goods you need or, conversely, an interested buyer.

Apparently, this is what has brought into being *futures contracts* traded at exchanges equipped with special settlement mechanisms, which in general can be described as follows.

Imagine that on March 1, you instruct your broker to buy some amount of wheat by, say, October 1 (and specify a prospective price). The broker passes your request to a produce exchange, which forwards it to a trader. The trader looks for a suitable price and, if successful, informs the potential sellers of his wish to buy a contract at this price. If another trader agrees, then the bargain is made. Otherwise the trader informs the broker and the latter informs you that there are goods at higher prices, and you must take one or another decision.

Assume that, finally, the price of the contract is agreed upon and you, the buyer, keep the *long* position, while the seller keeps the *short* position. Now, to make the contract effective each side must put certain amount, called the *initial margin*, into a special exchange account (this is usually 2%–10% of Φ_0 depending on the customers' history). Next, the settlement mechanism comes into force. Settlements are usually carried out at the end of each day and run as follows. (Here Φ_0 is the (futures) price, i.e., both sides agree that the wheat will be delivered on October 1, at the price Φ_0.)

Assume that at time $t = 0$ the price $\Phi_0 = \$\,10\,000$ is the (futures) price of wheat delivery at the delivery time $T = 3$, i.e., in three days. Assume also that at the end of the next day ($t = 1$) the market futures price of the delivery of wheat at the time $T = 3$ becomes $\$\,9\,900$. In that case the clearing house at the exchange transfers $\$\,100$ ($= 10\,000 - 9\,900$) into the supplier's (farmer's) account. Thus, the farmer earns some profit and is now in effect left with a futures contract worth $\$\,9\,900$ in place of $\$\,10\,000$.

We note that if the delivery were carried out at the end of the first day, at the new futures price \$ 9 900, then the total revenues of the farmer would be precisely the futures price Φ_0 because \$ 100 + \$ 9 900 = \$ 10 000 = Φ_0.

Of course, the clearing house writes these \$ 100 off the buyer's account, and he must additionally pay \$ 100 into the margin account. (Sometimes, additional payments are necessary only if the margin account falls below a certain level—the maintenance margin).

Assume that the same occurs at time $t = 2$ (see Fig. 7). Then the farmer takes \$ 200 onto his account and the same sum is written off the buyer's account.

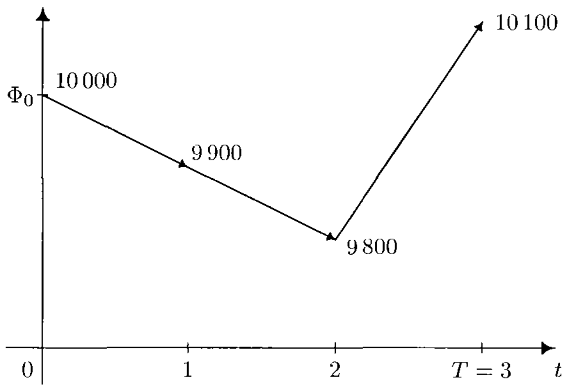


FIGURE 7. Evolution of a futures price

However, if the futures price of (instant) delivery (= the market price) becomes \$ 10 100 at time $t = 3$, then one must write \$ 300(= 10 100 − 9 800) off the farmer's account, and therefore, all in all, he loses \$ 100 (= 300 − 200), while the buyer gains \$ 100.

Note that if the buyer decides to buy wheat at time $T = 3$ on condition of instant delivery, then he must pay \$ 10 100 (= the instantaneous market price). However, bearing in mind that \$ 100 have already been transferred to his account, the delivery actually costs him 10 100 − 100 = \$ 10 000, i.e., it is precisely the same as the contract price Φ_0.

The same holds for the farmer: the money actually obtained for the delivery of wheat is precisely \$ 10 000, because he is paid the market price \$ 10 100 at time $T = 3$, which, combined with \$ 100 written off his account, makes up precisely 10 000 \$.

This example clearly depicts the role of the clearing house as a 'watchdog' keeping track of all the transactions and the state of the margin account, which is all very essential for smooth execution of contracts.

We have already noted that one of the cardinal problems of the theory of forward and futures contracts is to determine the 'fair' values of their prices.

We show below, in Chapter VI, § 1e, how the arguments based on the 'absence of arbitrage', combined with the martingale methods, enable one to derive formulas for the forward and futures prices of contracts with delivery time T sold at time $t < T$.

δ) *Options.* The theory and practice of option contracts has its own concepts and special vocabulary, and it is reasonable to become acquainted with them already at this early stage. This is all the more important as a large portion of the mathematical analysis of derivatives in what follows is related to options. This has several reasons.

First, the mathematical theory of options is the most developed one, and this example is convenient for a study of the main principles of derivatives transactions and, in particular, pricing and ***hedging*** (i.e., protecting) strategies.

Second, the actual number of options traded in the market runs into millions, therefore, there exists an impressive statistics, which is essential for the control of the quality of various probabilistic models of the evolution of option prices.

Although options are long known as financial instruments (see, e.g., the book [346]), L. Bachelier [12] must be the first who, in 1900, gave a rigorous mathematical analysis of option prices and provided arguments in favor of investment in options. Moreover, as already noted, trade in options became institutionalized not so long ago, in 1973 (see § 1c).

(*On the background of options.* According to N. Apostolu's "Keys to investing in options and futures" (Barron's Educational Series, 1991), "since the creation of the Chicago Board Options Exchange (CBOE) in 1973, trading volume in stock options has grown remarkably. The listed option has become a practical investment vehicle for institutions and individuals seeking financial profit or protection. The CBOE is the world's largest options marketplace and the nation's second-largest securities exchange. Options are also traded on the American Stock Exchange (AMEX), the New York Stock Exchange (NYSE), the Pacific Stock Exchange (PSE), and the Philadelphia Stock Exchange (PHLX). Options are not limited to common stock. They are written on bonds, currencies, and various indexes. The CBOE trades options on listed and over-the-counter stocks, on Standard & Poor's 100 and 500 market indexes, on U.S. Treasury bonds and notes, on long-term and short-term interest rates, and on seven different foreign currencies.")

ε) For definiteness, we assume in our discussion of options that transactions can occur at time

$$n = 0, 1, \dots, N$$

and all the trade stops after the instant N.

Assume also that we discuss options based on stock of value described by the random sequence

$$S = (S_n)_{0 \leqslant n \leqslant N}.$$

Using the standard vocabulary, we distinguish options of two kinds:

- *buyer's* options (*call options*) and
- *seller's* options (*put options*).

A call option gives one the *right to buy.*

A put option gives one the *right to sell.*

From the standpoint of financial engineering it is important that these two financial instruments 'work in opposite directions': as the gains from one of them increase, the gains from the other drop. This explains the widely used practice of *diversification*, of operating options in different classes, sometimes combined with other securities.

As regards their exercise, options fall into two types, *European* and *American.*

If an option can be shown for exercise only at some *fixed* time N, then we call N its *maturity time* (option's expiration date) and our option belongs to the *European type.*

On the other hand, if the option can be shown for exercise at an *arbitrary* (e.g., random) instant $\tau \leqslant N$, then we call it an option of *American type.*

In practice, most options are of American type. This gives the buyer more freedom, allowing him to *choose* the exercise time τ. Note that these two types of options are equivalent in *certain* cases (in the sense that the optimal exercise time τ of an American-type option is equal to N; see Chapter VI, § 5b and Chapter VIII, § 3c below for detail).

We now consider for definiteness a *standard call option* of *European type* with maturity time N. It can be characterized by the (*strike* or *exercise*) *price* K (*preset* at the instant of writing) at which the buyer will be able to buy, say, shares, whose market price S_N at time N may be distinct (sometimes, considerably distinct) from K.

If $S_N > K$, then the situation is favorable for the buyer because, by the contract terms, he has the right to buy shares at the price K. Once bought, they can be immediately sold at the market price S_N. In this case his gains from this operation will be equal to $S_N - K$.

On the other hand, if S_N turns out to be less than K, then the buyer's right of purchase (at the price K) is of no value because he will be able to buy shares at a lower price (S_N).

All combined, we can say that the buyer's gains f_N at time N can be expressed by the formula

$$f_N = (S_N - K)^+,$$

where $a^+ = \max(a, 0)$.

Of course, one must pay certain premium $\mathbb{C}_N$ for the acquisition of this financial instrument, so that the *net profit of the buyer of the call option* is

$$(S_N - K)^+ - \mathbb{C}_N,$$

i.e.,

$$\begin{aligned} (S_N - K) - \mathbb{C}_N &\quad \text{for} \quad S_N > K, \\ -\mathbb{C}_N &\quad \text{for} \quad S_N \leqslant K. \end{aligned}$$

Accordingly, the *writer's gain* is

$$\mathbb{C}_N - (S_N - K) \quad \text{for} \quad S_N > K$$

and

$$\mathbb{C}_N \quad \text{for} \quad S_N \leqslant K.$$

Hence it is clear that purchasing a call option one anticipates a *rise* of stock prices. (Note that, of course, the option premium $\mathbb{C}_N$ depends not only on N, but also on K, and, clearly, the less is K the larger must be $\mathbb{C}_N$.)

Remark. There exist special names for those acting on the assumption of a *rise* or *fall* of some article. The dealers expecting prices to go up are called *bulls*. A bull opens a long position expecting to sell with profit afterwards, when the market goes up. Those dealers who expect the market to move downwards are called *bears*. A bear tends to sell securities he has (or even has not—which is referred to as 'short selling'). He hopes to close his short position by buying the traded items afterwards, at lower prices. The difference between the current price and the purchase price in the future will be his premium.

Depending on the relation between the market price S_0 at $t = 0$ and K, options can be divided into three classes: options *bringing a gain* (in-the-money), ones *with gain zero* (at-the-money), and options *bringing losses* (out-of-money). In the case of a call option these classes correspond to the relations $S_0 > K$, $S_0 = K$, and $S_0 < K$, respectively.

We must point out straight away that there is an enormous difference between the positions of a buyer and a seller.

The buyer purchasing the option can simply wait till the maturity date N, watching—if he likes—the dynamics of the prices S_n, $n \geqslant 0$.

The position of the option writer is much more complicated because he must bear in mind his obligation to meet the terms of the contract, which requires him to not merely contemplate the changes in the prices S_n, $n \geqslant 0$, but to use all financial means available to him to build a portfolio of securities that ensures the final payment of $(S_N - K)^+$.

The following two questions are central here: what is the 'fair' price $\mathbb{C}_N$ of buying or selling an option and what must a seller do to carry out the contract.

In the case of a *standard put option* of *European type* with maturity date N the price K at which the option buyer *is entitled to sell* stock (at time N) is fixed.

Hence if the real price of the stock at time N is S_N and $S_N < K$, then selling it at the price K brings $K - S_N$ in revenues. His *net profit*, taking into account the premium $\mathbb{P}_N$ for the purchase of the option, say, is equal to

$$(K - S_N) - \mathbb{P}_N.$$

On the other hand, if $S_N > K$, i.e., the preset price is less than the market price, then there is no sense in showing the option for exercise.

Hence the *net profit of the buyer of a put option* is $(K - S_N)^+ - \mathbb{P}_N$.

As in the case of a call option, here we can also ask about a 'fair' price $\mathbb{P}_N$ suitable for both writer and buyer.

η) As an illustration we consider an example of a call option.

Assume that we buy 10 contracts on stock. As a rule, each contract involves 100 shares, so we discuss a purchase of 1 000 shares. Assume that the market price S_0 of a share is 30 (American dollars, say), $K = 35$, $N = 2$, and the premium for the total of 10 contracts is (\$) 250.

Further, let the market price S_2 at time $n = 2$ be \$ 40. In this case the option is shown for exercise and the corresponding net profit is *positive*:

$$(40 - 35) \times 1\,000 - 250 = (\$)\,4\,750.$$

However, if the market price S_2 is \$ 35.1, then we again show the option for exercise (since $S_2 > K = \$\,35$), but the corresponding net profit is now *negative*:

$$(35.1 - 35) \times 1\,000 - 250 = (\$) \ -150.$$

It is clear that our profit is *zero* if

$$(S_2 - K) \cdot 1\,000 = (\$)\,250,$$

i.e., $S_2 = \$\,35.25$ (because $K = \$\,35$).

Thus, each time the share price S_2 drops below \$ 35.25, the buyer of the call option takes losses.

Assume now that we consider an American-type call option, which gives us the right to exercise it at time $T = 1$ or $T = 2$. Imagine that the share price rises sharply at the instant $T = 1$, so that $S_1 = \$\,50$. Then the buyer of the call option can show it for exercise at this instant and pocket a huge net profit of

$$(50 - 35) \cdot 1\,000 - 250 = 15\,000 - 250 = (\$)\,14\,750.$$

It is however clear that the premium for such an option must be considerably larger than \$ 250 because 'greater opportunities should come more expensive'. The *actual* prices of American-type options are indeed higher than those of European ones. (See Chapter VI below.)

We have considered the above example from the buyer's standpoint. We now turn to the writer's position.

In principle, he has two options: to sell stock that he has already ('writing covered stock') or to sell stock he has not at the moment ('writing naked call').

The latter is very risky and can be literally ruinous: if the option is indeed shown for exercise (provided $S_2 > K$), then, to meet the terms of the contract, the seller must actually buy stock and sell it to the buyer at the price K.

If, e.g., $S_2 = \$\,40$, then he must pay $\$\,40\,000$ for $1\,000$ shares.

His premium was only $\$\,250$, therefore, his total losses are

$$40\,000 - 35\,000 - 250 = (\$)\ 4\,750.$$

It must be noted that the writer's net profit is in both cases at most $\$\,250$. This is a purely speculative profit made 'from swings of stock prices' and (in the case of 'naked stocks writing') in a rather risky way. We can say that here the '*writer's profit* is his *risk premium*'.

3. In practice, 'large' investors with big financial potentials reduce their risks by an extensive use of *diversification*, *hedging*, investing funds in most *various* securities (stock, bonds, options, ...), commodities, and so on. Very interesting and instructive in this respect is G. Soros's book [451], where he repeatedly describes (see, for instance, the tables in § § 11, 13 of [451]) the day-to-day *dynamics* (in 1968–1993) of the securities portfolio of his Quantum Fund (which contains various kinds of financial assets). For example, on August 4, 1968, this portfolio included stock, bonds, and various commodities ([451, p. 243]).

From the standpoint of financial engineering Soros wrote a masterpiece of a handbook for those willing to be active players in the securities market.

4. In our considerations of European **call options**, we have already seen that they can be characterized by

1) their maturity date N,
2) the pay-off function f_N.

• For a *standard call option* we have

$$f_N = (S_N - K)^+.$$

• For a *standard call option with aftereffect* we have

$$f_N = (S_N - K_N)^+,$$

where $K_N = a \cdot \min(S_0, S_1, \dots, S_N)$; $a = \text{const}$.

• For an *arithmetic Asian call option* we have

$$f_N = (\overline{S}_N - K)^+, \qquad \overline{S}_N = \frac{1}{N+1}\sum_{k=0}^{N} S_k.$$

We must note that the quantity K entering, for instance, the pay-off function $f_N = (S_N - K)^+$ of a standard call option, its *strike price*, is usually close to S_0. As a rule, one never writes options with large disparity between S_0 and K.

Pay-off functions for **put options** are as follows:

- for a *standard put option* we have

$$f_N = (K - S_N)^+;$$

- for a *standard put option with aftereffect* we have

$$f_N = (K_N - S_N)^+, \qquad K_N = a \cdot \max(S_0, S_1, \dots, S_N); a = \text{const}$$

- for an *arithmetic Asian put option* we have

$$f_N = (K - \overline{S}_N)^+, \qquad \overline{S}_N = \frac{1}{N+1} \sum_{k=0}^{N} S_k.$$

There exist many types of options, some of which have rather *exotic* names (see, for instance, [414] and below, Chapter VIII, § 4a). We also discuss several types of option-based strategies (combinations, spreads, etc.) in Chapter VI, § 4e.

One attraction of options for a buyer is that they are not very expensive, although the commission can be considerable. To give an idea of the calculation of the *price of an option* (the *premium* for the option, the non-reimbursable payment for its purchase), we consider the following, slightly idealized, situation

Assume that the stock price S_n, $0 \leqslant n \leqslant N$, satisfies the relation

$$S_n = S_0 + (\xi_1 + \dots + \xi_n),$$

where $S_0 > N$ is an integer and (ξ_k) is a sequence of independent identically distributed random variables with distribution

$$\mathsf{P}(\xi_k = \pm 1) = \frac{1}{2}\,.$$

Then, of course, $S_n > 0$, $0 \leqslant n \leqslant N$.

Assume also that we have at our disposal a bank account containing an amount $B = (B_n)_{0 \leqslant n \leqslant N}$, where $B_n \equiv 1$ (i.e., the interest rate $r = 0$ and $B_0 = 1$). We consider now a standard European call option with pay-off $f_N = (S_N - K)^+$.

We claim that the *rational* (or *fair, mutually appropriate*) price $\mathbb{C}_N$ of such an option can be expressed as follows:

$$\mathbb{C}_N = \mathsf{E}(S_N - K)^+,$$

i.e., the size of the premium is equal to the average gain of the buyer.

We enlarge on the formal definition of $\mathbb{C}_N$ and the methods of its calculation below (see Chapter VI), while here we can substantiate the formula $\mathbb{C}_N = \mathsf{E}(S_N - K)^+$ as follows.

Assume that the writer's ask price $\widetilde{\mathbb{C}}_N$ is larger than $\mathsf{E}(S_N - K)^+$ and the buyer agrees to purchase the option at this price. We claim that the writer has a *riskless profit* of $\widetilde{\mathbb{C}}_N - \mathbb{C}_N$ in this case.

In fact, the buyer acknowledges that the price must give the writer a possibility to meet the terms of the contract. It is clear that this price cannot be too low. But, understandably, the buyer also would not overpay: he would rather buy at the *lowest* price enabling the writer to meet the contract terms.

In other words, we must show that the option writer can use the premium $\mathbb{C}_N = \mathsf{E}(S_N - K)^+$ to meet the contract terms. For simplicity, we set $N = 1$ and $K = S_0$. Then $\mathbb{C}_1 = \mathsf{E}\xi_1^+ = \frac{1}{2}$. We now describe the ways in which the writer can operate with this premium in the securities market.

We represent $\frac{1}{2}$ as follows:

$$\frac{1}{2} = \left(\frac{1}{2} - \frac{S_0}{2}\right) + \frac{S_0}{2} \qquad (= \beta_0 \cdot 1 + \gamma_0 \cdot S_0).$$

Setting $X_0 = \beta_0 \cdot 1 + \gamma_0 \cdot S_0$ $(= \frac{1}{2})$ we can call this premium $\frac{1}{2}$ the (initial) capital of the writer; a part of it (β_0) is put into a bank account and another part is invested in (γ_0) shares. The fact that β_0 is negative in our case means that the seller merely *overdraws* his account, which, of course, must be repaid.

The pair (β_0, γ_0) forms the so-called writer's *investment portfolio* at time $n = 0$.

What is this portfolio worth at the time $n = 1$? Denoting this amount by X_1 we obtain

$$\begin{aligned} X_1 &= \beta_0 \cdot B_1 + \gamma_0 \cdot S_1 = \beta_0 + \gamma_0(S_0 + \xi_1) \\ &= \left(\frac{1}{2} - \frac{S_0}{2}\right) + \frac{1}{2}(S_0 + \xi_1) = \frac{1 + \xi_1}{2} \\ &= \begin{cases} 1 & \text{for } \xi_1 = 1, \\ 0 & \text{for } \xi_1 = -1. \end{cases} \end{aligned}$$

Since

$$\xi_1^+ = \frac{1 + \xi_1}{2},$$

it is obvious that

$$X_1 = f_1 \qquad (= (S_1 - K)^+).$$

In other terms, the portfolio (β_0, γ_0) ensures that the amount X_1 is *precisely* equal to the pay-off f_1, which enables the writer to meet the terms of the contract and repay his debt.

For if $\xi_1 = 1$, then he gets $\frac{1}{2}(S_0 + 1)$ for the stock. Since

$$\frac{1}{2}(S_0 + 1) = 1 + \left(\frac{S_0}{2} - \frac{1}{2}\right),$$

this money is sufficient to repay the debt $\left(\frac{S_0}{2} - \frac{1}{2}\right)$ and pay the buyer an amount of $(S_1 - K)^+ = (S_1 - S_0)^+ = \xi_1^+ = 1$.

On the other hand, if $\xi_1 = -1$, then he obtains $\frac{1}{2}(S_0 - 1)$ for the stock. The buyer does not exercise the option in this case (because $S_1 = S_0 + \xi_1 = S_0 - 1 < S_0 = K$), and therefore the writer must merely repay his debt $\frac{1}{2}(S_0 - 1)$, which is precisely equal to the money obtained for the stock (we now have $\xi_1 = -1$).

Hence, if the writer sets $\widetilde{\mathbb{C}}_1 > \mathbb{C}_1 = \mathsf{E}(S_1 - K)^+$, then upon meeting all the terms of the contract he gets a riskless profit of $\widetilde{\mathbb{C}}_1 - \mathbb{C}_1$.

We now claim that if the premium $\widetilde{\mathbb{C}}_1$ is smaller than $\mathbb{C}_1$ $(= \frac{1}{2})$, then the writer cannot meet the terms (without losses).

Indeed, choosing a portfolio (β_0, γ_0) we obtain

$$X_0 = \beta_0 + \gamma_0 S_0$$

and

$$X_1 = \beta_0 + \gamma_0(S_0 + \xi_1) = X_0 + \gamma_0 \xi_1.$$

If $\xi_1 = 1$,, then, under the terms of the contract, the writer must pay an amount of 1 to the buyer and, additionally, pays $-\beta_0$, i.e., he must obtain

$$\gamma_0(S_0 + 1) = 1 - \beta_0$$

for the stock, while if $\xi_1 = -1$, then he must get

$$\gamma_0(S_0 - 1) = -\beta_0$$

for the stock.

All in all, we must set

$$\gamma_0 = \frac{1}{2}, \qquad \beta_0 = \frac{1}{2} - \frac{S_0}{2}.$$

And yet, for these values of the parameters we have $\beta_0 + \gamma_0 S_0 = \frac{1}{2}$, therefore the equality $X_0 = \beta_0 + \gamma_0 S_0$ is impossible for $X_0 < \frac{1}{2}$.

Thus, we have established the formula $\mathbb{C}_N = \mathsf{E}(S_N - K)^+$ for $N = 1$ and $S_0 = K$. One can prove it in the general case by similar arguments. We are not doing that here because we shall deduce the same formula below (in Chapter VI) from general considerations.

Note that if $S_0 = K$, then

$$\mathbb{C}_N = \mathsf{E}(\xi_1 + \cdots + \xi_N)^+,$$

and

$$\mathbb{C}_N \sim \sqrt{\frac{N}{2\pi}}$$

by the Central limit theorem.

Hence the option premium increases as $\sqrt{N}$ with time N. This is perfectly consistent with the result following (in the case of $S_0 = K$) from Bachelier's formula for $\mathbb{C}_T$ at the end of § 1b:

$$\mathbb{C}_T = \sqrt{\frac{T}{2\pi}}$$

(for $r = 0$, $B_0 = 1$, $\sigma = 1$, and $S_0 = K$).

5. We have already mentioned that there is a great variety of option kinds on the market.

Options on indexes are quite common; for instance, options on the S&P100 and the S&P500 Indexes traded on the CBOE. (Options on S&P100 are of American type, while ones on S&P500 are of European type.) In view of the large volatility, the maturity time of these options is fairly short. See also Chapter VIII, § 4a.

In the case of options on futures the role of the prices (S_n) is played by the futures prices (Φ_n), whereas in the case of options on some index it is the value of this index that plays the same role of the prices (S_t).

2. Financial Markets under Uncertainty. Classical Theories of the Dynamics of Financial Indexes, their Critics and Revision. Neoclassical Theories

Here are just several questions that arise naturally once one has got acquainted with the theory and practice of financial markets:

- how do financial markets *operate* under uncertainty?
- how are the prices *set* and how can their dynamics in time be *described*?
- what *concepts* and *theories* must one use in calculations?
- can the future development of the prices be *predicted*?
- what are the *risks* of the use of some or other financial instruments?

In our descriptions of the price dynamics and in pricing derivative financial instruments we shall take the standpoint of a market without *opportunities for arbitrage*. *Mathematically*, this transparent *economic* assumption means that there exists a so-called 'martingale' (risk-neutral) probability measure such that the (discounted) prices are *martingales* with respect to it. This, in turn, gives one a possibility to use the well-developed machinery of stochastic calculus for the study of the evolution of prices and for various calculations.

Although we do not intend to give here a detailed account of the existing theories and concepts relating to *financial markets*, we shall discuss those that are closer to our approach, where the main (probabilistic) stress is put on the applications of stochastic calculus and statistics in financial mathematics and engineering. (We point out several handbooks and monographs: [79], [83], [112], [117], [151], [240], [268], [284], [332]–[334], [387], [460], where one can find the discussion of most diverse aspects of financial markets, economical theories and concepts, theories designed for the conditions of certainty or uncertainty, equilibrium models, optimality, utility, portfolios, risks, financial decisions, dividends, derivative securities, etc.).

We note briefly that back in the 1920s, considered to be the birth time of *financial theory*, its central interests were mainly concentrated around the problems of managing and raising funds, while its 'advanced mathematics' was virtually reduced to the calculation of compound interest.

The further development proceeded along two lines: under the assumption of a *complete certainty* (as regards the prices, the demand, etc.) and under the assumption of *uncertainty*.

Decisive for the first approach were the results of I. Fisher [159] and F. Modigliani and M. Miller [356], [350] who considered the issue of *optimal* solutions that must be taken by individual investors and corporations, respectively. Mathematically, this reduces to the problem of maximization under constraints for functions of several variables.

In the second field, we must first of all single out H. Markowitz [268] (1952) and M. Kendall [269] (1953).

Markowitz's paper, which established a basis of *investment portfolio theory*. concentrated on the *optimization of investment decisions* under *uncertinity*. The corresponding probabilistic method, the so-called *mean-variance analysis*, revealed a central role of the *covariance* of prices, which is a key ingredient determining the (unsystematic) *risks* of the investment portfolio in question. It was after that paper that one became fully aware of the importance of *diversification* (in making up a portfolio) to the reduction of unsystematic risks. This idea influenced later two *classical* theories developed by W. Sharpe [433] (1964) and S. Ross, [412] (1976),

- *CAPM—Capital Asset Pricing Model* [433] and
- *APT—Arbitrage Pricing Theory* [412],

These theories describe the components making up the yield from securities (for example, stock), explain their dependence on the state of the 'global market' of this kind of securities (*CAPM*-theory) and the factors influencing this yield (*APT*-theory), and describe the concepts that must underlie financial calculations. We discuss the central principles of Markowitz's theory, *CAPM*, and *APT* below, in § § 2b–2d.

It is already clear from this brief exposition that all the three theories are related to the *reduction of risks* in financial markets.

In the discussion of *risks*,[c] it should be noted that financial theory usually distinguishes two types of them:

- *unsystematic risks*, which can be reduced by a diversification (they are also called diversifiable, residual, specific, idiosyncratic risks, etc.), i.e., risks that investor can influence by his actions, and
- *systematic*, or market *risks* in the proper sense of the term (undiversifiable risks).

An example of a systematic risk is the one related to the stochastic nature of interest rates or stock indexes, which a ('small') investor cannot change on his own. This does not imply, of course, that one cannot withstand such risks. It is in effect to

[c] *All sorts* of risks, including insurance risks (see § § 3a, 3b below).

control possible systematic risks, to work out recommendations for rational financial solutions, to shelter from big and catastrophic risks (e.g., in insurance) that one develops (fairly complicated) systems for the collection and processing of statistical data, and for the prediction of the development of market prices. This is in fact the *raison d'être* of derivative financial instruments, such as futures contracts, options, combinations, spreads, etc. This is the purpose of *hedging*, more complicated techniques than diversification, which were developed for these instruments, take into account random changes of the future prices, and aim at the reduction of the risks of possible unfavorable effects of these changes. (One can find a detailed discussion of hedging and the corresponding pricing theory in Chapter VI.)

M. Kendall's lecture [269] (1953) at a session of the Royal Statistical Society (London) concentrates around another question, which is more basic in a certain sense: *how do market prices behave*, what stochastic processes can be used to describe their dynamics? The questions posed by that important work had ultimately brought about *Efficient Capital Market Theory* (the ECM-theory), which we discuss in the next section, and its many refinements and generalizations.

§ 2a. Random Walk Conjecture and Concept of Efficient Market

1. In several studies made in the 1930s their authors carried out empirical analysis of various financial characteristics in an attempt to answer the notorious question: *is the movement of prices, values*, and so on *predictable*? These papers were mostly written by statisticians; we single out among them A. Cowles [84] (1933) and [85] (1944), H. Working [480] (1934), and A. Cowles and H. Jones [86] (1937). A. Cowles considered data from stockmarkets, while H. Working discussed commodity prices.

Although one could find in these papers plenty of statistical data and also interesting—and unexpected—conclusions that the increments $h_n = \ln \dfrac{S_n}{S_{n-1}}$ of the logarithms of the prices S_n, $n \geqslant 1$, must be independent, neither economists nor practitioners paid much attention to this research.

As noted in [35; p. 93], one probable explanation can be that, first, economists regarded the price dynamics as a 'sideshow' of the economy and, second, there were not so many economists at that time who had a suitable background in mathematics and mastered statistical techniques.

As regards the practitioners, the conclusion made by these researchers that the sequence $(H_n)_{n\geqslant 1}$, where $H_n = h_1 + \cdots + h_n$, had the nature of a *random walk* (i.e., was the sum of independent random variables) was in contradiction with the prevailing idea of the time that the prices had their *rythms, cycles, trends*, ... , and *one could predict price movements by revealing them.*

There had been in fact no important research in this field since then and until 1953, when M. Kendall published the already mentioned paper [269], which opened up the modern era in the research of the evolution of financial indicators.

The starting point of Kendall's analysis was his intention to detect *cycles* in the behavior of the prices of stock and commodities. Analyzing factual data (the weekly records of the prices of 19 stocks in 1928–1938, the monthly average wheat prices on the Chicago market in 1883–1934, and the cotton prices at the New York Mercantile Exchange in 1816–1951) he, to his own surprise, could find no rhythms, trends, or cycles. More than that, he concluded that these series of data look as if "*... the Demon of Chance drew a random number ... and added it to the current price to determine the next ... price*". In other words, the logarithms of the prices $S = (S_n)$ appear to behave as a *random walk*: setting $h_n = \ln \dfrac{S_n}{S_{n-1}}$ we obtain

$$S_n = S_0 e^{H_n}, \qquad n \geqslant 1,$$

where H_n is the sum of *independent* random variables $h_1, \dots, h_n$.

Here it seems appropriate to recall again (cf. § 1b) that in fact the first author (before A. Cowles and H. Working) to put forward the idea to use a *random walk* to describe the evolution of prices was L. Bachelier [12]. He conjectured that the prices $S^{(\Delta)} = (S^{(\Delta)}_{k\Delta})$ (not logarithms, by the way) change their values at instants $\Delta, 2\Delta, \dots$ so that

$$S^{(\Delta)}_{k\Delta} = S_0 + \xi_\Delta + \xi_{2\Delta} + \cdots + \xi_{k\Delta},$$

where $(\xi_{i\Delta})$ are independent identically distributed random variables taking the values $\pm\sigma\sqrt{\Delta}$ with probability $\frac{1}{2}$. Hence

$$\mathsf{E} S^{(\Delta)}_{k\Delta} = S_0, \qquad \mathsf{D} S^{(\Delta)}_{k\Delta} = \sigma \cdot (k\Delta).$$

We have already mentioned that, setting $k = \left[\dfrac{t}{\Delta}\right]$, $t > 0$, and passing formally to the limit L. Bachelier discovered that the limiting process $S = (S_t)_{t\geqslant 0}$, where $S_t = \lim\limits_{\Delta\to 0} S^{(\Delta)}_{[\frac{t}{\Delta}]\Delta}$ (we must understand the limit here in a certain suitable probabilistic sense), had the following form:

$$S_t = S_0 + \sigma W_t,$$

where $W = (W_t)$ ($W_0 = 0$, $\mathsf{E} W_t = 0$, $\mathsf{E} W_t^2 = t$) was a process that is usually called now a *standard Brownian motion* or a *Wiener process*: a process with *independent Gaussian* (*normal*) *increments* and *continuous trajectories*. (See Chapter III, § § 3a, 3b for greater detail.)

2. The interest to a more thorough study of the dynamics of financial indexes and the construction of various probabilistic models explaining the phenomena revealed by observations (such as, e.g., the *cluster* property) has significantly grown after M. Kendall's paper. We single out two studies of the late 1950s: [405] by H. Roberts (1959) and [371] by M. F. M. Osborne (1959).

Roberts's paper, which drew upon the ideas of H. Working and M. Kendall, was addressed directly to practitioners and contained heuristic arguments in favor of the *random walk conjecture.* "Brownian Motion in the Stock Market" by the astrophysicist Osborne grew up, by his own words (see [35; p. 103]), from the desire to test his physical and statistical techniques on such 'earthly' items as stock prices. Unacquainted with the works of L. Bachelier, H. Working, and M. Kendall as he was, M. F. M. Osborne came in effect to the same conclusions; he pointed out, however (and this proved to be important for the subsequent development), that these were the logarithms of the prices S_t that varied in accordance with the law of a Brownian motion (with drift), not the prices themselves (which were the main point of Bachelier's analysis). This idea was later developed by P. Samuelson [420], who introduced into the theory and practice of finance a *geometric* (or, to use his own term, *economic*) *Brownian motion*

$$S_t = S_0 e^{\sigma W_t + (\mu - \sigma^2/2)t}, \qquad t \geqslant 0,$$

where $W = (W_t)$ is a standard Brownian motion.

3. It would be exaggeration to say that the use of the *random walk conjecture* for the description of the evolution of prices was accepted by economists then and there. But it was this conjecture that gave rise to the concept of *rational* (or, as one often says, *efficient*) *market*, whose initial destination was to provide arguments in favor of the use of *probabilistic* concepts and, in this context, to demonstrate the plausibility of the *random walk conjecture* and of the (more general) *martingale conjecture.*

In few words, 'efficiency' here means that the market responds *rationally* to new information. This implies that, on this market

1) *corrections of prices are instantaneous* and the market is always in 'equilibrium', the prices are 'fair' and leave the participants no room for arbitrage, i.e., for drawing profits from price differentials;
2) the *dealers* (traders, investors, etc.) are *uniform* in their interpretation of the obtained information and *correct their decisions instantaneously* as new information becomes available;
3) the participants are *homogeneous* in their goals; their actions are 'collectively rational'.

Incidentally, on the *formal* side the concept of 'efficiency' must be considered as related to and dependent upon the nature of the *information* flowing to the market and its participants.

One usually distinguishes between *three* kinds of accessible data:

1° the *past* values of the prices;
2° the information of a more broad character than the prices contained in *generally accesible sources* (newspapers, bulletins, TV, etc.);
3° all *conceivable* information.

For a suitable formalization of our concept of 'information' we start from the assumption that the 'uncertainty' in the market can be described (see § 1a in Chapter II for additional detail) as 'randomness' interpreted in the context of *some* probability space $(\Omega, \mathscr{F}, \mathsf{P})$. As usual, here

$\Omega = \{\omega\}$ is the *space of elementary outcomes*

$\mathscr{F}$ is some *σ-algebra of subsets of* Ω,

P is a *probability measure on* $(\Omega, \mathscr{F})$.

It is worthwhile to endow the probability space $(\Omega, \mathscr{F}, \mathsf{P})$ with a *flow* (*filtration*) $\mathbb{F} = (\mathscr{F}_n)_{n \geqslant 0}$ of σ-subalgebras $\mathscr{F}_n$ such that $\mathscr{F}_m \subseteq \mathscr{F}_n \subseteq \mathscr{F}$ for $m \leqslant n$.

We interpret the events in $\mathscr{F}_n$ as the '*information*' accessible to an observer up to the instant n.

4. *Remark.* In connection with the concept of an 'event' observable before time n, which is formally a subset in $\mathscr{F}_n$, we emphasize the following.

In experimenting (e.g., keeping record of the prices) we usually do not want that much to know of a *specific* outcome as we want to know whether this outcome belongs to one or another *subset* of all possible outcomes.

We mean by observable events sets $A \subseteq \Omega$ such that, under the conditions of our experiment (given the general state of the market in our case), there can be answers of two types, 'the outcome ω belongs to A' or 'the outcome ω does not belong to A'. For instance, if $\omega = (\omega_1, \omega_2, \omega_3)$, where we interpret the value $\omega_i = +1$ as a rise in prices at time i, while $\omega_i = -1$ means a drop in prices at the same instant, then the space of all conceivable outcomes (in this three-step model) consists of eight points:

$$\Omega = \big\{(-1,-1,-1), (-1,-1,+1), \dots, (+1,+1,+1)\big\}.$$

If we register *all* the values $\omega_1, \omega_2, \omega_3$, then, for instance, the set

$$A = \big\{(-1,+1,-1), (+1,-1,+1)\big\}$$

is an 'event' since we can say with confidence whether '$\omega \in A$' or '$\omega \notin A$'.

However, if we cannot record the result ω_2 at time $i = 2$, i.e., we have no information on this value, then A is not an 'event' any longer: knowing only ω_1 and ω_3 we cannot answer the question whether $\omega = (\omega_1, \omega_2, \omega_3)$ is in the set A.

5. In stochastic calculus one usually calls spaces $(\Omega, \mathscr{F}, \mathbb{F} = (\mathscr{F}_n), \mathsf{P})$ with distinguished flows $\mathbb{F} = (\mathscr{F}_n)$ of σ-algebras *filtered* probability spaces. In the framework of financial mathematics we shall also call $\mathbb{F} = (\mathscr{F}_n)$ an *information flow.* Using this concept we can describe various forms of *efficient* markets as follows.

Assume that there are three flows of σ-algebras

$$\mathbb{F}^1 = (\mathscr{F}_n^1), \qquad \mathbb{F}^2 = (\mathscr{F}_n^2), \qquad \mathbb{F}^3 = (\mathscr{F}_n^3),$$

in $(\Omega, \mathscr{F}, \mathsf{P})$, where $\mathscr{F}_n^1 \subseteq \mathscr{F}_n^2 \subseteq \mathscr{F}_n^3$, and we interpret each of the σ-algebras $\mathscr{F}_n^i$ as the data of the (i)th kind arriving at the instant n.

Following E. Fama [150] (1965) we shall say that a market is *weakly efficient* if there exists a discounting price $B = (B_n)_{n\geqslant 0}$ (usually, this is a risk-free bank account) and a probability measure $\widehat{\mathsf{P}}$ locally equivalent to P (i.e., its restrictions $\widehat{\mathsf{P}}_n = \widehat{\mathsf{P}} \,|\, \mathscr{F}_n^1$ to $\mathscr{F}_n^1$ are equivalent to $\mathsf{P}_n = \mathsf{P} \,|\, \mathscr{F}_n$ for each $n \geqslant 0$) such that each price $S = (S_n)$ (of a financial instrument on this market) satisfies the following condition: the ratio

$$\frac{S}{B} = \left(\frac{S_n}{B_n}\right)_{n\geqslant 0}$$

is a $\widehat{\mathsf{P}}$-martingale, i.e., the variables $\dfrac{S_n}{B_n}$ are $\mathscr{F}_n^1$-measurable and

$$\widehat{\mathsf{E}}\left|\frac{S_n}{B_n}\right| < \infty, \quad \widehat{\mathsf{E}}\left(\frac{S_{n+1}}{B_{n+1}} \,\Big|\, \mathscr{F}_n^1\right) = \frac{S_n}{B_n}, \qquad n \geqslant 0.$$

If we have the martingale property with respect to the information flow $\mathbb{F}^3 = (\mathscr{F}_n^3)$ then we have a *strongly efficient market*, while in the case of $\mathbb{F}^2 = (\mathscr{F}_n^2)$ we have a *semi-strongly efficient market*.

For simplicity, we shall assume throughout this section that $B_n \equiv 1$ and $\widehat{\mathsf{P}} = \mathsf{P}$.

Before a discussion of these definitions we point out the following relation between the classes of martingales $\mathscr{M}^1 = \mathscr{M}(\mathbb{F}^1)$, $\mathscr{M}^2 = \mathscr{M}(\mathbb{F}^2)$, and $\mathscr{M}^3 = \mathscr{M}(\mathbb{F}^3)$ with respect to the flows $\mathbb{F}^1$, $\mathbb{F}^2$, and $\mathbb{F}^3$:

$$\mathscr{M}^3 \subseteq \mathscr{M}^2 \subseteq \mathscr{M}^1.$$

In fact, the inclusion $S = (S_n)_{n\geqslant 0} \in \mathscr{M}^2$ (for one) means that the S_n are $\mathscr{F}_n^2$-measurable and $\mathsf{E}(S_{n+1} \,|\, \mathscr{F}_n^2) = S_n$. Hence, by the 'telescopic' property of conditional expectations

$$\mathsf{E}(S_{n+1} \,|\, \mathscr{F}_n^1) = \mathsf{E}\big(\mathsf{E}(S_{n+1} \,|\, \mathscr{F}_n^2) \,|\, \mathscr{F}_n^1\big)$$

and since the S_n are $\mathscr{F}_n^1$-measurable (we recall that $\mathscr{F}_n^1$ is generated by *all* prices up to time n, including the variables S_k, $k \leqslant n$), it follows that $\mathsf{E}(S_{n+1} \,|\, \mathscr{F}_n^1) = S_n$, i.e., $S \in \mathscr{M}^1$.

If $\xi_1, \xi_2, \dots$ is a sequence of *independent* random variables such that $\mathsf{E}|\xi_k| < \infty$, $\mathsf{E}\xi_k = 0$, $k \geqslant 1$, $\mathscr{F}_n^\xi = \sigma(\xi_1, \dots, \xi_n)$, $\mathscr{F}_0^\xi = \{\varnothing, \Omega\}$, and $\mathscr{F}_n^\xi \subseteq \mathscr{F}_n^1$, then, evidently, the sequence $S = (S_n)_{n\geqslant 0}$, where $S_n = \xi_1 + \cdots + \xi_n$ for $n \geqslant 1$ and $S_0 = 0$, is a martingale with respect to $\mathscr{F}^\xi = (\mathscr{F}_n^\xi)_{n\geqslant 0}$ and

$$\mathsf{E}(S_{n+1} \,|\, \mathscr{F}_n^i) = S_n + \mathsf{E}(\xi_{n+1} \,|\, \mathscr{F}_n^i), \qquad i = 1, 2, 3.$$

Hence the sequence $S = (S_n)_{n\geqslant 0}$ is clearly a martingale of class $\mathscr{M}^i$ if for each n the variables ξ_{n+1} are independent of $\mathscr{F}_n^i$ (in the sense that for each Borel set A

the event $\{\xi_{n+1} \in A\}$ is independent of the events in $\mathscr{F}_n^i$). Thus, if we treat ξ_{n+1} as 'entirely fresh data' relative to $\mathscr{F}_n^3$, then S is in the class $\mathscr{M}^3$.

It is important for what follows that if $X = (X_n)$ is a martingale with respect to the filtration $\mathbb{F} = (\mathscr{F}_n)$ and $X_n = x_1 + \cdots + x_n$ with $x_0 = 0$, then $x = (x_n)$ is a *martingale difference*, i.e.,

$$x_n \quad \text{is} \quad \mathscr{F}_n\text{-measurable},$$
$$\mathsf{E}|x_n| < \infty,$$
$$\mathsf{E}(x_n \mid \mathscr{F}_{n-1}) = 0.$$

It follows from the last property that, provided that $\mathsf{E}|x_n|^2 < \infty$ for $n \geqslant 1$, we have

$$\mathsf{E}\, x_n x_{n+k} = 0$$

for each $n \geqslant 0$ and $k \geqslant 1$, i.e., the variables $x = (x_n)$ are *uncorrelated*. In other words, *square-integrable* martingales belong to the class of random sequences with *orthogonal increments*:

$$\mathsf{E}\, \Delta X_n \Delta X_{n+k} = 0,$$

where $\Delta X_n \equiv X_n - X_{n-1} = x_n$ and $\Delta X_{n+k} = x_{n+k}$. Denoting the class of such sequences by OI_2 (here the subscript 2 corresponds to 'square integrability') we obtain

$$\mathscr{M}_2^3 \subseteq \mathscr{M}_2^3 \subseteq \mathscr{M}_2^1 \subseteq OI_2.$$

It follows from the above that, in the final analysis, the *efficiency* of a market is nothing else but the *martingal property* of prices in it. One example is market where prices are 'random walks'.[d]

6. Why do we believe the conjecture of the 'martingale property', which generalizes the 'random walk' conjecture and is inherent in the concept of 'efficient market', to be quite a natural one? There can be several answers. Arguably, the best explanation can be given in the framework of the *theory of arbitrage-free markets*, which directly associates the *efficiency* (or more specifically, the *rationality*) of a market with the absence of opportunities for arbitrage. It will be clear from what follows that the latter property immediately results in the appearance of martingales. (See Chapter V, § 2a for fuller detail.)

Now, to give an idea of the way in which martingales appear in this context, we present the following elementary arguments.

---

[d]In the probability and statistics literature one usually considers a 'random walk' to be a walk that can be described by the sum of *independent* random variables. On the other hand, the economists sometimes use this term in another sense, to emphasize merely the *random* nature of, say, price movement.

Let $S = (S_n)_{n\geqslant 1}$, where S_n is the price of, say, one share at the instant n. Let

$$\rho_n = \frac{\Delta S_n}{S_{n-1}}, \qquad n \geqslant 1,$$

(here $\Delta S_n \equiv S_n - S_{n-1}$) be the relative change in prices (the *interest rate*) and assume that the market is organized in such a way that, with respect to the flow $\mathbb{F} = (\mathscr{F}_n)$ of accessible data, the variables S_n are $\mathscr{F}_n$-measurable and (P-almost everywhere)

$$\mathsf{E}(\rho_n \mid \mathscr{F}_{n-1}) = r, \tag{1}$$

for some constant r. By the last two formulas,

$$S_n = (1 + \rho_n) S_{n-1} \tag{2}$$

and (assuming that $1 + r \neq 0$)

$$S_{n-1} = \frac{\mathsf{E}(S_n \mid \mathscr{F}_{n-1})}{1 + r}. \tag{3}$$

By assumption,

$$\Delta S_n = \rho_n S_{n-1}, \qquad n \geqslant 1.$$

We also assume that, along with our share, we consider a *bank account* $B = (B_n)_{n\geqslant 0}$ such that

$$\Delta B_n = r B_{n-1}, \qquad n \geqslant 1, \tag{4}$$

where r is the *interest rate* of the account and $B_0 > 0$.

Since $B_n = B_0(1 + r)^n$ by (4), we can see from (3) that

$$\frac{S_{n-1}}{B_{n-1}} = \mathsf{E}\biggl(\frac{S_n}{B_n} \biggm| \mathscr{F}_{n-1}\biggr).$$

This means precisely that the sequence $\left(\dfrac{S_n}{B_n}\right)_{n\geqslant 1}$ is a *martingale* with respect to the flow $\mathbb{F} = (\mathscr{F}_n)_{n\geqslant 1}$.

Our above assumption $\mathsf{E}(\rho_n \mid \mathscr{F}_{n-1}) = r$ (P-a.e.) seems to be fairly natural 'in an economist's view': otherwise (e.g., if $\mathsf{E}(\rho_n \mid \mathscr{F}_{n-1}) > r$ (P-almost everywhere) or $\mathsf{E}(\rho_n \mid \mathscr{F}_{n-1}) < r$ (P-almost everywhere) for $n \geqslant 1$) the investors will find out promptly that it is more profitable to restrict their investment to stock (in the first case) or to the bank account (in the second case). To put it another way, if one security 'dominates' another, then the less valuable one will swiftly disappear, as it should be in a 'well-organized', 'efficient' market.

7. We consider now a somewhat more complicated version of our model (2) of the evolution of stock prices.

Assuming that at time $n-1$ you *buy* one share at a price S_{n-1} and, at time n, *sell* it (at a price S_n) your (gross) 'profit' (which can be either positive or negative) is $\Delta S_n \equiv S_n - S_{n-1}$. It is, of course, more sensible to measure the 'profit' in the relative values $\left(\dfrac{\Delta S_n}{S_{n-1}}\right)$ (as we did it earlier), rather than in the absolute ones, ΔS_n, i.e., to compare ΔS_n and the money S_{n-1} paid for the share.

For example, if $S_{n-1} = 20$, while $S_n = 29$, then $\Delta S_n = 9$, which is not all that little compared with 20. But if $S_{n-1} = 200$ and $S_n = 209$, then the increment ΔS_n is 9 again; now however, compared with 200, this is not all that much.

Thus, $\rho_n = 9/20$ ($= 45\%$) in the first case, while $\rho_n = 9/200$ ($= 4.5\%$) in the second.

One often calls for brevity these *relative profits* the *returns* or the *growth coefficients* (alongside the already used term '(random) interest rate'). We shall occasionally use this terminology in what follows.

In accordance with our interpretation of the increments $\Delta S_n = S_n - S_{n-1}$ as the profits from buying (at time $n-1$) and selling (at time n), we now assume that we have an *additional source of revenue*, e.g., dividends on stock, which we assume to be $\mathscr{F}_n$-measurable and equal to δ_n at time n.

Then our total 'gross' profits are $\Delta S_n + \delta_n$, while their relative value is

$$\rho_n = \frac{\Delta S_n + \delta_n}{S_{n-1}}. \tag{5}$$

It would be interesting to have a notion of the possible 'global' pattern of the behavior of the prices (S_n), provided that, 'locally', this behavior can be described by the model (5). Clearly, to answer this question we must make certain assumptions about (ρ_n) and (δ_n).

With this in mind we now assume, for instance, that

$$\mathsf{E}(\rho_n \mid \mathscr{F}_{n-1}) \equiv r \geqslant 0$$

for all $n \geqslant 1$. If, in addition, $\mathsf{E}|S_n| < \infty$, and $\mathsf{E}|\delta_n| < \infty$, then

$$S_{n-1} = \frac{1}{1+r}\mathsf{E}(S_n \mid \mathscr{F}_{n-1}) + \frac{1}{1+r}\mathsf{E}(\delta_n \mid \mathscr{F}_{n-1}) \tag{6}$$

by (5), where $\mathsf{E}(\,\cdot \mid \mathscr{F}_{n-1})$ is a conditional expectation.

In a similar way,

$$S_n = \frac{1}{1+r}\mathsf{E}(S_{n+1} \mid \mathscr{F}_n) + \frac{1}{1+r}\mathsf{E}(\delta_{n+1} \mid \mathscr{F}_n),$$

which in view of (6), brings us to the equality

$$S_{n-1} = \frac{1}{(1+r)^2} \mathsf{E}(S_{n+1} \mid \mathscr{F}_{n-1}) + \frac{1}{(1+r)^2} \mathsf{E}(\delta_{n+1} \mid \mathscr{F}_{n-1}) + \frac{1}{1+r} \mathsf{E}(\delta_n \mid \mathscr{F}_{n-1}).$$

Continuing, we see that

$$S_n = \frac{1}{(1+r)^k} \mathsf{E}(S_{n+k} \mid \mathscr{F}_n) + \sum_{i=1}^{k} \frac{1}{(1+r)^i} \mathsf{E}(\delta_{n+i} \mid \mathscr{F}_n) \tag{7}$$

for each $k \geqslant 1$ and each $n \geqslant 1$.

Hence it is clear that each *bounded* solution ($|S_n| \leqslant$ const for $n \geqslant 1$) of (6) (for $n \geqslant 1$) has the following form (provided that $|\mathsf{E}(\delta_{n+i} \mid \mathscr{F}_n)| \leqslant$ const for $n \geqslant 0$, $i \geqslant 1$):

$$S_n = \sum_{i=1}^{\infty} \frac{1}{(1+r)^i} \mathsf{E}(\delta_{n+i} \mid \mathscr{F}_n). \tag{8}$$

In the economics literature this is called the *market fundamental solution* (see, e.g., [211]). In the particular case of dividends unchanged with time ($\delta_n \equiv \delta =$ const) and $\mathsf{E}(\rho_n \mid \mathscr{F}_{n-1}) \equiv r > 0$ it follows from (8) that the (bounded) prices S_n, $n \geqslant 1$, must also be constant:

$$S_n \equiv \frac{\delta}{r}, \qquad n \geqslant 1.$$

8. The class of martingales is fairly wide. For instance, it contains the 'random walk' considered above. Further, the martingale property

$$\mathsf{E}(X_n \mid \mathscr{F}_{n-1}) = X_{n-1}$$

shows that, as regards the *predictions* of the values of the increments $\Delta X_n = X_n - X_{n-1}$, the best we can get out of the 'data' $\mathscr{F}_{n-1}$ is that *the increment vanishes on average* (with respect to $\mathscr{F}_{n-1}$). This conforms with our innate perception that the conditional gains $\mathsf{E}(\Delta X_n \mid \mathscr{F}_{n-1})$ must vanish in a 'fair', 'well-organized' market, which, in turn, can be interpreted as the impossibility of riskless profits. It is in that connection that L. Bachelier wrote (in the English translation): "The mathematical expectation of the speculator is zero". (We recall that, in gambling, the system when one doubles one's stake after a loss and drops out after the first win is called the *martingale*; the conditional gains for this strategy are $\mathsf{E}(\Delta X_n \mid \mathscr{F}_{n-1}) = 0$; see [439; Chapter VII, § 1] for detail.)

Finally, we point out that, as shows an empirical analysis of price evolution (Chapter IV, § 3c), the autocorrelation of the variables

$$h_n = \ln \frac{S_n}{S_{n-1}}, \qquad n \geqslant 1,$$

is close to zero, which can be regarded as an argument (albeit indirect one) in favor of the martingale conjecture.

9. The conjecture of efficient markets gave an impetus to the development of new financial instruments, suitable for 'cautious' investors adhering to the idea of diversification (see § 2b).

Among these instruments we must first place a subvariety of 'Mutual Funds', the so-called 'Index Funds'.

The peculiarity of these funds is that they invest (their clients') money in shares of corporations included into one or another 'Index' of stock.

One of the first such funds was (and still is) Vanguard Index Trust—500 Portfolio (founded in 1976 by Vanguard Group; USA) that operates (buys and sells) shares of the firms included in Standard&Poor's-500 Index, which comprises the stock of 500 corporations (400 industrial companies, 20 transportation companies, 40 utilities, and 40 financial corporations).

According to the conjecture of an efficient market, prices change (and therefore financial decisions also change—and rather promptly at that) when the *information* is updated. On the other hand, commonplace investors (either individuals or institutions) do not have sufficient information and usually cannot quickly respond to the changes of prices. Moreover, the overheads of a 'lone trader' can 'eat up' all his profits.

For that reason investment in index funds can be attractive for those 'underinformed' investors who do not expect 'prompt-and-big' profits, but prefer (cautious) well-diversified investment in long-term securities instead.

Other examples of similar funds (issued by Vanguard) are Vanguard Index Trust—Extended Market Portfolio, Vanguard Index Trust—Small Capitalization Stock Portfolio, and Vanguard Bond Index Fund, based mostly on American securities, and also International Equity Fund—European Portfolio, Pacific Portfolio, and others based on foreign securities.

§ 2b. Investment Portfolio. Markowitz's Diversification

1. We already noted in § 2a that Markowitz's paper [332] (1952) was decisive for the development of the modern theory and practice of financial management and financial engineering. The most attractive point for investors in his theory was the idea of the *diversification* of an *investment portfolio*, because it did not merely demonstrate a theoretical possibility to *reduce* (unsystematic) investment *risks*, but also gave recommendations how one could achieve that in practice.

To clarify the basics and the main ideas of this theory we consider the following *single-step* investment problem.

Assume that the investor can allocate his starting capital x at the instant $n = 0$ among the stocks $A_1, \dots, A_N$ at the prices $S_0(A_1), \dots, S_0(A_N)$, respectively.

Let $X_0(b) = b_1 S_0(A_1) + \cdots + b_N S_0(A_N)$, where $b_i \geqslant 0$, $i = 1, \dots; N$. In other words, let

$$b = (b_1, \dots, b_N)$$

be the *investment portfolio*, where b_i is the number of shares A_i of value $S_0(A_i)$.

We assume the following law governing the evolution of the price of each share A_i: its price $S_1(A_i)$ at the instant $n = 1$ must satisfy the difference equation

$$\Delta S_1(A_i) = \rho(A_i) S_0(A_i),$$

or, equivalently,

$$S_1(A_i) = \big(1 + \rho(A_i)\big) S_0(A_i),$$

where $\rho(A_i)$ is the random interest rate of A_i, $\rho(A_i) > -1$.

If the investor has selected the portfolio $b = (b_1, \dots, b_N)$, then his initial capital $X_0(b) = x$ becomes

$$X_1(b) = b_1 S_1(A_1) + \cdots + b_N S_1(A_N),$$

and he would like to make the last value 'a bit larger'. However, his desire must be weighted against the 'risks' involved.

To this end Markowitz considers the following two characteristics of the capital $X_1(b)$:

$$\mathsf{E} X_1(b), \quad \text{its } \textit{expectation}$$

and

$$\mathsf{D} X_1(b), \quad \text{its } \textit{variance.}$$

Given these parameters, there are several ways to pose an optimization problem of the best portfolio choice depending on the optimality criteria.

For example, we can ask which portfolio b^* delivers the maximum value for some performance $f = f\big(\mathsf{E} X_1(b), \mathsf{D} X_1(b)\big)$ under the following 'budget constraint' on the class of admissible portfolios:

$$B(x) = \big\{ b = (b_1, \dots, b_N) : b_i \geqslant 0,\ X_0(b) = x \big\}, \qquad x > 0.$$

There exists also a natural *variational* setting: find

$$\inf \mathsf{D} X_1(b)$$

over the portfolios b satisfying the conditions

$$b \in B(x),$$
$$\mathsf{E} X_1(b) = m,$$

where m is a fixed constant.

Fig. 8 depicts a typical pattern of the set of points $\left(\mathsf{E}X_1(b), \sqrt{\mathsf{D}X_1(b)}\right)$ such that the portfolio b belongs to $B(x)$ and, maybe, satisfies also some additional constraints.

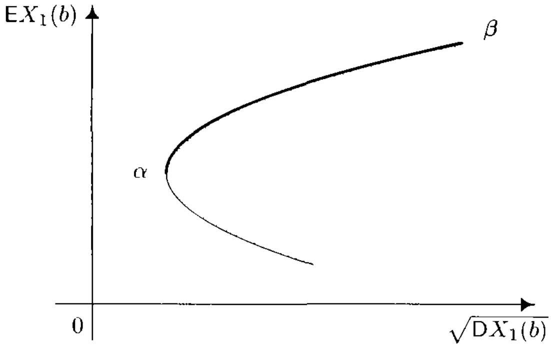


FIGURE 8. Illusration to Markowitz's mean-variance analysis

It is clear from this picture that if we are interested in the *maximum mean value* of the value of the portfolio with *minimum variance*, then we must choose portfolios such that the points $\left(\mathsf{E}X_1(b), \sqrt{\mathsf{D}X_1(b)}\right)$ lie on the thick piece of curve with end-points α and β. (Markowitz says that these are *efficient* portfolios and he terms the above kind of analysis of mean values and variances *Mean-variance analysis*.)

2. We now claim that in the *single-step* optimization problem for an investment portfolio, in place of the quantities $(S_1(A_1), \dots, S_1(A_N))$ we can directly consider the interest rates $\big(\rho(A_1), \dots, \rho(A_N)\big)$. This means the following.

Let $b \in B(x)$, i.e., let $x = b_1 S_0(A_1) + \dots + b_N S_0(A_N)$. We introduce the quantities $d = (d_1, \dots, d_N)$ by the equalities

$$d_i = \frac{b_i S_0(A_i)}{x}.$$

Since $b \in B(x)$, it follows that $d_i \geqslant 0$ and $\sum_{i=1}^{N} d_i = 1$. We can represent the portfolio value $X_1(b)$ as the product

$$X_1(b) = \big(1 + R(b)\big) X_0(b),$$

and let

$$\rho(d) = d_1 \rho(A_1) + \dots + d_N \rho(A_N).$$

Clearly,

$$\begin{aligned} R(b) &= \frac{X_1(b)}{X_0(b)} - 1 = \frac{X_1(b)}{x} - 1 = \frac{\sum b_i S_1(A_i)}{x} - 1 \\ &= \sum d_i \frac{S_1(A_i)}{S_0(A_i)} - 1 = \sum d_i \left(\frac{S_1(A_i)}{S_0(A_i)} - 1\right) = \sum d_i \rho(A_i) = \rho(d). \end{aligned}$$

Thus,

$$R(b) = \rho(d),$$

therefore if $d = (d_1, \dots, d_N)$ and $b = (b_1, \dots, b_N)$ satisfy the relations $d_i = \dfrac{b_i S_0(A_i)}{x}$, $i = 1, \dots, N$, then

$$X_1(b) = x\bigl(1 + \rho(d)\bigr)$$

for $b \in B(x)$. Hence, to solve an optimization problem for $X_1(b)$ we can consider the corresponding problem for $\rho(d)$.

3. We now discuss the issue of *diversification* as a means of reduction of the (unsystematic) risks to an arbitrarily low level, measured in terms of the variance or the standard error of $X_1(b)$.

To this end we consider first a pair of random variables ξ_1 and ξ_2 with finite second moments. If c_1 and c_2 are constants and $\sigma_i = \sqrt{\mathsf{D}\xi_i}$, $i = 1, 2$, then

$$\mathsf{D}(c_1\xi_1 + c_2\xi_2) = (c_1\sigma_1 - c_2\sigma_2)^2 + 2c_1c_2\sigma_1\sigma_2(1 + \sigma_{12}),$$

where $\sigma_{12} = \dfrac{\mathsf{Cov}(\xi_1, \xi_2)}{\sigma_1\,\sigma_2}$ and $\mathsf{Cov}(\xi_1, \xi_2) = \mathsf{E}\xi_1\xi_2 - \mathsf{E}\xi_1\mathsf{E}\xi_2$. Hence, if $c_1\sigma_1 = c_2\sigma_2$ and $\sigma_{12} = -1$, then, clearly, $\mathsf{D}(c_1\xi_1 + c_2\xi_2) = 0$.

Thus, if the variables ξ_1 and ξ_2 have *negative* correlation with coefficient $\sigma_{12} = -1$, then we can choose c_1 and c_2 such that $c_1\sigma_1 = c_2\sigma_2$ so as to obtain a combination $c_1\xi_1 + c_2\xi_2$ of variance *zero*. Of course, we can attain in this way a fairly small mean value $\mathsf{E}(c_1\xi_1 + c_2\xi_2)$. (The case of $c_1 = c_2 = 0$ is of no interest for optimization since $b \in B(x)$.)

It becomes clear from these elementary arguments that, given our constraints on (c_1, c_2) and the class of the variables (ξ_1, ξ_2), in the solution of the problem of making $\mathsf{E}(c_1\xi_1 + c_2\xi_2)$ possibly larger and $\mathsf{D}(c_1\xi_1 + c_2\xi_2)$ possibly smaller we must choose pairs (ξ_1, ξ_2) with covariance as close to -1 as possible.

The above phenomenon of *negative correlation*, which is also called the *Markowitz phenomenon*, is one of the basic ideas of investment diversification: *in building an investment portfolio one must invest in possibly many negatively correlated securities.*

Another concept at the heart of diversification is based on the following idea.

Let $\xi_1, \xi_2, \dots, \xi_N$ be a sequence of *uncorrelated* random variables with variances $\mathsf{D}\xi_i \leqslant C$, $i = 1, \dots, N$, where C is a constant. Then

$$\mathsf{D}(d_1\xi_1 + \dots + d_N\xi_N) = \sum_{i=1}^N d_i^2\,\mathsf{D}\xi_i \leqslant C\sum_{i=1}^N d_i^2.$$

Hence, setting, e.g., $d_i = 1/N$ we see that

$$\mathsf{D}(d_1\xi_1 + \dots + d_N\xi_N) \leqslant \frac{C}{N} \to 0 \quad \text{as} \quad N \to \infty.$$

This *phenomenon of the absence of correlation* indicates that, in the case of investment in uncorrelated securities, their number N must be *possibly larger* to reduce the risk (i.e., the variance $\mathsf{D}(d_1\xi_1 + \cdots + d_N\xi_N)$).

We return now to the question of the variance $\mathsf{D}\rho(d)$ of the variable

$$\rho(d) = d_1\rho(A_1) + \cdots + d_N\rho(A_N).$$

We have

$$\mathsf{D}\rho(d) = \sum_{i=1}^{N} d_i^2 \mathsf{D}\rho(A_i) + \sum_{\substack{i,j=1,\\ i\neq j}}^{N} d_i d_j \operatorname{Cov}(\rho(A_i), \rho(A_j)).$$

We now set $d_i = 1/N$. Then

$$\sum_{i=1}^{N} d_i^2 \mathsf{D}\rho(A_i) = \left(\frac{1}{N}\right)^2 \cdot N \cdot \frac{1}{N}\sum_{i=1}^{N} \mathsf{D}\rho(A_i) = \frac{1}{N}\cdot \overline{\sigma}_N^2,$$

where $\overline{\sigma}_N^2 = \dfrac{1}{N}\sum\limits_{i=1}^{N} \mathsf{D}\rho(A_i)$ is the mean variance. Further,

$$\sum_{\substack{i,j=1,\\ i\neq j}}^{N} d_i d_j \operatorname{Cov}(\rho(A_i), \rho(A_j)) = \left(\frac{1}{N}\right)^2 (N^2 - N)\cdot \overline{\operatorname{Cov}}_N,$$

where

$$\overline{\operatorname{Cov}}_N = \frac{1}{N^2 - N}\sum_{\substack{i,j=1,\\ i\neq j}}^{N} \operatorname{Cov}(\rho(A_i), \rho(A_j))$$

is the mean covariance. Thus,

$$\mathsf{D}\rho(d) = \frac{1}{N}\overline{\sigma}_N^2 + \left(1 - \frac{1}{N}\right)\overline{\operatorname{Cov}}_N, \tag{1}$$

and, clearly, if $\overline{\sigma}_N^2 \leqslant C$ and $\overline{\operatorname{Cov}}_N \to \overline{\operatorname{Cov}}$ as $N \to \infty$, then

$$\mathsf{D}\rho(d) \to \overline{\operatorname{Cov}} \quad \text{as} \quad N \to \infty. \tag{2}$$

We see from this formula that if $\overline{\operatorname{Cov}}$ is zero, then using diversification with N sufficiently large we can reduce the investment risk $\mathsf{D}\rho(d)$ to an arbitrarily low level. Unfortunately, the prices on, say, a stockmarket usually have positive correlation (their change in a fairly coordinated manner, in the same direction), and therefore $\overline{\operatorname{Cov}}_N$ *does not* approach zero as $N \to \infty$. The limit value of $\overline{\operatorname{Cov}}$ is just the *systematic* (or *market*) risk inherent to the market in question, which cannot be reduced by diversification, while the first term in (1) describes the *unsystematic risks*, which can be reduced, as shown, by a selection of a large number of stocks. (For a more detailed exposition of *Mean-variance analysis*, see [331]–[333] and [268].)

§ 2c. *CAPM*: Capital Asset Pricing Model

1. To evaluate the optimal portfolio using mean-variance analysis one must know $\mathsf{E}\rho(A_i)$ and $\mathsf{Cov}(\rho(A_i), \rho(A_j))$, and the theory *does not* explain how one can find these values. (In practice, one approximates them on the basis of conventional *statistical* means and covariances of the past data.)

CAPM-theory (W. F. Sharpe [433] and J. Lintner, [301]) and *APT*-theory that we consider below do not merely answer the questions on the values of $\mathsf{E}\rho(A_i)$ and $\mathsf{Cov}(\rho(A_i), \rho(A_j))$, but also exhibit the dependence of the (random) interest rates $\rho(A_i)$ of particular stocks A_i on the interest rate ρ of the 'large' market of A_i. Besides the covariance $\mathsf{Cov}(\rho(A_i), \rho(A_j))$, which plays a key role in Markowitz's mean-variance analysis, *CAPM* distinguishes another important parameter, the covariance $\mathsf{Cov}(\rho(A_i), \rho)$ between the interest rates of a stock A and the market interest rate.

CAPM bases its conclusions on the concept of *equilibrium* market, which implies, in particular, that there are no overheads and all the participants (investors) are 'uniform' in that they have the same capability of prediction of the future price movement on the basis of information available to everybody, the same time horizon, and all their decisions are based on the mean values and the covariances of the prices. It is also assumed that all the assets under consideration are 'infinitely divisible' and there exists a *risk-free* security (bank account, Treasury bills, ...) with interest rate r.

The existence of a risk-free security is of crucial importance because the interest r enters all formulas of *CAPM*-theory as a 'basis variable', a reference point.

It should be pointed out in this connection that long-term observations of the mean values $\mathsf{E}\rho(A_i)$ of interest rates $\rho(A_i)$ of risky securities A_i show that $\mathsf{E}\rho(A_i) > r$. Table 1, made up of the 1926–1985 yearly averages, enables one to compare the nominal and the real (with inflation taken into account) mean values of interest rates.

TABLE 1

Security	Nominal interest rate	Interest rate in real terms taking account of inflation
Common stock	12%	8.8%
Corporate bonds	5.1%	2.1%
Government bonds	4.4%	1.4%
Treasury bills	3.5%	0.4%

2. We now explain the fundamentals of *CAPM*-theory using an example of a market operating in one step.

Let $S_1 = S_0(1+\rho)$ be the value of some (random) price on some 'large' market (for example, we can consider the values of the S&P500 Index) at time $n = 1$. Let $S_1(A) = S_0(A)(1+\rho(A))$ be the price of the asset A (some stock from S&P500 Index) with interest rate $\rho(A)$ at time $n = 1$.

The evolution of the price of the risk-free asset can be described by the formula

$$B_1 = B_0(1+r).$$

On the basis of the built-in concept of equilibrium market, *CAPM*-theory establishes (see, e.g., [268] or [433]) that for each asset A there exists a quantity $\beta(A)$ (the *beta* of this asset[e]) such that

$$\mathsf{E}[\rho(A) - r] = \beta(A)\mathsf{E}[\rho - r], \tag{1}$$

where

$$\beta(A) = \frac{\mathsf{Cov}(\rho(A), \rho)}{\mathsf{D}\rho}. \tag{2}$$

In other words, the mean value of the 'premium' $\rho(A) - r$ (for using the risky asset A in place of the risk-free one) is *in proportion* to the mean value of the *premium* $\rho - r$ (of investments in some global characteristic of the market, e.g., the S&P500 Index).

Formula (2) shows that the value $\beta(A)$ of the 'beta' is defined by the correlation properties of the interest rates ρ and $\rho(A)$ or, equivalently, by the covariance properties of the corresponding prices S_1 and $S_1(A)$.

We now rewrite (1) as follows:

$$\mathsf{E}\rho(A) = r + \beta(A)\mathsf{E}(\rho - r); \tag{3}$$

let ρ_β be the value of the interest rate $\rho(A)$ of the asset A with $\beta(A) = \beta$.

If $\beta = 0$, then

$$\rho_0 = r,$$

while if $\beta = 1$, then

$$\rho_1 = \rho.$$

Bearing this in mind we see that (3) is the *equation of the CAPM line*

$$\mathsf{E}\rho_\beta = r + \beta\mathsf{E}(\rho - r), \tag{4}$$

[e] If each asset A has its 'beta' $\beta(A)$, then one would expect it to have also 'alpha' $\alpha(A)$. This is the name several authors (see, for example [267]) use for the *mean* value $\mathsf{E}\rho(A)$.

plotted in Fig. 9 and depicting the mean returns $\mathsf{E}\rho_\beta$ from assets in their dependence on β, the interest rate r, and the mean market interest rate $\mathsf{E}\rho$.

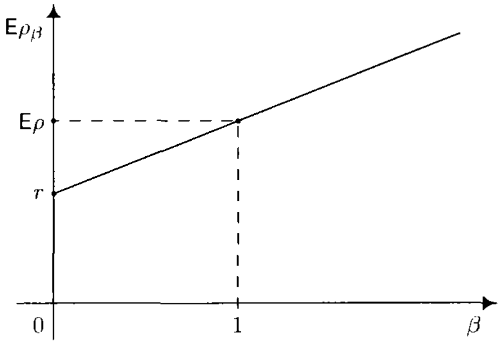


FIGURE 9. *CAPM* line

The value of $\beta = \beta(A)$ plays an important role in the selection of investment portfolios; this is 'a measure of sensitivity', 'a measure of the response' of the asset to changes on the market. For definitness we assume that we measure the state of the market in the values of the S&P500 Index and corporation A, whose shares we are now discussing, is one of the 500 firms in this index. If the index changes by 1% and the 'beta' of the stock A is 1.5, then the change in the price of A is (on the average) 1.5%.

In practice, one can evaluate the 'beta' of a certain asset using statistical data and conventional linear regression methods (since (3) is a linear relation).

3. For the asset A we now consider the variable

$$\eta(A) = \big(\rho(A) - \mathsf{E}\rho(A)\big) - \frac{\mathsf{Cov}(\rho(A), \rho)}{\mathsf{D}\rho}(\rho - \mathsf{E}\rho).$$

Clearly, $\mathsf{E}\eta(A) = 0$ and

$$\mathsf{E}\big(\eta(A)(\rho - \mathsf{E}\rho)\big) = 0,$$

i.e., the variables $\eta(A)$ and $\rho - \mathsf{E}\rho$ with means zero are *uncorrelated*. Hence

$$\rho(A) - \mathsf{E}\rho(A) = \beta(A)(\rho - \mathsf{E}\rho) + \eta(A); \tag{5}$$

this brings us by (3) to the following relation between the premiums $\rho(A) - r$ and $\rho - r$:

$$\rho(A) - r = \beta(A)(\rho - r) + \eta(A), \tag{6}$$

which shows that the premium $(\rho(A) - r)$ on the asset A is the sum of 'beta' $(\beta(A))$ times the market premium $(\rho - r)$ and the variable $\eta(A)$ uncorrelated with $\rho - \mathsf{E}\rho$.

By (5),

$$\mathsf{D}\rho(A) = \beta^2(A)\mathsf{D}\rho + \mathsf{D}\eta(A), \tag{7}$$

which means that the risk ($\mathsf{D}\rho(A)$) of an investment in A is a combination of two,

$$\text{the } \textit{systematic} \text{ risk } \quad \big(\beta^2(A)\mathsf{D}\rho\big)$$

inherent to the *market* and corresponding to the given 'beta' and the

$$\textit{unsystematic} \text{ risk } \quad \big(\mathsf{D}\eta(A)\big),$$

which is the property of the *asset* A itself.

As in the previous section, we claim that here, in the framework of *CAPM*, unsystematic risks can also be reduced by diversification. Namely, assume that there are N assets $A_1, \dots, A_N$ on the 'large' market such that the corresponding variables $\eta(A_1), \dots, \eta(A_N)$ are uncorrelated: $\mathsf{Cov}\big(\eta(A_i), \eta(A_j)\big) = 0$, $i \neq j$.

Let $d = (d_1, \dots, d_N)$ be an investment portfolio with $d_i \geqslant 0$, $\sum\limits_{i=1}^{N} d_i = 1$, and

$$\rho(d) = d_1 \cdot \rho(A_1) + \dots + d_N \cdot \rho(A_N).$$

Since

$$\rho(A_i) - r = \beta(A_i)\big[\rho - r\big] + \eta(A_i),$$

it follows that

$$\rho(d) - r = \sum_{i=1}^{N} d_i \beta(A_i)\big[\rho - r\big] + \sum_{i=1}^{N} d_i \eta(A_i).$$

Hence setting

$$\beta(d) = \sum_{i=1}^{N} d_i \beta(A_i) \quad \text{and} \quad \eta(d) = \sum_{i=1}^{N} d_i \eta(A_i),$$

we see (cf. (6)) that

$$\rho(d) - r = \beta(d)(\rho - r) + \eta(d).$$

Consequently, in a similar way to the previous section,

$$\mathsf{D}\rho(d) = \beta^2(d)\mathsf{D}\rho + \mathsf{D}\eta(d),$$

where $\mathsf{D}\eta(d) = \sum\limits_{i=1}^{N} d_i^2 \mathsf{D}\eta(A_i) \leqslant \dfrac{C}{N} \to 0$ as $N \to \infty$ if $\mathsf{D}\eta(A_i) \leqslant C$ and $d_i = \dfrac{1}{N}$.

Thus, in the framework of CAPM, *diversification enables one to reduce non-systematic risks to arbitrarily small values by a selection of sufficiently many* (N) *assets.*

We can supplement the above table of the mean values of interest rates by the table of their standard errors:

TABLE 2

Portfolio	Standard Deviation
Common stocks	21.2%
Corporate bonds	8.3%
Government bonds	8.2%
Treasury bills	3.4%

Taking the assumption that $\rho \sim \mathcal{N}(\mu, \sigma)$, i.e., we have a normal distribution with expectation μ and deviation σ, we obtain that the values of ρ lie with probability 0.9 in the interval $[\mu - 1.65\sigma, \mu + 1.65\sigma]$ and, with probability 0.67, in the interval $[\mu - \sigma, \mu + \sigma]$.

Thus, combining these two tables we obtain the following confidence intervals (on the percentage basis):

TABLE 3

Portfolio	Probability 0.67	Probability 0.9
Common stocks	$[-9.2, 33.2]$	$[-22.98, 46.98]$
Corporate bonds	$[-3.2, 13.4]$	$[-8.59, 118.79]$
Government bonds	$[-3.8, 12.6]$	$[-9.13, 17.93]$
Treasury bills	$[0.1, 6.9]$	$[-2.11, 9.11]$

(Recall that we presented the corresponding mean values in Table 1.)

We also present, for a few corporations A, the table of approximate values of their 'betas', the mean expected returns $\mathsf{E}\rho(A) = r + \beta(A)\mathsf{E}(\rho - r)$ and the standard errors $\sqrt{\mathsf{D}\rho(A)}$ (on the percentage basis):

TABLE 4

Stocks A	$\beta(A)$	$\mathsf{E}\rho(A)$	$\sqrt{\mathsf{D}\rho(A)}$
AT&T	0.81	12.4	23.1
Exxon	0.71	13.2	17.7
Compaq Computer	1.73	20.1	57.3

(The data are borrowed from [344] and [55]; the standard errors are calculated on the basis of the data for the period 1981–86; the estimates of $\beta(A)$ and $\mathsf{E}\rho(A)$ are presented as of the beginning of 1987 with $r = 5.6\%$ and $\mathsf{E}(\rho - r)$ taken to be 8.4%.

§ 2d. *APT*: Arbitrage Pricing Theory

1. In the foreground of the *CAPM*-theory there is the question on the relation between the return from a particular asset on an 'equilibrium' market and the return on it on the 'large' market where this asset is traded (see (1) in § 2c) and of the concomitant risks. Here (see formula (6) in the preceding section) the *return* (interest) $\rho(A)$ on an asset A is defined by the equality

$$\rho(A) = r + \beta(A)(\rho - r) + \eta(A). \tag{1}$$

A more recent theory of '*risks and returns*', *APT* (Arbitrage Pricing Theory; S. A. Ross, R. Roll and S. A. Ross [410]), is based on a *multiple-factor* model. It takes for granted that the value of $\rho(A)$ corresponding to A depends on several random factors $f_1, \dots, f_q$ (which can range in various domains and describe, say, the oil prices, the interest rates, etc.) and a 'noise' term $\zeta(A)$ so that

$$\rho(A) = a_0(A) + a_1(A)f_1 + \cdots + a_q(A)f_q + \zeta(A). \tag{2}$$

Here $\mathsf{E}f_i = 0$, $\mathsf{D}f_i = 1$, and $\mathsf{Cov}(f_i, f_j) = 0$ for $i \neq j$, while for the 'noise' term $\zeta(A)$ we have $\mathsf{E}\zeta(A) = 0$, and it does not correlate with $f_1, \dots, f_q$ or with the 'noise' terms corresponding to other assets.

Comparing (1) and (2) we see that (1) is a particular case of the single-factor model with factor $f_1 = \rho$. In this sense *APT* is a generalization of *CAPM*, although the methods of the latter theory remain one of the favorite tools of practitioners in security pricing, for they are clear, simple, and there exists already a well-established tradition of operating 'betas', the measures of assets' responses to market changes.

One of the central results of *CAPM*-theory, based on the *conjecture of an equilibium market*, is formula (1) in the preceding section, which describes the average premium $\mathsf{E}(\rho(A) - r)$ as a function of the average premium $\mathsf{E}(\rho - r)$.

Likewise, the central result of *APT*, which is based on the conjecture of the *absence of asymptotic arbitrage* in the market, is the (asymptotic formula) for the mean value $\mathsf{E}\rho(A)$ that we present below and that holds under the assumption that the behavior of $\rho(A)$ corresponding to an asset A in question can be described by the *multi-factor* model (2).

We recall that $\rho(A)$ is the (random) interest rate of the asset A in the *single-step* model $S_1(A) = S_0(A)(1 + \rho(A))$ (considered above).

2. Assume that we have an 'N-market' of N assets $A_1, \dots, A_N$ with q active factors $f_1, \dots, f_q$ such that

$$\rho(A_i) = a_0(A_i) + a_1(A_i)f_1 + \cdots + a_q(A_i)f_q + \zeta(A_i),$$

where $\mathsf{E}f_k = 0$, $\mathsf{E}\zeta(A_i) = 0$, the covariance $\mathsf{Cov}(f_k, f_l)$ is zero for $k \neq l$, $\mathsf{D}f_k = 1$, $\mathsf{Cov}\big(f_k, \zeta(A_i)\big) = 0$, and $\mathsf{Cov}\big(\zeta(A_i), \zeta(A_j)\big) = \sigma_{ij}$ for $k, l = 1, \dots, q$ and $i, j = 1, \dots, N$.

We now consider a portfolio $d = (d_1, \ldots, d_N)$. The corresponding return is

$$\begin{aligned}\rho(d) &= d_1\rho(A_i) + \cdots + d_N\rho(A_N)\\ &= \sum_{i=1}^N d_i a_{i0} + \left(\sum_{i=1}^N d_i a_{i1}\right) f_1 + \cdots + \left(\sum_{i=1}^N d_i a_{iq}\right) f_q + \sum_{i=1}^N d_i \zeta(A_i), \qquad (3)\end{aligned}$$

where $a_{ik} = a_k(A_i)$.

As shown below, under certain assumptions on the coefficients a_{ik} in (2) we can find a nontrivial portfolio $d = (d_1, \ldots, d_N)$ with $d_i = d_i(N)$ such that

$$d_1 + \cdots + d_N = 0, \qquad (4)$$

$$\sum_{i=1}^N d_i a_{ik} = 0, \qquad k = 1, \ldots, q, \qquad (5)$$

$$\sum_{i=1}^N d_i a_{i0} = \sum_{i=1}^N d_i^2. \qquad (6)$$

Then, for the portfolio $\theta d = (\theta d_1, \ldots, \theta d_N)$ (here θ is a constant) we have

$$\rho(\theta d) = \theta\rho(d),$$

and by (2)–(6),

$$\rho(\theta d) = \theta \sum_{i=1}^N d_i^2 + \theta \sum_{i=1}^N d_i \zeta(A_i). \qquad (7)$$

Hence

$$\mu(\theta d) = \mathsf{E}\rho(\theta d) = \theta \sum_{i=1}^N d_i^2,$$

$$\sigma^2(\theta d) = \mathsf{D}\rho(\theta d) = \theta^2 \sum_{i,j=1}^N d_i d_j \sigma_{ij}.$$

We now set

$$\theta = \left(\sum_{i=1}^N d_i^2\right)^{-2/3} \qquad \left(= \|d\|^{-4/3}, \text{ where } \|d\| = \left(\sum_{i=1}^N d_i^2\right)^{1/2}\right). \qquad (8)$$

Then

$$\mu(\theta d) = \left(\sum_{i=1}^N d_i^2\right)^{1/3}, \qquad (9)$$

$$\sigma^2(\theta d) = \frac{\sum_{i,j=1}^N d_i d_j \sigma_{ij}}{\left(\sum_{i=1}^N d_i^2\right)^{4/3}}. \qquad (10)$$

Assuming (for the simplicity of analysis; see, e.g., [240] or [268] for the general case) that $\sigma_{ij} = 0$ for $i \neq j$ and $\sigma_{ii} = 1$, we obtain

$$\sigma^2(\theta d) = \left(\sum_{i=1}^{N} d_i^2\right)^{-1/3}. \tag{11}$$

Formulas (9) and (11) play a key role in the asymptotic analysis below. They show that if $\sum\limits_{i=1}^{N} d_i^2 \to \infty$ as $N \to \infty$, then $\mu(\theta d) \to \infty$ and $\sigma^2(\theta d) \to 0$. However, if we set $S_0(A_1) = \cdots = S_0(A_N) = 1$, then by the condition $d_1 + \cdots + d_N = 0$ we obtain that the *initial* value of the portfolio θd is

$$X_0(\theta d) = \theta(d_1 + \cdots + d_N) = 0,$$

while its value at time $n = 1$ is

$$X_1(\theta d) = d_1 S_1(A_1) + \cdots + d_N S_1(A_N) = \theta\rho(d) = \rho(\theta d).$$

Further, if $\mathsf{E}X_1(\theta d) = \mu(\theta d) \to \infty$, while $\mathsf{D}X_1(\theta d) \to 0$ as $N \to \infty$, then for sufficiently large N we have $X_1(\theta d) \geqslant 0$ with large probability and $X_1(\theta d) > 0$ with non-zero probability. In other words, starting with initial capital *zero* and operating on an 'N-market' with assets $A_1, \dots, A_N$, $N \geqslant 1$, one can build a portfolio bringing one (asymptotically) nontrivial profit. This is just what is interpreted in APT as *the existence of asymptotic arbitrage.*

Thus, assuming that the 'N-markets' are *asymptotically* (as $N \to \infty$) *arbitrage-free* we arrive at the conclusion that one *must rule out* the possibility of $\sum_{i=1}^{N} d_i^2 \to \infty$, which would leave space for arbitrage. Of course, this imposes certain constraints on the coefficients of the multi-factor model (2) because the selection of the portfolio $d = (d_1, \dots, d_N)$, $d_i = d_i(N)$, $i \leqslant N$, with properties (4)–(6) described below proceeds with an eye to these coefficients.

We now consider the matrix

$$\mathscr{A} = \begin{pmatrix} 1 & a_{11} & a_{12} & \dots & a_{1q} \\ 1 & a_{21} & a_{22} & \dots & a_{2q} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & a_{N1} & a_{N2} & \dots & a_{Nq} \end{pmatrix}. \tag{12}$$

On its basis, we construct another matrix,

$$\mathscr{B} = \mathscr{A}(\mathscr{A}^*\mathscr{A})^{-1}\mathscr{A}^*, \tag{13}$$

which we assume to be well-defined (here '$*$' signifies transposition).

Let

$$\begin{aligned} d &= (I - \mathscr{B})a_0, \\ e &= \mathscr{B}a_0, \end{aligned} \tag{14}$$

where I is the identity matrix and a_0 is the column formed by $a_{10}, \dots, a_{N0}$. Then we have the orthogonal decomposition

$$a_0 = d + e \tag{15}$$

and

$$d^*\mathbf{1} = 0, \qquad d^*a_k = 0, \tag{16}$$

where a_k is the column formed by $a_{1k}, \dots, a_{Nk}$ and $\mathbf{1}$ is the column of ones.

Formulas (16) are just the above-discussed relations (4) and (5).

By (14) and (15) we also obtain

$$d^*a_0 = d^*d + d^*e = d^*d$$

which is the required formula (6).

We note now that, by (14), the column e can be represented as follows:

$$e = \lambda_0\mathbf{1} + \lambda_1 a_1 + \cdots + \lambda_q a_q,$$

where the numbers $\lambda_0, \dots, \lambda_q$ satisfy the relation

$$(\lambda_0, \dots, \lambda_q)^* = (\mathscr{A}^*\mathscr{A})^{-1}\mathscr{A}^*a_0.$$

Hence

$$d = a_0 - \lambda_0\mathbf{1} - \sum_{k=1}^{q} \lambda_k a_k$$

and

$$\sum_{i=1}^{N} d_i^2 = \sum_{i=1}^{N} \left(a_{i0} - \lambda_0 - \sum_{k=1}^{q} \lambda_k a_{ik} \right)^2. \tag{17}$$

Of course, for an 'N-market' all the coefficients $a_{i0}, \dots, a_{ik}$ and $\lambda_0, \dots, \lambda_k$ on the right-hand side of this formula depend on N.

Our assumption of the *absence* of asymptotic arbitrage rules out the possibility of

$$\lim_N \sum_{i=1}^{N} d_i^2(N) = \infty$$

(here we write $d_i(N)$ to emphasize the dependence of the number of shares of one or another asset on the total number of assets in the market), therefore by (17) the 'N-market' coefficients must ensure the inequality

$$\lim_N \sum_{i=1}^{N} \left[a_{i0}(N) - \lambda_0(N) - \sum_{k=1}^{q} \lambda_k(N) a_{ik}(N) \right]^2 < \infty. \tag{18}$$

This relation (deduced from the conjecture of the absence of asymptotic arbitrage) is interpreted in the framework of *APT* as follows: *if the number N of assets taken into account in the selection of a securities portfolio is sufficiently large, then for 'the majority of assets' we must ensure the 'almost linear' relation between the coefficients* $a_0(A_i), a_1(A_i), \dots, a_q(A_i)$:

$$a_0(A_i) \approx \lambda_0 + \sum_{k=1}^{q} \lambda_k a_k(A_i), \tag{19}$$

where all the variables under consideration depend on N and

$$a_0(A_i) = \mathsf{E}\rho(A_i).$$

Moreover, there exists a portfolio $d = (d_1, \dots, d_N)$ such that the *variance* of the return $\rho(d)$ is sufficiently small (in view of (11)), which means that the influence of the noise terms $\zeta(A_i)$ and individual active factors f_j *can be reduced by means of diversification* (*provided that there is no asymptotic arbitrage*). One should bear in mind, however, that all the above holds only for *large* N, i.e., for 'large' markets, while for small markets calculations of the return expectation $\mathsf{E}\rho(A_i)$ on the basis of the expression on the right-hand side of (19) can lead to grave errors. (As regards the corresponding precise assertion, see [231], [240], and [412]; for a *rigorous* mathematical theory of *asymptotic arbitrage* based on the concept of *contiguity*, see [250], [260], [261], [273], and Chapter VII, §§ 3a, b, c).

§ 2e. Analysis, Interpretation, and Revision of the Classical Concepts of Efficient Market. I

1. The central idea, the cornerstone of the concept of *efficient market* is the assumption that the prices instantaneously assimilate new data and are always set in a way that gives one no opportunity to 'buy cheap and sell immediately at a higher price elsewhere', i.e., as one usually says, there are no *opportunities for arbitrage.*

We have already shown that this idea of a 'rationally' organized, 'fair' market brings one to (normalized) market prices described by *martingales* (with respect to some measure equivalent to the initial probability measure).

We recall that if $X = (X_n)_{n \geqslant 0}$ is a martingale with respect to the filtration $(\mathscr{F}_n)_{n \geqslant 0}$, then $\mathsf{E}(X_{n+m} \mid \mathscr{F}_n) = X_n$. Hence the optimal (in the mean-square sense) estimator $\widehat{X}_{n+m;n}$ for the variable X_{n+m} based on the 'data' $\mathscr{F}_n$ is simply the variable X_n because $\widehat{X}_{n+m;n}$ is the same as $\mathsf{E}(X_{n+m} \mid \mathscr{F}_n)$.

Thus, we can say that our martingale conjecture for the prices (X_n) corresponds to the (economically conceivable) assumption that, on a 'well-organized' market, the best (in the mean-square sense, in any case) projection of the 'tomorrow' ('the day after tomorrow', etc.) prices that can be made on the basis of the 'today' information is the current price level.

In other words, the forecast is trivial. This seemingly rules out any chance of the prediction of 'the future dynamics of the observable prices'. (In his construction of a Brownian motion as a model for this evolution of prices L. Bachelier started essentially just from this idea of the impossibility of a better forecast of the 'tomorrow' prices than a mere repetition of their current levels.)

At the same time, it is well known that market operators (including such experts as 'fundamentalists', 'technicians', and quantitative analysts, 'quants') *are still trying to forecast* 'the future price movement', *foresee* the direction of changes and the future price levels, '*work out*' a timetable for buying and selling the stock of particular corporations, etc.

Remark. 'Fundamentalists' make their decisions by looking at the state of the 'economy at large', or some its sectors; the prospects of development are of particular interest for them; they build their analysis upon the assumption that the actions of market operators are 'rational'. Those tending to 'technical' analysis concentrate on the 'local' peculiarities of the market; they emphasize 'mass behavior' as a crucial factor.

As noted in [385; pp. 15–16], the 'fundamentalists' and the 'technicians' were the two major groups of financial market analysts in 1920–50. A third group emerged in the 1950s: 'quants', quantitative analysts, who are followers of L. Bachelier. This group is stronger biased towards 'fundamentalists' than towards the proponents of the 'technical' approach, who attach more weight to market *moods* than to the *rational* causes of investors' behavior.

2. We now return to the problem of building a basis under aspirations and attempts to forecast the 'future' price movement. Of course, we must start with the *analysis of empirical data*, look for explanations of several non-trivial features (for example, the cluster property) characteristic for the price movement, and clear up the probabilistic structure of prices as stochastic processes.

Let $H_n = \ln \dfrac{S_n}{S_0}$ be the logarithms of some (discounted) prices. We can represent the H_n as the sums $H_n = h_1 + \cdots + h_n$ with $h_k = \ln \dfrac{S_k}{S_{k-1}}$.

If the sequence (H_n) is a martingale with respect to some filtration $(\mathscr{F}_n)$, then the variables (h_n) form a *martingale-diference* $(\mathsf{E}(h_n \mid \mathscr{F}_{n-1}) = 0)$, and therefore the h_n (assuming that they are square-integrable) are *uncorrelated*, i.e., $\mathsf{E}h_{n+m}h_n = 0$, $m \geqslant 1$, $n \geqslant 1$.

As is well known, however, this *does not mean* that these variables are independent. It is not improbable that, say, h_{n+m}^2 and h_n^2 or $|h_{n+m}|$ and $|h_n|$ are positively *correlated*. The empirical analysis of much financial data shows that this is indeed the case (which does not contradict the assumption of the martingale nature of prices on an 'efficient' market!) It is also remarkable that this positive correlation, revealing itself in the behavior of the variables $(h_n)_{n\geqslant 1}$ as the phenomenon of their clustering into groups of large or small values, can be 'caught', understood, by means of several fairly simple models (e.g., *ARCH*, *GARCH*, models of stochastic volatility, etc.), which we discuss below in Chapter II. This indicates an opportunity (or at any rate, a chance) for a more non-trivial (and, in general, non-linear) prediction, e.g., of the *absolute* values $|h_{n+m}|$. There also arises a possibility of making more refined conclusions about the joint distributions of the sequence $(h_n)_{n\geqslant 1}$.

Taking the simplest assumption on the probabilistic nature of the h_n, let

$$h_n = \sigma_n \varepsilon_n,$$

where the ε_n are independent standard normally distributed variables such that $\varepsilon_n \sim \mathscr{N}(0,1)$ and the σ_n are some constants, the standard deviations of the h_n: $\sigma_n = +\sqrt{\mathsf{D}h_n}$.

However, this classical model of a Gaussian random walk has long been *considered inconsistent with the actual data*: the results of 'normality' tests show that the empirical distribution densities of h_n are more *extended*, more *leptokurtotic* around the mean value than it is characteristic for a normal distribution. The same analysis shows that the *tails* of the distributions of the h_n are more *heavy* than in the case of a normal distribution. (See Chapter IV for greater detail.)

In the finances literature one usually calls the coefficients σ_n in the relation $h_n = \sigma_n\varepsilon_n$ the *volatilities*. Here it is crucial that *the volatility is itself volatile*: $\sigma = (\sigma_n)$ is not only a function of time (a constant in the simplest case), but it is a *random* variable. (For greater detail on volatility see § 3a in Chapter IV.)

Mathematically, this assumption seems very appealing because it extends considerably the standard class of (*linear*) Gaussian model and brings into consideration (*non-linear*) *conditionally Gaussian* models, in which

$$h_n = \sigma_n \varepsilon_n,$$

where, as before, the (ε_n) are independent standard normally distributed variables ($\varepsilon_n \sim \mathscr{N}(0,1)$), but the $\sigma_n = \sigma_n(\omega)$ are $\mathscr{F}_{n-1}$-measurable non-negative *random* variables. In other terms,

$$\mathrm{Law}(h_n \mid \mathscr{F}_{n-1}) = \mathscr{N}(0, \sigma_n^2),$$

which means that $\mathrm{Law}(h_n)$ is a *suspension* (a *mixture*) of normal distributions averaged over some distribution of volatility ('random variance') σ_n^2.

It should be noted that, as is well known in mathematical statistics, 'mixtures' of distributions with fast decreasing tails can bring about distributions with heavy tails. Since such tails actually occur for empirical data (e.g., for many financial indexes), we may consider conditionally Gaussian schemes to be suitable probabilistic models.

Of course, turning to models with stochastic volatility (and, in particular, of the '$h_n = \sigma_n\varepsilon_n$' kind), it is crucial for their successful application to give an 'adequate' description of the evolution of the volatility (σ_n).

Already the first model, *ARCH* (Autoregressive Conditional Heteroskedasticity), proposed in the 1980s by R. F. Engle [140] and postulating that

$$\sigma_n^2 = \alpha_0 + \alpha_1 h_{n-1}^2 + \cdots + \alpha_q h_{n-q}^2,$$

enables one to 'grasp' the above-mentioned *cluster phenomenon* of the observable data revealed by the statistical analysis of financial time series.

The essence of this phenomenon is that large (small) values of h_n imply large (respectively, small) subsequent values (of uncertain sign, in general).

Of course, this becomes clearer once one makes a reference to the assumption of *ARCH*-theory about the dependence of σ_n^2 on the 'past' variables $h_{n-1}^2, \dots, h_{n-q}^2$. Thus, in conformity with observations, if we have a large value of $|h_n|$ in this model, then we should expect a large value of $|h_{n+1}|$. We point out, however, that the model makes *no* forecasts of the direction of the price movement, gives no information on the sign of h_{n+1}. (One practical implication is the following advice to operators on a stochastic market: considering a purchase of, say, an option, one better buy simultaneously a call and a put option; cf. § 4e in Chapter VI.)

Little surprise that the *ARCH* model gave birth to many similar models, constructed to 'catch on' to other empirical phenomena (besides the 'clusters'). The best known among these is the *GARCH* (Generalized *ARCH*) model developed by T. Bollerslev [48] (1986), in which

$$\sigma_n^2 = \alpha_0 + \alpha_1 h_{n-1}^2 + \cdots + \alpha_q h_{n-q}^2 + \beta_1 \sigma_{n-1}^2 + \cdots + \beta_p \sigma_{n-p}^2.$$

(One can judge the diversity of the generalizations of *ARCH* by the list of their names: *HARCH*, *EGARCH*, *AGARCH*, *NARCH*, *MARCH*, etc.)

One of the 'technical' advantages of the *GARCH* as compared with the *ARCH* models is as follows: whereas it takes one to bring into consideration large values of q to fit the *ARCH model* to real data, in *GARCH* models we may content ourselves with small q and p.

Having observed that the *ARCH* and the *GARCH* models explain the cluster phenomenon, we must also mention the existence of other empirical phenomena showing that the relation between prices and volatility is more subtle. Practitioners know very well that if the volatility is 'small', then prices tend to long-term rises or drops. On the other hand, if the volatility is 'large', then the rise or the decline of prices are visibly 'slowing'; the dynamics of prices tends to change its direction.

All in all, the inner (rather complex) structure of the financial market give one some hopes of a prediction of the price movement *as such*, or at least, of finding sufficiently reliable *boundaries* for this movement. ('Optimists' often preface such hopes with sententiae about market prices that 'must remember their past, after all', although this thesis is controversial and far from self-evident.)

In what follows (see Chapters II and III) we describe several probabilistic and statistical models describing the evolution of financial time series in greater detail. Now, we dwell on criticism towards and the revision of the assumption that the traders operate on an 'efficient' market.

3. As said in § 2a, it is built in the 'efficient market' concept that all participants are 'uniform' as regards their goals or the assimilation of new data, and they make decisions on a 'rational' basis. These postulates are objects of a certain criticism, however. Critics maintain that even if all participants have access to the entire information, they *do not* respond to it or interpret it *uniformly*, in a *homogeneous manner*; their goals can be utterly different; the periods when they are financially active can be various: from short ones for 'speculators' and 'technicians' to long periods of the central banks' involvement; the attitudes of the participants to various levels of riskiness can also be considerably different.

It has long been known that people are '*not linear*' in their decisions: they are less prone to take risks when they are anticipating profits and more so when they confront a prospect of losses [465]. The following dilemma formulated by A. Tversky of the Stanford University can serve an illustration of this quality (see [403]):

a) "Would you rather have $ 85 000 or an 85% chance of $ 100 000?"

b) "Would you rather lose $ 85 000 or run an 85% risk of losing $ 100 000?"

Most people would better take $ 85 000 and not try to get $ 100 000 in case a). In the second case b) the majority prefers a bet giving a chance (of 15%) to avoid losses.

One factor of prime importance in investment decisions is *time*. Given an opportunity to obtain either $ 5 000 today or $ 5 150 in a month, you would probably prefer money today. However, if the first opportunity will present itself in a year and the second in 13 months, then the majority would prefer 13 months. That is, investors can have different 'time horizons' depending on their specific aims, which, generally speaking, conforms badly with 'rational investment' models assuming (whether explicitly or not) that all investors have the same 'time horizon'.

We noted in § 2a.3 another inherent feature of the concept of efficient market: the participants must *correct* their decisions *instantaneously* once they get acquainted with new data. Everybody knows, however, that this is never the case: people need certain time (different for different persons) to ponder over new information and take one or another decision.

4. As G. Soros argues throughout his book [451], apart from 'equilibrium' or 'close to equilibrium' state, a market can also be 'far from equilibrium'. Addressing the idea that the operators on a market do not adequately perceive and interpret information, G. Soros writes: "We may distinguish between near-equilibrium conditions where certain corrective mechanisms prevent perceptions and reality from drifting too far apart, and far-from-equilibrium conditions where a reflexive double-feedback mechanism is at work and there is no tendency for perceptions and reality to come closer together without a significant change in the prevailing conditions, a change of regime. In the first case, classical economical theory applies and the divergence between perceptions and reality can be ignored as mere noise. In the second case, the theory of equilibrium becomes irrelevant and we are confronted with a one-directional historical process where changes in both perceptions and reality are irreversible. It is important to distinguish between these two different states of affairs because what is normal in one is abnormal in the other." [451, p. 6].

The views of R. B. Olsen [2], the founder of the Research Institute for Applied Economics ("Olsen & Associates", Zürich), fall in unison with these words: "There is a broad spectrum of market agents with different time horizons. These horizons range from one minute for short-term traders to several years for central bankers and corporations. The reactions of market agents to an outside event depend on the framework of his/her time horizon. Because time horizons are so different and vary by a factor of one million, economic agents take different decisions. This leads to a ripple effect, where the heterogeneous reactions of the agents are new events, requiring in turn secondary reactions by market participants."

It is apparently too early so far to speak of a rigorous mathematical theory treating financial markets as 'large complex systems' existing in the environment close to actually observed rather than in the classical 'equilibrium' one. We can define the current stage as that of 'data accumulating', 'model refining'. Here new methods of picking and storing information, its processing and assessing (which we discuss below, see Chapter IV) are of prime importance. All this provides necessary empirical data for the analysis of various conjectures pertaining to the operations of securities markets and for the correction of postulates underlying, say, the notion of an *efficient* market or assumptions about the distribution of prices, their dynamics, etc.

§ 2f. Analysis, Interpretation, and Revision of the Classical Concepts of Efficient Market. II

1. In this section we continue the discussion (or, rather, the description) of the assumptions underlying the concepts of *efficient* market and *rational* behavior of investors on it. We concentrate now on several new aspects, which we have not yet discussed.

The *concept of efficient market* was a remarkable achievement at its time, which has played (and plays still) a dominant role both in financial theory and the business of finance. It is therefore clear that pointing out its *strong* and *weak* points can help the understanding of neoclassical ideas (e.g., of a 'fractal' structure of the market) common nowadays in the economics and mathematics literature devoted to financial markets' properties and activities.

2. We have already seen that the concept of efficient market is based on the assumption that the 'today' prices are set with all the available information completely taken into account and that prices change only when this information is updated, when 'new', 'unexpected', 'unforeseen' data become available. Moreover, investors on such a market believe the established prices to be 'fair' because all the participants act in a 'uniform', 'collectively rational' way.

These assumptions naturally make the *random walk conjecture* (that the price is a sum of independent terms) and its generalization, the 'martingale conjecture' (which implies that the best forecast of the 'tomorrow' price is its current level) look quite plausible.

All this can be expressed by the phrase 'the market is a martingale', i.e., one plays *fairly* on an 'efficient market' (which is consistent with the traditional explanation of the word 'martingale'; see Chapter II, §§ 1b,c and, for example, [439; Chapter VII, § 1], where more details are given).

The reader must have already observed that, in fact, the concept of 'efficient market' (§ 2a) simply *postulates* that 'an efficient market is a martingale' (with respect to one or another 'information flow' and a certain probability measure). The corresponding arguments were not mathematically rigorous, but rather of the intuitive and descriptive nature.

In fact, this assertion ('the market is a martingale') has an irreproachable mathematical interpretation, provided that we start from the conjecture that (by definition) a 'fair', 'rationally' organized market is an *arbitrage-free market*. In other words, this is a market where no *riskless* profits are possible. (See § 2a in Chapter V for the formal definition.)

As we shall see below, one implication of this assumption of the absence of arbitrage is that there exists, generally speaking, an entire *spectrum* of ('martingale') measures such that the (discounted) prices are martingales with respect to these measures. This means more or less that the market can have an entire range of *stable* states, which, in its turn, is definitely related to the fact that market operators have various aims and different amounts of time to process and assimilate newly available information.

This presence of investors with different interests and potentials is a positive factor rather than a deficiency it may seem at the first glance.

The fact is that they reflect the 'diversification' of the market that ensures its liquidity, its capacity to transform assets promptly into means of payment (e.g., money), which is necessary for stability. We can support this thesis by the following well-known facts (see, e.g., [386; pp. 46–47]).

The day of President J. F. Kennedy's assassination (22.11.1963) markets immediately responded to the ensuing uncertainty: the long-term investors either suspended operations or turned to short-term investment. The exchanges were then closed for several days, and when they re-opened, the 'long-term' investors, guided by 'fundamental' information, returned to the market.

Although the complete picture of the well-known financial crash of October 19, 1978 in the USA is probably not yet understood, it is known, however, that just before that date 'long-term' investors were selling assets and switching to 'short-term' operations. The reasons lay in Federal Reserve System's tightening of monetary policy and the prospects of rises in property prices. As a result, the market was dominated by short-term activity, and in this environment 'technical' information (based, as it often is in instable times, on hearsay and speculations) came to the forefront.

In both examples, the long-term investors' 'run' from the market brought about a lack of liquidity and, therefore, instability. All this points to the fact that, for stability, a market must include operators with different 'investment horizons', there must be 'nonhomogeneity', 'fractionality' (or, as they put it, 'fractality') of the interests of the participants.

The fact that financial markets have the property of 'statistical fractality' (see the definition in Chapter III, § 5b) was explicitly pointed out by B. Mandelbrot as long ago as the 1960s. Later on, this question has attracted considerable attention, reinforced by newly discovered phenomena, such as the discovery of the statistical fractal structure in the currency cross rates or (short-term) variations of stock and bond prices.

As regards our understanding of what models of evolution of financial indicators are 'correct' and why 'stable' systems must have 'fractal' structure, it is worthwhile to compare *deterministic* and *statistical* fractal structures. In this connection, we provide an insight into several interconnected issues related to 'non-linear dynamic systems', 'chaos', and so on in Chapter II, § 4.

We have already mentioned that we adhere in this book to an irreproachable mathematical theory of the 'absence of arbitrage'. We must point out in this connection that none of such concepts as *efficiency*, *absence of arbitrage*, *fractality* can be a substitution for another. They supplement one another: for instance, many arbitrage-free models have fractal structure, while fractional processes can be martingales (with respect to some martingale measures; and then the corresponding market is arbitrage-free), but can also fail to be martingales, as does, e.g., a fractional Brownian motion (for the Hurst parameter H in $(0, \frac{1}{2}) \cup (\frac{1}{2}, 1)$).

3. In a similar way to § 2a, where we gave a descriptive definition of an 'efficient' market, it seems appropriate here, for a conclusion, to sum up the characteristic features of a market with fractal structure (a *fractal* market in the vocabulary of [386]) as follows:

1) at each instant of time prices on such a market are corrected by investors on the basis of the information relevant to their 'investment horizons'; investors do not necessarily respond to new information momentarily, but can react after it has been reaffirmed;
2) 'technical information' and 'technical analysis' are decisive for 'short' time horizons, while 'fundamental information' comes to the forefront with their extension;
3) the prices established are results of interaction between 'short-term' and 'long-term' investors;
4) the 'high-frequency' component in the prices is caused by the activities of 'short-term' investors, while the 'low-frequency', 'smooth' components reflect the activities of 'long-term' ones.
5) the market loses 'liquidity' and 'stability' if it sheds investors with various 'investment horizons' and loses its fractal character.

4. As follows from the above we mainly consider in this book models of ('efficient') markets *with no opportunities for arbitrage.*

Examples of such models that we thoroughly discuss in what follows are the *Bachelier model* (Chapter III, § 4b and Chapter VIII, § 1a), the *Black–Merton–Scholes model* (Chapter III, § 4b and Chapter VII, § 4c), and the *Cox–Ross–Rubinstein model* (Chapter II, § 1e and Chapter V, § 1d) based on a linear Brownian motion, a geometric Brownian motion, and a geometric random walk, respectively.

It seems fairly plausible after our qualitative description of 'fractal' markets that there may exist markets of this kind with *opportunities for arbitrage.*

Simplest examples of such models are (as recently shown by L. R. C. Rogers in "Arbitrage with fractional Brownian motion", Mathematical Finance, 7 (1997), 95–105) *modified* models of Bachelier and Black–Merton–Scholes, in which a *Brownian motion* is replaced by a *fractional Brownian motion* with $\mathbb{H} \in (0, \frac{1}{2}), \cup(\frac{1}{2}, 1)$. See also Chapter VII, § 2c below.

Remark. A fractional Brownian motion with $\mathbb{H} \in (0, \frac{1}{2}) \cup (\frac{1}{2}, 1)$ is *not a semimartingale*, therefore the corresponding martingale measures are nonexistent; see Chapter III, § 2c for details. This is an indirect indication (cf. the *First fundamental theorem* in Chapter V, § 2d) that there may (not necessarily!) exist *arbitrage* in such models.

3. Aims and Problems of Financial Theory, Engineering, and Actuarial Calculations

§ 3a. Role of Financial Theory and Financial Engineering. Financial Risks

1. Our discussion of financial structures and markets operating under uncertainty in the previous sections clearly points out RISKS as one of the central concepts both in *Financial theory* and *Insurance theory.*

This is a voluminous notion the contents of which is everybody's knowledge.

For instance, the *credit-lending risk* is related to the losses that the lender can incur if the borrower defaults.

Operational risk is related to errors possible, e.g., in payments.

Investment risk is a consequence of an unsatisfactory study of detail of an investment project, miscalculated investment decisions, possible changes of the economical or the political situation, and so on.

Financial mathematics and engineering (in its part relating to operations with securities) tackles mostly *market risks* brought about by the uncertainty of the development of market prices, interest rates, the unforeseeable nature of the actions and decisions of market operators, and so on.

The steep growth in attention paid by *Financial theory* to mathematical and engineering aspects (particularly ostensible during the last 2–3 decades) must have its explanation. The answer appears to be a simple one; it lies in the radical changes visible in the financial markets: the changes of their structure, greater volatility of prices, emergence of rather sophisticated financial instruments, new technologies used in price analysis, and so on.

All this imposes heavy weight on financial theory and puts forward new problems, whose solution requires a rather 'high-brow' mathematics.

2. At the first glance at financial mathematics and engineering, when one makes an emphasis on the 'game' aspect ('investor' versus 'market'), one can get an impression that it is the main aim of financial mathematics to work out recommendations and develop financial tools enabling the investor to '*gain over markets*'; at any rate, '*not to lose very much*'.

However, the role of financial theory (including financial mathematics) and financial engineering is much more prominent: they must help investors with the solution of a wide range of problems relating to *rational investment* by taking account of the *risks* unavoidable given the random character of the 'economic environment' and the resulting uncertainties of prices, trading volumes, or activities of market operators.

Financial mathematics and engineering are useful and important also in that their recommendations and the proposed financial instruments play a role of a 'regulator' in the *reallocation of funds*, which is necessary for better functioning of particular sectors and the whole of the economy.

The analysis of the securities market as a 'large' and 'complex' system calls for complicated, advanced mathematics, methods of data processing, numerical methods, and computing resources. Little wonder, therefore, that the finances literature draws on the most up-to-date results of stochastic calculus (on *Brownian motion*, *stochastic differential equations*, *local martingales*, *predictability*, ...), mathematical statistics (*bootstrap*, *jackknife*, ...), non-linear dynamics (*deterministic chaos*, *bifurcations*, *fractals*, ...), and, of course, it would be difficult to imagine modern financial markets without advanced computers and telecommunications.

3. Markowitz's theory was the first significant trial of the strength of probabilistic methods in the minimization of unavoidable commercial and financial risks by building an investment portfolio on a rational basis.

As regards 'risks', the fundamental economic postulates are as follows:

1) *'big' profits mean big risks*

and, on the way to these profits,

2) *the risks must be 'justifiable', 'reasonable', and 'calculated'.*

We can say in connection with 1) that 'big profits are a compensation for risks'; it seems appropriate to recall the proverb: "Nothing ventured, nothing gained". In connection with 2), we recall that 'to calculate' the risks in the framework of Markowitz's theory amounts to making up an 'efficient' portfolio of securities ensuring the maximum average profit that is possible under certain constraints on the size of the 'risk' measured in terms of the variance. We point out again that the idea of *diversification* in building an 'efficient' portfolio has gained solid ground in financial mathematics and became a starting point for the development of new strategies (e.g., hedging) and various financial instruments (options, futures, ...). It can be expressed by the well-known advice: "Don't put all your eggs in one basket."

Mean-variance analysis is based on a 'quadratic performance criterion' (it measures the risks in terms of the variance of return on a portfolio) and assumes that the market in question is 'static'. Modern analysts consider also more general *quality* *functions* and *utility functions* in their study of optimal investment, consumption, and allocation of resources. It is important to point out here a new aspect: the time *dynamics*, the need to take solutions 'successively', 'in stages'. Incidentally, decisions must be taken on the basis of generally accessible information and without anticipation. 'Statics' means that we are interested in the profits at time $N > 0$, while the portfolio is made up at time $n = 0$ (i.e., it is $\mathscr{F}_0$-measurable). Allowing for 'dynamics', we are tracing development in time, so that the returns at a particular instant $n \leqslant N$ are defined by the portfolio built on the basis of information obtained by time $n-1$ (inclusive; i.e., the components of the portfolio are $\mathscr{F}_{n-1}$-measurable), and so on.

Clearly, the incorporation of the *dynamical* aspect in the corresponding problems of financial mathematics calls for concepts and methods of *optimal stochastic control, stochastic optimization, dynamical programming, statistical sequential analysis*, and so on.

In our discussion of the risks relating to 'uncertainties' of all kinds we cannot bypass the issues of 'risk' *insurance*, which are the subject of *Insurance theory* or, somewhat broader, the *Theory of actuarial calculations.*

In the next section we describe briefly the formation of the insurance business providing a mechanism of compensation for financial losses, the history and the role of the 'actuarial trade'. We shall not use this information in what follows; nevertheless, we think it appropriate here, since it can give one an idea of a realm that is closely related to finance. Moreover, it becomes increasingly clear that financial and actuarial mathematics share their ideas and their method.

§ 3b. Insurance: a Social Mechanism of Compensation for Financial Losses

1. An *actuarius* in ancient Rome was a person making records in the Senate's Acta Publica or an army officer processing bills and supervising military supplies.

In its English version, the meaning of the word '*actuary*' has undergone several changes. First, it was a registrar, then a secretary or a counselor at a joint stock company. In due course, the trade of an *actuary* became associated with one who performed *mathematical calculations relating to life expectancies*, which are fundamental for pricing life insurance contracts, annuities, etc.

In modern usage an actuary is an expert in the *mathematics of insurance.* They are often called *social mathematicians*, for they keep key positions it working out the strategy not only of insurance companies, but also in pension and other kinds of funds; government actuaries are in charge of various issues of state insurance, pensions, and other entitlement schemes.

Insurance (*assurance*) is a social mechanism of compensation of individuals and organizations for financial losses incurred by some or other unfavorable circumstances.

The destination of *insurance* is to put *certainty* in place of the *uncertainties* of financial evaluations that are due to possible losses in the future.

Insurance can be defined as a social instrument that enables individuals or organizations to pay in advance to reduce (or rule out) certain share of the risk of losses.

Mankind was fairly quick at comprehending that the most efficient way to lessen the losses from uncertainties is a cooperation mechanism distributing the cost of the losses born by an individual over all participants. Individuals were smart enough to realize that it is difficult to forecast the timing, location, and the scope of events that can have implications unfavorable for their economic well-being. Insurance provided an instrument that could help a person to lessen, 'tame' the impact of the uncertain, the unforeseen, and the unknown.

It would be wrong to reduce insurance (and the relevant mathematics) to, say, property and life insurance. It must be treated on a broader scale, as *risk* insurance, which means the inclusion of, e.g., betting in securities markets or investments (foreign as well as domestic). We shall see below that the current state of *insurance theory* (in the standard sense of the term '*insurance*') is distinguished by extremely close interrelations with *financial theory*. One spectacular example here are futures reinsurance contracts floated at the Chicago Board of Trade in December, 1992.

Not all kinds of 'uncertainty', not all risks are subject to insurance. One usually uses here the following vocabulary and classification, which is helpful in outlining the realm of insurance.

2. All 'uncertainties' are usually classified with one of the two groups: *pure* uncertainties and *speculative* uncertainties.

A *speculative uncertainity* is the uncertainty of possible financial *gains* or *losses* (generally speaking, one does not sell insurance against such uncertainties).

A *pure uncertainity* means that there can be only *losses* (for example, from fire); many of these can be insured against.

Often, one uses the same word '*risk*' in the discussion of uncertainties of both types. (Note that, in insurance theory, one often identifies risk and pure uncertainty; on the other hand, a popular finances journal 'Risk' is mostly devoted to speculative uncertainties.)

The sources of risks and of the losses incurred by them are *accidents*, which are also classified in insurance with one of the two groups: *physical* accidents and *moral* accidents.

Moral accidents are about dishonesty, unfairness, negligence, vile intentions, etc.

With *physical* accidents one ranks, for instance, earthquakes, economic cycles, weather, various natural phenomena, and so on.

There are different ways to 'combat' risks:

1) One can deliberately avert risks, avoid them by rational behavior, decisions, and actions.

2) One can reduce risks by transferring possible losses to other persons or institutions.
3) One can try to reduce risks by making forecasts. Statistical methods are an important weapon of an actuary in forecasting possible losses. *Forecast-making* and *accumulation of funds* are decisive for a sustained successful insurance business.

Although insurance is a logical and in many respects remarkable 'remedy' against risks, not all uncertainties and accompanying financial losses can be covered by it. Insurance risks must satisfy certain conditions; namely,

1) there must exist a sufficiently large group of 'uniformly' insurable participants with 'characteristics' stable in time;
2) insurance cases should not affect many participants at one time;
3) these cases and the range of losses should not be consequences of deliberate actions of those insured: they must be accidental; damages are paid only if the cause of losses can be determined;
4) the potential losses ensuing from the risks in question must be easily identifiable ('difficult to counterfeit');
5) the potential losses must be sufficiently large ('there is no point in the insurance against small, easily recoupable losses');
6) the probability of losses must be sufficiently small (insurance cases must be rare), so as to make the insurance economically affordable;
7) statistical data must be accessible and can serve a basis for the calculations of the probabilities of losses ('representativeness and the feasibility of statistical estimation').

These and similar requirements are standard, they make up a minimum dose of conditions enabling insurance.

Several types (kinds) of insurance are known. Their diversity is easier to sort out if one considers

(a) classes of insured objects
(b) unfavorable factors that can lead to insurance risks
(c) modes of payment (premiums, benefits)
(d) types of insurance (social versus commercial).

Group (a) includes various classes of objects, for instance, life, health, property, cars, ships, stocks of merchandise, and so on.

Group (b) includes the above-mentioned *physical* and *moral* accidents.

The modes of payment (of premiums and benefits) in group (c) vary with insurance policies; life insurance is subject to most diversification.

Different types of insurance (d)—either free (commercial) or obligatory (social)—are based on similar principles but differ in their philosophy and ways of organization.

3. The content of *insurance*, its aims and objectives are difficult to understand without a knowledge of both its structure and history.

We can say that, in a sense, insurance is as old as the mankind itself.

The most ancient forms of insurance known are the so-called *bottomry* and *respondencia* contracts (for sea carriages) discovered in Babylonian texts of 4–3 millennia BC.

Bottomry, which was essentially a *pledge* contract, was formally a loan (e.g., in the form of goods that must be delivered to some other place) made to the ship owner who could give not only the ship or another 'tangible' property as collateral, but also his own life and the lives of his dependents (which means that they could be reduced to slavery). In the case of a *respondencia* the collateral was always secured by goods.

The Babylonians developed also a system of insurance contracts in which the supplier of goods, in case of a risky carriage, agreed to write off the loan if the carrier were robbed, kidnapped for ransom, and so on.

The Hammurabe code (2100 BC) legalized this practice. It also provided for damages and compensations (by the government) to individuals who had suffered from fire, robbery, violence, and so on.

Later on, the practices of making such contracts were adopted (through the Phoenicians) by the Greeks, the Romans, and the Indians. It was mentioned in early Roman codes and Byzantine laws; it is reflected by the modern insurance legislation.

The origins of life insurance can be traced back to Greek *thiasoi* and *eranoi* or to Roman *collegia*, from 600 BC to the end of the Roman empire (which is traditionally put at 476 AC). Originally, the *collegia* (guilds) were religious associations, but with time, they took upon themselves more utilitarian functions, e.g., funeral services. Through their funeral collegia the Romans paid for funerals; they also developed some rudimentary forms of life insurance. The lawyer Ulpian (around 220 AD) made up mortality tables (rather crude ones).

The practice of premium insurance has apparently started in Italian city republics (Venice, Pisa, Florence, Genoa), around 1250. The first precisely dated insurance contract was made in Genoa, in 1347. The first 'proper' life insurance contract, which related to pregnant women and slaves, was also signed there (1430).

Annuity contracts were known to the Romans as long ago as the 1 century AD. (These contracts stipulate that the insurance company is paid some money and it pays back periodic benefits during a certain period of time or till the end of life. This is in a certain sense a converse of life insurance contracts, where the insured pays premium on a regular basis and obtains certain sum in bulk.)

The text of Rome's Palcidian law included also a table of the mean values of life expectancy necessary in the calculations of annuity payments and the like. In 225, Ulpian made up more precise tables that were in use in Tuscany through the 18th century.

Three hundred years ago, in 1693, Edmund Galley (whose name is associated with a well-known comet) improved Ulpian's actuarial tables by assuming that the mortality rate in a fixed group of people is a *deterministic* function of time. In 1756, Joseph Podson extended and corrected Galley's tables, which made it possible to draw up a year-to-year 'scale of premiums'.

With the growth of cities and trade in the medieval Europe, the guilds extended their practice of helping their members in the case of a fire, a shipwrecks, an attack of pirates, etc.; provided them aid in funerals, in the case of disability, and so on. Following the step of the 14th-century Genoa, sea insurance contracts spread over virtually all European marine nations.

The modern history of *sea insurance* is primarily 'the history of Lloyd's', a corporation of insurers and insurance brokers founded in 1689, on the basis of Edward Lloyd's coffee shop, where ship owners, salesmen, and marine insurers used to gather and make deals. It was incorporated by an act of British Parliament in 1871. Established for the aims of marine insurance, Lloyd's presently operates almost all kinds of risks.

Since 1974, Lloyd's publishes a *daily* Gazette providing the details of sea travel (and, nowadays, also of air flights) and information about accidents, natural disasters, shipwrecks, etc.

Lloyd's also publishes *weekly* accounts of the ships loaded in British and continental ports and the dates of the end of shipment. General information on the insurance market can also be found there.

In 1760, Lloyd's gave birth to a society for the inspection and the classification of all sea-going ships of at least 100 tons' capacity. Lloyd's surveyors examine and classify vessels according to the state of their hulls, engines, safety facilities, and so on. This society also provides technical advice.

The *yearly* "Lloyd's Register of British and Foreign Shipping" provides the data necessary for Lloyd's underwriters to negotiate marine insurance contracts even if the ship in question is thousands miles away.

The great London fire of 1666 gave an impetus to the development of *fire insurance*. (The first fire insurance company was founded in 1667.) Insurance against break-downs of steam generators was launched in England in 1854; one can insure employers' liability since 1880, transport liability since 1895, and against collisions of vehicles since 1899.

The first fire insurance companies in the United States emerged in New York in 1787 and in Philadelphia in 1794. Virtually from their first days, these companies began tackling also the issues of fire prevention and extinguishing. The first life insurance company in the United States was founded in 1759.

The big New York fire of 1835 brought to the foreground the necessity to have reserves to pay unexpected huge ('catastrophic') damages. The great Chicago fire of 1871 showed that fire-insurance payments in modern, 'densely built' cities can be immense. The first examples of *reinsurance* (when the losses are covered by several firms) related just to fire damages of *catastrophic* size; nowadays, this is standard practice in various kinds of insurance. As regards other first American examples of modern-type insurance, we can point out the following ones: casualty insurance (1864), liability insurance (1880), insurance against burglary (1885), and so on.

The first Russian joint-stock company devoted to fire insurance was set up in Siberia, in 1827, and the first Russian firm insuring life and incomes (The Russian Society for Capital and Income Insurance) was founded in 1835.

In the 20th century, we are eyewitnesses to the extension of the sphere of insurance related to 'inland marine' and covering a great variety of transported items including tourist baggages, express mail, parcels, means of transportation, transit goods, and even bridges, tunnels, and the like.

These days one can get insured against any conceivable insurable risk. Such firms as Lloyd's insure dancers' legs and pianists' fingers, outdoor parties against the consequences of bad weather, and so on.

Since the end of 19th century one could see a growing tendency of government involvement in insurance, in particular, in domains related to the protection of employees against illness, disability (temporal or perpetual), old age insurance, and unemployment insurance. Germany was apparently a pioneer in the so-called social security (laws of 1883–89).

In the middle of this century, a tendency to mergers and consolidations in insurance business became evident. For instance, there was a consolidation of American life and property insurance companies in 1955–65. A new form of 'merger' became widespread: a holding company that owns shares of other firms, not only insurance companies, but also with businesses in banking, computer services, and so on. The strong point of these firms is their diversification, the variety of their potentialities. The burden of taxes imposed on an insurance firm is lighter if it is a part of a holding company. A holding company can get involved with foreign stock, which is sometimes impossible for insurance firms. This gives insurance companies greater leverage: they have access to larger resources while investing less themselves.

4. It is impossible to consider now the insurance practices and theory separated from the practices and the theory of finance and investment in securities.

It is as good as established that derivative instruments (futures, options, swaps, warrants, straddles, spreads, and so on) will be in the focus of the future global financial system. Certainly, the pricing methods used in finance will penetrate the actuarial science ever more deeply. In this connection, it is worthwhile and reasonable not to separate actuarial and financial problems related, one way or another, to various forms of risks, but to tackle them in one package. In favor of this opinion counts also the following division of the history of insurance mathematics into periods, proposed by H. Bühlmann, the well-known Swiss expert in actuarial science.

The *first period* ('insurance of the first kind') dates back to E. Galley who, as already mentioned, drew up insurance tables (1693) based on the assumption that the mortality rate in a fixed group of people is a deterministic function of time.

The *second period* ('insurance of the second kind') is connected with the introduction of probabilistic ideas and the methods of probability theory and statistics into life insurance and other types of insurance.

The *third period* ('insurance of the third kind') can be characterized by the use on a large scale of financial instruments and financial engineering to reduce insurance risks.

Mathematics of the insurance of the second kind is based on the Law of large numbers, the Central limit theorem, and Poisson-type processes. The theory of insurance of the third kind is more sophisticated: it requires the knowledge of stochastic calculus, stochastic differential equations, martingales and related concepts, as well as new methods, such as bootstrap, resampling, simulation, neural networking, and so on.

A good illustration of the above idea of the reasonableness and advantageousness of an integrated approach to actuarial and financial problems of securities markets is the efficiency of purely financial, 'optional' method in actuarial calculations related to reinsurance of 'catastrophic events' [87].

In the global market, one says that an accident is 'catastrophic' if 'the damages exceed $ 5m and a large number of insurers and insured are affected' (an extract from "Property Claims Services", 1993). The actual size of the damages in some 'catastrophic cases' is such that no single insurer is able (or willing) to insure against such accidents. This explains the fact that insurance against such events becomes not merely a joint but an international undertaking.

Here are several examples of damages incurred by 'catastrophic' accidents.

From 1970 through 1993 there occurred on average 34 catastrophes each year, with annual losses amounting to $ 2.5 billion. In most cases, a 'catastrophic' event brought damages of less than $ 250m. However, the losses from the hurricane 'Andrew' (August 1992) are estimated at $ 13.7 billion, of which only about $ 3 billion of damages were reimbursed by insurance.

In view of large sums payable in 'catastrophic' cases: hurricanes, earthquakes, floods, the CBOT (Chicago Board of Trade) started in December of 1992 the *futures* trade in catastrophe insurance contracts as an *alternative* to catastrophe reinsurance. These contracts are easy to sell ('liquid'), anonymous; their operational costs are low, and the supervision of all transactions by a clearing house gives one the confidence important in this kind of deals. Moreover, the pricing of such futures turned out in effect to reduce to the calculation of the rational price of *arithmetic Asian call options* (see the definition in § 1c). See [87] for greater detail.

§ 3c. A Classical Example of Actuarial Calculations: the Lundberg–Cramér Theorem

1. Oddly, around the same time when L. Bachelier introduced a *Brownian motion* to describe share prices and laid in this way the basis of *stochastic financial mathematics*, Ph. Lundberg in Uppsala (Sweden) published his thesis "Approximerad framsröllning av sannolikhetsfunctionen. Atersförsäkring av kollektivrisker" (1903), which became the cornerstone of the *theory of insurance* (of the second kind) and where he developed systematically *Poisson processes*, which are, together with

Brownian motions, central objects of the general theory of stochastic processes.

In 1929, on initiative of several Swedish insurers, the Stockholm University established a chair in actuarial mathematics. Its first holder was H. Cramér and this marked the starting point in the activities of the '*Stockholm group*', renowned for their results both in actuarial mathematics and general probability theory, statistics, and the theory of stochastic processes.

We now formulate the classical result of the theory of actuarial calculations, the *Lundberg-Cramér fundamental theorem of risk theory.*

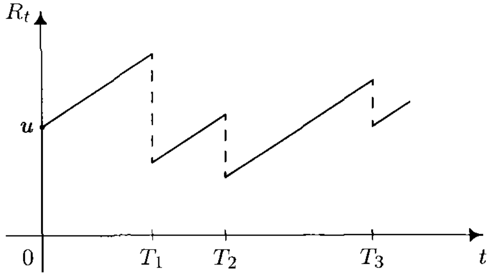


FIGURE 10. Risk process R_t

We define the risk process (see Fig. 10) of, say, an insurance business as follows:

$$R_t = u + ct - \sum_{k=1}^{N_t} \xi_k,$$

where

- u is the initial capital,
- c is the rate of the collection of premiums,
- (ξ_k) is a sequence of independent, identically distributed random variables with distribution $F(x) = \mathsf{P}\{\xi_1 \leqslant x\}$, $F(0) = 0$, and expectation $\mu = \mathsf{E}\xi_1 < \infty$,
- $N = (N_t)_{t \geqslant 0}$ is a Poisson process:

$$N_t = \sum_k I(T_k \leqslant t),$$

where insurance claims are submitted at the instants $T_1, T_2, \dots$ and it is assumed that $(T_{k+1} - T_k)_{k \geqslant 1}$ are independent variables having exponential distribution with parameter λ:

$$\mathsf{P}\{T_{k+1} - T_k \geqslant t\} = e^{-\lambda t}.$$

Clearly,

$$\mathsf{E}R_t = u + (c - \lambda\mu)t = u + \rho\lambda\mu t,$$

where the relative safety loading $\rho = c/(\lambda\mu) - 1$ is assumed to be *positive* (the condition of *positive net profit*).

One of the first natural problems arising in connection with this model is the calculation of the *probability* $\mathsf{P}(\tau < \infty)$ of a *ruin* in general or the *probability* $\mathsf{P}(\tau \leqslant t)$ of a *ruin before time* t, where

$$\tau = \inf\{t\colon R_t \leqslant 0\}.$$

THEOREM (Lundberg–Cramér). *Assume that there exists a constant $R > 0$ such that*

$$\int_0^\infty e^{Rx}(1 - F(x))dx = \frac{c}{\lambda}.$$

Then

$$\mathsf{P}(\tau < \infty) \leqslant e^{-Ru},$$

where u is the initial capital.

The assumptions in the *Lundberg–Cramér* model can be weakened, while the model itself can be made more complicated. For instance, we can assume that the risk process has the following form:

$$R_t = u + (ct + \sigma B_t) - \sum_{k=1}^{N_t} \xi_k,$$

where (B_t) is a Brownian motion and (N_t) is a Cox process (that is, a 'counting process' with random intensity; see, e.g., [250]).

In conclusion, we briefly dwell on the question of the nature of the distributions $F = F(x)$ of insurance benefits. As a matter of a pure convention, one often classifies insurance cases leading to payments with one of the following three types:

- '*normal*',
- '*extremal*',
- '*catastrophic*'.

To describe 'normal' events, one uses distributions with rapidly decreasing tails (e.g., an exponential distribution satisfying the condition $1 - F(x) \sim e^{-x}$ as $x \to \infty$).

One describes 'extremal' events by distributions $F = F(x)$ with heavy tails; e.g., $1 - F(x) \sim x^{-\alpha}$, $\alpha > 0$, as $x \to \infty$ (Pareto-type distributions) or

$$1 - F(x) = \exp\Big\{-\Big(\frac{x-\mu}{\sigma}\Big)^p\Big\}, \qquad x > \mu,$$

with $p \in (0,1)$ (a Weibull distribution).

We note that the Lundberg–Cramér theorem relates to the 'normal' case and cannot be applied to large payments. (One cannot even define the 'Lundberg coefficient' R in the latter case; for the proof of the Lundberg–Cramér theorem see, e.g., [439].)

Chapter II. Stochastic Models. Discrete Time

1. Necessary Probabilistic Concepts and Several Models of the Dynamics of Market Prices

§ 1a. Uncertainty and Irregularity in the Behavior of Prices. Their Description and Representation in Probabilistic Terms

1. Assume that we measure time in days $n = 0, 1, 2, \dots$, and let

$$S = (S_n)_{n \geqslant 0}$$

be the market price of a share, or the exchange rate of two currencies, or another financial index (of unlimited 'life span', by contrast to, say, bond prices). An empirical study of the S_n, $n \geqslant 0$, shows that they vary in a highly irregular way; they fluctuate as if (using the words of M. Kendall; see Chapter I, § 2a) " ... *the Demon of Chance drew a random number ... and added it to the current price to determine the next ... price*".

Beyond all doubts, L. Bachelier was the first to describe the prices $(S_n)_{n \geqslant 0}$ using the concepts and the methods of *probability theory*, which provides a framework for the study of empirical phenomena featured by both statistical uncertainty and stability of statistical frequencies.

Taking the probabilistic approach and using *A. N. Kolmogorov's axiomatics of probability theory*, which is generally accepted now, we shall assume that all our considerations are carried out with respect to some *probability space*

$$(\Omega, \mathscr{F}, \mathsf{P}),$$

where

- Ω is the space of *elementary events* ω ('*market situations*', in the present context);
- $\mathscr{F}$ is some σ-algebra of *subsets* of Ω (the set of '*observable market events*');
- P is a *probability* (or *probability measure*) on $\mathscr{F}$.

As pointed out in Chapter I, § 1a, *time* and *dynamics* are integral parts of the financial theory. For that reason, it seems worthwhile to define our probability space $(\Omega, \mathscr{F}, \mathsf{P})$ more specifically, by assuming that we have a *flow* $\mathbb{F} = (\mathscr{F}_n)_{n\geqslant 0}$ of *σ-algebras* such that

$$\mathscr{F}_0 \subseteq \mathscr{F}_1 \subseteq \cdots \subseteq \mathscr{F}_n \subseteq \cdots \subseteq \mathscr{F}.$$

The point in introducing this flow of nondecreasing σ-subalgebras of $\mathscr{F}$, which is also called a *filtration*, becomes clear once one have accepted the following interpretation:

$\mathscr{F}_n$ is the set of events *observable through time n.*

We can express it otherwise by saying that $\mathscr{F}_n$ is the '*information*' on the market situation that is available to an observer up to time n inclusive. (In the framework of the concept of an 'efficient' market this can be, e.g., one of the three σ-algebras $\mathscr{F}_n^1$, $\mathscr{F}_n^2$, and $\mathscr{F}_n^3$; see Chapter I, § 2a.)

Thus, we assume that our underlying probabilistic model is a *filtered probability space*

$$\big(\Omega, \mathscr{F}, (\mathscr{F}_n)_{n\geqslant 0}, \mathsf{P}\big),$$

which is also called a *stochastic basis.*

In many cases it seems reasonable to *generalize* the concept of a stochastic basis by assuming that, instead of a *single* probability measure P, we have an entire *family* $\mathscr{P} = \{\mathsf{P}\}$ of probability measures. (The reasons are that it is often difficult to single out a particular measure P.) Using the vocabulary of statistical decision theory we can call the collection $(\Omega, \mathscr{F}, (\mathscr{F}_n)_{n\geqslant 0}, \mathscr{P})$ a *filtered stochastic (statistical) experiment.*

2. Regarding $\mathscr{F}_n$ as the information that had been accessible to observation through time n, it is natural to assume that

$$S_n \text{ is } \mathscr{F}_n\text{-measurable},$$

or that (using a more descriptive language), prices are formed on the basis of the developments observable on the market up to time n (inclusive).

Bearing in mind our interpretation of S_n as the 'price' (of some stock, say) at time n, we shall assume that $S_n > 0$, $n \geqslant 0$.

We now present the two most common methods for the description of the prices $S = (S_n)_{n\geqslant 0}$.

The first method, which is similar to the formula for *compound interest* (see Chapter I, § 1b) uses the representation

$$S_n = S_0 e^{H_n}, \tag{1}$$

where $H_n = h_0 + h_1 + \cdots + h_n$ with $h_0 = 0$ and the random variables $h_n = h_n(\omega)$, $n \geqslant 0$, are $\mathscr{F}_n$-measurable. Hence

$$H_n = \ln \frac{S_n}{S_0} \tag{2}$$

and the 'logarithmic returns' can be evaluated by the formula

$$h_n = \ln \frac{S_n}{S_{n-1}} = \ln\left(1 + \frac{\Delta S_n}{S_{n-1}}\right), \tag{3}$$

where $\Delta S_n = S_n - S_{n-1}$.

We now set

$$\widehat{h}_n = \frac{\Delta S_n}{S_{n-1}} \quad \text{and} \quad \widehat{H}_n = \sum_{1\leqslant k\leqslant n} \widehat{h}_k. \tag{4}$$

Then we can rewrite (1) as

$$S_n = S_0 \prod_{1\leqslant k\leqslant n} (1 + \widehat{h}_k), \tag{5}$$

or, equivalently, as

$$S_n = S_0 \prod_{1\leqslant k\leqslant n} (1 + \Delta\widehat{H}_k) = S_0 e^{\widehat{H}_n} \prod_{1\leqslant k\leqslant n} (1 + \Delta\widehat{H}_k) e^{-\Delta\widehat{H}_k}. \tag{6}$$

The representation (5) is just the second method of the description of prices. It is equivalent to the 'simple interest' formula.

Let $\mathscr{E}(\widehat{H})_n$ be the expression on the right-hand side of (6):

$$\mathscr{E}(\widehat{H})_n = e^{\widehat{H}_n} \prod_{1\leqslant k\leqslant n} (1 + \Delta\widehat{H}_k) e^{-\Delta\widehat{H}_k}. \tag{7}$$

We call the stochastic sequence

$$\mathscr{E}(\widehat{H}) = \big(\mathscr{E}(\widehat{H})_n\big)_{n\geqslant 0}, \qquad \mathscr{E}(\widehat{H})_0 = 1,$$

defined by this expression the *stochastic exponential* generated by the variables $\widehat{H} = (\widehat{H}_n)_{n\geqslant 0}$, $\widehat{H}_0 = 1$, or the *Doléans exponential*.

Thus, we can say that the first method for the description of prices uses the *usual* exponential

$$S_n = S_0 e^{H_n},$$

while the second method involves the *stochastic* exponential:

$$S_n = S_0 \mathscr{E}(\widehat{H})_n, \tag{8}$$

where

$$\widehat{H}_n = \sum_{1\leqslant k\leqslant n} (e^{\Delta H_k} - 1),$$

which is equivalent to the following representation:

$$\widehat{H}_n = H_n + \sum_{1 \leqslant k \leqslant n} (e^{\Delta H_k} - \Delta H_k - 1). \tag{9}$$

It is also clear from (3) and (4) that

$$H_n = \sum_{1 \leqslant k \leqslant n} \ln(1 + \Delta \widehat{H}_k), \tag{10}$$

where $\widehat{h}_n = \Delta \widehat{H}_n > -1$ by our assumption that $S_n > 0$.

It is worth noting that the stochastic exponential satisfies the stochastic *difference equation*

$$\Delta \mathscr{E}(\widehat{H})_n = \mathscr{E}(\widehat{H})_{n-1} \Delta \widehat{H}_n, \tag{11}$$

which is an immediate consequence of (7).

Remark 1. Formulas (1) and (8) are related to discrete time $n = 0, 1, \dots$. In the case when the prices $S = (S_t)_{t \geqslant 0}$ evolve in continuous time $t \geqslant 0$ one usually assumes that the processes $H = (H_t)_{t \geqslant 0}$ and $\widehat{H} = (\widehat{H}_t)_{t \geqslant 0}$ are *semimartingales* (see Chapter III, § 5b). Then *by the Itô formula* (Chapter III, § 5c; see also [250; Chapter I, § 4e]) we obtain

$$e^{H_t} = \mathscr{E}(\widehat{H})_t,$$

where

$$\widehat{H}_t = H_t + \frac{1}{2}\langle H^c \rangle_t + \sum_{0 < s \leqslant t} (e^{\Delta H_s} - 1 - \Delta H_s) \tag{12}$$

and $\mathscr{E}(\widehat{H}) = \left(\mathscr{E}(\widehat{H})_t\right)_{t \geqslant 0}$ is the *stochastic exponential*:

$$\mathscr{E}(\widehat{H})_t = e^{\widehat{H}_t - \frac{1}{2}\langle \widehat{H}^c \rangle_t} \prod_{0 < s \leqslant t} (1 + \Delta \widehat{H}_s) e^{-\Delta \widehat{H}_s}, \tag{13}$$

satisfying (cf. (11)) the linear stochastic *differential equation*

$$d\mathscr{E}(\widehat{H})_t = \mathscr{E}(\widehat{H})_{t-} d\widehat{H}_t. \tag{14}$$

(In (12) and (13) we denote by $\langle H^c \rangle$ and $\langle \widehat{H}^c \rangle$ the quadratic characteristics of the continuous martingale components of the semimartingales H and $\widehat{H}$; see Chapter III, § 5a. As regards stochastic differential equations in the case when $\widehat{H}$ is a Brownian motion, see Chapter III, § 3e.)

Thus, we have the following representations that are the counterparts of (1) and (8), respectively, in the continuous-time case:

$$S_t = S_0 e^{H_t} \tag{15}$$

and

$$S_t = S_0 \mathscr{E}(\widehat{H})_t, \tag{16}$$

where the process $\widehat{H} = (\widehat{H}_t)_{t \geqslant 0}$ is related to $H = (H_t)_{t \geqslant 0}$ by (12). The series in (12) is absolutely convergent because, with probability one, a semimartingale makes only finitely many 'large' jumps $\left(|\Delta H_s| > \frac{1}{2}\right)$ on each interval $[0, t]$ and $\sum_{0<s\leqslant t} |\Delta H_s|^2 < \infty$. (See Remark 3 in Chapter III, § 5b.)

Remark 2. By (3) and (4),

$$h_n = \ln(1 + \widehat{h}_n) \tag{17}$$

and

$$\widehat{h}_n = e^{h_n} - 1. \tag{18}$$

Clearly, we have

$$\widehat{h}_n \approx h_n \tag{19}$$

for small values of h_n; moreover,

$$\widehat{h}_n - h_n = \frac{1}{2} h_n^2 + \frac{1}{6} h_n^3 + \cdots. \tag{20}$$

3. We now dwell on the problem of the description of the probability distributions for the sequences $S = (S_n)_{n \geqslant 0}$ and $H = (H_n)_{n \geqslant 0}$.

Taking the viewpoint of classical probability theory and the well-developed 'statistics of normal distribution' it would be nice if $H = (H_n)_{n \geqslant 0}$ could be a *Gaussian* (*normally distributed*) sequence. If we set

$$H_n = h_1 + \cdots + h_n, \qquad n \geqslant 1, \tag{21}$$

then the properties of such a sequence are completely determined by the *two-dimensional* distributions of the sequence $h = (h_n)_{n \geqslant 1}$, which can be characterized by the expectations

$$\mu_n \equiv \mathsf{E} h_n, \qquad n \geqslant 1,$$

and the covariances

$$\mathsf{Cov}(h_n, h_m) \equiv \mathsf{E}h_n h_m - \mathsf{E}h_n \mathsf{E}h_m, \qquad m, n \geqslant 1.$$

This assumption of normality can considerably facilitate the solution of many problems relating to the properties of *distributions*. For instance, the '*Theorem on normal correlation*' (see, for example, [303; Chapter 13]) delivers an *explicit* formula for the conditional expectation $\widetilde{h}_{n+1} = \mathsf{E}(h_{n+1} \mid h_1, \dots, h_n)$, which is the *optimal* (in the mean-square sense) *estimator* for h_{n+1} in terms of $h_1, \dots, h_n$. Namely,

$$\widetilde{h}_{n+1} = \mu_{n+1} + \sum_{i=1}^{n} a_i(h_i - \mu_i), \tag{22}$$

where the coefficients a_i are evaluated in terms of the covariance matrix (see [303; Chapter 13] and also [439; Chapter II, § 13]).

Formula (22) becomes particularly simple if $h_1, \dots, h_n$ are *independent*. In this case

$$a_i = \frac{\mathsf{Cov}(h_{n+1}, h_i)}{\mathsf{D}h_i},$$

and we obtain the estimator

$$\widetilde{h}_{n+1} = \mathsf{E}h_{n+1} + \sum_{i=1}^{n} \frac{\mathsf{Cov}(h_{n+1}, h_i)}{\mathsf{D}h_i}(h_i - \mathsf{E}h_i). \tag{23}$$

The *estimation error*

$$\Delta_{n+1} = \mathsf{E}(\widetilde{h}_{n+1} - h_{n+1})^2$$

can be expressed by the formula

$$\Delta_{n+1} = \mathsf{D}h_{n+1} - \sum_{i=1}^{n} \frac{\mathsf{Cov}^2(h_{n+1}, h_i)}{\mathsf{D}h_i}. \tag{24}$$

We note that if

$$\varphi_{(\mu,\sigma^2)}(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

is the density of the normal distribution with parameters (μ, σ^2), then (see Fig. 11)

$$\int_{\mu-\sigma}^{\mu+\sigma} \varphi_{(\mu,\sigma^2)}(x)\, dx = 0.6827\dots.$$

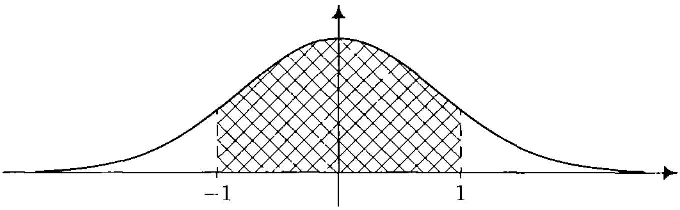


FIGURE 11. Graph of the density $\varphi_{(0,1)}(x)$ of the standard normal distribution. The area of the shaded part is approximately 0.6827

In the same way,

$$\int_{\mu-1.65\sigma}^{\mu+1.65\sigma} \varphi_{(\mu,\sigma^2)}(x)\,dx \approx 0.90. \tag{25}$$

By the Gaussian property we obtain

$$h_{n+1} - \widetilde{h}_{n+1} \sim \mathscr{N}(0, \Delta_{n+1}),$$

therefore

$$\mathsf{P}\{|h_{n+1} - \widetilde{h}_{n+1}| \leqslant 1.65\sqrt{\Delta_{n+1}}\,\} \approx 0.90$$

by (25). Hence we can say that the expected value of h_{n+1} lies in the confidence interval

$$\big[\,\widetilde{h}_{n+1} - 1.65\sqrt{\Delta_{n+1}},\ \widetilde{h}_{n+1} + 1.65\sqrt{\Delta_{n+1}}\,\big]$$

with probability close to 0.90. This means that, in 90% cases the predicted value $\widetilde{S}_{n+1}$ of the market price S_{n+1} (calculated from the observations $h_1, \dots, h_n$) lies in the interval

$$\big[\,S_n e^{\widetilde{h}_{n+1} - 1.65\sqrt{\Delta_{n+1}}},\ S_n e^{\widetilde{h}_{n+1} + 1.65\sqrt{\Delta_{n+1}}}\,\big].$$

4. However attractive, this conjecture on the 'normality' of the distribution of the variables h_n, $n \geqslant 1$, must be taken with caution, because the empirical analysis of much financial data shows (see Chapter IV) that

(a) the number of values in a sample that lie outside the 'confidence' intervals $\big[\,\overline{h}_n - k\widehat{\sigma}_n, \overline{h}_n + k\widehat{\sigma}_n\big]$ with $k = 1, 2, 3$ (here $\overline{h}_n = \dfrac{1}{n}\sum\limits_{i=1}^{n} h_i$ is the sample mean and $\widehat{\sigma}_n$ is the standard deviation:

$$\widehat{\sigma}_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(h_i - \overline{h}_n)^2)$$

is *considerably larger* than it should be if this conjecture were true. More geometrically this means that the 'tails' of the empirical densities are 'heavy', i.e., they decrease at a much lower rate than it should occur for Gaussian distributions;

(b) the *kurtosis*

$$\widehat{k}_n = \frac{\widehat{m}_4}{\widehat{m}_2^2} - 3$$

(here $\widehat{m}_2$ and $\widehat{m}_4$ are the empirical second and fourth moments) is markedly *positive* (the kurtosis of a normal distribution is equal to zero), which shows that the distribution density has a high peak around the mean value (is *leptokurtotic*):

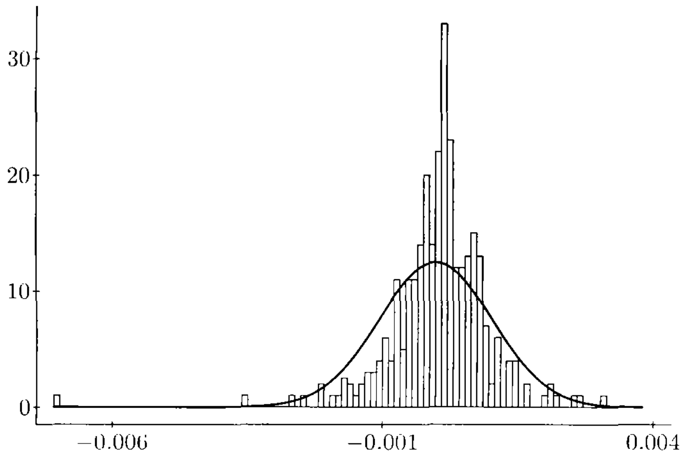


FIGURE 12. Empirical density of the one-dimensional distribution of the variables h_n, $n \leqslant 300$, governed by the $HARCH(16)$ model (see § 3b). The continuous curve is the density of the corresponding normal distribution $\mathscr{N}(m, \sigma^2)$ with $m = \overline{h}_{300}$ and $\sigma^2 = \widehat{\sigma}^2_{300}$

Arguably, the most strong assumption about the general properties of the distribution of the h_n, apart from the Gaussian property, is the following one:

these variables are independent and
identically distributed.

Under this assumption it is easy to carry out the analysis of the prices $S_n = S_0 e^{H_n}$, where $H_n = h_1 + \cdots + h_n$, by the standard methods of probability theory designed expressly for such situations. It is clear, however, that the assumption of the *independence* of the h_n undermines all the hopes that the 'past' information could be of any use in the prediction of the 'future values'.

In practice, the situation is more favorable, since numerous studies of financial time series enable one to say that, as mentioned already, the variables (h_n) are *non-Gaussian* and *dependent*, although they can be uncorrelated and their dependence can be rather weak. The easiest way to demonstrate certain degree of dependence is to consider the empirical correlations for the $|h_n|$ or h_n^2, rather then for the h_n. (In the model of stochastic volatility discussed below we have $\mathsf{Cov}(h_n, h_m) = 0$ for $n \neq m$, but the values of $\mathsf{Cov}(h_n^2, h_m^2)$ and $\mathsf{Cov}(|h_n|, |h_m|)$ are far from zero; see § 3c.)

§ 1b. Doob Decomposition. Canonical Representations

1. We shall assume that the variables h_n, $n \geqslant 1$, in the model

$$S_n = S_0 e^{H_n}, \qquad H_n = h_1 + \cdots + h_n, \tag{1}$$

have finite absolute first moments, i.e., $\mathsf{E}|h_n| < \infty$ for $n \geqslant 1$.

The Doob decomposition that we discuss later in this section indicates that one should study the sequence $H = (H_n)$ in its dependence on the properties of the filtration $(\mathscr{F}_n)$, i.e., of the flow of 'information' $\mathscr{F}_n$ accessible to an 'observer' (of the securities market, in the context adopted here).

Since $\mathsf{E}|h_n| < \infty$ for $n \geqslant 1$, the *conditional expectations* $\mathsf{E}(h_n \,|\, \mathscr{F}_{n-1})$ are well defined and ($H_0 = 0$, $\mathscr{F}_0 = \{\varnothing, \Omega\}$)

$$H_n = \sum_{k \leqslant n} \mathsf{E}(h_k \,|\, \mathscr{F}_{k-1}) + \sum_{k \leqslant n} \bigl[h_k - \mathsf{E}(h_k \,|\, \mathscr{F}_{k-1})\bigr]. \tag{2}$$

In other words, setting

$$A_n = \sum_{k \leqslant n} \mathsf{E}(h_k \,|\, \mathscr{F}_{k-1}), \tag{3}$$

and

$$M_n = \sum_{k \leqslant n} \bigl[h_k - \mathsf{E}(h_k \,|\, \mathscr{F}_{k-1})\bigr] \tag{4}$$

we obtain the following *Doob decomposition* for $H = (H_n)$:

$$H_n = A_n + M_n, \qquad n \geqslant 1, \tag{5}$$

where

a) the sequence $A = (A_n)_{n \geqslant 0}$, where $A_0 = 0$, is *predictable*, i.e.,

$$A_n \text{ are } \mathscr{F}_{n-1}\text{-measurable}, \quad n \geqslant 1;$$

b) the sequence $M = (M_n)_{n \geqslant 0}$, where $M_0 = 0$, is a *martingale*, i.e.,

$$\mathsf{E}(M_n \,|\, \mathscr{F}_{n-1}) = M_{n-1} \quad (\mathsf{P}\text{-a.s.}), \quad n \geqslant 1,$$

the M_n are $\mathscr{F}_n$-measurable and $\mathsf{E}|M_n| < \infty$ for $n \geqslant 1$.

Remark. Assume that, besides the filtration $(\mathscr{F}_n)$, we have a *subfiltration* $(\mathscr{G}_n)$, where $\mathscr{G}_n \subseteq \mathscr{F}_n$ and $\mathscr{G}_n \subseteq \mathscr{G}_{n+1}$. Then we can decompose $H = (H_n)$ with respect to the flow $(\mathscr{G}_n)$ in a similar way to (5):

$$H_n = \sum_{k=1}^{n} \mathsf{E}(h_k \mid \mathscr{G}_{k-1}) + \sum_{k=1}^{n} (h_k - \mathsf{E}(h_k \mid \mathscr{G}_{k-1})).$$

The sequence $A = (A_n)$ of the variables

$$A_n = \sum_{i=1}^{n} \mathsf{E}(h_k \mid \mathscr{G}_{k-1})$$

is $(\mathscr{G}_n)$-predictable (i.e., the A_n are $\mathscr{G}_{n-1}$-measurable), but $M = (M_n)$, where

$$M_n = \sum_{k=1}^{n} (h_k - \mathsf{E}(h_k \mid \mathscr{G}_{k-1}))$$

is, generally speaking, not a martingale with respect to $(\mathscr{G}_n)$ since the h_k are measurable with respect to the $\mathscr{F}_k$, but not necessarily with respect to the $\mathscr{G}_k$.

It must be pointed out that if, besides (5), we have another decomposition,

$$H_n = A'_n + M'_n,$$

where the sequence $A' = (A'_n, \mathscr{F}_n)$ is predictable (with respect to the flow $(\mathscr{F}_n)$), $A'_0 = 0$, and $M' = (M'_n, \mathscr{F}_n)$ is a martingale, then $A'_n = A_n$ and $M'_n = M_n$ for all $n \geqslant 0$.

For we have

$$A'_{n+1} - A'_n = (A_{n+1} - A_n) + (M_{n+1} - M_n) - (M'_{n+1} - M'_n).$$

Hence, considering the conditional expectations $\mathsf{E}(\cdot \mid \mathscr{F}_n)$ of both sides we see that $A'_{n+1} - A'_n = A_{n+1} - A_n$ (since the A'_{n+1} and A_{n+1} are $\mathscr{F}_n$-measurable). However, $A'_0 = A_0 = 0$, therefore $A'_n = A_n$ and $M'_n = M_n$ for all $n \geqslant 0$. Hence the decomposition (1) with *predictable* sequence $A = (A_n)$ is *unique.*

We note also that if $\mathsf{E}(h_k \mid \mathscr{F}_{k-1}) = 0$ for $k \geqslant 1$ in the model under consideration, then the sequence $H = (H_n)$ is itself a *martingale* by (2).

We now present an example of the Doob decomposition, which shows clearly that, for all its simplicity, it is 'nontrivial'.

EXAMPLE. Let $X_0 = 0$ and $X_n = \xi_1 + \cdots + \xi_n$, $n \geqslant 1$, where the ξ_n are independent Bernoulli variables such that

$$\mathsf{P}(\xi_n = \pm 1) = \frac{1}{2}.$$

We consider the Doob decomposition for $H_0 = 0$ and $H_n = |X_n|$, $n \geqslant 1$.

In this case we have

$$h_n = \Delta H_n = \Delta|X_n| = |X_n| - |X_{n-1}| = |X_{n-1} + \xi_n| - |X_{n-1}|,$$

and it is clear that

$$\begin{aligned}\Delta M_n \equiv h_n - \mathsf{E}(h_n \,|\, \mathscr{F}_{n-1}) &= |X_{n-1} + \xi_n| - \mathsf{E}(|X_{n-1} + \xi_n| \,|\, \mathscr{F}_{n-1}) \\ &= |X_{n-1} + \xi_n| - \mathsf{E}(|X_{n-1} + \xi_n| \,|\, X_{n-1}) \\ &= (\operatorname{Sgn} X_{n-1})\xi_n,\end{aligned}$$

where

$$\operatorname{Sgn} x = \begin{cases} 1, & x > 0, \\ 0, & x = 0, \\ -1, & x < 0. \end{cases}$$

Thus, the *martingale* $M = (M_n)_{n\geqslant 1}$ in (5) is now as follows:

$$M_n = \sum_{1\leqslant k\leqslant n} (\operatorname{Sgn} X_{k-1})\Delta X_k.$$

Further,

$$\mathsf{E}(h_n \,|\, \mathscr{F}_{n-1}) = \mathsf{E}(|X_{n-1} + \xi_n| \,|\, X_{n-1}) - |X_{n-1}|.$$

The right-hand side vanishes on the set $\{\omega\colon X_{n-1} = i\}$ with $i \neq 0$, while if $i = 0$, then it is equal to one. Hence

$$\sum_{i=1}^{n} \mathsf{E}(h_i \,|\, \mathscr{F}_{i-1}) = N\{1 \leqslant k \leqslant n\colon X_{k-1} = 0\},$$

where $N\{1 \leqslant k \leqslant n\colon X_{k-1} = 0\}$ is the number of k, $1 \leqslant k \leqslant n$, such that $X_{k-1} = 0$.

Let $L_n(0) = N\{0 \leqslant k \leqslant n-1\colon X_k = 0\}$ be the number of zeros in the sequence $(X_k)_{0\leqslant k\leqslant n-1}$. Then

$$|X_n| = \sum_{1\leqslant k\leqslant n} (\operatorname{Sgn} X_{k-1})\Delta X_k + L_n(0),$$

which is a *discrete analogue of the well-known Tanaka formula* for the absolute value of a Brownian motion (see Chapter III, § 5c). Hence, in particular,

$$\mathsf{E}L_n(0) = \mathsf{E}|X_n|.$$

Since $\dfrac{X_n}{\sqrt{n}} \sim \mathscr{N}(0,1)$, it follows that $\mathsf{E}|X_n| \sim \sqrt{\dfrac{2}{\pi}n}$ and therefore

$$\mathsf{E}L_n(0) \sim \sqrt{\frac{2}{\pi}n}.$$

This is a known result of the *average number of zeros* in a symmetric random Bernoulli walk (see, e.g., [156]).

2. Let $M = (M_n)_{n\geqslant 1}$ be a *square integrable* martingale ($\mathsf{E}M_n^2 < \infty$, $n \geqslant 1$) with $M_0 = 0$. Then setting $H_n = M_n^2$ in the decomposition (2) we obtain the following representation:

$$M_n^2 = \sum_{1\leqslant k\leqslant n} \mathsf{E}(\Delta M_k^2 \,|\, \mathscr{F}_{k-1}) + \sum_{1\leqslant k\leqslant n} \bigl(\Delta M_k^2 - \mathsf{E}(\Delta M_k^2 \,|\, \mathscr{F}_{k-1})\bigr), \tag{6}$$

where $\Delta M_k^2 = M_k^2 - M_{k-1}^2$.

We now set

$$\begin{aligned} \langle M\rangle_n &= \sum_{1\leqslant k\leqslant n} \mathsf{E}(\Delta M_k^2 \,|\, \mathscr{F}_{k-1}), \\ m_n &= \sum_{1\leqslant k\leqslant n} \bigl(\Delta M_k^2 - \mathsf{E}(\Delta M_k^2 \,|\, \mathscr{F}_{k-1})\bigr). \end{aligned} \tag{7}$$

Using this notation we can rewrite (6) as follows:

$$M_n^2 = \langle M\rangle_n + m_n; \tag{8}$$

the (predictable) sequence $\langle M\rangle = \bigl(\langle M\rangle_n\bigr)_{n\geqslant 1}$ is called here the *quadratic characteristic* of the martingale M (cf. Chapter III, § 5b).

We note that since $M = (M_n)$ is a martingale, it follows that

$$\mathsf{E}\bigl(\Delta M_k^2 \,|\, \mathscr{F}_{k-1}\bigr) = \mathsf{E}\bigl((\Delta M_k)^2 \,|\, \mathscr{F}_{k-1}\bigr). \tag{9}$$

This property explains why one often calls the quadratic characteristic $\langle M\rangle$ also the *predictable quadratic variation* of the (square integrable) martingale M. In that case one reserves the term *quadratic variation* for the (in general, unpredictable) sequence $[M] = ([M]_n)_{n\geqslant 1}$ of the variables (cf. Chapter III, § 5b)

$$[M]_n = \sum_{k\leqslant n} (\Delta M_k)^2. \tag{10}$$

3. We now assume that the sequence $H = (H_n)$ is itself a martingale, and even a square integrable one, i.e., $\mathsf{E}(\Delta H_k \,|\, \mathscr{F}_{k-1}) = \mathsf{E}(h_k \,|\, \mathscr{F}_{k-1}) = 0$, $\mathsf{E}h_k^2 < \infty$, $k \geqslant 1$. Then

$$\langle H\rangle_n = \sum_{k\leqslant n} \mathsf{E}(h_k^2 \,|\, \mathscr{F}_{k-1}). \tag{11}$$

The variables $\mathsf{E}(h_k^2 \,|\, \mathscr{F}_{k-1})$ contributing to the quadratic characteristic $\langle H\rangle_n$ define the *volatility* of H and, in large measure, its properties. For example, if $\langle H\rangle_n \to \infty$ with probability one, then the square integrable martingale H satisfies the *strong law of large numbers*:

$$\frac{H_n}{\langle H\rangle_n} \to 0 \quad (\mathsf{P}\text{-a.s.}) \tag{12}$$

as $n \to \infty$. (See [439; Chapter VII, § 5].)

The collection of variables $\big(\mathsf{E}(h_k^2 \mid \mathscr{F}_{k-1})\big)_{k\geqslant 1}$ will play an important role in our subsequent analysis of 'financial' time series $S = (S_n)$ with $S_n = S_0 e^{H_n}$. In our considerations of these series, using the vocabulary of the *financial theory* we shall call the sequence $\big(\mathsf{E}(h_k^2 \mid \mathscr{F}_{k-1})\big)_{k\geqslant 1}$ the *stochastic volatility*. (See § 3 for a closer look at volatility.)

If the conditional expectations $\mathsf{E}(h_k^2 \mid \mathscr{F}_{k-1})$ coincide with the unconditional ones (e.g., if (h_n) is a sequence of independent random variables and $\mathscr{F}_{k-1} = \sigma(h_1, \dots, h_{k-1})$ is the σ-algebra generated by $h_1, \dots, h_{n-1}$), then the volatility is a mere collection of variances $\sigma_k^2 = \mathsf{E}h_k^2$, $k \geqslant 1$, that are the standard *measures for the dispersion* (*changeability*) of the h_k (here we assume that $\mathsf{E}h_k = 0$).

4. In deducing the Doob decomposition (2) or (5) we assumed that $\mathsf{E}|h_k| < \infty$ for $k \geqslant 1$. Actually, we required this assumption only to ensure that the conditional expectations $\mathsf{E}(h_k \mid \mathscr{F}_{k-1})$, $k \geqslant 1$, were *well defined*. Thus, there arises a natural idea of using the decomposition (2) also in the (more general) case when the *conditional expectations* $\mathsf{E}(h_k \mid \mathscr{F}_{k-1})$ are well defined and finite and the condition $\mathsf{E}|h_k| < \infty$ is not necessarily satisfied.

To this end we recall that if $\mathsf{E}|h_k| < \infty$, then the conditional expectation $\mathsf{E}(h_k \mid \mathscr{F}_{k-1})$ is (according to A. N. Kolmogorov) the $\mathscr{F}_{k-1}$-measurable random variable such that

$$\int_A \mathsf{E}(h_k \mid \mathscr{F}_{k-1})\, d\mathsf{P} = \int_A h_k\, d\mathsf{P} \tag{13}$$

for each $A \in \mathscr{F}_{k-1}$. The existence of such a variable is a consequence of the Radon–Nikodým theorem (see, e.g., [439; Chapter II, § 7]).

Note, however, that the condition $\mathsf{E}|h_k| < \infty$ is not at all *necessary* for the existence of an $\mathscr{F}_{k-1}$-measurable variable $\mathsf{E}(h_k \mid \mathscr{F}_{k-1})$ satisfying (13). For instance, if $h_k \geqslant 0$ (P-a.s.), then we can define it without assuming that $\mathsf{E}h_k < \infty$. Hence the idea of the following definition of a *generalized* expectation, which we shall also denote by $\mathsf{E}(h_k \mid \mathscr{F}_{k-1})$.

We represent h_k as the following sum:

$$h_k = h_k^+ - h_k^-,$$

where $h_k^+ = \max(h_k, 0)$ and $h_k^- = -\min(h_k, 0)$. We assume that we have already defined $\mathsf{E}(h_k^+ \mid \mathscr{F}_{k-1})(\omega)$ and $\mathsf{E}(h_k^- \mid \mathscr{F}_{k-1})(\omega)$ so that

$$\min\big\{\mathsf{E}(h_k^+ \mid \mathscr{F}_{k-1})(\omega), \mathsf{E}(h_k^- \mid \mathscr{F}_{k-1})(\omega)\big\} < \infty \tag{14}$$

for all $\omega \in \Omega$. Then we set

$$\mathsf{E}(h_k \mid \mathscr{F}_{k-1})(\omega) \equiv \mathsf{E}(h_k^+ \mid \mathscr{F}_{k-1})(\omega) - \mathsf{E}(h_k^- \mid \mathscr{F}_{k-1})(\omega) \tag{15}$$

(here and throughout "$\equiv$" means *by definition*), and we call $\mathsf{E}(h_k \mid \mathscr{F}_{k-1})$ the *generalized conditional expectation*.

If $\mathsf{E}|h_k| < \infty$, then this generalized expectation is clearly the same as the usual conditional expectation.

If $\mathsf{E}(|h_k| \,|\, \mathscr{F}_{k-1})(\omega) < \infty$ for $\omega \in \Omega$, then (14) obviously holds and $\mathsf{E}(h_k \,|\, \mathscr{F}_{k-1})(\omega)$ is not merely well defined, but also *finite* for each $\omega \in \Omega$. In this case we say that the generalized conditional expectation $\mathsf{E}(h_k \,|\, \mathscr{F}_{k-1})$ is *well defined* and *finite*.

Remark. Proceeding in accordance with the general spirit of probability theory, which usually puts weight on the verification of particular properties '*for almost all* $\omega \in \Omega$', rather then 'for all' of them, we can easily construct an 'almost-all' version of the above definition of the generalized conditional expectation by setting $\mathsf{E}(h_k \,|\, \mathscr{F}_{k-1})(\omega)$ to be arbitrary on the zero-probability set where (14) fails.

We now consider the representation (2). The right-hand side of (2) is surely well defined if $\mathsf{E}(|h_k| \,|\, \mathscr{F}_{k-1}) < \infty$ for $k \geqslant 1$ (and for all or almost all $\omega \in \Omega$). In that case we shall say that (2) is a *generalized Doob decomposition* of the sequence $H = (H_n)_{n \geqslant 1}$.

5. We now discuss a similar decomposition (or, as we shall also call it, a *representation*) in the case when the conditional expectations $\mathsf{E}(h_k \,|\, \mathscr{F}_{k-1})$ (either 'usual' or generalized) are *not defined*. Then we can proceed as follows.

We represent h_k as a sum:

$$h_k = h_k I(|h_k| \leqslant a) + h_k I(|h_k| > a),$$

where a is a certain positive constant (one usually sets $a = 1$) and $I(A)$ (we shall also write I_A or $I_A(\omega)$) is the indicator of the set A (i.e., $I_A(\omega) = 1$ if $\omega \in A$ and $I_A(\omega) = 0$ otherwise).

Now, the variables $h_k I(|h_k| \leqslant a)$ have well-defined first absolute moments, therefore

$$\begin{aligned} H_n^{(\leqslant a)} &\equiv \sum_{1 \leqslant k \leqslant n} h_k I(|h_k| \leqslant a) \\ &= \sum_{1 \leqslant k \leqslant n} \mathsf{E}\big(h_k I(|h_k| \leqslant a) \,|\, \mathscr{F}_{k-1}\big) \\ &\qquad + \sum_{1 \leqslant k \leqslant n} \big[h_k I(|h_k| \leqslant a) - \mathsf{E}\big(h_k I(|h_k| \leqslant a) \,|\, \mathscr{F}_{k-1}\big)\big] \\ &\big(= A_n^{(\leqslant a)} + M_n^{(\leqslant a)}\big). \end{aligned} \tag{16}$$

Hence

$$H_n = A_n^{(\leqslant a)} + M_n^{(\leqslant a)} + \sum_{1 \leqslant k \leqslant n} h_k I\big(|h_k| > a\big), \tag{17}$$

where $\big(A_n^{(\leqslant a)}\big)_{n \geqslant 1}$ is a predictable sequence, $\big(M_n^{(\leqslant a)}\big)_{n \geqslant 1}$ is a martingale, and $\Big(\sum_{1 \leqslant k \leqslant n} h_k I(|h_k| > a)\Big)_{n \geqslant 1}$ is the sequence of 'large' jumps.

Using the vocabulary of the 'general theory of stochastic processes' (see Chapter III, § 5 and [250; Chapter I, § 4c]) we call (17) the *canonical representation* of H.

We note that if, besides (17), we have another representation of H in the form

$$H_n = A'_n + M'_n + \sum_{1 \leqslant k \leqslant n} h_k I(|h_k| > a) \tag{18}$$

with predictable sequence (A'_n) and martingale (M'_n), then, of necessity, $A'_n = A_n^{(\leqslant a)}$ and $M'_n = M_n^{(\leqslant a)}$.

In other words, there exists a *unique* representation of the form (18). This justifies the name *canonical* given to this representation (17).

§ 1c. Local Martingales. Martingale transformations. Generalized Martingales

1. In the above analysis of the sequence $H = (H_n)$ based on the Doob decomposition (5) and its generalization these were the concepts of a '*martingale*' and '*predictability*' (and, accordingly, the *martingale* $M = (M_n)$ and the *predictable* sequence $A = (A_n)$ involved in the representation of H) that played a key role.

This explains why one often calls the subsequent stochastic analysis *martingale* or *stochastic calculus*, meaning here analysis in filtered probability spaces, i.e., probability spaces distinguished by a special structure, a flow of σ-algebras $(\mathscr{F}_n)$ (a 'filtration'). It is precisely with this structure that stopping times, martingales, predictability, sub- and supermartingales, and some other concepts are connected.

Perhaps, an even more important position than that of *martingales* in the modern stochastic calculus is occupied by the concept of *local martingale*. Remarkably, local martingales form a wider class than martingales, but retain many important properties of the latter.

We now present several definitions.

Let $(\Omega, \mathscr{F}, (\mathscr{F}_n), \mathsf{P})$ be a stochastic basis, i.e., a filtered probability space with discrete time $n \geqslant 0$.

DEFINITION 1. We call a sequence of random variables $X = (X_n)$ defined on the stochastic basis a *stochastic sequence* if X_n is $\mathscr{F}_n$-measurable for each $n \geqslant 0$.

To emphasize this property of *measurability* one writes stochastic sequences as $X = (X_n, \mathscr{F}_n)$, thus incorporating in the notation the σ-algebras $\mathscr{F}_n$ with respect to which the X_n are measurable.

DEFINITION 2. We call a stochastic sequence $X = (X_n, \mathscr{F}_n)_{n \geqslant 0}$

a *martingale*,
a *supermartingale*,
a *submartingale*

if $\mathsf{E}|X_n| < \infty$ for each $n \geqslant 0$ and if (P-a.s.)

$$\mathsf{E}(X_n \mid \mathscr{F}_{n-1}) = X_{n-1},$$
$$\mathsf{E}(X_n \mid \mathscr{F}_{n-1}) \leqslant X_{n-1},$$
$$\mathsf{E}(X_n \mid \mathscr{F}_{n-1}) \geqslant X_{n-1},$$

respectively, for all $n \geqslant 1$.

Clearly, $\mathsf{E}X_n = \text{Const}\ (= \mathsf{E}X_0)$ for a martingale, the expectations are nonincreasing ($\mathsf{E}X_n \leqslant \mathsf{E}X_{n-1}$) for a supermartingale, and they are nondecreasing ($\mathsf{E}X_n \geqslant \mathsf{E}X_{n-1}$) for a submartingale.

A classical example of a martingale is delivered by a *Lévy martingale* $X = (X_n)$ with $X_n = \mathsf{E}(\xi \mid \mathscr{F}_n)$, where ξ is a $\mathscr{F}$-measurable random variable such that $\mathsf{E}|\xi| < \infty$.

This is a *uniformly integrable* martingale, i.e., the family $\{X_n\}$ is uniformly integrable:

$$\sup_n \mathsf{E}\big(|X_n| I(|X_n| > C)\big) \to 0 \quad \text{as} \quad C \to \infty.$$

In what follows, we denote by $\mathscr{M}_{\mathrm{UI}}$ the class of *all uniformly integrable martingales.* We denote the class of *all martingales* by $\mathscr{M}$.

In the case when the martingales in question are defined only for $n \leqslant N < \infty$, the concepts of a martingale and of a uniformly integrable martingale are clearly the same ($\mathscr{M} = \mathscr{M}_{\mathrm{UI}}$).

Sometimes, when there is a need to point out the measure P and the flow $(\mathscr{F}_n)$ with respect to which the *martingale* property is considered, one also denotes the classes $\mathscr{M}_{\mathrm{UI}}$ and $\mathscr{M}$ by $\mathscr{M}_{\mathrm{UI}}(\mathsf{P}, (\mathscr{F}_n))$ and $\mathscr{M}(\mathsf{P}, (\mathscr{F}_n))$.

DEFINITION 3. We call a stochastic sequence $x = (x_n, \mathscr{F}_n)_{n \geqslant 1}$ with $\mathsf{E}|x_n| < \infty$ a *martingale difference* if (here we set $\mathscr{F}_0 = \{\varnothing, \Omega\}$)

$$\mathsf{E}(x_n \mid \mathscr{F}_{n-1}) = 0 \quad (\mathsf{P}\text{-a.s.}), \qquad n \geqslant 1.$$

Clearly, if $x = (x_n)$ is such a sequence, then the corresponding sequence of sums $X = (X_n, \mathscr{F}_n)$, where $X_n = X_0 + x_1 + \cdots + x_n$, is a *martingale.* Conversely, with each martingale $X = (X_n, \mathscr{F}_n)$ we can associate the *martingale difference* $x = (x_n, \mathscr{F}_n)$ with $x_n = \Delta X_n$, where $\Delta X_n = X_n - X_{n-1}$ for $n \geqslant 1$ and $\Delta X_0 = X_0$ for $n = 0$.

DEFINITION 4. We call a stochastic sequence $X = (X_n, \mathscr{F}_n)$ a *local martingale* (*submartingale, supermartingale*) if there exists a (localizing) sequence $(\tau_k)_{k \geqslant 1}$ of Markov times (i.e., of variables satisfying the condition $\{\omega \colon \tau_k \leqslant n\} \in \mathscr{F}_n$, $n \geqslant 1$; see also Definition 1 in § 1f) such that $\tau_k \leqslant \tau_{k+1}$ (P-a.s.), $\tau_k \uparrow \infty$ (P-a.s.) as $k \to \infty$, and each '*stopped*' sequence

$$X^{\tau_k} = \big(X_{\tau_k \wedge n}, \mathscr{F}_n\big)$$

is a martingale (submartingale, supermartingale).

Remark 1. One often includes in the definition of a local martingale the requirement that the sequence X^{τ_k} be not merely a martingale for each $k \geqslant 1$, but a uniformly integrable martingale (see, e.g., [250]). We note also that sometimes, intending to consider also sequences $X = (X_n, \mathscr{F}_n)$ with *nonintegrable* 'initial' random value X_0, one defines stopped sequences X^{τ_k} in a somewhat different way:

$$X^{\tau_k} = \bigl(X_{\tau_k \wedge n} I(\tau_k > 0), \mathscr{F}_n\bigr).$$

We shall write $\mathscr{M}_{\mathrm{loc}}$ or $\mathscr{M}_{\mathrm{loc}}(\mathsf{P}, (\mathscr{F}_n))$ for the class of local martingales.

By Definition 4, each martingale is a local martingale, so that

$$\mathscr{M} \subseteq \mathscr{M}_{\mathrm{loc}}.$$

If $X \in \mathscr{M}_{\mathrm{loc}}$ and the family of random variables

$$\Sigma = \bigl\{X_\tau : \tau \text{ is a finite stopping time}\bigr\}$$

is uniformly integrable $\Bigl($i.e., $\sup\limits_{X_\tau \in \Sigma} \mathsf{E}\bigl\{|X_\tau| I(|X_\tau| \geqslant C)\bigr\} \to 0$ as $C \to \infty\Bigr)$, then X is a martingale ($X \in \mathscr{M}$); moreover, it is a Lévy martingale: there exists an integrable, $\mathscr{F}$-measurable random variable X_∞ such that $X_n = \mathsf{E}(X_\infty \,|\, \mathscr{F}_n)$. Thus, $X \in \mathscr{M}_{\mathrm{UI}}$ in this case. (See [250; Chapter 1, § 1e] or [439; Chapter VII, § 4] for greater detail.)

DEFINITION 5. We say that a stochastic sequence $X = (X_n, \mathscr{F}_n)_{n \geqslant 0}$ is a *generalized martingale* (*submartingale*, *supermartingale*) if $\mathsf{E}|X_0| < \infty$, the generalized conditional expectations $\mathsf{E}(X_n \,|\, \mathscr{F}_{n-1})$ are *well defined* for each $n \geqslant 1$ and

$$\mathsf{E}(X_n \,|\, \mathscr{F}_{n-1}) = X_{n-1} \qquad (\mathsf{P}\text{-a.s.})$$

(accordingly, $\mathsf{E}(X_n \,|\, \mathscr{F}_{n-1}) \geqslant X_{n-1}$ or $\mathsf{E}(X_n \,|\, \mathscr{F}_{n-1}) \leqslant X_{n-1}$).

Remark 2. By the *definition* (see § 1b) of the generalized expectation $\mathsf{E}(X_n \,|\, \mathscr{F}_{n-1})$ and the 'martingale' equality $\mathsf{E}(X_n \,|\, \mathscr{F}_{n-1}) = X_{n-1}$ we obtain *automatically* that $\mathsf{E}(|X_n| \,|\, \mathscr{F}_{n-1}) < \infty$ (P-a.s.). This means that the conditional expectation $\mathsf{E}(X_n \,|\, \mathscr{F}_{n-1})$ is not merely well defined, but also *finite*. Hence we can assume in Definition 5 that $\mathsf{E}(|X_n| \,|\, \mathscr{F}_{n-1}) < \infty$ (P-a.s.) for $n \geqslant 0$.

DEFINITION 6. We call a stochastic sequence $x = (x_n, \mathscr{F}_n)_{n \geqslant 1}$ a *generalized martingale difference* (*submartingale difference*, *supermartingale difference*) if the generalized conditional expectations $\mathsf{E}(x_n \,|\, \mathscr{F}_{n-1})$ are well defined for each $n \geqslant 1$ and

$$\mathsf{E}(x_n \,|\, \mathscr{F}_{n-1}) = 0 \quad (\mathsf{P}\text{-a.s.})$$

(respectively, if $\mathsf{E}(x_n \,|\, \mathscr{F}_{n-1}) \geqslant 0$ or $\mathsf{E}(x_n \,|\, \mathscr{F}_{n-1}) \leqslant 0$ (P-a.s.)).

Remark 3. As in Remark 2, it can be requested in the definition of a generalized martingale difference that $\mathsf{E}(|x_n| \,|\, \mathscr{F}_{n-1}) < \infty$ (P-a.s.) for $n \geqslant 1$.

DEFINITION 7. Let $M = (M_n, \mathscr{F}_n)$ be a stochastic sequence and let $Y = (Y_n, \mathscr{F}_{n-1})$ be a predictable sequence (the Y_n are $\mathscr{F}_{n-1}$-measurable for $n \geqslant 1$ and Y_0 is $\mathscr{F}_0$-measurable).

Then we call the stochastic sequence

$$Y \cdot M = \big((Y \cdot M)_n, \mathscr{F}_n\big),$$

where

$$(Y \cdot M)_n = Y_0 \cdot M_0 + \sum_{1 \leqslant k \leqslant n} Y_k \Delta M_k,$$

the *transformation* of M by means of Y. If, in addition, M is a *martingale*, then we call $X = Y \cdot M$ a *martingale transformation* (of the *martingale* M by means of the (predictable) sequence Y).

The following result shows that the concepts introduced by Definitions 4, 5, and 7, are closely related in the *discrete-time* case.

THEOREM. *Let* $X = (X_n, \mathscr{F}_n)_{n \geqslant 0}$ *be a stochastic sequence with* $\mathsf{E}|X_0| < \infty$. *Then the following conditions are equivalent:*

(a) X *is a local martingale* ($X \in \mathscr{M}_{\mathrm{loc}}$);

(b) X *is a generalized martingale* ($X \in G\mathscr{M}$);

(c) X *is a martingale transformation* ($X \in \mathscr{M}T$), *i.e.,* $X = Y \cdot M$ *for some predictable sequence* $Y = (Y_n, \mathscr{F}_{n-1})$ *and some martingale* $M = (M_n, \mathscr{F}_n)$.

Proof. (c)$\Longrightarrow$(a). Let $X \in \mathscr{M}T$ and let

$$X_n = X_0 + \sum_{k=1}^{n} Y_k \Delta M_k, \tag{1}$$

where Y is a predictable sequence and M is a martingale. If $|Y_k| \leqslant C$ for $k \geqslant 1$, then X is clearly a martingale. Otherwise we set $\tau_j = \inf\{n-1\colon |Y_n| > j\}$. Since the Y_n are $\mathscr{F}_{n-1}$-measurable, the τ_j are stopping times, $\tau_j \uparrow \infty$ as $j \to \infty$, and the 'stopped sequences' X^{τ_j} are again of the form (1) with bounded $Y_k^{\tau_j} = Y_k I\{k \leqslant \tau_j\}$. Hence $X \in \mathscr{M}_{\mathrm{loc}}$.

(a)$\Longrightarrow$(b). Let $X \in \mathscr{M}_{\mathrm{loc}}$ and let (τ_k) be the corresponding localizing sequence. Then $\mathsf{E}|X_{n+1}^{\tau_k}| < \infty$ and $\mathsf{E}(|X_{n+1}| \mid \mathscr{F}_n) = \mathsf{E}(|X_{n+1}^{\tau_k}| \mid \mathscr{F}_n)$ on the set $\{\tau_k > n\} \in \mathscr{F}_n$. Hence $\mathsf{E}(|X_{n+1}| \mid \mathscr{F}_n) < \infty$ (P-a.s.).

In a similar way, on the same set $\{\tau_k > n\}$ we have

$$\mathsf{E}(X_{n+1} \mid \mathscr{F}_n) = \mathsf{E}(X_{n+1}^{\tau_k} \mid \mathscr{F}_n) = X_n^{\tau_k} = X_n,$$

so that $X \in G\mathscr{M}$.

(b)$\Longrightarrow$(c). Let $X \in G\mathscr{M}$. We set

$$A_n(k) = \big\{\omega\colon \mathsf{E}(|X_{n+1}| \mid \mathscr{F}_n) \in [k, k+1)\big\}.$$

Then

$$u_n = \sum_{k\geqslant 0} (k+1)^{-3} \Delta X_n I_{A_{n-1}(k)}$$

is a $\mathscr{F}_n$-measurable random variable with $\mathsf{E}(u_n \mid \mathscr{F}_{n-1}) = 0$. Hence $M_n = \sum_{i=1}^{n} u_i$ is a martingale (we set $M_0 = 0$) and (1) holds for $Y = (Y_n)$, where

$$Y_n = \sum_{k\geqslant 0} (k+1)^3 I_{A_{n-1}(k)},$$

so that $X \in \mathscr{M}T$.

2. We shall demonstrate the full extent of the importance of the concepts of a local martingale, a martingale transformation, and a generalized martingale in financial mathematics in Chapter V. These concepts play an important role also in stochastic calculus, which can be shown, for example, as follows.

Let $X = (X_n, \mathscr{F}_n)_{n\geqslant 0}$ be a local submartingale with localizing sequence (τ_k), where $\tau_k > 0$ (P-a.s.). Then for each k we obtain the following decomposition for $X^{(\tau_k)} = (X_{n\wedge\tau_k}, \mathscr{F}_n)$:

$$X_{n\wedge\tau_k} = A_n^{(\tau_k)} + M_n^{(\tau_k)}$$

with predictable sequences $(A_n^{(\tau_k)})_{n\geqslant 0}$ and martingales $(M_n^{(\tau_k)})_{n\geqslant 0}$.

This decomposition is unique (provided that the sequences $(A_n^{(\tau_k)})_{n\geqslant 0}$ are predictable), therefore it is easy to see that

$$A_{n\wedge\tau_k}^{(\tau_{k+1})} = A_n^{(\tau_k)}.$$

Setting $A_n = A_n^{(\tau_k)}$ for $n \leqslant \tau_k$ we see that $(X_n - A_n)_{n\geqslant 0}$ is a local martingale because the 'stopped' sequences

$$(X_n^{(\tau_k)} - A_n^{(\tau_k)})_{n\geqslant 0}$$

are martingales.

Hence *if* $X = (X_n, \mathscr{F}_n)_{n\geqslant 0}$ *is a local submartingale, then*

$$X_n = A_n + M_n, \qquad n \geqslant 0, \tag{2}$$

where $A = (A_n)$ *is a predictable sequence with* $A_0 = 0$, *and* $M = (M_n)$ *is a local martingale.*

It should be noted that the sequence $A = (A_n)$ is *increasing* (more precisely, nondecreasing) in this case, which is a consequence of the following explicit formula for the $A_n^{(\tau_k)}$:

$$A_n^{(\tau_k)} = \sum_{i\leqslant n\wedge\tau_k} \mathsf{E}(\Delta X_i \mid \mathscr{F}_{i-1}),$$

and of the submartingale property $\mathsf{E}(\Delta X_i \mid \mathscr{F}_{i-1}) \geqslant 0$.

We note also that the decomposition (2) with predictable process $A = (A_n)$ is *unique*.

DEFINITION 8. Let $X = (X_n, \mathscr{F}_n)$ be a stochastic sequence admitting a representation $X_n = A_n + M_n$ with measurable sequence $A = (A_n, \mathscr{F}_{n-1})$ and local martingale $M = (M_n, \mathscr{F}_n)$. Then we say that X admits a *generalized Doob decomposition* and A is the *compensator* (or the *predictable compensator*, or else the *dual predictable projection*) of the sequence X.

(We call A a '*compensator*' since it *compensates* X to a local martingale.)

3. For conclusion, we present a simple but useful result of [251] describing conditions ensuring that a local martingale is a (usual) martingale.

LEMMA. 1) *Let $X = (X_n, \mathscr{F}_n)_{n\geqslant 0}$ be a local martingale such that $\mathsf{E}|X_0| < \infty$ and either*

$$\mathsf{E}X_n^- < \infty, \qquad n \geqslant 0, \tag{3}$$

or

$$\mathsf{E}X_n^+ < \infty, \qquad n \geqslant 0. \tag{4}$$

Then $X = (X_n, \mathscr{F}_n)_{n\geqslant 0}$ is a martingale.

2) *Let $X = (X_n, \mathscr{F}_n)_{0\leqslant n\leqslant N}$ be a local martingale and assume that $N < \infty$, $\mathsf{E}|X_0| < \infty$, and either $\mathsf{E}X_N^- < \infty$, or $\mathsf{E}X_N^+ < \infty$. Then (3) and (4) hold for each $n \leqslant N$ and $X = (X_n, \mathscr{F}_n)_{0\leqslant n\leqslant N}$ is a martingale.*

Proof. 1) We claim that each of conditions (3) and (4) implies the other and, therefore, the inequality $\mathsf{E}|X_n| < \infty$, $n \geqslant 0$.

For if, say, (3) holds, then by Fatou lemma ([439; Chapter II, § 6]),

$$\begin{aligned}\mathsf{E}X_n^+ &= \mathsf{E}\varliminf_k X^+_{n\wedge\tau_k} \leqslant \varliminf_k \mathsf{E}X^+_{n\wedge\tau_k} = \varliminf_k \big[\mathsf{E}X_{n\wedge\tau_k} + \mathsf{E}X^-_{n\wedge\tau_k}\big]\\ &= \mathsf{E}X_0 + \varliminf_k \mathsf{E}X^-_{n\wedge\tau_k} \leqslant |\mathsf{E}X_0| + \sum_{i=0}^{n} \mathsf{E}X_i^- < \infty.\end{aligned}$$

Consequently, $\mathsf{E}|X_n| < \infty$, $n \geqslant 0$.

Further, we have $|X_{(n+1)\wedge\tau_k}| \leqslant \sum_{i=0}^{n+1} |X_i|$ with $\mathsf{E}\sum_{i=0}^{n+1} |X_i| < \infty$, therefore by Lebesgue theorem on dominated convergence ([439; Chapter II, §§ 6 and 7]), passing to the limit (as $k \to \infty$) in the relation

$$\mathsf{E}\big(X_{\tau_k\wedge(n+1)} I(\tau_k > 0) \,|\, \mathscr{F}_n\big) = X_{\tau_k\wedge n}$$

we obtain $\mathsf{E}(X_{n+1} \,|\, \mathscr{F}_n) = X_n$, $n \geqslant 0$.

2) We note that if $\mathsf{E}X_N^- < \infty$, then also $\mathsf{E}X_n^- < \infty$ for $n \leqslant N$, because local martingales are generalized martingales. Hence $X_n = \mathsf{E}(X_{n+1} \,|\, \mathscr{F}_n)$, and therefore $X_n^- \leqslant \mathsf{E}(X_{n+1}^- \,|\, \mathscr{F}_n)$ and $\mathsf{E}X_n^- \leqslant \mathsf{E}X_{n+1}^- \leqslant \mathsf{E}X_N^-$ for all $n \leqslant N-1$.

Thus, it follows from 1) that the sequence $X = (X_n, \mathscr{F}_n)_{0\leqslant n\leqslant N}$ is a martingale.

In a similar way we can consider the case of $\mathsf{E}X_N^+ < \infty$ thus completing the proof of the lemma.

COROLLARY. *Each local martingale $X = (X_n)_{n\geqslant 0}$ that is bounded below* $\big(\inf_n X_n(\omega) \geqslant C > -\infty$, P-a.s.$\big)$ *or above* $\big(\sup_n X_n(\omega) \leqslant C < \infty$, P-a.s.$\big)$ *is a martingale.*

4. It is instructive to compare the result of the lemma with the corresponding result in the continuous-time case.

If $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geqslant 0}, \mathsf{P})$ is a filtered probability space with a nondecreasing family of σ-algebras $(\mathscr{F}_t)_{t\geqslant 0}$ ($\mathscr{F}_s \subseteq \mathscr{F}_t \subseteq \mathscr{F}$, $s \leqslant t$), then we call a stochastic process $X = (X_t)_{t\geqslant 0}$ a *martingale* (a *supermartingale*, a *submartingale*) if the X_t are $\mathscr{F}_t$-measurable, $\mathsf{E}|X_t| < \infty$ for $t \geqslant 0$, and $\mathsf{E}(X_t \mid \mathscr{F}_s) = X_s$ $\big(\mathsf{E}(X_t \mid \mathscr{F}_s) \leqslant X_s$ or $\mathsf{E}(X_t \mid \mathscr{F}_s) \geqslant X_s\big)$ for $s \leqslant t$.

We call the process $X = (X_t)_{t\geqslant 0}$ a *local martingale* if we can find a nondecreasing sequence of stopping times (τ_k) such that $\tau_k \uparrow \infty$ (P-a.s.) and the stopped processes $X^{\tau_k} = \big(X_{t\wedge\tau_k} I(\tau_k > 0), \mathscr{F}_t\big)$ are uniformly integrable martingales for all k.

Using the same arguments as in the proof of the lemma, which are based on Fatou's lemma and Lebesgue's theorem on dominated convergence ([439; Chapter II, § § 6 and 7]) we can prove the following results.

I. Each local martingale $X = (X_t)_{t\geqslant 0}$ satisfying the condition

$$\mathsf{E} \sup_{s\leqslant t} X_s^- < \infty, \qquad t \geqslant 0,$$

is a *supermartingale.*

II. Each local martingale $X = (X_t)_{t\geqslant 0}$ satisfying the condition

$$\mathsf{E} \sup_{s\leqslant t} |X_s| < \infty, \qquad t \geqslant 0,$$

is a *martingale.*

III. Each local martingale $X = (X_t)_{t\geqslant 0}$ satisfying the condition

$$\mathsf{E} \sup_{s<\infty} |X_s| < \infty, \qquad t \geqslant 0,$$

is a *uniformly integrable martingale.*

It is useful to bear in mind that, in the discrete-time case, if $\mathsf{E}X_n^- < \infty$ for $n \leqslant N$, then $\mathsf{E} \max_{n\leqslant N} X_n^- < \infty$. However, for continuous time, the inequality $\mathsf{E}X_t^- < \infty$, $t \leqslant T$, does not in general mean that $\mathsf{E} \sup_{t\leqslant T} X_t^- < \infty$. Essentially, this is the main reason explaining why the result of the above lemma cannot be automatically extended to the case of continuous time.

We note also that the local martingale in assertion I can indeed be a 'proper' supermartingale and not a martingale.

The following example is well known.

EXAMPLE. Let $B = (B_t^1, B_t^2, B_t^3)_{t\geqslant 0}$ be a three-dimensional Brownian motion. Let

$$R_t = \sqrt{(B_t^1)^2 + (B_t^2)^2 + (B_t^3)^2} \quad \text{with} \quad R_0 = \sqrt{(B_0^1)^2 + (B_0^2)^2 + (B_0^3)^2} = 1.$$

Then the process R_t (called the *Bessel process of order* 3) has the stochastic differential

$$dR_t = \frac{dt}{R_t} + d\beta_t, \qquad R_0 = 1,$$

where $\beta = (\beta_t)_{t\geqslant 0}$ is a standard Brownian motion (see § 3a, Chapter III, and, e.g., [402]).

By the Itô formula

$$df(R_t) = f'(R_t)\,dR_t + \frac{1}{2}f''(R_t)\,dt$$

(see Chapter III, § 3d and also [250], [402]) used for $f(R_t) = 1/R_t$ (which is possible because the zero value is unattainable for the *three-dimensional* Bessel process R with $R_0 = 1$) we obtain

$$d\left(\frac{1}{R_t}\right) = -\frac{d\beta_t}{R_t^2},$$

or, in the integral form,

$$\frac{1}{R_t} = 1 - \int_0^t \frac{d\beta_s}{R_s^2}.$$

The stochastic integral $\left(\int_0^t \frac{d\beta_s}{R_s^2}\right)_{t\geqslant 0}$ is a local martingale (see [402]), therefore the process $X = (X_t)_{t\geqslant 0}$ with $X_t = \dfrac{1}{R_t}$ and $X_0 = 1$ is a *local* martingale, but not a *martingale*.

Indeed, by the self-similarity property (see Chapter III, § 3a) of the Brownian motions B^1, B^2, B^3 we obtain

$$\begin{aligned}\mathsf{E}X_t &= \mathsf{E}\,\frac{1}{\sqrt{1 + (B_t^1 - B_0^1)^2 + (B_t^2 - B_0^2)^2 + (B_t^3 - B_0^3)^2}}\\ &= \mathsf{E}\,\frac{1}{\sqrt{1 + t(\xi_1^2 + \xi_2^2 + \xi_3^2)}} \downarrow 0 \quad \text{as} \quad t \to \infty,\end{aligned}$$

where the ξ_i, $i = 1, 2, 3$, are independent standard normally distributed random variables ($\xi_i \sim \mathcal{N}(0,1)$). On the other hand, if we had the *martingale* property, then the expectations $\mathsf{E}X_t$ would be *constant*.

§ 1d. Gaussian and Conditionally Gaussian Models

1. The concept of an 'efficient' market substantiates the 'martingale conjecture' for (discounted) prices. This makes 'martingale' a central notion in the analysis of the dynamics of prices regarded as stochastic sequences or processes having distributions with specific properties. However, the mere fact that the distributions have the martingale property is not sufficient for concrete calculations; one must know a 'finer' structure of these distributions, which brings one to the necessity of a thorough study of most diverse probabilistic and statistical models in order to find ones with distributions better conforming to the properties of the empirical distributions constructed on the basis of statistical data. The rest of this chapter is in fact devoted to achieving this aim. We present models enabling one to explain some or other properties discovered by the analysis of statistical 'stock' provided, in particular, by financial time series.

The assumption that the distributions $\mathrm{Law}(h_1, \dots, h_n)$ of the variables $h_1, \dots, h_n$ are Gaussian is, of course, most attractive from the viewpoints of both theoretical analysis and the well-developed 'statistics of the normal distribution'. However, one must take into account the fact (already pointed out in this book) that, as shows the statistical analysis of many financial series, this conjecture does not always reflect the behavior of prices properly.

Seeking an alternative to the conjecture that the unconditional distributions $\mathrm{Law}(h_1, \dots, h_n)$ for the sequence $h = (h_n)_{n\geqslant 1}$ are Gaussian and bearing in mind the *Doob decomposition*, which was introduced in terms of the conditional expectations $\mathsf{E}(h_n \mid \mathscr{F}_{n-1})$, it seems to be fairly natural to assume that these are *conditional* (rather than the unconditional) probability distributions $\mathrm{Law}(h_n \mid \mathscr{F}_{n-1})$ that are Gaussian. In other words,

$$\mathrm{Law}(h_n \mid \mathscr{F}_{n-1}) = \mathscr{N}(\mu_n, \sigma_n^2) \tag{1}$$

for some $\mathscr{F}_{n-1}$-measurable variables $\mu_n = \mu_n(\omega)$ and $\sigma_n^2 = \sigma_n^2(\omega)$.[a]

More precisely, (1) means that the (regular) conditional distribution $\mathsf{P}(h_n \leqslant x \mid \mathscr{F}_{n-1})$ can be described by the formula

$$\mathsf{P}(h_n \leqslant x \mid \mathscr{F}_{n-1})(\omega) = \frac{1}{\sqrt{2\pi\sigma_n^2}} \int_{-\infty}^{x} e^{-\frac{(y-\mu_n(\omega))^2}{2\sigma_n^2(\omega)}}\, dy$$

for all $x \in \mathbb{R}$ and $\omega \in \Omega$.

In view of the regularity, we can (see [439; Chapter II, § 7]) evaluate the expectations $\mathsf{E}(h_n \mid \mathscr{F}_{n-1})(\omega)$ by mere integration (for each fixed $\omega \in \Omega$):

$$\mathsf{E}(h_n \mid \mathscr{F}_{n-1})(\omega) = \int_{-\infty}^{\infty} x \, d\mathsf{P}(h_n \leqslant x \mid \mathscr{F}_{n-1})(\omega),$$

[a] To avoid the consideration of trivial cases in which some special explanations are nevertheless necessary we shall assume throughout that $\sigma_n(\omega) \neq 0$ for all n and ω.

which, in our case, brings us to the formula

$$\mathsf{E}(h_n \mid \mathscr{F}_{n-1}) = \mu_n. \tag{2}$$

In a similar way,

$$\mathsf{D}(h_n \mid \mathscr{F}_{n-1}) = \sigma_n^2. \tag{3}$$

Thus, the 'parameters' μ_n and σ_n^2 have a simple, 'traditional' meaning: they are the *conditinal expectation* and the *conditional variance* of the (conditional) distribution $\mathrm{Law}(h_n \mid \mathscr{F}_{n-1})$.

The distribution $\mathrm{Law}(h_n)$ itself is, therefore, a 'suspension' (a 'mixture') of the conditionally Gaussian distributions $\mathrm{Law}(h_n \mid \mathscr{F}_{n-1})$ averaged over the distribution of μ_n and σ_n^2.

We note that 'mixtures' of normal distributions $\mathscr{N}(\mu, \sigma^2)$ with 'random' parameters $\mu = \mu(\omega)$ and $\sigma^2 = \sigma^2(\omega)$ form a rather broad class of distributions. We shall repeatedly come across particular cases of such distributions in what follows.

Besides the sequence $h = (h_n)$, it is useful to consider the 'standard' conditionally Gaussian sequence $\varepsilon = (\varepsilon_n)_{n \geqslant 1}$ of $\mathscr{F}_n$-measurable random variables such that

$$\mathrm{Law}(\varepsilon_n \mid \mathscr{F}_{n-1}) = \mathscr{N}(0,1), \quad \text{where} \quad \mathscr{F}_0 = \{\varnothing, \Omega\}.$$

This sequence is obviously a *martingale difference* since $\mathsf{E}(\varepsilon_n \mid \mathscr{F}_{n-1}) = 0$. More than that, this is a sequence of *independent* variables with standard normal distribution $\mathscr{N}(0,1)$ because

$$\mathrm{Law}(\varepsilon_n \mid \varepsilon_1, \dots, \varepsilon_{n-1}) = \mathscr{N}(0,1).$$

By our assumption above, $\sigma_n(\omega) \neq 0$ ($n \geqslant 1$, $\omega \in \Omega$), so that the variables $\varepsilon_n \equiv (h_n - \mu_n)/\sigma_n$, $n \geqslant 1$, form a standard Gaussian sequence. Hence we can assume that the conditionally Gaussian (with respect to the flow $(\mathscr{F}_n)$ and the probability P) sequences $h = (h_n), n \geqslant 1$, that we consider can be represented as sums

$$h_n = \mu_n + \sigma_n \varepsilon_n, \tag{4}$$

where $\varepsilon = (\varepsilon_n)$ is a sequence of independent $\mathscr{F}_n$-measurable random variables with standard normal distribution $\mathscr{N}(0,1)$. (As regards the representation of $h = (h_n)$ in the general case, when σ_n may vanish, see [303; Chapter 13, § 1].)

It is clear that to study the probabilistic properties of the sequence $h = (h_n)$ (and therefore of $S = (S_n)$) we must specify further the structure of the variables μ_n and σ_n^2. This is what we do in the models that follow.

We note that in the study of the distributions of $h = (h_n)$, given that we prefer to deal with conditionally Gaussian ones, it is often reasonable to consider this property in the following framework.

Let $(\mathscr{G}_n)$ be a subfiltration of $(\mathscr{F}_n)$, i.e., assume that $\mathscr{G}_n \subseteq \mathscr{F}_n$ and $\mathscr{G}_n \subseteq \mathscr{G}_{n+1}$; for instance, let $\mathscr{G}_n = \mathscr{F}_{n-1}$. We assume now that $\operatorname{Law}(h_n \mid \mathscr{G}_n) = \mathscr{N}(\mu_n, \sigma_n^2)$ with $\mu_n = \mathsf{E}(h_n \mid \mathscr{G}_n)$ and $\sigma_n^2 = \mathsf{D}(h_n \mid \mathscr{G}_n)$. Then, again, the distribution $\operatorname{Law}(h_n)$ is a mixture of Gaussian distributions.

We proceed now to several particular (linear and nonlinear) *Gaussian* and *conditionally Gaussian* models the variables μ_n and σ_n, $n \geqslant 1$, are described more concretely in their dependence on the 'initial values' $(\dots, h_{-1}, h_0)$ and $(\dots, \varepsilon_{-1}, \varepsilon_0)$ of h and ε that must be set separately.

2. AutoRegressive model of order p ($AR(p)$). In this model, it is assumed that

$$\mathscr{F}_n = \sigma(\varepsilon_1, \dots, \varepsilon_n) \tag{5}$$

and

$$\mu_n = a_0 + a_1 h_{n-1} + \cdots + a_p h_{n-p}, \tag{6}$$

$$\sigma_n \equiv \sigma = \text{Const} \quad (\sigma > 0). \tag{7}$$

Thus, here we have

$$h_n = \mu_n + \sigma_n \varepsilon_n = a_0 + a_1 h_{n-1} + \cdots + a_p h_{n-p} + \sigma \varepsilon_n.$$

To define the sequence $h = (h_n)_{n \geqslant 1}$, which is called the *autoregressive model of order* p, we must fix 'initial values', the variables $h_{1-p}, \dots, h_0$. If they are constants, then the sequence $(h_n)_{n \geqslant 1}$ is not merely conditionally Gaussian but truly Gaussian. In § 2b, we consider the properties of the autoregressive model of order one ($p = 1$) in greater detail.

3. Moving Average model ($MA(q)$). In this model, we fix the 'initial values' $(\varepsilon_{1-q}, \dots, \varepsilon_{-1}, \varepsilon_0)$ and set $\mathscr{F}_n = \sigma(\varepsilon_1, \dots, \varepsilon_n)$,

$$\begin{aligned} \mu_n &= b_0 + b_1 \varepsilon_{n-1} + b_2 \varepsilon_{n-2} + \cdots + b_q \varepsilon_{n-q}, \\ \sigma_n &\equiv \sigma = \text{Const} \end{aligned} \tag{8}$$

so that

$$h_n = b_0 + b_1 \varepsilon_{n-1} + b_2 \varepsilon_{n-2} + \cdots + b_q \varepsilon_{n-q} + \sigma \varepsilon_n. \tag{9}$$

4. AutoRegressive Moving Average model of order (p, q) ($ARMA(p, q)$). We set $\mathscr{F}_n = \sigma(\varepsilon_1, \dots, \varepsilon_n)$, define the initial conditions $(\varepsilon_{1-q}, \dots, \varepsilon_{-1}, \varepsilon_0)$ and $(h_{1-p}, \dots, h_{-1}, h_0)$, and also set

$$\mu_n = (a_0 + a_1 h_{n-1} + \cdots + a_p h_{n-p}) + (b_1 \varepsilon_{n-1} + b_2 \varepsilon_{n-2} + \cdots + b_q \varepsilon_{n-q}) \tag{10}$$

and

$$\sigma_n \equiv \sigma = \text{Const}\,.$$

Hence, in the $ARMA(p, q)$ model we obtain

$$h_n = (a_0 + a_1 h_{n-1} + \cdots + a_p h_{n-p}) + (b_1 \varepsilon_{n-1} + b_2 \varepsilon_{n-2} + \cdots + b_q \varepsilon_{n-q}) + \sigma \varepsilon_n. \tag{11}$$

All these models, $AR(p)$, $MA(q)$, and $ARMA(p, q)$, are *linear Gaussian* models (provided that the 'initial values' are, say, constants).

We proceed now to certain interesting *conditionally Gaussian models* that (by contrast to the above ones) are *nonlinear*.

5. AutoRegressive Conditional Heteroskedastic model ($ARCH(p)$). Once again, we assume that the sequence $\varepsilon = (\varepsilon_n)_{n\geqslant 1}$ is the (only) source of randomness and set $\mathscr{F}_n = \sigma(\varepsilon_1, \dots, \varepsilon_n)$. Now let

$$\mu_n = \mathsf{E}(h_n \mid \mathscr{F}_{n-1}) = 0, \tag{12}$$

and

$$\sigma_n^2 = \mathsf{E}(h_n^2 \mid \mathscr{F}_{n-1}) = \alpha_0 + \sum_{i=1}^{p} \alpha_i h_{n-i}^2, \tag{13}$$

where $\alpha_0 > 0$, $\alpha_i \geqslant 0$, $i = 1, \dots, p$, $\mathscr{F}_0 = \{\varnothing, \Omega\}$, and $h_{1-p}, \dots, h_0$ are certain initial constants.

In other words, the conditional variance σ_n^2 is a function of $h_{n-1}^2, \dots, h_{n-p}^2$.

This model, which was introduced (as already mentioned in Chapter I, § 2e) by R. F. Engle [140] (1982), proved to be rather successful in the explanation of several nontrivial properties of financial time series, such as the phenomenon of the *cluster* property for the values of the variables h_n.

Thus,

$$h_n = \sigma_n \varepsilon_n, \qquad n \geqslant 1, \tag{14}$$

where $\varepsilon = (\varepsilon_n)$ is a sequence of independent, normally distributed random variables, $\varepsilon_n \sim \mathscr{N}(0, 1)$, and the σ_n^2 are defined by (13).

If

$$\mu_n = a_0 + a_1 h_{n-1} + \cdots + a_r h_{n-r} \tag{15}$$

in place of (12) and the σ_n^2 satisfy (13), then (4) takes the following form:

$$h_n = a_0 + a_1 h_{n-1} + \cdots + a_r h_{n-r} + \sigma_n \varepsilon_n. \tag{16}$$

Such models are sometimes denoted by $AR(r)/ARCH(p)$.

We now set (assuming that $\mathsf{E}h_n^2 < \infty$)

$$\nu_n = h_n^2 - \sigma_n^2. \tag{17}$$

Then

$$h_n^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i h_{n-i}^2 + \nu_n, \tag{18}$$

where

$$\mathsf{E}(\nu_n \mid \mathscr{F}_{n-1}) = \mathsf{E}(h_n^2 \mid \mathscr{F}_{n-1}) - \sigma_n^2 = 0$$

by (13), i.e., the sequence $\nu = (\nu_n)$ is a *martingale difference.*

Thus, we can regard the $ARCH(p)$ model as the autoregressive model $AR(p)$ for the sequence (h_n^2) with 'noise' $\nu = (\nu_n)$ that is a *martingale difference.*

6. Generalized AutoRegressive Conditional Heteroskedastic model ($GARCH(p,q)$). The successes with the application of $ARCH(p)$ encouraged the development of various generalizations, refinements, and modifications of this model.

The $GARCH(p,q)$ model described below and introduced by T. Bollerslev [48] (1986) is one such modification.

As before, let

$$\mu_n = \mathsf{E}(h_n \mid \mathscr{F}_{n-1}) = 0,$$

but now, in place of (13) we assume that

$$\sigma_n^2 = \mathsf{E}(h_n^2 \mid \mathscr{F}_{n-1}) = \alpha_0 + \sum_{i=1}^{p} \alpha_i h_{n-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{n-j}^2, \tag{19}$$

where $\alpha_0 > 0$, $\alpha_i, \beta_j \geqslant 0$ and where $(h_{1-p}, \dots, h_0)$, $(\sigma_{1-q}^2, \dots, \sigma_0^2)$ are the 'initial values', which we can for simplicity set to be constants.

In the $GARCH(p,q)$ model we set

$$h_n = \sigma_n \varepsilon_n, \tag{20}$$

where $(\varepsilon_1, \varepsilon_2, \dots)$ is a sequence of independent identically distributed random variables with distribution $\mathscr{N}(0,1)$ and the σ_n^2 satisfy (19).

Now let

$$\alpha(L) h_{n-1}^2 = \sum_{i=1}^{p} \alpha_i h_{n-i}^2, \tag{21}$$

where L is the lag operator ($L^i h_{n-1}^2 = h_{(n-1)-i}^2$; see § 2a.2 below), and

$$\beta(L) \sigma_{n-1}^2 = \sum_{j=1}^{q} \beta_j \sigma_{n-j}^2. \tag{22}$$

In this notation

$$\sigma_n^2 = \alpha_0 + \alpha(L) h_{n-1}^2 + \beta(L) \sigma_{n-1}^2.$$

Setting $\nu_n = h_n^2 - \sigma_n^2$ as above, we now obtain

$$\begin{aligned} h_n^2 &= \nu_n + \sigma_n^2 = \nu_n + \alpha_0 + \alpha(L) h_{n-1}^2 + \beta(L)(h_{n-1}^2 - \nu_{n-1}) \\ &= \alpha_0 + \big(\alpha(L) + \beta(L)\big) h_{n-1}^2 - \beta(L)\nu_{n-1} + \nu_n. \end{aligned}$$

In other terms,

$$h_n^2 = \alpha_0 + \big(\alpha(L) + \beta(L)\big) h_{n-1}^2 + \nu_n - \beta(L)\nu_{n-1}. \tag{23}$$

Hence we can regard the $GARCH(p,q)$ model as the *autoregressive moving average model* $ARMA(\max(p,q),q)$ for the sequence (h_n^2) with 'noise' (ν_n) that is a *martingale difference.*

In particular, setting $\nu_n \equiv h_n^2 - \sigma_n^2$ in $ARCH(1)$ with

$$h_n = \sigma_n \varepsilon_n \quad \text{and} \quad \sigma_n^2 = \alpha_0 + \alpha_1 h_{n-1}^2$$

we see that

$$h_n^2 = \alpha_0 + \alpha_1 h_{n-1}^2 + \nu_n,$$

where the 'noise' (ν_n) is a martingale difference.

There exist various generalizations of the $ARCH$ and $GARCH$ models (e.g., $EGARCH$, $AGARCH$, $STARCH$, $NARCH$, $MARCH$, $HARCH$) related, in the final analysis, to one or another description of the variables $\sigma_n^2 = \mathsf{E}(h_n^2 \mid \mathscr{F}_{n-1})$ as measurable functions with respect to the σ-algebras $\mathscr{F}_{n-1} = \sigma(\varepsilon_1, \ldots, \varepsilon_{n-1})$.

7. Stochastic volatility model. In the previous models, we had a *single* source of randomness, a Gaussian sequence of independent variables $\varepsilon = (\varepsilon_n)$. The *stochastic volatility* models involve *two* sources of randomness, $\varepsilon = (\varepsilon_n)$ and $\delta = (\delta_n)$. In the most elementary case they are assumed to be independent standard Gaussian sequences, that is, sequences of independent, $\mathscr{N}(0,1)$-distributed random variables.

Let $\mathscr{G}_n = \sigma(\delta_1, \ldots, \delta_n)$. We set

$$h_n = \sigma_n \varepsilon_n, \tag{24}$$

where the σ_n are $\mathscr{G}_n$-measurable. Then it is clear that

$$\operatorname{Law}(h_n \mid \mathscr{G}_n) = \mathscr{N}(0, \sigma_n^2), \tag{25}$$

i.e., the $\mathscr{G}_n$-conditional distribution of h_n is Gaussian, with parameters 0 and σ_n^2.

We now set

$$\sigma_n = e^{\frac{1}{2}\Delta_n}. \tag{26}$$

Then $\sigma_n^2 = e^{\Delta_n}$, where the Δ_n are $\mathscr{G}_n$-measurable. There exist highly popular models in which the sequence (Δ_n) is governed by an autoregressive model (we shall write $(\Delta_n) \in AR(p)$) as follows:

$$\Delta_n = a_0 + a_1 \Delta_{n-1} + \cdots + a_p \Delta_{n-p} + c\delta_n.$$

A natural generalization of (24) is the scheme with

$$h_n = \mu_n + \sigma_n \varepsilon_n, \tag{27}$$

where μ_n and σ_n are $\mathscr{G}_n$-measurable.

If $\varepsilon = (\varepsilon_n)$ in (27) is a normally distributed *stationary* sequence with $\mathsf{E}\varepsilon_n = 0$ and $\mathsf{E}\varepsilon_n^2 = 1$, while $\sigma = (\sigma_n)$ is independent of $\varepsilon = (\varepsilon_n)$, then we arrive at the so-called *Taylor model.*

We complete our *brief* discussion of several Gaussian and conditionally Gaussian models used in financial mathematics and financial engineering at this point. We study the properties of these models *more closely* below, in § § 2 and 3.

§ 1e. Binomial Model of Price Evolution

1. An exceptional role in probability theory is played by the *Bernoulli scheme*, the sequence

$$\delta = (\delta_1, \delta_2, \dots)$$

of independent identically distributed random variables that take only two values, say, 1 and 0, with probabilities p and q, $p + q = 1$.

It was for this scheme that the first limit theorem of probability theory, the *Law of large numbers*, was proved (J. Bernoulli "Ars Conjectandi", 1713). It states that for each $\varepsilon > 0$,

$$\mathsf{P}\left(\left|\frac{S_n}{n} - p\right| > \varepsilon\right) \to 0 \quad \text{as} \quad n \to \infty,$$

where $\dfrac{S_n}{n}$ $\left(= \dfrac{\delta_1 + \cdots + \delta_n}{n}\right)$ is the frequency of 'ones' in the sequence $\delta_1, \dots, \delta_n$.

Likewise, it was for this scheme that several other remarkable results of probability theory (the *de Moivre–Laplace limit theorem*, the *Strong law of large numbers*, the *Law of the iterated logaritm*, the *Arcsine law*, ...) were first proved. Later on, all these results turned out to have a much broader range for applications.

The role in financial mathematics of the *Cox–Ross–Rubinstein binomial model* [82] described below is close in this sense to that of the Bernoulli scheme in classical probability theory; very simple as it is, this model enables complete calculations of many financial characteristics and instruments: rational option prices, hedging strategies, and so on (see Chapter VII below).

2. We assume that all the financial operations proceed on a (B, S)-*market* formed a *bank account* $B = (B_n)_{n \geqslant 0}$ and some *stock* of value $S = (S_n)_{n \geqslant 0}$.

We can represent the evolution of B and S as follows:

$$B_n = (1 + r_n) B_{n-1}, \tag{1}$$

$$S_n = (1 + \rho_n) S_{n-1}, \tag{2}$$

or, equivalently,

$$\Delta B_n = r_n B_{n-1},$$
$$\Delta S_n = \rho_n S_{n-1}$$

where $B_0 > 0$ and $S_0 > 0$.

The main *distinction* between a bank account and stock is the fact that the bank interest rate

$$r_n \quad \text{is} \quad \mathscr{F}_{n-1}\text{-measurable},$$

while the 'market' interest on the stock

$$\rho_n \quad \text{is} \quad \mathscr{F}_n\text{-measurable},$$

where $(\mathscr{F}_n)$ is the filtration (information flow) on a given probability space $(\Omega, \mathscr{F}, \mathsf{P})$.

In the framework of the *Cox–Ross–Rubinstein binomial model* (the 'CRR-model') of a *(B,S)-market* one assumes that

$$r_n \equiv r = \text{Const}$$

and that $\rho = (\rho_n)_{n\geqslant 1}$ is a Bernoulli sequence of independent identically distributed random variables $\rho_1, \rho_2, \dots$ taking two values, a or b, $a < b$.

We can write each ρ_n as the following sum:

$$\rho_n = \frac{a+b}{2} + \frac{b-a}{2}\,\varepsilon_n, \tag{3}$$

or as

$$\rho_n = a + (b-a)\delta_n, \tag{4}$$

to find out that

$$\rho_n = \begin{cases} b \\ a \end{cases} \iff \varepsilon_n = \begin{cases} +1 \\ -1 \end{cases} \iff \delta_n = \begin{cases} +1 \\ 0 \end{cases}.$$

Our assumptions that $r_n \equiv \text{Const}$, while the ρ_n take only two values, enable us to set the original probability space Ω from the very beginning to be the space of *binary* sequences:

$$\Omega = \{a,b\}^\infty \quad \text{or} \quad \Omega = \{-1,1\}^\infty, \quad \text{or} \quad \Omega = \{0,1\}^\infty.$$

By (2),

$$S_n = S_0 \prod_{k\leqslant n} (1+\rho_k). \tag{5}$$

Comparing this expression and (5) in § 1a we see that the ρ_k play the role of the variables $\widehat{h}_k$ introduced there. Clearly, we can also represent the S_n as follows:

$$S_n = S_0 e^{H_n} = S_0 e^{h_1+\cdots+h_n},$$

where $h_n = \ln(1+\rho_k)$ (cf. (1) and (10) in § 1a).

3. We now discuss the particular case of a and b satisfying the relations

$$a = \lambda^{-1} - 1, \qquad b = \lambda - 1$$

for some $\lambda > 1$.

In this case

$$S_n = \begin{cases} \lambda S_{n-1} & \text{if } \rho_n = b, \\ \lambda^{-1} S_{n-1} & \text{if } \rho_n = a. \end{cases} \tag{6}$$

Regarding (3) as the definition of $\varepsilon_n (= \pm 1)$, we can represent the S_n as follows:

$$S_n = S_0 \lambda^{\varepsilon_1 + \cdots + \varepsilon_n}, \tag{7}$$

which is the same as

$$S_n = S_0 e^{h_1 + \cdots + h_n} \tag{8}$$

with $h_k = \varepsilon_k \ln \lambda$.

This random sequence $S = (S_n)$ is called a *geometric random walk* on the set

$$E_{S_0} = \{S_0 \lambda^k : k = 0, \pm 1, \dots\}.$$

If $S_0 \in E = \{\lambda^k : k = 0, \pm 1, \dots\}$, then $E_{S_0} = E$. In this case $S = (S_n)$ is called a *Markov walk* on the phase space $E = \{\lambda^k : k = 0, \pm 1, \dots\}$.

The binomial model in question is a discrete analogue of a *geometric Brownian motion* $S = (S_t)_{t \geqslant 0}$, i.e., of a random process having the following representation (cf. (8)):

$$S_t = S_0 e^{\sigma W_t + (\mu - \sigma^2/2)t},$$

where $W = (W_t)_{t \geqslant 0}$ is a standard Wiener process or a standard Brownian motion (see Chapter III, § 3a).

It seems appropriate to recall in this connection that it is an *arithmetic* random walk $S_n = S_{n-1} + \xi_n$ with a Bernoulli sequence $\xi = (\xi_n)_{n \geqslant 1}$ that is a discrete counterpart to *usual* Brownian motion.

4. We started our previous discussion with the assumption that all considerations proceed in the framework of some filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_n), \mathsf{P})$ with probability measure. The question of the properties of this probability measure P and, therefore, of the quantities $p = \mathsf{P}(\rho_n = b)$ and $q = \mathsf{P}(\rho_n = a)$ is not so easy in general. It would be more realistic in a certain sense to assume that there is an entire family $\mathscr{P} = \{\mathsf{P}\}$ of probability measures in $(\Omega, \mathscr{F})$, rather than a single one, and that the corresponding values of $p = \mathsf{P}(\rho_n = b)$ belong to the interval $(0, 1)$.

As regards the issue of possible generalizations of the binomial model just discussed, it would be reasonable to assume that the variables ρ_n can take all values in the *interval* $[a, b]$ rather than only the two values a and b. In that case, the probability distribution for ρ_n can be an arbitrary distribution on $[a, b]$. This is the model we shall consider in Chapter V, § 1c in connection with the calculations of rational option prices on so-called *incomplete* markets. In the same section we shall also consider a *nonprobabilistic* approach based on the assumption that the ρ_n are 'chaotic' variables. (As regards the description of the evolution of prices involving models of 'dynamical chaos', see § § 4a, b below.)

§ 1f. Models with Discrete Intervention of Chance

1. In the study of sequences $H = (H_n)_{n\geqslant 0}$, it is often convenient to 'embed' them (in the sense that we explain below) in certain schemes with continuous time.

Given a stochastic basis $\mathscr{B} = (\Omega, \mathscr{F}, (\mathscr{F}_n)_{n\geqslant 0}, \mathsf{P})$ we consider an associated basis $\widetilde{\mathscr{B}}$ with continuous time ($t \geqslant 0$) defines as follows:

$$\widetilde{\mathscr{B}} = \bigl(\Omega, \mathscr{F}, (\widetilde{\mathscr{F}}_t)_{t\geqslant 0}, \mathsf{P}\bigr),$$

where $\widetilde{\mathscr{F}}_t = \mathscr{F}_{[t]}$ and $[t]$ is the entire part of t.

Given a stochastic sequence $H = (H_n, \mathscr{F}_n)$, we introduce the process $\widetilde{H} = (\widetilde{H}_t, \widetilde{\mathscr{F}}_t)$ (with continuous time) by setting[b]

$$\widetilde{H}_t = H_{[t]}.$$

Thus (see Fig. 13), the trajectories $\widetilde{H}_t$, $t \geqslant 0$, are piecewise constant and right-continuous, making jumps $\Delta\widetilde{H}_t \equiv \widetilde{H}_t - \widetilde{H}_{t-}$ for $t = 1, 2, \dots$. Moreover, $\Delta\widetilde{H}_n = \Delta H_n = h_n$.

Clearly, the converse is also true: a random process $\widetilde{H} = (\widetilde{H}_t, \widetilde{\mathscr{F}}_t)_{t\geqslant 0}$ on a probability space $\widetilde{\mathscr{B}} = \bigl(\Omega, \mathscr{F}, (\widetilde{\mathscr{F}}_t)_{t\geqslant 0}, \mathsf{P}\bigr)$ that has piecewise constant right-continuous trajectories making jumps for $t = 1, 2, \dots$, is actually a process with discrete time of the above form.

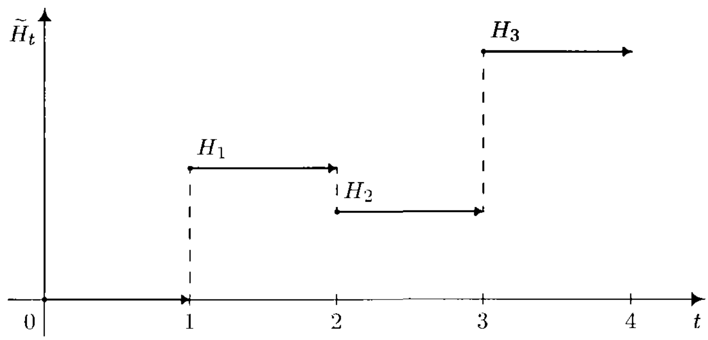


FIGURE 13. Embedding a sequence (H_n) with discrete time in a scheme with continuous time

2. So far, in our discussion of various models for the dynamics of prices we have either (in most cases) considered models with prices $S = (S_n)$ registered at discrete instants of time $n = 0, 1, \dots$, or (as in the Bachelier case) models with prices

[b] As in the dicrete-time case, we write $\widetilde{H} = (\widetilde{H}_t, \widetilde{\mathscr{F}}_t)$ to indicate that the variables $\widetilde{H}_t$ are $\widetilde{\mathscr{F}}_t$-measurable for each $t \geqslant 0$.

$S = (S_t)_{t \geqslant 0}$ described by a continuous random process (e.g., the Brownian motion) with continuous time $t \geqslant 0$.

In practice, the statistical analysis (see Chapter IV) of the evolution of actual prices $S = (S_t)_{t \geqslant 0}$ shows that they have a hybrid structure in a certain sense.

More precisely, this means the following. According to a large stock of data, the trajectories of the prices $S = (S_t)_{t \geqslant 0}$ look like in the picture below:

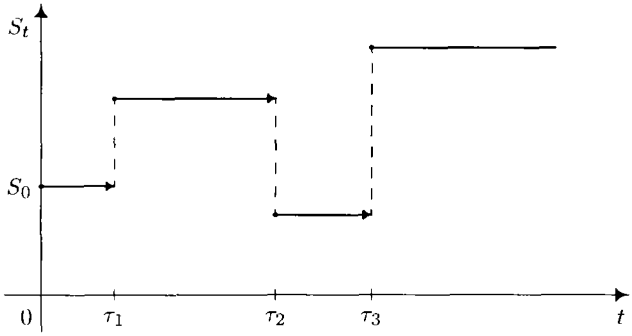


FIGURE 14. A process with discrete intervention of chance (at times $\tau_1, \tau_2, \tau_3, \dots$)

In other words, representing S_t by the formula $S_t = S_0 e^{H_t}$ we see that

$$H_t = \sum_{k \geqslant 1} h_k I(\tau_k \leqslant t), \tag{1}$$

where $\tau_1, \tau_2, \dots$ are the instants of jumps and h_k are their amplitudes $(\Delta H_{\tau_k} \equiv H_{\tau_k} - H_{\tau_k -} = h_k)$.

Assuming that all our considerations proceed with respect to a stochastic basis $(\Omega, \mathscr{F}, (\widetilde{\mathscr{F}}_t)_{t \geqslant 0}, \mathsf{P})$, it would be natural to impose certain 'measurability' conditions on the (random) instants of jumps τ_k, $k \geqslant 1$, and the variables h_k, $k \geqslant 1$, so as to ensure, at any rate, that H_t be $\mathscr{F}_t$-measurable for each $t \geqslant 0$, i.e., can be defined by the 'data' accessible to an observer over the period $[0, t]$.

3. To this end we introduce several definitions in which by 'an *extended* random variable' $\tau = \tau(\omega)$ we mean a map[c]

$$\Omega \to \overline{\mathbb{R}}_+ \equiv [0, \infty].$$

[c]Recall that, in accordance with the traditional probabilistic definitions, a *real-valued* random variable can take only *finite* values, i.e., values in $\mathbb{R} = (-\infty, \infty)$.

DEFINITION 1. We call a nonnegative random value $\tau = \tau(\omega)$ a *Markov time* or a *random valriable independent of the 'future'* if

$$\{\omega\colon \tau(\omega) \leqslant t\} \in \mathscr{F}_t \tag{2}$$

for each $t \geqslant 0$.

Markov times are also called *stopping times*, though sometimes one reserves the latter term for the Markov times $\tau = \tau(\omega)$ such that either $\tau(\omega) < \infty$ for all $\omega \in \Omega$, or $\mathsf{P}(\tau(\omega) < \infty) = 1$.

The meaning of condition (2) becomes fairly obvious if we interpret $\tau(\omega)$ as the instant when one must make certain 'decisions' (e.g., to buy or to sell stock). Since $\mathscr{F}_t$ is a σ-algebra, (2) is equivalent to the condition $\{\tau(\omega) > t\} \in \mathscr{F}_t$, meaning that one's intention at time t to postpone this 'decision' until later is determined by the information $\mathscr{F}_t$ accessible over the period $[0,t]$, and one cannot take into consideration the 'future' (i.e., what might happen after time t).

DEFINITION 2. Let $\tau = \tau(\omega)$ be some stopping time. Then we denote by $\mathscr{F}_\tau$ the collection of sets $A \in \mathscr{F}$ such that

$$A \cap \{\omega\colon \tau(\omega) \leqslant t\} \in \mathscr{F}_t \tag{3}$$

for each $t \in \mathbb{R}_+ = [0,\infty)$.

By $\mathscr{F}_{\tau-}$ we denote the σ-algebra generated by $\mathscr{F}_0$ and all the sets $A \cap \{\omega\colon \tau(\omega) < t\}$, where $A \in \mathscr{F}_t$ and $t \in \mathbb{R}_+$.

It is easy to see that $\mathscr{F}_\tau$ is a σ-algebra.

If we interpret $\mathscr{F}_t$ as the collection of all events occurred not later than time t, then it is reasonable to interpret $\mathscr{F}_\tau$ and $\mathscr{F}_{\tau-}$ as the collections of events observable over the periods $[0,\tau]$ and $[0,\tau)$, respectively.

4. We now turn back to the representation (1) of the process $H = (H_t)_{t\geqslant 0}$ as a *process with discrete intervention of chance* (at instants of time $\tau_1, \tau_2, \dots$; see Fig. 14). We assume that

$$0 < \tau_1(\omega) < \tau_2(\omega) < \cdots$$

for all (or P-almost all) $\omega \in \Omega$ and that $\tau_k = \tau_k(\omega)$ is a stopping time for each $k \geqslant 1$, while the variables $h_k = h_k(\omega)$ are $\mathscr{F}_{\tau_k}$-measurable.

It follows, in particular, from these assumptions and conditions (2) and (3) that the process $H = (H_t)_{t\geqslant 0}$ is *adapted* to the flow $(\mathscr{F}_t)_{t\geqslant 0}$, i.e.,

$$H_t \quad \text{is} \quad \mathscr{F}_t\text{-measurable}$$

for each $t \geqslant 0$.

Thus, in accordance with the above conventions, we can write $H = (H_t, \mathscr{F}_t)_{t\geqslant 0}$.

The real-valued process $H = (H_t)_{t\geqslant 0}$ defined by (1) is a particular case of so-called *multivariate point processes* ranging in the phase space $\mathbb{E}$ ([250]; in our case we have $\mathbb{E} = \mathbb{R}$), while the term '*point* (or *counting*) *process*' in the proper sense is related to the case of $h_n \equiv 1$, i.e., to the process

$$N_t = \sum_k I(\tau_k \leqslant t), \qquad t \geqslant 0. \tag{4}$$

Remark. Sometimes, one relates 'point' processes only to a sequence of instants $\tau = (\tau_1, \tau_2, \dots)$ and reserves the attribute 'counting' for the process $N = (N_t)_{t\geqslant 0}$, corresponding to this sequence in accordance with formula (4). It is clear that there exists a one-to-one correspondence between τ and N: we can determine N by τ in accordance with (4), and we can determine τ by N because

$$\tau_k = \inf\{t\colon N_t = k\}.$$

We note that, as usual, we set here $\tau_k(\omega)$ to be equal to $+\infty$ if $\{t\colon N_t = k\} = \varnothing$. For instance, considering the stopping times corresponding to the 'one-point' point process with trajectories as in Fig. 15 we can say that $\tau_2 = \tau_3 = \cdots = +\infty$.

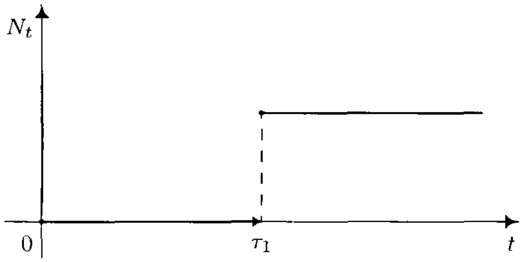


FIGURE 15. 'One-point' point process

5. The point processes $N = (N_t)_{t\geqslant 0}$ defined in (4) can be described in an equivalent way using the concept of a *random change of time* [250; Chapter II, § 3b].

DEFINITION 3. We say that a family of random variables $\sigma = (\sigma_t)_{t\geqslant 0}$ defined on a stochastic basis $\mathscr{B} = (\Omega, \mathscr{F}, (\mathscr{F}_n)_{n\in\mathbb{N}}, \mathsf{P})$ and ranging in $\mathbb{N} = \{0, 1, \dots\}$ or $\overline{\mathbb{N}} = \{0, 1, \dots, \infty\}$ is a *random change of time* if

(i) the σ_t are stopping times (with respect to $(\mathscr{F}_n)_{n\in\mathbb{N}}$) for all t;
(ii) $\sigma_0 = 0$;
(iii) each trajectory $\sigma_t(\omega)$, $t \geqslant 0$, is increasing, right-continuous, with jumps of amplitude one.

We now associate with $\sigma = (\sigma_t)_{t\geqslant 0}$ the stochastic basis $\mathscr{B}^\sigma$

$$\mathscr{B}^\sigma = (\Omega, \mathscr{F}, (\mathscr{F}_{\sigma_t})_{t\geqslant 0}, \mathsf{P})$$

with continuous time and set

$$\tau_k = \inf\{t : \sigma_t \geqslant k\}, \qquad k \in \mathbb{N}.$$

It is easy to show that the $\tau_k = \tau_k(\omega)$ are stopping times with respect to the flow $(\mathscr{G}_t)_{t\geqslant 0}$ such that $\mathscr{G}_t = \mathscr{F}_{\sigma_t}$. Moreover, if $\sigma_t < \infty$ for all $t \in \mathbb{N}$, then

$$H_t = \sum_{k\geqslant 1} h_k I(\tau_k \leqslant t) = \sum_{1\leqslant k\leqslant \sigma_t} h_k. \tag{5}$$

It is also clear that

$$\sigma_t = \sum_{k\geqslant 1} I(\tau_k \leqslant t),$$

i.e., σ_t is the number of decisions taken over the period $[0, t]$. To put it another way, the random change of time $\sigma = (\sigma_t)_{t\geqslant 0}$ constructed by the sequence $(\tau_1, \tau_2, \dots)$ is just the *counting* process $N = (N_t)$ associated with this sequence, i.e., the process

$$N_t = \sum_{k\geqslant 1} I(\tau_k \leqslant t),$$

In all the above formulas t plays the role of real ('physical') time, while σ_t plays the role of *operational* time counting the 'decisions' taken in 'physical' time t. (We return to the issue of operational time in the empirical analysis of financial time series in Chapter IV, § 3d.)

We note also that if

$$\sigma_t = [t],$$

then $\mathscr{B}^\sigma = \widetilde{\mathscr{B}}$, where $\widetilde{\mathscr{B}}$ is the above-mentioned stochastic basis with discrete time $(\Omega, \mathscr{F}, (\widetilde{\mathscr{F}}_t)_{t\geqslant 0}, \mathsf{P})$, where $\widetilde{\mathscr{F}}_t = \mathscr{F}_{[t]}$.

6. As follows from § 1b, the Doob decomposition plays a key role in the stochastic analysis of the sequence $H = (H_n)_{n\geqslant 0}$ by enabling one to distinguish the 'martingale' and the 'predictable' components of H in their dependence on the flow $(\mathscr{F}_n)_{n\geqslant 0}$ of incoming data (on the market situation, in the financial framework).

A similar decomposition (the Doob–Meyer decomposition) can be constructed both for the counting process N and for the multivariate point process H. This (as in the discrete-time case) is the main starting point in the stochastic analysis of these processes involving the concepts of 'marginal' and 'predictability' (see Chapter III, § 5b and, in greater detail, [250]).

2. Linear Stochastic Models

The empirical analysis of the evolution of financial (and, of course, many other: economic, social, and so on) indexes, characteristics, ... must start with constructing an appropriate probabilistic, statistical (or some other) model; its proper choice is a rather complicated task.

The *General theory of time series* has various 'standard' *linear* models in its store. The first to mention are $MA(q)$ (the *moving average model of order* q), $AR(p)$ (the *autoregressive model of order* p), and $ARMA(p, q)$ (the *mixed autoregressive moving average model of order* (p, q)) considered in § 1d. These models are thoroughly studied in the theory or time series, especially if one assumes that they are *stationary*.

The reasons for the popularity of these models are their simplicity on the one hand and, on the other, the fact that, even with few parameters involved, these models deliver good approximations of stationary sequences in a rather broad class.

However, the class of 'econometric' time series is far from being exhausted by stationary series. Often, one can fairly clearly discern the following *three* ingredients of statistical data:

- a *slowly changing* (*e.g.*, '*inflationary*') *trend component* (x);
- *periodic or aperiodic cycles* (y),
- an *irregular*, *fluctuating* ('*stochastic*' or '*chaotic*') *component* (z).

These components can be combined in the observable data (h) in diverse ways. Schematically, we shall write this down as

$$h = x * y * z,$$

where the symbol '$*$' can denote, e.g., addition '$+$', multiplication'$\times$', and so on.

There are many monographs devoted to the theory of time series and their applications to the analysis of financial data (see, e.g., [62], [193], [202], [211], [212], [351], or [460].)

In what follows, we discuss several *linear* (and after that, nonlinear) models with the intention to give an idea of their structure, peculiarities, and properties used in the empirical analysis of financial data.

It should be mentioned here that, in the long run, one of the important objectives in the empirical analysis of the statistics of financial indexes is to *predict*, *forecast* the 'future dynamics of prices':

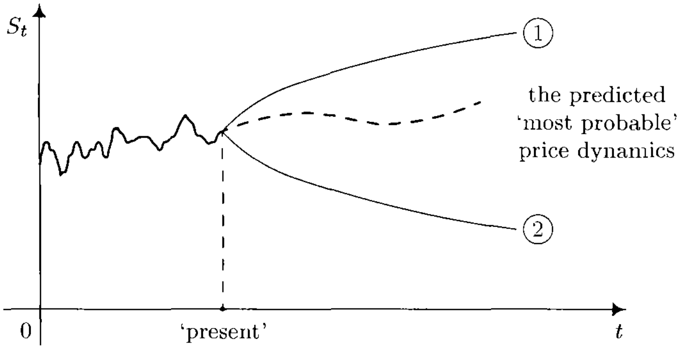


FIGURE 16. The region between the curves 1 and 2 is the confidence domain (corresponding to a certain degree of confidence), in which the prices are supposed to evolve in the future

The reliability of these predictions depends, of course, on a successful choice of a model, the precision in the evaluation of the parameters of this model, and the quality of extrapolation (linear or nonlinear).

The analysis of time series that we carry out in Chapter IV is expressive in that respect. We show there how, starting from *simple* linear Gaussian models, one is forced to *modify* them, make them *more complex* in order to obtain, finally, a model 'capturing' the phenomena discovered by the empirical analysis (e.g., the failure of the Gaussian property, the 'cluster' property, and the effect of 'long memory' in prices).

Recalling our scheme '$h = x * y * z$' of the interplay between the three ingredients x, y, and z of the prices h, we can say that most weight in our discussion will be put on the last, 'fluctuating' component z, and then (in the case of exchange rates) on the periodic (seasonal) component y.

We shall not discuss the details of the analysis of the trend component x, but we point out that it is often responsible for the 'nonstationary' behavior of the models under consideration. A classical example is provided by the so-called $ARIMA(p,d,q)$ models (*AutoRegressive Integrated Moving Average* models; here d is the order of integration) that are widely used and promoted by G. E. P. Box and G. M. Jenkins ([53]; see § 2c below for greater detail).

§ 2a. Moving Average Model $MA(q)$

1. It is assumed in all the models that follow (either linear or nonlinear) that we have a certain '*basic*' sequence $\varepsilon = (\varepsilon_n)$, which is assumed to be *white noise* (see Fig. 17) in the theory of time series and is identified with the *source of randomness* responsible for the stochastic behavior of the statistical objects in question.

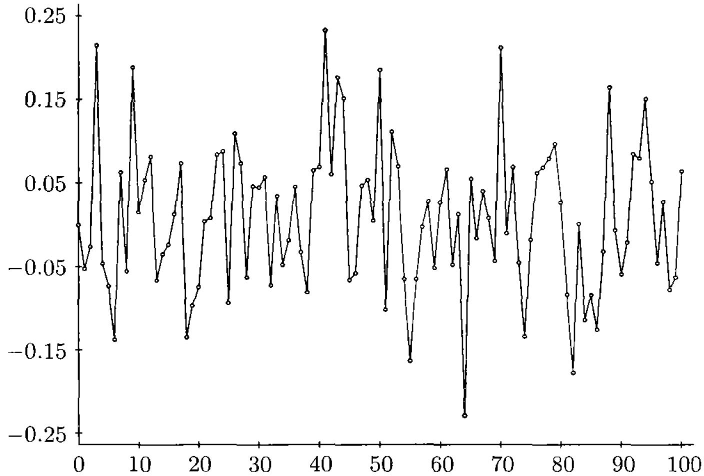


FIGURE 17. Computer simulation of 'white noise' $h_n = \sigma\varepsilon_n$ with $\sigma = 0.1$ and $\varepsilon_n \sim \mathcal{N}(0, 1)$

Alternatively, it is assumed (in the 'L^2-theory') that a sequence $\varepsilon = (\varepsilon_n)$ is *white noise in the wide sense*, i.e., $\mathsf{E}\varepsilon_n = 0$, $\mathsf{E}\varepsilon_n^2 < \infty$, and

$$\mathsf{E}\varepsilon_n\varepsilon_m = 0 \tag{1}$$

for all $n \neq m$. (It is convenient to assume at this point that the time parameter n can take the values $0, \pm 1, \pm 2, \dots$.)

In other words, *white noise in the wide sense* is a square integrable sequence of uncorrelated random variables with zero expectations.

If we require additionally that these variables be *Gaussian* (normal), then we call the sequence $\varepsilon = (\varepsilon_n)$ *white noise in the strict sense* or simply *white noise.* This is equivalent to the condition that $\varepsilon = (\varepsilon_n)$ is a sequence of independent normally distributed ($\varepsilon_n \sim \mathcal{N}(0, \sigma_n^2)$) random variables. In what follows we assume that $\sigma_n^2 \equiv 1$. (One says in this case that $\varepsilon = (\varepsilon_n)$ is a *standard* Gaussian sequence; it is instructive to compare this concept with that of a fractal Gaussian noise, which is also used in the financial data statistics, see Chapter III, § 2d.)

2. In the *moving average* scheme $MA(q)$ describing the evolution of the sequence $h = (h_n)$, one assumes that the h_n can be constructed from the 'basic' sequence ε

as follows:

$$h_n = (\mu + b_1\varepsilon_{n-1} + \cdots + b_q\varepsilon_{n-q}) + b_0\varepsilon_n. \tag{2}$$

Here q is a parameter describing the degree of our dependence on the 'past' while ε_n 'updates' the information contained in the σ-algebra $\mathscr{F}_{n-1} = \sigma(\varepsilon_{n-1}, \varepsilon_{n-2}, \dots)$; see Fig. 18.

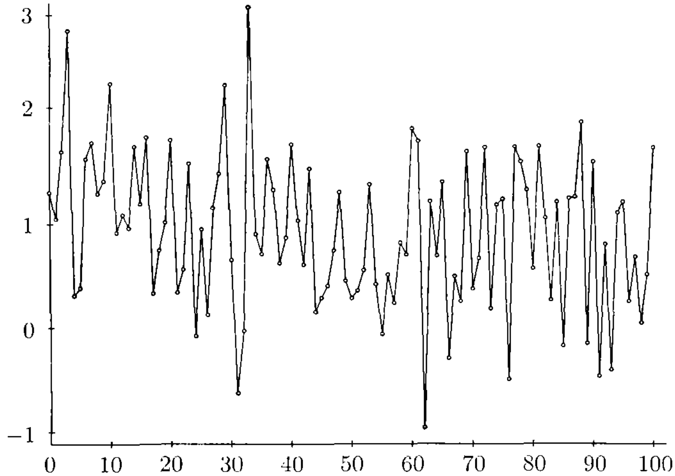


FIGURE 18. Computer simulation a sequence $h = (h_n)$ governed by the $MA(1)$ model with $h_n = \mu + b_1\varepsilon_{n-1} + b_0\varepsilon_n$, where $\mu = 1$, $b_1 = 1$, $b_0 = 0.1$, and $\varepsilon_n \sim \mathscr{N}(0, 1)$

For brevity, it is convenient to introduce the lag operator L acting on number sequences $x = (x_n)$ by the formula

$$Lx_n = x_{n-1}. \tag{3}$$

Since $L(Lx_n) = Lx_{n-1} = x_{n-2}$, it is natural to use the notation L^2 for the operator

$$L^2x_n = x_{n-2}.$$

In general, we set $L^kx_n = x_{n-k}$ for all $k \geqslant 0$.

We point out the following useful, albeit simple, properties of the operator L (here c, c_1, and c_2 are arbitrary constants):

$$\begin{aligned}
L(cx_n) &= cLx_n,\\
L(x_n + y_n) &= Lx_n + Ly_n,\\
(c_1L + c_2L^2)x_n &= c_1Lx_n + c_2L^2x_n = c_1x_{n-1} + c_2x_{n-2},\\
(1 - \lambda_1L)(1 - \lambda_2L)x_n &= x_n - (\lambda_1 + \lambda_2)x_{n-1} + (\lambda_1\lambda_2)x_{n-2}.
\end{aligned}$$

Using L we can write (2) in the following compact form:

$$h_n = \mu + \beta(L)\varepsilon_n, \tag{4}$$

where

$$\beta(L) = b_0 + b_1 L + \cdots + b_q L^q.$$

3. We now consider the question on the *probabilistic characteristics* of the sequence $h = (h_n)$.

Let $q = 1$. Then

$$h_n = \mu + b_0\varepsilon_n + b_1\varepsilon_{n-1}, \tag{5}$$

and we immediately see that

$$\mathsf{E}h_n = \mu, \qquad \mathsf{D}h_n = b_0^2 + b_1^2, \tag{6}$$

$$\mathsf{Cov}(h_n, h_{n+1}) = b_0 b_1, \qquad \mathsf{Cov}(h_n, h_{n+k}) = 0, \quad k > 1. \tag{7}$$

The last two properties mean that $h = (h_n)$ is a sequence with correlated consecutive elements (h_n and h_{n+1}), while for $k \geqslant 2$ the correlation between the h_n and the h_{n+k} is zero.

We note also that if $b_0 b_1 > 0$, then the elements h_n and h_{n+1} have *positive* correlation, while if $b_0 b_1 < 0$, then their correlation is *negative*. (We come across a similar situation in Chapter IV, § 3c, when explaining the phenomenon of negative correlation in the case of exchange rates.)

By (6) and (7), the elements of $h = (h_n)$ have mean values, variances, and covariances independent of n. (Of course, this is already incorporated into our assumption that $\varepsilon = (\varepsilon_n)$ is *white noise in the wide sense* with $\mathsf{E}\varepsilon_n = 0$ and $\mathsf{D}\varepsilon_n^2 = 1$ and that the coefficients in (5) are independent of n.) Thus, the sequence $h = (h_n)$ is (just by definition) stationary in the wide sense. If we assume in addition that $\varepsilon = (\varepsilon_n)$ is a Gaussian sequence, then the sequence $h = (h_n)$ is also Gaussian and, therefore, all its probabilistic properties can be expressed in terms of expectation, variance, and covariance. In this case, the sequence $h = (h_n)$ is stationary in the strict sense, i.e.,

$$\mathrm{Law}(h_{i_1}, \ldots, h_{i_n}) = \mathrm{Law}(h_{i_1+k}, \ldots, h_{i_n+k})$$

for all $n \geqslant 1$, $i_1, \ldots, i_n$, and arbitrary k.

4. Let a trajectory $(h_1, \ldots, h_n)$ be a result of observations of the variables h_k at instants of time $k = 1, \ldots, n$ and let

$$\overline{h}_n = \frac{1}{n}\sum_{k=1}^{n} h_k \tag{8}$$

be the time average.

From the statistical viewpoint, the point of considering the 'statistics' $\overline{h}_n$ lies in the fact that belief that it is a natural 'choice' for the role of an estimator of the expectation μ.

If we measure the quality of this estimator by the value of the standard deviation $\Delta_n^2 = \mathsf{E}\,|\overline{h}_n - \mu|^2$, then the following test is useful: as $n \to \infty$ we have

$$\Delta_n^2 \to 0 \iff \frac{1}{n}\sum_{k=1}^{n} \mathsf{R}(k) \to 0, \tag{9}$$

where $\mathsf{R}(k) = \mathsf{Cov}(h_n, h_{n+k})$ and $h = (h_n)$ is an arbitrary sequence that is stationary in the wide sense.

Indeed, let (for simplicity) $\mu = \mathsf{E}h_n = 0$. Then

$$\left|\frac{1}{n}\sum_{k=1}^{n} \mathsf{R}(k)\right|^2 = \left|\mathsf{E}\left(\frac{1}{n}\sum_{k=1}^{n} h_k\right)h_0\right|^2 \leqslant \mathsf{E}h_0^2\,\mathsf{E}\left|\frac{1}{n}\sum_{k=1}^{n} h_k\right|^2,$$

therefore the implication $\Longrightarrow$ in (9) holds.

On the other hand,

$$\begin{aligned}\mathsf{E}\left|\frac{1}{n}\sum_{k=1}^{n} h_k\right|^2 &= \frac{1}{n^2}\mathsf{E}\left(\sum_{k=1}^{n} h_k^2 + 2\sum_{k\neq l}^{n} h_k h_l\right)\\ &= \frac{2}{n^2}\sum_{l=1}^{n}\sum_{k=0}^{l-1}\mathsf{R}(k) - \frac{1}{n}\mathsf{R}(0).\end{aligned}$$

We now choose $\delta > 0$ and we choose $n = n(\delta)$ such that

$$\left|\frac{1}{l}\sum_{k=0}^{l-1}\mathsf{R}(k)\right| \leqslant \delta$$

for all $l \geqslant n(\delta)$. Then

$$\begin{aligned}\left|\frac{1}{n^2}\sum_{l=1}^{n}\sum_{k=0}^{l-1}\mathsf{R}(k)\right| &= \left|\frac{1}{n^2}\sum_{l=1}^{n(\delta)}\sum_{k=0}^{l-1}\mathsf{R}(k) + \frac{1}{n^2}\sum_{l=n(\varepsilon)+1}^{n}\sum_{k=0}^{l-1}\mathsf{R}(k)\right|\\ &\leqslant \frac{1}{n^2}\left|\sum_{l=1}^{n(\delta)}\sum_{k=0}^{l-1}\mathsf{R}(k)\right| + \frac{1}{n^2}\left|\sum_{l=n(\delta)+1}^{n} l\cdot\frac{1}{l}\sum_{k=0}^{l-1}\mathsf{R}(k)\right|\\ &\leqslant \frac{1}{n^2}\left|\sum_{l=1}^{n(\delta)}\sum_{k=0}^{l-1}\mathsf{R}(k)\right| + \delta\end{aligned}$$

for $n \geqslant n(\delta)$, and since $n(\delta) < \infty$ and $|\mathsf{R}(0)| \leqslant$ Const, it follows that

$$\overline{\lim_n} \, \mathsf{E}\left|\frac{1}{n}\sum_{k=1}^{n} h_k\right|^2 \leqslant \delta.$$

Since $\delta > 0$ was chosen arbitrary, this proves the implication $\Longleftarrow$.

Thus, the model $MA(1)$ is 'ergodic' in the following sense: the *time averages* $\overline{h}_n$ converge (in L^2) to the mean value μ, or, as it is alternatively called, to the *ensemble average*. (As regards the general concepts of 'ergodicity', 'mixing', 'ergodic theorems', see, e.g., [439; Chapter V].)

The importance of this 'ergodic' result from the statistical viewpoint is completely clear: it *substantiates* the estimation of mean values in terms of time averages calculated from the observable values of $h_1, h_2, \dots$. It is clear that 'ergodicity' plays a key role in the substantiation of the statistical methods of estimation from samples not only in the case of mean values, but also for other characteristics: moments of various orders, covariances, and so on.

5. We recall that, given a covariance $\mathsf{Cov}(h_n, h_m)$, we can define the *correlation* $\mathsf{Corr}(h_n, h_m)$ by the formula

$$\mathsf{Corr}(h_n, h_m) = \frac{\mathsf{Cov}(h_n, h_m)}{\sqrt{\mathsf{D}h_n \mathsf{D}h_m}}. \tag{10}$$

By the Cauchy–Schwartz–Bunyakovskiĭ inequality, $|\mathsf{Corr}(h_n, h_m)| \leqslant 1$.

In the stationary case, $\mathsf{Cov}(h_n, h_{n+k})$ is independent of n. We denote this value by $\mathsf{R}(k)$ and set

$$\rho(k) = \mathsf{Corr}(h_n, h_{n+k}) = \frac{\mathsf{R}(k)}{\mathsf{R}(0)}. \tag{11}$$

Then we obtain by (7) that, in the model $MA(1)$,

$$\rho(k) = \begin{cases} 1, & k = 0, \\ \dfrac{b_0 b_1}{b_0^2 + b_1^2}, & k = 1, \\ 0, & k > 1. \end{cases} \tag{12}$$

It is interesting that if we set $\theta_1 = b_1/b_0$, then

$$\rho(1) = \frac{\theta_1}{1 + \theta_1^2} = \frac{(1/\theta_1)}{1 + (1/\theta_1)^2}. \tag{13}$$

6. We now consider the model $MA(q)$:

$$h_n = \mu + \beta(L)\varepsilon_n, \quad \text{where} \quad \beta(L) = b_0 + b_1 L + \cdots + b_q L^q.$$

It is easy to see that

$$\mathsf{E}h_n = \mu,$$
$$\mathsf{D}h_n = b_0^2 + b_1^2 + \cdots + b_q^2$$

and

$$\mathsf{R}(k) = \begin{cases} \sum\limits_{j=0}^{q-k} b_j b_{k+j}, & k = 1, \dots, q, \\ 0, & k > q. \end{cases} \tag{14}$$

It is clear from (14) that schemes of type $MA(q)$ can be used to simulate (by varying the coefficients b_i) the behavior of sequences $h = (h_n)$ with no correlation between the variables h_n and h_{n+k} for $k > q$.

Remark. In our discussion of the adjustment of one or another model to empirical data, or, as we put it above, 'the simulation of the behavior of the sequences $h = (h_n)$', we should mention that the general pattern here is as follows.

Given sample values $h_1, h_2, \dots$, we determine first certain empirical characteristics, e.g., the sample mean

$$\overline{h}_n = \frac{1}{n} \sum_{k=1}^{n} h_k,$$

the sample variance

$$\widehat{\sigma}_n^2 = \frac{1}{n} \sum_{k=1}^{n} (h_k - \overline{h}_n)^2,$$

the sample correlation (of order k)

$$r_n(k) = \frac{\sum_{i=k+1}^{n} (h_i - \overline{h}_n)(h_{i-k} - \overline{h}_n)}{n\widehat{\sigma}_n^2},$$

the sample partial correlations, and so on.

After that, using the formulas for the corresponding theoretical characteristics (as, e.g., (12) or (14)) in the models that we are approximating, we vary the parameters (such as, for instance, the b_i in (12) and (14)) to *adjust* the theoretical characteristics to the empirical. Finally, at the last stage, we *estimate the quality* of this adjustment using our knowledge about the distributions of the empirical characteristics and their deviations from the theoretical distributions.

7. The next natural step after the models $MA(q)$ is to consider $MA(\infty)$, i.e., models with

$$h_n = \mu + \sum_{j=0}^{\infty} b_j \varepsilon_{n-j}. \tag{15}$$

Of course, we must impose certain conditions on the coefficients to ensure that the sum in (15) is convergent. If we require that

$$\sum_{j=0}^{\infty} b_j^2 < \infty, \tag{16}$$

than the series in (15) converges in mean square.

Under this assumption,

$$\mathsf{E} h_n = \mu, \qquad \mathsf{D} h_n = \sum_{j=0}^{\infty} b_j^2, \tag{17}$$

and

$$\mathsf{R}(k) = \sum_{j=0}^{\infty} b_{k+j} b_j, \qquad k \geqslant 0. \tag{18}$$

Considering the representation (15) for $h = (h_n)$ it is usually said in the theory of stationary stochastic processes that the h is the 'output of a physically realizable filter with impulse response $b = (b_j)$ and with input $\varepsilon = (\varepsilon_n)$'.

It is remarkable that, in a way, each 'regular', stationary (in the wide sense) sequence $h = (h_n)$ can be represented as a sum (15) with property (16). As regards the precise formulation and all the totality of the problems relating to the *Wold expansion* of a stationary sequence in a sum of 'singular' and 'regular' components such that the latter are representable as in (15), see § 2d below and, in greater detail, e.g., [439; Chapter VI, § 5].

§ 2b. Autoregressive Model $AR(p)$

1. By the definition in § 1d, a sequence $h = (h_n)_{n \geqslant 1}$ is said to be governed by the *autoregressive model (scheme) $AR(p)$ of order p* if

$$h_n = \mu_n + \sigma \varepsilon_n, \tag{1}$$

where

$$\mu_n = a_0 + a_1 h_{n-1} + \cdots + a_p h_{n-p}. \tag{2}$$

We can put it another way by saying that $h = (h_n)$ satisfies the *difference* equation of order p

$$h_n = a_0 + a_1 h_{n-1} + \cdots + a_p h_{n-p} + \sigma \varepsilon_n, \tag{3}$$

which, using the operator L introduced in § 2a, can be rewritten as

$$(1 - a_1 L - \cdots - a_p L^p) h_n = a_0 + \sigma\varepsilon_n, \tag{4}$$

or, in a more compact form,

$$\alpha(L) h_n = w_n, \tag{5}$$

where $\alpha(L) = 1 - a_1 L - \cdots - a_p L^p$, $w_n = a_0 + \sigma\varepsilon_n$.

As already mentioned in § 1d, for a complete description of the evolution of the sequence $h = (h_n)$ governed by the difference equation (3) we must also set 'initial values' $(h_{1-p}, h_{2-p}, \dots, h_0)$.

One often sets $h_{1-p} = \cdots = h_0 = 0$, or one can assume that these are random variables independent of the sequence $\varepsilon_1, \varepsilon_2, \dots$. In 'ergodic' cases the asymptotic behavior of the h_n as $n \to \infty$ is independent on the 'initial' conditions, and in this sense, the initial data are not that important. Still, we describe precisely all our assumptions concerning the 'initial' conditions in what follows.

We present the results of a computer simulation of the autoregressive model $AR(2)$ in Fig. 19.

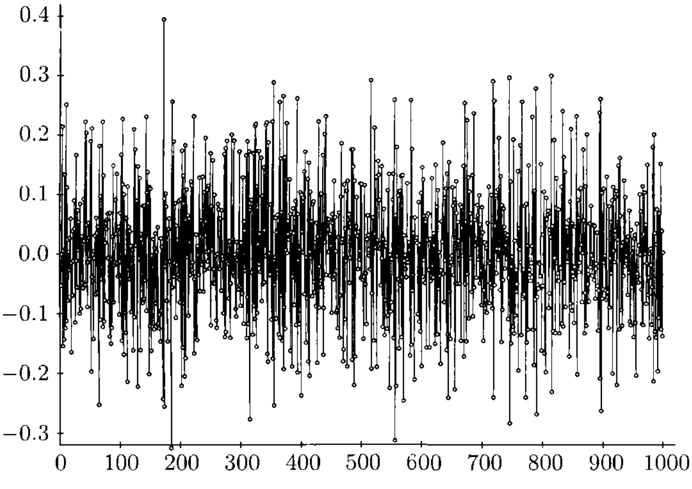


FIGURE 19. Computer simulation of a sequence $h = (h_n)$ $(2 \leqslant n \leqslant 1000)$ governed by the $AR(2)$ model with $h_n = a_0 + a_1 h_{n-1} + a_2 h_{n-2} + \sigma\varepsilon_n$, where $a_0 = 0$, $a_1 = -0.5$, $a_2 = 0.01$, $\sigma = 0.1$, and $h_0 = h_1 = 0$

2. First, we consider in detail the simple case of $p = 1$, where

$$h_n = a_0 + a_1 h_{n-1} + \sigma\varepsilon_n. \tag{6}$$

It can be distinguished from the general class of models $AR(p)$ by the property that, of all the 'past-related' variables $h_{n-1}, h_{n-2}, \dots, h_{n-p}$ involved in (3), the contribution in h_n is made only by the closest in time variable h_{n-1}.

If $\varepsilon = (\varepsilon_n)_{n\geqslant 1}$ is a sequence of independent random variables and h_0 is independent of ε, then the sequence $h = (h_n)_{n\geqslant 1}$ becomes a classical example of a *constructively* defined *Markov chain.*

From (6) we find recursively that

$$h_n = a_0(1 + a_1 + \cdots + a_1^{n-1}) + a_1^n h_0 + \sigma(\varepsilon_n + a_1\varepsilon_{n-1} + \cdots + a_1^{n-1}\varepsilon_1). \tag{7}$$

Hence the properties of the sequence h essentially depend on the values of the parameter a_1. We must distinguish between the three cases of $|a_1| < 1$, $|a_1| = 1$, and $|a_1| > 1$, where the case of $|a_1| = 1$ plays the role of a 'boundary' of sorts, which we explain in what follows.

By (7) we obtain

$$\begin{aligned}\mathsf{E}h_n &= a_1^n \mathsf{E}h_0 + a_0\bigl(1 + a_1 + \cdots + a_1^{n-1}\bigr),\\ \mathsf{D}h_n &= a_1^{2n}\mathsf{D}h_0 + \sigma^2\bigl(1 + a_1^2 + \cdots + a_1^{2(n-1)}\bigr),\\ \operatorname{Cov}(h_n, h_{n-k}) &= a_1^{2n-k}\mathsf{D}h_0 + \sigma^2 a_1^k\bigl(1 + a_1^2 + \cdots + a_1^{2(n-k-1)}\bigr)\end{aligned}$$

for $n - k \geqslant 1$.

It is clear from these relations that if $|a_1| < 1$ and $\mathsf{E}|h_0| < \infty$, then

$$\mathsf{E}h_n = a_1^n \mathsf{E}h_0 + \frac{a_0(1 - a_1^n)}{1 - a_1} \longrightarrow \frac{a_0}{1 - a_1}$$

as $n \to \infty$ and (if $\mathsf{D}h_0 < \infty$)

$$\begin{aligned}\mathsf{D}h_n = a_1^{2n}\mathsf{D}h_0 + \frac{\sigma^2(1 - a_1^{2n})}{1 - a_1^2} &\longrightarrow \frac{\sigma^2}{1 - a_1^2},\\ \operatorname{Cov}(h_n, h_{n-k}) &\longrightarrow \frac{\sigma^2 a_1^k}{1 - a_1^2}.\end{aligned}$$

Hence, the sequence $h = (h_n)_{n\geqslant 0}$ approaches in this case ($|a_1| < 1$) a steady state as $n \to \infty$. Moreover, if the initial distribution (the distribution of h_0) is normal, i.e.,

$$h_0 \sim \mathscr{N}\left(\frac{a_0}{1 - a_1}, \frac{\sigma^2}{1 - a_1^2}\right),$$

then $h = (h_n)_{n\geqslant 0}$ is a *stationary* Gaussian sequence (both in the wide and the strict sense) with

$$\mathsf{E}h_n = \frac{a_0}{1 - a_1}, \qquad \mathsf{D}h_n = \frac{\sigma^2}{1 - a_1^2}, \tag{8}$$

and

$$\operatorname{Cov}(h_n, h_{n+k}) = \frac{\sigma^2 a_1^k}{1 - a_1^2}. \tag{9}$$

We recall that a sequence is stationary in the *strict* sense if

$$\operatorname{Law}(h_0, h_1, \dots, h_m) = \operatorname{Law}(h_k, h_{1+k}, \dots, h_{m+k})$$

for all admissible values of m and k, while it is stationary in the *wide* sense if

$$\operatorname{Law}(h_i, h_j) = \operatorname{Law}(h_{i+k}, h_{j+k}).$$

If

$$\rho(k) = \operatorname{Corr}(h_n, h_{n+k}) = \frac{\operatorname{Cov}(h_n, h_{n+k})}{\sqrt{\mathsf{D}h_n \, \mathsf{D}h_{n+k}}},$$

then we obtain by (8) and (9) that

$$\rho(k) = a_1^k \tag{10}$$

for $|a_1| < 1$, i.e., the correlation between the variables h_n and h_{n+k} decreases *geometrically*.

Comparing the representation (7) and formula (2) in the preceding section, we observe that for each fixed n the variable h_n in the $AR(1)$ model can be treated as its counterpart h_n in the $MA(q)$ model with $q = n - 1$. Slightly abusing the language one says sometimes that 'the $AR(1)$ model can be regarded as the $MA(\infty)$ model'.

In $AR(1)$, the case $|a_1| = 1$ corresponds to a *classical random walk* (cf. Chapter I, § 2a, where we discuss the random walk conjecture and the concept of an 'efficient' market). For example, if $a_1 = 1$, then

$$h_n = a_0 n + h_0 + \sigma(\varepsilon_1 + \dots + \varepsilon_n).$$

Hence

$$\mathsf{E}h_n = a_0 n + \mathsf{E}h_0$$

and

$$\mathsf{D}h_n = \sigma^2 n \to \infty \quad \text{as} \quad n \to \infty.$$

The case of $|a_1| > 1$ is '*explosive*' (in the sense that both the expectation $\mathsf{E}h_n$ and variance $\mathsf{D}h_n$ increase *exponentially* with n).

3. We now consider the case $p = 2$:

$$h_n = a_0 + a_1 h_{n-1} + a_2 h_{n-2} + \sigma\varepsilon_n. \tag{11}$$

Using the above-mentioned operator L we can rewrite this difference equation as follows:

$$(1 - a_1 L - a_2 L^2) h_n = a_0 + \sigma\varepsilon_n. \tag{12}$$

If $a_2 = 0$, then the corresponding equation

$$(1 - a_1 L) h_n = a_0 + \sigma\varepsilon_n \tag{13}$$

fits into the case $AR(1)$ we have already discussed.

We now set $w_n = a_0 + \sigma\varepsilon_n$. Then (13) assumes the following form:

$$(1 - a_1 L) h_n = w_n. \tag{14}$$

It is natural to look for a 'conversion' of this relation enabling one to determine the h_n from the 'input' (w_n).

Based on the properties of L (see § 2a.2) we can see that

$$(1 + a_1 L + a_1^2 L^2 + \cdots + a_1^k L^k)(1 - a_1 L) = (1 - a_1^{k+1} L^{k+1}). \tag{15}$$

We now apply the operator $1 + a_1 L + a_1^2 L^2 + \cdots + a_1^k L^k$ to both sides of (13). In view of (15),

$$h_n = (1 + a_1 L + a_1^2 L^2 + \cdots + a_1^k L^k) w_n + a_1^{k+1} L^{k+1} h_n. \tag{16}$$

Setting here $k = n - 1$ and $w_n = a_0 + \sigma\varepsilon_n$ we obtain

$$h_n = (a_0 + \sigma\varepsilon_n) + a_1(a_0 + \sigma\varepsilon_{n-1}) + \cdots + a_1^{n-1}(a_0 + \sigma\varepsilon_1) + a_1^n h_0, \tag{17}$$

which is precisely the representation (7) above.

By (16), bearing in mind (14) with $k = n - 1$ we obtain

$$h_n = (1 + a_1 L + a_1^2 L^2 + \cdots + a_1^{n-1} L^{n-1})(1 - a_1 L) h_n + a_1^n h_0. \tag{18}$$

If $|a_1| < 1$ and n is sufficiently large, then we have the approximate equality

$$h_n \approx (1 + a_1 L + a_1^2 L^2 + \cdots + a_1^{n-1} L^{n-1})(1 - a_1 L) h_n. \tag{19}$$

This suggests that a natural way to define the inverse operator $(1 - a_1 L)^{-1}$ is to consider the limit (in an appropriate sense) of the sequence of operators $1 + a_1 L + a_1^2 L^2 + \cdots + a_1^n L^n$ as $n \to \infty$. (Cf. the algebraic representation $(1 - z)^{-1} = 1 + z + z^2 + \cdots$ for $|z| < 1$.)

These heuristic arguments can be formalized, e.g., as follows.

We consider a stationary sequence $\widetilde{h} = (\widetilde{h}_n)$, where

$$\widetilde{h}_n = \sum_{j=0}^{\infty} a_1^j w_{n-j} \tag{20}$$

(the series converges in mean square). It is easy to see that $(\widetilde{h}_n)$ is a solution of (14). We claim that this is the *unique* stationary solution with finite second moment.

Let $h = (h_n)$ be another stationary solution. Then

$$h_n = \sum_{j=0}^{k} a_1^j w_{n-j} + a_1^{k+1} h_{n-(k+1)} \tag{21}$$

by (16), therefore

$$\mathsf{E}\left| h_n - \sum_{j=0}^{k} a_1^j w_{n-j} \right|^2 = a_1^{2k+1} \mathsf{E} h_{n-(k+1)}^2 = a_1^{2k+1} \mathsf{E} h_0^2 \to 0 \tag{22}$$

as $k \to \infty$.

Hence we obtain by (20) that equation (13) has a unique stationary solution (with finite second moment).

4. The above arguments show the way to the 'conversion' of difference equation (12) enabling one to determine the elements of the sequence (h_n) by the elements of (w_n).

Since

$$(1 - \lambda_1 L)(1 - \lambda_2 L) = 1 - (\lambda_1 + \lambda_2)L + \lambda_1 \lambda_2 L^2 \tag{23}$$

for any λ_1 and λ_2, it follows that if we choose λ_1 and λ_2 such that

$$\begin{aligned} \lambda_1 + \lambda_2 &= a_1, \\ \lambda_1 \lambda_2 &= -a_2, \end{aligned} \tag{24}$$

then

$$1 - a_1 L - a_2 L^2 = (1 - \lambda_1 L)(1 - \lambda_2 L). \tag{25}$$

It is clear from (24) that λ_1 and λ_2 are the roots of the quadratic equation

$$\lambda^2 - a_1 \lambda - a_2 = 0, \tag{26}$$

i.e.,

$$\lambda_1 = \frac{a_1 + \sqrt{a_1^2 + 4a_2}}{2} \quad \text{and} \quad \lambda_2 = \frac{a_1 - \sqrt{a_1^2 + 4a_2}}{2}.$$

To put it another way, $\lambda_1 = z_1^{-1}$ and $\lambda_2 = z_2^{-1}$, where z_1 and z_2 are the roots of the equation

$$1 - a_1 z - a_2 z^2 = 0, \tag{27}$$

while the polynomial $1 - a_1 z - a_2 z^2$ if the result of the substitution $L \to z$ in the operator expression $1 - a_1 L - a_2 L^2$.

In view of (25), we can rewrite (12) as follows:

$$(1 - \lambda_1 L)(1 - \lambda_2 L) h_n = w_n.$$

After the formal multiplication of both sides by $(1 - \lambda_2 L)^{-1}(1 - \lambda_1 L)^{-1}$ we see that

$$h_n = (1 - \lambda_2 L)^{-1}(1 - \lambda_1 L)^{-1} w_n. \tag{28}$$

If $\lambda_1 \neq \lambda_2$, then we formally obtain

$$\frac{1}{(1 - \lambda_1 L)(1 - \lambda_2 L)} = (\lambda_1 - \lambda_2)^{-1} \left[\frac{\lambda_1}{1 - \lambda_1 L} - \frac{\lambda_2}{1 - \lambda_2 L} \right], \tag{29}$$

therefore

$$h_n = \frac{\lambda_1}{\lambda_1 - \lambda_2}(1 - \lambda_1 L)^{-1} w_n - \frac{\lambda_2}{\lambda_1 - \lambda_2}(1 - \lambda_2 L)^{-1} w_n. \tag{30}$$

If $|\lambda_i| < 1$, $i = 1, 2$, i.e., if the roots of the characteristic equation (27) lie *outside* the unit disc, then

$$(1 - \lambda_i L)^{-1} = 1 + \lambda_i L + \lambda_i^2 L^2 + \cdots, \qquad i = 1, 2, \tag{31}$$

and therefore, in view of (30), the stationary solution (with finite second moment) of (12) has the following form:

$$h_n = \sum_{j=0}^{\infty} (c_1 \lambda_1^j + c_2 \lambda_2^j) w_{n-j}, \tag{32}$$

where

$$c_1 = \frac{\lambda_1}{\lambda_1 - \lambda_2}, \qquad c_2 = \frac{\lambda_2}{\lambda_2 - \lambda_1}.$$

(The uniqueness of the stationary solution to (12) can be proved in the same way as for (13).)

5. Finally, we proceed to general $AR(p)$ models, in which

$$h_n = a_0 + a_1 h_{n-1} + \cdots + a_p h_{n-p} + \sigma\varepsilon_n, \tag{33}$$

that is (setting $w_n = a_0 + \sigma\varepsilon_n$),

$$(1 - a_1 L - a_2 L^2 - \cdots - a_p L^p) h_n = w_n. \tag{34}$$

Using the same method as for $p = 1$ or $p = 2$, we consider the factorization

$$1 - a_1 L - a_2 L^2 - \cdots - a_p L^p = (1 - \lambda_1 L)(1 - \lambda_2 L) \cdots (1 - \lambda_p L) \tag{35}$$

and assume that the λ_i, $i = 1, \dots, p$, are all distinct.

If $|\lambda_i| < 1$, $i = 1, \dots, p$, then we can express the stationary solution of (34), which is the unique solution with finite second moment starting at $-\infty$ in time, as follows:

$$h_n = (1 - \lambda_1 L)^{-1} \cdots (1 - \lambda_p L)^{-1} w_n. \tag{36}$$

We note that the numbers $\lambda_1, \lambda_2, \dots, \lambda_p$ are the roots of the equation

$$\lambda^p - a_1 \lambda^{p-1} - \cdots - a_{p-1}\lambda - a_p = 0 \tag{37}$$

(cf. (26)). Putting that another way, we can say that $\lambda_i = z_i^{-1}$, where the z_i are the roots of the equation $1 - a_1 z - a_2 z^2 - \cdots - a_p z^p = 0$ (cf. (27)). Hence we can obtain a stationary solution $h = (h_n)$ if all the roots of this equation lie *outside* the unit disc.

To obtain a representation of type (30) or (32) we consider the following analogue of the expansion (29):

$$\frac{1}{(1 - \lambda_1 z) \cdots (1 - \lambda_p z)} = \frac{c_1}{1 - \lambda_1 z} + \cdots + \frac{c_p}{1 - \lambda_p z}, \tag{38}$$

where $c_1, \dots, c_p$ are constants to be determined.

Multiplying both sides of (38) by $(1 - \lambda_1 z) \cdots (1 - \lambda_p z)$, we see that for all z we must have

$$1 = \sum_{i=1}^{p} c_i \prod_{\substack{1 \leqslant k \leqslant p \\ k \neq i}} (1 - \lambda_k z). \tag{39}$$

This equality must hold, in particular, for $z = \lambda_1^{-1}, \dots, z = \lambda_p^{-1}$, which gives us the following values of $c_1, \dots, c_p$:

$$c_i = \frac{\lambda_i^{p-1}}{\prod_{\substack{1 \leqslant k \leqslant p \\ k \neq i}} (\lambda_i - \lambda_k)}. \tag{40}$$

(We note that $c_1 + \cdots + c_p = 1$.)

By (36), (38), and (40),

$$h_n = \sum_{l=0}^{\infty}(c_1\lambda_1^l + \cdots + c_p\lambda_p^l)w_{n-l}, \tag{41}$$

which is a generalization of (32) to the case $p \geqslant 2$.

The representation (41) enables one to evaluate various characteristics of the sequence $h = (h_n)$ such as the moments $\mathsf{E}h_n^k$, the covariances, the expectations $\mathsf{E}(h_{n+k} \mid \mathscr{F}_n)$ (where $\mathscr{F}_n = \sigma(\dots, w_{-1}, w_0, \dots, w_n)$), and so on.

Assuming that we have a stationary case it is easy to find the moments $\mathsf{E}h_n \equiv \mu$ directly from (33):

$$\mu = \frac{a_0}{1 - (a_1 + \cdots + a_p)}\,. \tag{42}$$

As regards the covariances $\mathsf{R}(k) = \mathsf{Cov}(h_n, h_{n+k})$, we see easily from (33) that

$$\mathsf{R}(k) = a_1\mathsf{R}(k-1) + \cdots + a_p\mathsf{R}(k-p) \tag{43}$$

for $k = 1, 2, \dots$. If $k = 0$, then

$$\mathsf{R}(0) = a_1\mathsf{R}(1) + \cdots + a_p\mathsf{R}(p) + \sigma^2. \tag{44}$$

The correlation functions $\rho(k)$, $k \geqslant 0$, satisfy the same equations (43) and (44) (they are called the *Yule–Walker equations*, after G. U. Yule and G. T. Walker; [271]).

6. One of the central issues in the statistics of the autoregressive schemes $AR(q)$ is the *estimation* of the parameters $\theta = (a_0, a_1, \dots, a_p, \sigma)$ involved in (1) and (2), where we now assume for definiteness that $h_0, h_{-1}, \dots$ are known constants.

If we assume that the white noise $\varepsilon = (\varepsilon_n)$ is Gaussian, then the most important estimation tools is the *maximum likelihood method* in which the quantity

$$\widehat{\theta}_n = \arg\max_{\theta} p_\theta(h_1, h_2, \dots, h_n)$$

is the estimator for the parameter θ defined in terms of the observations $h_1, h_2, \dots, h_n$ (here $p_\theta(h_1, h_2, \dots, h_n)$ is the joint probability density of the (Gaussian) vector $(h_1, h_2, \dots, h_n)$).

We shall illustrate the general principle pertaining to the maximum likelihood method by the example of the $AR(1)$ model with $h_n = a_0 + a_1 h_{n-1} + \sigma\varepsilon_n$; here we regard $\sigma > 0$ as a known parameter, $h_0 = 0$, and $n \geqslant 1$.

Since $\theta = (a_0, a_1)$, it follows that

$$p_\theta(h_1, \dots, h_n) = \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n \exp\left\{-\frac{1}{2}\sum_{k=1}^{n}\frac{(h_k - a_0 - a_1 h_{k-1})^2}{2\sigma^2}\right\}.$$

It is easy to see that the value of the estimator $\theta = (\widehat{a}_0, \widehat{a}_1)$ can be found from the minimality condition for the function

$$\psi(a_0, a_1) = \sum_{k=1}^{n} (h_k - a_0 - a_1 h_{k-1})^2.$$

Letting $a_0 \in \mathbb{R}$ and $a_1 \in \mathbb{R}$, we obtain

$$\begin{aligned} \frac{\partial\psi}{\partial a_0} = 0 &\iff 2\sum_{k=1}^{n} (h_k - a_0 - a_1 h_{k-1}) = 0, \\ \frac{\partial\psi}{\partial a_1} = 0 &\iff 2\sum_{k=1}^{n} (h_k - a_0 - a_1 h_{k-1}) h_{k-1} = 0. \end{aligned} \tag{45}$$

Solving this linear system one finds the values of the estimators $\widehat{a}_0$ and $\widehat{a}_1$.

We now concentrate on the properties of the estimators $\widehat{a}_1 = \widehat{a}_1(h_1, \dots, h_n)$, $n \geqslant 1$, for a_1. We assume for simplicity that a_0 is a known parameter ($a_0 = 0$) and $\sigma = 1$.

Under these assumptions, $h_n = a_1 h_{n-1} + \varepsilon_n$ and

$$\widehat{a}_1 = \frac{\sum_{k=1}^{n} h_{k-1} h_k}{\sum_{k=1}^{n} h_{k-1}^2}$$

by (45). Hence

$$\widehat{a}_1 = a_1 + \frac{\sum_{k=1}^{n} h_{k-1}\varepsilon_k}{\sum_{k=1}^{n} h_{k-1}^2}.$$

We now set

$$M_n = \sum_{k=1}^{n} h_{k-1}\varepsilon_k.$$

The sequence $M = (M_n)$ is (for each value of the parameter a_1) a *martingale* with quadratic characteristic

$$\langle M \rangle_n = \sum_{k=1}^{n} h_{k-1}^2.$$

Hence if a_1 is an actual value of this unknown parameter, then

$$\widehat{a}_1 = a_1 + \frac{M_n}{\langle M \rangle_n}. \tag{46}$$

From this representation we see that the maximum likelihood estimator $\widehat{a}_1$ is *strongly consistent*: $\widehat{a}_1 \to a_1$ as $n \to \infty$ with probability one because $\langle M \rangle_n \to \infty$ (P-a.s.), and by the *strong law of large numbers* for square integrable martingales (see (12) in § 1b and [439; Chapter VII, § 5]),

$$\frac{M_n}{\langle M \rangle_n} \to 0 \quad (\mathsf{P}\text{-a.s.}).$$

Calculating in our case ($a_0 = 0$, $\sigma = 1$) the *Fisher information*

$$I_n(a_1) = \mathsf{E}_{a_1}\left\{-\frac{\partial^2 \log \mathsf{P}_{a_1}(h_1,\dots,h_n)}{\partial a_1^2}\right\},$$

where E_{a_1} is the expectation with respect to the measure

$$\mathsf{P}_{a_1} = \mathrm{Law}(h_1,\dots,h_n \mid \theta = a_1),$$

we obtain

$$I_n(a_1) = \mathsf{E}_{a_1}\langle M\rangle_n = \mathsf{E}_{a_1}\sum_{k=1}^{n} h_{k-1}^2.$$

Using the above formulas for the $\mathsf{E}h_{k-1}$ and $\mathsf{D}h_{k-1}$, we see that

$$I_n(a_1) \sim \begin{cases} \dfrac{n}{1-a_1^2}, & |a_1| < 1, \\ \dfrac{n^2}{2}, & |a_1| = 1, \\ \dfrac{a_1^{2n}}{(a_1^2-1)^2}, & |a_1| > 1. \end{cases} \tag{47}$$

We see that $|a_1| = 1$ is some distinguished, 'boundary' case, when $h = (h_n)$ is a *random walk* (cf. Chapter I, § 2a "Random walk conjecture and concept of efficient market"). For $|a_1| \neq 1$ the corresponding sequence $h = (h_n)$ is Markovian. Moreover, if $|a_1| < 1$, then the sequence $h = (h_n)$ 'approaches a steady state' as $n \to \infty$.

This fact is responsible for the considerable attention that is paid in the statistics of the sequences $h = (h_n)$ to the solution of the following problem: which conjecture of the two,

$$H_0\colon |a_1| = 1 \quad \text{or} \quad H_1\colon |a_1| > 1,$$

is more likely?

In the econometrics literature and, in particular, in the literature on financial models the range of issues touching upon the validity of the equality $|a_1| = 1$ is called 'the unit root problem'.

We now present (without proofs; referring instead to [424], [445], and the literature cited there for greater detail) several results concerning the limit distribution for the deviations $\widehat{a}_1 - a_1$.

THEOREM 1. *As $n \to \infty$ we have*

$$\lim_n \mathsf{P}_{a_1}\left\{\sqrt{I_n(a_1)}\,(\widehat{a}_1 - a_1) \leqslant x\right\} = \begin{cases} \Phi(x), & |a_1| < 1, \\ \mathrm{H}_{a_1}(x), & |a_1| = 1, \\ \mathrm{Ch}(x), & |a_1| > 1, \end{cases} \tag{48}$$

where $\Phi(x) = \int_{-\infty}^{x} \varphi_{(0,1)}(y)\,dy$ *is the standard normal distribution,* $\mathrm{Ch}(x)$ *is the Cauchy distribution with density* $\dfrac{1}{\pi(1+x^2)}$, *and* $\mathrm{H}_{a_1}(x)$ *is the distribution of the random variable*

$$a_1 \frac{W^2(1)-1}{2\sqrt{2}\displaystyle\int_0^1 W^2(s)\,ds},$$

where $(W(s))_{s\leqslant 1}$ *is a standard Wiener process (Brownian motion).*

It is interesting that if $|a_1| \neq 1$, then the densities of the limit distributions are *symmetric* with respect to the origin. However, the density of the distribution $\mathrm{H}_{a_1}(x)$ is *asymmetric* if $|a_1| = 1$. (This is an easy consequence of the observation that $\mathsf{P}(W^2(1) - 1 > 0) \neq \frac{1}{2}$.)

The next result shows that, considering a *random normalization of the deviation* $(\widehat{a}_1 - a_1)$ (which means that we use the *stochastic* Fisher information $\langle M\rangle_n$ in place of the *Fisher information* $I_n(a_1)$) we can obtain only *two* distinct limit distributions rather than *three* as in (48).

THEOREM 2. *As* $n \to \infty$ *we have*

$$\lim \mathsf{P}_{a_1}\bigl\{\sqrt{\langle M\rangle_n}\,(\widehat{a}_1 - a_1) \leqslant x\bigr\} = \begin{cases} \Phi(x), & |a_1| \neq 1, \\ \mathrm{H}'_{a_1}(x), & |a_1| = 1, \end{cases} \tag{49}$$

where $\mathrm{H}'_{a_1}(x)$ *is the probability distribution of the random variable*

$$a_1 \frac{W^2(1)-1}{2\sqrt{\displaystyle\int_0^1 W^2(s)\,ds}}.$$

Finally, the *sequential* maximum likelihood estimators together with random normalization bring us to a *unique* limit distribution.

THEOREM 3. *Assume that* $\theta > 0$, *let*

$$\tau(\theta) = \inf\bigl\{n \geqslant 1 \colon \langle M\rangle_n \geqslant \theta\bigr\}, \tag{50}$$

and let

$$\widehat{a}_1\bigl(\tau(\theta)\bigr) = \frac{\sum_{k=1}^{\tau(\theta)} h_{k-1}h_k}{\sum_{k=1}^{\tau(\theta)} h_{k-1}^2} \tag{51}$$

be the sequential maximum likelihood estimator. Then

$$\lim_{n\to\infty} \mathsf{P}_{a_1}\bigl\{\sqrt{\langle M\rangle_{\tau(\theta)}}\,\bigl[\widehat{a}_1(\tau(\theta)) - a_1\bigr] \leqslant x\bigr\} = \Phi(x) \tag{52}$$

for each $a_1 \in \mathbb{R}$.

We now compare this result with assertion (48).

If n was our original time parameter, then we can regard θ as 'new', 'operational' time defined in terms of stochastic Fisher information $\langle M\rangle$. We note that if Fisher information changes only slightly on (large) time intervals, then they correspond to small intervals of 'new time' θ and conversely. Thus, using this 'new', 'operational' time we make the flow of information more uniform, homogeneous, the incoming data are now of 'equal worth', and are 'identically distributed' in a sense for all the values of a_1. Eventually, this results in the uniqueness and normality of the limit distribution.

We come across problems related to this 'new' time below, in Chapter IV, § 3d, where we explain in close detail one such change of time aimed at 'flattening' the statistical data on currency cross rates, in the dynamics of which 'geographic' components of the periodic nature are clearly visible.

7. We now present several additional results on the properties of maximum likelihood estimators (see [258], [424], and [445]).

First,

$$\sup_{a_1\in\mathbb{R}} \mathsf{E}_{a_1}|\widehat{a}_1 - a_1| \to 0 \quad \text{as} \quad n\to\infty.$$

Second, let $U(a_1)$ be the class of estimators $\widetilde{a}_1$ with bias $b_{a_1}(\widetilde{a}_1) \equiv \mathsf{E}_{a_1}(\widetilde{a}_1 - a_1)$ satisfying the conditions

$$b_{a_1}(\widetilde{a}_1) \to 0 \quad \text{and} \quad \frac{db_{a_1}}{da_1}(\widetilde{a}_1) \to 0 \quad \text{as} \quad n\to\infty.$$

(The maximum likelihood estimators $\widehat{a}_1$ are in the class $U(a_1)$ if $|a_1| \neq 1$.)

If $|a_1| \neq 1$, then the maximum likelihood estimators $\widehat{a}_1$ are *asymptotically efficient* in $U(a_1)$ in the following sense: for $\widetilde{a}_1 \in U(a_1)$ we have

$$\overline{\lim_{n}}\, \frac{\mathsf{E}_{a_1}\langle M\rangle_n(\widehat{a}_1 - a_1)^2}{\mathsf{E}_{a_1}\langle M\rangle_n(\widetilde{a}_1 - a_1)^2} \leqslant 1.$$

If $|a_1| < 1$ (the 'stationary' case), then the estimators $\widehat{a}_1$ are also *asymptotically efficient* in the usual sense, i.e., for all $\widetilde{a}_1 \in U(a_1)$,

$$\overline{\lim_{n}}\, \frac{\mathsf{D}_{a_1}\widehat{a}_1}{\mathsf{D}_{a_1}\widetilde{a}_1} \leqslant 1.$$

Sequential maximum likelihood estimators have also the property of *asymptotic uniformity*: as $\theta\to\infty$, we have

$$\sup_{|a_1|\leqslant 1}\, \sup_x \Bigl|\mathsf{P}_{a_1}\bigl\{\sqrt{\langle M\rangle_{\tau(\theta)}}\,\bigl[\widehat{a}_1(\tau(\theta)) - a_1\bigr] \leqslant x\bigr\} - \Phi(x)\Bigr| \to 0,$$

$$\sup_{1<r\leqslant|a_1|\leqslant R}\, \sup_x \Bigl|\mathsf{P}_{a_1}\bigl\{\sqrt{\langle M\rangle_{\tau(\theta)}}\,\bigl[\widehat{a}_1(\tau(\theta)) - a_1\bigr] \leqslant x\bigr\} - \Phi(x)\Bigr| \to 0.$$

§ 2c. Mixed Autoregressive Moving Average Model $ARMA(p,q)$ and Integrated Model $ARIMA(p,d,q)$

1. These models combine the properties of the already considered $MA(q)$ and $AR(p)$ models, thus often providing a fair opportunity to find a model that can well explain the probabilistic backgrounds of statistical 'stock'.

As above, we consider a filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_n), \mathsf{P})$. It is now convenient to assume that $\mathscr{F}_n = \sigma(\dots, \varepsilon_{-1}, \varepsilon_0, \varepsilon_1, \dots, \varepsilon_n)$, where $\varepsilon = (\varepsilon_n)$ is 'white noise' (in the strict sense).

By definition, (see § 1d), a sequence $h = (h_n)$ is governed by the model $ARMA$ if

$$h_n = \mu_n + \sigma\varepsilon_n, \tag{1}$$

where

$$\mu_n = (a_0 + a_1 h_{n-1} + \cdots + a_p h_{n-p}) + (b_1\varepsilon_{n-1} + b_2\varepsilon_{n-2} + \cdots + b_q\varepsilon_{n-q}). \tag{2}$$

Without loss of generality we can assume that the value of σ is known to be equal to one: $\sigma = 1$. Then by (1) and (2) we obtain

$$h_n - (a_1 h_{n-1} + \cdots + a_p h_{n-p}) = a_0 + [\varepsilon_n + b_1\varepsilon_{n-1} + b_2\varepsilon_{n-2} + \cdots + b_q\varepsilon_{n-q}] \tag{3}$$

or

$$\alpha(L)h_n = a_0 + \beta(L)\varepsilon_n, \tag{4}$$

where

$$\alpha(L) = 1 - a_1 L - \cdots - a_p L^p \tag{5}$$

and

$$\beta(L) = 1 + b_1 L + \cdots + b_q L^q. \tag{6}$$

We note that if $q = 0$, then

$$\alpha(L)h_n = w_n,$$

where $w_n = a_0 + \varepsilon_n$, i.e., we arrive at the $AR(p)$ model (cf. (5) in § 2b).

On the other hand, if $p = 0$, then (3) takes the following form:

$$h_n = a_0 + \beta(L)\varepsilon_n, \tag{7}$$

i.e., we obtain the $MA(q)$ model (cf. (4) in § 2a).

Considering a formal conversion of (4) we see that

$$h_n = \mu + \frac{\beta(L)}{\alpha(L)}\varepsilon_n, \tag{8}$$

where

$$\mu = \frac{a_0}{1 - (a_1 + \cdots + a_p)} \tag{9}$$

(under the assumption that $a_1 + \cdots + a_p \neq 1$).

We now consider the question on the existence of a stationary solution of equation (3) (in the class L^2). By (8) and in view of our previous discussions (in § 2b.5), the answer to this question depends on the properties of the operator $\alpha(L)$, that is, of the autoregressive components of our model $ARMA(p,q)$.

If all the roots of equation (37) in § 2b are smaller than one in absolute values (and in this case $a_1+\dots+a_p \neq 1$), then this model has a unique stationary solution $h=(h_n)$ (in the class L^2).

For this stationary solution,

$$\mathsf{E}h_n = \frac{a_0}{1-(a_1+\dots+a_p)} \tag{10}$$

by (8) (cf. (42) in § 2b).

It is easy to conclude from (3) that the covariance $\mathsf{R}(k)=\mathsf{Cov}(h_n,h_{n+k})$ satisfies for $k>q$ the same relations

$$\mathsf{R}(k) = a_1\mathsf{R}(k-1)+\dots+a_p\mathsf{R}(k-p) \tag{11}$$

as in the $AR(p)$ case (cf. formula (43) in § 2b).

If $k<q$, then the corresponding representation of $\mathsf{R}(k)$ has a more complicated form than (11), since we must also take into account the correlation dependence between ε_{n-k} and h_{n-k}.

2. As an illustration, we consider the model $ARMA(1,1)$, which is a combination of $AR(1)$ and $MA(1)$:

$$h_n - a_1h_{n-1} = a_0+\varepsilon_n+b_1\varepsilon_{n-1}. \tag{12}$$

We assume that $|a_1|<1$ (the 'stationary' case). Then

$$\alpha(L)=1-a_1L, \qquad \beta(L)=1+b_1L \tag{13}$$

and (8) can be written as follows:

$$\begin{aligned} h_n &= \frac{a_0}{1-a_1}+\frac{1+b_1L}{1-a_1L}\varepsilon_n \\ &= \frac{a_0}{1-a_1}+\Bigl(\sum_{k=0}^{\infty}a_1^kL^k\Bigr)(1+b_1L)\varepsilon_n \\ &= \frac{a_0}{1-a_1}+(a_1+b_1)\sum_{k=1}^{\infty}a_1^{k-1}\varepsilon_{n-k}+\varepsilon_n. \end{aligned} \tag{14}$$

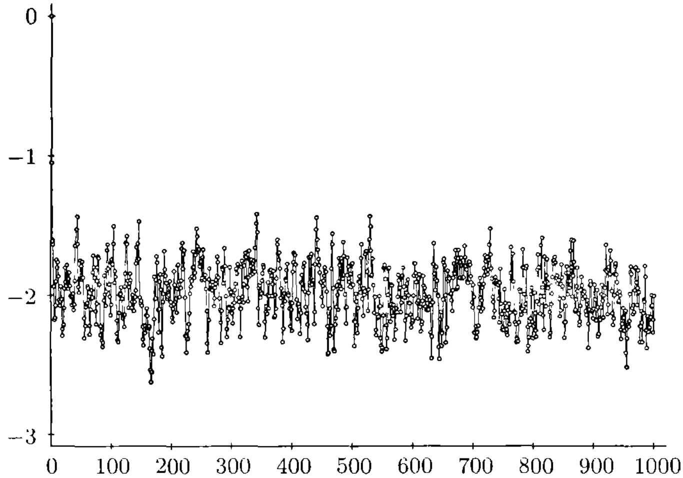


FIGURE 20. Computer simulation of a sequence $h = (h_n)$ governed by the $ARMA(1,1)$ model with $h_n = a_0 + a_1 h_{n-1} + b_1\varepsilon_{n-1} + \sigma\varepsilon_n$ $(0 \leqslant n \leqslant 1000)$, where $a_0 = -1$, $a_1 = 0.5$, $b = 0.1$, $\sigma = 0.1$, and $h_0 = 0$

Hence we immediately see that the covariance $\mathsf{R}(k) = \mathsf{Cov}(h_n, h_{n+k})$ satisfies the following relations:

$$\begin{aligned} \mathsf{R}(k) &= a_1\mathsf{R}(k-1), \quad k \geqslant 2, \\ \mathsf{R}(1) &= a_1\mathsf{R}(0) + b_1, \\ \mathsf{R}(0) &= a_1\mathsf{R}(1) + (1 + a_1 b_1 + b_1^2), \end{aligned} \tag{15}$$

and therefore

$$\mathsf{R}(0) = \mathsf{D}h_n = \frac{1 + 2a_1b_1 + b_1^2}{1 - a_1^2}$$

and

$$\rho(k) = \frac{\mathsf{R}(k)}{\mathsf{R}(0)} = \frac{(1 + a_1b_1)(a_1 + b_1)}{1 + 2a_1b_1 + b_1^2}\, a_1^{k-1}. \tag{16}$$

It should be mentioned that the correlation decreases *geometrically* as $k \to \infty$ for $|a_1| < 1$. This must be taken into account in the adjustment of the $ARMA(1,1)$ models (or the more general $ARMA(p,q)$ models) to particular statistical data.

3. The above-considered models $ARMA(p,q)$ are well understood and used mostly in the description of stationary time series. On the other hand, .if a time series $x = (x_n)$ is *nonstationary*, then the consideration of the differences $\Delta x_n = x_n - x_{n-1}$ or the d-order differences $\Delta^d x_n$ brings one to the (occasionally) 'more' stationary sequence $\Delta^d x = (\Delta^d x_n)$.

There exist special expressions describing this situation: one says that a sequence $x = (x_n)$ is governed by the *ARIMA*(p, d, q) *model* if $\Delta^d x = (\Delta^d x_n)$ is governed by the model *ARMA*(p, q). (In a symbolic form, we write $\Delta^d ARIMA(p, d, q) = ARMA(p, q)$.)

For a deeper insight into the meaning of these models we consider the particular case of *ARIMA*$(0, 1, 1)$. Here $\Delta x_n = h_n$, where (h_n) is a sequence governed by *MA*(1), i.e.,

$$\Delta x_n = \mu + (b_0 + b_1 L)\varepsilon_n.$$

Let S be the operator of summation ('integration') defined by the formula $S = \Delta^{-1}$, or, equivalently,

$$S = 1 + L + L^2 + \cdots = (1 - L)^{-1}.$$

Then we can formally write

$$x_n = (Sh)_n,$$

where $h_n = \mu + (b_0 + b_1 L)\varepsilon_n = \mu + b_0\varepsilon_n + b_1\varepsilon_{n-1}$.

Hence $x = (x_n)$ can be regarded as the result of the 'integration' of some sequence $h = (h_n)$ governed by the *MA*(1) model, which explains the name $ARIMA = AR + I + MA$. (Cf. Fig. 18 and Fig. 21.)

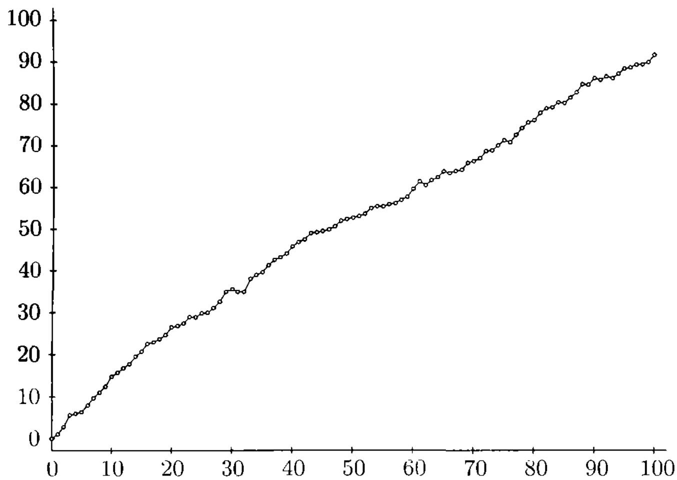


FIGURE 21. Computer simulation of a sequence $x = (x_n)$ governed by the *ARIMA*$(0, 1, 1)$ model with $\Delta x_n = \mu + b_1\varepsilon_{n-1} + b_0\varepsilon_n$, where $\mu = 1$, $b_1 = 1$, $b_0 = 0.1$, and $x_0 = 0$

We have already mentioned that these models are widely used in the Box–Jenkins theory [53]. For information about their applications in financial data statistics, see, e.g., [351].

§ 2d. Prediction in Linear Models

1. We pointed out in the introduction to this section that the construction of probabilistic and statistical models (based on the 'past' data) is not an end in itself; it is necessary, in the long run, to predict the 'future' price movements.

It is very seldom that one can give an 'error-proof' forecast using the 'past' data. (This is characteristic, e.g., of the so-called singular stationary sequences; see subsection 4 below and, in more detail, e.g., [439; Chapter VI].)

The typical situation is, of course, the one when making a forecast we are inevitably making an error, the size of which determines the risks involved in the solutions based on this forecast.

2. In the case of stationary linear models there exists a well-developed (and beautiful) theory of the construction of *optimal linear estimators* (in the mean-square sense), which is mainly due to A. N. Kolmogorov and N. Wiener.

We have already seen that many of the sequences $h = (h_n)$ considered can be represented as *one-sided moving averages*

$$h_n = \sum_{k=0}^{\infty} a_k \varepsilon_{n-k}, \tag{1}$$

where $\sum\limits_{k=0}^{\infty} |a_k|^2 < \infty$ and $\varepsilon = (\varepsilon_n)$ is some 'basic' sequence, white noise. (See formula (15) in § 2a for the $MA(q)$ and $MA(\infty)$ models, formula (41) in § 2b for the $AR(p)$ models, and (8) in § 2c for $ARMA(p,q)$.)

To describe results on the extrapolation of these sequences we require certain concepts and notation.

If $\xi = (\xi_n)$ is a stochastic sequence, then let $\mathscr{F}_n^\xi = \sigma(\dots, \xi_{n-1}, \xi_n)$ be the σ-algebra generated by the 'past' $\{\xi_k,\ k \leqslant n\}$, let $\mathscr{F}_{\infty-}^\xi = \bigvee \mathscr{F}_n^\xi$ be the σ-algebra generated by all the variables ξ_n, let $H_n^\xi = \overline{\mathscr{L}}(\dots, \xi_{n-1}, \xi_n)$ be the closed (in L^2) linear manifold spanned by the variables $\{\xi_k,\ k \leqslant n\}$, and let $H_\infty^\xi = \overline{\mathscr{L}}(\xi_k,\ k < \infty)$ be the closed linear manifold spanned by all the variables ξ_n.

Let $\eta = \eta(\omega)$ be a random variable with finite second moment $\mathsf{E}\eta^2(\omega)$. We now formulate the problem of finding an estimate of η in terms of our observations of the sequence ξ.

The following two approaches are most widely used here.

If these are the variables $(\dots, \xi_{n-1}, \xi_n)$ that we must observe, then in the framework of the first approach, we consider the class of *all* $\mathscr{F}_n^\xi$-measurable estimators $\widetilde{\eta}_n$ and the one said to be the *best* (*optimal*) is the estimator $\widehat{\eta}_n$ delivering the smallest mean-square deviation, i.e.,

$$\mathsf{E}|\eta - \widehat{\eta}_n|^2 = \inf_{\widetilde{\eta}_n} \mathsf{E}|\eta - \widetilde{\eta}_n|^2. \tag{2}$$

As is well known, the estimator optimal in this sense must have the following form:

$$\widehat{\eta}_n = \mathsf{E}(\eta \mid \mathscr{F}_n^{\xi}), \tag{3}$$

where $\mathsf{E}(\,\cdot \mid \cdot\,)$ is the conditional expectation, which is, generally speaking, a *nonlinear* function of the observations $(\dots, \xi_{n-1}, \xi_n)$. (The issues of the *nonlinear optimal filtering*, *extrapolation*, and *interpolation* for fairly wide classes of stochastic processes are considered, e.g., in [303].)

In the framework of the other approach, on which we concentrate now, one considers only the set of *linear* estimators together with its closure (in L^2), i.e., functions of $(\dots, \xi_{n-1}, \xi_n)$ belonging to H_n^{ξ}.

In a similar way to (2), by the *best* (*optimal*) *linear* estimator of the variable η we mean $\widehat{\lambda}_n \in H_n^{\xi}$ such that

$$\mathsf{E}|\eta - \widehat{\lambda}_n|^2 = \inf_{\widetilde{\lambda}_n \in H_n^{\xi}} \mathsf{E}|\eta - \widetilde{\lambda}_n|^2. \tag{4}$$

In this case we use the following notation for $\widehat{\lambda}_n$ (cf. (3)):

$$\widehat{\lambda}_n = \widehat{\mathsf{E}}(\eta \mid H_n^{\xi}), \tag{5}$$

where $\widehat{\mathsf{E}}(\,\cdot \mid \cdot\,)$ is called conditional expectation in the *wide sense*.

3. We now return to the issue of (linear) prediction on the basis of the 'information' available by time 0 for the variables in a sequence $h = (h_n)$, $n \geqslant 1$, described by (1). This problem can be solved relatively easy if the 'information' in question means all the 'past' generated by $\varepsilon = (\varepsilon_n)_{n\leqslant 0}$ (rather than by $h = (h_n)_{n\leqslant 0}$), i.e.,

$$H_0^{\varepsilon} = \overline{\mathscr{L}}(\dots, \varepsilon_{-1}, \varepsilon_0).$$

Indeed,

$$\begin{aligned} \widehat{h}_n^{(\varepsilon)} \equiv \widehat{\mathsf{E}}(h_n | H_0^{\varepsilon}) &= \widehat{\mathsf{E}}\biggl(\sum_{k=0}^{n-1} a_k\varepsilon_{n-k} + \sum_{k=n}^{\infty} a_k\varepsilon_{n-k} \biggm| H_0^{\varepsilon}\biggr) \\ &= \sum_{k=n}^{\infty} a_k\varepsilon_{n-k} + \widehat{\mathsf{E}}\biggl(\sum_{k=0}^{n-1} a_k\varepsilon_{n-k} \biggm| H_0^{\varepsilon}\biggr) = \sum_{k=n}^{\infty} a_k\varepsilon_{n-k}, \end{aligned} \tag{6}$$

since

$$\widehat{\mathsf{E}}(\varepsilon_i \mid H_0^{\varepsilon}) = \mathsf{E}\varepsilon_i \ (= 0)$$

for $i \geqslant 1$, which follows easily from the orthogonality of the components of the white noise $\varepsilon = (\varepsilon_i)$.

Moreover, it is not difficult to find the extrapolation error

$$\sigma_n^2 = \mathsf{E}\big|h_n - \widehat{\mathsf{E}}(h_n \mid H_0^\varepsilon)\big|^2 = \sum_{k=0}^{n-1} |a_k|^2.$$

Clearly, $\sigma_n^2 \leqslant \sigma_{n+1}^2$ and

$$\lim_n \sigma_n^2 = \sum_{k=0}^{\infty} |a_k|^2 \ \left(= \mathsf{E}h_n^2\right).$$

Hence the role of the 'past data' $H_0^\varepsilon = \overline{\mathscr{L}}(\dots, \varepsilon_{n-1}, \varepsilon_0)$ in our prediction of the values of h_n diminishes with the growth of n and, in the limit (as $n \to \infty$), we should take the mere expectation (i.e., 0 in our case) as the best linear estimator.

Of course, the problem of extrapolation on the basis of not necessarily *infinitely many* variables $\{\varepsilon_k,\ k \leqslant 0\}$, but *finitely many* ones (e.g., the variables $\{\varepsilon_k,\ l \leqslant k \leqslant 0\}$) is also of interest. It should be noted, however, that this latter problem is technically more complicated that the one under consideration now.

The representation

$$\widehat{h}_n^{(\varepsilon)} = \sum_{k=n}^{\infty} a_k \varepsilon_{n-k} \tag{7}$$

so obtained solves the extrapolation problem for the h_n on the basis of the (linear) 'information' H_0^ε rather than the 'information' H_0^h contained in $\overline{\mathscr{L}}(\dots, h_{-1}, h_0)$.

Clearly, if we assume that

$$H_0^\varepsilon = H_0^h, \tag{8}$$

then using (7) we can (*in principle*, at any rate) express the corresponding estimator $\widehat{\mathsf{E}}(h_n | H_0^h)$ in the form of a sum $\sum\limits_{k=n}^{\infty} \widehat{a}_k h_{n-k}$.

4. In view of our assumptions (1) and (8), it would be useful to recall several general principles of the theory of stationary stochastic sequences.

Let $\xi = (\xi_n)$ be a stationary sequence (in the wide sense).

We can represent each element $\eta \in H(\xi)$ as a sum of two orthogonal components:

$$\eta = \widehat{\mathsf{E}}(\eta \mid S(\xi)) + \big[\eta - \widehat{\mathsf{E}}(\eta \mid S(\xi))\big],$$

where $S(\xi) = \bigcap_n H_n^\xi$ is the (linear) information contained in the 'infinitely remote past'. Let $R(\xi)$ be the set of elements of the form $\eta - \widehat{\mathsf{E}}(\eta \mid S(\xi))$, where $\eta \in H(\xi)$. Then we see that $H(\xi)$ can itself be represented as an orthogonal sum: $H(\xi) = S(\xi) \oplus R(\xi)$.

We say that a stationary sequence $\xi = (\xi_n)$ is *regular* if $H(\xi) = R(\xi)$ and *singular* if $H(\xi) = S(\xi)$.

The meaning of the condition $H(\xi) = S(\xi)$ is fairly clear: this indicates that all the information provided by the variables in the sequence ξ 'lies in the infinitely remote past'. This explains why singular sequences are also said to be *purely* or *completely deterministic*. If $S(\xi) = \varnothing$ (i.e., $H(\xi) = R(\xi)$), then one says that the sequence ξ is *purely* or *completely nondeterministic*.

The following result sheds light on the concepts that we have just introduced.

PROPOSITION 1. *Each nondegenerate stationary (in the wide sense) random sequence $\xi = (\xi_n)$ has a unique decomposition*

$$\xi_n = \xi_n^r + \xi_n^s,$$

where $\xi^r = (\xi_n^r)$ is a regular sequence and $\xi^s = (\xi_n^s)$ is a singular sequence; moreover, ξ^r and ξ^s are orthogonal ($\mathsf{Cov}(\xi_n^r, \xi_m^s) = 0$ for all n and m).

(See [439; Chapter IV, § 5] for greater detail.)

We now introduce another important concept.

DEFINITION. We call a sequence $\varepsilon = (\varepsilon_n)$ an *innovation* sequence for $\xi = (\xi_n)$ if

a) $\varepsilon = (\varepsilon_n)$ is 'white noise' in the wide sense, i.e., $\mathsf{E}\varepsilon_n = 0$, $\mathsf{E}\varepsilon_n\varepsilon_m = 0$ for $n \neq m$, and $\mathsf{E}\varepsilon_n^2 = 1$;

b) $H_n^\xi = H_n^\varepsilon$ for each n.

The meaning of the term 'innovation' is clear: the elements of the sequence $\varepsilon = (\varepsilon_n)$ are orthogonal, therefore we can regard ε_{n+1} as ('new') information that 'updates', 'innovates' the data in H_n^ε and enables one to construct H_{n+1}^ε. Since $H_n^\xi = H_n^\varepsilon$ for all n, ε_{n+1} updates also the information in H_n^ξ. (Cf. Chapter I, § 2a where we consider the 'random walk' conjecture and discuss the concept of 'efficient market'.)

The next result establishes a connection between the concepts introduced above.

PROPOSITION 2. *A nondegenerate sequence $\xi = (\xi_n)$ is regular if and only if there exist an innovation sequence $\varepsilon = (\varepsilon_n)$ and numbers $(a_k)_{k\geqslant 0}$ such that $\sum_{k=0}^{\infty} |a_k|^2 < \infty$ and* (P-a.s.)

$$\xi_n = \sum_{k=0}^{\infty} a_k \varepsilon_{n-k}. \tag{9}$$

(See the proof, e.g., in [439; Chapter VI, § 5].)

The following result is an immediate consequence of Propositions 1 and 2.

PROPOSITION 3. *If $\xi = (\xi_n)$ is a nondegenerate stationary sequence, then it has a Wold expansion*

$$\xi_n = \xi_n^s + \sum_{k=0}^{\infty} a_k \varepsilon_{n-k}, \tag{10}$$

where $\sum_{k=0}^{\infty} a_k^2 < \infty$ *and* $\varepsilon = (\varepsilon_n)$ *is some innovation sequence (for* ξ^r*).*

5. Assuming that the sequence $h = (h_n)$ has a representation (1) with innovation sequence (for h) $\varepsilon = (\varepsilon_n)$ we obtain

$$\widehat{h}_n \equiv \widehat{\mathsf{E}}(h_n \mid H_0^h) = \widehat{\mathsf{E}}(h_n \mid H_0^\varepsilon) \ \ (= \widehat{h}_n^{(\varepsilon)}).$$

Hence

$$\widehat{h}_n = \sum_{k=n}^{\infty} a_k \varepsilon_{n-k} = \sum_{k=n}^{\infty} \widehat{a}_k^{(n)} h_{n-k} \tag{11}$$

by (6), where the last equality is a consequence of the relation $H_n^\varepsilon = H_n^h$ holding for all n.

Formula (11) solves the problem of optimal linear extrapolation of the variables h_n on the basis of the 'past' information $\{h_k,\ k \leqslant 0\}$ *in principle.* However, the following two questions come naturally: when does a sequence $h = (h_n)$ admits a representation (1) with innovation sequence $\varepsilon = (\varepsilon_n)$ and how can one find the coefficients $\widehat{a}_k$ in (11)?

The answer to the first question is contained in Proposition 2: the sequence $h = (h_n)$ must be regular. It is here that the well-known Kolmogorov test is working: *if* $h = (h_n)$ *is a stationary sequence in the wide sense with spectral density* $f = f(\lambda)$ *such that*

$$\int_{-\pi}^{\pi} \ln f(\lambda)\, d\lambda > -\infty, \tag{12}$$

then this sequence is regular (see, e.g., [439; Chapter VI, § 5]).

6. We recall that each stationary sequence in the wide sense $\xi = (\xi_n)$ with $\mathsf{E}\xi_n = 0$ has a *spectral representation*

$$\xi_n = \int_{-\pi}^{\pi} e^{i\lambda n}\, z(d\lambda), \tag{13}$$

where $z = z(\Delta)$ (with $\Delta \in \mathscr{B}([-\pi,\pi])$) is a complex-valued stochastic measure with orthogonal values: $\mathsf{E}z(\Delta_1)\overline{z}(\Delta_2) = 0$ if $\Delta_1 \cap \Delta_2 = \varnothing$ (see, e.g., [439; Chapter VI, § 3]).

In addition, the covariance function $R(n) = \mathsf{Cov}(\xi_{k+n}, \xi_k)$ has the representation

$$\mathsf{R}(n) = \int_{-\pi}^{\pi} e^{i\lambda n}\, F(d\lambda) \tag{14}$$

with *spectral measure* $F(\Delta) = \mathsf{E}|z(\Delta)|^2$.

If there exists a (nonnegative) function $f = f(\lambda)$ such that

$$F(\Delta) = \int_{\Delta} f(\lambda)\, d\lambda, \tag{15}$$

then we call it the *spectral density*, while the function $F(\lambda) = F((-\infty, \lambda])$ is called the *spectral function*.

Hence if we know a priori that the initial sequence $h = (h_n)$ is stationary in the wide sense with spectral density $f = f(\lambda)$ satisfying condition (12), then this sequence is *regular*, and admits the representation (9) with innovation sequence $\varepsilon = (\varepsilon_n)$.

7. In the linear models considered above we did not define the sequence $h = (h_n)$ in terms of spectral representations; we used instead the series (1) with an orthonormal sequence $\varepsilon = (\varepsilon_n)$ such that $\mathsf{E}\varepsilon_n = 0$ and $\mathsf{E}\varepsilon_i\varepsilon_j = \delta_{ij}$, where δ_{ij} is the Kronecker delta:

$$\delta_{ij} = 1 \quad \text{for } i = j \qquad \text{and} \qquad \delta_{ij} = 0 \quad \text{for } i \neq j.$$

Clearly, the sequence $\varepsilon = (\varepsilon_n)$ is stationary and we have

$$\mathsf{R}_\varepsilon(n) = \begin{cases} 1, & n = 0, \\ 0, & n \neq 0 \end{cases} \tag{16}$$

for its correlation (= covariance) function and

$$f_\varepsilon(\lambda) = \frac{1}{2\pi}, \qquad -\pi \leqslant \lambda < \pi \tag{17}$$

for the spectral density. $\left(\text{Because } \mathsf{R}_\varepsilon(n) = \int_{-\pi}^{\pi} e^{i\lambda n} \frac{1}{2\pi}\, d\lambda.\right)$

If $h_n = \sum_{k=0}^{\infty} a_k\varepsilon_{n-k}$ and $\sum_{k=0}^{\infty} |a_k|^2 < \infty$, then it is easy to see that the covariance function of the sequence $h = (h_n)$ can be defined as follows:

$$\mathsf{R}_h(n) = \int_{-\pi}^{\pi} e^{i\lambda n} f_h(\lambda)\, d\lambda, \tag{18}$$

where

$$f_h(\lambda) = \frac{1}{2\pi}|\Phi(e^{-i\lambda})|^2 \tag{19}$$

and

$$\Phi(z) = \sum_{k=0}^{\infty} a_k z^k. \tag{20}$$

We now assume that the power series (20) has *convergence radius* $r > 1$ and Φ *does not vanish for* $|z| \leqslant 1$. Under this assumption we can answer the question on the values of the coefficients $\widehat{a}_k$ in the representation (11) for the optimal linear prediction $\widehat{h}_n$ as follows (see, e.g., [439; Chapter VI, § 6] for greater detail).

Let

$$\widehat{\varphi}_n(\lambda) = e^{i\lambda n} \frac{\Phi_n(e^{-i\lambda})}{\Phi(e^{-i\lambda})}, \tag{21}$$

where

$$\Phi_n(z) = \sum_{k=n}^{\infty} a_k z^k. \tag{22}$$

We expand $\widehat{\varphi}_n(\lambda)$ in the Fourier series:

$$\widehat{\varphi}_n(\lambda) = \widehat{a}_n^{(n)} + \widehat{a}_{n+1}^{(n)} e^{-i\lambda} + \widehat{a}_{n+2}^{(n)} e^{(-2i\lambda)} + \cdots. \tag{23}$$

Then the coefficients $\widehat{a}_n^{(n)}, \widehat{a}_{n+1}^{(n)}, \dots$ are *precisely* those delivering the optimal linear prediction $\widehat{h}_n$ of h_n:

$$\widehat{h}_n \equiv \widehat{\mathsf{E}}(h_n \mid H_0^h) = \sum_{k=n}^{\infty} \widehat{a}_k^{(n)} h_{n-k}. \tag{24}$$

Thus, we have a method for constructing a prediction of h_n on the basis of the 'past' variables $\{\dots, h_{-1}, h_0\}$, provided that we can find the representation (23).

8. To illustrate the above method we consider now several examples for the stationary models: $MA(q)$, $AR(p)$, and $ARMA(p,q)$. We shall indicate explicit formulas for the predictions only for several values of p and q. As regards general formulas, see, e.g., [439; Chapter IV, § 6].

EXAMPLE 1 (*The model $MA(q)$*). Let

$$h_n = \beta(L)\varepsilon_n, \tag{25}$$

where

$$\beta(L) = b_0 + b_1 L + \cdots + b_q L^q, \tag{26}$$

i.e.,

$$h_n = \sum_{k=0}^{q} b_k \varepsilon_{n-k}. \tag{27}$$

Comparing this representation with (1) we see that $a_k = b_k$ for $0 \leqslant k \leqslant q$ and $a_k = 0$, $k > q$.

Consequently,

$$\Phi(z) = \sum_{k=0}^{q} b_k z^k$$

and

$$\Phi_n(z) = \begin{cases} \displaystyle\sum_{k=n}^{q} b_k z^k, & n \leqslant q, \\ 0, & n > q. \end{cases}$$

Hence

$$\widehat{\varphi}_n(\lambda) = e^{i\lambda n} \frac{\sum_{k=n}^{q} e^{-i\lambda k} b_k}{\sum_{k=0}^{q} e^{-i\lambda k} b_k} \tag{28}$$

for $n \leqslant q$ and $\widehat{\varphi}_n(\lambda) = 0$ for $n > q$.

Thus, if $n > q$, then $\widehat{a}_n = 0$ for all the coefficients in the expansion (23). therefore the optimal prediction in this case is $\widehat{h}_n = 0$. This is far from surprising. for the correlation between h_n and each of $h_0, h_{-1}, \dots$ is equal to zero for $n > q$.

If $q = 1$. then

$$h_n = b_0 \varepsilon_n + b_1 \varepsilon_{n-1}$$

and

$$\widehat{\varphi}_1(\lambda) = e^{i\lambda} \frac{e^{-i\lambda} b_1}{b_0 + e^{-i\lambda} b_1} = \frac{b_1}{b_0 + e^{-i\lambda} b_1}.$$

Of course, we can assume from the very beginning that $b_0 = 1$ and can set $b_1 = \theta$, where $|\theta| < 1$, which ensures that the function $\Phi(z) = 1 + \theta z$ does not vanish for $|z| \leqslant 1$. Then

$$\widehat{\varphi}_1(\lambda) = \frac{\theta}{1 + \theta e^{-i\lambda}} = \theta \sum_{k=0}^{\infty} (-1)^k (\theta e^{-i\lambda})^k = \theta \sum_{k=0}^{\infty} e^{-i\lambda k} (-1)^k \theta^k.$$

Comparing this with (23) we see that, for $q = 1$ and $n = 1$,

$$\begin{aligned}
\widehat{a}_1^{(1)} &= \theta, \\
\widehat{a}_2^{(1)} &= (-1) \cdot \theta^2, \\
\widehat{a}_3^{(1)} &= \theta^3, \\
&\dots \\
\widehat{a}_k^{(1)} &= (-1)^{k-1} \theta^k, \\
&\dots
\end{aligned}$$

Hence we obtain the following formula for the prediction of h_1 on the basis of the 'past' information $(\dots, h_{-1}, h_0)$ in the model $MA(1)$:

$$\begin{aligned}
\widehat{h}_1 &= \widehat{a}_1^{(1)} h_0 + \widehat{a}_2^{(1)} h_{-1} + \cdots + \widehat{a}_k^{(1)} h_{1-k} + \cdots \\
&= \theta\big(h_0 - \theta h_{-1} + \theta^2 h_{-2} + \cdots + (-1)^{k-1} \theta^{k-1} h_{1-k} + \cdots\big) \\
&= \sum_{k=0}^{\infty} (-1)^k \theta^{k+1} h_{-k}.
\end{aligned}$$

It is now clear that the *largest contribution* to out prediction of h_1 is the contribution of the 'most recent past', the variable h_0; while the contribution from the 'preceding' variables decreases geometrically (h_{-k} enters the formula with coefficient θ^k, where $|\theta| < 1$).

EXAMPLE 2 (*The model $AR(p)$*). We assume that (cf. (33) in § 2b)

$$h_n = a_1 h_{n-1} + \cdots + a_p h_{n-p} + \varepsilon_n \tag{29}$$

for $-\infty < n < \infty$; moreover, all the λ_i $(i = 1, \dots, p)$ in the factorization (35) (§ 2b) are distinct and $|\lambda_i| < 1$.

According to formula (41) in § 2b, the stationary solution of (29) can be represented as follows:

$$h_n = \sum_{k=0}^{\infty} (c_1 \lambda_1^k + \cdots + c_p \lambda_p^k) \varepsilon_{n-k}, \tag{30}$$

i.e., in the representation (1) we have

$$a_k = c_1 \lambda_1^k + \cdots + c_p \lambda_p^k. \tag{31}$$

In principle, the prediction $\widehat{h}_n$ of h_n on the basis of the variables h_k, $k \leqslant 0$, is described by (24), where the coefficients $\widehat{a}_k^{(n)}$ can be found from the Fourier expansion of $\widehat{\varphi}_n(\lambda)$ (see (21)–(23)) by taking into account formula (31) for the coefficients a_k in the definition of $\widehat{\varphi}_n(\lambda)$.

Fairly many various methods described in the literature produce formulas for the predictions $\widehat{h}_n$ that are convenient in applications (see, e.g., [211], where one can find a description of a recursion procedure enabling one to construct $\widehat{h}_n$ from $\widehat{h}_1, \dots, \widehat{h}_{n-1}$).

We restrict ourselves to an illustration by carrying out a calculation for the stationary model $AR(1)$:

$$h_n = \theta h_{n-1} + \varepsilon_n \tag{32}$$

with $|\theta| < 1$.

We recall that $\mathsf{R}(n) = \mathsf{Cov}(h_k, h_{k+n})$ can be defined in this case by the formula

$$\mathsf{R}(n) = \frac{\theta^k}{1 - \theta^2}.$$

Hence it is easy to see that the spectral function $f = f(\lambda)$ is well defined and

$$f(\lambda) = \frac{1}{2\pi} \cdot \frac{1}{|1 - \theta e^{-i\lambda}|^2}. \tag{33}$$

Comparing (33) and (19) we see that

$$\Phi(z) = \frac{1}{1 - \theta z} = \sum_{k=0}^{\infty} (\theta z)^k.$$

Hence

$$\widehat{\varphi}_n(\lambda) = \theta^n$$

by (21), and therefore (see (23) and (24))

$$\widehat{h}_n = \theta^n h_0, \qquad n \geqslant 1,$$

i.e., for a prediction of h_n on the basis of $\{h_k, k \leqslant 0\}$ one requires only the value of h_0. This is not surprising, given that $h = (h_n)$ is a Markov sequence.

EXAMPLE 3 (*The model ARMA(p, q) with parameters* $p = q = 1$). Setting $a_0 = 0$ in equation (12) (§ 2c) we find the following formula for a stationary solution:

$$h_n = (a_1 + b_1) \sum_{k=1}^{\infty} a_1^{k-1} \varepsilon_{n-k} + \varepsilon_n$$

where $|a_1| < 1$. Hence the coefficients in the representation $h_n = \sum_{k=0}^{\infty} \tilde{a}_k \varepsilon_{n-k}$ are as follows:

$$\tilde{a}_0 = 1, \qquad \tilde{a}_k = (a_1 + b_1) a_1^{k-1}. \tag{34}$$

Consequently,

$$\Phi(z) = \sum_{k=0}^{\infty} \tilde{a}_k z^k = 1 + \frac{(a_1 + b_1)}{a_1} \sum_{k=1}^{\infty} (a_1 z)^k = 1 + \frac{a_1 + b_1}{a_1} \cdot \frac{a_1 z}{1 - a_1 z}$$

and

$$\Phi_n(z) = \sum_{k=n}^{\infty} \tilde{a}_k z^k = \frac{a_1 + b_1}{a_1} \sum_{k=n}^{\infty} (a_1 z)^k = \frac{a_1 + b_1}{a_1} (a_1 z)^n \frac{1}{1 - a_1 z}$$

for $n \geqslant 1$. In view of (21) and (22), setting $z = e^{-i\lambda}$ we obtain (for $|b_1| < 1$)

$$\begin{aligned}\widehat{\varphi}_n(\lambda) &= e^{i\lambda n} \frac{\frac{a_1+b_1}{a_1} \cdot \frac{(a_1 z)^n}{1-a_1 z}}{1 + \frac{a_1+b_1}{a_1} \cdot \frac{a_1 z}{1-a_1 z}} = \frac{(a_1 + b_1) a_1^{n-1}}{1 + b_1 z} \\ &= a_1^{n-1} (a_1 + b_1) \sum_{k=0}^{\infty} (-1)^k (b_1 z)^k\end{aligned}$$

and (see (23))

$$\widehat{a}_k^{(n)} = a_1^{n-1} (a_1 + b_1) (-1)^k b_1^k.$$

Hence

$$\begin{aligned}\widehat{h}_n &= a_1^{n-1} (a_1 + b_1) \{ h_0 + (-1) b_1 h_{-1} + b_1^2 h_{-2} + \cdots \} \\ &= a_1^{n-1} (a_1 + b_1) \sum_{k=n}^{\infty} (-1)^k b_1^k h_{n-k}\end{aligned}$$

by (24).

Remark. As regards the prediction formulas for the general models $ARMA(p, q)$ and $ARIMA(p, d, q)$, see, e.g., [211].

3. Nonlinear Stochastic Conditionally Gaussian Models

The interest in *nonlinear* models has its origin in the quest for explanations of several phenomena (apparent both in financial statistics and economics as the whole) such as the 'cluster property' of prices, their 'disastrous' jumps and downfalls, 'heavy tails' of the distributions of the variables $h_n = \ln \frac{S_n}{S_{n-1}}$, 'the long memory' in prices, and some other, which cannot be understood in the framework of *linear* models.

On the other hand, there is no unanimity as to which of the nonlinear models—stochastic, chaotic ('dynamical chaos'), or other—should be used. Plenty of advocators are adducing arguments in favor of one or another approach.

There can be no doubt that economic indicators, including financial indexes, are prone to *fluctuate*.

Many *macroeconomic* indicators (the volumes of production, consumption, or investment, the general level of prices, interest rates, government reserves, and so on), which describe the state of the economy 'on the average', 'at large', do fluctuate, and so do also *microeconomic* indexes (current prices, the volume of traded stocks, and so on). Moreover, these fluctuations can have a very *high frequency* or be extremely *irregular*. This is known to occur also in stochastic and chaotic models; hence the researchers' attempts to describe fluctuation dynamics, abrupt transitions, 'catastrophic' outbursts, the grouping of the values (the cluster property), and so on, by means of these models.

Considering the behavior of many economic indexes (production volumes, the sizes of particular populations, or government reserves) 'on the average' one can discern certain *trends*, but this movement can accelerate or slow down: the growth can occur in *cycles* of sorts (periodic or aperiodic).

Thus, an analyst of statistical data relating to the economy, finances, or some area of natural sciences or technology, finds himself in front of a rather nontrivial problem of the choice of a 'right' model.

Below we describe several *nonlinear stochastic* and *chaotic models* that are popular in financial mathematics and financial statistics. We are making no claims of a comprehensive exposition, but are willing instead to present an 'introduction' into this range of problems. (As regards some nonlinear models, we can recommend, e.g., the monographs [193], [202], [461], and [462].)

§ 3a. *ARCH* and *GARCH* Models

1. Let $(\Omega, \mathscr{F}, \mathsf{P})$ be the original probability space and let $\varepsilon = (\varepsilon_n)_{n\geqslant 1}$ be a sequence of independent, normally distributed random variables ($\varepsilon_n \sim \mathscr{N}(0,1)$) simulating the 'randomness', 'uncertainty' in the models that we consider below.

By $\mathscr{F}_n$ we shall mean the σ-algebra $\sigma(\varepsilon_1, \dots, \varepsilon_n)$; we set $\mathscr{F}_0 = \{\varnothing, \Omega\}$.

We shall interpret $S_n = S_n(\omega)$ as the price (of a share or a bond, or an exchange rate, etc.) at time $n = 0, 1, \dots$. Here time can be measured in years, months, ..., minutes, and so on.

As already mentioned (§ 1d), to describe the evolution of the variables

$$h_n = \ln \frac{S_n}{S_{n-1}}, \tag{1}$$

R. Engle [140] considered the *conditionally Gaussian model* with

$$h_n = \sigma_n \varepsilon_n. \tag{2}$$

Here the volatilities σ_n are defined as follows:

$$\sigma_n^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i h_{n-i}^2, \tag{3}$$

where $\alpha_0 > 0$, $\alpha_i \geqslant 0$, and $h_0 = h_0(\omega)$ is a random variable independent of $\varepsilon = (\varepsilon_n)_{n\geqslant 1}$. (One often sets h_0 to be a constant or a random variable with expectation of its square chosen so that the expectations Eh_n^2, $n \geqslant 0$, are constant.)

We see from (3) that the σ_n are (predictable) functions of $h_{n-1}^2, \dots, h_{n-p}^2$. Moreover, it is clear that large (small) values of the h_{n-i}^2 imply large (respectively, small) values of σ_n^2. On the other hand, if h_n^2 turns out to be large, while the values of the 'preceding' variables $h_{n-1}^2, \dots, h_{n-p}^2$ are small, then this can be explained by a large value of ε_n. This gives an insight into how (nonlinear) models (1)–(3) can explain such phenomena as the 'cluster property', i.e., the grouping of the values of the h_n into batches of 'large' or 'small' values.

These considerations justify (as already mentioned in § 1d) the name of *ARCH*(p) (*AutoRegressive Conditional Heteroskedastic model*) given to this model by R. Engle [140]. The conditional variance ('volatility') σ_n^2 behaves in this model in an extremely uneven way because, by (3), it depends on the 'past' variables $h_{n-1}^2, h_{n-2}^2, \dots$.

2. We now consider several properties of sequences $h = (h_n)_{n \geqslant 1}$, governed by the $ARCH(p)$ model. For simplicity, we restrict ourselves to the case $p = 1$. (See, e.g., [202], [193], and [393] for a thorough study of the properties of the $ARCH(p)$ models and their applications; we present one result concerning the occurrence of 'heavy tails' in such models in § 3c (subsection 6).)

For $p = 1$ (see Fig. 22) we have

$$\sigma_n^2 = \alpha_0 + \alpha_1 h_{n-1}^2, \tag{4}$$

The following simple properties of the $h_n = \sigma_n \varepsilon_n$ are obvious:

$$\mathsf{E} h_n = 0, \tag{5}$$

$$\mathsf{E} h_n^2 = \alpha_0 + \alpha_1 \mathsf{E} h_{n-1}^2, \tag{6}$$

$$\mathsf{E}(h_n^2 \mid \mathscr{F}_{n-1}) = \sigma_n^2 = \alpha_0 + \alpha_1 h_{n-1}^2. \tag{7}$$

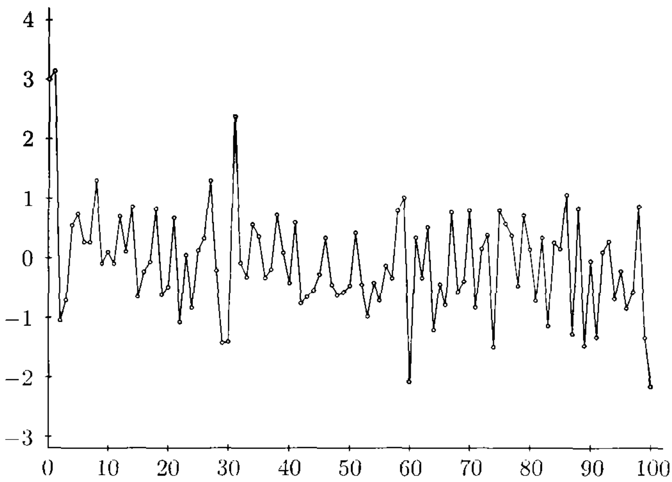


FIGURE 22. Computer simulation of a sequence $h = (h_n)$ governed by the $ARCH(1)$ model with $h_n = \sqrt{\alpha_0 + \alpha_1 h_{n-1}^2}\, \varepsilon_n$ $(0 \leqslant n \leqslant 100)$, where $\alpha_0 = 0.9$, $\alpha_1 = 0.2$; $h_0 = 3$

If

$$0 < \alpha_1 < 1,$$

then the recursion relation (6) has the *unique* stationary solution

$$\mathsf{E} h_n^2 \equiv \frac{\alpha_0}{1 - \alpha_1}, \qquad n \geqslant 0, \tag{8}$$

therefore setting $h_0^2 = \dfrac{\alpha_0}{1 - \alpha_1}$ we obtain formula (8) for the $\mathsf{E} h_n^2$, $n \geqslant 1$.

Next, a simple calculation shows that

$$\begin{aligned}
\mathsf{E}h_n^4 &= \mathsf{E}\sigma_n^4\,\mathsf{E}\varepsilon_n^4 = 3\mathsf{E}\sigma_n^4 = 3\mathsf{E}\big(\alpha_0 + \alpha_1 h_{n-1}^2\big)^2 \\
&= 3\big(\alpha_0^2 + 2\alpha_0\alpha_1\mathsf{E}h_{n-1}^2 + \alpha_1^2\mathsf{E}h_{n-1}^4\big) \\
&= \frac{3\alpha_0^2(1+\alpha_1)}{1-\alpha_1} + 3\alpha_1^2\mathsf{E}h_{n-1}^4.
\end{aligned} \tag{9}$$

Hence, assuming that $0 < \alpha_1 < 1$ and $3\alpha_1^2 < 1$, we can obtain the following 'stationary' solution ($\mathsf{E}h_n^4 \equiv \text{Const}$):

$$\mathsf{E}h_n^4 = \frac{3\alpha_0^2(1+\alpha_1)}{(1-\alpha_1)(1-3\alpha_1^2)}. \tag{10}$$

By (8) and (10), the 'stationary' value of the kurtosis is

$$K \equiv \frac{\mathsf{E}h_n^4}{(\mathsf{E}h_n^2)^2} - 3 = \frac{6\alpha_1^2}{1-3\alpha_1^2}.$$

It is *positive*, which means that the density of the 'steady-state' distribution of the variables h_n has a peak around the mean value (and the larger α_1^2, the more distinguished is this peak). We recall that $K = 0$ for a normal distribution.

Remark. The empirical value $\widehat{K}_N$ of the kurtosis can be calculated from the values of $h_1, h_2, \dots, h_N$ by the formula

$$\widehat{K}_N = \frac{1}{N}\sum_{k=1}^{N}(h_k - \overline{h}_N)^4 \Big/ \Big(\frac{1}{N}\sum_{k=1}^{N}(h_k - \overline{h}_N)^2\Big)^2 - 3,$$

where $\overline{h}_N = \dfrac{1}{N}(h_1 + \cdots + h_N)$. Judging by the data from S. Taylor's book [460], it is a rule rather than an exception that the kurtosis of financial indexes is positive. The cases of negative kurtosis are very rare in practice. As regards the values of the empirical kurtosis $\widehat{K}_N$ in the cases when it is positive, one can consider the following table relating to *gold* and *silver* prices, the exchange rate of the British pound against the US dollar, and General Motors' share price (calculated from the monthly averages):

TABLE 1.

Gold	1975–82	$\widehat{K}_N = 8.4$
Silver	1970–74	$\widehat{K}_N = 8.4$
GBP/USD	1974–82	$\widehat{K}_N = 5.4$
General Motors	1966–76	$\widehat{K}_N = 4.2$

3. For $0 < \alpha_1 < 1$ the sequence $h = (h_n)$ with $h_n = \sigma_n \varepsilon_n$ is a square integrable martingale difference, therefore it is a sequence of orthogonal variables:

$$\mathsf{Cov}(h_n, h_m) = 0, \qquad n \neq m.$$

Of course, this does not mean that h_n and h_m are *independent*, because, as we see, their joint distribution $\mathrm{Law}(h_n, h_m)$ is not Gaussian for $\alpha_1 > 0$.

One can get an idea of the character of the *dependence* between h_n and h_m by considering the *correlation* dependence between their squares h_n^2 and h_m^2 or their absolute values $|h_n|$ and $|h_m|$.

A simple calculation shows that

$$\mathsf{D}h_n^2 = \frac{2}{1 - 3\alpha_1^2} \left(\frac{\alpha_0}{1 - \alpha_1} \right)^2, \tag{11}$$

and

$$\mathsf{E}h_n^2 h_{n-1}^2 = \frac{1 + 3\alpha_1}{1 - 3\alpha_1^2} \cdot \frac{\alpha_0^2}{1 - \alpha_1}, \tag{12}$$

therefore

$$\rho(1) \equiv \mathsf{Corr}(h_n^2, h_{n-1}^2) = \frac{\mathsf{Cov}(h_n^2, h_{n-1}^2)}{\sqrt{\mathsf{D}h_n^2\, \mathsf{D}h_{n-1}^2}} = \alpha_1.$$

Further,

$$\begin{aligned} \mathsf{E}h_n^2 h_{n-k}^2 &= \mathsf{E}\big[h_{n-k}^2\, \mathsf{E}(h_n^2 \mid \mathscr{F}_{n-1})\big] = \mathsf{E}\big[h_{n-k}^2 \mathsf{E}(\sigma_n^2 \varepsilon_n^2 \mid \mathscr{F}_{n-1})\big] \\ &= \mathsf{E}\big[h_{n-k}^2(\alpha_0 + \alpha_1 h_{n-1}^2)\big] = \alpha_0 \mathsf{E}h_{n-k}^2 + \alpha_1 \mathsf{E}h_{n-1}^2 h_{n-k}^2 \end{aligned}$$

for $k < n$, which, gives the following simple recursion relation for $\rho(k) \equiv \dfrac{\mathsf{Cov}(h_n^2, h_{n-k}^2)}{\sqrt{\mathsf{D}h_n^2\, \mathsf{D}h_{n-k}^2}}$ in the 'stationary case':

$$\rho(k) = \alpha_0 \mathsf{E}h_{n-k}^2 + \alpha_1 \rho(k-1),$$

so that

$$\rho(k) = \alpha_1^k. \tag{13}$$

4. We pointed out in § 1d that the $ARCH(p)$ models are intimately connected with (general) autoregressive schemes $AR(p)$. For assume that we are considering the $ARCH(p)$ model and let $\nu_n = h_n^2 - \sigma_n^2$. If $\mathsf{E}h_n^2 < \infty$, then the sequence $\nu = (\nu_n)$ is a martingale difference (with respect to the flow $(\mathscr{F}_n)$) and it follows from (3) that the variables $x_n = h_n^2$ are governed by the autoregressive model $AR(p)$:

$$x_n = \alpha_0 + \alpha_1 x_{n-1} + \cdots + \alpha_p x_{n-p} + \nu_n, \tag{14}$$

with noise $\nu = (\nu_n)$ that is a *martingale difference.*

If $p = 1$, then

$$x_n = \alpha_0 + \alpha_1 x_{n-1} + \nu_n, \tag{15}$$

therefore the above formula (13) is the same as (10) in § 2b.

5. The $ARCH(p)$ models are also closely connected with autoregressive models *with random coefficients* used in the description of 'random walks in a random medium'.

To give an idea, we restrict ourselves again to $p = 1$.

Then it follows from the relations $h_n = \sigma_n \varepsilon_n$ and $\sigma_n^2 = \alpha_0 + \alpha_1 h_{n-1}^2$ that

$$h_n = \sqrt{\alpha_0 + \alpha_1 h_{n-1}^2}\, \varepsilon_n. \tag{16}$$

We now consider the following first-order autoregressive model with *random* coefficients:

$$x_n = B_1 \eta_n x_{n-1} + B_0 \delta_n, \tag{17}$$

where (η_n) and (δ_n) are two independent standard Gaussian sequences.

From the standpoint of finite-dimensional distributions the sequence $x = (x_n)$ where we (for definiteness) set $x_0 = 0$, has the same structure as the sequence $\widetilde{x} = (\widetilde{x}_n)$ with

$$\widetilde{x}_n = \sqrt{B_0^2 + B_1^2 \widetilde{x}_{n-1}^2}\, \widetilde{\varepsilon}_n, \qquad \widetilde{x}_0 = 0, \tag{18}$$

and $\widetilde{\varepsilon} = (\widetilde{\varepsilon}_n)$ a standard Gaussian sequence.

Comparing (16) and (18) we see that $h = (h_n)$ and $\widetilde{x} = (\widetilde{x}_n)$ are constructed following the same *pattern.* Hence if $B_0^2 = \alpha_0$ and $B_1^2 = \alpha_1$, then the probabilistic laws governing the sequences $h = (h_n)$ and $\widetilde{x} = (\widetilde{x}_n)$ with $h_0 = \widetilde{x}_0 = 0$ are the same.

6. We now make the $ARCH(1)$ model slightly more complicated by assuming that the relations between the variables $h_n, n \geqslant 1$ are as follows:

$$h_n = \beta_0 + \beta_1 h_{n-1} + \sqrt{\alpha_0 + \alpha_1 h_{n-1}^2}\, \varepsilon_n. \tag{19}$$

In this case, it is said that $h = (h_n)$ is governed by the $AR(1)/ARCH(1)$ model or that $h = (h_n)$ satisfies the autoregressive scheme $AR(1)$ with $ARCH(1)$ noise $\big(\sqrt{\alpha_0 + \alpha_1 h_{n-1}^2}\, \varepsilon_n\big)_{n \geqslant 1}$.

This model is conditionally Gaussian, therefore we can represent the density $p_\theta(h_1, \dots, h_n)$ of the joint distribution P_θ of the variables $h_1, \dots, h_n$ for a fixed value of the parameter $\theta = (\alpha_0, \alpha_1, \beta_0, \beta_1)$ as follows ($h_0 = 0$):

$$p_\theta(h_1, \dots, h_n) = (2\pi)^{-n/2} \prod_{k=1}^{n} \left(\alpha_0 + \alpha_1 h_{k-1}^2\right)^{-1/2} \times \exp\left\{-\frac{1}{2} \sum_{k=1}^{n} \frac{(h_k - \beta_0 - \beta_1 h_{k-1})^2}{\alpha_0 + \alpha_1 h_{k-1}^2}\right\}. \tag{20}$$

As an example of the application of this representation we consider the problem of *estimation* (using the maximum likelihood method) of the unknown value of β_1 under the assumption that all other parameters, β_0, α_0, and α_1, *are known.*

The maximum likelihood estimator $\widehat{\beta}_1$ of β_1 is the root of the equation

$$\frac{d\mathsf{P}_{(\alpha_0,\alpha_1,\beta_0,\beta_1)}}{d\beta_1}(h_1, \dots, h_n) = 0.$$

In view of (20) and (19), we obtain

$$\widehat{\beta}_1 = \frac{\displaystyle\sum_{k=1}^{n} \frac{(h_k - \beta_0) h_{k-1}}{\alpha_0 + \alpha_1 h_{k-1}^2}}{\displaystyle\sum_{k=1}^{n} \frac{h_{k-1}^2}{\alpha_0 + \alpha_1 h_{k-1}^2}} \tag{21}$$

and

$$\widehat{\beta}_1 = \beta_1 + \frac{M_n}{\langle M \rangle_n}, \tag{22}$$

where

$$M_n = \sum_{k=1}^{n} \frac{h_{k-1}\varepsilon_k}{\sqrt{\alpha_0 + \alpha_1 h_{k-1}^2}}$$

is a martingale and

$$\langle M \rangle_n = \sum_{k=1}^{n} \frac{h_{k-1}^2}{\alpha_0 + \alpha_1 h_{k-1}^2} \tag{23}$$

is its quadratic characteristic (cf. (46) in § 2b).

Here $\langle M \rangle_n \to \infty$ (P-a.s.), therefore $\dfrac{M_n}{\langle M \rangle_n} \to 0$ (P-a.s.) by the strong law of large numbers for square integrable martingales (see (12) in § 1b and [439; Chapter VII, § 5]). Hence the estimators $\widehat{\beta}_1$ so obtained are strongly consistent in the following sense: $\mathsf{P}_\theta(\widehat{\beta}_1 \to \beta) = 1$ for $\theta = (\alpha_0, \alpha_1, \beta_0, \beta)$ with $\beta \in \mathbb{R}$.

7. We now consider the issue of the *prediction* of the price movement in the future under the assumption that the sequence $h = (h_n)$ is governed by the model $ARCH(p)$.

Since the sequence $h = (h_n)$ is a martingale difference, it follows that $\mathsf{E}(h_{n+m} \mid \mathscr{F}_n) = 0$. Hence the optimal (in the mean-square sense) estimator

$$\widehat{h}_{n+m} \equiv \mathsf{E}(h_{n+m} \mid \mathscr{F}_n^h) = \mathsf{E}\big(\mathsf{E}(h_{n+m} \mid \mathscr{F}_n) \mid \mathscr{F}_n^h\big)$$

is equal to 0, so that it seems reasonable to consider the prediction problem for the future values of *nonlinear* functions of h_{n+m}; e.g., of h_{n+m}^2 and $|h_{n+m}|$.

We have

$$\begin{aligned}
\widehat{h_{n+m}^2} &\equiv \mathsf{E}(h_{n+m}^2 \mid \mathscr{F}_n^h) = \mathsf{E}(\sigma_{n+m}^2 \varepsilon_{n+m}^2 \mid \mathscr{F}_n^h) \\
&= \mathsf{E}\big[\mathsf{E}(\sigma_{n+m}^2 \varepsilon_{n+m}^2 \mid \mathscr{F}_{n+m-1}^\varepsilon) \mid \mathscr{F}_n^h\big] \\
&= \mathsf{E}[\sigma_{n+m}^2 \mid \mathscr{F}_n^h] \quad \big(\equiv \widehat{\sigma_{n+m}^2}\,\big)
\end{aligned} \tag{24}$$

(here $\mathscr{F}_n^h = \sigma(h_1, \dots, h_n)$), therefore it is clear that this question on the prediction of the future values of h_{n+m}^2 reduces to the problem of the prediction of the 'volatility' σ_{n+m}^2 on the basis of the past observations $h_0, h_1, \dots, h_n$.

Since

$$\sigma_{n+m}^2 = \alpha_0 + \alpha_1 \sigma_{n+m-1}^2 \varepsilon_{n+m-1}^2,$$

it follows by induction that

$$\begin{aligned}
\sigma_{n+m}^2 &= \alpha_0 + \alpha_1[\alpha_0 + \alpha_1 \sigma_{n+m-2}^2 \varepsilon_{n+m-2}^2]\varepsilon_{n+m-1}^2 \\
&= \cdots = \\
&= \alpha_0 + \alpha_0 \sum_{j=1}^{m-1} \prod_{i=1}^{j} \alpha_1 \varepsilon_{n+j-i+1}^2 + \sigma_n^2 \prod_{i=1}^{m} \alpha_1 \varepsilon_{n+m-i}^2.
\end{aligned}$$

Hence, considering the conditional expectation $\mathsf{E}(\,\cdot \mid \mathscr{F}_n^h)$ and taking into account the independence of the variables in the sequence (ε_n), we obtain

$$\begin{aligned}
\widehat{h_{n+m}^2} = \widehat{\sigma_{n+m}^2} &\equiv \mathsf{E}(\sigma_{n+m}^2 \mid \mathscr{F}_n^h) = \alpha_0 + \alpha_0 \sum_{j=1}^{m-1} \alpha_1^j + h_n^2 \alpha_1^m \\
&= \alpha_0 \frac{1 - \alpha_1^m}{1 - \alpha_1} + \alpha_1^m h_n^2.
\end{aligned} \tag{25}$$

As we might expect, the estimators $\widehat{h_{n+m}^2}$ converge as $m \to \infty$ (with probability one) to the 'stationary' value $\mathsf{E}h_n^2 \equiv \dfrac{\alpha_0}{1 - \alpha_1}$.

8. We recall that the intervals $(\mu-\sigma,\ \mu+\sigma)$ and $(\mu-1.65\sigma,\ \mu+1.65\sigma)$ are (with a certain degree of precision) confidence intervals for the normal distribution $\mathscr{N}(\mu,\sigma^2)$ with confidence levels 68 % and 90 %, respectively.

Since

$$S_{n+m} = S_n e^{h_{n+1}+\cdots+h_{n+m}} \tag{26}$$

and

$$\begin{aligned} \mathsf{E}\big[(h_{n+1}+\cdots+h_{n+m})^2 \mid \mathscr{F}_n^h\big] &= \mathsf{E}(h_{n+1}^2 \mid \mathscr{F}_n^h) + \cdots + \mathsf{E}(h_{n+m}^2 \mid \mathscr{F}_n^h) \\ &= \widehat{\sigma_{n+1}^2} + \cdots + \widehat{\sigma_{n+m}^2}, \end{aligned}$$

it follows that we can *as a first approximation* take the intervals

$$\Big(S_n e^{-\sqrt{\widehat{\sigma_{n+1}^2}+\cdots+\widehat{\sigma_{n+m}^2}}},\ S_n e^{+\sqrt{\widehat{\sigma_{n+1}^2}+\cdots+\widehat{\sigma_{n+m}^2}}}\Big)$$

and

$$\Big(S_n e^{-1.65\sqrt{\widehat{\sigma_{n+1}^2}+\cdots+\widehat{\sigma_{n+m}^2}}},\ S_n e^{+1.65\sqrt{\widehat{\sigma_{n+1}^2}+\cdots+\widehat{\sigma_{n+m}^2}}}\Big),$$

as confidence intervals (with confidence levels 68 % and 90 %, respectively).

Here it must be clear that writing about 'first approximation' we mean that the variables h_k are not normally distributed in general, and the question of the difference between the actual confidence levels of the above confidence intervals and 68 % (or 90 %) calls for an additional investigation of the precision of the normal approximation.

9. The success of the conditionally Gaussian $ARCH(p)$ model, which provided explanations for a variety of phenomena that could be distinguished in the behavior of the financial indexes (the 'cluster property', the 'heavy tails', the peaks (*leptokurtosis*) of the distribution densities of the variables $h_n,\dots$) inspired an avalanche of its generalizations trying to 'capture', explain several other phenomena discovered by means of statistical analysis.

Historically, one of the first generalizations of the $ARCH(p)$ model was (as already mentioned in § 1d) the generalized *ARCH model* of T. Bollerslev [48] (1986). Characterized by two parameters p and q, it is called the $GARCH(p,q)$ model.

In this model, just as in $ARCH(p)$, we set $h_n = \sigma_n\varepsilon_n$. As regards the 'volatility' σ_n, however, we assume that

$$\sigma_n^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i h_{n-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{n-j}^2, \tag{27}$$

where $\alpha_0 > 0$, $\alpha_i \geqslant 0$, and $\beta_j \geqslant 0$. (If all the β_j vanish, then we obtain the $ARCH(p)$ model.)

The central advantage of the $GARCH(p,q)$ models, as compared with their forefather, the $ARCH(p)$ model, is the following (experimental) fact: in the adjustments of the $GARCH(p,q)$ model to the statistical data one can restrict oneself to *small* values of p and q, while in the framework of the $ARCH(p)$ model we must often consider uncomfortably *large* values of p. (In [431] the author had to use autoregressive models of order twelve, $AR(12)$, to describe the monthly averages of S&P500 Index. See also the review paper [141].)

One can carry out the analysis of the $GARCH(p,q)$ models with 'volatility' σ_n that is assumed to depend (in a predictable manner) both on the h^2_{n-i}, $i \leqslant p$, and the σ^2_{n-j}, $j \leqslant q$, in a similar way to our analysis of the $ARCH(p)$ models.

Omitting the details we present several simple formulas relating to the $GARCH(1,1)$ model, in which

$$h_n = \sigma_n \varepsilon_n \quad \text{and} \quad \sigma^2_n = \alpha_0 + \alpha_1 h^2_{n-1} + \beta_1 \sigma^2_{n-1} \tag{28}$$

with $\alpha_0 > 0$, $\alpha_1 \geqslant 0$, and $\beta_1 \geqslant 0$.

It is clear that

$$\mathsf{E} h^2_n = \alpha_0 + (\alpha_1 + \beta_1) \mathsf{E} h^2_{n-1}$$

and the 'stationary' value $\mathsf{E} h^2_n$ is well defined for $\alpha_1 + \beta_1 < 1$; namely,

$$\mathsf{E} h^2_n \equiv \frac{\alpha_0}{1 - \alpha_1 - \beta_1}. \tag{29}$$

If $3\alpha_1^2 + 2\alpha_1\beta_1 + \beta_1^2 < 1$, then we have a well-defined 'stationary' value

$$\mathsf{E} h^4_n \equiv \frac{3\alpha_0^2(1 + \alpha_1 + \beta_1)}{(1 - \alpha_1 - \beta_1)(1 - \beta_1^2 - 2\alpha_1\beta_1 - 3\alpha_1^2)}, \tag{30}$$

and therefore, for the 'stationary kurtosis' we obtain

$$K \equiv \frac{\mathsf{E} h^4_n}{(\mathsf{E} h^2_n)^2} - 3 = \frac{6\alpha_1^2}{1 - \beta_1^2 - 2\alpha_1\beta_1 - 3\alpha_1^2}. \tag{31}$$

It is also easy to find the 'stationary' values of the autocorrelation function $\rho(k)$ (cf. (13)):

$$\rho(1) = \frac{\alpha_1(1 - \alpha_1\beta_1 - \beta_1^2)}{1 - 2\alpha_1\beta_1 - \beta_1^2}, \tag{32}$$

$$\rho(k) = (\alpha_1 + \beta_1)^{k-1} \rho(1), \qquad k > 1. \tag{33}$$

Finally, we point out that we can generalize (25) to the case of the $GARCH(1,1)$-models as follows:

$$\begin{aligned} \widehat{h^2_{n+m}} = \widehat{\sigma^2_{n+m}} &\equiv \mathsf{E}(\sigma^2_{n+m} \mid \mathscr{F}^h_n) \\ &= \alpha_0 \frac{1 - \gamma^m}{1 - \gamma} + \gamma^{m-1}(\alpha_1 h^2_n + \beta_1 \sigma^2_n), \end{aligned}$$

where $\gamma = \alpha_1 + \beta_1$.

10. Models in the *ARCH* family, which evolve in discrete time, have counterparts in the continuous time case. Moreover, after a suitable normalization, we obtain a (weak) convergence of the solutions of the stochastic difference equations characterizing *ARCH*, *GARCH*, and other models to the solutions of the corresponding differential equations.

For definiteness, we now consider the following modification of the $GARCH(1,1)$ model (which is called the $GARCH(1,1)$-M model in [15]).

Let Δ be the time step, and let $H^{(\Delta)} = (H^{(\Delta)}_{k\Delta})$, $k = 0, 1, \dots$, where $H^{(\Delta)}_{k\Delta} = H^{(\Delta)}_0 + h_\Delta{}^{(\Delta)} + \dots + h^{(\Delta)}_{k\Delta}$ and

$$h^{(\Delta)}_{k\Delta} = c(\sigma^{(\Delta)}_{k\Delta})^2 + (\sigma^{(\Delta)}_{k\Delta})\varepsilon_{k\Delta} \tag{34}$$

with $\varepsilon_{k\Delta} \sim \mathscr{N}(0, \Delta)$, a constant c, and

$$(\sigma^{(\Delta)}_{k\Delta})^2 = \alpha_0(\Delta) + (\sigma^{(\Delta)}_{(k-1)\Delta})^2(\beta(\Delta) + \alpha_1(\Delta)\varepsilon^2_{(k-1)\Delta}). \tag{35}$$

We set the initial condition $H^{(\Delta)}_0 = H_0$ and $\sigma^{(\Delta)}_0 = \sigma_0$ for all $\Delta > 0$, where (H_0, σ_0) is a pair of random variables independent of the Gaussian sequences $(\varepsilon_{k\Delta})$, $\Delta > 0$, of independent random variables.

We now embed the sequence $(H^{(\Delta)}, \sigma^{(\Delta)}) = (H^{(\Delta)}_{k\Delta}, \sigma^{(\Delta)}_{k\Delta})_{k\geqslant 0}$ in a scheme with continuous time $t \geqslant 0$ by setting

$$H^{(\Delta)}_t = H^{(\Delta)}_{k\Delta} \quad \text{and} \quad \sigma^{(\Delta)}_t = \sigma^{(\Delta)}_{k\Delta} \tag{36}$$

for $k\Delta \leqslant t < (k+1)\Delta$.

In view of the general results of the theory of weak convergence of random processes (see, e.g., [250] and [304]) it seems natural to expect that, under certain conditions on the coefficients in (34) and (35), the sequence of processes $(H^{(\Delta)}, \sigma^{(\Delta)})$ weakly converges (as $\Delta \to 0$, in the Skorokhod space D) to some diffusion process $(H, \sigma) = (H_t, \sigma_t)_{t\geqslant 0}$.

As shown in [364], for

$$\alpha_0(\Delta) = \alpha_0\Delta, \quad \alpha_1(\Delta) = \alpha\sqrt{\frac{\Delta}{2}}, \quad \beta(\Delta) = 1 - \alpha\sqrt{\left(\frac{\Delta}{2}\right)} - \beta\Delta,$$

the limiting process (H, σ) satisfies the following stochastic differential equations (see Chapter III, § 3e):

$$dH_t = c\sigma_t^2\,dt + \sigma_t\,dW^{(1)}_t, \tag{37}$$

$$d\sigma_t^2 = (\alpha_0 - \beta\sigma_t^2)\,dt + \alpha\sigma_t^2\,dW^{(2)}_t, \tag{38}$$

where $(W^{(1)}, W^{(2)})$ are two independent standard Brownian motions that are also independent of the initial values $(H_0, \sigma_0) \equiv (H^{(\Delta)}_0, \sigma^{(\Delta)}_0)$.

§ 3b. *EGARCH*, *TGARCH*, *HARCH*, and Other Models

1. In 1976, F. Black noticed the following phenomenon in the behavior of financial indexes: the variables h_{n-1} and σ_n are *negatively correlated*; namely, the empiric covariance $\mathsf{Cov}(h_{n-1}, \sigma_n)$ is negative.

This phenomenon, which is called the *leverage effect* (or also the *asymmetry effect*), is responsible for the trend of growth in volatility after a drop in prices (i.e., when the logarithmic returns become negative). This cannot be understood in the framework of *ARCH* or *GARCH* models, where the volatility σ_n^2, which depends on the *squares* of the h_{n-i}^2, $i \geqslant 1$, is indifferent to the *signs* of the h_{n-j}, so that in the *GARCH* models the values $h_{n-j} = a$ and $h_{n-j} = -a$ result in the *same* value of future volatility σ_n^2.

To explain Black's discovery D. B. Nelson [366] put forward (1990) the so-called *EGARCH*(p,q) (*Exponential GARCH*(p,q)) model, in which the 'asymmetry' was taken into account by means of the replacement of $h_{n-i}^2 = \sigma_{n-i}^2 \varepsilon_{n-i}^2$ in the *GARCH* models with linear combinations of the variables ε_{n-i} and $|\varepsilon_{n-i}|$. Namely, we assume again that $h_n = \sigma_n \varepsilon_n$, but the σ_n must now satisfy the following relations:

$$\ln \sigma_n^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i \left[\theta \varepsilon_{n-i} + \gamma \left(|\varepsilon_{n-i}| - \sqrt{\frac{2}{\pi}} \right) \right] + \sum_{j=1}^{q} \beta_j \ln \sigma_{n-j}^2. \tag{1}$$

(We note that $\sqrt{2/\pi} = \mathsf{E}|\varepsilon_{n-i}|$.)

Since $h_{n-i} = \sigma_{n-i}\varepsilon_{n-i}$ and $\sigma_{n-i} \geqslant 0$, the signs of h_{n-i} and ε_{n-i} are the same. Hence if $\varepsilon_{n-i} = b > 0$, then the corresponding term in σ_n^2 is equal to $b(\theta + \gamma)$, while if $\varepsilon_{n-i} = -b < 0$, then it is equal to $b(-\theta + \gamma)$.

2. The *EGARCH* models are not unique in capturing the asymmetry effect while retaining the main properties of the *GARCH* family. Another example is the *TGARCH*(p,q) model ('T' as in 'threshold'), which was suggested by the threshold models of *TAR* (Threshold AR) type. In this model,

$$h_n = \sum_{i=1}^{k} I_{A_i}(h_{n-d})(\alpha_0^i + \alpha_1^i h_{n-1} + \cdots + \alpha_p^i h_{n-p}), \tag{2}$$

where d is a lag parameter and $A_1, \ldots, A_k$ are disjoint subsets of $\mathbb{R}$ such that $\sum\limits_{i=1}^{k} A_i = \mathbb{R}$.

For instance, we can set

$$h_n = \begin{cases} \alpha_0^1 + \alpha_1^1 h_{n-1} + \alpha_2^1 h_{n-2} & \text{if } h_{n-2} > 0, \\ \alpha_0^2 + \alpha_1^2 h_{n-1} + \alpha_2^2 h_{n-2} & \text{if } h_{n-2} \leqslant 0. \end{cases} \tag{3}$$

(Such threshold models were thoroughly investigated in the monograph [461].)

By definition (see [399]), a sequence $h = (h_n)$ is described by the $TGARCH(p, q)$ model if $h_n = \sigma_n \varepsilon_n$, where

$$\sigma_n = a_0 + \sum_{i=1}^{p} \left[a_i h_{n-i}^+ + b_i h_{n-i}^-\right] + \sum_{j=1}^{q} \left[c_j \sigma_{n-j}^+ + d_j \sigma_{n-j}^-\right] \tag{4}$$

and, as usual, $x^+ = \max(x, 0)$ and $x^- = -\min(x, 0)$. We do not assume in this model that the coefficients (and, therefore, the volatilities σ_n) are positive, although σ_n^2 retains its meaning of the conditional variance $\mathsf{E}(h_n^2 \mid \mathscr{F}_{n-1}^h)$.

Since

$$h_n = \sigma_n \varepsilon_n = (\sigma_n^+ - \sigma_n^-)(\varepsilon_n^+ - \varepsilon_n^-) = \left[\sigma_n^+ \varepsilon_n^+ + \sigma_n^- \varepsilon_n^-\right] - \left[\sigma_n^- \varepsilon_n^+ + \sigma_n^+ \varepsilon_n^-\right],$$

it follows that

$$h_n^+ = \left[\sigma_n^+ \varepsilon_n^+ + \sigma_n^- \varepsilon_n^-\right] \quad \text{and} \quad h_n^- = \left[\sigma_n^- \varepsilon_n^+ + \sigma_n^+ \varepsilon_n^-\right].$$

These relations enable one to rewrite (4) as follows:

$$\sigma_n = \alpha_0 + \sum_{i=1}^{p^*} \alpha_i(\varepsilon_{n-i}) \sigma_{n-i}^+ + \sum_{i=1}^{p^*} \beta_i(\varepsilon_{n-i}) \sigma_{n-i}^-, \tag{5}$$

where $p^* = \max(p, q)$ and the functions $\alpha_i(\varepsilon_{n-i})$ and $\beta_i(\varepsilon_{n-i})$, are linear combinations of ε_{n-i}^+ and ε_{n-i}^-.

The study of such models runs into certain technical difficulties rooted in the lack of the Markov property. Nevertheless, in simple cases (say, in the case of $p = q = 1$) one can analyze the properties of these models fairly completely.

Indeed, let $p = q = 1$. Then

$$\sigma_n = a_0 + \left[a_1 h_{n-1}^+ + b_1 h_{n-1}^-\right] + \left[c_1 \sigma_{n-1}^+ + d_1 \sigma_{n-1}^-\right], \tag{6}$$

or, equivalently,

$$\sigma_n = \alpha_0 + \alpha_1(\varepsilon_{n-1}) \sigma_{n-1}^+ + \beta_1(\varepsilon_{n-1}) \sigma_{n-1}^-, \tag{7}$$

where

$$\begin{aligned} \alpha_0 &= a_0, \\ \alpha_1(\varepsilon_{n-1}) &= a_1 \varepsilon_{n-1}^+ + b_1 \varepsilon_{n-1}^- + c_1, \\ \beta_1(\varepsilon_{n-1}) &= a_1 \varepsilon_{n-1}^- + b_1 \varepsilon_{n-1}^+ + d_1. \end{aligned} \tag{8}$$

If $\alpha_0 = 0$, then

$$\begin{aligned} \sigma_n^+ &= \left(\alpha_1(\varepsilon_{n-1})\right)^+ \sigma_{n-1}^+ + \left(\beta_1(\varepsilon_{n-1})\right)^+ \sigma_{n-1}^-, \\ \sigma_n^- &= \left(\alpha_1(\varepsilon_{n-1})\right)^- \sigma_{n-1}^+ + \left(\beta_1(\varepsilon_{n-1})\right)^- \sigma_{n-1}^- \end{aligned} \tag{9}$$

by (7). Hence it is clear that, with respect to the flow $(\mathscr{F}_n)$, the sequence $(\sigma_n^+, \sigma_n^-, \varepsilon_n)_{n \geqslant 1}$ is Markov, which enables one to study it by usual 'Markovian' methods. (See [399] for greater detail.)

3. We now consider another phenomenon, the 'long memory' or 'strong aftereffect' in the evolution of prices $S = (S_n)_{n\geqslant 0}$.

There exists several ways to describe the *dependence* on the 'past' of the variables in a random sequence. In probability theory one has various measures of this dependence: *ergodicity coefficients*, *mixing coefficients*, and so on.

For instance, we can measure the rate at which the dependence on the past in a stationary sequence of real-valued variables $X = (X_n)$ fades away by the rate of the convergence to zero (as $m \to \infty$) of the supremum

$$\sup_A \mathsf{E}\,\big|\mathsf{P}(X_{n+m} \in A \,|\, X_1, \dots, X_n) - \mathsf{P}(X_{n+m} \in A)\big|,$$

taken over all Borel sets $A \subseteq \mathbb{R}$.

Of course, the (auto)correlation function is the standard measure of this dependence.

It should be noted that, as shown by many statistical studies, financial time series exhibit a *stronger correlation dependence* between the variables in the sequences $|h| = (|h_n|)_{n\geqslant 1}$ and $h^2 = (h_n^2)_{n\geqslant 1}$ than the one attainable in the framework of *ARCH* or *GARCH* (not to mention *MA*, *AR*, or *ARMA*) models.

We recall that, by formula (13) in § 3a,

$$\mathsf{Corr}(h_{n-k}^2, h_n^2) = \alpha_1^k \quad \text{with } \alpha_1 < 1,$$

in the $ARCH(1)$ model, while the autocorrelation function $\rho(k)$ for the $GARCH(1,1)$ model is described by expressions (32) and (33) in the same § 3a. According to these formulas, the correlation in these models approaches zero at geometric rate ('the past is quickly forgotten').

One often says that a stationary (in the wide sense) sequence $Y = (Y_n)$ is a sequence with *'long memory'* or *'strong aftereffect'* if its autocorrelation function $\rho(k)$ approaches zero at *hyperbolic* rate, i.e.,

$$\rho(k) \sim c\,k^{-\rho}, \qquad k \to \infty, \tag{10}$$

for some $\rho > 0$.

This rate of decrease is characteristic, e.g., of the autocorrelation function of fractal Gaussian noise (see Chapter III, § 2d) $Y = (Y_n)_{n\geqslant 1}$ with elements

$$Y_n = X_n - X_{n-1},$$

where $X = (X_t)_{t\geqslant 0}$ is a fractal Brownian motion with parameter $\mathbb{H}$, $0 < \mathbb{H} < 1$ (see Chapter III, § 2c). For this motion we have (see (3) in Chapter III, § 2c)

$$\mathsf{Cov}(X_s, X_t) = \frac{1}{2}\big\{|t|^{2\mathbb{H}} + |s|^{2\mathbb{H}} - |t-s|^{2\mathbb{H}}\big\}\mathsf{E}X_1^2$$

and (see (3) in Chapter III, § 2d)

$$\mathsf{Cov}(Y_n, Y_{n+k}) = \frac{\sigma^2}{2}\{|k+1|^{2\mathbb{H}} - 2|k|^{2\mathbb{H}} + |k-1|^{2\mathbb{H}}\}, \tag{11}$$

where $\sigma^2 = \mathsf{D}Y_n$. Hence the autocorrelation function $\rho(k) = \mathsf{Corr}(Y_n, Y_{n+k})$ decreases hyperbolically as $k \to \infty$:

$$\rho(k) \sim \mathbb{H}(2\mathbb{H}-1)k^{2\mathbb{H}-2}.$$

We note that

$$\sum_{k=1}^{\infty} \rho(k) = \infty$$

for $\frac{1}{2} < \mathbb{H} < 1$.

For $\mathbb{H} = \frac{1}{2}$ (a usual Brownian motion) the variables $Y = (Y_n)$ form Gaussian 'white noise' with $\rho(k) = 0$, $k \geqslant 1$.

On the other hand, if $0 < \mathbb{H} < \frac{1}{2}$, then

$$\sum_{k=1}^{\infty} |\rho(k)| < \infty, \qquad \sum_{k=1}^{\infty} \rho(k) = 0.$$

Remark. In Chapter 7 of the monograph [202] one can find a discussion of various models of processes with 'strong aftereffect' and much information about the applications of these models in economics, biology, hydrology, and so on. See also [418].

4. Another model, $HARCH(p)$, was introduced and analyzed in [360] and [89]. This is a model from the $ARCH$ family in which the autocorrelation functions for the absolute values and the squares of the variables h_n decrease *slower* than in the case of models of $ARCH(p)$ or $GARCH(p,q)$ kinds. The same 'long memory' phenomenon is characteristic of the $FIGARCH$ models introduced in [15].

By definition, the $HARCH(p)$ (*Heterogeneous AutoRegressive Conditional Heteroskedastic*) model (of order p) is defined by the relation

$$h_n = \sigma_n \varepsilon_n,$$

where

$$\sigma_n^2 = \alpha_0 + \sum_{j=1}^{p} \alpha_j \left(\sum_{i=1}^{j} h_{n-i}\right)^2$$

with $\alpha_0 > 0$, $\alpha_p > 0$, and $\alpha_j \geqslant 0$, $j = 1, \dots, p-1$.

In particular, for $p = 1$ we have

$$\sigma_n^2 = \alpha_0 + \alpha_1 h_{n-1}^2,$$

i.e., $HARCH(1) = ARCH(1)$.

For $p = 2$,

$$\sigma_n^2 = \alpha_0 + \alpha_1 h_{n-1}^2 + \alpha_2 (h_{n-1} + h_{n-2})^2. \tag{12}$$

We now consider several properties of this model.

First, we point out that the presence of the term $(h_{n-1} + h_{n-2})^2$ enables the model to 'capture' the above-mentioned asymmetry effects.

Further, if $\alpha_0 + \alpha_1 + \alpha_2 < 1$, then it follows from (12) that there exists a 'stationary' value

$$\mathsf{E}h_n^2 = \mathsf{E}\sigma_n^2 = \frac{\alpha_0}{1 - \alpha_1 - 2\alpha_2}. \tag{13}$$

In a similar way, considering $\mathsf{E}\sigma_n^4$ and using the equalities

$$\mathsf{E}h_{n-1}h_{n-2} = \mathsf{E}h_{n-1}^3 h_{n-2} = \mathsf{E}h_{n-1}h_{n-2}^3 = 0,$$

we obtain by (12) that for $(\alpha_1 + \alpha_2)^2 + \alpha_2^2 < \frac{1}{3}$, the 'stationary value'

$$\mathsf{E}h_n^4 = \frac{C}{\frac{1}{3} - (\alpha_1 + \alpha_2)^2 - \alpha_2^2}, \tag{14}$$

is well defined, where

$$C = \frac{\alpha_0^2[1 + 2\alpha_2(\alpha_1 + 3\alpha_2) - (\alpha_1 + 2\alpha_2)^2]}{[1 - (\alpha_1 + 2\alpha_2)]^2}. \tag{15}$$

(We note that $\mathsf{E}\sigma_n^4 = \frac{1}{3}\mathsf{E}h_n^4$.)

We now find the autocorrelation function for (h_n^2).

Let $R(k) = \mathsf{E}h_n^2 h_{n-k}^2$. Then in the 'stationary' case, for $k = 1$ we have

$$\begin{aligned} R(1) &= \mathsf{E}\sigma_n^2 \varepsilon_n^2 h_{n-1}^2 = \mathsf{E}\sigma_n^2 h_{n-1}^2 \\ &= \mathsf{E}\big(h_{n-1}^2\big[\alpha_0 + \alpha_1 h_{n-1}^2 + \alpha_2(h_{n-1} + h_{n-2})^2\big]\big) \\ &= \alpha_0 \mathsf{E}h_{n-1}^2 + \alpha_1 \mathsf{E}h_{n-1}^4 + \alpha_2 \mathsf{E}h_{n-1}^4 + 2\alpha_2 \mathsf{E}h_{n-1}^3 h_{n-2} + \alpha_2 \mathsf{E}h_{n-1}^2 h_{n-2}^2. \end{aligned}$$

Consequently, if $\alpha_2 < 1$, then

$$R(1) = \frac{\alpha_0 \mathsf{E}h_{n-1}^2 + (\alpha_1 + \alpha_2)\mathsf{E}h_{n-1}^4}{1 - \alpha_2}. \tag{16}$$

Further,

$$\begin{aligned} R(k) &= \mathsf{E}h_n^2 h_{n-k}^2 = \mathsf{E}\sigma_n^2 h_{n-k}^2 \\ &= \mathsf{E}\big[\alpha_0 + \alpha_1 h_{n-1}^2 + \alpha_2(h_{n-1} + h_{n-2})^2\big] h_{n-k}^2 \\ &= \alpha_0 \mathsf{E}h_{n-k}^2 + (\alpha_1 + \alpha_2)R(k-1) + \alpha_2 R(k-2), \end{aligned}$$

where $R(0) = \mathsf{E}h_n^4$.

Hence the autocorrelation function $\rho(k) = \mathrm{Corr}(h_n^2, h_{n-k}^2)$ clearly satisfies in the stationary case the equation

$$\rho(k) = A + B\rho(k-1) + C\rho(k-2), \quad k \geqslant 2, \tag{17}$$

where

$$A = \frac{\alpha_0 \mathsf{E} h_n^2}{\mathsf{D} h_n^2}, \qquad B = \frac{(\alpha_1 + \alpha_2)(\mathsf{E} h_n^2)^2}{\mathsf{D} h_n^2}, \qquad C = \frac{(\alpha_2 - 1)(\mathsf{E} h_n^2)^2}{\mathsf{D} h_n^2},$$

and

$$\rho(0) = 1, \qquad \rho(1) = \frac{R(1) - (\mathsf{E} h_n^2)^2}{\mathsf{D} h_n^2}.$$

We continue our analysis of the 'strong aftereffect' in Chapter IV, § 3e, while discussing exchange rates.

§ 3c. Stochastic Volatility Models

1. A characteristic feature of these models, introduced in § 1d, is the existence of *two sources* of randomness, $\varepsilon = (\varepsilon_n)$ and $\delta = (\delta_n)$, governing the behavior of the sequence $h = (h_n)$ so that

$$h_n = \sigma_n \varepsilon_n, \tag{1}$$

where $\sigma_n = e^{\frac{1}{2}\Delta_n}$ and the sequence (Δ_n) is in the class $AR(p)$, i.e.,

$$\Delta_n = a_0 + \sum_{i=1}^{p} a_i \Delta_{n-i} + c\delta_n. \tag{2}$$

We shall assume that $\varepsilon = (\varepsilon_n)$ and $\delta = (\delta_n)$ are independent standard Gaussian sequences. Then we shall say that $h = (h_n)$ is governed by the $SV(p)$ (Stochastic Volatility) model.

We now consider the properties of this model in the case of $p = 1$ and $|a_1| < 1$. We have

$$h_n = \sigma_n \varepsilon_n, \qquad \ln \sigma_n^2 = a_0 + a_1 \ln \sigma_{n-1}^2 + c\delta_n. \tag{3}$$

Let $\mathscr{F}_n^{\varepsilon,\delta} = \sigma(\varepsilon_1, \dots, \varepsilon_n; \delta_1, \dots, \delta_n)$ and let $\mathscr{F}_n^{\delta} = \sigma(\delta_1, \dots, \delta_n)$. Clearly,

$$\mathsf{E}(h_n \mid \mathscr{F}_n^{\delta}) = \sigma_n \mathsf{E}\varepsilon_n = 0$$

and

$$\begin{aligned}
\mathsf{E}(h_n \mid \mathscr{F}_{n-1}^{\varepsilon,\delta}) &= \mathsf{E}(\sigma_n \varepsilon_n \mid \mathscr{F}_{n-1}^{\varepsilon,\delta}) \\
&= \mathsf{E}\big(\sigma_n \mathsf{E}(\varepsilon_n \mid \mathscr{F}_{n-1}^{\varepsilon,\delta} \vee \sigma(\delta_n)) \mid \mathscr{F}_{n-1}^{\varepsilon,\delta}\big) \\
&= \mathsf{E}\big(\sigma_n \mathsf{E}(\varepsilon_n \mid \mathscr{F}_{n-1}^{\varepsilon,\delta})\big) = 0
\end{aligned}$$

because $\mathsf{E}\varepsilon_n = 0$.

Hence the sequence $h = (h_n)$ is a martingale difference with respect to the flow $(\mathscr{F}_{n-1}^{\varepsilon,\delta})$. (Although not with respect to $(\mathscr{F}_n^{\delta})$ because the h_n are not $\mathscr{F}_n^{\delta}$-measurable.)

Further,

$$\mathsf{E}h_n^2 = \mathsf{E}\sigma_n^2\,\mathsf{E}\varepsilon_n^2 = \mathsf{E}\sigma_n^2 = \mathsf{E}e^{\Delta_n}.$$

We shall assume that

$$\Delta_0 \sim \mathscr{N}\left(\frac{a_0}{1-a_1}, \frac{c^2}{1-a_1^2}\right). \tag{4}$$

By (3), the sequence $\Delta = (\Delta_n)$ fits into the autoregressive scheme $AR(1)$ (i.e., $\Delta_n = a_0 + a_1\Delta_{n-1} + c\delta_n$) and is stationary (see § 2b).

By (4),

$$\mathsf{E}h_n^2 = \mathsf{E}e^{\Delta_n} = e^{\frac{a_0}{1-a_1}}\, e^{\frac{c^2}{2(1-a_1^2)}},$$

where we evaluate $\mathsf{E}e^{\Delta_n}$ using the fact that

$$\mathsf{E}e^{\sigma\xi - \frac{1}{2}\sigma^2} = 1$$

for each σ and a random variable $\xi \sim \mathscr{N}(0,1)$.

In a similar way,

$$\mathsf{E}|h_n| = \mathsf{E}|\varepsilon_n|\,\mathsf{E}\sigma_n = \sqrt{\frac{2}{\pi}}\,\mathsf{E}e^{\frac{1}{2}\Delta_n} = \sqrt{\frac{2}{\pi}}\, e^{\frac{a_0}{2(1-a_1)}}\, e^{\frac{1}{8}\cdot\frac{c^2}{1-a_1^2}}.$$

We now consider the covariance properties of the sequences $h = (h_n)$ and $h^2 = (h_n^2)$.

We have

$$\mathsf{E}h_n h_{n+1} = 0$$

and, more generally,

$$\mathsf{E}h_n h_{n+k} = 0$$

for each $k \geqslant 1$. Hence $h = (h_n)$ is a sequence of uncorrelated random variables: if $R_h(k) = \mathsf{E}h_n h_{n+k}$, then

$$R_h(k) = \begin{cases} \mathsf{E}h_n^2, & k = 0, \\ 0, & k > 0. \end{cases}$$

Further,

$$\begin{aligned}
\mathsf{E}h_n^2 h_{n-1}^2 &= \mathsf{E}\sigma_n^2\sigma_{n-1}^2 = \mathsf{E}e^{\Delta_n+\Delta_{n-1}} \\
&= \mathsf{E}\big(e^{\Delta_{n-1}}\mathsf{E}(e^{a_0+a_1\Delta_{n-1}+c\delta_n} \mid \Delta_{n-1})\big) \\
&= \mathsf{E}e^{a_0+\Delta_{n-1}(1+a_1)}\mathsf{E}e^{c\delta_n} = e^{a_0+\frac{c^2}{2}}\,\mathsf{E}e^{(1+a_1)\Delta_{n-1}} \\
&= e^{a_0+\frac{c^2}{2}}\, e^{\frac{a_0(1+a_1)}{1-a_1}}\,\mathsf{E}e^{(1+a_1)(\Delta_{n-1}-\frac{a_0}{1-a_1})} \\
&= e^{\frac{2a_0}{1-a_1}+\frac{c^2}{2}}\, e^{\frac{(1+a_1)^2}{2}\cdot\frac{c^2}{1-a_1^2}} = e^{\frac{2a_0}{1-a_1}+\frac{c^2}{2}}\, e^{\frac{c^2}{2}\cdot\frac{1+a_1}{1-a_1}} \\
&= e^{\frac{2a_0}{1-a_1}+\frac{c^2}{2}\cdot\frac{2}{1-a_1}} = e^{\frac{2a_0+c^2}{1-a_1}}.
\end{aligned}$$

Hence

$$\begin{aligned}
\mathsf{Cov}(h_n^2, h_{n-1}^2) &= e^{\frac{2a_0+c^2}{1-a_1}} - e^{\frac{2a_0}{1-a_1}}\, e^{\frac{c^2}{1-a_1^2}} \\
&= e^{\frac{2a_0}{1-a_1}}\, e^{\frac{c^2}{1-a_1^2}}\Big(e^{\frac{c^2a_1}{1-a_1^2}} - 1\Big).
\end{aligned}$$

As might be expected, the variables h_n^2 and h_{n-1}^2 are *positively* correlated for $a_1 > 0$ and *negatively* correlated for $a_1 < 0$.

Besides the above formulas

$$\mathsf{E}\sigma_n^2 = e^{\frac{a_0}{1-a_1}}\, e^{\frac{c^2}{2(1-a_1^2)}},$$

$$\mathsf{E}\sigma_n^2\sigma_{n-1}^2 = e^{\frac{2a_0}{1-a_1}}\, e^{\frac{c^2}{1-a_1}},$$

we present also more general ones. Namely, for positive constants r and s we have

$$\mathsf{E}\sigma_n^r = e^{\frac{ra_0}{2(1-a_1)}+\frac{r^2}{8}\cdot\frac{c^2}{1-a_1^2}},$$

$$\mathsf{E}\sigma_n^r\sigma_{n-1}^s = \mathsf{E}\sigma_n^r\,\mathsf{E}\sigma_n^s e^{rs\frac{c^2}{4(1-a_1^2)}}.$$

These formulas can be used in the calculations of various moments of the h_n. For example,

$$\mathsf{E}|h_n| = \sqrt{\frac{2}{\pi}}\,\mathsf{E}\sigma_n = \sqrt{\frac{2}{\pi}}\, e^{\frac{a_0}{2(1-a_1)}+\frac{c^2}{8(1-a_1^2)}},$$

$$\mathsf{E}h_n^4 = 3\mathsf{E}\sigma_n^4,$$

$$\mathsf{E}h_n^2h_{n-k}^2 = \mathsf{E}\sigma_n^2\sigma_{n-k}^2,$$

$$\mathsf{E}|h_nh_{n-k}| = \frac{2}{\pi}\,\mathsf{E}\sigma_n\sigma_{n-k}.$$

In particular, these relations yield the following expression for the 'stationary' kurtosis:

$$K = \frac{\mathsf{E}h_n^4}{(\mathsf{E}h_n^2)^2} - 3 = 3\left(\frac{\mathsf{E}\sigma_n^4 - (\mathsf{E}\sigma_n^2)^2}{(\mathsf{E}\sigma_n^2)^2}\right) = 3\,\frac{\mathsf{D}\sigma_n^2}{(\mathsf{E}\sigma_n^2)^2} \geqslant 0,$$

which shows that the 'stochastic volatility' models with two sources of randomness $\varepsilon = (\varepsilon_n)$ and $\delta = (\delta_n)$ enable one (in a similar way to the *ARCH* family) to generate sequences $h = (h_n)$ such that with the distribution densities for the h_n have peaks around the mean value $\mathsf{E}h_n = 0$ (the *leptokurtosis* phenomenon).

2. We now dwell on the issue of constructing volatility estimators $\widehat{\sigma}_n$ on the basis of the observations $h_1, \dots, h_n$.

If $h_n = \mu + \sigma_n\varepsilon_n$, then $\mathsf{E}h_n = \mu$ and

$$\mathsf{E}|h_n - \mu| = \mathsf{E}|\sigma_n\varepsilon_n| = \sqrt{\frac{2}{\pi}}\,\mathsf{E}\sigma_n.$$

It is natural to make this relation the starting point in our construction of the estimators $\widehat{\sigma}_n$ of the volatilities σ_n by adopting the formula

$$\widehat{\sigma}_n = \sqrt{\frac{\pi}{2}}\,|h_n - \overline{h}_n| \tag{5}$$

with

$$\overline{h}_n = \frac{1}{n}\sum_{k=1}^{n} h_k$$

if μ is unknown and the formula

$$\widehat{\sigma}_n = \sqrt{\frac{\pi}{2}}\,|h_n - \mu| \tag{6}$$

if μ is known.

Another estimation method for the σ_n^2 is based on the fact that $\mathsf{E}h_n^2 = \mathsf{E}\sigma_n^2$, i.e., on the properties of second-order moments.

Clearly, we could choose $\widehat{\sigma_n^2} = h_n^2$ as an estimator for σ_n^2. This is a nonbiased estimator, but its mean squared error

$$\begin{aligned}\mathsf{E}\big|\widehat{\sigma_n^2} - \sigma_n^2\big|^2 &= \mathsf{E}\big|h_n^2 - \sigma_n^2\big|^2 = \mathsf{E}h_n^4 - 2\mathsf{E}h_n^2\sigma_n^2 + \mathsf{E}\sigma_n^4 \\ &= 4\mathsf{E}\sigma_n^4 - 2\mathsf{E}\sigma_n^4 = 2\mathsf{E}\sigma_n^4 = 2\exp\left\{\frac{2a_0}{1-a_1} + \frac{2c^2}{1-a_1^2}\right\}\end{aligned}$$

can be fairly large.

Of course, if the variables σ_k^2, $k \leqslant n$, are correlated, then we can use not only h_n^2, but also the preceding observations $h_{n-1}^2, h_{n-2}^2, \dots$ to construct estimators for σ_n^2.

It is clear that if the σ_k^2, $k \leqslant n$, are *weakly* correlated, then the past variables h_{n-1}^2, $h_{n-2}^2, \dots$ must enter our formulas with small, decreasing weights. On the other hand, if the σ_k^2, $k \leqslant n$, are *strongly* correlated, then the variables h_{n-1}^2, $h_{n-2}^2, \dots$ can be a source of important additional information about σ_n^2 (compared with the information contained in h_n^2).

This idea brings one to the consideration of *exponentially weighted* estimators

$$\widetilde{\sigma_n^2} = (1-\lambda)\sum_{k=0}^{\infty} \lambda^k h_{n-k}^2, \qquad 0 < \lambda < 1, \tag{7}$$

in which, as we see, $-\infty$, rather than time 0, is the starting point. Making this assumption we obtain the following formula for the stationary solution of the autoregressive recursion relation (2):

$$\Delta_n = \frac{a_0}{1-a_1} + c\sum_{k=0}^{\infty} a_1^k \delta_{n-k}, \tag{8}$$

where the series converges in mean square. As shown in § 2b, this is the unique stationary solution.

We note that $(1-\lambda)\sum\limits_{k=0}^{\infty} \lambda^k = 1$, i.e., the sum of the weighted coefficients involved in the construction of $\widetilde{\sigma_n^2}$ is equal to one.

Since $\mathsf{E}h_{n-k}^2 = \mathsf{E}\sigma_{n-k}^2 = \mathsf{E}\sigma_n^2$, it follows that $\mathsf{E}\widetilde{\sigma_n^2} = \mathsf{E}\sigma_n^2$, which means that the estimator $\widetilde{\sigma_n^2}$ is unbiased (as is $\widehat{\sigma_n^2}$).

We also point out that the quality of the estimator $\widetilde{\sigma_n^2}$ considerably depends on the chosen value of the parameter λ, so that there arises the (fairly complicated) problem of the 'optimal' choice of λ.

By (7) we see that the $\widetilde{\sigma_n^2}$ satisfy the recursion relations

$$\widetilde{\sigma_n^2} = \lambda\widetilde{\sigma_{n-1}^2} + (1-\lambda)h_n^2, \tag{9}$$

which is convenient in the construction of estimators by means of statistical analysis and model-building.

3. In our considerations of the model

$$h_n = e^{\frac{1}{2}\Delta_n}\,\varepsilon_n, \qquad (\Delta_n = \ln\sigma_n^2), \tag{10}$$

where $\Delta_n = a_0 + a_1\Delta_{n-1} + c\delta_n$, it might seem reasonable to use

$$m_n = \mathsf{E}(\Delta_n \mid h_1, \dots, h_n)$$

as an estimator for Δ_n. This estimator would be optimal in the mean square sense. Unfortunately, this is a *nonlinear* scheme, which makes the problem of finding an *explicit* formula for m_n almost hopeless. The first idea coming naturally to mind is to 'linearize' this problem and, after that, to use the theory of 'Gaussian linear filtering' of R. Kalman and R. Bucy (see, e.g., [303]).

For example, we can proceed as follows.

By (10) we obtain

$$\ln h_n^2 = \ln \varepsilon_n^2 + \ln \sigma_n^2 = \mathsf{E} \ln \varepsilon_n^2 + \Delta_n + \left(\ln \varepsilon_n^2 - \mathsf{E} \ln \varepsilon_n^2\right),$$

where $\mathsf{E} \ln \varepsilon_n^2 \approx -1.27$ and $\mathsf{D} \ln \varepsilon_n^2 = \pi^2/2 \approx 4.93$. Hence setting $x_n = \ln^2 h_n$ we arrive at the following *linear* system:

$$\Delta_n = a_0 + a_1 \Delta_{n-1} + c\delta_n, \tag{11}$$

$$x_n = \mathsf{E} \ln \varepsilon_n^2 + \Delta_n + \frac{\pi}{\sqrt{2}} \xi_n, \tag{12}$$

where

$$\xi_n = \frac{\sqrt{2}}{\pi} \left(\ln \varepsilon_n^2 - \mathsf{E} \ln \varepsilon_n^2\right) \tag{13}$$

with $\mathsf{E}\xi_n = 0$ and $\mathsf{D}\xi_n = 1$, and we can regard the above quantity -1.27 as an approximation to $\mathsf{E} \ln \varepsilon_n^2$.

Thus, we can assume that we have a *linear* system (11)–(12), in which the distributions of the ξ_n are *non*-Gaussian. This rules out a direct use of the Kalman–Bucy linear filtering.

Nevertheless, we shall consider the Kalman–Bucy filter *as if the* ξ_n, $n \geqslant 1$, *were normally distributed variables*, $\xi_n \sim \mathcal{N}(0,1)$, *independent of the sequence* (δ_n).

Let $\mu_n = \mathsf{E}(\Delta_n \mid x_1, \dots, x_n)$ and let $\gamma_n = \mathsf{D}\Delta_n$ under this assumption. Then the evolution of the μ_n and γ_n is described by the following system (see, e.g., [303; Chapter VI, § 7, Theorem 1]):

$$\mu_{n+1} = (a_0 + a_1 \mu_n) + \frac{a_1 \gamma_n}{\frac{\pi^2}{2} + \gamma_n} (x_{n+1} + 1.27 - \mu_n), \tag{14}$$

$$\gamma_{n+1} = (a_1^2 \gamma_n + c^2) - \frac{(a_1 \gamma_n)^2}{\frac{\pi^2}{2} + \gamma_n}. \tag{15}$$

Here $\mu_0 = \mathsf{E}\Delta_0$ and $\gamma_0 = \mathsf{D}\Delta_0$.

We note that if the parameter c in (11) is large, then Δ_n is the dominating term in the formula for x_n. Hence one can expect the μ_n to be 'good' approximations of the m_n.

4. We point out that if the parameters $\theta = (a_0, a_1, c)$ of the initial model (11) are unknown, then one often uses the Bayes method to find an estimator $\widehat{\sigma}_n$ for σ_n on the basis of $(h_1, \dots, h_n)$. Its cornerstone is the assumption that we are given a *prior* distribution $\operatorname{Law}(\theta)$ of θ. Then, in principle, we can find, first, the posterior distribution $\operatorname{Law}(\theta, \sigma_n \mid h_1, \dots, h_n)$ and then the posterior distributions $\operatorname{Law}(\theta \mid h_1, \dots, h_n)$ and $\operatorname{Law}(\sigma_n \mid h_1, \dots, h_n)$. This would enable us to construct the estimators $\widehat{\theta}_n$ and $\widehat{\sigma}_n$, e.g., as the posterior means or the values delivering maxima of the posterior densities. For a more comprehensive treatment of the Bayesian approach, see, for instance, [252] and the comments to this paper on pp. 395–417 of the same issue (v. 12, no. 4, 1994) of *Journal of Business and Economic Statistics*.

5. We have described the models in the *GARCH* family and 'stochastic volatility' models in the framework of the *conditional* approach. The conditional distribution $\operatorname{Law}(h_n \mid \sigma_n)$ has always been a normal distribution, $\mathcal{N}(0, \sigma_n^2)$, with σ_n^2 dependent on the 'past' in a 'predictable' way. Taking this approach it is natural to ask about the unconditional distributions $\operatorname{Law}(h_n)$ and $\operatorname{Law}(h_1, \dots, h_n)$.

To give a notion of the results possible here we consider now the following model (see [105]).

Let $h_n = \sigma_n \varepsilon_n$, where (ε_n) is again a standard Gaussian sequence,

$$\sigma_n^2 = a\sigma_{n-1}^2 + b\delta_n, \qquad -\infty < n < \infty, \tag{16}$$

and (δ_n) is the sequence of *nonnegative* independent stable random variables with exponent α, $0 < \alpha < 1$ (cf. Chapter III, § 1c.4). We assume that the sequences (ε_n) and (δ_n) are independent.

If $0 \leqslant a < 1$, then we obtain recursively by (16) that

$$\sigma_n^2 = b \sum_{k=0}^{\infty} a^k \delta_{n-k} + \lim_{m \to \infty} a^m \sigma_{n-m-1}^2. \tag{17}$$

By the self-similarity properties of stable distributions,

$$\sum_{k=0}^{\infty} a^k \delta_{n-k} \stackrel{\mathrm{d}}{=} \left(\sum_{k=0}^{\infty} a^{\alpha k} \right)^{1/\alpha} \delta_1.$$

Consequently, if $0 < \alpha < 1$ and $0 \leqslant a < 1$, then (16) has a (finite) nonnegative 'stationary' solution (σ_n^2), where

$$\sigma_n^2 \stackrel{\mathrm{d}}{=} b \left(\frac{1}{1 - a^\alpha} \right)^{1/\alpha} \delta_1. \tag{18}$$

Hence we see from the definition ($h_n = \sigma_n \varepsilon_n$) that the stationary one-dimensional distribution $\operatorname{Law}(h_n)$ is stable, with stability exponent 2α.

6. To complete this section, devoted to nonlinear stochastic models and their properties, we dwell on the above-mentioned phenomenon of 'heavy tails' observable in these models. (See also Chapter IV, § 2c.)

We consider the $ARCH(1)$-model of variables $h = (h_n)_{n\geqslant 0}$ with initial condition h_0 independent on the standard Gaussian sequence $\varepsilon = (\varepsilon_n)_{n\geqslant 1}$, $h_n = \sqrt{\alpha_0 + \alpha_1 h_{n-1}^2}\,\varepsilon_n$ for $n \geqslant 1$, $\alpha_0 > 0$, and $0 < \alpha_1 < 1$.

For an appropriate choice of the distribution of h_0 this model turns out to have a solution $h = (h_n)_{n\geqslant 0}$ that is a strictly stationary process with phenomenon of 'heavy tails' (observable for sufficiently small $\alpha_1 > 0$): $\mathsf{P}(h_n > x) \sim cx^{-\gamma}$, where $c > 0$ and $\gamma > 0$.

The corresponding (fairly tricky) proof can be found in the recent monograph of P. Embrechts, C. Klueppelberg, T. Mikosch "Modelling extremal events for insurance and finance", Berlin, Springer-Verlag, 1997 (Theorems 8.4.9 and 8.4.12). One can also find there a thorough analysis of many models of $ARCH$, $GARCH$ and related kinds and a large list of literature devoted to these models.

4. Supplement: Dynamical Chaos Models

§ 4a. Nonlinear Chaotic Models

1. So far, in our descriptions of the evolution of the sequences $h = (h_n)$ with $h_n = \ln \dfrac{S_n}{S_{n-1}}$, where S_n is the level of some 'price' at time n, we were based on the conjecture that these variables were *stochastic*, i.e., the $S_n = S_n(\omega)$ and $h_n = h_n(\omega)$ were *random* variables defined on some filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_n)_{n\geqslant 1}, \mathsf{P})$ and simulating the statistical uncertainty of 'real-life' situations.

On the other hand, it is well known that even very simple nonlinear *deterministic* systems of the type

$$x_{n+1} = f(x_n; \lambda) \tag{1}$$

or

$$x_{n+1} = f(x_n, x_{n-1}, \dots, x_{n-k}; \lambda), \tag{2}$$

where λ is a parameter, can produce (for appropriate initial conditions) sequences with behavior vary similar to that of stochastic sequences.

This justifies the following question: is it likely that many economic, including financial, series are actually realizations of *chaotic* (rather than stochastic) systems, i.e., systems described by deterministic nonlinear systems? It is known that such systems can bring about phenomena (e.g., the 'cluster property') observable in the statistical analysis of financial data (see Chapter IV).

Referring to a rather extensive special literature for the formal definitions (see, e.g., [59], [71], [104], [198], [378], [379], [383],[385], [386], [428], or [456]), we now present several examples of nonlinear chaotic systems in order to provide the reader with a notion of their behavior. We shall also consider the natural question as to how one can guess the kind of the system (stochastic or chaotic) that has generated a particular realization.

Discussing the forecasts of the future price movements, the predictability problem is also of considerable interest in nonlinear chaotic models. As we shall see below, the situation here does not inspire one with much optimism, because, all the determinism notwithstanding, the behavior of the trajectories of chaotic systems can considerably vary after a small change in the initial data and the value of λ.

2. EXAMPLE 1. We consider the so-called *logistic* map

$$x \rightsquigarrow Tx \equiv \lambda x(1-x)$$

and the corresponding (one-dimensional) dynamical system

$$x_n = \lambda x_{n-1}(1-x_{n-1}), \qquad n \geqslant 1, \quad 0 < x_0 < 1. \tag{3}$$

(Apparently, logistic equations (3) occurred first in the models of population dynamics that imposed constraints on the growth of a population.)

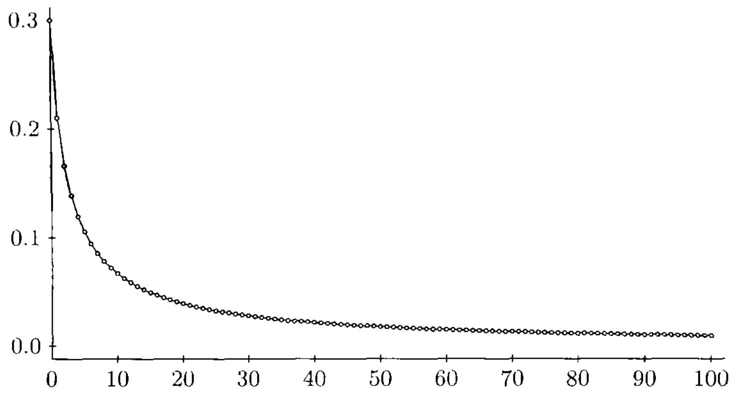


FIGURE 23a. Case $\lambda = 1$

For $\lambda \leqslant 1$ the solutions $x_n = x_n(\lambda)$ converge monotonically to 0 as $n \to \infty$ for all $0 < x_0 < 1$ (Fig. 23a). Thus, the state $x_\infty = 0$ is the *unique stable* state in this case, and it is the limit point of the x_n as $n \to \infty$.

For $\lambda = 2$ we have $x_n \uparrow \frac{1}{2}$ (Fig. 23b). Hence there also exists in this case a *unique stable* state ($x_\infty = \frac{1}{2}$) attracting the x_n as $n \to \infty$.

We now consider larger values of λ. For $\lambda < 3$ the system (3) still has a unique stable state. However, an entirely new phenomenon occurs for $\lambda = 3$: as n grows, one can distinguish *two* states x_∞ (Fig. 23c), and the system alternates between these states.

This pattern is retained as λ increases, until something new happens for $\lambda = 3.4494\ldots$: the system has now *four* distinguished states x_∞ and leaps from one to another (Fig. 23d).

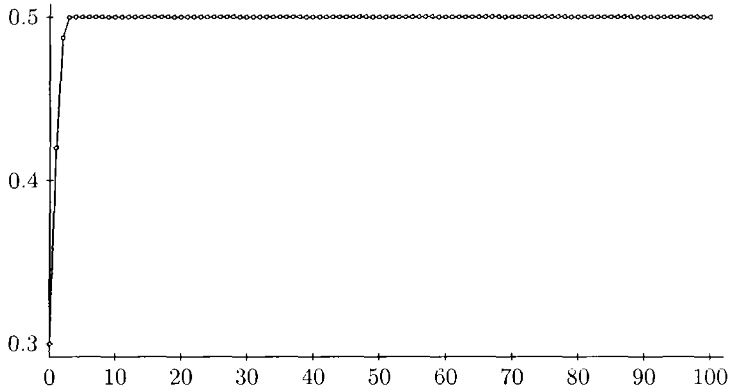


FIGURE 23b. Case $\lambda = 2$

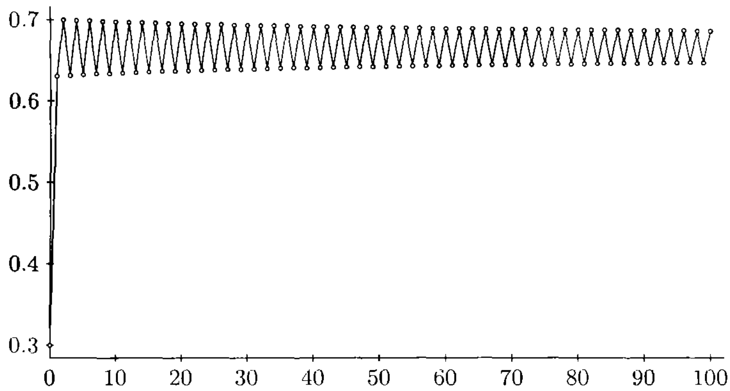


FIGURE 23c. Case $\lambda = 3$

New distinguished states come into being with further increases in λ: there are 8 such states for $\lambda = 3.5360\ldots$, 16 for $\lambda = 3.5644\ldots$, and so on. For $\lambda = 3.6$, there exists infinitely many such states, which is usually interpreted as a *loss of stability* and a transition into a *chaotic* state.

Now the periodic character of the movements between different states is completely lost; the system wanders over an infinite set of states jumping from one to another. It should be pointed out that, although our system is deterministic, it is impossible *in practice* to predict its position at some later time because the *limited precision in our knowledge of the values of the x_n and λ can considerably influence the results.*

It is clear from this brief description already that the values (λ_k) of λ at which the system 'branches', 'bifurcates' draw closer together in the process (Fig. 24).

As conjectured by M. Feigenbaum and proved by O. Lanford [294], for all par-

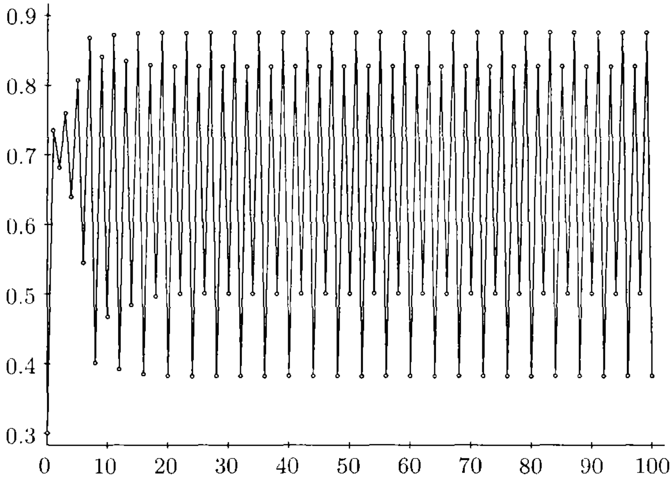


FIGURE 23d. Case $\lambda = 3.5$

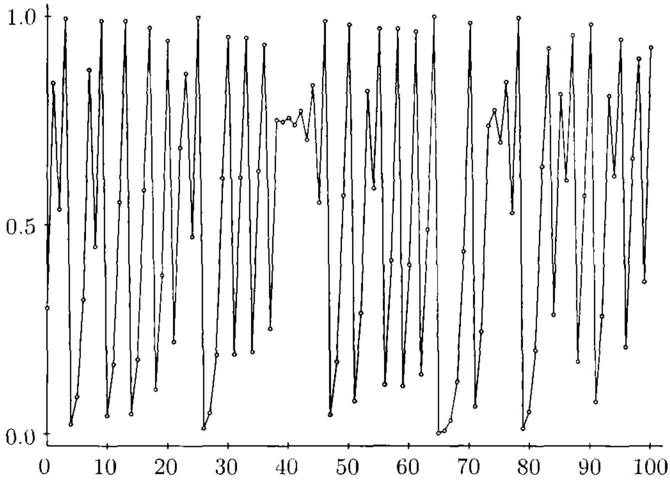


FIGURE 23e. Case $\lambda = 4$

abolic systems we have

$$\frac{\lambda_k - \lambda_{k-1}}{\lambda_{k+1} - \lambda_k} \to F, \qquad k \to \infty,$$

where $F = 4.669201\ldots$ is a universal constant (the *Feigenbaum constant*).

The value $\lambda = 4$ is of particular importance for (3): it is for this value of the

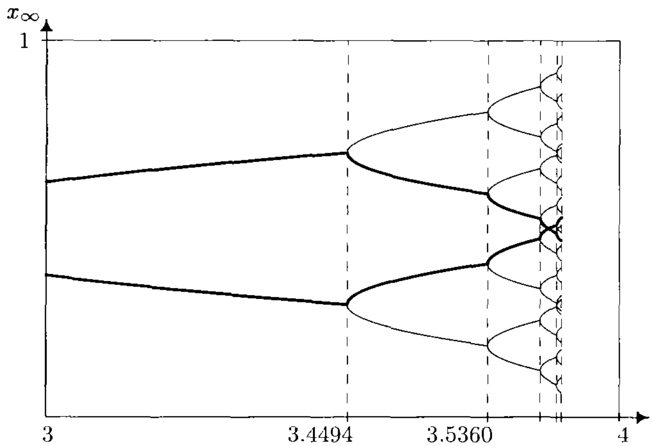


FIGURE 24. *Doubling* process for the states x_∞ in the logistic system as $\lambda \uparrow 4$

parameter that the sequence of observations (x_n) of our (chaotic) system is similar to a realization of a stochastic sequence of 'white noise' type.

Indeed, let $x_0 = 0.1$. We now calculate recursively the values of $x_1, x_2, \dots, x_{1000}$ using (3). The (empirical) mean value and the standard deviation evaluated on the basis of these 1000 numbers are 0.48887 and 0.35742, respectively (up to the fifth digit).

TABLE 2.

1	−0.033	11	−0.046	21	−0.008	31	0.038
2	−0.058	12	0.002	22	0.009	32	−0.017
3	−0.025	13	−0.011	23	−0.039	33	0.014
4	−0.035	14	0.040	24	−0.020	34	0.001
5	−0.012	15	0.014	25	−0.008	35	0.017
6	−0.032	16	−0.023	26	0.017	36	−0.052
7	−0.048	17	−0.030	27	0.006	37	0.004
8	0.027	18	0.037	28	−0.004	38	0.053
9	−0.020	19	0.078	29	−0.019	39	−0.021
10	−0.013	20	0.017	30	−0.076	40	0.007

In Table 2 we present the values of the (empirical) correlation function $\widehat{\rho}(k)$ calculated from $x_0, x_1, \dots, x_{1000}$. It is clearly visible from this table that the values x_n of the logistic map with $\lambda = 4$ can in practice be assumed to be *uncorrelated*.

In this sense, the sequence (x_n) can be called '*chaotic white noise*'.

It is worth noting that the system $x_n = 4x_{n-1}(1-x_{n-1})$, $n \geqslant 1$, with $x_0 \in (0,1)$ has an *invariant* distribution P (which means that P satisfies the equality $\mathsf{P}(T^{-1}A) = \mathsf{P}(A)$ for each Borel subset A of $(0,1)$) with density

$$p(x) = \frac{1}{\pi[x(1-x)]^{1/2}}, \qquad x \in (0,1). \tag{4}$$

Thus, assuming that the initial value x_0 is a *random* variable with density $p = p(x)$ of the probability distribution, we can see that the random variables x_n, $n \geqslant 1$, have the same distribution as x_0. We point out that all the 'randomness' of the resulting stochastic dynamic system (x_n) is related to the *random* initial value x_0, while the dynamics of the transitions $x_n \to x_{n+1}$ is *deterministic* and described by (3).

If (4) holds, then it is easy to see that $\mathsf{E}x_0 = \frac{1}{2}$, $\mathsf{E}x_0^2 = \frac{3}{8}$, and $\mathsf{D}x_0 = \frac{1}{8} = (0.35355\dots)^2$. (Cf. the values 0.48887 and 0.35742 presented above.) As regards the correlation function

$$\rho(k) = \frac{\mathsf{E}x_0x_k - \mathsf{E}x_0\mathsf{E}x_k}{\sqrt{\mathsf{D}x_0\,\mathsf{D}x_k}},$$

we have

$$\rho(k) = \begin{cases} 1 & \text{if } k = 0 \\ 0 & \text{if } k \neq 0 \end{cases}.$$

EXAMPLE 2 (*the Bernoulli transformation*). We set

$$x_n = 2x_{n-1} \pmod 1, \qquad x_0 \in (0,1).$$

Here the uniform distribution with density $p(x) = 1$, $x \in (0,1)$, is invariant, and we have $\mathsf{E}x_0 = \frac{1}{2}$, $\mathsf{E}x_0^2 = \frac{1}{3}$, $\mathsf{D}x_0 = \frac{1}{12}$, and $\rho(k) = 2^{-k}$, $k = 1, 2, \dots$.

EXAMPLE 3 (*the 'tent' map*). We set

$$x_n = 1 - |1 - 2x_{n-1}|, \qquad x_0 \in (0,1).$$

Here, as in Example 2, the uniform distribution on $(0,1)$ is invariant. Hence $\mathsf{E}x_0 = \frac{1}{2}$, $\mathsf{E}x_0^2 = \frac{1}{3}$, $\mathsf{D}x_0 = \frac{1}{12}$, and $\rho(k) = 0$ for $k \neq 0$.

EXAMPLE 4. Let

$$x_n = 1 - 2\sqrt{|x_{n-1}|}, \qquad x_0 \in (-1,1).$$

Then the distribution with density $p(x) = (1-x)/2$ on $(-1,1)$ is invariant. For this distribution we have $\mathsf{E}x_0 = -\frac{1}{3}$, $\mathsf{E}x_0^2 = \frac{1}{3}$, and $\mathsf{D}x_0 = \frac{2}{9}$.

The behavior of the sequences $(x_n)_{n \leqslant N}$ for $x_0 = 0.2$ and $N = 100$ or $N = 1000$ is depicted in Fig. 25a,b.

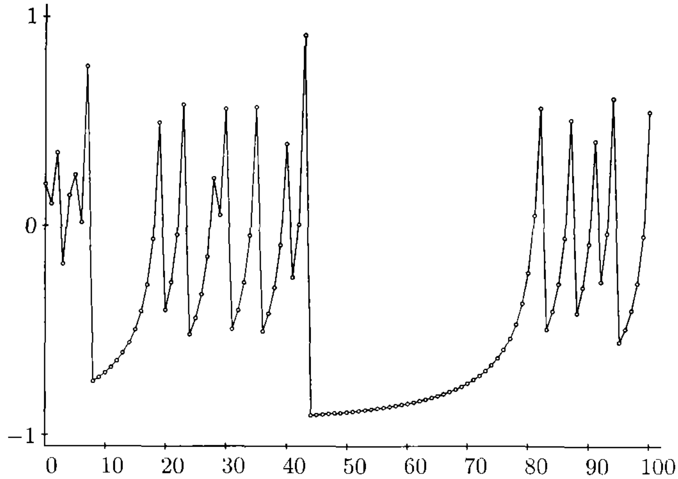


FIGURE 25a. Graph of the sequence $x = (x_n)_{n\geqslant 0}$ with $x_n = 1 - 2\sqrt{|x_{n-1}|}$ and $x_0 = 0.2$ for $N = 100$

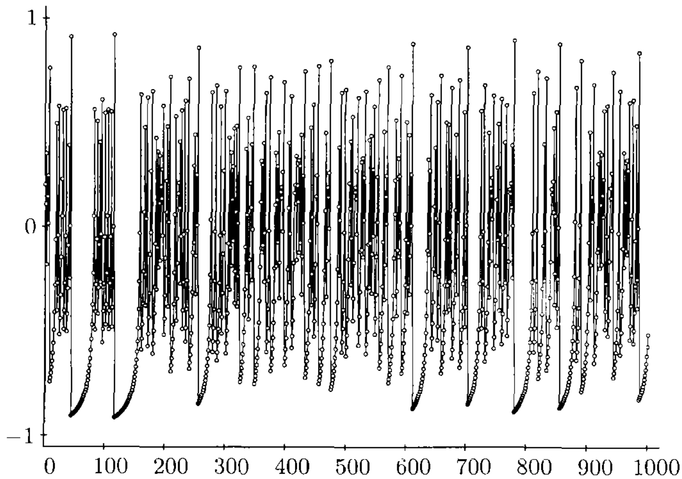


FIGURE 25b. Graph of the sequence $x = (x_n)_{n\geqslant 0}$ with $x_n = 1 - 2\sqrt{|x_{n-1}|}$ and $x_0 = 0.2$ for $N = 1000$

3. The above examples of nonlinear dynamical systems are of interest from various viewpoints. First, considering the example of, say, the logistic system, which develops in accordance with a 'binary' pattern, one can get a clear idea of fractality discussed in Chapter III, § 2. Second, the behavior of such 'chaotic' systems suggests one to use them in the construction of models simulating the evolution of financial indexes, in particular, in *times of crashes*, which are featured by 'chaotic' (rather than 'stochastic') behavior.

§ 4b. Distinguishing between 'Chaotic' and 'Stochastic' Sequences

1. The fact that purely deterministic dynamical systems can have properties of 'stochastic white noise' is not unexpected. It has been fairly long known, although this still turns out to be surprising for many. The more interesting, for that reason, are the questions on how one can draw a line between 'stochastic' and 'chaotic' sequences, whether it is possible in principle, and whether the true nature of 'irregularities' in the financial data is 'stochastic' or 'chaotic'. (Presumably, an appropriate approach here could be based on the concept of 'complexity in the sense of A. N. Kolmogorov, P. Martin-Löf, and V. A. Uspenskiĭ' adapted for particular realizations.)

Below we discuss the approach taken in [305] and [448]. In it, the central role in distinguishing between 'chaotic' and 'stochastic' is assigned to the function

$$C(\varepsilon) = \lim_{N\to\infty} \frac{\psi(N,\varepsilon)}{N^2}, \tag{1}$$

where, given a sequence (x_n), the function $\psi(N,\varepsilon)$ is the number of pairs (i,j), $i,j \leqslant N$, such that

$$|x_i - x_j| < \varepsilon.$$

Besides $C(\varepsilon)$, we shall also consider the functions

$$C_m(\varepsilon) = \lim_{N\to\infty} \frac{\psi_m(N,\varepsilon)}{N^2},$$

where $\psi_m(N,\varepsilon)$ is the number of (i,j), $i,j \leqslant N$, such that all the differences between the corresponding components of the vectors $(x_i, x_{i+1}, \dots, x_{i+m-1})$ and $(x_j, x_{j+1}, \dots, x_{j+m-1})$ are at most ε. (For $m = 1$ we have $\psi_1(N,\varepsilon) = \psi(N,\varepsilon)$.)

For stochastic sequences (x_n) of 'white noise' type we have

$$C_m(\varepsilon) \sim \varepsilon^{v_m} \tag{2}$$

for small ε, where the *fractal exponent* v_m is equal to m. Many deterministic systems also have property (2) (e.g., the logistic system (3) in the preceding section [305]). The exponent v_m is also called the *correlation dimension* and is closely connected with the Hausdorff dimension and Kolmogorov's information dimension.

Distinguishing between 'chaotic' and 'stochastic' sequences in [305] and [448] is based on the following observation: these sequences have different correlation dimensions. As seen from what follows, this dimension is larger for 'stochastic' sequences than for 'chaotic' ones.

By [448] and [305], the quantities

$$\widetilde{v}_{m,j} = \frac{\ln C_m(\varepsilon_j) - \ln C_m(\varepsilon_{j+1})}{\ln \varepsilon_j - \ln \varepsilon_{j+1}}$$

and

$$\widetilde{v}_m(k) = \frac{1}{k} \sum_{j=1}^{k} \widetilde{v}_{m,j},$$

where $\varepsilon_j = \varphi^j$ and $0 < \varphi < 1$, can be taken as estimators of the correlation dimension v_m.

In Table 3a one can find the values of the $\widetilde{v}_{m,j}$ corresponding to $\varphi = 0.9$, $m = 1, 2, 3, 4, 5, 10$, and several values of j in the case of the *logistic sequence* $(x_n)_{n \leqslant N}$, where $N = 5\,900$ and $\varepsilon_j = \varphi^j$ (these data are borrowed from [305])

TABLE 3a. Values of $\widetilde{v}_{m,j}$ for the logistic system

$j \backslash m$	1	2	3	4	5	10
20	0.78	0.90	0.96	0.98	1.02	1.19
30	0.81	0.89	0.95	0.98	0.98	1.11
35	0.83	0.90	0.94	0.97	0.95	1.01
40	0.83	0.91	0.97	0.99	1.05	1.20

We now compare the data in this table with the estimates for $\widetilde{v}_{m,j}$ obtained by a simulation of *Gaussian white noise* with parameters characteristic of the logistic map (3) in § 6a (Table 3b; the data from [305]):

TABLE 3b. Values of the $\widetilde{v}_{m,j}$ for Gaussian white noise

$j \backslash m$	1	2	3	4	5	10
20	0.84	1.68	2.52	3.35	4.20	8.43
30	0.98	1.97	2.95	3.98	4.98	–
35	0.99	1.97	2.93	4.00	5.53	–
40	1.00	2.02	3.03	4.15	5.38	–

Comparing these tables we see that it is fairly difficult to distinguish between 'chaotic' and 'stochastic' cases on the basis of the correlation dimension $\widetilde{v}_{1,j}$ (corresponding to $m = 1$). However, if m is larger, then a considerable difference between the values of the $\widetilde{v}_{m,j}$ for these two cases is apparent. This is a rather solid argument in favor of the conjecture on distinct natures of the corresponding sequences (x_n), although there is virtually no difference between their empirical mean values, variances, or correlations.

2. To illustrate the problems of distinguishing between 'stochastic' and 'chaotic' cases in the case of *financial* series we present the tables of the correlation dimensions calculated for the daily values of the 'returns' $h_n = \ln \dfrac{S_n}{S_{n-1}}$, $n \geqslant 1$, corresponding to IBM stock price and S&P500 Index (Tables 4a, and 4b are compiled from 5903 observations carried out between June 2, 1962 and December 31, 1985; the data are borrowed from [305]):

TABLE 4a. Values of $\widetilde{v}_{m,j}$ for IBM stock

$j \backslash m$	1	2	3	4	5	10
20	0.46	0.90	1.31	1.68	2.05	3.63
30	0.83	1.76	2.61	3.44	4.27	8.44
35	0.97	1.93	2.88	3.82	4.79	9.84
40	0.98	1.96	2.94	3.86	4.94	–

TABLE 4b. Values of $\widetilde{v}_{m,j}$ for the S&P500 Index

$j \backslash m$	1	2	3	4	5	10
20	0.58	1.10	1.58	2.03	2.43	3.93
30	0.93	1.82	2.07	3.49	4.25	6.93
35	0.98	1.94	2.88	3.79	4.75	11.00
40	0.99	1.98	2.92	3.84	4.81	–

Comparing these tables we see, first, that the estimates of the 'correlation dimension' of the IBM and the S&P500 returns in this two tables are very close. Second, comparing the data in Tables 4a, b and Tables 3a, b we see that the sequences (h_n) corresponding to these two indexes (where $h_n = \ln \dfrac{S_n}{S_{n-1}}$, $n \geqslant 1$) are closer to 'stochastic white noise'. Of course, this cannot disprove the conjecture that other 'chaotic sequences', of larger correlation dimensions, can also have similar properties. (For greater detail on the issue of distinguishing and for an economist's commentary, see [305].)

3. We now consider briefly another approach to discovering distinctions between 'chaotic' and 'stochastic' cases, which was suggested in [17].

Let $x = (x_n)$ be a 'chaotic' sequence that is a realization of some dynamical system in which x_0 is a random variable with probability distribution $F = F(x)$ invariant with respect to this system.

Next, let $\widetilde{x} = (\widetilde{x}_n)$ be a 'stochastic' sequence of independent identically distributed variables with (one-dimensional) distribution $F = F(x)$.

We consider now the variables

$$M_n = \max(x_0, x_1, \dots, x_n) \quad \text{and} \quad \widetilde{M}_n = \max(\widetilde{x}_0, \widetilde{x}_1, \dots, \widetilde{x}_n).$$

Let $F_n(x) = \mathsf{P}(M_n \leqslant x)$ and let $\widetilde{F}_n(x) = \mathsf{P}(\widetilde{M}_n \leqslant x)$.

The approach in [17] is based on the observation that the *maximum* is a characteristic well suited for capturing the difference between 'stochastic' and 'chaotic' sequences.

To substantiate this approach, the authors of [17] proceed as follows.

In the theory of limit theorems for extremum values there are well-known necessary and sufficient conditions ensuring that the variables $a_n(\widetilde{M}_n - b_n)$, where $a_n > 0$ and b_n are some constants and $n \geqslant 1$, have a (nontrivial) limit distribution

$$\lim_n \mathsf{P}\big(a_n(\widetilde{M}_n - b_n) \leqslant x\big) = \widetilde{G}(x)$$

Referring to [187], [206], [156], [124], and [137] for details, we now present several examples.

If $F(x) = 1 - x^{-\rho}$, $x \geqslant 1$, and $\rho > 0$, then

$$\mathsf{P}\left(\frac{\widetilde{M}_n}{n^{1/\rho}} \leqslant x\right) \to \exp(-x^{-\rho}), \qquad x > 0.$$

If $F(x) = 1 - (-x)^{\rho}$, $-1 \leqslant x \leqslant 0$, and $\rho > 0$, then (for $x < 0$)

$$\mathsf{P}\big(n^{1/\rho}\widetilde{M}_n \leqslant x\big) \to \exp\big(-|x|^{\rho}\big).$$

If $F(x) = 1 - e^{-x}$, $x \geqslant 0$, then

$$\mathsf{P}\big(\widetilde{M}_n - \log n \leqslant x\big) \to \exp\big(-e^{-x}\big)$$

for $x \in \mathbb{R}$.

If $F(x) = \Phi(x)$ is a standard normal distribution, then

$$\mathsf{P}\big((2\log n)^{1/2}(\widetilde{M}_n - b_n) \leqslant x\big) \to \big(-e^{-x}\big), \qquad x \in \mathbb{R},$$

where we choose the b_n such that $\mathsf{P}(x_0 > b_n) = \frac{1}{n}$. (In this case $b_n \sim (2\log n)^{1/2}$.)

For the Bernoulli transformation (Example 2 in § 4a) we have the invariant distribution $F(x) = x$, $x \in (0,1)$. Setting $a_n = n$ and $b_n = 1 - n^{-1}$ we obtain the limit distribution

$$\widetilde{G}(x) = \exp(x-1), \qquad x \leqslant 1.$$

In Example 4 from § 4a we have $F(x) = 1 - p^2(x)$, where $p(x) = (1-x)/2$, $x \in (-1,1)$, therefore setting $a_n = \sqrt{n}$ and $b_n = 1 - 2/\sqrt{n}$ we see that

$$\widetilde{G}(x) = \exp\biggl(-\Bigl(\frac{x}{2} - 1\Bigr)^2\biggr).$$

For Example 1 in § 4a the invariant distribution is $F(x) = \dfrac{2}{\pi} \arcsin\sqrt{x}$ (Example 1 in § 4a), and after the corresponding renormalization we obtain

$$\widetilde{G}(x) = \exp(-(1-2x)^{1/2}).$$

Given the distributions $\widetilde{F}_n(x) = (F(x))^n$ and the limit distribution $\widetilde{G}(x)$, it would be reasonable to compare them with the corresponding distributions $F_n(x)$ and, if possible, with their limits, say, $G(x)$. However, as pointed out in [17], there exists a serious technical problem, because there are no analytic expressions for the $F_n(x)$ in the examples in § 4a that are convenient for further analysis. For that reason, the approach taken in [17] consists in the numerical analysis of the distributions $F_n(x)$ for large n and their comparison with the corresponding distributions $\widetilde{F}_n(x)$.

For the dynamical systems in § 4a this analysis shows that, *globally*, the behavior of the $F_n(x)$ (for chaotic systems with invariant distribution $F(x)$) *has a character distinct* from the behavior of the $\widetilde{F}_n(x)$ (for stochastic systems of independent identically distributed variables with one-dimensional distribution $F(x)$). This indicates that, for the models under consideration, the *maximum* value is a 'good' statistics for our problem of distinguishing between chaotic and stochastic cases. But of course, this does not rule out the possibility that there exists a chaotic system $x_{n+1} = f(x_n, x_{n-1}, \dots, x_{n-k}; \lambda)$ with k sufficiently large that is difficult to distinguish from stochastic white noise on the basis of a large (but finite) number of observations.

Chapter III. Stochastic Models. Continuous Time

1. Non-Gaussian Models of Distributions and Processes

§ 1a. Stable and Infinitely Divisible Distributions

1. In the next chapter we discuss the results of the statistical analysis of prospective models of distribution and evolution for such financial indexes as currency exchange rates, share prices, and so on. This analysis will demonstrate the importance of *stable distributions* and *stable processes* as natural and fairly likely candidates in the construction of probabilistic models of this kind.

For that reason, we provide the reader in this section with necessary information on both these distributions and processes and on more general, *infininely divisible* ones: without the latter our description and the discussion of the properties of financial indexes will lack of completeness.

In our exposition of basic concepts and properties of *stability* and *inifinite divisibility* we shall (to a certain extent) follow the chronological order. First, we shall discuss one-dimensional stable distributions considered in the 1920s by P. Lévy, G. Pólya, A. Ya. Khintchine, and then proceed to one and several-dimensional infinitely divisible distributions studied by B. de Finetti, A. N. Kolmogorov, P. Lévy, and A. Ya. Khintchine in the 1930s. After that, in § 1b we introduce the reader to the main concepts and the properties of Lévy processes and stable processes.

The monographs [156], [188], [418], and [484] belong among widely used textbooks on stable and infinitely divisible distributions and processes.

2. DEFINITION 1. We say that a nondegenerate random variable X is *stable* or has a *stable distribution* if for any positive numbers a and b there exists a positive c and a number d such that

$$\mathrm{Law}(aX_1 + bX_2) = \mathrm{Law}(cX + d) \tag{1}$$

for independent random copies X_1 and X_2 of X (this means that $\mathrm{Law}(X_i) = \mathrm{Law}(X)$, $i = 1, 2$; we shall assume without loss of generality that all the random variables under consideration are defined on the same probability space $(\Omega, \mathscr{F}, \mathsf{P})$).

It can be proved (see the above-mentioned monographs) that in (1) we necessarily have

$$c^\alpha = a^\alpha + b^\alpha \tag{2}$$

for some $\alpha \in (0, 2]$ independent of a and b.

One often uses another, equivalent, definition.

DEFINITION 2. A random variable X is said to be *stable* if for each $n \geqslant 2$ there exist a positive number C_n and a number D_n such that

$$\operatorname{Law}(X_1 + X_2 + \cdots + X_n) = \operatorname{Law}(C_n X + D_n), \tag{3}$$

where $X_1, X_2, \dots, X_n$ are independent copies of X.

If $D_n = 0$ in (3) for $n \geqslant 2$, i.e.,

$$\operatorname{Law}(X_1 + X_2 + \cdots + X_n) = \operatorname{Law}(C_n X), \tag{4}$$

then X is said to be a strictly *stable* variable.

Remarkably,

$$C_n = n^{1/\alpha}$$

in (3) and (4) for some α, $0 < \alpha \leqslant 2$, which is, of course, the same parameter as in (2).

To emphasize the role and importance of α, one often uses the term '*α-stability*' in place of '*stability*'.

For completeness, it is reasonable to add a third definition to the above two. It brings forward the role of stable distributions as the only ones that can be the limit distributions for (suitably normalized and centered) sums of *independent identically distributed* random variables.

DEFINITION 3. We say that a random variable X has a *stable distribution* (or simply, is *stable*) if the distribution of X has a domain of *attraction* in the following sense: there exist sequences of independent identically distributed random variables (Y_n), positive numbers (d_n), and real numbers (a_n) such that

$$\frac{Y_1 + \cdots + Y_n}{d_n} + a_n \xrightarrow{\mathrm{d}} X \quad \text{as} \quad n \to \infty; \tag{5}$$

here "$\xrightarrow{\mathrm{d}}$" means convergence *in distribution*, i.e.,

$$\operatorname{Law}\left(\frac{Y_1 + \cdots + Y_n}{d_n} + a_n\right) \to \operatorname{Law}(X)$$

in the sense of *weak* convergence of the corresponding measures.

This definition is equivalent to the above two because *a random variable X can be the limit in distribution of the variables $\dfrac{Y_1 + \cdots + Y_n}{d_n} + a_n$ for some sequence (Y_n) of independent identically distributed random variables if and only if X is stable* (in the sense of Definition 1 or Definition 2; see the proof in [188] or [439; Chapter III, § 5]).

3. According to a remarkable result of probability theory (P. Lévy, A. Ya. Khintchine) the characteristic function

$$\varphi(\theta) = \mathsf{E}e^{i\theta X}$$

of a stable random variable X has the following representation:

$$\varphi(\theta) = \begin{cases} \exp\Big\{i\mu\theta - \sigma^\alpha|\theta|^\alpha\Big(1 - i\beta(\operatorname{Sgn}\theta)\tan\dfrac{\pi\alpha}{2}\Big)\Big\} & \text{if } \alpha \neq 1, \\ \exp\Big\{i\mu\theta - \sigma|\theta|\Big(1 + i\beta\dfrac{2}{\pi}(\operatorname{Sgn}\theta)\ln|\theta|\Big)\Big\} & \text{if } \alpha = 1, \end{cases} \tag{6}$$

where $0 < \alpha \leqslant 2$, $|\beta| \leqslant 1$, $\sigma > 0$, and $\mu \in \mathbb{R}$.

Here the four parameters $(\alpha, \beta, \sigma, \mu)$ have the following meaning:

- α (which is, of course, the same as in (2) and (4)) is the *stability exponent* or the characteristic parameter;
- β is the *skewness* parameter of the distribution density;
- σ is the *scale* parameter;
- μ is the *location* parameter.

The parameter α 'controls' the decrease of the 'tails' of distributions. If $0 < \alpha < 2$, then

$$\lim_{x\to\infty} x^\alpha \mathsf{P}(X > x) = C_\alpha \frac{1+\beta}{2}\sigma^\alpha, \tag{7}$$

$$\lim_{x\to\infty} x^\alpha \mathsf{P}(X < -x) = C_\alpha \frac{1-\beta}{2}\sigma^\alpha, \tag{8}$$

where

$$C_\alpha = \left(\int_0^\infty x^{-\alpha}\sin x\,dx\right)^{-1} = \begin{cases} \dfrac{1-\alpha}{\Gamma(2-\alpha)\cos\frac{\pi\alpha}{2}}, & \alpha \neq 1, \\ \dfrac{2}{\pi}, & \alpha = 1. \end{cases} \tag{9}$$

If $\alpha = 2$, than by (6) we obtain

$$\varphi(\theta) = e^{i\mu\theta - \sigma^2\theta^2} = e^{i\mu\theta - \frac{\theta^2}{2}(2\sigma^2)}, \tag{10}$$

which shows that $\varphi(\theta)$ is the characteristic function of the normal distribution $\mathcal{N}(\mu, 2\sigma^2)$, i.e., of a normally distributed random variable X such that

$$\mathsf{E}X = \mu \quad \text{and} \quad \mathsf{D}X = 2\sigma^2.$$

The value of β in (6) is not well defined in this case (because if $\alpha = 2$, then this parameter enters the term $\beta\tan\pi$, which is equal to zero). One usually sets $\beta = 0$.

Clearly, as regards the behavior of the tails of distributions, the case $\alpha =$ significantly differs from $\alpha < 2$. For instance, if $\mu = 0$ and $2\sigma^2 = 1$, then

$$\mathsf{P}(|X| > x) \sim \sqrt{\frac{2}{\pi}}\,\frac{e^{-x^2/2}}{x} \quad \text{as} \quad x \to \infty. \tag{11}$$

Comparing this with (7) and (8) we see that for $\alpha < 2$ the tails are 'heavier': th decrease in the normal case is faster. (This is a right place to point out that, a seen in statistical analysis, for many series of financial indexes $S = (S_n)_{n\geqslant 0}$ th distributions of the 'returns' $h_n = \ln \dfrac{S_n}{S_{n-1}}$ have 'heavy tails'. It is natural, for thi reason, to try the class of stable distributions as a source of probabilistic model governing the sequences $h = (h_n)$.)

It important to note that, as seen from (7) and (8), the expectation $\mathsf{E}|X|$ i finite if and only if $\alpha > 1$. In general, $\mathsf{E}|X|^p < \infty$ if and only if $p < \alpha$.

In connection with asymptotic formulas (7) and (8) the *Pareto distribution* i worth recalling. Its distribution density is

$$f_{\alpha,b}(x) = \begin{cases} \dfrac{\alpha b^{\alpha}}{x^{\alpha+1}}, & x \geqslant b, \\ 0, & x < b, \end{cases} \tag{12}$$

with $\alpha > 0$ and $b > 0$, therefore its distribution function $F_{\alpha,b}(x)$ satisfies the relatio

$$1 - F_{\alpha,b}(x) = \left(\frac{b}{x}\right)^{\alpha}, \qquad x \geqslant b. \tag{13}$$

Comparing with (7) and (8) we see that the behavior at infinity of stable distri butions is similar to that of the Pareto distribution. One can say that the tails c stable distributions fall within the *Pareto type*.

The ***skewness*** parameter $\beta \in [-1, 1]$ in (6) characterizes the asymmetry of distribution. If $\beta = 0$, then the distribution is symmetric. If $\beta > 0$, then th distribution is skewed on the left, and the closer β approaches one, the greater thi skewness. The case of $\beta < 0$ corresponds to a right skewness.

The parameter σ plays the role of a *scale* coefficient. For a normal distributio ($\alpha = 2$) we have $\mathsf{D}X = 2\sigma^2$ (note that the variance here is $2\sigma^2$, not σ^2 as in th standard notation). On the other hand, if $\alpha < 2$, then $\mathsf{D}X$ is not defined.

We call μ the *location* parameter, because for $\alpha > 1$ we have $\mu = \mathsf{E}X$ (so tha $\mathsf{E}|X| < \infty$). There is no such interpretation in the general case for the mere reaso that $\mathsf{E}X$ does not necessarily exist.

4. Following an established tradition, we shall denote the stable distribution wit parameters α, β, σ, and μ by

$$S_{\alpha}(\sigma, \beta, \mu);$$

we shall write $X \sim S_\alpha(\sigma, \beta, \mu)$ to indicate that X has a stable distribution with parameters α, β, σ, and μ.

Note that $S_\alpha(\sigma, \beta, \mu)$ is a *symmetric* distribution if and only if $\beta = \mu = 0$. (It is obvious from the formula for the characteristic function that the constant D_n in (3) is equal to zero in that case.) This distribution is *symmetric relative* to μ (which can be *arbitrary*) if and only if $\beta = 0$.

In the symmetric case ($\beta = \mu = 0$) one often uses the notation

$$X \sim S\alpha S.$$

In this case the characteristic function is

$$\varphi(\theta) = e^{-\sigma^\alpha |\theta|^\alpha}. \tag{14}$$

5. Unfortunately, explicit formulas for the densities of stable distributions are known only for some values of the parameters. Among these distributions are

the normal distribution $S_2(\sigma, 0, \mu) = \mathscr{N}(\mu, 2\sigma^2)$ with density

$$\frac{1}{2\sigma\sqrt{\pi}} e^{-\frac{(x-\mu)^2}{4\sigma^2}}; \tag{15}$$

the Cauchy distribution $S_1(\sigma, 0, \mu)$ with density

$$\frac{\sigma}{\pi((x-\mu)^2 + \sigma^2)}; \tag{16}$$

the one-sided stable distribution (also called the Lévy or Smirnov distribution) $S_{1/2}(\sigma, 1, \mu)$ with exponent $\alpha = 1/2$ on (μ, ∞) with density

$$\left(\frac{\sigma}{2\pi}\right)^{1/2} \frac{1}{(x-\mu)^{3/2}} \exp\left(-\frac{\sigma}{2(x-\mu)}\right). \tag{17}$$

We point out two interesting and useful particular cases of (16) and (17):

if $X \sim S_1(\sigma, 0, 0)$, *then*

$$\mathsf{P}(X \leqslant x) = \frac{1}{2} + \frac{1}{\pi} \arctan \frac{x}{\sigma} \tag{18}$$

for $x > 0$;

if $X \sim S_{1/2}(\sigma, 1, 0)$, *then*

$$\mathsf{P}(X \leqslant x) = 2\left(1 - \Phi\left(\sqrt{\frac{\sigma}{x}}\right)\right) \tag{19}$$

for $x > 0$.

As regards the representation of the densities of stable distributions by *series*, see [156], [225], [418], and [484].

6. We now assume that $X_1, X_2, \ldots, X_n$ are independent random variables such that

$$X_i \sim S_\alpha(\sigma_i, \beta_i, \mu_i), \qquad i = 1, 2, \ldots, n.$$

Although, in general, they are *not identically distributed*, the fact that they all have the same stability exponent α indicates (see formula (6) for the characteristic function) that their sum $X = X_1 + \cdots + X_n$ has a distribution of the same type, $S_\alpha(\sigma, \beta, \mu)$, with parameters

$$\begin{aligned} \sigma &= \left(\sigma_1^\alpha + \cdots + \sigma_n^\alpha\right)^{1/\alpha}, \\ \beta &= \frac{\beta_1 \sigma_1^\alpha + \cdots + \beta_n \sigma_n^\alpha}{\sigma_1^\alpha + \cdots + \sigma_n^\alpha}, \\ \mu &= \mu_1 + \cdots + \mu_n. \end{aligned}$$

7. We now proceed to the more general class of so-called 'infinitely divisible' distributions, which includes stable distributions.

DEFINITION 4. We say that a random variable X and its probability distribution are *infinitely divisible* if *for each* $n \geqslant 1$ there exist *independent identically distributed* random variables $X_{n1}, \ldots, X_{nn}$ such that $X \stackrel{\mathrm{d}}{=} X_{n1} + \cdots + X_{nn}$.

The practical importance of this class is based on the following property: *these and only these* distributions can be the limits of the distributions of the sums $\left(\sum_{k=1}^{n} X_{nk}\right)$ of variables in the scheme of series

$$\begin{array}{l} X_{11} \\ X_{21}, X_{22} \\ \cdots \\ X_{n1}, X_{n2}, \ldots, X_{nn} \\ \cdots \end{array} \tag{20}$$

every row of which consists of independent identically distributed random variables $X_{n1}, X_{n2}, \ldots, X_{nn}$. Note that no connection between the variables in *distinct* rows of (20) is assumed. (For greater detail, see [188] or [439; Chapter III, § 5].)

A more restricted class of *stable* distributions corresponds to the case when the variables X_{nk} in (20) can be constructed by means of a *special* procedure from the variables in a fixed sequence of independent identically distributed variables $Y_1, Y_2, \ldots$ (see the end of subsection 2):

$$X_{nk} = \frac{Y_k}{d_n} + \frac{a_n}{n}, \qquad 1 \leqslant k \leqslant n, \quad n \geqslant 1. \tag{21}$$

(We point out that the class of infinitely divisible distributions includes also the hyperbolic and the Gaussian\\inverse Gaussian distributions discussed below, in § 1d.)

Definition 4 relates to the scalar case ($X \in \mathbb{R}$). It can be immediately extended to the vector-valued case ($X \in \mathbb{R}^d$) with no significant modifications.

Let $\mathsf{P} = \mathsf{P}(dx)$ be the probability distribution of an *infinitely divisible random vector* $X \in \mathbb{R}^d$ and let

$$\varphi(\theta) = \mathsf{E}e^{i(\theta,X)} = \int_{\mathbb{R}^d} e^{i(\theta,x)}\,\mathsf{P}(dx)$$

be its characteristic function; let (θ, x) be the scalar product of two vectors $\theta = (\theta_1, \dots, \theta_d)$ and $x = (x_1, \dots, x_d)$.

The work of B. de Finetti, followed by A. N. Kolmogorov (for $\mathsf{E}|X|^2 < \infty$) and finally P. Lévy and A. Ya. Khintchine in the 1930s resulted in the following *Lévy–Khintchine formula* for the characteristic function of $X \in \mathbb{R}^d$:

$$\varphi(\theta) = \exp\left\{i(\theta, B) - \frac{1}{2}(\theta, C\theta) + \int_{\mathbb{R}^d} \left(e^{i(\theta,x)} - 1 - i(\theta, x)I(|x| \leqslant 1)\right)\nu(dx)\right\}, \quad (22)$$

where $B \in \mathbb{R}^d$, $C = C(d \times d)$ is a symmetric nonnegative definite matrix, and $\nu = \nu(dx)$ is a *Lévy measure*: a positive measure in $\mathbb{R}^d$ such that $\nu(\{0\}) = 0$ and

$$\int_{\mathbb{R}^d} (|x|^2 \wedge 1)\,\nu(dx) < \infty. \quad (23)$$

(Note that both cases $\nu(\mathbb{R}^d) < \infty$ and $\nu(\mathbb{R}^d) = \infty$ are possible here.)

It should be pointed out that $\varphi(\theta)$ is specified by the three characteristics B, C, and ν, and the triplet (B, C, ν) in (22) is *unambigously* defined.

EXAMPLES.

1. If X is a degenerate random variable ($\mathsf{P}(X = a) = 1$), then $B = a$. $C = 0$, $\nu = 0$, and

$$\varphi(\theta) = e^{i\theta a}.$$

2. If X is a random variable having the Poisson distribution with parameter λ, then $\nu(dx) = \lambda I_{\{1\}}(dx)$ is a measure concentrated at the point $x = 1$, $B = \lambda$, and

$$\varphi(\theta) = e^{\lambda(e^{i\theta} - 1)}.$$

3. If $X \sim \mathcal{N}(m, \sigma^2)$, then $B = m$, $C = \sigma^2$, $\nu = 0$, and

$$\varphi(\theta) = e^{im\theta - \frac{\sigma^2}{2}\theta^2}.$$

4. If X is a random variable having the Cauchy distribution with density (16), then

$$\varphi(\theta) = e^{i\mu\theta - \sigma|\theta|}.$$

5. If X is a random variable with density (17) (a one-sided stable distribution with exponent $\alpha = 1/2$), then

$$\varphi(\theta) = e^{i\mu\theta - \sigma|\theta|^{1/2}(1 - i\,\mathrm{Sgn}\,\theta)}.$$

8. A representation of the characteristic function $\varphi(\theta)$ by formula (22) (using the traditional, 'canonical' truncation function $h(x) = x\,I(|x| \leqslant 1)$) is not unique. For instance, in place of $I(|x| \leqslant 1)$, we could use a representation with $I(|x| \leqslant a)$, where $a > 0$. Of course, the corresponding triplet of characteristics would also be different. Notably, C and ν do not change in that case; these are 'intrinsic' characteristics independent of our choice of the truncation function. Only the first characteristic, B, actually changes.

For precise statements, we shall need the following definition.

DEFINITION 5. We call a bounded function $h = h(x)$, $x \in \mathbb{R}^d$, with compact support satisfying the equality $h(x) = x$ in a neighborhood of the origin a *truncation function*.

Besides (22), the characteristic function $\varphi(\theta)$ has the following representations (valid for an arbitrary truncation function $h = h(x)$):

$$\varphi(\theta) = \exp\left\{ i(\theta, B(h)) - \frac{1}{2}(\theta, C\theta) + \int_{\mathbb{R}^d} \left(e^{i(\theta,x)} - 1 - i(\theta, h(x))\right) \nu(dx) \right\}, \tag{24}$$

where C and ν are independent of h and are the same as in (22), while the value of $B(h)$ varies with h in accordance with the following equality:

$$B(h) - B(h') = \int_{\mathbb{R}^d} \left(h(x) - h'(x)\right) \nu(dx). \tag{25}$$

We point out that the integrals on the right-hand sides of (22) and (24) are well defined in view of (23), because the function

$$e^{i(\theta,x)} - 1 - i\big(\theta, h(x)\big)$$

is bounded and it is $O(|x|^2)$ as $|x| \to 0$.

If we replace (23) by the stronger inequality

$$\int_{\mathbb{R}^d} (|x| \wedge 1)\, \nu(dx) < \infty, \tag{26}$$

then we can set $h(x) = 0$ in (24), so that

$$\varphi(\theta) = \exp\left\{ i(\theta, B(0)) - \frac{1}{2}(\theta, C\theta) + \int_{\mathbb{R}^d} \left(e^{i(\theta,x)} - 1\right) \nu(dx) \right\}. \tag{27}$$

The constant $B(0)$ in this representation is called the *drift component* of the random variable X.

On the other hand, if we replace (23) by the stronger inequality

$$\int_{\mathbb{R}^d} \left(|x|^2 \wedge |x|\right) \nu(dx) < \infty, \tag{28}$$

then (24) holds for $h(x) = x$, i.e.,

$$\varphi(\theta) = \exp\left\{ i(\theta, \widetilde{B}) - \frac{1}{2}(\theta, C\theta) + \int_{\mathbb{R}^d} \left(e^{i(\theta,x)} - 1 - i(\theta, x)\right) \nu(dx) \right\}. \tag{29}$$

In this case the parameter $\widetilde{B}$ (the *center*) is in fact the expectation $\widetilde{B} = \mathsf{E}X$.

We note that the condition $\mathsf{E}|X| < \infty$ is equivalent to the inequality

$$\int_{|x|>1} |x| \, \nu(dx) < \infty.$$

9. We have already mentioned that there are only three cases when stable distributions are explicitly known. These are (see subsection 5):

normal distribution ($\alpha = 2$),
Cauchy distribution ($\alpha = 1$),
Lévy–Smirnov distribution ($\alpha = 1/2$).

The class of infinitely divisible distribution is much broader; it covers, in addition, the following distributions (although this is not always easy to prove):

Poisson distribution,
gamma-distribution,
geometric distribution,
negative binomial distribution,
t-distribution (*Student distribution*),
F-distribution (*Fisher distribution*),
log-normal distribution
logistic distribution,
Pareto distribution,
two-sided exponential (*Laplace*) *distribution*,
hyperbolic distribution,
Gaussian\\inverse Gaussian distribution ...

TABLE 1
(discrete distributions)

Distribution	Probabilities p_k	Parameters
Poisson	$\dfrac{e^{-\lambda}\lambda^k}{k!}$, $k = 0, 1, \dots$	$\lambda > 0$
Geometric	pq^{k-1}, $k = 1, 2, \dots$	$0 < p \leqslant 1$, $q = 1 - p$
Negative binomial	$C_{k-1}^{r-1} p^r q^{k-r}$, $k = r, r+1, \dots$	$0 < p \leqslant 1$, $q = 1 - p$, $r = 1, 2, \dots$
Binomial	$C_n^k p^k q^{n-k}$, $k = 0, 1, \dots, n$	$0 \leqslant p \leqslant 1$, $q = 1 - p$, $n = 1, 2, \dots$

However, many well-known distributions are *not* infinitely divisible. These include: *binomial* and *uniform* distributions, each *nondegenerate distribution* with finite support, *distributions* with densities $f(x) = Ce^{-|x|^\alpha}$, where $\alpha > 2$.

Some of these distributions are *discrete*, while other have distribution *densities*. For a complete picture and the convenience of references we list these distributions explicitly in Tables 1 and 2.

10. The notion of 'stable' random variable can be naturally extended to the vector-valued case (cf. Definitions 1 and 2).

DEFINITION 6. We call a random vector

$$X = (X_1, X_2, \dots, X_d)$$

a *stable random vector in* $\mathbb{R}^d$ or a *vector with stable d-dimensional distribution* if for each pair of positive numbers A, B there exists a positive number C and a vector $D \in \mathbb{R}^d$ such that

$$\operatorname{Law}\big(AX^{(1)} + BX^{(2)}\big) = \operatorname{Law}(CX + D), \tag{30}$$

where $X^{(1)}$ and $X^{(2)}$ are independent copies of X.

It can be shown (see, e.g., [418; p. 58]) that a nondegenerate random vector $X = (X_1, X_2, \dots, X_d)$ is stable if and only if for each $n \geqslant 2$ there exist $\alpha \in (0, 2]$ and a vector D_n such that

$$\operatorname{Law}\big(X^{(1)} + X^{(2)} + \cdots + X^{(n)}\big) = \operatorname{Law}\big(n^{1/\alpha}X + D_n\big), \qquad 31)$$

where $X^{(1)}, X^{(2)}, \dots, X^{(n)}$ are independent copies of X.

TABLE 2
(distributions with densities)

Distribution	Density $p = p(x)$	Parameters
Uniform on $[a,b]$	$\dfrac{1}{b-a},\ a \leqslant x \leqslant b$	$a, b \in \mathbb{R},\ a < b$
Normal (Gaussian)	$\dfrac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}},\ x \in \mathbb{R}$	$\mu \in \mathbb{R},\ \sigma > 0$
Gamma (Γ-distribution)	$\dfrac{x^{\alpha-1} e^{-x/\beta}}{\Gamma(\alpha)\beta^\alpha},\ x \geqslant 0$	$\alpha > 0,\ \beta > 0$
Exponential (gamma distribution with $\alpha = 1,\ \beta = 1/\lambda$)	$\lambda e^{-\lambda x},\ x \geqslant 0$	$\lambda > 0$
t-distribution (Student distribution)	$\dfrac{\Gamma(\frac{n+1}{2})}{\sqrt{\pi n}\Gamma(\frac{n}{2})} \left(1 + \dfrac{x^2}{n}\right)^{-\frac{n+1}{2}},\ x \in \mathbb{R}$	$n = 1, 2, \dots$
Beta (β-distribution)	$\dfrac{x^{r-1}(1-x)^{s-1}}{\beta(r,s)},\ 0 \leqslant x \leqslant 1$	$r > 0,\ s > 0$
Two-sided exponential (Laplace distribution)	$\dfrac{\lambda}{2} e^{-\lambda \lvert x \rvert},\ x \in \mathbb{R}$	$\lambda > 0$
Chi-squared χ^2 distribution; gamma distribution with $\alpha = n/2,\ \beta = 2$)	$\dfrac{1}{2^{n/2}\Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}},\ x \geqslant 0$	$n = 1, 2, \dots$
Cauchy	$\dfrac{\sigma}{\pi((x-\mu)^2 + \sigma^2)},\ x \in \mathbb{R}$	$\mu \in \mathbb{R},\ \sigma > 0$
Pareto	$\dfrac{\alpha b^\alpha}{x^{\alpha+1}},\ x \geqslant b$	$\alpha > 0, b > 0$
Log-normal	$\dfrac{1}{\sigma x \sqrt{2\pi}} e^{-\frac{(\log x - \mu)^2}{2\sigma^2}},\ x > 0$	$\mu \in \mathbb{R},\ \sigma > 0$
Logistic	$\dfrac{\beta e^{-(\alpha+\beta x)}}{(1 + e^{-(\alpha+\beta x)})^2},\ x \in \mathbb{R}$	$\alpha \in \mathbb{R},\ \beta > 0$
Hyperbolic	see (2) in§ 1d	$\alpha, \beta, \mu, \delta$ see (5) in § 1d
Gaussian\\\\ inverse Gaussian	see (14) in § 1d	$\alpha, \beta, \mu, \delta$. see (5) in § 1d

If $D_n = 0$, i.e.,

$$\operatorname{Law}\bigl(X^{(1)} + X^{(2)} + \cdots + X^{(n)}\bigr) = \operatorname{Law}\bigl(n^{1/\alpha} X\bigr), \tag{32}$$

then we call X a *strictly stable* random vector with exponent α or a *strictly α-stable* random vector.

Remark. Along with using the notation $\operatorname{Law}(X) = \operatorname{Law}(Y)$, one often writes $X \overset{\mathrm{d}}{=} Y$ to indicate that the distributions of X and Y are the same; one says that X and Y coincide *in distribution*. The notation $X^n \xrightarrow{\mathrm{d}} X$ or $\operatorname{Law}(X^n) \to \operatorname{Law}(X)$ indicates convergence in distribution (as already pointed out in subsection 2), i.e., weak convergence of the corresponding measures. (For more detail, see [439; Chapter III].) If $X = (X_t)_{t\geqslant 0}$ and $Y = (Y_t)_{t\geqslant 0}$ are two stochastic processes, then we write

$$\{X_t, t \geqslant 0\} \overset{\mathrm{d}}{=} \{Y_t, t \geqslant 0\}$$

or

$$\operatorname{Law}(X_t, t \geqslant 0) = \operatorname{Law}(Y_t, t \geqslant 0)$$

to denote that all the *finite-dimensional* distributions of X and Y are the same, and we shall say that X and Y coincide *in distribution*.

§ 1b. Lévy Processes

1. The Lévy processes, which we discuss below, are certain stochastic processes with independent increments. They form one of the most important classes of stochastic processes, which includes such basic objects of probability theory as Brownian motions and Poisson processes.

DEFINITION 1. We call a stochastic process $X = (X_t)_{t\geqslant 0}$ with state space $\mathbb{R}^d$ defined on the probability space $(\Omega, \mathscr{F}, \mathsf{P})$ a (d-dimensional) *Lévy process* if

1) $X_0 = 0$ (P-a.s.);
2) for each $n \geqslant 1$ and each collection $t_0, t_1, \dots,\ 0 \leqslant t_0 < t_1 < \cdots < t_n$, the variables $X_{t_0}, X_{t_1} - X_{t_0}, \dots, X_{t_n} - X_{t_{n-1}}$ are independent (the property of *independent increments*);
3) for all $s \geqslant 0$ and $t \geqslant 0$,

$$X_{t+s} - X_s \overset{\mathrm{d}}{=} X_t - X_0$$

(the 'homogeneity' property of increments);

4) for all $t \geqslant 0$ and $\varepsilon > 0$,

$$\lim_{s\to t} \mathsf{P}\bigl(|X_s - X_t| > \varepsilon\bigr) = 0$$

(the property of *stochastic continuity*);

5) for P-almost all $\omega \in \Omega$ the trajectories $(X_t(\omega))_{t\geqslant 0}$ belong to the space D^d of (vector-valued) functions $f = (f_t)_{t\geqslant 0}$, $f_t = (f_t^1, f_t^2, \dots, f_t^d)$ with components $f^i = (f_t^i)_{t\geqslant 0}$, $i = 1, \dots, d$, that are right-continuous and with limits from the left for all $t > 0$.

Remark 1. If we ask that the process $X = (X_t)_{t\geqslant 0}$ in this definition have only properties 1)–4), then it can be shown that *there exists a modification* $X' = (X'_t)_{t\geqslant 0}$ of $X = (X_t)_{t\geqslant 0}$ (i.e., $\mathsf{P}(X'_t \neq X_t) = 0$ for $t \geqslant 0$) with property 5). Thus, the process X' is the same as X as regards the properties 1)–4), but its trajectories are 'regular' in a certain sense. For that reason, one incorporates property 5) of the trajectories in the definition of *Lévy processes* from the very beginning (without loss of generality).

Remark 2. Duly interpreting conditions 1)–5) we can reformulate the definition of a Lévy process as follows: this is a *stochastically continuous process with homogeneous independent increments that starts from the origin and has right-continuous trajectories with limits from the left.*

A classical example of such a process is a d-dimensional *Brownian motion* $X = (X^1, X^2, \dots, X^d)$, the components of which are *independent* standard Brownian motions $X^i = (X_t^i)_{t\geqslant 0}$, $i = 1, \dots, d$.

It is instructive to define a (one-dimensional) Brownian motion separately, outside the general framework of *Lévy processes.*

DEFINITION 2. We call a continuous Gaussian random process $X = (X_t)_{t\geqslant 0}$ a (*standard*) *Brownian motion* or a *Wiener process* if $X_0 = 0$ and

$$\begin{aligned} \mathsf{E}X_t &= 0, \\ \mathsf{E}X_s X_t &= \min(s,t). \end{aligned} \tag{1}$$

By the Gaussian property and (1) we immediately obtain that this is a process with homogeneous independent (Gaussian) increments. Since

$$X_t - X_s \sim \mathcal{N}(0, t-s), \qquad t \geqslant s,$$

it follows that $\mathsf{E}|X_t - X_s|^3 = 3|t-s|^2$, and by the well-known Kolmogorov test ([470]; see also (7) in § 2c below) there exists a continuous modification of this process. Hence the Wiener process (the Brownian motion) is a Lévy process with an additional important feature of *continuous* trajectories.

2. Lévy processes $X = (X_t)_{t\geqslant 0}$ have homogeneous independent increments, therefore their distributions are completely determined by the one-dimensional distributions $\mathsf{P}_t(dx) = \mathsf{P}(X_t \in dx)$. (Recall that $X_0 = 0$.) By the mere definition of these processes, the *distribution* $\mathsf{P}_t(dx)$ *is infinitely divisible for each* t.

Let

$$\varphi_t(\theta) = \mathsf{E}e^{i(\theta,X_t)} = \int_{\mathbb{R}^d} e^{i(\theta,x)}\,\mathsf{P}_t(dx) \tag{2}$$

be the characteristic function. Then, by formula (22) in § 1a (see also (24) in the same § 1a),

$$\varphi_t(\theta) = \exp\left\{ i(\theta, B_t) - \frac{1}{2}(\theta, C_t\theta) + \int_{\mathbb{R}^d} \bigl(e^{i(\theta,x)} - 1 - i(\theta,x)I(|x| \leqslant 1)\bigr)\,\nu_t(dx)\right\}, \tag{3}$$

where $B_t \in \mathbb{R}^d$, C_t is a symmetric nonnegative definite matrix of order $d \times d$, and $\nu_t(dx)$ is the Lévy measure (for each t) with property (23) in § 1a.

The increments of Lévy processes are homogeneous and independent, therefore

$$\varphi_{t+s}(\theta) = \varphi_t(\theta)\varphi_s(\theta), \tag{4}$$

so that

$$\varphi_t(\theta) = \exp\{t\psi(\theta)\}. \tag{5}$$

(The function $\psi = \psi(\theta)$ is called the *cumulant* or the *cumulant function.*)

Since the triplets (B_t, C_t, ν_t) are *unambigously* defined by the characteristic function, it follows by (5) (see, e.g., [250; Chapter II, 4.19] for greater detail) that

$$B_t = t \cdot B, \qquad C_t = t \cdot C, \qquad \nu_t(dx) = t \cdot \nu(dx), \tag{6}$$

where $B = B_1$, $C = C_1$, and $\nu = \nu_1$.

Hence it is clear that in (5) we have

$$\psi(\theta) = i(\theta, B) - \frac{1}{2}(\theta, C\theta) + \int_{\mathbb{R}^d} \bigl(e^{i(\theta,x)} - 1 - i(\theta,x)I(|x| \leqslant 1)\bigr)\,\nu(dx). \tag{7}$$

3. The representation (5) with cumulant $\psi(\theta)$ as in (7) is the main tool in the study of the *analytic* properties of Lévy processes. As regards the properties of their trajectories, on the other hand, the so-called *canonical representation* (see § 3a, Chapter VI and [250; Chapter II, § 2c] for greater detail) is of importance. It generalizes the canonical representations of Chapter II, § 1b for stochastic sequences $H = (H_n)_{n\geqslant 0}$ (see (16) in § 1b and also Chapter IV, § 3e) to the continuous-time case.

4. We now discuss the meaning of components in the triplet $(B_t, C_t, \nu_t)_{t\geqslant 0}$. Figuratively, $(B_t)_{t\geqslant 0}$ is the *trend component* responsible for the development of the process $X = (X_t)_{t\geqslant 0}$ *on the average.* The component $(C_t)_{t\geqslant 0}$ defines the *variance* of the continuous Gaussian component of X, while the *Lévy measures* $(\nu_t)_{t\geqslant 0}$ are responsible for the behavior of the *'jump' component* of X by exhibiting the *frequency and magnitudes of 'jumps'.*

Of course, this (fairly liberal) interpretation must be substantiated by precise results. Here is one such result (see [250; Chapter II, 2.21] as regards the general case).

Let $X = (X_t)_{t\geqslant 0}$ be a process with 'jumps' $|\Delta X_t| \leqslant 1$, $t \geqslant 0$, let $X_0 = 0$, and let $(B_t, C_t, \nu_t)_{t\geqslant 0}$ be the corresponding triplet. Then $\mathsf{E}X_t^2 < \infty$, $t \geqslant 0$, and the following processes are *martingales* (see Chapter II, § 1c):

a) $M_t \equiv X_t - B_t - X_0, \quad t \geqslant 0;$

b) $M_t^2 - C_t, \quad t \geqslant 0;$

c) $\displaystyle\int_0^t \int_{|x|\leqslant 1} g(x)\, \mu_t(dx) - \int_0^t \int_{|x|\leqslant 1} g(x)\, \nu_t(dx), \quad t \geqslant 0,$

where $\mu_t(A) = \sum\limits_{0<s\leqslant t} I(\Delta X_s \in A, \Delta X_s \neq 0)$ is the *measure* of the '*jumps*' of X on the interval $(0, t]$ and $g = g(x)$ are continuous functions vanishing in a neighborhood of the origin.

As already pointed out, a standard Brownian motion is a classical example of a *continuous* Lévy process (with $B_t = 0$, $C_t = t$, and $\nu_t \equiv 0$).

We now consider examples of *discontinuous* Lévy processes, which, at the same time, will give us a better insight into the concept of *Lévy measure* $\nu = \nu(dx)$.

5. The case of a finite Lévy measure ($\nu(\mathbb{R}) < \infty$). A classical example here is, of course, a *Poisson process* $X = (X_t)_{t\geqslant 0}$ with $\lambda > 0$, i.e. (by definition), a Lévy process with $X_0 = 0$ such that X_t has the Poisson distribution with parameter λt:

$$\mathsf{P}(X_t = k) = \frac{e^{-\lambda t}(\lambda t)^k}{k!}, \qquad k = 0, 1, \dots .$$

In this case we have $B_t = \lambda t$ $(= \mathsf{E}X_t)$, $C_t = 0$ and the *Lévy measure* is concentrated at a single point:

$$\nu(dx) = \lambda I_{\{1\}}(dx).$$

The representation (3) takes now the following form:

$$\begin{aligned}
\varphi_t(\theta) &= \exp\Big\{ i\theta(\lambda t) + \int_{\mathbb{R}^d} \big(e^{i\theta x} - 1 - i\theta x I(x \leqslant 1)\big)\, \nu_t(dx) \Big\} \\
&= \exp\Big\{ i\theta(\lambda t) + \int_{\mathbb{R}^d} \big(e^{i\theta x} - 1 - i\theta x I(x = 1)\big)\, \lambda t I_{\{1\}}(dx) \Big\} \\
&= \exp\big\{ \lambda t(e^{i\theta} - 1) \big\}. \qquad (8)
\end{aligned}$$

It is worth noting that, starting form the Poisson process one can arrive at a wide class of *purely jump* Lévy processes.

Namely, let $N = (N_t)_{t\geqslant 0}$ be a Poisson process with parameter $\lambda > 0$ and let $\xi = (\xi_j)_{j\geqslant 1}$ be a sequence of independent identically distributed random variables (that must also be independent of N) with distribution

$$\mathsf{P}(\xi_j \in A) = \frac{\nu(A)}{\lambda}, \qquad A \in \mathscr{B}(\mathbb{R}),$$

where $\lambda = \nu(\mathbb{R}) < \infty$ and $\nu(\{0\}) = 0$.

We consider now the process $X = (X_t)_{t\geqslant 0}$ such that $X_0 = 0$ and

$$X_t = \sum_{j=1}^{N_t} \xi_j, \qquad t > 0. \tag{9}$$

Alternatively, this can be written as follows:

$$X_t = \sum_{j=1}^{\infty} \xi_j I(\tau_j \leqslant t), \tag{10}$$

where $0 < \tau_1 < \tau_2 < \cdots$ are the times of jumps of the process $N = (N_t)_{t\geqslant 0}$.

A direct calculation shows that

$$\begin{aligned}\varphi_t(\theta) = \mathsf{E}e^{i\theta X_t} &= \sum_{k=0}^{\infty} \mathsf{E}\big(e^{i\theta X_t} \,|\, N_t = k\big)\, \mathsf{P}(N_t = k)\\ &= \sum_{k=0}^{\infty} (\mathsf{E}e^{i\theta\xi_1})^k \,\frac{e^{-\lambda t}(\lambda t)^k}{k!} = \exp\Big\{t \int (e^{i\theta x} - 1)\,\nu(dx)\Big\}.\end{aligned} \tag{11}$$

The process $X = (X_t)_{t\geqslant 0}$ defined by (10) is called a *compound Poisson process.* It is easy to see that it is a Lévy process. The 'standard' Poisson process corresponds to the case $\xi_j \equiv 1$, $j \geqslant 1$.

6. The case of infinite Lévy measure ($\nu(\mathbb{R}) = \infty$). To construct the simplest example of a Lévy process with $\nu(\mathbb{R}) = \infty$ we can proceed as follows.

Let $\lambda = (\lambda_k)_{k\geqslant 1}$ be a sequence of positive numbers and let $\beta = (\beta_k)_{k\geqslant 1}$ be a sequence of nonzero real numbers such that

$$\sum_{k=1}^{\infty} \lambda_k \beta_k^2 < \infty. \tag{12}$$

We set

$$\nu(dx) = \sum_{k=1}^{\infty} \lambda_k\, I_{\{\beta_k\}}(dx) \tag{13}$$

and let $N^{(k)} = (N_t^{(k)})_{t\geqslant 0}$, $k \geqslant 1$, be a sequence of independent Poisson processes with parameters λ_k, $k \geqslant 1$, respectively.

Setting

$$X_t^{(n)} = \sum_{k=1}^{n} \beta_k \bigl(N_t^{(k)} - \lambda_k t\bigr), \tag{14}$$

it is easy to see that for each $n \geqslant 1$ the process $X^{(n)} = (X_t^{(n)})_{t\geqslant 0}$ is a Lévy process with Lévy measure

$$\nu^{(n)}(dx) = \sum_{k=1}^{n} \lambda_k \, I_{\{\beta_k\}}(dx) \tag{15}$$

and

$$\varphi_t^{(n)}(\theta) = \mathsf{E} e^{i\theta X_t^{(n)}} = \exp\biggl\{ t \int \bigl(e^{i\theta x} - 1 - i\theta x\bigr)\, \nu^{(n)}(dx) \biggr\}. \tag{16}$$

The limit process $X = (X_t)_{t\geqslant 0}$,

$$X_t = \sum_{k=1}^{\infty} \beta_k \bigl(N_t^{(k)} - \lambda_k t\bigr) \tag{17}$$

(X_t is the L^2-limit of the $X_t^{(n)}$ as $n \to \infty$) is also a Lévy process with 'Lévy measure' defined by (13).

Remark 3. In this case we have property 5) from Definition 1 because the $X^{(n)}$ are square integrable martingales, for which, by the *Doob inequality* (see formula (36) in § 3b and also [250; Chapter I, 1.43] or [439; Chapter VII, § 3]), we have

$$\mathsf{E} \max_{s\leqslant t} |X_s^{(n)} - X_s|^2 \to 0 \quad \text{as} \quad n \to \infty.$$

Remark 4. Since $\nu(\mathbb{R}) = \sum_{k=1}^{\infty} \lambda_k$, $\nu(\{0\}) = 0$ and

$$\int_{\mathbb{R}} (x^2 \wedge 1)\, \nu(dx) \leqslant \sum_{k=1}^{\infty} \lambda_k \beta_k^2 < \infty,$$

the measure $\nu = \nu(dx)$ in (13) satisfies all the conditions imposed on a Lévy measure (see formulas (22)–(23) in § 1a).

If $\sum_{k=1}^{\infty} \lambda_k = \infty$, but (12) holds, then we have an example of a Lévy process with Lévy measure ν satisfying the relation $\nu(\mathbb{R}) = \infty$.

We now present another well-known example of a Lévy process with $\nu(\mathbb{R}) = \infty$. We mean the so-called *gamma process* $X = (X_t)_{t\geqslant 0}$ with $X_0 = 0$ and probability distribution $\mathsf{P}(X_t \leqslant x)$ with density

$$p_t(x) = \frac{x^{t-1} e^{-x/\beta}}{\Gamma(t)\beta^t}\, I_{(0,\infty)}(x) \tag{18}$$

(cf. Table 2 in § 1a).

Now,

$$\varphi_t(\theta) = (1 - i\theta\beta)^{-t} \tag{19}$$

We claim that the characteristic function $\varphi_t(\theta)$ can be represented as follows:

$$\varphi_t(\theta) = \exp\left\{t\int_0^\infty (e^{i\theta x} - 1)\frac{e^{-x/\beta}}{x}\,dx\right\}, \tag{20}$$

which yields the equality $\nu_t(dx) = t\,\nu(dx)$, where

$$\nu(dx) = I_{(0,\infty)}(x)\,\frac{e^{-x/\beta}}{x}\,dx. \tag{21}$$

Clearly, $\nu(0,\infty) = \infty$, however

$$\int_0^\infty (x^2 \wedge 1)\,\nu(dx) < \infty. \tag{22}$$

We now proceed to the proof of the representation (20).

We consider the Laplace transform

$$\begin{aligned} L_t(u) = \mathsf{E}e^{-uX_t} &= \int_0^\infty e^{-ux}p_t(x)\,dx = (1+\beta u)^{-t} \\ &= \exp\{-t\ln(1+\beta u)\} = \exp\left\{-t\int_0^u \frac{dy}{\frac{1}{\beta}+y}\right\} \\ &= \exp\left\{-t\int_0^u dy\int_0^\infty e^{-\frac{x}{\beta}-yx}\,dx\right\} \\ &= \exp\left\{t\int_0^\infty (e^{-ux}-1)\frac{e^{-x/\beta}}{x}\,dx\right\}. \end{aligned}$$

Using analytical continuation in the complex half-plane $\{z = a + ib,\ a \leqslant 0\}$ we obtain

$$\int_{\mathbb{R}} e^{zx}\,p_t(dx) = \exp\left\{t\int_0^\infty (e^{zx}-1)\frac{e^{-x/\beta}}{x}\,dx\right\}.$$

Setting here $z = i\theta$, we arrive at the required representation (20).

7. Now that we have obtained 'explicit' representations (10), (14), and (17) for several (jump) Lévy processes, we have at our disposal a method of their *simulation*. We need only simulate the random variables ξ_j and β_k and the exponentially distributed variables $\Delta_i = \tau_i - \tau_{i-1}$ (the time intervals between two jumps of the Poisson process at instants τ_{i-1} and τ_i). For a simulation of infinitely divisible random variables in their turn, the question of their representations as functions of 'elementary', 'standard' random variables becomes now of interest. Look at an example demonstrating opportunities emerging here. Let X and Y be two independent random variables such that $X \geqslant 0$ (and X is arbitrary in the rest), while Y has an exponential distribution. Then, as shown by Ch. Goldie, their product XY is an infinitely divisible random variable.

In the next section we shall show how one can 'combine' processes of simple structure and obtain stable processes as result.

§ 1c. Stable Processes

1. We start with the case of a *real-valued* process $X = (X_t)_{t\in T}$ with an (arbitrary) parameter set T, which is defined on some probability space $(\Omega, \mathscr{F}, \mathsf{P})$.

Generalizing the definition of a stable random vector (Definition 6 in § 1a) we arrive in a natural way to the following concept.

DEFINITION 1. We say that a real random process $X = (X_t)_{t\in T}$ is *stable* if for each $k \geqslant 1$ and all $t_1, \dots, t_k$ in T the random vectors $(X_{t_1}, \dots, X_{t_k})$ are stable, i.e., all finite-dimensional distributions of X are stable.

If all these finite-dimensional distributions are stable, then it is easy to see from their compatibility that they have the same stability exponent α. This explains the name *α-stable* that one often gives to such processes when one wants to point out some particular value of α.

In what follows, we shall be mainly interested in α-stable processes $X = (X_t)_{t\in T}$ that are incidentally Lévy processes (§ 1b). A suitable name for them is *α-stable Lévy processes.*

In § 1a.10 we presented two equivalent definitions of stable nondegenerate random vectors (see formulas (30) and (31) in § 1a). In the case of stable (one-dimensional) processes $X = (X_t)_{t\geqslant 0}$ that are also Lévy processes, this equivalence brings us to the following result: a *nondegenerate* (see Definition 2 below) *one-dimensional Lévy process $X = (X_t)_{t\geqslant 0}$ is α-stable* ($\alpha \in (0,2]$) *if and only if for each $a > 0$ there exists a number D* (depending on a in general) *such that*

$$\{X_{at},\, t \geqslant 0\} \stackrel{\mathrm{d}}{=} \{a^{1/\alpha} X_t + Dt\}. \tag{1}$$

We now present several definitions relating to *multidimensional* processes $X = (X_t)_{t\geqslant 0}$.

DEFINITION 2. We say that a random process $(X_t)_{t\geqslant 0}$ taking values in $\mathbb{R}^d$ is *degenerate* if $X_t = \gamma t$ (P-a.s.) for some $\gamma \in \mathbb{R}^d$ and all $t \geqslant 0$. Otherwise we say that the process X is *nondegenerate.*

DEFINITION 3. We call a nondegenerate random process $X = (X_t)_{t\geqslant 0}$ taking values in $\mathbb{R}^d$ an *α-stable Lévy process* ($\alpha \in (0,2]$) if

1) X is a Lévy process

and

2) for each $a > 0$ there exists $D \in \mathbb{R}^d$ (dependent on a in general) such that

$$\{X_{at},\, t \geqslant 0\} \stackrel{\mathrm{d}}{=} \{a^{1/\alpha} X_t + Dt,\, t \geqslant 0\} \tag{2}$$

or, equivalently,

$$\operatorname{Law}(X_{at},\, t \geqslant 0) = \operatorname{Law}(a^{1/\alpha} X_t + Dt,\, t \geqslant 0). \tag{3}$$

DEFINITION 4. We call an α-stable Lévy process $X = (X_t)_{t\geqslant 0}$ a *strictly α-stable Lévy process* if $D = 0$ in (2) and (3), i.e.,

$$\{X_{at},\, t \geqslant 0\} \stackrel{\mathrm{d}}{=} \{a^{1/\alpha} X_t,\, t \geqslant 0\} \tag{4}$$

or, equivalently,

$$\operatorname{Law}(X_{at},\, t \geqslant 0) = \operatorname{Law}(a^{1/\alpha} X_t,\, t \geqslant 0). \tag{5}$$

Remark 1. Sometimes (see, e.g., [423]), a stable vector-valued Lévy random process $X = (X_t)_{t\geqslant 0}$ is defined as follows: for each $a > 0$ there exist a number c and $D \in \mathbb{R}^n$ such that

$$\{X_{at},\, t \geqslant 0\} \stackrel{\mathrm{d}}{=} \{cX_t + Dt,\, t \geqslant 0\} \tag{6}$$

or, equivalently,

$$\operatorname{Law}(X_{at},\, t \geqslant 0) = \operatorname{Law}(cX_t + Dt,\, t \geqslant 0). \tag{7}$$

(If $D = 0$, then one talks about strict stability.) It is remarkable that, as in the case of stable variables and vectors, we have $c = a^{1/\alpha}$ for nondegenerate Lévy processes, where α is a universal (i.e., independent of a) parameter belonging to $(0, 2]$. This result, the proof of which can be found, e.g., in [423], explains the explicit involvement of the coefficient $a^{1/\alpha}$ in Definitions 3 and 4.

Remark 2. It is useful to note that the condition

$$\operatorname{Law}(X_{at},\, t \geqslant 0) = \operatorname{Law}(cX_t,\, t \geqslant 0)$$

is precisely that of *self-similarity*. Thus, α-stable Lévy processes, for which $c = a^{1/\alpha}$, are self-similar. The quantity $\mathbb{H} = 1/\alpha$ here is called the *Hurst parameter* or the *Hurst exponent*. See § 2c for greater detail.

Remark 3. If a process $X = (X_t)_{t\geqslant 0}$ is α-stable $(0 < \alpha \leqslant 2)$ and has simultaneously the self-similarity property

$$\operatorname{Law}(X_{at},\, t \geqslant 0) = \operatorname{Law}(a^{\mathbb{H}} X_t,\, t \geqslant 0), \qquad a > 0,$$

but is not a Lévy process, then the formula $\mathbb{H} = 1/\alpha$ fails. Various pairs $(\alpha, \mathbb{H})$ can correspond to such processes, provided that

$$\alpha < 1 \quad \text{and} \quad 0 < \mathbb{H} \leqslant 1/\alpha,$$

or

$$\alpha \geqslant 1 \quad \text{and} \quad 0 < \mathbb{H} \leqslant 1.$$

(See [418; Corollary 7.1.11 and Figure 7.1].)

2. Since α-stable processes are particular cases of Lévy processes $X = (X_t)_{t\geqslant 0}$, the characteristic functions of which $\varphi_t(\theta) = \mathsf{E}e^{i(\theta,X_t)}$ have the representation $\varphi_t(\theta) = \exp\{t\psi(\theta)\}$ with cumulant $\psi(\theta)$ defined by formula (7) in § 1b, one may well ask what are the parameters B, C, and ν corresponding to these (α-stable) processes. Of a particular interest here is the Lévy measure $\nu = \nu(dx)$, which is 'responsible' for the magnitudes of jumps $\Delta X_t \equiv X_t - X_{t-}$ of $X = (X_t)_{t\geqslant 0}$.

We now present, following [423], the main results obtained in this direction.

THEOREM 1. *Let $X = (X_t)_{t\geqslant 0}$ be a nondegenerate Lévy process in $\mathbb{R}^d$ with triplet (B, C, ν). Then*

1) *the process X is a 2-stable (Lévy) process if and only if $\nu = 0$, i.e., if and only if it is Gaussian;*
2) *the process X is strictly 2-stable (Lévy) process if and only if it is a Gaussian process with zero mean ($B = 0$).*

THEOREM 2. *Let $X = (X_t)_{t\geqslant 0}$ be a nondegenerate Lévy process in $\mathbb{R}^d$ with triplet (B, C, ν). Assume that $0 < \alpha < 2$. Then X is an α-stable (Lévy) process if and only if $C = 0$ and the Lévy measure ν is as follows:*

$$\nu(A) = \int_S \lambda(d\xi) \int_0^\infty I_A(r\xi) r^{-(1+\alpha)}\, dr, \qquad A \in \mathscr{B}(\mathbb{R}^d \setminus \{0\}), \tag{8}$$

where λ is some nonzero finite measure in $S = \{x \in \mathbb{R}^d : |x| = 1\}$.

The cumulant $\psi(\theta)$ of such a process has the following representation:

$$\psi(\theta) = i(\theta, B) + \int_S \lambda(d\xi) \int_0^\infty \big(e^{i(\theta,r\xi)} - 1 - i(\theta, r\xi) I_{(0,1]}(r)\big) r^{-(1+\alpha)}\, dr. \tag{9}$$

Noteworthy, if $0 < \alpha < 2$, then the radial part of the Lévy measure is $r^{-(1+\alpha)}dr$. With a decrease of α, $r^{-(1+\alpha)}$ decreases for $0 < r < 1$ and increases for $1 < r < \infty$. Thus, we can say that *large* jumps of the process trajectories prevail for α close to zero, but the process is developing in *small* jumps if α is close to two. Good illustrations to this property can be found in [253].

THEOREM 3. *Let $X = (X_t)_{t\geqslant 0}$ be a nondegenerate Lévy process in $\mathbb{R}^d$ with triplet (B, C, ν).*

1) *Let $\alpha \in (0, 1)$. Then X is a strictly α-stable (Lévy) process if and only if it has the cumulant*

$$\psi(\theta) = \int_S \lambda(d\xi) \int_0^\infty \big(e^{i(\theta,r\xi)} - 1\big) r^{-(1+\alpha)}\, dr \tag{10}$$

with some nonzero finite measure λ in S, i.e., if and only if $C = 0$ and the 'drift' (see (27) in § 1a) is zero.

2) *Let $\alpha \in (1,2)$. Then X is a strictly α-stable (Lévy) process if and only if $C = 0$ and the 'center' (see (29) in § 1a) is zero (i.e., $\mathsf{E}X = 0$), or, equivalently, if and only if its cumulant is*

$$\psi(\theta) = \int_S \lambda(d\xi) \int_0^\infty \big(e^{i(\theta,r\xi)} - 1 - i(\theta, r\xi)\big) r^{-(1+\alpha)}\, dr \tag{11}$$

with some nonzero finite measure λ in S.

3) *Let $\alpha = 1$. Then X is a strictly 1-stable (Lévy) process if and only if*

$$\psi(\theta) = i(\theta, B) + \int_S \lambda(d\xi) \int_0^\infty \big(e^{i(\theta,r\xi)} - 1 - i(\theta, r\xi) I_{(0,1]}\big) r^{-2}\, dr \tag{12}$$

for some finite measure λ in S and some constant B such that

$$\int_S x\, \lambda(d\xi) = 0 \tag{13}$$

and $\lambda(S) + |B| > 0$.

COROLLARY. *Let $X = (X_t)_{t\geqslant 0}$ be an α-stable Lévy process in $\mathbb{R}^d$. If $\alpha \neq 1$, then there always exists $\gamma \in \mathbb{R}^d$ such that the centered process $X = (X_t - \gamma t)_{t\geqslant 0}$ is strictly α-stable.*

On the other hand if $\alpha = 1$ and (13) holds, then the process $X = (X_t)_{t\geqslant 0}$ is itself strictly 1-stable.

For $d = 1$ we can explicitly calculate integrals in the representations (9)–(11) for cumulants to obtain

$$\psi(\theta) = \begin{cases} i\mu\theta - \sigma^\alpha|\theta|^\alpha\Big(1 - i\beta(\operatorname{Sgn}\theta)\tan\dfrac{\pi\alpha}{2}\Big), & \alpha \neq 1, \qquad (14)\\[2mm] i\mu\theta - \sigma|\theta|\Big(1 + i\beta\dfrac{2}{\pi}(\operatorname{Sgn}\theta)\ln|\theta|\Big), & \alpha = 1, \qquad (15)\end{cases}$$

for $0 < \alpha \leqslant 2$ (cf. formula (6) in § 1a), where $\beta \in [-1,1]$, $\sigma > 0$, and $\mu \in \mathbb{R}$.

A (nonzero) α-stable Lévy process is *strictly* stable if and only if

$$\mu = 0 \quad \text{for} \quad \alpha \neq 1,$$

and

$$\beta = 0, \quad \sigma + |\mu| > 0 \quad \text{for} \quad \alpha = 1.$$

3. We shall now discuss formulas (2) and (3) in the definition of an α-stable process. For $t = 1$ formula (1) can be written as follows:

$$X_a \overset{\mathrm{d}}{=} a^{1/\alpha} X_1 + D_a, \tag{16}$$

where D_a is a constant. Using the representations (14) and (15) for the characteristic function of this process we obtain

$$D_a = \begin{cases} (a - a^{1/\alpha})\mu, & \alpha \neq 1,\\ \beta\dfrac{2}{\pi}\sigma^\alpha a \ln a, & \alpha = 1. \end{cases} \tag{17}$$

4. We see from the above that operating the distributions of stable process can be difficult because it is only in three cases that we have explicit formulas for the densities (see § 1a.5).

However, in certain situations we can show a way for constructing stable processes, e.g., from a Brownian motion, by means of a *random* (and independent on the motion) *change of time.*

We present an interesting example in this direction concerning the simulation of symmetric α-stable distributions using *three* independent random variables: *uniformly distributied, Gaussian, and exponentially distributed.*

Let $Z = (Z_t)_{t\geqslant 0}$ be a symmetric α-stable Lévy process with characteristic function

$$\varphi_t(\theta) = \mathsf{E}e^{i\theta Z_t} = e^{-t|\theta|^\alpha}, \tag{18}$$

where $0 < \alpha < 2$.

It will be clear from what follows that the process Z can be realized as

$$Z_t = B_{T_t}, \qquad t \geqslant 0, \tag{19}$$

where $B = (B_t)_{t\geqslant 0}$ is a Brownian motion with $\mathsf{E}B_t = 0$ and $\mathsf{E}B_t^2 = 2t$, while $T = (T_t)_{t\geqslant 0}$ is a nonnegative nondecreasing $\frac{\alpha}{2}$-stable random process, which is called a stable subordinator. We say of the process Z obtained by a transformation (19) that it is constructed from a Brownian motion by means of a *random change of time (subordination)* $T = (T_t)_{t\geqslant 0}$.

The process $T = (T_t)_{t\geqslant 0}$ required for (19) can be constructed as follows.

Let $U^{(\alpha)} = U^{(\alpha)}(\omega)$ be a nonnegative stable random variable with Laplace transform

$$\mathsf{E}e^{-\lambda U^{(\alpha)}} = e^{-\lambda^\alpha}, \qquad \lambda > 0, \tag{20}$$

where $0 < \alpha < 1$.

Note that if $U^{(\alpha)}$ and $U_1, \dots, U_n$ are independent identically distributed random variables, then the sum

$$n^{-1/\alpha} \sum_{j=1}^{n} U_j \tag{21}$$

has the same Laplace transform as $U^{(\alpha)}$, so that $U^{(\alpha)}$ is indeed a stable random variable.

Assume that $0 < \alpha < 2$. We now construct a nonnegative nondecreasing $\frac{\alpha}{2}$-stable process $T = (T_t)_{t\geqslant 0}$ such that $\mathrm{Law}(T_1) = \mathrm{Law}(U^{(\alpha/2)})$.

By the 'self-similarity' property (5),

$$\mathrm{Law}(T_t) = \mathrm{Law}(t^{2/\alpha} T_1) = \mathrm{Law}(t^{2/\alpha} U^{(\alpha/2)}). \tag{22}$$

therefore

$$\begin{aligned}\mathsf{E}e^{i\theta Z_t} &= \mathsf{E}e^{i\theta B_{T_t}} = \mathsf{E}\big[\mathsf{E}\big(e^{i\theta B_{T_t}} \mid T_t\big)\big] = \mathsf{E}e^{-\frac{\theta^2 2T_t}{2}} = \mathsf{E}e^{-\theta^2 T_t} \\ &= \mathsf{E}e^{-\theta^2 t^{2/\alpha} U^{(\alpha/2)}} = e^{-(\theta^2 t^{2/\alpha})^{\alpha/2}} = e^{-t|\theta|^\alpha}, \end{aligned} \tag{23}$$

which delivers the required representation (19).

Let $p = p(x;\alpha)$ be the distribution density of the variable $U = U^{(\alpha)}$, where $0 < \alpha < 1$. By [238] and [264], this density, which is concentrated on the half-axis $x \geqslant 0$, has the following 'explicit' representation:

$$p(x;\alpha) = \frac{1}{\pi}\left(\frac{\alpha}{1-\alpha}\right)\left(\frac{1}{x}\right)^{\frac{1}{1-\alpha}} \int_0^\pi a(z;\alpha)\exp\left\{-\left(\frac{1}{x}\right)^{\frac{\alpha}{1-\alpha}} a(z;\alpha)\right\} dz, \tag{24}$$

where

$$a(z;\alpha) = \left(\frac{\sin\alpha z}{\sin z}\right)^{\frac{1}{1-\alpha}} \frac{\sin(1-\alpha)z}{\sin\alpha z}. \tag{25}$$

As observed by H. Rubin (see [264; Corollary 4.1]), the density $p(x;\alpha)$ is the distribution density of the random variable

$$\zeta = \left(\frac{a(\xi;\alpha)}{\eta}\right)^{\frac{1-\alpha}{\alpha}}, \tag{26}$$

where $a = a(z;\alpha)$ is as in (25), ξ and η are independent random variables, ξ is uniformly distributed on $[0,\pi]$, and η has the exponential distribution with parameter one.

Remark. The fact that $p(x;\alpha)$ is the distribution density of ζ can be verified with no difficulty. For let

$$h(x) = \frac{1}{\pi}\int_0^\pi a(z;\alpha)\exp\big(-xa(z;\alpha)\big)\,dz.$$

Then $h = h(x)$ is obviously the density for $\eta/a(\xi;\alpha)$. A simple change of variables shows now that $p(x;\alpha)$ is the distribution density for ζ.

Thus,

$$\operatorname{Law}(T_1) = \operatorname{Law}(U^{(\alpha/2)}) = \operatorname{Law}\left(\left(\frac{a(\xi;\alpha/2)}{\eta}\right)^{\frac{2-\alpha}{\alpha}}\right), \tag{27}$$

and, by (22),

$$\operatorname{Law}(T_t) = \operatorname{Law}(t^{2/\alpha} T_1). \tag{28}$$

From (27) and (28), denoting a Gaussian random variable with zero mean and variance 2 by $\gamma(0,2)$, we see that

$$\begin{aligned}
\operatorname{Law}(Z_t - Z_s) &= \operatorname{Law}\bigl(B_{T_t} - B_{T_s}\bigr) = \operatorname{Law}\bigl(B_{T_t - T_s}\bigr) \\
&= \operatorname{Law}\Bigl(\sqrt{T_t - T_s}\,\gamma(0,2)\Bigr) = \operatorname{Law}\Bigl(\sqrt{T_{t-s}}\,\gamma(0,2)\Bigr) \\
&= \operatorname{Law}\Bigl((t-s)^{1/\alpha}\sqrt{T_1}\,\gamma(0,2)\Bigr) \\
&= \operatorname{Law}\biggl((t-s)^{1/\alpha}\Bigl(\frac{a(\xi;\alpha/2)}{\eta}\Bigr)^{\frac{2-\alpha}{2\alpha}}\gamma(0,2)\biggr).
\end{aligned}$$

This representation for $\operatorname{Law}(Z_t - Z_s)$ shows a way in which, simulating *three* independent random variables ξ, η, and $\gamma = \gamma(0,2)$, one can obtain an observation sample for the increments $Z_t - Z_s$ of a symmetric α-stable random process.

Remark. For general results on the construction of Lévy processes from other Lévy processes (of a simpler structure) by means of subordinators, see [47], [239], [409], and [483].

The process $Z = (Z_t)_{t\geqslant 0}$ is used in [327] for the description of the behavior of prices (the *Mandelbrot–Taylor model*). It is worth noting that if t is real, '*physical*' time, then T_t can be interpreted as '*operational*' time (see Chapter IV, § 3d) or as the random 'number of transitions' that occurred before time t. (This, slightly loose, interpretation is inspired by similarities with the sums $\sum_{k=1}^{T_n} \xi_k$ *of a random number* T_n *of random variables* ξ_k, $k \geqslant 1$.)

We must emphasize that for each t the distribution of $Z_t = B_{T_t}$ is a *mixture* of Gaussian distributions. In other words we can say that the distribution of the Z_t is *conditionally Gaussian*. We have already discussed such distributions (see Chapter II, § 1d and § 3a). Below, in § 1d, we shall consider models based on *hyperbolic* distributions, which are also conditionally Gaussian and belong to the class of infinitely divisible distributions, but are not stable. All this indicates that in our search for an adequate description of the evolution of prices we must, in a certain sense, look towards conditionally Gaussian distributions and processes.

5. In conclusion we consider three cases of stable Lévy processes corresponding to the three cases of known explicit formulas for stable densities (see § 1a.5).

EXAMPLE 1. A standard Brownian motion $X = (X_t)_{t\geqslant 0}$ in $\mathbb{R}^d$ is a strictly 2-stable Lévy process. The corresponding probability distribution $\mathsf{P}_1 = \mathsf{P}_1(dx)$ of X_1 is as follows:

$$\mathsf{P}_1(dx) = (2\pi)^{-d/2} e^{-|x|^2/2}\, dx, \qquad x \in \mathbb{R}^d. \tag{29}$$

The characteristic function of X_t is

$$\varphi_{X_t}(\theta) = \mathsf{E} e^{i(\theta, X_t)} = e^{-\frac{t}{2}|\theta|^2}, \tag{30}$$

and one immediately sees that (cf. (14))

$$\varphi_{X_{at}}(\theta) = e^{-\frac{at}{2}|\theta|^2} = e^{-\frac{t}{2}|\sqrt{a}\theta|^2} = \varphi_{\sqrt{a}X_t}(\theta). \tag{31}$$

EXAMPLE 2. A standard Cauchy process in $\mathbb{R}^d$ is a strictly 1-stable Lévy process with

$$\mathsf{P}_1(dx) = \pi^{-\frac{d+1}{2}}\Gamma\left(\frac{d+1}{2}\right)\left(1+|x|^2\right)^{-\frac{d+1}{2}}dx, \qquad x \in \mathbb{R}^d. \tag{32}$$

The characteristic function of X_t is

$$\varphi_{X_t}(\theta) = e^{-t|\theta|}, \tag{33}$$

and, clearly (cf. (14)),

$$\varphi_{X_{at}}(\theta) = e^{-at|\theta|} = e^{-t|a\theta|} = \varphi_{aX_t}(\theta). \tag{34}$$

EXAMPLE 3. For a one-sided strictly $\frac{1}{2}$-stable Lévy process in $(0,\infty)$ we have

$$\mathsf{P}_1(dx) = (2\pi)^{-1/2} I_{(0,\infty)}(x) e^{1/(2x)} x^{-3/2}\,dx, \qquad x \in \mathbb{R}, \tag{35}$$

and

$$\varphi_{X_t}(\theta) = \exp\left\{-t\sqrt{|\theta|}\,(1 - i\operatorname{Sgn}\theta)\right\}. \tag{36}$$

We immediately see that

$$\begin{aligned}\varphi_{aX_t}(\theta) &= \exp\left\{-at\sqrt{|\theta|}\,(1-\operatorname{Sgn}\theta)\right\}\\ &= \exp\left\{-t\sqrt{|a^2\theta|}\,(1-i\operatorname{Sgn}(a^2\theta))\right\} = \varphi_{a^2X_t}(\theta).\end{aligned} \tag{37}$$

These examples virtually exhaust all known cases when the one-dimensional distributions of X_1 (and therefore, of X_t) can be expressed in terms of *elementary* functions.

§ 1d. Hyperbolic Distributions and Processes

1. In 1977, O. Barndorff-Nielsen [21] introduced a class of distributions that is interesting in many respects, the so-called *generalized hyperbolic distributions*. His motivation was the desire to find an adequate explanation for some empirical laws in geology; subsequently, these distributions found applications to geomorphology, turbulence theory, ..., and also to financial mathematics.

Generalized hyperbolic distributions are not stable, but they are close to stable ones in that they can be characterized by several *parameters* of a close meaning (see subsection 2).

We distinguish two distributions in this class that are most frequent in applications:

1) the *hyperbolic distribution* in the proper sense;
2) the *Gaussian**inverse Gaussian distribution*.

It should be mentioned that these distributions are *mixtures* of Gaussian ones. Hence they are naturally consistent with the idea of the use of *conditionally Gaussian* distributions. At the same time, the distributions in question are infinitely divisible and form a fairly wide subclass of infinitely divisible distributions. Judging by the behavior of their 'tails', they are intermediate between stable (with exponent $\alpha < 2$) and Gaussian distributions (with $\alpha = 2$): their 'tails' decrease faster than for stable ($\alpha < 2$), but slower than for Gaussian distributions.

2. The attribute 'hyperbolic' is due to the following observation.

For a *normal* (Gaussian) density

$$\varphi(x) = \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{1}$$

the graph of its logarithm $\ln\varphi(x)$ is a *parabola*, while the graph of the logarithm $\ln h_1(x)$ of the density

$$h_1(x) = C_1(\alpha,\beta,\delta)\exp\Bigl\{-\alpha\sqrt{\delta^2 + (x-\mu)^2} + \beta(x-\mu)\Bigr\} \tag{2}$$

of a *hyperbolic* distribution is a *hyperbola*

$$f(x) = \ln C_1(\alpha,\beta,\delta) - \alpha\sqrt{\delta^2 + (x-\mu)^2} + \beta(x-\mu) \tag{3}$$

with asymptotes

$$a(x) = -\alpha|x-\mu| + \beta(x-\mu). \tag{4}$$

The *four* parameters $(\alpha,\beta,\mu,\delta)$ in the definition (2) of a hyperbolic distribution are assumed to satisfy the conditions

$$\alpha > 0, \quad 0 \leqslant |\beta| < \alpha, \quad \mu \in \mathbb{R}, \quad \delta \geqslant 0. \tag{5}$$

The parameters α and β 'govern' the form of the density graph, μ is the location parameter, and δ is the scale parameter. The constant C_1 has the representation

$$C_1(\alpha,\beta,\delta) = \frac{\sqrt{\alpha^2-\beta^2}}{2\alpha\delta K_1(\delta\sqrt{\alpha^2-\beta^2}\,)}, \tag{6}$$

where $K_1(x)$ is the modified third-kind Bessel function of index 1 (see [23]).

One often uses another set of parameters to describe a hyperbolic distribution. Namely, one sets

$$\alpha = \frac{1}{2}(\varphi+\gamma) \quad \text{and} \quad \beta = \frac{1}{2}(\varphi-\gamma), \tag{7}$$

so that $\varphi\gamma = \alpha^2 - \beta^2$.

We shall denote the density $h_1(x) \equiv h_1(x; \alpha, \beta, \mu, \delta)$ expressed in terms of these new parameters by $h_2(x) \equiv h_2(x; \varphi, \gamma, \mu, \delta)$. It has the following representation:

$$h_2(x) = C_2(\varphi, \gamma, \delta) \exp\Bigl\{-\frac{1}{2}(\varphi + \gamma)\sqrt{\delta^2 + (x-\mu)^2} + \frac{1}{2}(\varphi - \gamma)(x - \mu)\Bigr\}, \tag{8}$$

where

$$C_2(\varphi, \gamma, \delta) = \frac{\omega}{\delta \varkappa K_1(\delta \varkappa)}$$

with $\varkappa = (\varphi\gamma)^{1/2}$ and $\omega^{-1} = (\varphi^{-1} + \gamma^{-1})$. (The last parameters are used in [289].)

It is clear from (8) that if a random variable X has density $h_2(x; \varphi, \gamma, \mu, \delta)$, then the variable $Y = (X - a)/b$, where $a \in \mathbb{R}$ and $b > 0$, has the distribution density $h_2\bigl(x; b\varphi, b\gamma, \delta/b, (\mu - a)/b\bigr)$. Thus, the class of hyperbolic distributions is invariant under *shifts* and *scaling*.

It is also clear from (8) that $h_2(x) > 0$ for all $x \in \mathbb{R}$, and the 'tails' of $h_2(x)$ decrease *exponentially* at the 'rate' φ as $x \to -\infty$ and at the 'rate' γ as $x \to \infty$.

As $\delta \to \infty$, $\delta/\varkappa \to \sigma^2$, and $\varphi - \gamma \to 0$, we have

$$h_2(x; \varphi, \gamma, \mu, \delta) \to \varphi(x),$$

where $\varphi(x) \equiv \varphi(x; \mu, \sigma^2)$ is the normal density.

As $\delta \to 0$, we obtain in the limit the Laplace distribution (which is asymmetric for $\varphi \neq \gamma$) with density

$$\lambda(x; \varphi, \gamma, \mu) = \omega^{-1} \exp\Bigl\{-\frac{1}{2}(\varphi + \gamma)|x - \mu| + \frac{1}{2}(\varphi - \gamma)(x - \mu)\Bigr\}.$$

Introducing the parameters

$$\xi = (1 + \delta\sqrt{\varphi\gamma}\,)^{-1/2} \quad \Bigl(= \bigl(1 + \delta\sqrt{\alpha^2 - \beta^2}\,\bigr)^{-1/2}\Bigr)$$

and

$$\chi = (\varphi - \gamma)(\varphi + \gamma)^{-1}\xi \quad \Bigl(= \frac{\beta}{\alpha}\xi\Bigr),$$

we notice that they do not change when one passes from a random variable X with hyperbolic density $h_2(x; \varphi, \gamma, \mu, \delta)$ to a random variable $Y = X - a$ with density $h_2(x; \varphi, \gamma, \mu - a, \delta)$ indicated above. These parameters ξ and χ, which have the meaning of skewness and kurtosis, are good indicators of the deviation of a distribution from normality (see Chapter IV, § 2b for greater detail).

We note that the range of (χ, ξ) is the interior of the triangle

$$\nabla = \bigl\{(\chi, \xi) \colon 0 \leqslant |\chi| < \xi < 1\bigr\}$$

(see Fig. 26; here $\mathscr{N}$ corresponds to the normal distribution, $\mathscr{E}$ to the exponential, and $\mathscr{L}$ to the Laplace distribution).

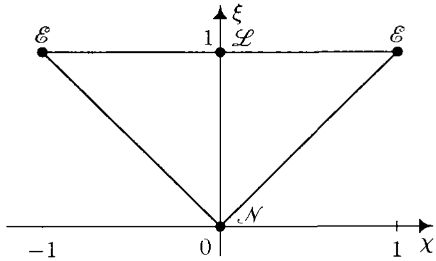


FIGURE 26

The boundary point $(0,0) \notin \nabla$ corresponds to the normal distribution; the points $(-1,1)$ and $(1,1)$, also lying outside ∇, to the exponential distribution, and $(0,1) \notin \nabla$ to the Laplace distribution. Passing to the limit as $\chi \to \pm\xi$ we obtain (see [21]–[23], [25], [26]) the so-called generalized inverse Gaussian distribution.

It was mentioned in [21], [23], [25], [26], [22] that a hyperbolic distribution is a *mixture* of Gaussian ones: if a random variable X has density $h_1(x;\alpha,\beta,\mu,\delta)$, then

$$\operatorname{Law} X = \mathsf{E}'_{\sigma^2}\mathscr{N}(\mu+\beta\sigma^2,\sigma^2), \tag{9}$$

where by E'_{σ^2} we mean the result of averaging with respect to the parameter σ^2 that has the inverse Gaussian distribution with density

$$p'_{\sigma^2}(x) = \frac{\sqrt{a/b}}{2K_1(\sqrt{ab}\,)}\exp\left\{-\frac{1}{2}\left(ax+\frac{b}{x}\right)\right\}; \tag{10}$$

here $a=\alpha^2-\beta^2$ and $b=\delta^2$.

3. We now consider another representative of Barndorff-Nielsen's class of generalized hyperbolic distributions [21], the so-called *Gaussian\\inverse Gaussian* distribution (*GIG-distribution*). (Later on, Barndorff-Nielsen started using also the name 'normal inverse Gaussian distributions' [22].)

In the spirit of symbolical formula (9) we can define the *Gaussian\\inverse Gaussian* distribution of a random variable Y as follows:

$$\operatorname{Law} Y = \mathsf{E}''_{\sigma^2}\mathscr{N}(\mu+\beta\sigma^2,\sigma^2), \tag{11}$$

where we consider averaging E''_{σ^2} of the *normal* distribution $\mathscr{N}(\mu+\beta\sigma^2,\sigma^2)$ with respect to the inverse Gaussian distribution with density

$$p''_{\sigma^2}(x) = \sqrt{\frac{b}{2\pi}}\,e^{\sqrt{ab}}\frac{1}{x^{3/2}}\exp\left\{-\frac{1}{2}\left(ax+\frac{b}{x}\right)\right\}, \tag{12}$$

where $a = \alpha^2 - \beta^2$, $b = \delta^2$. (The parameters α, β, μ, and δ must satisfy conditions (5).)

It is interesting that if $W = (W_t)_{t\geqslant 0}$ is a standard Brownian motion (a Wiener process) and

$$T(t) = \inf\{s \geqslant 0\colon W_s + \sqrt{a}\, s \geqslant \sqrt{b}\, t\}$$

is the first time when the process $(W_s + \sqrt{a}\, s)_{s\geqslant 0}$ reaches the level $\sqrt{b}\, t$ then $T(1)$ has just a distribution with density (12). Hence if $B = (B_t)_{t\geqslant 0}$ is a standard Brownian motion independent of W, then Y has the same distribution as the variable

$$B_{T(1)} + \bigl(\mu + \beta T(1)\bigr). \tag{13}$$

(Cf. formula (19) in § 1c.)

Now let $g(x) = g(x; \alpha, \beta, \mu, \delta)$ be the density of a GIG-distribution.

By (11) and (12),

$$g(x) = C_3(\alpha, \beta, \mu, \delta)\left[q\Bigl(\frac{x-\mu}{\delta}\Bigr)\right]^{-1} K_1\left(\alpha\delta\, q\Bigl(\frac{x-\mu}{\delta}\Bigr)\right) e^{\beta(x-\mu)}, \tag{14}$$

where

$$C_3(\alpha, \beta, \mu, \delta) = \frac{\alpha}{\pi}\, e^{\delta\sqrt{\alpha^2-\beta^2}}$$

and $q(x) = \sqrt{1+x^2}$.

Since

$$K_1(x) \sim \sqrt{\frac{\pi}{2}}\, x^{-1/2} e^{-x} \quad \text{as} \quad x \to \infty, \tag{15}$$

it follows that

$$g(x) \sim \left(\frac{\alpha}{2\pi\delta}\right)^{1/2} \frac{1}{\bigl[1 + (\frac{x-\mu}{\delta})^2\bigr]^{3/4}} \exp\Bigl\{-\alpha\sqrt{\delta^2 + (x-\mu)^2} + \beta(x-\mu)\Bigr\} \tag{16}$$

as $|x| \to \infty$, therefore

$$\ln \frac{h_1(x)}{g(x)} \sim \frac{3}{4} \ln\left(1 + \Bigl(\frac{x-\mu}{\delta}\Bigr)^2\right), \qquad |x| \to \infty. \tag{17}$$

The last relation shows that $h_1(x)$ has 'heavier' tails than $g(x)$.

4. A hyperbolic distribution (with density $h_1(x)$) has a simpler structure than a Gaussian\\inverse Gaussian distribution (GIG-distribution) with density $g(x)$. However, one crucial feature of the latter distribution makes it advantageous in certain respects.

Let Y be a random variable with density $g(x) = g(x;\alpha,\beta,\mu,\delta)$. Then its probability generating function is

$$\mathsf{E}e^{\lambda Y} = \exp\Bigl\{\delta\Bigl[\sqrt{\alpha^2+\beta^2} - \sqrt{\alpha^2-(\beta+\lambda)^2}\Bigr] + \mu\lambda\Bigr\}. \tag{18}$$

Hence if $Y_1,\dots,Y_m$ are independent GIG-distributed variables with *the same values of* α and β, but with (generally speaking) distinct μ_i and δ_i, then the sum $Y = Y_1+\dots+Y_m$ is also a GIG-distributed variable with the same parameters α and β, but with $\mu = \mu_1+\dots+\mu_m$ and $\delta = \delta_1+\dots+\delta_m$.

In other words, the GIG-distributions are *closed* (in the above sense) with respect to convolutions.

On the other hand, if X has a hyperbolic distribution, then setting for simplicity $\beta = \mu = 0$ we obtain

$$\mathsf{E}e^{\lambda X} = \frac{\alpha}{K_1(\alpha\delta)}\,\frac{K_1(\delta\sqrt{\alpha^2-\lambda^2})}{\sqrt{\alpha^2-\lambda^2}}\,. \tag{19}$$

Hence the class of hyperbolic distributions is not closed in a way similar to GIG-distributions.

It is important to point out that both GIG-distribution and hyperbolic distribution are *infinitely divisible.* For GIG-distributions this is clearly visible from (18), while for hyperbolic distribution this is proved in [21]–[23], [25], [26]. By (18) we obtain also the following simple formulae for the expectation $\mathsf{E}Y$ and the variance $\mathsf{D}Y$:

$$\mathsf{E}Y = \mu + \frac{\delta\,\frac{\alpha}{\beta}}{\bigl[1-\bigl(\frac{\alpha}{\beta}\bigr)^2\bigr]^{1/2}}\,, \qquad \mathsf{D}Y = \frac{\delta}{\alpha\bigl[1-\bigl(\frac{\alpha}{\beta}\bigr)^2\bigr]^{3/2}}\,.$$

Remark. As regards the applications of these distributions to the analysis of financial indexes, see E. Eberlein and U. Keller [127], where one can find impressive results of the statistical processing of several financial indexes (see also Chapter II, § 2b).

5. Since hyperbolic distributions are infinitely divisible, we can define a Lévy process (i.e., a process with independent homogeneous increments) with hyperbolic distribution of increments.

We restrict ourselves to the case of symmetric centered densities $h_1(x) = h_1(x;\alpha,\beta,\mu,\delta)$ with parameters $\beta = \mu = 0$. In this case $h_1(x)$ has the following representation:

$$h_1(x) = \frac{1}{2\delta K_1(\alpha\delta)}\exp\Biggl\{-\alpha\delta\sqrt{1+\Bigl(\frac{x}{\delta}\Bigr)^2}\Biggr\}. \tag{20}$$

Let $Z = (Z_t)_{t\geqslant 0}$ be the Lévy process such that Z_1 has the distributions with density (20).

It is worth noting that since $\mathsf{E}|Z_t| < \infty$, $Z_0 = 0$, and $Z = (Z_t)$ has independent increments, this process is a *martingale* (with respect to the natural flow of σ-algebras $(\mathscr{F}_t)$, $\mathscr{F}_t = \sigma(Z_s, s \leqslant t)$), i.e.,

$$\mathsf{E}(Z_t \mid \mathscr{F}_s) = Z_s, \qquad t \geqslant s. \tag{21}$$

(Actually, $\mathsf{E}|Z_t|^p < \infty$ for all $p \geqslant 1$.)

Let $\varphi_t(\theta) = \mathsf{E}\exp(i\theta Z_t)$ be the characteristic function of the hyperbolic Lévy process $Z = (Z_t)_{t\geqslant 0}$. Then (cf. formulas (4) and (5) in § 1b)

$$\varphi_t(\theta) = \bigl(\varphi_1(\theta)\bigr)^t. \tag{22}$$

By (9) and (10) (with $X = Z_1$) we obtain

$$\operatorname{Law} Z_1 = \mathsf{E}'_{\sigma^2}\mathscr{N}(0, \sigma^2) \tag{23}$$

(because $\beta = \mu = 0$), where

$$p'_{\sigma^2}(x) = \frac{\left(\frac{\alpha}{\delta}\right)}{2K_1(\alpha\delta)} \exp\left\{-\frac{1}{2}\left(\alpha^2 x + \frac{\delta^2}{x}\right)\right\}. \tag{24}$$

Based on this representation, the following '*Lévy–Khintchine*' formula for $\varphi_t(\theta)$ was put forward in [127] (cf. formulas (22) and (29) in § 1a):

$$\varphi_t(\theta) = \exp\left\{t \int_{-\infty}^{\infty} (e^{i\theta x} - 1 - i\theta x)\, \nu(dx)\right\}, \tag{25}$$

where the Lévy measure ν has the density $p_\nu(x)$ (with respect to Lebesgue measure) with the following (fairly complicated) representation:

$$p_\nu(x) = \frac{1}{\pi^2|x|} \int_0^\infty \frac{\exp(-|x|\sqrt{2y + \alpha^2}\,)}{y\bigl(J_1^2(\delta\sqrt{2y}) + Y_1^2(\delta\sqrt{2y})\bigr)}\, dy + \frac{\exp(-|x|)}{|x|}. \tag{26}$$

Here J_1 and Y_1 are the Bessel functions of the first and second kinds, respectively.

In view of the asymptotic properties of J_1 and Y_1 (formulas 9.1.7,9 and 9.2.1,2 in [1]), one can show (see [127]) that the denominator of the integrand in (26) is asymptotically a constant as $y \to 0$ and is $y^{-1/2}$ as $y \to \infty$. Hence

$$p_\nu(x) \sim \frac{1}{x^2} \quad \text{as} \quad x \to 0, \tag{27}$$

which shows that the process $Z = (Z_t)_{t\geqslant 0}$ makes *infinitely many* small jumps on each arbitrarily small time interval.

Indeed, let

$$\mu^Z(\omega; \Delta, B) = \sum_{s\in\Delta} I\bigl(Z_s(\omega) - Z_{s-}(\omega) \in B\bigr), \qquad B \in \mathscr{B}(\mathbb{R} \setminus \{0\}),$$

be the jump measure of Z, i.e., the number of $s \in \Delta$ such that the difference $Z_s(\omega) - Z_{s-}(\omega)$ lies in the set B. Then (see, e.g., [250; Chapter II, 1.8])

$$\mathsf{E}\mu^Z(\omega; \Delta, B) = |\Delta|\, \nu(B),$$

therefore $\int_{(0,\varepsilon]} \nu(dx) = \infty$ and $\int_{[-\varepsilon,0)} \nu(dx) = \infty$ for each $\varepsilon > 0$.

2. Models with Self-Similarity. Fractality

It has long been observed in the statistical analysis of financial time series that many of these series have the property of (statistical) *self-similarity*; namely, the structure of their parts 'is the same' as the structure of the whole object. For instance, if the S_n, $n \geqslant 0$, are the *daily* values of S&P500 Index, then the empirical densities $\widehat{f}_1(x)$ and $\widehat{f}_k(x)$, $k > 1$, of the distributions of the variables

$$h_n \quad \left(= \ln \frac{S_n}{S_{n-1}}\right) \quad \text{and} \quad h_{kn} \quad \left(= \ln \frac{S_{kn}}{S_{k(n-1)}}\right), \qquad n \geqslant 1,$$

calculated for large stocks of data, satisfy the relation

$$\widehat{f}_1(x) \approx k^{\mathbb{H}} \widehat{f}_k(k^{\mathbb{H}} x),$$

where $\mathbb{H}$ is some constant (which, by the way, is significantly *larger* than 1/2—by contrast to what one would expect in accordance with the *central limit theorem*).

Of course, such properties call for explanations. As will be clear from what follows, such an explanation can be provided in the framework of the general concept of (statistical) *self-similarity*, which not only paved way to a *fractional Brownian motion*, *fractional Gaussian noise*, and other important notions, but was also crucial for the development of *fractal geometry* (B. Mandelbrot). This concept of self-similarity is intimately connected with nonprobabilistic concepts and theories, such as *chaos* or *nonlinear dynamical systems*, which (with an eye to their possible applications in financial mathematics) were discussed in Chapter II, § 4.

§ 2a. Hurst's Statistical Phenomenon of Self-Similarity

1. In 1951, a British climatologist H. E. Hurst, who spent more than 60 years in Egypt as a participant of Nile hydrology projects, published a paper [236] devoted to his discovery of the following surprising phenomenon in the fluctuations of yearly run-offs of Nile and several other rivers.

Let $x_1, x_2, \dots, x_n$ be the values of n successive yearly water run-offs (of Nile, say, in some its part). Then the value of $\frac{1}{n}X_n$, where $X_n = \sum_{k=1}^{n} x_k$, is a 'good' estimate for the expectation of the x_k.

The deviation of the cumulative value X_k corresponding to k successive years from the (empirical) mean as calculated using the data for n years is

$$X_k - \frac{k}{n}X_n,$$

and

$$\min_{k\leqslant n}\left(X_k - \frac{k}{n}X_n\right) \quad \text{and} \quad \max_{k\leqslant n}\left(X_k - \frac{k}{n}X_n\right)$$

are the smallest and the biggest deviations. Let

$$\mathcal{R}_n = \max_{k\leqslant n}\left(X_k - \frac{k}{n}X_n\right) - \min_{k\leqslant n}\left(X_k - \frac{k}{n}X_n\right)$$

be the '*range*' characterizing the amplitude of the deviation of the cumulative values X_k from their mean value $\frac{k}{n}X_n$ over n successive years.

However, Hurst did not operate with the values of the $\mathcal{R}_n$ themselves; he considered instead the normalized values $Q_n = \mathcal{R}_n/\mathcal{S}_n$, where

$$\mathcal{S}_n = \sqrt{\frac{1}{n}\sum_{k=1}^{n} x_k^2 - \left(\frac{1}{n}\sum_{k=1}^{n} x_k\right)^2}$$

is the empirical mean deviation introduced in order to make the statistics *invariant* under the change

$$x_k \to c(x_k + m), \qquad k \geqslant 1.$$

This is a desirable property because even the expectation and the variance of the x_k are usually unknown.

Based on the large volume of factual data, the records of observations of Nile flows in 622–1469 (i.e., over a period of 847 years), H. Hurst discovered the following behavior of the statistics $\mathcal{R}_n/\mathcal{S}_n$ for large n:

$$\frac{\mathcal{R}_n}{\mathcal{S}_n} \sim cn^{\mathbb{H}}, \tag{1}$$

where c is a certain constant, the equivalence '$\sim$' is interpreted in some suitable sense, and the parameter $\mathbb{H}$, which is now called the *Hurst parameter* or the *Hurst exponent*, is approximately equal to 0.7. (H. Hurst obtained close values of this parameter also for other rivers.) This was an *unexpected* result; he had anticipated that $\mathbb{H} = 0.5$ for the reason that we explain now following a later paper by W. Feller [157].

Let $x_1, x_2, \dots$ be a sequence of *independent identically distributed* random variables with $\mathsf{E}x_n = 0$ and $\mathsf{E}x_n^2 = 1$. (This is what Hurst had expected.) Then, as shown by Feller,

$$\mathsf{E}\,\mathcal{R}_n \sim \sqrt{\frac{\pi}{2}}\,n^{1/2} \quad \text{and} \quad \mathsf{D}\,\mathcal{R}_n \sim \Big(\frac{\pi^2}{6} - \frac{\pi}{2}\Big)n$$

for large n. Since, moreover, $\mathcal{S}_n \to 1$ (with probability one) in this case, the values of Q_n must increase (on the average, at any rate) as $n^{1/2}$ for n large.

In a study of the statistical properties of the sequence $(x_n)_{n\geqslant 1}$ it makes sense to ask about the structure of the empirical distribution function $\widehat{\mathrm{Law}}(x_1 + \dots + x_n)$ calculated from a (large) number of samples, say, $(x_1, \dots, x_n)$, $(x_{n+1}, \dots, x_{2n}), \dots$. In the case when the x_k are the deviations of the water level from some 'mean' value, one finds out (e.g., for Nile again) that

$$\widehat{\mathrm{Law}}(x_1 + \dots + x_n) \approx \widehat{\mathrm{Law}}(n^{\mathbb{H}} x_1), \tag{2}$$

where $\mathbb{H} > 1/2$.

2. How and thanks to what probabilistic and statistical properties of the sequence (x_n) in (2) can the parameter $\mathbb{H}$ be distinct from 0.5?

Looking at formula (4) in § 1a we see one of the possible explanations for the relation $\mathbb{H} \sim 0.7$: the x_j can be *independent stable* random variables with stability exponent $\alpha = \frac{1}{\mathbb{H}} \sim 1.48$.

There exists another possible explanation: relation (2) with $\mathbb{H} \neq 1/2$ can occur even in the case of normally distributed, but *dependent* variables $x_1, x_2, \dots$! In that case the stationary sequence (x_n) is necessarily a sequence with *strong aftereffect* (see § 2c below.)

3. Properties (1) and (2), which can be regarded as a peculiar form of self-similarity, can be also observed for many financial indexes (with the h_n in place of the x_n). Unsurprisingly, the above observation that the x_n can be 'independent and stable' or 'dependent and normal' has found numerous applications in financial mathematics, and in particular, in the analysis of the 'fractal' structure of 'volatility'.

Hurst's results and the above observations served as the starting point for B. Mandelbrot, who suggested that, in the *Hurst model* (considered by Mandelbrot himself) and in many other probabilistic models (e.g., in financial mathematics), one could use strictly stable processes (§ 1c) and *fractional Brownian motions* (§ 2c), which have the property of self-similarity.

It should be pointed out that a variety of real-life systems with *nonlinear dynamics* (occurring in physics, geophysics, biology, economics, ...) are featured by self-similarity of kinds (1) and (2). It is this property of self-similarity that occupies the central place in *fractal geometry*, the founder of which, B. Mandelbrot, has chosen the title "*The Fractal Geometry of Nature*" for his book [320], so as to emphasize the universal character of self-similarity.

We present the necessary definitions of statistical self-similarity and fractional Brownian motion in § 2c. The aim of the next section, which has no direct relation to financial mathematics and was inspired by [104], [379], [385], [386], [428], [456], and books by some other authors, must give one a *general* notion of self-similarity.

§ 2b. A Digression on Fractal Geometry

1. It is well known that the emergence of the Euclidean geometry in Ancient Greece was the result of an attempt to reduce the variety of natural forms to several 'simple', 'pure', 'symmetric' objects. That gave rise to *points*, *lines*, *planes*, and most simple three-dimensional objects (*spheres*, *cones*, *cylinders*, ...).

However, as B. Mandelbrot has observed (1984), "*clouds are not spheres, mountaines are not cones, coastlines are not circles, and bark is not smooth, nor does lightning travel in a straight line* ..." Mandelbrot has developed the so-called *fractal geometry* with the precise aim to describe objects, forms, phenomena that were far from 'simple' and 'symmetric'. On the contrary, they could have a rather complex structure, but exhibited at the same time some properties of self-similarity, self-reproducibility.

We have no intention of giving a formal definition of *fractal geometry* or of a *fractal*, its central concept. Instead, we wish to call the reader's attention to the importance of *the idea of fractality in general, and in financial mathematics in particular* and give for that reason only an illustrative description of this subject.

The following 'working definition' is very common: "*A fractal is an object whose portions have the same structure as the whole.*" (The word '*fractal*' introduced by Mandelbrot presumably in 1975 [315] is a derivative of the Latin verb *fractio* meaning '*fracture, break up*' [296]).

A classical example of a three-dimensional object of *fractal* structure is a tree. The branches with their twigs ('portions') are similar to the 'whole', the main trunk with branches.

Another graphic example is the *Sierpiński gasket*[a] (see Fig. 27) that can be obtained from a solid ('black') triangle by removing the interior of the central triangle, and by removing subsequently the interiors of the central triangles from each of the resulting 'black' triangles.

The resulting 'black' sets converge to a set called the *Sierpiński gasket*. This limit set is an example of so-called *attractors*.

---

[a] W. Sierpiński (1882–1969), a Polish mathematician, invented the 'Sierpiński gasket' and the 'Sierpiński carpet' as long ago as 1916.

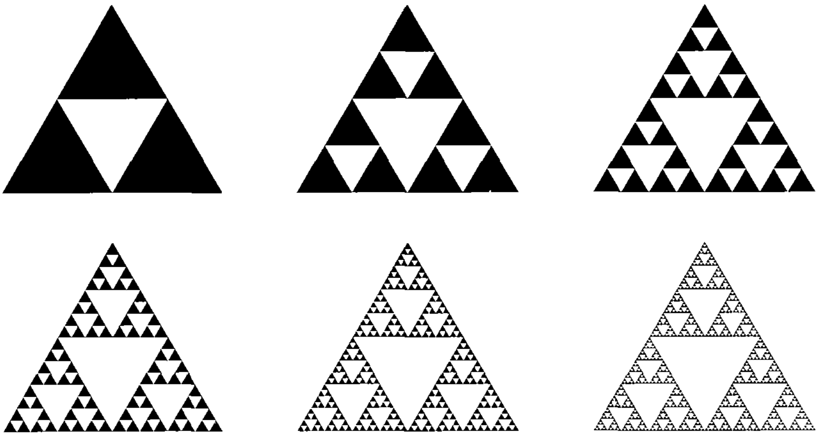


FIGURE 27. Consequtive steps of the construction of the *Sierpiński gasket*

Obviously, there are many 'cavities' in the Sierpiński gasket. Hence there arises a natural question on the 'dimension' of this object.

Strictly speaking, it is not two-dimensional due to these cavities. At the same time, it is certainly not one-dimensional. Presumably, one can assign some fractal (i.e., fractional) dimension to the Sierpiński gasket. (Indeed, taking a suitable definition one can see that this 'dimension' is $1.58\ldots$; see, e.g. [104], [428], or [456].)

The Sierpiński gasket is an example of a *symmetric* fractal object. In the nature, of course, it is 'asymmetric fractals' that dominate. The reason lies in the *local randomness* of their development. This, however, does not rule out its *determinism* on the global level; for a support we refer to the construction of the Sierpiński gasket by means of the following stochastic procedure [386].

We consider the equilateral triangle with vertices $A = A(1,2)$, $B = B(3,4)$, and $C = C(5,6)$ in the plane (our use of the integers $1, 2, \ldots, 6$ will be clear from the construction that follows).

In this triangle, we choose an arbitrary point a and then roll a 'fair' die with faces marked by the integers $1, 2, \ldots, 6$. If the number thrown is, e.g., '5' or '6', then we join with the vertex $C = C(5,6)$ and let be the middle of the joining segment. We roll the dice again and, depending on the result, obtain another point, and so on.

Remarkably, the limit set of the points so obtained is ('almost always') the Sierpiński gasket ('black' points in Fig. 27).

Another classical, well known to mathematicians example of a set with fractal structure is the *Cantor set* introduced by G. Cantor (1845–1918) in 1883 as an example of a set of a special structure (this is a nowhere dense *perfect* set, i.e., a closed set without isolated points, which has the cardinal number of the continuum).

We recall that this is the subset of the closed interval $[0,1]$ consisting of the numbers representable as $\sum_{i=1}^{\infty} \frac{\varepsilon_i}{3^i}$ with $\varepsilon_i = 0$ or 2. Geometrically, the Cantor set can be obtained from $[0,1]$ by deleting first the central (open) interval $\left(\frac{1}{3}, \frac{2}{3}\right)$, then the central subintervals $\left(\frac{1}{9}, \frac{2}{9}\right)$ and $\left(\frac{7}{9}, \frac{8}{9}\right)$ of the resulting intervals $\left[0, \frac{1}{3}\right]$ and $\left[\frac{2}{3}, 1\right]$, and so on, ad infinitum. The total length of the deleted intervals is 1; nevertheless, the remaining 'sparse' set has the cardinal number of the continuum. The self-similarity of the Cantor set (i.e., the fact that 'its portions have the same structure as the whole set') is clear from our geometric construction: this set is the union of subsets looking each like a reduced copy of the whole object.

Among other well-known objects with property of self-similarity we mention the '*Pascal triangle*' (after B. Pascal), the '*Koch snowflakes*' (after H. van Koch), the *Peano curve* (G. Peano), and the *Julia sets* (G. Julia); see, e.g., [379].

2. In our above discussion of 'dimension' we gave no precise definition (F. Hausdorff, who invented the *Hausdorff dimension*, pointed out that the problem of an adequate definition of 'dimension' is a very difficult one). Referring the reader to literature devoted to this subject (e.g., [104], [428], [456]), we note only that the notion of a 'fractal dimension' of, say, a plane curve is fairly transparent: it must show how the curve *sweeps* the plane. If the curve in question is a *realization* $X = (X_t)_{t\geqslant 0}$ of some process, then its fractal dimension increases with the proportion of 'high-frequency' components in $(X_t)_{t\geqslant 0}$.

3. Let now $X = (X_t)_{t\geqslant 0}$ be a *stochastic* (*random*) process. In this case, it is reasonable to define the 'fractal dimension' for all the totality of realizations, rather than for separate realizations. This brings us to the concept of *statistical fractal dimension*, which we introduce in the next section.

§ 2c. Statistical Self-Similarity. Fractional Brownian Motion

1. DEFINITION 1. We say that a random process $X = (X_t)_{t\geqslant 0}$ with state space $\mathbb{R}^d$ is *self-similar* or satisfies the *property of* (*statistical*) *self-similarity* if for each $a > 0$ there exists $b > 0$ such that

$$\operatorname{Law}(X_{at},\, t \geqslant 0) = \operatorname{Law}(bX_t,\, t \geqslant 0). \tag{1}$$

In other words *changes of the time scale* $(t \to at)$ *produce the same results as changes of the phase scale* $(x \to bx)$.

We saw in § 1c that for (nonzero) *strictly stable* processes there exists a constant $\mathbb{H}$ such that $b = a^{\mathbb{H}}$. In addition, for strictly α-stable processes we have

$$\mathbb{H} = \frac{1}{\alpha}. \tag{2}$$

In the case of (general) *stable* processes, in place of (1), we have the property

$$\operatorname{Law}(X_{at},\, t \geqslant 0) = \operatorname{Law}(a^{\mathbb{H}} X_t + tD_a,\, t \geqslant 0) \tag{3}$$

(see formula (2) in § 1c), which means that, for these processes, a change of the time scale produces the same results as a change of the phase scale and a subsequent '*translation*' defined by the vector tD_a, $t \geqslant 0$. Moreover, $\mathbb{H} = 1/\alpha$ for α-stable processes.

2. It follows from the above, that it would be reasonable to introduce the following definition.

DEFINITION 2. If $b = a^{\mathbb{H}}$ in Definition 1 for each $a > 0$, then we call $X = (X_t)_{t \geqslant 0}$ a *self-similar process* with *Hurst exponent* $\mathbb{H}$ or we say that this process has the *property of statistical self-similarity* with *Hurst exponent* $\mathbb{H}$. The quantity $\mathbb{D} = \frac{1}{\mathbb{H}}$ is called the *statistical fractal dimension* of X.

A classical example of a self-similar process is a *Brownian motion* $X = (X_t)_{t \geqslant 0}$. We recall that for this (Gaussian) process we have $\mathsf{E}X_t = 0$ and $\mathsf{E}X_s X_t = \min(s,t)$. Hence

$$\mathsf{E}X_{as}X_{at} = \min(as, at) = a\min(s,t) = \mathsf{E}(a^{1/2}X_s)(a^{1/2}X_t),$$

so that the two-dimensional distributions $\operatorname{Law}(X_s, X_t)$ have the property

$$\operatorname{Law}\bigl(X_{as}, X_{at}\bigr) = \operatorname{Law}\bigl(a^{1/2}X_s, a^{1/2}X_t\bigr).$$

By the Gaussian property it follows that a Brownian motion has the property of statistical self-similarity with *Hurst exponent* $\mathbb{H} = 1/2$.

Another example is a *strictly α-stable Lévy motion* $X = (X_t)_{t \geqslant 0}$, which satisfies the relation

$$X_t - X_s \sim S_\alpha\bigl((t-s)^{1/2}, 0, 0\bigr), \qquad \alpha \in (0,2].$$

For this process with homogeneous independent increments we have

$$X_{at} - X_{as} \overset{\mathrm{d}}{=} a^{1/\alpha}(X_t - X_s),$$

so that the corresponding Hurst exponent $\mathbb{H}$ is $1/\alpha$ and $\mathbb{D} = \alpha$. For $\alpha = 2$ we obtain a Brownian motion.

We emphasize that the processes in both examples have *independent* increments.

The next example relates to the case of processes with *dependent increments*.

3. Fractional Brownian motion. We consider the function

$$A(s,t) = |s|^{2\mathbb{H}} + |t|^{2\mathbb{H}} - |t-s|^{2\mathbb{H}}, \qquad s,t \in \mathbb{R}. \tag{4}$$

For $0 < \mathbb{H} \leqslant 1$ this function is *nonnegative definite* (see, e.g., [439; Chapter II, § 9]), therefore there exists a *Gaussian* process on some probability space (e.g., on the space of real functions $\omega = (\omega_t)$, $t \in \mathbb{R}$) that has the zero mean and the autocovariance function

$$\mathsf{Cov}(X_s, X_t) = \frac{1}{2} A(s,t),$$

i.e., a process such that

$$\mathsf{E} X_s X_t = \frac{1}{2}\{|s|^{2\mathbb{H}} + |t|^{2\mathbb{H}} - |t-s|^{2\mathbb{H}}\}. \tag{5}$$

Hence

$$\mathsf{E} X_{as} X_{at} = a^{2\mathbb{H}} \mathsf{E} X_s X_t = \mathsf{E}(a^{\mathbb{H}} X_s)(a^{\mathbb{H}} X_t),$$

so that

$$\mathrm{Law}(X_{as}, X_{at}) = \mathrm{Law}(a^{\mathbb{H}} X_s, a^{\mathbb{H}} X_t).$$

As in the case of a Brownian motion, whose distribution is completely determined by the two-dimensional distributions, we conclude that X is a self-similar process with Hurst exponent $\mathbb{H}$.

By (5),

$$\mathsf{E}|X_t - X_s|^2 = |t-s|^{2\mathbb{H}}. \tag{6}$$

We recall that, in accordance with the *Kolmogorov test* ([470]), *a random process $X = (X_t)_{t \geqslant 0}$ has a continuous modification if there exist constants $\alpha > 0$, $\beta > 0$, and $c > 0$ such that*

$$\mathsf{E}|X_t - X_s|^{\alpha} \leqslant c\,|t-s|^{1+\beta} \tag{7}$$

for all $s, t \geqslant 0$. Hence if $\mathbb{H} > 1/2$, then it immediately follows from (6) (regarded as (7) with $\alpha = 2$ and $\beta = 2\mathbb{H} - 1$) that the process $X = (X_t)_{t \geqslant 0}$ under consideration has a continuous modification. Further, if $0 < \mathbb{H} \leqslant 1/2$, then by the Gaussian property, for each $0 < k < \mathbb{H}$ we have

$$\mathsf{E}|X_t - X_s|^{1/k} \leqslant c\,|t-s|^{\mathbb{H}/k}$$

with some constant $c > 0$. Hence we can apply the Kolmogorov test again (with $\alpha = 1/k$ and $\beta = \mathbb{H}/k - 1$).

Thus, our Gaussian process $X = (X_t)_{t \geqslant 0}$ has a *continuous* modification for all $\mathbb{H}$, $0 < \mathbb{H} \leqslant 1$.

DEFINITION 3. We call a continuous Gaussian process $X = (X_t)_{t \geqslant 0}$ with zero mean and the covariance function (5) a (standard) *fractional Brownian motion with Hurst self-similarity exponent* $0 < \mathbb{H} \leqslant 1$. (We shall often denote such a process by $B_{\mathbb{H}} = (B_{\mathbb{H}}(t))_{t \geqslant 0}$ in what follows.)

By this definition, a (standard) fractional Brownian motion $X = (X_t)_{t \geqslant 0}$ has the following properties (which could also be taken as a basis of its definition):

1) $X_0 = 0$ and $\mathsf{E} X_t = 0$ for all $t \geqslant 0$;

2) X has homogeneous increments, i.e.,

$$\mathrm{Law}(X_{t+s} - X_s) = \mathrm{Law}(X_t), \qquad s, t \geqslant 0;$$

3) X is a Gaussian process and

$$\mathsf{E}X_t^2 = |t|^{2\mathbb{H}}, \qquad t \geqslant 0,$$

where $0 < \mathbb{H} \leqslant 1$;

4) X has continuous trajectories.

These properties show again that a fractional Brownian motion has the self-similarity property.

It is worth pointing out that a converse result is also true in a certain sense ([418; pp. 318–319]): *if a nondegenerate process* $X = (X_t)_{t\geqslant 0}$, $X_0 = 0$, *has finite variance, homogeneous increments, and is a self-similar process with Hurst exponent* $\mathbb{H}$, *then* $0 < \mathbb{H} \leqslant 1$ *and the autocovariance function of this process satisfies the equality* $\mathrm{Cov}(X_s, X_t) = \mathsf{E}X_1^2 A(s,t)$, *where* $A(s,t)$ *is defined by* (4). Moreover, if $0 < \mathbb{H} < 1$, then the expectation $\mathsf{E}X_t$ is zero, and if $\mathbb{H} = 1$, then $X_t \overset{\mathrm{d}}{=} tX_1$ (P-a.s.).

We note also that, besides Gaussian processes, one knows also of some *nonGaussian processes* with these properties (see [418; p. 320]).

4. If $\mathbb{H} = 1/2$, then a (standard) fractional Brownian motion is precisely a (standard) Brownian motion (a Wiener process).

The processes $B_{\mathbb{H}}$ so introduced were first considered by A. N. Kolmogorov in [278] (1940), where they were called *Wiener helices*. The name '*fractional Brownian motion*' was introduced in 1968, by B. Mandelbrot and J. van Ness [328]. By contrast to Kolmogorov, who had constructed the process $B_{\mathbb{H}}$ starting from the covariance function (4), Mandelbrot and van Ness used an 'explicit' representation by means of stochastic integrals with respect to a (certain) Wiener process $W = (W_t)_{t\in\mathbb{R}}$ with $W_0 = 0$: for $0 < \mathbb{H} < 1$ they set

$$B_{\mathbb{H}}(t) = c_{\mathbb{H}} \left\{ \int_{-\infty}^{0} \left[(t-s)^{\mathbb{H}-1/2} - (-s)^{\mathbb{H}-1/2} \right] dW_s + \int_0^t (t-s)^{\mathbb{H}-1/2}\, dW_s \right\}, \tag{8}$$

where the (normalizing) constant

$$c_{\mathbb{H}} = \sqrt{\frac{2\mathbb{H}\,\Gamma(\frac{3}{2} - \mathbb{H})}{\Gamma(\frac{1}{2} + \mathbb{H})\Gamma(2 - 2\mathbb{H})}} \tag{9}$$

is such that $\mathsf{E}B_{\mathbb{H}}^2(1) = 1$.

Remark 1. As regards various representations for the right-hand side of (8), see [328]. In this paper one can also find a comprehensive discourse on the origins of the term *fractional Brownian motion* and its relation to the *fractional integral*

$$\int_0^t (t-s)^{\mathbb{H}-1/2}\, dW_s \tag{10}$$

of Holmgren–Riemann–Liouville, where $\mathbb{H}$ can be an *arbitrary* positive number (see also Weyl [475]).

Remark 2. Almost surely, the trajectories of a fractional Brownian motion $B_{\mathbb{H}}$, $0 < \mathbb{H} \leqslant 1$, satisfy the Hölder condition with exponent $\beta < \mathbb{H}$. They are incidentally nowhere differentiable and

$$\varlimsup_{t\to t_0} \left| \frac{B_{\mathbb{H}}(t) - B_{\mathbb{H}}(t_0)}{t - t_0} \right| = \infty \quad (\textsf{P}\text{-a.s.})$$

for each $t_0 \geqslant 0$. (The corresponding proof can be carried out in the same way as for the standard Brownian motion, i.e., for $\mathbb{H} = 1/2$; see, e.g., [123]).

Remark 3. Considering a Hölder function $\mathbb{H}_t$ ($|\mathbb{H}_t - \mathbb{H}_s| \leqslant c|t-s|^\alpha$, $\alpha > 0$) with values in $(0,1)$ in place of $\mathbb{H}$ in (8) we obtain a random process that is called a *multifractional Brownian motion*. In was introduced and thoroughly investigated in [381].

Remark 4. In the theory of stochastic processes one assigns a major role to *semimartingales*, a class for which there exists well-developed stochastic calculus (see § 5 and, for more detail, [250] and [304]). It is worth noting in this connection that a fractional Brownian motion $B_{\mathbb{H}}$, $0 < \mathbb{H} \leqslant 1$, *is not a semimartingale* (except for the case of $\mathbb{H} = 1/2$, i.e., the case of a Brownian motion, and $\mathbb{H} = 1$). See [304; Chapter 4, § 9, Example 2] for the corresponding proof for the case of $1/2 < \mathbb{H} < 1$.

Remark 5. In connection with *Hurst's $\mathcal{R}/\mathcal{S}$-analysis* (i.e., analysis based on the study of the properties of the range, empirical standard deviation, and their ratio), it could be instructive to note following [328] that if $X = (X_t)_{t\geqslant 0}$ is a continuous self-similar process with Hurst parameter $\mathbb{H}$, $X_0 = 0$, and $\widetilde{\mathcal{R}}_t = \sup\limits_{0\leqslant s\leqslant t} X_s - \inf\limits_{0\leqslant s\leqslant t} X_s$, then $\operatorname{Law}(\widetilde{\mathcal{R}}_t) = \operatorname{Law}(t^{\mathbb{H}}\widetilde{\mathcal{R}}_1)$, $t > 0$. In the case of a Brownian motion ($\mathbb{H} = 1/2$) W. Feller found a precise formula of the distribution of $\widetilde{\mathcal{R}}_1$. (Its density is $8 \sum\limits_{k=1}^{\infty} (-1)^{k-1} k^2 \varphi(kx)$, $x \geqslant 0$, where $\varphi(x) = (2\pi)^{-1/2} \exp(-x^2/2)$.)

5. There exist various generalization of self-similarity property (1). For example, let $X(\alpha) = (X_t(\alpha))_{t\geqslant 0}$ be a process of Ornstein–Uhlenbeck type with parameter $\alpha \in \mathbb{R}$, i.e., a Gaussian Markov process defined by the formula

$$X_t(\alpha) = \int_0^t e^{\alpha(t-s)}\, dW_s, \qquad t \geqslant 0, \tag{11}$$

where $W = (W_s)_{s\geqslant 0}$ is a standard Brownian motion; see § 3a. (Note that $X(\alpha) = (X_t(\alpha))_{t\geqslant 0}$ is the solution of the stochastic linear differential equation $dX_t(\alpha) = \alpha X_t(\alpha)\, dt + dW_t$, $X_0(\alpha) = 0$.) From (11) it easily follows that

$$\operatorname{Law}(X_{at}(\alpha),\ t \in \mathbb{R}) = \operatorname{Law}(a^{1/2} X_t(\alpha a),\ t \in \mathbb{R})$$

for each $a \in \mathbb{R}$, which can be regarded as a self-similarity (of a sort) of the *family* of processes $\{X(\alpha),\ \alpha \in \mathbb{R}\}$.

6. We now dwell upon one crucially important method of *statistical inference* that is based on self-similarity properties.

Let $X = (X_t)_{t \geqslant 0}$ be a self-similar process with self-similarity exponent Assume that $\Delta > 0$. Then

$$\operatorname{Law}(X_\Delta) = \operatorname{Law}(\Delta^{\mathbb{H}} X_1), \tag{12}$$

so that if $f_\Delta(x) = \dfrac{d\mathsf{P}(X_\Delta \leqslant x)}{dx}$ is the density of the probability distribution of X_Δ, then

$$f_1(x) = \Delta^{\mathbb{H}} f_\Delta(x\Delta^{\mathbb{H}}). \tag{13}$$

It is common in the usual statistical analysis that, based on considerations of the general nature, one can assume X to be self-similar with some, *unknown* in general, value of the parameter $\mathbb{H}$. Assume that we have managed to find a 'likely' estimate $\widehat{\mathbb{H}}$ for $\mathbb{H}$. Then the verification of the conjecture that X is indeed a self-similarity process with parameter $\widehat{\mathbb{H}}$ can be carried out as follows.

Assume that, based on independent observations of the values of X_1 and X_Δ, we can construct empirical densities $\widehat{f}_1(x)$ and $\widehat{f}_\Delta(x)$ for sufficiently many values of Δ. Then if

$$\widehat{f}_1(x) \approx \Delta^{\widehat{\mathbb{H}}} \widehat{f}_\Delta(x\Delta^{\mathbb{H}}) \tag{14}$$

for a wide range of values of x and Δ, then we have a fairly solid argument in favor of the conjecture that X is a self-similar process with exponent $\widehat{\mathbb{H}}$.

Certainly, if we know the theoretical density $f_1(x)$ (which, of course, depends on $\mathbb{H}$), then instead of verifying (14) we should look if the graph of $\Delta^{\widehat{\mathbb{H}}} \widehat{f}_\Delta(x\Delta^{\widehat{\mathbb{H}}})$ is close to that of $f_1(x)$ plotted for the Hurst exponent $\widehat{\mathbb{H}}$ (or for the true value of this parameter if it is known a priori).

7. For α-stable Lévy processes the Hurst exponent $\mathbb{H}$ is equal to $1/\alpha$, therefore the estimation of $\mathbb{H}$ reduces to the estimation of α. For a fractional Brownian motion $B_{\mathbb{H}} = (B_{\mathbb{H}}(t))_{t \geqslant 0}$ we can estimate $\mathbb{H}$ from the results of *discrete* observations, for instance, as follows.

We consider the time interval $[0,1]$ and partition it into n equal parts of length $\Delta = 1/n$. Let

$$\nu_1(B_{\mathbb{H}}; \Delta) = \frac{\sum_{k=1}^n |B_{\mathbb{H}}(k\Delta) - B_{\mathbb{H}}((k-1)\Delta)|}{n}$$

(cf. formula (19) in Chapter IV, § 3a).

Since

$$\mathsf{E}|B_{\mathbb{H}}(t+s) - B_{\mathbb{H}}(t)| = \sqrt{\frac{2}{\pi}}\, s^{\mathbb{H}},$$

it follows that

$$\mathsf{E}\,\nu_1(B_{\mathbb{H}}; \Delta) = \sqrt{\frac{2}{\pi}}\, \Delta^{\mathbb{H}}.$$

Hence we infer the natural conclusion: we should take the statistics

$$\widehat{\mathbb{H}}_n = \frac{\log\bigl[\sqrt{\pi/2}\,\nu_1(B_{\mathbb{H}};\Delta)\bigr]}{\log\Delta}$$

as an estimator for $\mathbb{H}$. (It was shown in [380] that $\widehat{\mathbb{H}}_n \to \mathbb{H}$ with probability one.)

§ 2d. Fractional Gaussian Noise: a Process with Strong Aftereffect

1. In many domains of applied probability theory a Brownian motion $B = (B_t)_{t\geqslant 0}$ is regarded as an easy way to obtain *white noise*.

Setting

$$\beta_n = B_n - B_{n-1}, \qquad n \geqslant 1, \tag{1}$$

we obtain a sequence $\beta = (\beta_n)_{n\geqslant 1}$ of independent identically distributed random Gaussian variables with $\mathsf{E}\beta_n = 0$ and $\mathsf{E}\beta_n^2 = 1$. In accordance with Chapter II, § 2a, we call such a sequence *white* (*Gaussian*) *noise*; we have used it as a *source of randomness* in the construction of various random processes, both linear (MA, AR, $ARMA$, ...), and nonlinear ($ARCH$, $GARCH$, ...).

In a similar way, a fractional Brownian motion $B_{\mathbb{H}}$ is useful for the construction of both *stationary Gaussian sequences with strong aftereffect* (systems with *long memory, persistent systems*), and sequences with *intermittency, antipersistence* (*relaxation processes*).

By analogy with (1) we now set

$$\beta_n = B_{\mathbb{H}}(n) - B_{\mathbb{H}}(n-1), \qquad n \geqslant 1, \tag{2}$$

and we shall call the sequence $\beta = (\beta_n)_{n\geqslant 1}$ *fractional* (*Gaussian*) *noise with Hurst parameter* $\mathbb{H}$, $0 < \mathbb{H} < 1$.

By formula (5) in § 2c for the covariance function of a (standard) process $B_{\mathbb{H}}$ we obtain that the covariance function $\rho_{\mathbb{H}}(n) = \operatorname{Cov}(\beta_k, \beta_{k+n})$ is as follows:

$$\rho_{\mathbb{H}}(n) = \frac{1}{2}\bigl\{|n+1|^{2\mathbb{H}} - 2|n|^{2\mathbb{H}} + |n-1|^{2\mathbb{H}}\bigr\}. \tag{3}$$

Hence

$$\rho_{\mathbb{H}}(n) \sim \mathbb{H}(2\mathbb{H}-1)|n|^{2\mathbb{H}-2} \tag{4}$$

as $n \to \infty$.

Thus, if $\mathbb{H} = 1/2$, then $\rho_{\mathbb{H}}(n) = 0$ for $n \neq 0$, and $(\beta_n)_{n\geqslant 1}$ is (as already mentioned) a Gaussian sequence of independent random variables. On the other hand if $\mathbb{H} \neq 1/2$, then we see from (4) that the covariance decreases fairly slowly (as $|n|^{-(2-2\mathbb{H})}$) with the increase of n, which is usually interpreted as a 'long memory' or a 'strong aftereffect'.

We should note a crucial difference between the cases of $0 < \mathbb{H} < 1/2$ and $1/2 < \mathbb{H} < 1$.

If $0 < \mathbb{H} < 1/2$, then the covariance is *negative*: $\rho_{\mathbb{H}}(n) < 0$ for $n \neq 0$, moreover, $\sum\limits_{n=0}^{\infty} |\rho_{\mathbb{H}}(n)| < \infty$.

On the other hand, if $1/2 < \mathbb{H} < 1$, then the covariance is *positive*: $\rho_{\mathbb{H}}(n) > 0$ for $n \neq 0$, and $\sum\limits_{n=0}^{\infty} \rho_{\mathbb{H}}(n) = \infty$.

A positive covariance means that positive (negative) values of β_n are usually followed also by positive (respectively, negative) values, so that fractional Gaussian noise with $1/2 < \mathbb{H} < 1$ can serve a suitable model in the description of the 'cluster' phenomena (Chapter IV, § 3e), which one observes in practice, in the empirical analysis of the returns $h_n = \ln \dfrac{S_n}{S_{n-1}}$ for many financial indexes $S = (S_n)$.

On the other hand, a negative covariance means that positive (negative) values are usually followed by negative (respectively, positive) ones. Such *strong intermittency* ('up and down and up ... ') is indeed revealed by the analysis of the behavior of volatilities (see § § 3 and 4 in Chapter IV).

2. The sequence $\beta = (\beta_n)$ is a Gaussian *stationary* sequence with correlation function $\rho_{\mathbb{H}}(n)$ defined by (3).

By a direct verification one can see that the spectral density $f_{\mathbb{H}}(\lambda)$ of the spectral representation

$$\rho_{\mathbb{H}}(n) = \int_{-\pi}^{\pi} e^{i\lambda n} f_{\mathbb{H}}(\lambda)\, d\lambda \tag{5}$$

can be expressed as follows:

$$f_{\mathbb{H}}(\lambda) = \frac{\displaystyle\int_0^{\infty} (\cos x\lambda) \left(\sin^2 \frac{x}{2}\right) x^{-2\mathbb{H}-1}\, dx}{\displaystyle\int_0^{\infty} \left(\sin^2 \frac{x}{2}\right) x^{-2\mathbb{H}-1}\, dx}. \tag{6}$$

Calculating the corresponding integrals we obtain (see [418] for detail) that

$$f_{\mathbb{H}}(\lambda) = K(\mathbb{H}) |e^{i\lambda} - 1|^2 \sum_{k=-\infty}^{\infty} \frac{1}{|\lambda + 2\pi k|^{2\mathbb{H}+1}}, \qquad |\lambda| \leqslant \pi, \tag{7}$$

where $K(\mathbb{H}) = \left(\dfrac{\pi}{\mathbb{H}\Gamma(2\mathbb{H}) \sin(\mathbb{H}\pi)}\right)^{-1}$

3. The case $\mathbb{H} = 1/2$. With an eye to the applications of fractional Brownian motions to the description of the dynamics of financial indexes, we now choose a unit of time $n = 0, \pm 1, \pm 2, \ldots$ and set

$$H_n = B_{\mathbb{H}}(n), \qquad h_n = H_n - H_{n-1}.$$

Clearly, $\mathsf{E}h_n = 0$ and $\mathsf{D}h_n^2 = 1$.

The corresponding Gaussian sequence $h = (h_n)$ of independent identically distributed variables is, as already pointed out in this section, *white* (*Gaussian*) *noise*.

4. The case of $1/2 < \mathbb{H} < 1$. The corresponding noise $h = (h_n)$ is often said to be 'black'. It is featured by a *strong aftereffect*, *long memory* (a *persistent* system).

Phenomena of this kind are encountered in the behavior of the levels of rivers, or the character of solar activity, the widths of consecutive annual rings of trees, and, finally (which is the most interesting from our present standpoint), in the values of the returns $h_n = \ln \dfrac{S_n}{S_{n-1}}$, $n \geqslant 1$, for stock prices, currency cross rates, and other financial indexes (see Chapter IV).

If $\mathbb{H} = 1/2$, then the standard deviation $\sqrt{\mathsf{D}(h_1 + \cdots + h_n)}$ increases with n as $\sqrt{n}$, while if $\mathbb{H} > 1/2$, then the growth is faster, of order $n^{\mathbb{H}}$. To put it otherwise, the dispersion of the values of the resulting variable $H_n = h_1 + \cdots + h_n$ is larger than for *white noise* ($\mathbb{H} = 1/2$).

It is instructive to note that if $\mathbb{H} = 1$ then we may take $B_{\mathbb{H}}(t) = tB_{\mathbb{H}}(1)$ (P-a.s.) for a fractional Brownian motion. Hence the sequence $h = (h_n)$ of increments $h_n = B_{\mathbb{H}}(n) - B_{\mathbb{H}}(n-1)$, $n \geqslant 1$, is trivial in this case: we always have $h_n \equiv B_{\mathbb{H}}(1)$, $n \geqslant 1$, which one could call the 'perfect persistence'.

5. The case of $0 < \mathbb{H} < 1/2$. Typical examples of systems with such values of the Hurst parameter $\mathbb{H}$ are provided by turbulence. The famous Kolmogorov's *Law of* 2/3 ([276], 1941) says that in the case of an incompressible viscous fluid with very large Reynolds number the mean square of the difference of the velocities at two points lying at a distance r that is neither too large nor too small is proportional to $r^{2\mathbb{H}}$, where $\mathbb{H} = 1/3$.

Fractional noise $h = (h_n)$ with $0 < \mathbb{H} < 1/2$ ('*pink noise*') has a *negative covariance*, which, as already mentioned, corresponds to *fast alternation* of the values of the h_n. This is also characteristic of *turbulence* phenomena, which (coupled with self-similarity) indicates that a fractional Brownian motion with $0 < \mathbb{H} < 1/2$ can serve a fair model in a description of turbulence.

An example of 'financial turbulence', with Hurst parameter $0 < \mathbb{H} < 1/2$, is provided by the sequence $\widehat{r} = (\widehat{r}_n)$ with $\widehat{r}_n = \ln \dfrac{\widehat{\sigma}_n}{\widehat{\sigma}_{n-1}}$, where

$$\widehat{\sigma}_n^2 = \frac{1}{n-1} \sum_{k=1}^{n} (h_k - \overline{h}_n)^2$$

is the empirical variance (volatility) of the sequence of logarithmic returns $h = (h_n)$, $h_n = \ln \dfrac{S_n}{S_{n-1}}$, calculated for stock prices, DJIA, S&P500 Index, etc. (see Chapter IV, § 3a).

Many authors (see, e.g., [385], [386], and [180]) see big similarities between hydrodynamic turbulence and the behavior of prices on financial markets. This analogy brings, e.g., the authors of [180] to the conclusion that "*in any case, we have reason to believe that the qualitative picture of turbulence that has developed during the past* 70 *years will help our understanding of the apparently remote field of financial markets*".

3. Models Based on a Brownian Motion

§ 3a. Brownian Motion and its Role of a Basic Process

1. In a constructive definition of random sequences $h = (h_n)$ describing the dynamics of the 'returns' $h_n = \ln \dfrac{S_n}{S_{n-1}}$ (corresponding, say, to some stock of price S_n at time n), one usually assumes, both in linear and nonlinear models, that we have some *basic* sequence $\varepsilon = (\varepsilon_n)$, which is the 'carrier' of randomness and which generates $h = (h_n)$. Usually, $\varepsilon = (\varepsilon_n)$ is assumed to be *white* (*Gaussian*) *noise*.

The choice of such a sequence $\varepsilon = (\varepsilon_n)$ as a basis reflects the natural wish for building 'complex' objects (such as, in general, the variables h_n) from 'simple' bricks.

The sequence $\varepsilon = (\varepsilon_n)$ can indeed be considered 'simple', for it consists of *independent identically distributed random variables with classical normal* (*Gaussian*) *distribution* $\mathscr{N}(0, 1)$.

2. In the continuous-time case a similar role in the construction of many models of 'complex' structure is played by a *Brownian motion*, introduced as a mathematical concept for the first time by L. Bachelier ([12], 1900) and A. Einstein ([132], 1905). A rigorous mathematical theory of a Brownian motion, as well as the corresponding measure in the function space were constructed by N. Wiener ([476], 1923) and one calls them also a *Wiener process*, and the *Wiener measure*.

By Definition 2 in § 1b, the *standard Brownian motion* $B = (B_t)_{t\geqslant 0}$ is a *continuous Gaussian random process with homogeneous independent increments* such that $B_0 = 0$, $\mathsf{E}B_t = 0$, and $\mathsf{E}B_t^2 = t$. Its covariance function is $\mathsf{E}B_sB_t = \min(s, t)$.

We have already pointed out several times the *self-similarity* property of a Brownian motion: for each $a > 0$ we have

$$\mathrm{Law}(B_{at}; t \geqslant 0) = \mathrm{Law}(a^{1/2}B_t; t \geqslant 0).$$

It follows from this property that the process $\left(\frac{1}{\sqrt{a}} B_{at}\right)_{t\geqslant 0}$ is also a Brownian motion. In addition, we mention several other transformations bringing about new processes $B^{(i)}$ ($i = 1, 2, 3, 4$) that are also Brownian motions: $B_t^{(1)} = -B_t$; $B_t^{(2)} = tB_{1/t}$ for $t > 0$ with $B_0^{(2)} = 0$; $B_t^{(3)} = B_{t+s} - B_s$ for $s > 0$; $B_t^{(4)} = B_T - B_{T-t}$ for $0 \leqslant t \leqslant T$, $T > 0$.

A multivariate process $B = (B^1, \dots, B^d)$ formed by d independent standard Brownian motions $B^i = (B_t^i)_{t\geqslant 0}$, $i = 1, \dots, d$, is called a *d-dimensional standard Brownian motion.*

Endowed with a rich structure, Brownian motion can be useful in the construction of various classes of random processes.

For instance, Brownian motion plays the role of a 'basic' process in the construction of *diffusion* Markov processes $X = (X_t)_{t\geqslant 0}$ as solutions of stochastic differential equations

$$dX_t = a(t, X_t)\, dt + \sigma(t, X_t)\, dB_t, \tag{1}$$

interpreted in the following (integral) sense:

$$X_t = X_0 + \int_0^t a(s, X_s)\, ds + \int_0^t \sigma(s, X_s)\, dB_s \tag{2}$$

for each $t > 0$.

The integral

$$I_t = \int_0^t \sigma(s, X_s)\, dB_s \tag{3}$$

involved in this expression is treated as a *stochastic Itô integral* with respect to the Brownian motion. (We consider the issues of stochastic integration and stochastic differential equations below, in § 3c.)

An important position in financial mathematics is occupied by a *geometric Brownian motion* $S = (S_t)_{t\geqslant 0}$ satisfying the stochastic differential equation

$$dS_t = S_t(a\, dt + \sigma\, dB_t) \tag{4}$$

with coefficients $a \in \mathbb{R}$ and $\sigma > 0$.

Setting an initial value S_0 independent of the Brownian motion $B = (B_t)_{t\geqslant 0}$ we can find the explicit solution

$$S_t = S_0\, e^{at}\, e^{\sigma B_t - \frac{\sigma^2}{2} t} \tag{5}$$

of this equation, which can be also written as

$$S_t = S_0 e^{H_t}$$

(cf. formula (1) in Chapter II, § 1a) with

$$H_t = \left(a - \frac{\sigma^2}{2}\right) t + \sigma B_t. \tag{6}$$

We call the process $H = (H_t)_{t\geqslant 0}$ a *Brownian motion with local drift* $(a - \sigma^2/2)$ *and diffusion* σ^2. We see from (6) that the local drift characterizes the average rate of change of $H = (H_t)_{t\geqslant 0}$; the diffusion σ^2 is often called the *differential variance* or (in the literature on finances) the *volatility*.

Probably, it was P. Samuelson ([420], 1965) who understood for the first time the importance of a geometric Brownian motion in the description of price dynamics; he called it also *economic* Brownian motion.

We present now other well-known examples of processes obtainable as solutions of stochastic differential equations (1) with suitably chosen coefficients $a(t, x)$ and $\sigma(t, x)$.

A Brownian bridge $X = (X_t)_{0\leqslant t\leqslant T}$ with $X_0 = \alpha$ and $X_T = \beta$ is a process governed by the equation

$$dX_t = \frac{\beta - X_t}{T - t}\,dt + dB_t, \qquad 0 < t < T, \tag{7}$$

where $B = (B_t)_{t\geqslant 0}$ is a Brownian motion.

Using, e.g., Itô's formula (see § 3d below) one can verify that the process $X = (X_t)_{0\leqslant t\leqslant T}$ with

$$X_t = \alpha\left(1 - \frac{t}{T}\right) + \beta\frac{t}{T} + (T - t)\int_0^t \frac{dB_s}{T - s} \tag{8}$$

is a solution of (2) (here we treat the integral as a stochastic integral with respect to a Brownian motion). Since this equation is uniquely soluble (see § 3e), formula (8) defines a Brownian bridge issuing from the point α at time $t = 0$ and arriving at β for $t = T$.

For a standard Brownian motion, its autocovariance function $\rho(s, t)$ is equal to $\min(s, t)$, and for a Brownian bridge it is $\rho(s, t) = \min(s, t) - \dfrac{st}{T}$. The corresponding expectation is $\mathsf{E} X_t = \alpha\left(1 - \dfrac{t}{T}\right) + \beta\dfrac{t}{T}$.

It is easy to verify that for each standard Brownian motion $W = (W_t)_{t\geqslant 0}$, the process $W^0 = (W_t^0)_{0\leqslant t\leqslant T}$ defined by the formula

$$W_t^0 = W_t - \frac{t}{T}W_T \tag{9}$$

has the covariance function $\rho(s, t) = \min(s, t) - \dfrac{st}{T}$. Hence the process $Y = (Y_t)_{0\leqslant t\leqslant T}$ with

$$Y_t = \alpha\left(1 - \frac{t}{T}\right) + \beta\frac{t}{T} + W_t^0, \tag{10}$$

has the same finite-dimensional distributions as $X = (X_t)_{0\leqslant t\leqslant T}$ defined by (8) ($\mathrm{Law}(Y_t, t \leqslant T) = \mathrm{Law}(X_t, t \leqslant T)$), and, thus, it can be regarded as a version of a Brownian bridge.

An Ornstein–Uhlenbeck process ([466], 1930) is a solution of the (linear) stochastic differential equation

$$dX_t = -\alpha X_t\,dt + \sigma\,dB_t, \qquad \alpha > 0. \tag{11}$$

Again, using Itô's formula we can verify that the process $X = (X_t)_{t\geqslant 0}$ with

$$X_t = X_0 e^{-\alpha t} + \sigma \int_0^t e^{-\alpha(t-s)}\,dB_s \tag{12}$$

is the (unique; see § 3e) solution of (11).

If the initial value X_0 is independent of the Brownian motion $B = (B_t)_{t\geqslant 0}$ and has finite second moment, then

$$\mathsf{E}X_t = e^{-\alpha t}\mathsf{E}X_0, \tag{13}$$

$$\mathsf{D}X_t = \frac{\sigma^2}{2\alpha} + \left(\mathsf{D}X_0 - \frac{\sigma^2}{2\alpha}\right)e^{2\alpha t}, \tag{14}$$

$$\mathsf{Cov}(X_s, X_t) = \left[\mathsf{D}X_0 + \frac{\sigma^2}{2\alpha}\left(e^{2\alpha\min(s,t)} - 1\right)\right]e^{-\alpha(s+t)}. \tag{15}$$

If X_0 has a normal (Gaussian) distribution with $\mathsf{E}X_0 = 0$ and $\mathsf{D}X_0 = \sigma^2/(2\alpha)$, then $X = (X_t)_{t\geqslant 0}$ is a stationary Gaussian process with expectation zero and covariance function

$$\rho(s,t) = \frac{\sigma^2}{2\alpha}\,e^{-\alpha|t-s|}. \tag{16}$$

It should be noted in connection with the Ornstein–Uhlenbeck equation (11) that it is a well-posed version of *Langevin's equation* ([295], 1908)

$$m\,\frac{dV_t}{dt} = -\beta V_t + \sigma\,\frac{dB_t}{dt}, \tag{17}$$

describing the evolution of the velocity V_t of a particle of mass m put in a fluid and pushed ahead in the presence of friction $(-\beta V_t)$ by clashes with molecules described by a Brownian motion.

Written as in (17), this equation has in general no sense if one understands derivatives in the usual sense, because (for almost all realizations) of a Brownian motion the derivative dB_t/dt is nonexistent (see § 3b.7 below).

However, we can assign a precise meaning to this equation if we treat it as the Ornstein–Uhlenbeck equation

$$dV_t = -\frac{\beta}{m}V_t\,dt + \frac{\sigma}{m}\,dB_t, \tag{18}$$

which has for $V_0 = 0$ a solution that can be described, in accordance with (12), by the equality

$$V_t = \frac{\sigma}{m} \int_0^t e^{\frac{\beta}{m}(t-s)}\, dB_s. \tag{19}$$

The Bessel process of order $\alpha > 1$ is, by definition, a process $X = (X_t)_{t\geqslant 0}$, governed by the (nonlinear) stochastic differential equation

$$dX_t = \frac{\alpha - 1}{2} \frac{dt}{X_t} + dB_t \tag{20}$$

with initial value $X_0 = x \geqslant 0$, where $B = (B_t)_{t\geqslant 0}$ is a Brownian motion. (This equation has a—unique—strong solution; see § 3e).

If $\alpha = d$, where $d = 2, 3, \dots$, then X can be realized [402] as the radial component $R = (R_t)_{t\geqslant 0}$ of a d-dimensional Brownian motion

$$B^{(x)} = \left(x_1 + B_t^1, \dots, x_d + B_t^d\right)_{t\geqslant 0}$$

with independent standard Brownian motions $B^i = (B_t^i)_{t\geqslant 0}$ and $x_1^2 + \cdots + x_d^2 = x^2$, i.e.,

$$R_t = \sqrt{(x_1 + B_t^1)^2 + \cdots + (x_d + B_t^d)^2}\,. \tag{21}$$

We discuss several other interesting processes governed by stochastic differential equations in § 4, in connection with the construction of models describing the dynamics of bond prices $\mathsf{P}(t, T)$ (see Chapter I, § 1b).

§ 3b. Brownian Motion: a Compendium of Classical Results

1. Brownian motion as a limit of random walks. As described by many authors (see, for instance, [201: p. 254] or [266; p. 47]), circa 1827, a botanist R. Brown discovered that pollen particles put in a liquid make chaotic, irregular movements. (He described this phenomenon in his pamphlet "A Brief Account of Microscopical Observation ..." published in 1828.)

These movements were called a *Brownian motion*. They, as became clear later, were brought on by clashes of liquid molecules with the immersed particles. The corresponding *mathematical model* of this physical phenomenon was built by A. Einstein ([132], 1905). However, it should be pointed out for fairness sake that even earlier, in 1900, a similar model had been built by L. Bachelier [12] in connection with the description of the evolution of stock prices and other financial indexes on the Paris securities market.

As pointed out in Chapter I (§ 1b), a Brownian motion arose in Bachelier's analysis as a (formal) limit of simplest random walks.

Namely, let $(\xi_k)_{k\geqslant 1}$ be a sequence of independent identically distributed random variable assuming two values, ± 1, with probabilities $\frac{1}{2}$ (the *Bernoulli scheme*).

We consider the half-axis $\mathbb{R}_+ = [0, \infty)$, and for each $\Delta > 0$ we construct the process $S^{(\Delta)} = (S_t^{(\Delta)})_{t \geqslant 0}$,

$$S_t^{(\Delta)} = x + \sum_{k=1}^{[t/\Delta]} \sqrt{\Delta}\, \xi_k, \tag{1}$$

with *piecewise constant* trajectories.

Starting from the processes $S^{(\Delta)}$ we can also construct random processes $\overline{S}^{(\Delta)} = (\overline{S}_t^{(\Delta)})_{t \geqslant 0}$ with *continuous* trajectories by setting

$$\overline{S}_t^{(\Delta)} = S_{k\Delta}^{(\Delta)} + \frac{1}{\Delta}(t - k\Delta)\bigl(S_{(k+1)\Delta}^{(\Delta)} - S_{k\Delta}^{(\Delta)}\bigr). \tag{2}$$

By the multidimensional *central limit theorem* (see, e.g., [51; Chapter 8] or [439; Chapter VII, § 8]) we can conclude that for all $t_1, \dots, t_k$, $k \geqslant 1$, the finite-dimensional distributions $\operatorname{Law}(S_{t_1}^{(\Delta)}, \dots, S_{t_k}^{(\Delta)})$ and $\operatorname{Law}(\overline{S}_{t_1}^{(\Delta)}, \dots, \overline{S}_{t_k}^{(\Delta)})$ converge (weakly) to the finite-dimensional distributions $\operatorname{Law}(B_{t_1}, \dots, B_{t_k})$, where $B = (B_t)_{t \geqslant 0}$ is a standard Brownian motion.

We can actually say even more; namely,

$$\operatorname{Law}\bigl(S_t^{(\Delta)},\ t \geqslant 0\bigr) \to \operatorname{Law}(B_t, t \geqslant 0)$$

and

$$\operatorname{Law}\bigl(\overline{S}_t^{(\Delta)},\ t \geqslant 0\bigr) \to \operatorname{Law}(B_t, t \geqslant 0)$$

in the sense of *weak* convergence of distributions in the (Skorokhod) spaces D (of right-continuous functions having limits from the left) and the space C (of continuous functions); see, e.g., [39] and [250] for greater detail.

2. Brownian motion as a Markov process. Let $(\Omega, \mathscr{F}, \mathsf{P})$ be some fixed probability space. Let $B = (B_t(\omega))_{t \geqslant 0}$ be a Brownian motion defined on this space.

By $\mathscr{F}_t^0 = \sigma(B_s, s \leqslant t)$ we shall mean the σ-algebra of events generated by the variables B_s, $s \leqslant t$; let

$$\mathscr{F}_t^+ = \bigcap_{s>t} \mathscr{F}_s^0 \tag{3}$$

be the σ-algebra of events observable not only on the interval $[0, t]$, but also 'in the infinitesimal future' relative to the instant t.

We note that, by contrast to $(\mathscr{F}_t^0)_{t \geqslant 0}$, the family $(\mathscr{F}_t^+)_{t \geqslant 0}$ has an important property of *right continuity*:

$$\bigcap_{s>t} \mathscr{F}_s^+ = \mathscr{F}_t^+. \tag{4}$$

Still, the difference between the σ-algebras $\mathscr{F}_t^0$ and $\mathscr{F}_t^+$ is not that essential (in the following sense). Let $\mathscr{N} = \{A \in \mathscr{F} : \mathsf{P}(A) = 0\}$ be the set of zero-probability events in $\mathscr{F}$. Then the σ-algebra $\sigma(\mathscr{F}_t^+ \cup \mathscr{N})$ generated by the events in $\mathscr{F}_t^+$ and $\mathscr{N}$ coincides with the σ-algebra $\sigma(\mathscr{F}_t^0 \cup \mathscr{N})$ generated by the events in $\mathscr{F}_t^0$ and $\mathscr{N}$:

$$\sigma(\mathscr{F}_t^+ \cup \mathscr{N}) = \sigma(\mathscr{F}_t^0 \cup \mathscr{N}). \tag{5}$$

This is an argument in favor of the introduction of another family of σ-algebras $(\mathscr{F}_t)_{t\geqslant 0}$ with $\mathscr{F}_t = \sigma\big(\mathscr{F}_t^0 \cup \mathscr{N}\big) \equiv \sigma\big(\mathscr{F}_t^+ \cup \mathscr{N}\big)$ such that each σ-algebra is clearly completed with all sets of probability zero, and the whole family is *right-continuous*, i.e., $\mathscr{F}_t = \bigcap_{s>t} \mathscr{F}_s$. (One says that the corresponding basis $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geqslant 0}, \mathsf{P})$ satisfies *usual conditions*; see subsection 3 below.)

Assume that $T > 0$ and let $\widetilde{B}(T) = (\widetilde{B}_t(T;\omega))_{t\geqslant 0}$ be the process constructed from the Brownian motion $B = (B_t(\omega))_{t\geqslant 0}$ by the formula

$$\widetilde{B}_t(T;\omega) = B_{t+T}(\omega) - B_T(\omega).$$

We already mentioned in § 3a.2 that 1) $\widetilde{B}(T)$ is also a *Brownian motion*. Moreover, it is easy to show that 2) the σ-algebras $\mathscr{F}_T^0 = \sigma(B_s, s \leqslant T)$ and $\widetilde{\mathscr{F}}_\infty^0(T) = \sigma(\widetilde{B}_s(T), s \geqslant 0)$ are *independent*. (As usual, we prove first that the events in the corresponding cylindrical algebras are independent and then use the 'monotonic classes method'; see, e.g., [439; Chapter II, § 2].)

It is the combination of these properties that one often calls the *Markov property* of the Brownian motion (see, e.g., [288; Chapter II]) inferring from it, say, the standard Markov property of the independence of the 'future' and the 'past' for any fixed 'present'. Namely, *if $f = f(x)$ is a bounded Borel function and $\sigma(B_T)$ is the σ-algebra generated by B_T, then for each $t > 0$ we have*

$$\mathsf{E}\big(f(B_{T+t}) \,|\, \mathscr{F}_T^0\big) = \mathsf{E}\big(f(B_{T+t}) \,|\, \sigma(B_T)\big) \qquad (\mathsf{P}\text{-a.s.}). \tag{6}$$

This analytic form of the Markov property leaves place for various generalizations. For instance, one can consider the σ-algebra $\mathscr{F}_T$ in place of $\mathscr{F}_T^0$, and bounded *trajectory functionals* $f(B_{T+t}, t \geqslant 0)$ in place of $f(B_{T+t})$. See, e.g., [123] and [126] for greater detail.

The following generalization, which brings one to the *strong Markov property*, relates to the extension of the above Markov properties to the case when, in place of (deterministic) time T, one considers random Markov times $\tau = \tau(\omega)$.

To this end we assume that $\tau = \tau(\omega)$ is a finite Markov time (relative to the flow $(\mathscr{F}_t)_{t\geqslant 0}$).

By analogy with $\widetilde{B}(T)$ we now consider the process $\widetilde{B}(\tau) = (\widetilde{B}_t(\tau(\omega);\omega))_{t\geqslant 0}$, where

$$\widetilde{B}_t(\tau(\omega);\omega) = B_{t+\tau(\omega)}(\omega) - B_{\tau(\omega)}(\omega). \tag{7}$$

The most elementary version of the *strong Markov property* requires that the process $\widetilde{B}(\tau)$ be also a Brownian motion and the σ-algebras $\mathscr{F}_\tau$ (see Definition 2 in Chapter II, § 1b) and $\widetilde{\mathscr{F}}_\infty^0(\tau) \equiv \sigma(\widetilde{\mathscr{F}}_\infty^0(\tau) \cup \mathscr{N})$ be independent.

The analytic property (6) has the following, perfectly natural, generalization:

$$\mathsf{E}\big(f(B_{\tau(\omega)+t}(\omega))\,|\,\mathscr{F}_\tau\big) = \mathsf{E}\big(f(B_{\tau(\omega)+t}(\omega))\,|\,\sigma(B_\tau)\big) \qquad (\mathsf{P}\text{-a.s.}). \tag{8}$$

(As regards generalizations of this property from the case of a function f to functionals, see, e.g., the books [123], [126], and [288].)

3. Brownian motion and square integrable martingales. The following properties follow immediately from the definition of the Brownian motion $B = (B_t)_{t\geqslant 0}$: for $t \geqslant 0$

$$B_t \text{ is } \mathscr{F}_t\text{-measurable}, \tag{9}$$

$$\mathsf{E}|B_t| < \infty, \tag{10}$$

$$\mathsf{E}(B_t\,|\,\mathscr{F}_s) = B_s \quad (\mathsf{P}\text{-a.s.}) \quad \text{for} \quad s \leqslant t. \tag{11}$$

These three properties are precisely the ones from the definition of a *martingale* $B = (B_t)_{t\geqslant 0}$ with respect to the flow of σ-algebras $(\mathscr{F}_t)$ and the probability measure P (cf. Definition 2 in Chapter II, § 1c).

Further, since

$$\mathsf{E}\big(B_t^2 - B_s^2\,|\,\mathscr{F}_s\big) = t - s, \tag{12}$$

the process $(B_t^2 - t)_{t\geqslant 0}$ is also a martingale.

Now let $B = (B_t)_{t\geqslant 0}$ be *some* process satisfying (9)–(12). This is a remarkable fact that, in effect, these properties unambiguously specify the probabilistic structure of this process.

Namely, let $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geqslant 0}, \mathsf{P})$ be a filtered probability space with flow of σ-algebras $(\mathscr{F}_t)_{t\geqslant 0}$ satisfying the *usual conditions* ([250; Chapter I, § 1]) of right-continuity and completeness with respect to the measure P. (We point out that the $\mathscr{F}_t$ here are not necessarily the same as the earlier introduced algebra $\sigma(\mathscr{F}_t^+ \cup \mathscr{N})$.)

Each process $B = (B_t)_{t\geqslant 0}$ with properties (9)–(11) is called a *martingale*. To emphasize the property of measurability with respect to the flow $(\mathscr{F}_t)_{t\geqslant 0}$ and the measure P, one often writes $B = (B_t, \mathscr{F}_t)$ or $B = (B_t, \mathscr{F}_t, \mathsf{P})$. (Cf. definitions in Chapter II, § 1c for the discrete-time case.)

THEOREM (P. Lévy, [298]). *Let $B = (B_t, \mathscr{F}_t)_{t\geqslant 0}$ be a continuous square integrable martingale defined on a filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geqslant 0}, \mathsf{P})$. Assume that* (12) *holds, i.e., $(B_t^2 - t, \mathscr{F}_t)_{t\geqslant 0}$ is also a martingale. Then $B = (B_t)_{t\geqslant 0}$ is a standard Brownian motion.*

See the proof in § 5c below.

4. Wald's identities. Convergence and optional stopping theorems for uniformly integrable martingales. For a Brownian motion we have

$$\mathsf{E}B_t = 0 \quad \text{and} \quad \mathsf{E}B_t^2 = t.$$

In many problems of stochastic calculus it is often necessary to find $\mathsf{E}B_\tau$ and $\mathsf{E}B_\tau^2$ for Markov times τ (with respect to the flow $(\mathscr{F}_t)_{t\geqslant 0}$).

The following relations are extended versions of *Wald's identities* for $B = (B_t, \mathscr{F}_t)_{t\geqslant 0}$:

$$\mathsf{E}\sqrt{\tau} < \infty \implies \mathsf{E}B_\tau = 0,$$
$$\mathsf{E}\tau < \infty \implies \mathsf{E}B_\tau^2 = \mathsf{E}\tau.$$

In the particular case when τ is a bounded Markov time ($\mathsf{P}(\tau \leqslant c) = 1$ for some constant $c > 0$), the equalities $\mathsf{E}B_\tau = 0$ and $\mathsf{E}B_\tau^2 = \mathsf{E}\tau$ are immediate consequences of the following result.

THEOREM (J. Doob, [109]). *Let $X = (X_t, \mathscr{F}_t)_{t\geqslant 0}$ be a uniformly integrable martingale (i.e., a martingale such that $\sup\limits_t \mathsf{E}\big(|X_t| I(|X_t| > N)\big) \to 0$ as $N \to \infty$). Then*

1) *there exists an integrable random variable X_∞ such that*

$$X_t \to X_\infty \quad (\mathsf{P}\text{-a.s.}),$$
$$\mathsf{E}|X_t - X_\infty| \to 0$$

as $t \to \infty$ and

$$\mathsf{E}(X_\infty \,|\, \mathscr{F}_t) = X_t \quad (\mathsf{P}\text{-a.s.})$$

for all $t \geqslant 0$;

2) *for all Markov times σ and τ we have*

$$X_{\tau\wedge\sigma} = \mathsf{E}(X_\sigma \,|\, \mathscr{F}_\tau) \quad (\mathsf{P}\text{-a.s.}),$$

where $\tau \wedge \sigma = \min(\tau, \sigma)$.

(Cf. *Doob's convergence* and *optional stopping theorems* in the discrete-time case in Chapter V, § 3a.)

5. Stochastic exponential. In Chapter II, § 1a we gave the definition of a *stochastic exponential* $\mathscr{E}(\widehat{H})_t$ for processes $\widehat{H} = (\widehat{H}_t)_{t\geqslant 0}$ that are semimartingales.

In the case of $\widehat{H}_t = \lambda B_t$ this stochastic exponential $\mathscr{E}(\lambda B)_t$ can be defined by the equality

$$\mathscr{E}(\lambda B)_t = e^{\lambda B_t - \frac{\lambda^2}{2}t}. \tag{13}$$

It immediately follows from Itô's formula (§ 3d) that $X_t = \mathscr{E}(\lambda B)_t$ satisfies the stochastic differential equation

$$dX_t = \lambda X_t\, dB_t \tag{14}$$

with initial condition $X_0 = 1$.

If ξ has the distribution $\mathscr{N}(0,1)$, then

$$\mathsf{E}e^{\lambda\xi - \frac{\lambda^2}{2}} = 1.$$

By this property and the self-similarity of a Brownian motion, meaning that

$$\operatorname{Law}(\lambda B_t) = \operatorname{Law}(\lambda\sqrt{t}\,B_1),$$

we obtain

$$\mathsf{E}\exp\left\{\lambda B_t - \frac{\lambda^2}{2}t\right\} = \mathsf{E}\exp\left\{\lambda B_t - \frac{(\lambda\sqrt{t}\,)^2}{2}\right\} = \mathsf{E}\exp\left\{(\lambda\sqrt{t})B_1 - \frac{(\lambda\sqrt{t}\,)^2}{2}\right\} = 1.$$

In a similar way we can show that for all $s \leqslant t$

$$\mathsf{E}\big(\mathscr{E}(\lambda B)_t \,|\, \mathscr{F}_s\big) = \mathscr{E}(\lambda B)_s \qquad (\mathsf{P}\text{-a.s.}). \tag{15}$$

That is, the stochastic exponent $\mathscr{E}(\lambda B) = (\mathscr{E}(\lambda B)_t)_{t\geqslant 0}$ is a martingale.

6. Constructions of a Brownian motion. Let $\varepsilon = (\varepsilon_k)_{k\geqslant 0}$ be Gaussian white noise, i.e., a sequence of independent normally ($\mathscr{N}(0,1)$) distributed random variables. Let $H_k = H_k(t)$, $k \geqslant 0$, be the *Haar functions* (see, e.g., [439; Chapter II, § 11]) defined on the time interval $[0,1]$, and let

$$S_k(t) = \int_0^t H_k(s)\,ds$$

be the *Schauder functions*.

We now set

$$B_t^{(n)} = \sum_{k=0}^{n} \varepsilon_k S_k(t).$$

It follows from the results of P. Lévy [298] and Z. Ciesielski [76] that the $B_t^{(n)}$ converge (P-a.s.) *uniformly* in $t \in [0,1]$ and their P-a.s. *continuous* limit is a standard Brownian motion

An earlier construction of R. Paley and N. Wiener [374] (1934) is the (uniformly convergent) series

$$B_t \equiv \varepsilon_0 t + \sum_{n=1}^{\infty}\left(\sum_{k=2^{n-1}}^{2^n-1} \sqrt{2}\,\varepsilon_k\,\frac{\sin k\pi t}{k\pi}\right).$$

7. Local properties of the trajectories. The following results are well known and their proofs can be found in many monographs and textbooks (e.g., [124], [245], [266], [470]).

With probability one, the trajectories of a Brownian motion

a) *satisfy* the Hölder condition

$$|B_t - B_s| \leqslant c\,|t-s|^\gamma$$

for each $\gamma < \frac{1}{2}$;

b) *do not satisfy* the Lipschitz condition and therefore are not differentiable at any $t > 0$;

c) have unbounded variation on each interval (a,b): $\int_{(a,b)} |dB_s| = \infty$.

8. The zeros of the trajectories of a Brownian motion. Let $(B_t(\omega))_{t\geqslant 0}$ be a trajectory of a Brownian motion corresponding to an elementary outcome $\omega \in \Omega$, and let

$$\mathfrak{N}(\omega) = \{0 \leqslant t < \infty \colon B_t(\omega) = 0\}$$

be its zero set.

Then we have the following results ([124], [245], [266], [470]): P-almost surely,

a) Lebesgue measure $\lambda(\mathfrak{N}(\omega))$ is zero;
b) the point $t = 0$ is a condensation point of the zeros;
c) there are no isolated zeros in $(0, \infty)$, so that $\mathfrak{N}(\omega)$ is dense in itself;
d) the set $\mathfrak{N}(\omega)$ is closed and unbounded.

9. Behavior at the origin. The *Local law of the iterated logarithm* states that (P-a.s.)

$$\overline{\lim_{t\downarrow 0}} \frac{|B_t|}{\sqrt{2t \ln|\ln t|}} = 1.$$

By this property, as applied to the Brownian motions $(B_{t+h} - B_t)_{t\geqslant 0}$, we obtain that (P-a.s.)

$$\overline{\lim_{h\downarrow 0}} \frac{|B_{t+h} - B_t|}{\sqrt{h}} = \infty$$

for each $t \geqslant 0$, i.e., the Lipschitz condition, as already pointed out, fails on Brownian trajectories.

10. The modulus of continuity is a geometrically transparent measure for the oscillations of functions, trajectories, and so on. A well-known result of P. Lévy [298] about the modulus of continuity of the trajectories of a Brownian motion states that, with probability one,

$$\overline{\lim_{h\downarrow 0}} \frac{\max\limits_{0\leqslant s<t\leqslant 1,\, t-s\leqslant h} |B_t - B_s|}{\sqrt{2h \ln(1/h)}} = 1.$$

11. Behavior as $t \to \infty$. With probability one,

$$\frac{B_t}{t} \to 0 \quad \text{as} \quad t \to \infty,$$

(*Strong law of large numbers*).

Moreover,

$$\frac{B_t}{\sqrt{t \ln t}} \to 0 \quad \text{as} \quad t \to \infty \quad \text{(P-a.s.)},$$

but

$$\overline{\lim_{t\to\infty}} \frac{|B_t|}{\sqrt{t}} = \infty \quad \text{(P-a.s.)}.$$

The precise asymptotic behavior of the trajectories of a Brownian motion as $t \to \infty$ can be described by the *Law of the iterated logarithm*. Namely,

$$\varlimsup_{t\to\infty} \frac{|B_t|}{\sqrt{2t \ln \ln t}} = 1 \quad (\mathsf{P}\text{-a.s.}). \tag{16}$$

12. Quadratic variation. Although the trajectories of a Brownian motion have (P-a.s.) unbounded variation, i.e., $\int_{(a,b)} |dB_s| = \infty$, we can assert that $\int_{(a,b)} |dB_s|^2 = b - a$ in a certain sense.

The corresponding result, which plays a key role in many issues of stochastic calculus (e.g., in the proof of Itô's formula; § 3d), can be stated as follows.

Let $T^{(n)} = (t_0^{(n)}, \dots, t_{k_n}^{(n)})$ be a partitioning of the interval $[a, b]$ such that

$$a = t_0^{(n)} \leqslant \dots \leqslant t_{k_n}^{(n)} = b.$$

Let

$$\|T^{(n)}\| = \sup_{0 \leqslant k < k_n} \left| t_{k+1}^{(n)} - t_k^{(n)} \right|. \tag{17}$$

Then we have:

a) if $\|T^{(n)}\| \to 0$ as $n \to \infty$, then

$$\sum_{k=0}^{k_n - 1} \left| B_{t_{k+1}^{(n)}} - B_{t_k^{(n)}} \right|^2 \xrightarrow{\mathsf{P}} b - a; \tag{18}$$

b) if $\sum\limits_{n=1}^{\infty} \|T^{(n)}\| < \infty$, then we have convergence in (18) with probability one;

c) if $B^{(1)}$ and $B^{(2)}$ are two independent Brownian motions and $\|T^{(n)}\| \to 0$ as $n \to \infty$, then

$$\sum_{k=0}^{k_n - 1} \left(B^{(1)}_{t_{k+1}^{(n)}} - B^{(1)}_{t_k^{(n)}} \right) \left(B^{(2)}_{t_{k+1}^{(n)}} - B^{(2)}_{t_k^{(n)}} \right) \xrightarrow{\mathsf{P}} 0. \tag{19}$$

In a symbolic form, relations (18) and (19) are often written as follows:

$$(dB_t)^2 = dt \quad \text{and} \quad dB_t^{(1)} dB_t^{(2)} = 0. \tag{20}$$

13. Passage times of levels. a) Assume that $a > 0$ and let $T_a = \inf\{t \geqslant 0 : B_t = a\}$. Then it is clear that

$$\mathsf{P}(T_a < t) = \mathsf{P}\Big(\sup_{s \leqslant t} B_s > a\Big). \tag{21}$$

Using the *reflection principle* of D. André we see that

$$\mathsf{P}(T_a < t) = 2\mathsf{P}(B_t \geqslant a). \tag{22}$$

(See, e.g., [124], [266], and [439].)

Since

$$\mathsf{P}(B_t \geqslant a) = \frac{1}{\sqrt{2\pi t}} \int_a^\infty e^{-\frac{x^2}{2t}}\, dx, \tag{23}$$

it follows that

$$\mathsf{P}(T_a < t) = \int_0^t \frac{a}{\sqrt{2\pi s^3}}\, e^{-\frac{a^2}{2s}}\, ds, \tag{24}$$

therefore the distribution density $p_a(t) = \dfrac{\partial \mathsf{P}(T_a < t)}{\partial t}$ is defined by the formula

$$p_a(t) = \frac{a}{\sqrt{2\pi t^3}}\, e^{-\frac{a^2}{2t}}. \tag{25}$$

Hence, in particular, $\mathsf{P}(T_a < \infty) = 1$, $\mathsf{E}T_a = \infty$, and the corresponding Laplace transform is

$$\mathsf{E}e^{-\lambda T_a} = e^{-a\sqrt{2\lambda}}. \tag{26}$$

It should be noted that $T = (T_a)_{a \geqslant 0}$ is a process with independent homogeneous increments (by the strong Markov property of a Brownian motion). Moreover, this is a stable process with parameter $\alpha = \frac{1}{2}$, i.e., $\mathrm{Law}(T_a) = \mathrm{Law}(a^2 T_1)$ (cf. § 1a.5).

b) Assume that $a > 0$ and let $S_a = \inf\{t \geqslant 0 : |B_t| = a\}$. We claim that $\mathsf{E}S_a = a^2$ and

$$\mathsf{E}e^{-\lambda S_a} = \frac{1}{\cosh(a\sqrt{2\lambda})}. \tag{27}$$

By Wald's identity, $\mathsf{E}B^2_{S_a \wedge t} = \mathsf{E}(S_a \wedge t)$ for each $t > 0$, therefore $\mathsf{E}(S_a \wedge t) \leqslant a^2$. Hence $\mathsf{E}S_a = \lim_{t \to \infty} \mathsf{E}(S_a \wedge t) \leqslant a^2$ by the monotone convergence theorem. (Of course, it follows from this that $S_a < \infty$ with probability one.) On the other hand, $\mathsf{E}S_a < \infty$, therefore, using Wald's identity again we obtain that $\mathsf{E}B^2_{S_a} = \mathsf{E}S_a$. Bearing in mind that $B^2_{S_a} = a$ we see that $\mathsf{E}S_a = a^2$. Bearing in mind that $B^2_{S_a} = a$ we see therefore that $\mathsf{E}S_a = a^2$.

To prove (27) we now consider the martingale $X^{(a)} = (X_{t \wedge S_a}, \mathscr{F}_t)$, where

$$X_{t \wedge S_a} = \exp\Big\{\lambda B_{t \wedge S_a} - \frac{\lambda^2}{2}(t \wedge S_a)\Big\}. \tag{28}$$

Since $|B_{t\wedge S_a}| \leqslant a$, this is a uniformly integrable martingale and, by Doob's theorem in subsection 4,

$$\mathsf{E}X_{t\wedge S_a} = 1. \tag{29}$$

By the theorem on dominated convergence, we can pass to the limit as $t \to \infty$ in this equality to obtain

$$\mathsf{E}X_{S_a} = 1,$$

i.e.,

$$\mathsf{E}\exp\left\{\lambda B_{S_a} - \frac{\lambda^2}{2}S_a\right\} = 1. \tag{30}$$

Since $\mathsf{P}(S_a < \infty) = 1$ and since $\mathsf{P}(B_{S_a} = a) = \mathsf{P}(B_{S_a} = -a) = \frac{1}{2}$ by symmetry reasons, the required equality (27) follows by (30).

c) Let $T_{a,b} = \inf\{t\colon B_t = a + bt\}$, $a > 0$. If $b \leqslant 0$, then $\mathsf{P}(T_{a,b} < \infty) = 1$. Using the fact that the process $\left(e^{\theta B_t - \frac{\theta^2}{2}t}\right)_{t\geqslant 0}$ is a martingale, and setting $\theta = b+\sqrt{b^2+2\lambda}$ we can find the Laplace transform

$$\mathsf{E}e^{-\lambda T_{a,b}} = \exp\bigl\{-a\bigl[b + \sqrt{b^2+2\lambda}\bigr]\bigr\}. \tag{31}$$

By this formula or directly from Wald's identity $0 = \mathsf{E}B_{T_{a,b}}$ $(= a + b\,\mathsf{E}T_{a,b})$ we obtain

$$\mathsf{E}T_{a,b} = -\frac{a}{b}\,.$$

(Using the trick described after formula (27) we can prove that $\mathsf{E}T_{a,b} < \infty$.)

If $b > 0$, then we consider the martingale $\left(e^{\theta B_t - \frac{\theta^2}{2}t}\right)_{t\geqslant 0}$ with $\theta = 2b$ to obtain

$$\exp\left\{2bB_t - \frac{(2b)^2 t}{2}\right\} \leqslant \exp\left\{2b(a+bt) - \frac{(2b)^2}{2}t\right\} \leqslant e^{2ab},$$

so that this martingale is uniformly integrable. Hence

$$\begin{aligned} 1 &= \mathsf{E}\exp\left\{\theta B_{T_{a,b}} - \frac{\theta^2}{2}T_{a,b}\right\} \\ &= \mathsf{E}\exp\left\{\theta B_{T_{a,b}} - \frac{\theta^2}{2}T_{a,b}\right\} I(T_{a,b} < \infty) = \mathsf{P}(T_{a,b} < \infty)e^{2ab}, \end{aligned}$$

therefore if $a > 0$ and $b > 0$ (or $a < 0$ and $b < 0$), then

$$\mathsf{P}(T_{a,b} < \infty) = e^{-2ab}. \tag{32}$$

14. Maximal inequalities. Let $B = (B_t)_{t\geqslant 0}$ be a Brownian motion. Then for $\lambda > 0$, $p \geqslant 1$, and a finite Markov time T we have

$$\mathsf{P}\Big(\max_{t\leqslant T} |B_t| \geqslant \lambda\Big) \leqslant \frac{\mathsf{E}|B_T|^p}{\lambda^p}\,; \tag{33}$$

and

$$\mathsf{E}|B_T|^p \leqslant \mathsf{E}\max_{t\leqslant T} |B_t|^p \leqslant \Big(\frac{p}{p-1}\Big)^p \mathsf{E}|B_T|^p \tag{34}$$

for $p > 1$.

In particular, if $p = 2$, then

$$\mathsf{P}\Big(\max_{t\leqslant T} |B_t| \geqslant \lambda\Big) \leqslant \frac{\mathsf{E}B_T^2}{\lambda^2}\,, \tag{35}$$

$$\mathsf{E}\max_{t\leqslant T} B_t^2 \leqslant 4\mathsf{E}B_T^2. \tag{36}$$

Inequalities (33) and (35) are called the *Kolmogorov–Doob inequalities*, while (34) and (36) are called *Doob's inequalities* (see, e.g., [109], [110], [124], [303], [304], or [402]).

By (36) we obtain

$$\mathsf{E}\max_{t\leqslant T} |B_t| \leqslant 2\sqrt{\mathsf{E}B_T^2}\,. \tag{37}$$

If $\mathsf{E}T < \infty$, then $\mathsf{E}B_T^2 = \mathsf{E}T$, so that

$$\mathsf{E}\max_{t\leqslant T} |B_t| \leqslant 2\sqrt{\mathsf{E}T}\,. \tag{38}$$

As shown in [116], one can refine inequality (38); actually,

$$\mathsf{E}\max_{t\leqslant T} |B_t| \leqslant \sqrt{2}\,\sqrt{\mathsf{E}T}\,, \tag{39}$$

where the constant $\sqrt{2}$ is best possible.

Setting $T = 1$ in (39) we obtain

$$\mathsf{E}\max_{t\leqslant 1} |B_t| \leqslant \sqrt{2}\,. \tag{40}$$

Of course, it would be interesting to find the precise value of $\mathsf{E}\max\limits_{t\leqslant 1} |B_t|$.

The following argument shows that

$$\mathsf{E}\max_{t\leqslant 1} |B_t| = \sqrt{\frac{\pi}{2}}\,. \tag{41}$$

By the self-similarity of a Brownian motion, with $S_1 = \inf\{t \geqslant 0 \colon |B_t| = 1\}$, we find

$$\Bigl\{\sup_{t\leqslant 1}|B_t| \leqslant x\Bigr\} = \Bigl\{\sup_{t\leqslant 1}\frac{1}{x}|B_t| \leqslant 1\Bigr\} = \Bigl\{\sup_{t\leqslant 1}|B_{t/x^2}| \leqslant 1\Bigr\}$$
$$= \Bigl\{\sup_{t\leqslant 1/x^2}|B_t| \leqslant 1\Bigr\} = \Bigl\{S_1 \geqslant \frac{1}{x^2}\Bigr\} = \Bigl\{\frac{1}{\sqrt{S_1}} \leqslant x\Bigr\}.$$

Hence $\operatorname{Law}\Bigl(\sup_{t\leqslant 1}|B_t|\Bigr) = \operatorname{Law}\Bigl(\dfrac{1}{\sqrt{S_1}}\Bigr)$.

Further, since

$$\sigma = \sqrt{\frac{2}{\pi}}\int_0^\infty e^{-\frac{x^2}{2\sigma^2}}\,dx$$

for each $\sigma > 0$ by the properties of the normal integral, setting $\sigma = \dfrac{1}{\sqrt{S_1}}$ we see from (27) that

$$\mathsf{E}\max_{t\leqslant 1}|B_t| = \mathsf{E}\frac{1}{\sqrt{S_1}} = \sqrt{\frac{2}{\pi}}\int_0^\infty \mathsf{E}e^{-\frac{x^2 S_1}{2}}\,dx = \sqrt{\frac{2}{\pi}}\int_0^\infty \frac{dx}{\cosh x}$$
$$= 2\sqrt{\frac{2}{\pi}}\int_0^\infty \frac{e^x}{e^{2x}+1}\,dx = 2\sqrt{\frac{2}{\pi}}\int_1^\infty \frac{dy}{1+y^2}$$
$$= 2\sqrt{\frac{2}{\pi}}\arctan x\Big|_1^\infty = 2\sqrt{\frac{2}{\pi}}\cdot\frac{\pi}{4} = \sqrt{\frac{\pi}{2}},$$

which proves the required relation (41).

§ 3c. Stochastic Integration with respect to a Brownian Motion

1. Classical analysis has at its disposal various approaches to the 'operation of integration', which bring about such (generally speaking, different) concepts as Riemann, Lebesgue, Riemann–Stieltjes, Lebesgue–Stieltjes, Denjoy, and other integrals (see A. N. Kolmogorov's paper "A study of the concept of integral" in [277]).

In stochastic analysis one also considers various approaches to integration of random functions with respect to stochastic processes, stochastic measures, and so on, which brings about various construction of 'stochastic integrals'.

Apparently, N. Wiener was the first to define the stochastic integral

$$I_t(f) \equiv \int_{(0,t]} f(s)\,dB_s \tag{1}$$

for smooth deterministic functions $f = f(s)$, $s \geqslant 0$, using the idea of 'integration by parts' (see [375] and [476].)

Namely, starting from the 'natural' formula $d(fB) = f dB + B df$, one sets by definition

$$I_t(f) = f(t)B_t - \int_0^t f'(s)B_s\,ds, \tag{2}$$

where the integral $\int_0^t f'(s)B_s\,ds$ is treated as the trajectory-wise (i.e., for each $\omega \in \Omega$) Riemann integral of the continuous functions $f'(s)B_s(\omega)$, $s \geqslant 0$.

2. In 1944, K. Itô [244] made a significant step forward in the extension of the concept of a 'stochastic integral' and laid on this way the foundations of *modern stochastic calculus*, a powerful and efficient tool of the investigation of stochastic processes.

Itô's construction is as follows.

Let $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geqslant 0}, \mathsf{P})$ be a filtered probability space satisfying the usual conditions (see § 3b.2 and details in [250]). Let $B = (B_t, \mathscr{F}_t)_{t\geqslant 0}$ be a standard Brownian motion and let $f = (f(t,\omega))_{t\geqslant 0,\,\omega\in\Omega}$ be a random function that is measurable with respect to (t,ω) and *nonanticipating* (independent of the 'future'), i.e.,

$$f(t,\omega) \text{ is } \mathscr{F}_t\text{-measurable}$$

for each $t \geqslant 0$.

Such functions $f = f(t,\omega)$ are also said to be *adapted* (to the family of σ-algebras $(\mathscr{F}_t)_{t\geqslant 0}$).

Examples are provided by *elementary* functions

$$f(t,\omega) = Y(\omega)I_{\{0\}}(t), \tag{3}$$

where $Y(\omega)$ is a $\mathscr{F}_0$-measurable random variable.

Another example is the function

$$f(t,\omega) = Y(\omega)I_{(r,s]}(t) \tag{4}$$

(also said to be *elementary*), where $0 \leqslant r < s$ and $Y(\omega)$ is a $\mathscr{F}_r$-measurable random variable.

For functions of type (3) 'concentrated' at $t = 0$ (with respect to the time variable) the 'natural' value of the stochastic integral

$$I_t(f) = \int_{(0,t]} f(s,\omega)\,dB_s$$

is zero. If f is a function of type (4), then the 'natural' value of $I_t(f)$ is $Y(\omega)[B_{s\wedge t} - B_{r\wedge t}]$.

For a *simple* function

$$f(t,\omega) = Y_0(\omega)I_{\{0\}}(t) + \sum Y_i(\omega)I_{(r_i,s_i]}(t) \tag{5}$$

that is a linear combination of elementary functions we set by definition

$$I_t(f) = \int_{(0,t]} f(s,\omega)\,dB_s = \sum_i Y_i(\omega)\big(B_{s_i\wedge t} - B_{r_i\wedge t}\big). \tag{6}$$

Remark 1. We point out that one need not assume that $B = (B_t)_{t\geqslant 0}$ is a Brownian motion in the definition of such integrals of elementary functions. Any process can play this role. The peculiarities of a Brownian motion, which is the object of our current interests, become essential if one wants to define a 'stochastic integral' with simple properties for a *broader store* of functions $f = f(t,\omega)$, not merely for elementary ones and their linear combinations, simple functions.

Let us now agree to treat integrals $\int_0^t$ (usual or stochastic) as integrals $\int_{(0,t]}$ over the set $(0,t]$. Since the $Y_i(\omega)$ are $\mathscr{F}_{r_i}$-measurable, it follows that

$$\begin{aligned}\mathsf{E}\big[Y_i(\omega)(B_{r_i\wedge t} - B_{s_i\wedge t})\big] &= \mathsf{E}\,\mathsf{E}\big[Y_i(\omega)(B_{r_i\wedge t} - B_{s_i\wedge t}) \mid \mathscr{F}_{r_i}\big]\\ &= \mathsf{E}\big[Y_i(\omega)\,\mathsf{E}\big((B_{r_i\wedge t} - B_{s_i\wedge t}) \mid \mathscr{F}_{r_i}\big)\big] = 0.\end{aligned}$$

In a similar way,

$$\mathsf{E}\big[Y_i(\omega)(B_{r_i\wedge t} - B_{s_i\wedge t})\big]^2 = \mathsf{E}Y_i^2(r_i\wedge t - s_i\wedge t).$$

Hence if $f = f(t,\omega)$ is a simple function, then

$$\mathsf{E}\int_0^t f(s,\omega)\,dB_s = 0 \tag{7}$$

and

$$\mathsf{E}\left(\int_0^t f(s,\omega)\,dB_s\right)^2 = \mathsf{E}\int_0^t f^2(s,\omega)\,ds. \tag{8}$$

In a more compact form,

$$\mathsf{E}I_t(f) = 0, \tag{7$'$}$$

$$\mathsf{E}I_t^2(f) = \mathsf{E}\int_0^t f^2(s,\omega)\,ds. \tag{8$'$}$$

3. We shall now describe the classes of functions $f = f(t,\omega)$ to which one can extend 'stochastic integration' preserving 'natural' properties (e.g., (7) and (8)).

We shall assume that all the functions $f = f(t,\omega)$ under consideration are defined on $\mathbb{R}_t \times \Omega$ and are nonanticipating.

If

$$\mathsf{P}\left(\int_0^t f^2(s,\omega)\,ds < \infty\right) = 1 \tag{9}$$

for each $t > 0$, then we say that f belongs to the class J_1.

If, in addition,

$$\mathsf{E}\int_0^t f^2(s,\omega)\,ds < \infty \tag{10}$$

for all $t > 0$, then f is said to belong to J_2.

It is for these function classes that Itô [244] gave his 'natural' definition of a stochastic integral $I_t(f)$ based on the following observations.

Since the trajectories of a Brownian motion (P-a.s.) are of unbounded variation (see § 3b.7), the integral $I_t(f) = \int_0^t f(s,\omega)\,dB_s$ cannot be defined as the trajectory-wise Lebesgue–Stieltjes integral. Itô's idea was to define it as the limit (in a suitable probabilistic sense) of integrals $I_t(f_n)$ of simple functions f_n, $n \geqslant 1$, approximating the integrand f.

It turned out (see [244] or, for greater detail, [303; Chapter 4]) that if $f \in J_2$, then there exists a sequence of simple functions $f_n = f_n(t,\omega)$ such that

$$\mathsf{E}\int_0^t [f(s,\omega) - f_n(s,\omega)]^2\,ds \to 0 \tag{11}$$

for each $t \geqslant 0$. Hence if $f \in J_2$, then

$$\mathsf{E}\int_0^t [f_n(s,\omega) - f_m(s,\omega)]^2\,ds \to 0 \tag{12}$$

as $m,n \to \infty$.

By (8) we obtain the *isometry* relation

$$\mathsf{E}\big[I_t(f_n) - I_t(f_m)\big]^2 = \mathsf{E}\int_0^t [f_n(s,\omega) - f_m(s,\omega)]^2\,ds. \tag{13}$$

Taken together with (12), it shows that the random variables $\{I_t(f_n)\}_{n\geqslant 1}$ form a Cauchy sequence in the sense of convergence in mean square (i.e., in L^2).

Hence, by the Cauchy criterion for L^2-convergence (see, e.g., [439; Chapter II, § 10]) there exists a random variable in L^2, which we denote by $I_t(f)$, such that

$$I_t(f) = \text{l.i.m.}\, I_t(f_n),$$

i.e.,

$$\mathsf{E}\big[I_t(f) - I_t(f_n)\big]^2 \to 0 \quad \text{as} \quad n \to \infty.$$

This limit, $I_t(f)$, which is easily seen to be independent of the choice of an approximating sequence $(f_n)_{n\geqslant 1}$, is also denoted by $\int_0^t f(s,\omega)\,dB_s$ and is called the *stochastic integral* of the (nonanticipating) function $f = f(s,\omega)$ with respect to the Brownian motion $B = (B_s)_{s\geqslant 0}$ over the time interval $(0,t]$.

We now discuss the properties of the integrals $I_t(f)$ $(t > 0)$ so defined for functions $f \in J_2$; the details can be found, e.g., in [123], [250], [288], or [303].

(a) If $f, g \in J_2$ and a and b are constants, then

$$I_t(af + bg) = aI_t(f) + bI_t(g).$$

(b) One can choose modifications of the variables $I_t(f)$, $t > 0$, coordinated so that $I(f) = (I_t(f))_{t\geqslant 0}$ with $I_0(f) = 0$ is a continuous stochastic process (this is precisely the modification we shall consider throughout). Moreover,

$$I_s(f) = I_t(fI_{(0,s]}), \qquad s \leqslant t. \tag{14}$$

(c) If $\tau = \tau(\omega)$ is a Markov time such that $\tau(\omega) \leqslant T$, then

$$I_\tau(f) = I_T(fI_{(0,\tau]}), \tag{15}$$

where, by definition, $I_\tau(f) = I_{\tau(\omega)}(f)$.

(d) The process $I(f) = (I_t(f))_{t\geqslant 0}$ is a square integrable martingale, i.e.,

$$\begin{aligned}
&I_t(f) \text{ are } \mathscr{F}_t\text{-measurable}, \quad t > 0;\\
&\mathsf{E}I_t^2(f) < \infty, \quad t > 0;\\
&\mathsf{E}\big(I_t(f) \,|\, \mathscr{F}_s\big) = I_s(f).
\end{aligned}$$

In addition, if $f, g \in J_2$, then

$$\mathsf{E}I_t(f)I_t(g) = \mathsf{E}\int_0^t f(s,\omega)g(s,\omega)\,ds. \tag{16}$$

Remark 2. By analogy with the notation used for discrete time (see Definition 7 in Chapter II, § 1c), one often denotes $I_t(f)$ also by $(f \cdot B)_t$ (cf. [250; Chapter I, § 4d]).

We now proceed to the definition of stochastic integrals $I_t(f)$ for functions in the class J_1; we refer to [303; Chapter 4, § 4] for more detail.

Let $f \in J_1$, i.e., let $\mathsf{P}\Big(\int_0^t f^2(s,\omega)\,ds < \infty\Big) = 1$, $t > 0$. Then there exists a sequence of functions $f^{(m)} \in J_2$, $m \geqslant 1$, such that

$$\int_0^t \big[f(s,\omega) - f^{(m)}(s,\omega)\big]^2\,ds \xrightarrow{\mathsf{P}} 0$$

for each $t > 0$ (the symbol '$\xrightarrow{\mathsf{P}}$' indicates convergence in probability).

Since

$$\mathsf{P}\big\{|I_t(f^{(m)}) - I_t(f^{(n)})| > \delta\big\} \leqslant \frac{\varepsilon}{\delta^2} + \mathsf{P}\bigg(\int_0^t \big[f^{(m)}(s,\omega) - f^{(n)}(s,\omega)\big]^2\,ds > \varepsilon\bigg)$$

for $\varepsilon > 0$ and $\delta > 0$, letting first $m, n \to \infty$ and then $\varepsilon \downarrow 0$ we obtain

$$\lim_{n,m\to\infty} \mathsf{P}\big\{|I_t(f^{(m)}) - I_t(f^{(n)})| > \delta\big\} = 0$$

for each $\delta > 0$. Hence the sequence $I_t(f^{(m)})$, $m \geqslant 1$, is *fundamental in probability*, and by the corresponding Cauchy criterion ([439; Chapter II, § 10]) there exists a random variable $I_t(f)$ such that

$$I_t(f^{(m)}) \xrightarrow{\mathsf{P}} I_t(f) \quad \text{as} \quad m \to \infty.$$

The variable $I_t(f)$ (denoted also by $(f \cdot B)_t$, $\int_{(0,t]} f(s,\omega)\,dB_s$, or $\int_0^t f(s,\omega)\,dB_s$) is called the stochastic integral of f over $(0,t]$.

We now mention several properties of the stochastic integrals $I_t(f)$, $t > 0$, for $f \in J_1$.

It can be shown that, again, we can define the stochastic integrals $I_t(f)$, in a coordinated manner for different $t > 0$ so that the process $I(f) = (I_t(f))_{t\geqslant 0}$ has (P-a.s.) continuous trajectories.

The above-mentioned properties (a), (b), (c) holding in the case of $f \in J_2$ also persist for f in J_1. However, (d) does not hold any more in general. It can be replaced now by the following property:

(d′) for $f \in J_1$ the process $I(f) = (I_t(f))_{t\geqslant 0}$ is a *local martingale*, i.e., there exists a sequence $(\tau_n)_{n\geqslant 1}$ of Markov times such that $\tau_n \uparrow \infty$ as $n \to \infty$ and the 'stopped' processes

$$I^{\tau_n}(f) = \big(I_{t\wedge\tau_n}(f)\big)_{t\geqslant 0}$$

are martingales for each $n \geqslant 1$ (cf. Definition 4 in Chapter II, § 1c.)

4. Let $B = (B_t)_{t\geqslant 0}$ be a Brownian motion defined on a probability space $(\Omega, \mathscr{F}, \mathsf{P})$ and let $(\mathscr{F}_t)_{t\geqslant 0}$ be the family of σ-algebras generated by this process (see § 3b.2; for more precision we shall also denote $\mathscr{F}_t$ by $\mathscr{F}_t^B$, $t \geqslant 0$).

The following theorem, based on the concept of a stochastic integral, describes the structure of *Brownian functionals*.

THEOREM 1. *Let $X = X(\omega)$ be a $\mathscr{F}_T^B$-measurable random variable.*

1. *If $\mathsf{E}X^2 < \infty$, then there exists a stochastic process $f = (f_t(\omega), \mathscr{F}_t^B)_{t\leqslant T}$ such that*

$$\mathsf{E}\int_0^T f_t^2(\omega)\,dt < \infty \tag{17}$$

and (P-a.s.)

$$X = \mathsf{E}X + \int_0^T f_t(\omega)\,dB_t. \tag{18}$$

2. *If* $\mathsf{E}|X| < \infty$, *then the representation* (18) *holds for some process* $f = (f_t(\omega), \mathscr{F}_t^B)_{t \leqslant T}$ *such that*

$$\mathsf{P}\biggl(\int_0^T f_t^2(\omega)\,dt < \infty\biggr) = 1. \tag{19}$$

3. *Let* $X = X(\omega)$ *be a positive random variable* ($\mathsf{P}(X>0) = 1$) *such that* $\mathsf{E}X < \infty$. *Then there exists a process* $\varphi = (\varphi_t(\omega), \mathscr{F}_t^B)_{t \leqslant T}$ *with* $\mathsf{P}\Bigl(\int_0^T \varphi_t^2(\omega)\,dt < \infty\Bigr) = 1$ *such that* (P-a.s.)

$$X = \mathsf{E}X \cdot \exp\biggl\{\int_0^T \varphi_t(\omega)\,dB_t - \frac{1}{2}\int_0^T \varphi_t^2(\omega)\,dt\biggr\}. \tag{20}$$

From this theorem we derive the following result on the structure of *Brownian martingales*.

THEOREM 2. 1. *Let* $M = (M_t, \mathscr{F}_t^B)_{t \leqslant T}$ *be a square integrable martingale. Then there exists a process* $f = (f_t(\omega), \mathscr{F}_t^B)_{t \leqslant T}$ *satisfying* (17) *such that*

$$M_t = M_0 + \int_0^t f_s(\omega)\,dB_s. \tag{21}$$

2. *Let* $M = (M_t, \mathscr{F}_t^B)_{t \leqslant T}$ *be a local martingale. Then the representation* (21) *holds for some process* $f = (f_t(\omega), \mathscr{F}_t^B)_{t \leqslant T}$ *satisfying* (19).

3. *Let* $M = (M_t, \mathscr{F}_t^B)_{t \leqslant T}$ *be a positive local martingale. Then there exists a process* $\varphi = (\varphi_t(\omega), \mathscr{F}_t^B)_{t \leqslant T}$ *such that* $\mathsf{P}\Bigl(\int_0^T \varphi_t^2(\omega)\,dt < \infty\Bigr) = 1$ *and*

$$M_t = M_0 \exp\biggl\{\int_0^t \varphi_s(\omega)\,dB_s - \frac{1}{2}\int_0^t \varphi_s^2(\omega)\,ds\biggr\}.$$

The proofs of these theorems, which are mainly due to J. M. C. Clark [77], but were also in different versions proved by K. Itô [244] and J. Doob [109], [110], can be found in many books; see, e.g., [266], [303], or [402].

§ 3d. Itô Processes and Itô's Formula

1. The above definition of a stochastic integral plays a key role in distinguishing the following important class of stochastic processes.

We shall say that a stochastic process $X = (X_t)_{t \geqslant 0}$ defined on a filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t \geqslant 0}, \mathsf{P})$ satisfying the *usual* conditions (see § 3b.3), is an *Itô process* if there exist two nonanticipating processes $a = (a(t,\omega))_{t \geqslant 0}$ and $b = (b(t,\omega))_{t \geqslant 0}$

such that

$$\mathsf{P}\left(\int_0^t |a(s,\omega)|\,ds < \infty\right) = 1, \qquad t > 0, \tag{1}$$

$$\mathsf{P}\left(\int_0^t b^2(s,\omega)\,ds < \infty\right) = 1, \qquad t > 0, \tag{2}$$

and

$$X_t = X_0 + \int_0^t a(s,\omega)\,ds + \int_0^t b(s,\omega)\,dB_s, \tag{3}$$

where $B = (B_t, \mathscr{F}_t)_{t\geqslant 0}$ is a Brownian motion and X_0 is a $\mathscr{F}_0$-measurable random variable.

For brevity, one uses in the discussion of Itô processes the following (formal) differential notation in place of the integral notation (3):

$$dX_t = a(t,\omega)\,dt + b(t,\omega)\,dB_t; \tag{4}$$

here one says that the process $X = (X_t)_{t\geqslant 0}$ has the *stochastic differential* (4).

2. Now let $F(t,x)$ be a function from the class $C^{1,2}$ defined in $\mathbb{R}_+ \times \mathbb{R}$ $\left(\text{i.e., } F \text{ is a function with continuous derivatives } \dfrac{\partial F}{\partial t}, \dfrac{\partial F}{\partial x}, \text{ and } \dfrac{\partial^2 F}{\partial x^2}\right)$ and let $X = (X_t)_{t\geqslant 0}$ be a process with differential (4).

Under these assumptions, as proved by Itô, the process $F = (F(t, X_t))_{t\geqslant 0}$ also has a stochastic differential and

$$dF(t,X_t) = \left[\frac{\partial F}{\partial t} + a(t,\omega)\frac{\partial F}{\partial x} + \frac{1}{2}b^2(t,\omega)\frac{\partial^2 F}{\partial x^2}\right]dt + \frac{\partial F}{\partial x}b(t,\omega)\,dB_t. \tag{5}$$

More precisely, for each $t > 0$ we have the following *Itô's formula* (*formula of the change of variables*) for $F(t, X_t)$:

$$F(t,X_t) = F(0,X_0) + \int_0^t \left[\frac{\partial F}{\partial s} + a(s,\omega)\frac{\partial F}{\partial x} + \frac{1}{2}b^2(s,\omega)\frac{\partial^2 F}{\partial x^2}\right]ds + \int_0^t \frac{\partial F}{\partial x}b(s,\omega)\,dB_s. \tag{6}$$

(The proof can be found in many places; see, e.g., [123], [250; Chapter I, § 4e], or [303; Chapter 4, § 3].)

3. We also present a generalization of (6) to several dimensions.

We assume that $B = (B^1, \dots, B^d)$ is a d-dimensional Brownian motion with independent (one-dimensional) Brownian components $B^i = (B^i_t)_{t\geqslant 0}$, $i = 1, \dots, d$.

We say that the process $X = (X^1, \dots, X^d)$ with $X^i = (X^i_t)_{t\geqslant 0}$ is a *d-dimensional Itô process* if there exists a vector $a = (a^1, \dots, a^d)$ and a matrix $b = \|b^{ij}\|$ of order $d \times d$ with nonanticipating components $a^i = a^i(t,\omega)$ and entries $b^{ij} = b^{ij}(t,\omega)$ such that

$$\mathsf{P}\left(\int_0^t |a^i(s,\omega)|\,ds < \infty\right) = 1,$$
$$\mathsf{P}\left(\int_0^t \left(b^{ij}(s,\omega)\right)^2 ds < \infty\right) = 1$$

for $t > 0$ and

$$dX^i_t = a^i(t,\omega)\,dt + \sum_{j=1}^d b^{ij}(t,\omega)\,dB^j_t \tag{7}$$

for $i = 1, \dots, d$, or, in the vector notation,

$$dX_t = a(t,\omega)\,dt + b(t,\omega)\,dB_t,$$

where $a(t,\omega) = (a^1(t,\omega), \dots, a^d(t,\omega))$ and $B_t = (B^1_t, \dots, B^d_t)$ are column vectors.

Now let $F(t, x_1, \dots, x_d)$ be a continuous function with continuous derivatives $\dfrac{\partial F}{\partial t}$, $\dfrac{\partial F}{\partial x_i}$, and $\dfrac{\partial^2 F}{\partial x_i \partial x_j}$; $i, j = 1, \dots, d$.

Then we have the following *d-dimensional version of Itô's formula*:

$$\begin{aligned}
F(t, X^1_t, \dots, X^d_t) &= F(0, X^1_0, \dots, X^d_0) \\
&\quad + \int_0^t \left[\frac{\partial F}{\partial s}(s, X^1_s, \dots, X^d_s) + \sum_{i=1}^d \frac{\partial F}{\partial x_i}(s, X^1_s, \dots, X^d_s)\, a^i(s,\omega)\right. \\
&\quad \left. + \frac{1}{2}\sum_{i,j=1}^d \left(\frac{\partial^2 F}{\partial x_i \partial x_j}(s, X^1_s, \dots, X^d_s) \sum_{k=1}^d b^{ik}(s,\omega)\, b^{jk}(s,\omega)\right)\right] ds \\
&\quad + \int_0^t \sum_{i,j=1}^d \frac{\partial F}{\partial x_i}(s, X^1_s, \dots, X^d_s)\, b^{ij}(s,\omega)\, dB^j_s.
\end{aligned} \tag{8}$$

For continuous processes $X^i = (X^i_t)_{t\geqslant 0}$ we shall use the notation $\langle X^i, X^j\rangle = \left(\langle X^i, X^j\rangle_t\right)_{t\geqslant 0}$, where

$$\langle X^i, X^j\rangle_t = \sum_{k=1}^d \int_0^t b^{ik}(s,\omega)\, b^{jk}(s,\omega)\, ds.$$

Then (8) can be rewritten in the following compact form:

$$F(t, X_t) = F(0, X_0) + \int_0^t \frac{\partial F}{\partial s}(s, X_s)\, ds$$
$$+ \sum_{i=1}^d \int_0^t \frac{\partial F}{\partial x_i}(s, X_s)\, dX_s^i + \frac{1}{2} \sum_{i,j=1}^d \int_0^t \frac{\partial^2 F}{\partial x_i \partial x_j}(s, X_s)\, d\langle X^i, X^j \rangle_s. \quad (9)$$

Using differentials, this can be rewritten again, as

$$dF = \frac{\partial F}{\partial t}\, dt + \sum_{i=1}^d \frac{\partial F}{\partial x_i}\, dX_t^i + \frac{1}{2} \sum_{i,j=1}^d \frac{\partial^2 F}{\partial x_i \partial x_j}\, d\langle X^i, X^j \rangle_t \quad (10)$$

(here $F = F(t, X_t)$).

It is worth noting that if we formally write (using Taylor's formula)

$$dF = \frac{\partial F}{\partial s}\, dt + \sum_{i=1}^d \frac{\partial F}{\partial x_i}\, dX_t^i + \frac{1}{2} \sum_{i,j=1}^d \frac{\partial^2 F}{\partial x_i \partial x_j}\, dX_t^i\, dX_t^j, \quad (11)$$

and agree that

$$(dB_t^i)^2 = dt, \quad (12)$$

$$dB_t^i\, dt = 0, \quad (13)$$

$$dB_t^i\, dB_t^j = 0, \quad i \neq j, \quad (14)$$

then we obtain (10) from (11) because

$$dX_t^i\, dX_t^j = d\langle X^i, X^j \rangle_t. \quad (15)$$

Remark 1. The formal expressions (15) and (14) can be interpreted quite meaningfully, as symbolically written limit relations (2) and (3) from the preceding section, § 3b. We can also interpret relation (13) in a similar way.

For the proof of Itô's formula in several dimensions see, e.g., the monograph [250; Chapter I, § 4e]. As regards generalizations of this formula to functions $F \notin C^{1,2}$, see, e.g., [166] and [402].

4. We now present several examples based on Itô's formula.

Example 1. Let $F(x) = x^2$ and let $X_t = B_t$. Then, by (11), we formally obtain

$$dB_t^2 = 2B_t\, dB_t + (dB_t)^2.$$

In view of (12), we see that

$$dB_t^2 = 2B_t\, dB_t + dt, \quad (16)$$

or, in the integral form,

$$B_t^2 = 2\int_0^t B_s\, dB_s + t. \quad (17)$$

EXAMPLE 2. Let $F(x) = e^x$ and $X_t = B_t$. Then

$$d(e^{B_t}) = e^{B_t}\,dB_t + \frac{1}{2}e^{B_t}(dB_t)^2,$$

i.e., bearing in mind (11),

$$d(e^{B_t}) = e^{B_t}\Big(dB_t + \frac{1}{2}\,dt\Big). \tag{18}$$

Let $F(t,x) = e^{x-\frac{1}{2}t}$ and let $X_t = B_t$. Then we formally obtain

$$dF(t,B_t) = -\frac{1}{2}F(t,B_t)\,dt + F(t,B_t)\,dB_t + \frac{1}{2}F(t,B_t)(dB_t)^2.$$

In view of (12), we obtain

$$dF(t,B_t) = F(t,B_t)\,dB_t.$$

Considering the *stochastic exponential*

$$\mathscr{E}(B)_t = e^{B_t-\frac{1}{2}t} \tag{19}$$

(cf. formula (13) in Chapter II, § 1a), we see that it has the stochastic differential

$$d\mathscr{E}(B)_t = \mathscr{E}(B)_t\,dB_t. \tag{20}$$

This relation can be treated as a stochastic differential equation (see § 3a and § 3e further on), with a solution delivered by (19).

EXAMPLE 3. Putting the preceding example in a broader context we consider now the process

$$Z_t = \exp\bigg\{\int_0^t b(s,\omega)\,dB_s - \frac{1}{2}\int_0^t b^2(s,\omega)\,ds\bigg\}, \tag{21}$$

where $b = (b(t,\omega))_{t\geqslant 0}$ is a nonanticipating process with

$$\mathsf{P}\bigg(\int_0^t b^2(s,\omega)\,ds < \infty\bigg) = 1, \qquad t > 0.$$

Setting

$$X_t = \int_0^t b(s,\omega)\,dB_s - \frac{1}{2}\int_0^t b^2(s,\omega)\,ds$$

and $F(x) = e^x$ we can use Itô's formula (5) to see that the process $Z = (Z_t)_{t\geqslant 0}$ (the 'Girsanov exponential') has the stochastic differential

$$dZ_t = Z_t b(t,\omega)\,dB_t. \tag{22}$$

EXAMPLE 4. Let $X_t = B_t$ and let $Y_t = t$. Then

$$d(B_t\, t) = t\, dB_t + B_t\, dt,$$

or, in the integral form,

$$B_t\, t = \int_0^t s\, dB_s + \int_0^t B_s\, ds \tag{23}$$

$\Big($cf. N. Wiener's definition of the stochastic integral $\int_0^t s\, dB_s$, in § 3c.1.$\Big)$

EXAMPLE 5. Let $F(x_1, x_2) = x_1 x_2$ and let $X^1 = (X_t^1)_{t \geqslant 0}$, $X^2 = (X_t^2)_{t \geqslant 0}$ be two processes having Itô differentials. Then we formally obtain

$$d(X_t^1 X_t^2) = X_t^1\, dX_t^2 + X_t^2\, dX_t^1 + dX_t^1\, dX_t^2. \tag{24}$$

In particular, if

$$dX_t^i = a^i(t, \omega)\, dt + b^i(t, \omega)\, dB_t^i, \qquad i = 1, 2,$$

then

$$d(X_t^1 X_t^2) = X_t^1\, dX_t^2 + X_t^2\, dX_t^1. \tag{25}$$

On the other hand, if

$$dX_t^i = a^i(t, \omega)\, dt + b^i(t, \omega)\, dB_t, \qquad i = 1, 2,$$

that

$$d(X_t^1 X_t^2) = X_t^1\, dX_t^2 + X_t^2\, dX_t^1 + b^1(t, \omega) b^2(t, \omega)\, dt. \tag{26}$$

EXAMPLE 6. Let $X = (X^1, \dots, X^d)$ be a d-dimensional Itô process whose components X^i have stochastic differentials (7).

Let $V = V(x)$ be a real-valued function of $x = (x_1, \dots, x_d)$ with continuous second derivatives and let

$$(L_t V)(x, \omega) = \sum_{i=1}^d a^i(t, \omega) \frac{\partial V}{\partial x_i} + \sum_{i,j=1}^d \left(\sum_{k=1}^d b^{ik}(t, \omega)\, b^{jk}(t, \omega) \right) \frac{\partial^2 V}{\partial x_i \partial x_j}. \tag{27}$$

Then the process $(V(X_t))_{t \geqslant 0}$ has the stochastic differential

$$dV(X_t) = (L_t V)(X_t, \omega)\, dt + \frac{\partial V}{\partial x}(X_t)\, b(t, \omega)\, dB_t \tag{28}$$

(we use the matrix notation), where

$$\frac{\partial V}{\partial x}(X_t)\, b(t, \omega)\, dB_t = \sum_{i,j=1}^d \frac{\partial V}{\partial x_i}(X_t)\, b^{ij}(t, \omega)\, dB_t^j. \tag{29}$$

EXAMPLE 7. Let $V = V(t,x)$ be a continuous real-valued function in $[0,\infty) \times \mathbb{R}^d$ with continuous derivatives $\dfrac{\partial V}{\partial t}$, $\dfrac{\partial V}{\partial x_i}$, and $\dfrac{\partial^2 V}{\partial x_i \partial x_j}$. Also, let

$$\varphi_t = \int_0^t C(s,\omega)\,ds,$$

where $C = C(t,\omega)$ is a nonanticipating function with

$$\mathsf{P}\biggl(\int_0^t |C(s,\omega)|\,ds < \omega\biggr) = 1, \qquad t > 0.$$

Then $(e^{-\varphi_t} V(t,X_t))_{t\geqslant 0}$ is an Itô process with stochastic differential

$$d\bigl(e^{-\varphi_t} V(t,X_t)\bigr) = e^{-\varphi_t}\biggl[\frac{\partial V}{\partial t}(t,X_t) + (L_t V)(X_t,\omega) - C(t,\omega)V(t,X_t)\biggr]\,dt + e^{-\varphi_t}\,\frac{\partial V}{\partial x}(t,X_t)\,b(t,\omega)\,dB_t. \tag{30}$$

5. *Remark* 2. Let $X = (X_t)_{t\geqslant 0}$ be a *diffusion Markov process* with stochastic differential

$$dX_t = a(t,X_t)\,dt + b(t,X_t)\,dB_t,$$

where $\int_0^t \bigl|a(s,X_s)\bigr|\,ds < \infty$, $\int_0^t b^2(s,X_s)\,ds < \infty$ (P-a.s.), and $t > 0$ (cf. formulas (1) in § 3a and (9) in § 3e).

If $Y_t = F(t,X_t)$, where $F = F(t,x) \in C^{1,2}$ and $\frac{\partial F}{\partial x} > 0$, then $Y = (Y_t)_{t\geqslant 0}$ is also a diffusion Markov process with

$$dY_t = \alpha(t,Y_t)\,dt + \beta(t,Y_t)\,dB_t,$$

where

$$\alpha(t,y) = \frac{\partial F}{\partial t}(t,x) + \frac{\partial F}{\partial x}(t,x)\,a(t,x) + \frac{1}{2}\,\frac{\partial^2 F}{\partial x^2}(t,x)\,b^2(t,x), \tag{31}$$

$$\beta(t,y) = \frac{\partial F(t,x)}{\partial x}\,b(t,x) \tag{32}$$

and t, x, and y are related by the equality $y = F(t,x)$.

These formulas, which describe the transformation of the (local) characteristics $a(t,x)$ and $b(t,x)$ of the Markov process X into the (local) characteristics $\alpha(t,y)$ and $\beta(t,y)$ of the Markov process Y, have been obtained by A. N. Kolmogorov [280; § 17] as long ago as 1931. They are consequences of Itô's formula (5). It would be natural for that reason to call (31) and (32) the *Kolmogorov–Itô formulas.*

§ 3e. Stochastic Differential Equations

1. Among Itô processes $X = (X_t)_{t \geqslant 0}$ with stochastic differentials

$$dX_t = \alpha(t, \omega)\, dt + \beta(t, \omega)\, dB_t, \tag{1}$$

a major role is played by processes such that the dependence of the corresponding coefficients $\alpha(t, \omega)$ and $\beta(t, \omega)$ on ω factors through the process $X_t(\omega)$ itself, i.e.,

$$\alpha(t, \omega) = a\big(t, X_t(\omega)\big), \qquad \beta(t, \omega) = b\big(t, X_t(\omega)\big), \tag{2}$$

where $a = a(t, x)$ and $b = b(t, x)$ are measurable functions on $\mathbb{R}_+ \times \mathbb{R}$.

For instance, the process

$$S_t = S_0 e^{at}\, e^{\sigma B_t - \frac{\sigma^2}{2} t}, \tag{3}$$

which is called a geometric (or economic) Brownian motion (see § 3a) has (in accordance with Itô's formula) the stochastic differential

$$dS_t = aS_t\, dt + \sigma S_t\, dB_t. \tag{4}$$

It is easy to verify using the same Itô formula that the process

$$Y_t = \int_0^t \frac{S_t}{S_u}\, du \tag{5}$$

has the differential

$$dY_t = (1 + aY_t)\, dt + \sigma Y_t\, dB_t. \tag{6}$$

(This process, $Y = (Y_t)_{t \geqslant 0}$, plays an important role in problems of the *quickest detection* of changes in the local drift of a Brownian motion; see [440], [441].)

If

$$Z_t = S_t \left[Z_0 + (c_1 - \sigma c_2) \int_0^t \frac{du}{S_u} + c_2 \int_0^t \frac{dB_u}{S_u} \right] \tag{7}$$

for some constants c_1 and c_2, then, using Itô's formula again, we can verify that

$$dZ_t = (c_1 + aZ_t)\, dt + (c_2 + \sigma Z_t)\, dB_t. \tag{8}$$

In the above examples we started from 'explicit' formulas for the processes $S = (S_t)$, $Y = (Y_t)$, and $Z = (Z_t)$ and found their stochastic differentials (4), (6), and (8) using Itô's formula.

However, one can change the standpoint; namely, one can regard (4), (6), and (8) as stochastic differential equations with respect to unknown processes $S = (S_t)$, $Y = (Y_t)$, $Z = (Z_t)$ and can attempt to prove that their solutions (3), (5), and (7) are unique (in one or another sense).

Of course, we must assign a precise meaning to the concept of 'stochastic differential equation', define its 'solution', and explain what the 'uniqueness' of a solution means. The above-considered notion of a stochastic integral will play a key role in our introduction of all these concepts.

2. Let $B = (B_t, \mathscr{F}_t)_{t \geqslant 0}$ be a Brownian motion defined on a filtered probability space (stochastic basis) $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t \geqslant 0}, \mathsf{P})$ satisfying the usual conditions (§ 7a.2). Let $a = a(t, x)$ and $b = b(t, x)$ be measurable functions on $\mathbb{R}_+ \times \mathbb{R}$.

DEFINITION 1. We say that a stochastic differential equation

$$dX_t = a(t, X_t)\,dt + b(t, X_t)\,dB_t \tag{9}$$

with $\mathscr{F}_0$-measurable initial condition X_0 has a continuous *strong solution* (or simply a *solution*) $X = (X_t)_{t \geqslant 0}$ if for each $t > 0$

$$X_t \text{ are } \mathscr{F}\text{-measurable},$$

$$\mathsf{P}\Big(\int_0^t |a(s, X_s)|\,ds < \infty\Big) = 1, \tag{10}$$

$$\mathsf{P}\Big(\int_0^t b^2(s, X_s)\,ds < \infty\Big) = 1 \tag{11}$$

and (P-a.s.)

$$X_t = X_0 + \int_0^t a(s, X_s)\,ds + \int_0^t b(s, X_s)\,dB_s. \tag{12}$$

DEFINITION 2. We say that two continuous stochastic processes $X = (X_t)_{t \geqslant 0}$ and $Y = (Y_t)_{t \geqslant 0}$ are *stochastically indistinguishable* if

$$\mathsf{P}\Big(\sup_{s \leqslant t} |X_s - Y_s| > 0\Big) = 0 \tag{13}$$

for each $t > 0$.

DEFINITION 3. We say that a measurable function $f = f(t, x)$ on $\mathbb{R}_+ \times \mathbb{R}$ satisfies the *local Lipschitz condition* (with respect to the phase variable x) if for each $n \geqslant 1$ there exists a quantity $K(n)$ such that

$$|a(t, x) - a(t, y)| + |b(t, x) - b(t, y)| \leqslant K(n)|x - y| \tag{14}$$

for all $t \geqslant 0$ and x, y such that $|x| \leqslant n$ and $|y| \leqslant n$.

THEOREM 1 (K. Itô [242], [243]; see also, e.g., [123: Chapter 9], [288; Chapter V], or [303; Chapter 4]). *Assume that the coefficients $a(t, x)$ and $b(t, x)$ satisfy the local Lipschitz condition and the condition of linear growth*

$$|a(t, x)| + |b(t, x)| \leqslant K(1)|x|, \tag{14$'$}$$

and that the initial value X_0 is $\mathscr{F}_0$-measurable.

Then the stochastic differential equation (9) *has a (unique, up to stochastic indistinguishability) continuous solution $X = (X_t, \mathscr{F}_t)$, which is a Markov process.*

This result can be generalized in various directions: the local Lipschitz conditions can be weakened, the coefficients may depend on ω (although in a special way), one can consider the case when the coefficients $a = a(t, X_t)$ and $b = b(t, X_t)$ depend on the 'past' (slightly abusing the notation, we write in this case $a = a(t; X_s,\ s \leqslant t)$ and $b = b(t; X_s,\ s \leqslant t)$).

There also exist generalization to several dimensions, when $X = (X^1, \dots, X^d)$ is a multivariate process, $a = a(t, x)$ is a vector, $b = b(t, x)$ is a matrix, and $B = (B^1, \dots, B^d)$ is a d-dimensional Brownian motion; see, e.g., [123], [288], and [303] on this subject.

Of all the generalizations we present here only one, somewhat unexpected result of A.K. Zvonkin [485], which shows that the local Lipschitz condition is not necessary for the existence of a strong solution of a stochastic differential equation

$$dX_t = a(t, X_t)\, dt + dB_t; \tag{15}$$

the mere measurability with respect to (t, x) and the uniform boundedness of the coefficient $a(t, x)$ suffice. (A multidimensional generalization of this result was proved by A. Yu. Veretennikov [471].)

Hence, for example, the stochastic differential equation

$$dX_t = \sigma(X_t)\, dt + dB_t, \qquad X_0 = 0, \tag{16}$$

with 'bad' coefficient

$$\sigma(x) = \begin{cases} 1, & x \geqslant 0, \\ -1, & x < 0, \end{cases} \tag{17}$$

has a strong solution (moreover, a unique one).

Note, however, that if in place of (16) we consider the equation

$$dX_t = \sigma(X_t)\, dB_t, \qquad X_0 = 0 \tag{18}$$

with the same $\sigma(x)$, then the situation changes drastically, because, first, this equation is known to have *at least two strong solutions* on some probability spaces. Second, this equation has no strong solutions whatsoever on *some* other probability spaces.

To prove the first assertion we consider a *coordinate* Wiener process $W = (W_t)_{t \geqslant 0}$ in the space of continuous functions $\omega = (\omega_t)_{t \geqslant 0}$ on $[0, +\infty)$ endowed with Wiener measure, i.e., a process defined as $W_t(\omega) = \omega_t, t \geqslant 0$.

By Lévy's theorem (see § 3b.3), if

$$B_t = \int_0^t \sigma(W_s)\, dW_s,$$

then $B = (B_t)_{t \geqslant 0}$ is also a Wiener process (a Brownian motion). Moreover, it is easy to see that

$$\int_0^t \sigma(W_s)\, dB_s = \int_0^t \sigma^2(W_s)\, dW_s = W_t,$$

because $\sigma^2(x) = 1$.

Hence the process $W = (W_t)_{t\geqslant 0}$ is a solution of equation (18) (in our coordinate probability space) with a specially designed Brownian motion B. However, $\sigma(-x) = -\sigma(x)$, so that

$$\int_0^t \sigma(-W_s)\,dB_s = -\int_0^t \sigma(W_s)\,dB_s = -W_t,$$

i.e., the process $-W = (-W_t)_{t\geqslant 0}$ is also a solution of (18).

As regards the second assertion, assume that the equation

$$X_t = \int_0^t \sigma(X_s)\,dB_s$$

has a strong solution (with respect to the flow of σ-algebras $(\mathscr{F}_t^B)_{t\geqslant 0}$ generated by a Brownian motion B). By Lévy's theorem, this process $X = (X_t, \mathscr{F}_t^B)_{t\geqslant 0}$ is a Brownian motion.

By Tanaka's formula (see § 5c or [402] and compare with the example in Chapter II, § 1b)

$$|X_t| = \int_0^t \sigma(X_s)\,dX_s + L_t(0),$$

where

$$L_t(0) = \lim_{\varepsilon\downarrow 0} \frac{1}{2\varepsilon} \int_0^t I\big(|X_s| \leqslant \varepsilon\big)\,ds$$

is the *local time* (P. Lévy) of the Brownian motion X at the origin over the period $[0,t]$.

Hence (P-a.s.)

$$B_t = \int_0^t \sigma(X_s)\,dX_s = |X_t| - L_t(0),$$

so that $\mathscr{F}_t^B \subseteq \mathscr{F}_t^{|X|}$.

The above assumption that X is adapted to the flow $\mathscr{F}^B = (\mathscr{F}_t^B)_{t\geqslant 0}$ ensures the inclusion $\mathscr{F}_t^X \subseteq \mathscr{F}_t^{|X|}$, which, of course, cannot occur for a Brownian motion X. All this shows that an equation does not necessarily have a solution for an *arbitrarily* chosen probability space and an arbitrary Brownian motion. (M. Barlow [20] showed that (18) does not necessarily have a strong solution even in the case of a bounded continuous function $\sigma = \sigma(X) > 0$.)

3. Note that, in fact, the two above-obtained solutions of (18), W and $-W$, have the same distribution, i.e.,

$$\operatorname{Law}(W_s, s \geqslant 0) = \operatorname{Law}(-W_s, s \geqslant 0).$$

This can be regarded as a justification for our introducing below the concept of *weak solutions* of stochastic differential equations.

DEFINITION 4. Let μ be a probability Borel measure on the real line $\mathbb{R}$. We say that a stochastic differential equation (9) with initial data X_0 satisfying the relation $\operatorname{Law}(X_0) = \mu$ has a *weak solution* if *there exist* a filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geqslant 0}, \mathsf{P})$, a Brownian motion $B = (B_t, \mathscr{F}_t)_{t\geqslant 0}$ on this space, and a continuous stochastic process $X = (X_t, \mathscr{F}_t)_{t\geqslant 0}$ such that $\operatorname{Law}(X_0 \,|\, \mathsf{P}) = \mu$ and (12) holds (P-a.s.) for each $t > 0$.

It should be pointed out that, by contrast to a strong solution, which is sought on a particular filtered probability space endowed with a particular Brownian motion, we do not fix these objects (the probability space and the Brownian motion) in the definition of a weak solution. We only require that they exist.

It is clear from the above definitions that one can expect a weak solution to exist under less restrictive conditions on the coefficients of (9).

One of the first results in this direction (see [446], [457]) is as follows.

We consider a stochastic differential equation

$$dX_t = a(X_t)\,dt + b(X_t)\,dB_t \tag{19}$$

with initial distribution $\operatorname{Law}(X_0) = \mu$ such that $\int x^{2(1+\varepsilon)}\mu(dx) < \infty$ for some $\varepsilon > 0$.

If the coefficients $a = a(x)$ and $b = b(x)$ are *bounded continuous* functions, then equation (19) has a weak solution.

If, in addition, $b^2(x) > 0$ for $x \in \mathbb{R}$, then we have the uniqueness (in distribution) of the weak solution.

Remark 1. In fact, if $b(x)$ is bounded, continuous, and nowhere vanishing, then there exists a unique weak solution even if $a(x)$ is merely *bounded* and *measurable* (see [457]).

4. The above results concerning weak solutions can be generalized in various ways: to the multidimensional case, to the case of coefficients depending on the past, and so on.

One of the most transparent results in this direction is based on *Girsanov's theorem* on an absolutely continuous change of measure. In view of the importance of this theorem in many other questions, we present it here. (As regards the applications of this theorem in the discrete-time case, see Chapter V, § § 3b and 3d.)

Let $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geqslant 0}, \mathsf{P})$ be a filtered probability space and let $B = (B_t, \mathscr{F}_t)_{t\geqslant 0}$, $B = (B^1, \dots, B^d)$, be a d-dimensional Brownian motion. Let also $a = (a_t, \mathscr{F}_t)$, $a = (a^1, \dots, a^d)$, be a d-dimensional stochastic process such that

$$\mathsf{P}\left(\int_0^t \|a_t\|^2 < \infty\right) = 1, \qquad t \leqslant T, \tag{20}$$

where $\|a_t\|^2 = (a_t^1)^2 + \cdots + (a_t^d)^2$ and $T < \infty$.

We now construct a process $Z = (Z_t, \mathscr{F}_t)_{t\leqslant T}$ by setting

$$Z_t = \exp\left\{\int_0^t (a_s, dB_s) - \frac{1}{2}\int_0^t \|a_s\|^2\,ds\right\}, \tag{21}$$

where

$$(a_s, dB_s) = \sum_{i=1}^d a_s^i\,dB_s^i \quad (= a_s^*\,dB_s)$$

is the *scalar product.*

If

$$\mathsf{E}\exp\left\{\frac{1}{2}\int_0^t \|a_s\|^2\,ds\right\} < \infty$$

(the *Novikov condition*; cf. also the corresponding condition in the discrete-time case in Chapter V, § 3b), then

$$\mathsf{E}Z_T = 1, \tag{22}$$

so that $Z = (Z_t, \mathscr{F}_t)_{t\leqslant T}$ is a uniformly integrable martingale.

Since Z_T is positive (P-a.s.) and (22) holds, we can define a probability measure $\widetilde{\mathsf{P}}_T$ on $(\Omega, \mathscr{F}_T)$ by setting

$$\widetilde{\mathsf{P}}_T(A) = \mathsf{E}[I_A Z_T], \qquad A \in \mathscr{F}_T.$$

Clearly, if $\mathsf{P}_T = \mathsf{P} \,|\, \mathscr{F}_T$ is the restriction of the measure P to $\mathscr{F}_T$, then $\widetilde{\mathsf{P}}_T \sim \mathsf{P}_T$.

THEOREM 2 (I. V. Girsanov [183]). *Let*

$$\widetilde{B}_t = B_t - \int_0^t a_s\,ds, \qquad t \leqslant T.$$

Then $\widetilde{B} = (\widetilde{B}_t, \mathscr{F}_t, \widetilde{\mathsf{P}}_T)_{t\leqslant T}$ is a Brownian motion.

The proof can be found in [183] and in the present book, in Chapter VII, § 3b; see also in [266] or [303].

We now consider the question of the *existence* of a weak solution of the one-dimensional stochastic differential equation

$$dX_t = a(t, X)\,dt + dB_t, \tag{23}$$

where the coefficient $a = a(t, X)$ is in general assumed to depend on the 'past' variables X_s, $s \leqslant t$. (The case of $d > 1$ can be considered in a similar way.)

Let C be the space of continuous functions $x = (x_t)_{t \geqslant 0}$, $x_0 = 0$, let $\mathscr{C}_t = \sigma(x\colon x_s,\ s \leqslant t)$, and let $\mathscr{C} = \sigma\Big(\bigcup\limits_{t \geqslant 0} \mathscr{C}_t\Big)$. Also, let P^W be the Wiener measure in $(C, \mathscr{C})$.

We say that a functional $a = a(t, x)$, where $t \in \mathbb{R}_+$ and $x \in C$, is *measurable* if it is a measurable *map* from $\mathbb{R}_+ \times C$ into $\mathbb{R}$, and it is said to be *progressively measurable* if, in addition, for each $t > 0$ and each Borel set A we have the inclusion

$$\big\{(s \leqslant t, x \in C)\colon a(s, x) \in A\big\} \in \mathscr{B}([0, t]) \otimes \mathscr{C}_t.$$

We shall consider equation (23) for $t \leqslant T$ with $X_0 = 0$ (for simplicity) and shall assume that: $a = a(t, x)$ is a progressively measurable functional,

$$\mathsf{P}^W\bigg\{x\colon \int_0^t a^2(t, x)\,dt < \infty\bigg\} = 1, \tag{24}$$

and

$$\mathsf{E}^W \exp\bigg\{\int_0^T a(t, x)\,dW_t(x) - \frac{1}{2}\int_0^T a^2(t, x)\,dt\bigg\} = 1, \tag{25}$$

where $W = (W_t(x))_{t \geqslant 0}$ is the canonical Wiener process ($W_t(x) = x_t$) and $\mathsf{E}^W(\,\cdot\,)$ is averaging with respect to the measure P^W.

In accordance with our definition of a weak solution, we must *construct* a probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t \leqslant T}, \mathsf{P})$ and processes $X = (X_t, \mathscr{F}_t)$ and $B = (B_t, \mathscr{F}_t)$ on it such that B is a Brownian motion with respect to the measure P and

$$X_t = \int_0^t a(s, X)\,ds + B_t \tag{26}$$

(P-a.s.) for each $t \leqslant T$.

We now set

$$\Omega = C, \qquad \mathscr{F} = \mathscr{C}, \qquad \mathscr{F}_t = \mathscr{C}_t$$

and define a measure P in $\mathscr{F}_T$ by setting

$$\mathsf{P}_T(dx) = Z_T(x)\,\mathsf{P}^W(dx),$$

where

$$Z_T(x) = \exp\left\{\int_0^T a(t,x)\,dW_t(x) - \frac{1}{2}\int_0^T a^2(t,x)\,dt\right\}.$$

The process

$$B_t(x) = W_t(x) - \int_0^t a(s,W(x))\,ds, \qquad t \leqslant T$$

on the probability space $(\Omega, \mathscr{F}_T, \mathsf{P}_T)$ is a Brownian motion by Girsanov's theorem. Hence, setting $X_t(x) = W_t(x)$ we obtain

$$X_t(x) = \int_0^t a(s,X(x))\,ds + B_t(x), \qquad t \leqslant T,$$

which precisely proves the existence of a weak solution to the stochastic differential equation (23) (under assumptions (24) and (25)).

Remark 2. Conditions (24) and (25) hold for sure if $|a(t,x)| \leqslant c$ for all $t \leqslant T$ and $x \in C$. Hence equation (23) has a weak solution in this case. We point out, however, that such an equation does not necessarily have a strong solution, as shows an example suggested by B. Tsirelson (see, for instance, [303; § 4.4]). In this connection we recall that an equation $dX_t = a(t,X_t)\,dt + dB_t$ with coefficient $a(t,X_t)$ dependent *only* on the 'present' X_t, rather than on the entire 'past' X_s, $s \leqslant t$, (as in (23)) has not merely a weak solution, but a strong one (see subsection 2 above for a description of A. K. Zvonkin's result [485]).

§ 3f. Forward and Backward Kolmogorov's Equations. Probabilistic Representation of Solutions

1. Below we expose several results and methods of the theory of diffusion Markov processes that have their origin in A. N. Kolmogorov's fundamental paper "*Über die analitishen Methoden in der Warscheinlichkeitsrechnung*" [280] (1931).

P. S. Aleksandrov and A. Ya. Khintchine [5] wrote about this work, which revealed close relations of the theory of stochastic processes to mathematical analysis in general and the theory of differential equations (both ordinary and partial) in particular:

> "*In entire probability theory of the 20th century it is difficult to find another study as fundamental for the further development of the science ....*"

In [280] Kolmogorov does not consider the trajectories of, say, Markov processes $X = (X_t)_{t\geqslant 0}$. He studies instead the properties of the transition probabilities

$$\mathsf{P}(s,x;t,A) = \mathsf{P}(X_t \in A \mid X_s = x), \qquad x \in \mathbb{R}, \quad A \in \mathscr{B}(\mathbb{R}),$$

i.e., the probability of the event that a trajectory of X arrives in the set A at time t, provided that $X_s = x$ at time s.

The starting point for his analysis is the equation

$$\mathsf{P}(s, x; t, A) = \int_{\mathbb{R}} \mathsf{P}(s, x; u, dy)\, \mathsf{P}(u, y; t, A) \tag{1}$$

$(0 \leqslant s < u < t)$, which expresses the Markov property. (One usually calls (1) the *Kolmogorov–Chapman equation.*)

Assuming that there exist densities

$$f(s, x; t, y) = \frac{\partial F(s, x; t, y)}{\partial y}, \tag{2}$$

where

$$F(s, x; t, y) = \mathsf{P}(s, x; t, (-\infty, y]),$$

and limits

$$a(s, x) = \lim_{\Delta \downarrow 0} \frac{1}{\Delta} \int_{-\infty}^{\infty} (y - x) f(s, x; s + \Delta, y)\, dy, \tag{3}$$

$$b^2(s, x) = \lim_{\Delta \downarrow 0} \frac{1}{\Delta} \int_{-\infty}^{\infty} (y - x)^2 f(s, x; s + \Delta, y)\, dy, \tag{4}$$

and imposing certain regularity conditions on the functions in question and also the condition

$$\lim_{\Delta \downarrow 0} \frac{1}{\Delta} \int_{\mathbb{R}} |y - x|^{2+\delta} f(s, x; s + \Delta, y)\, dy = 0 \tag{5}$$

with $\delta > 0$, Kolmogorov derives from (1) the following *backward parabolic differential equations* (with respect to $x \in \mathbb{R}$ and $s < t$) for such *diffusion* processes (see [280] or, e.g., [170] and [182] for detail):

$$-\frac{\partial f}{\partial s} = a(s, x)\, \frac{\partial f}{\partial x} + \frac{1}{2}\, b^2(s, x) \frac{\partial^2 f}{\partial x^2}. \tag{6}$$

He also obtains the *forward parabolic equations* (with respect to $y \in \mathbb{R}$ and $t > s$):

$$\frac{\partial f}{\partial t} = -\frac{\partial}{\partial y}\big[a(t, y) f\big] + \frac{1}{2} \frac{\partial^2}{\partial y^2}\big[b^2(t, y) f\big]. \tag{7}$$

He discusses the existence, the uniqueness, and the regularity of the solutions to these equations. (We note that equations (7) had been considered before by A. D. Fokker [161] and M. Planck [389] in their analysis of diffusion from a physical standpoint.)

2. The theory and the methods of Markov processes made the next major step in the 1940s and the 1950s, in papers by K. Itô [242]–[244], who sought an 'explicit', effective construction of diffusion processes (and also diffusion processes with jumps) with well-defined local characteristics $a(s,x)$ and $b^2(s,x)$ (see (3) and (4)).

He succeeded by the representing the desired processes in terms of some 'basic' Brownian motion $B = (B_t)_{t\geqslant 0}$, as solutions of stochastic differential equations

$$dX_t = a(t, X_t)\,dt + b(t, X_t)\,dB_t. \tag{8}$$

In accordance with [242]–[244] (see also § 3e), equation (8) with $X_0 =$ Const and with coefficients satisfying the local Lipschitz condition and linear growth condition with respect to the phase variable has a strong solution, which, moreover, is unique. If, in addition, the coefficients $a(t,x)$ and $b(t,x)$ are continuous in (t,x), then X is a Markov diffusion process, which has, in particular, properties (3)–(5). Hence, if we add certain conditions on the smoothness of the transitional densities and the coefficients $a(t,x)$ and $b(t,x)$ (see, e.g., § 14 in [182] for more detail), then both forward and backward Kolmogorov equations are satisfied.

In a similar way one can treat the case of d-dimensional processes $X = (X^1,\dots,X^d)$ satisfying the equations

$$dX_t^i = a^i(t, X_t)\,dt + \sum_{j=1}^{d} b^{ij}(t, X_t)\,dB_t^j. \tag{9}$$

Setting

$$\sigma^{ij} = \sum_{k=1}^{d} b^{ik}b^{jk}, \tag{10}$$

$$L(s,x)f = \sum_{i=1}^{d} a^i(s,x)\frac{\partial f}{\partial x_i} + \frac{1}{2}\sum_{i,j=1}^{d}\sigma^{ij}(s,x)\,\frac{\partial^2 f}{\partial x_i\partial x_j}; \tag{11}$$

and

$$L^*(t,y)f = -\sum_{i=1}^{d}\frac{\partial}{\partial y_i}\bigl[a^i(t,y)f\bigr] + \frac{1}{2}\sum_{i,j=1}^{d}\frac{\partial^2}{\partial y_i\partial y_j}\bigl[\sigma^{ij}(t,y)f\bigr] \tag{12}$$

(cf. formula (27) in § 3d) we can write the *backward* and the *forward Kolmogorov parabolic equations* in the following form:

$$-\frac{\partial f}{\partial s} = L(s,x)f \tag{13}$$

and

$$\frac{\partial f}{\partial t} = L^*(t, y) f. \tag{14}$$

We point out the case when the a^i and b^{ij} are independent of time, i.e.,

$$a^i = a^i(x) \quad \text{and} \quad b^{ij} = b^{ij}(x).$$

In this case, for $0 \leqslant s < t$ we have

$$f(s, x; t, y) = f(0, x; t - s, y).$$

Introducing the function $g = g(x, t; y) = f(0, x; t, y)$ we see from (13) that it satisfies the following parabolic equation with respect to (t, x):

$$\frac{\partial g}{\partial t} = L(x) g \tag{15}$$

where $L(x)g = \sum_{i=1}^{d} a^i(x) \dfrac{\partial g}{\partial x_i} + \dfrac{1}{2} \sum_{i,j=1}^{d} \sigma^{ij}(x) \dfrac{\partial^2 g}{\partial x_i \partial X_j}$.

3. We proceed now to a discussion of several well-known results showing how one can obtain probabilistic representations (in terms of Brownian motions and solutions of stochastic differential equations) for the solutions of several classical problems of partial differential equations.

A. Cauchy problem

Find a continuous function $u = u(t, x)$ in the domain $[0, \infty) \times \mathbb{R}^d$ such that

$$u(0, x) = \varphi(x) \tag{16}$$

where $\varphi = \varphi(x)$ is some fixed function satisfying one of the parabolic equations below (cf. (15)).

A1. The heat equation.

$$\frac{\partial u}{\partial t} = \frac{1}{2} \Delta u, \tag{17}$$

where Δ is the Laplace operator $\Delta = \sum_{i=1}^{d} \dfrac{\partial^2}{\partial x_i^2}$.

The function

$$v(t, x) = \mathsf{E}_x f(B_t) \tag{17$'$}$$

is a solution of the Cauchy problem for (17); it is called the *probabilistic solution*. By E_x in (17$'$) we mean averaging with respect to the measure P_x corresponding to a Brownian motion starting from x; i.e., $B_0 = x$.)

A2. The inhomogeneous heat equation.

$$\frac{\partial u}{\partial t} = \frac{1}{2}\Delta u + \psi, \tag{18}$$

where $\psi = \psi(t, x)$.

The probabilistic solution is as follows:

$$v(t,x) = \mathsf{E}_x\biggl(\varphi(B_t) + \int_0^t \psi(t-s, B_t)\, ds\biggr). \tag{18$'$}$$

A3. The Feynman–Kac equation.

$$\frac{\partial u}{\partial t} = \frac{1}{2}\Delta u + cu, \tag{19}$$

where $c = c(x)$.

Its probabilistic solution is

$$v(t,x) = \mathsf{E}_x\biggl(\varphi(B_t)\exp\biggl\{\int_0^t c(B_s)\, ds\biggr\}\biggr). \tag{19$'$}$$

A4. The Cameron–Martin equation.

$$\frac{\partial u}{\partial t} = \frac{1}{2}\Delta u + (a, \nabla u), \tag{20}$$

where $a = \bigl(a^1(x), \dots, a^d(x)\bigr)$ and $\nabla = \Bigl(\dfrac{\partial}{\partial x_1}, \dots, \dfrac{\partial}{\partial x_d}\Bigr)$.

Its probabilistic solution is

$$v(t,x) = \mathsf{E}_x\biggl(\varphi(B_t)\exp\biggl\{\int_0^t a(B_s)\, dB_s - \frac{1}{2}\int_0^t |a(B_s)|^2\, ds\biggr\}\biggr). \tag{20$'$}$$

Remark 1. Of course, if one requires that the functions $v(t,x)$ in (17$'$)–(20$'$) be solutions, then one needs first of all that the expectations in these formulas be well defined. To this end it suffices, for instance, that the functions $\varphi(x)$, $\psi(t,x)$, $c(x)$, and $a(x)$ be bounded. (These assumptions turn out to be excessively restrictive in many problems; in that case the question should be studied more carefully.)

Remark 2. In our discussion of 'probabilistic solutions' we do not mean solutions in some special 'probabilistic' class; these solutions are merely defined in probabilistic terms such as averaging with respect to a Wiener measure and so on.

B. Dirichlet problem

One looks for a function $u = u(x)$ in the class C^2 in a bounded domain $G \subseteq \mathbb{R}^d$ that satisfies the equation

$$\Delta u = 0, \qquad x \in G \tag{21}$$

(a harmonic function) and the condition

$$u(x) = \varphi(x), \qquad x \in \overline{\partial G}. \tag{22}$$

Let

$$\tau = \inf\{t \colon B_t \notin G\}.$$

Then the probabilistic solution is

$$v(x) = \mathsf{E}_x \varphi(B_\tau). \tag{23}$$

Remark 3. Here one must also impose certain conditions on the domain G for $\tau = \tau(\omega)$ to be a finite Markov time; it is also necessary that $\varphi(B_\tau)$ be integrable.

4. We now discuss the main steps in the proof that $v(x)$ in (23) is indeed the solution of (21) (under several additional assumptions).

Let G be a bounded open subset of $\mathbb{R}^d$ and let $\varphi = \varphi(x)$ be a bounded function. We shall also assume that $v(x) = \mathsf{E}_x \varphi(B_\tau)$ is a C^2-function. Then by Itô's formula we obtain (on $[\![0, \tau[\![\equiv \{(\omega, t) \colon t < \tau(\omega)\}$)

$$v(B_t) = v(B_0) + \frac{1}{2} \int_0^t (\Delta v)(B_s)\, ds + \int_0^t (\nabla v)(B_s)\, dB_s. \tag{24}$$

If v satisfies the condition

$$\mathsf{P}\biggl(\int_0^\tau \sum_{i=1} \Bigl(\frac{\partial v}{\partial x_i}(B_s)\Bigr)^2 ds < \infty\biggr) = 1,$$

then the last integral in (24) is a *local martingale* on $[\![0, \tau[\![$ (see [250; p. 88]).

By the Markov property of a Brownian motion,

$$\mathsf{E}_x\bigl(\varphi(B_\tau) \mid \mathscr{F}_t\bigr) = v(B_t) \tag{25}$$

(P-a.s.) on the set $[\![0, \tau[\![$, i.e., $(v(B_t))$ is a martingale on this set.

Hence it follows from (24) that the integral $\Bigl(\int_0^t \Delta v(B_s)\, ds\Bigr)$ is a local martingale on $[\![0, \tau[\![$. At the same time it is a continuous process of bounded variation, therefore this is a zero process. (This implication, which is a consequence of the *Doob–Meyer decomposition* for submartingales, is routinely used when one must verify that a probabilistic solution really satisfies one or another equation; see [250; Chapter I, § 3b] and § 5b below.)

Hence we conclude that $\Delta v(x) = 0$, $x \in G$, because if this equality fails at a point $x_0 \in G$, then $\Delta v(x) \neq 0$ in a neighborhood of x_0 by the continuity of $\Delta v(x)$, so that, with positive probability, the process $\left(\int_0^t \Delta v(B_s)\, ds\right)_{t\leqslant\tau}$ is nonzero on $[\![0,\tau[\![$.

Remark 4. As regards the unique solvability of Dirichlet problems for elliptic equations and the representation of an arbitrary solution $u(x)$ of such a problem as $\mathsf{E}_x\varphi(B_\tau)$ see, e.g., [123; 8.5], or [170; vol. 1, Chapter 6, § 2].

5. On the conceptual level, the issue of probabilistic solutions to a Cauchy problem can be treated conceptually much the same way.

For example we consider now the heat equation (17); we claim that the function $v(t,x) = \mathsf{E}_x\varphi(B_t)$ is a solution to this equation (under several additional assumptions) and satisfies the initial condition $u(0,x) = \varphi(x)$

Indeed, by the Markov property of a Brownian motion,

$$\mathsf{E}_x(f(B_t)\,|\,\mathscr{F}_s) = v(t-s, B_s). \tag{26}$$

Since $\big(\mathsf{E}_x(f(B_t)\,|\,\mathscr{F}_s)\big)_{s\leqslant t}$ is a martingale (for $|f| \leqslant c$, at any rate), the process $\big(v(t-s,B_s)\big)_{s\leqslant t}$ is also a martingale.

Further, if $v \in C^2$, then we can use Itô's formula to obtain

$$\begin{aligned} v(t-s,B_s) = v(t,B_0) + \int_0^t \left(-\frac{\partial v}{\partial t} + \frac{1}{2}\Delta v\right)(t-r,B_r)\,dr \\ + \int_0^t \nabla v(t-r,B_r)\,dB_r. \end{aligned} \tag{27}$$

If

$$\mathsf{P}\left\{\int_0^t \sum_{i=1}^d \left(\frac{\partial v}{\partial x_i}(x_i - r, B_r)\right)^2 dr < \infty\right\} = 1 \tag{28}$$

for each t, then the last integral in (27) is a local martingale. Hence the process

$$\left(\int_0^t \left(-\frac{\partial v}{\partial t} + \frac{1}{2}\Delta v\right)(t-r,B_r)\,dr\right)_{t\geqslant 0}$$

is indistinguishable (P-a.s.) from a zero process, because it is both a local martingale and a continuous process of bounded variation.

If, in addition $\varphi = \varphi(x)$, is a continuous function, then $v(t,x) \to \varphi(x)$ as $t \to 0$.

In a similar way, using Itô's formula, one can prove that, indeed, formulas (17′)–(20′) describe probabilistic solutions to the problems (17)–(20), respectively.

Remark 5. As regards a more thorough discussion of the above-stated Cauchy and Dirichlet problems for parabolic equations and also the corresponding problem for the forward and backward Kolmogorov equations, see, e.g., [123], [170], [182], and [288].

4. Diffusion Models of the Evolution of Interest Rates, Stock and Bond Prices

§ 4a. Stochastic Interest Rates

1. The simplest model where one encounters *stochastic interest rates* $r = (r_n)_{n \geqslant 1}$, is the model of a bank account $B = (B_n)_{n \geqslant 0}$, in which (by definition)

$$r_n = \frac{\Delta B_n}{B_{n-1}}. \tag{1}$$

If $(\Omega, \mathscr{F}, (\mathscr{F}_n)_{n \geqslant 0}, \mathsf{P})$ is a stochastic basis (a filtered probability space) describing the stochastics of a financial market and the available 'information' $(\mathscr{F}_n)_{n \geqslant 0}$ about this market, then, as already pointed out (Chapter II, § 1e), it is natural to assume that the state B_n of the bank account is $\mathscr{F}_{n-1}$-measurable.

Hence both sequences $B = (B_n)_{n \geqslant 0}$ and $r = (r_n)_{n \geqslant 1}$ are predictable (see Chapter II, § 1a).

This feature explains why one usually imposes the requirement of the *predictability* of the interest rates $r = (r(t))_{t \geqslant 0}$ in the continuous-time case (see § 5a and, for greater detail, [250; Chapter I]); this requirement is automatically satisfied if $r = (r(t))_{t \geqslant 0}$ is a continuous (or only left-continuous) process.

In what follows we consider only models in which the *interest rates* $r = (r(t))_{t \geqslant 0}$ are diffusion processes and, therefore, have continuous trajectories (so that the additional requirement of predictability becomes superfluous).

In the case of continuous time $t \geqslant 0$ the standard definition of the bank *interest rate* $r = (r(t))_{t \geqslant 0}$ is based on the relation

$$dB_t = r(t) B_t \, dt, \tag{2}$$

which is a natural continuous analog of (1).

Clearly,

$$r(t) = (\ln B_t)' \tag{3}$$

and

$$B_t = B_0 \exp\left\{ \int_0^t r(s)\, ds \right\}. \tag{4}$$

The concept of interest rate (short rate of interest, spot rate, instantaneous interest rate) plays an even more important role in the 'indirect data' of the evolution of share prices (see § 4c below). This explains the existence of a variety of models with interest rates $r = (r(t))_{t \geqslant 0}$ described by diffusion equations

$$dr(t) = a\big(t, r(t)\big)\, dt + b\big(t, r(t)\big)\, dW_t \tag{5}$$

or, say, by equations of 'diffusion-with-jumps' type

$$dr(t) = a\big(t, r(t)\big)\, dt + b\big(t, r(t)\big)\, dW_t + \int c\big(t, r(t-), x\big)\big(\mu(dt, dx) - \nu(dt, dx)\big). \tag{6}$$

$\mu = \mu(dt, dx)$ is a random Poisson measure in $\mathbb{R}_t \times E \times \mathbb{R}^d$, and $\nu = \nu(dt, dx)$ is its compensator (see [250; Chapter III, § 2c] for greater detail).

2. We discuss now several popular models of interest rates $r = (r(t))_{t \geqslant 0}$ falling in the class of diffusion models (5) with a standard Wiener process (a Brownian motion) $W = (W_t)_{t \geqslant 0}$ defined on some stochastic basis $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t \geqslant 0}, \mathsf{P})$.

The *Merton model* (R. C. Merton, [346] (1973)):

$$dr(t) = \alpha\, dt + \gamma\, dW_t. \tag{7}$$

The *Vasiček model* (O. Vasiček [472] (1977)):

$$dr(t) = (\alpha - \beta r(t))\, dt + \gamma\, dW_t. \tag{8}$$

The *Dothan model* (M. Dothan [111] (1978)):

$$dr(t) = \alpha r(t)\, dt + \gamma r(t)\, dW_t. \tag{9}$$

The Cox–Ingersoll–Ross models (J. C. Cox, J. E. J. Ingersoll, and S. A. Ross [80] (1980) and [81] (1985)):

$$dr(t) = \beta\big(r(t)\big)^{3/2}\, dW_t, \tag{10}$$

and

$$dr(t) = \big(\alpha - \beta r(t)\big)\, dt + \gamma\big(r(t)\big)^{1/2}\, dW_t. \tag{11}$$

The Ho–Lee model (T. Ho and S. Lee [224] (1986)):

$$dr(t) = \alpha(t)\, dt + \gamma\, dW_t. \tag{12}$$

The Black–Derman–Toy model (F. Black, E. Derman, and W. Toy [42] (1990)):

$$dr(t) = \alpha(t)r(t)\,dt + \gamma(t)\,dW_t. \tag{13}$$

The Hull–White models (J. Hull and A. White, [234] (1990)):

$$dr(t) = \big(\alpha(t) - \beta(t)r(t)\big)\,dt + \gamma(t)\,dW_t, \tag{14}$$

and

$$dr(t) = \big(\alpha(t) - \beta(t)r(t)\big)\,dt + \gamma(t)\big(r(t)\big)^{1/2}\,dW_t. \tag{15}$$

The Black–Karasinski model (F. Black and P. Karasinski [43] (1991)):

$$dr(t) = r(t)\big(\alpha(t) - \beta(t)\ln r(t)\big)\,dt + \gamma(t)r(t)\,dW_t. \tag{16}$$

The Sandmann–Sondermann model (K. Sandmann and D. Sondermann [422] (1993)):

$$r(t) = \ln\big(1 + \xi(t)\big), \tag{17}$$

where

$$d\xi(t) = \xi(t)\big(\alpha(t)\,dt + \gamma(t)\,dW_t\big). \tag{18}$$

3. Introducing these models their authors also explain the motivations behind them.

For instance, the Vasiček model (8) comes naturally with the assumption that the interest rate oscillates near a fixed level α/β. (It is clear from (8) that the process has a positive drift for $r(t) < \alpha/\beta$, and a negative drift for $r(t) > \alpha/\beta$; if $\alpha = 0$, then (8) turns to the Ornstein–Uhlenbeck equation discussed in § 3a.)

It should be noted that, as shown by many empirical studies of bond rates (see, e.g., [69] and [70]), one cannot in general say that interest rates display a trend of returning to some *fixed* mean value (α/β) (*mean reversion*).

This is taken into account in the Hull–White models, in which the *variable* level $\alpha(t)/\beta(t)$, $t \geqslant 0$, replaces the constant α/β.

Moving ever further, we can assume that this variable level is itself a stochastic process. The following model provides a suitable example here.

The *Chen model* (L. Chen ([70], 1995)):

$$dr(t) = \big(\alpha(t) - r(t)\big)\,dt + \big(\gamma(t)r(t)\big)^{1/2}\,dW_t^1, \tag{19}$$

where $\alpha(t)$ and $\gamma(t)$ are stochastic processes of diffusion type:

$$d\alpha(t) = \big(\alpha - \alpha(t)\big)\,dt + \big(\alpha(t)\big)^{1/2}\,dW_t^2, \tag{20}$$

$$d\gamma(t) = \big(\gamma - \gamma(t)\big)\,dt + \big(\gamma(t)\big)^{1/2}\,dW_t^3; \tag{21}$$

(here α and γ are constants, while W^1, W^2, and W^3 are independent Wiener processes).

In many of the above models the diffusion coefficient ('volatility') depends on the current level of the interest rate $r(t)$. This can be explained, for instance, as follows: if the interest rate strongly increases, then the risks of investing into the assets in question are also higher. These risks are reflected by the fluctuation terms (e.g., $(r(t))^{1/2}\,dW_t$) in the equations for $r(t)$.

4. We now present another (fairly simplified) model of the evolution of interest rates, which was inspired by the following considerations.

Plausibly enough, the stochastic process $r = (r(t))_{t \geqslant 0}$ is, to a certain extent, a reflection of the state of the 'economy', a judgment on this state.

Taking this as a starting point, we shall now assume that the state of the 'economy' can be simulated by, say, a homogeneous Markov jump process $\theta = (\theta(t))_{t \geqslant 0}$ that has (for simplicity) only two states, $i = 0$ and $i = 1$. Let $\mathsf{P}(\theta(0) = 0) = \mathsf{P}(\theta(0) = 1) = \frac{1}{2}$ and assume that the transition probability densities λ_{ij} satisfy the equalities $\lambda_{ii} = -\lambda$ and $\lambda_{ij} = \lambda$ for $i \neq j$.

Thus, the 'economy' switches between the states $i = 0$ and $i = 1$, and the time of each stay is distributed exponentially, with parameter λ.

We also assume that we can judge the state $\theta = (\theta(t))_{t \geqslant 0}$ of the 'economy' only by *indirect data*; namely, we can observe the process $X = (X_t)_{t \geqslant 0}$ with differential

$$dX_t = \theta(t)\, dt + d\widetilde{W}_t, \tag{22}$$

where $\widetilde{W} = (\widetilde{W}_t)_{t \geqslant 0}$ is some Wiener process.

Now let

$$r(t) = \mathsf{E}(\theta(t) \mid \mathscr{F}_t^X) \tag{23}$$

be the optimal (in the mean square sense) estimator for $\theta(t)$ on the basis of the observations X_s, $s \leqslant t$, $(\mathscr{F}_t^X = \sigma(X_s,\, s \leqslant t))$.

By general nonlinear filtering theory (see [303; formula (9.86)]),

$$dr(t) = \lambda\bigl(1 - 2r(t)\bigr)\, dt + r(t)\bigl(1 - r(t)\bigr)\bigl(dX_t - r(t)\, dt\bigr). \tag{24}$$

We note that the process $W = (W_t)_{t \geqslant 0}$, where

$$W_t = X_t - \int_0^t r(s)\, ds, \tag{25}$$

is a Wiener process with respect to the flow $(\mathscr{F}_t^X)_{t \geqslant 0}$ (see Chapter VII, § 3b and [303; Theorem 7.12]; the process W is called an *innovation* process). Hence (24) can be rewritten as

$$dr(t) = \lambda\bigl(1 - 2r(t)\bigr)\, dt + r(t)\bigl(1 - r(t)\bigr)\, dW_t, \tag{26}$$

which is an equation of type (5). It is also interesting that, in view of (22), by (24) we obtain

$$dr(t) = a\bigl(r(t), \theta(t)\bigr)\, dt + b\bigl(r(t)\bigr)\, d\widetilde{W}_t, \tag{27}$$

where

$$a(r, \theta) = \lambda(1 - 2r) + r(1 - r)(\theta - r) \quad \text{and} \quad b(r) = r(1 - r)$$

(cf. the Chen model (19)).

5. The above models of the dynamics of interest rates $r = (r(t))_{t\geqslant 0}$ rely on stochastic differential equations that involve some basic Wiener process.

Incidentally, many of these equations admit 'explicit' solutions that are functionals of this Wiener process. For instance, the solution of the equation (8) in the Vasiček model and of its generalization, equation (14) in the Hull–White model, has the representation

$$r(t) = g(t)\left[r(0) + \int_0^t \frac{\alpha(s)}{g(s)}\,ds + \int_0^t \frac{\gamma(s)}{g(s)}\,dW_s\right] \tag{28}$$

(because (8) and (14) are linear equations), where

$$g(t) = \exp\left\{-\int_0^t \beta(s)\,ds\right\}. \tag{29}$$

One can see this easily using Itô's formula; cf. formula (7) in § 3e.

We now set

$$T(t) = \int_0^t \left(\frac{\gamma(s)}{g(s)}\right)^2 ds. \tag{30}$$

Assuming that $T(t) < \infty$ for all $t > 0$ and $T(t) \to \infty$ as $t \to \infty$, we now introduce 'new' time θ by the equality $\theta = T(t)$ (see Chapter IV, § 3d for a more detailed discussion of this time as 'operational').

As is well known (see, e.g., [303; Lemma 17.4]), under these assumptions there exists a (new) Wiener process $W^* = (W_t^*)_{t\geqslant 0}$ such that

$$\int_0^t \left(\frac{\gamma(s)}{g(s)}\right) dW_s = W^*_{T(t)}. \tag{31}$$

Hence we obtain the following representation for $r(t)$ in (28):

$$r(t) = f(t) + g(t)W^*_{T(t)}, \tag{32}$$

where

$$f(t) = g(t)\left[r(0) + \int_0^t \frac{\alpha(s)}{g(s)}\,ds\right]. \tag{33}$$

If

$$T^*(\theta) = \inf\{t \colon T(t) = \theta\},$$

then we can pass back from 'new' time $\theta = T(t)$ to 'old' time using the formula $t = T^*(\theta)$. This shows that we can write our process as $r^*(\theta) = r(T^*(\theta))$ and the process $r^* = (r^*(\theta))_{\theta\geqslant 0}$ has a very simple structure. Namely,

$$r^*(\theta) = f^*(\theta) + g^*(\theta)W^*_\theta,$$

where $f^*(\theta) = f(T^*(\theta))$ and $g^*(\theta) = g(T^*(\theta))$.

In the Black–Karasinski model (16) we have

$$d\ln r(t) = \bigl(\alpha(t) - \gamma^2(t) - \beta(t)\ln r(t)\bigr)\,dt + \gamma(t)\,dW_t. \tag{34}$$

Defining $T(t)$ by the same formula (30) we see that

$$r(t) = F\bigl(f(t) + g(t)W^*_{T(t)}\bigr) \tag{35}$$

(for some other Wiener process, $W^* = (W^*_t)_{t\geqslant 0}$) where $g(t)$ is as in (29),

$$f(t) = g(t)\left[r(0) + \int_0^t \frac{\alpha(s) - \gamma^2(s)}{g(s)}\,ds\right], \tag{36}$$

and $F(x) = e^x$.

In a similar way, for the Sandmann–Sondermann model (17)–(18) we obtain

$$r(t) = F\bigl(f(t) + W^*_{T(t)}\bigr), \tag{37}$$

where $F(x) = \ln(1+e^x)$, $T(t) = \int_0^t \gamma^2(s)\,ds$, and

$$f(t) = \ln\xi(0) + \int_0^t \Bigl(\alpha(s) - \frac{1}{2}\gamma^2(s)\Bigr)\,ds.$$

W. Schmidt [426] has observed that in all the above 'explicit' representations the interest rates $r(t)$ have the form (35). This brings one to the following, rather general, model.

The *Schmidt model* (W. Schmidt [426] (1997)):

$$r(t) = F\bigl(f(t) + g(t)W_{T(t)}\bigr), \tag{38}$$

where $W = (W_t)_{t\geqslant 0}$ is a Wiener process, $T(t)$ and $F(x)$ are continuous nonnegative strictly increasing functions of $t \geqslant 0$ and $x \in \mathbb{R}$, respectively, while $f = f(t)$ and $g = g(t) > 0$ are continuous functions.

Note that 'new' time $\theta = T(t)$ in this model is a deterministic function of 'old' time t, so that $(X_t = f(t) + g(t)W_{T(t)})$ is a Gaussian process. Choosing suitable functions $F(x)$ one can obtain various probability distributions of the interest rates $r(t)$.

6. The Schmidt model (38) is also attractive in the following respect: its 'discrete' version enables one to build easily *discrete models* of the evolution of interest rates, using one or another random-walk approximation to a Wiener process.

For instance, if

$$T_i^{(n)} = \inf\Bigl\{t \geqslant 0 \colon T(t) > \frac{i}{n}\Bigr\}$$

for $n \geqslant 1$, where $i = 0, 1, \ldots$ and $(\xi_i^{(n)})$ is a sequence of Bernoulli random variables with $\mathsf{P}(\xi_i^{(n)} = \pm 1/\sqrt{n}) = \frac{1}{2}$, then the (piecewise constant) process $W^{(n)} = (W_t^{(n)})_{t \geqslant 0}$ with

$$W_t^{(n)} = \sum_{i=1}^{[nT(t)]} \xi_i^{(n)}, \qquad W_0^{(n)} = 0,$$

weakly converges (as $n \to \infty$) to a Wiener process $W = (W_t)_{t \geqslant 0}$.

We now set

$$r_i^{(n)}(j) = F\left(f\big(t_i^{(n)}\big) + g\big(t_i^{(n)}\big) \frac{j}{\sqrt{n}} \right), \qquad j = 0, \pm 1, \ldots, \pm i.$$

To obtain a discrete-time model of the evolution of the interest rates $r_i^{(n)}$ ranging in the sets $\{r_i^{(n)}(j),\ j = 0, \pm 1, \ldots, \pm i\}$, let $r_0^{(n)} = r_0$ and assume that, with probability $\frac{1}{2}$, the state $r_i^{(n)}(j)$, $j = 0, \pm 1, \ldots, \pm i$, changes either to $r_{i+1}^{(n)}(j+1)$ or to $r_{i+1}^{(n)}(j-1)$.

Clearly, this construction brings one (see [426]) to the *binomial* model of the evolution of interest rates, which can be depicted as follows (for a given value of $n = 1, 2, \ldots$):

$$\begin{tikzcd}[column sep=small]
 & & r_0 \arrow[dl] \arrow[dr] & & \\
 & r_1^{(n)}(-1) \arrow[dl] \arrow[dr] & & r_1^{(n)}(+1) \arrow[dl] \arrow[dr] & \\
r_2^{(n)}(-2) & & r_2^{(n)}(0) & & r_2^{(n)}(+2)
\end{tikzcd}$$

....

§ 4b. Standard Diffusion Model of Stock Prices (Geometric Brownian Motion) and its Generalizations

1. We have already pointed out (Chapter I, § 1b) that the first model developed for the description of the evolution of stock prices $S = (S_t)_{t \geqslant 0}$ had been the *linear* model of L. Bachelier (1900)

$$S_t = S_0 + \mu t + \sigma W_t, \tag{1}$$

where $W = (W_t)_{t \geqslant 0}$ was a standard Brownian motion (a Wiener process).

Although, in principle, it was a major step in the application of probabilistic concepts to the analysis of financial markets, it was clear from the very beginning that the model (1) had many deficiencies. The first of them was that the variables S_t (that were supposed to represent stock *prices*) could take *negative* values.

Essential in this connection was the next step, made by P. Samuelson [420]. He suggested to describe stock prices in terms of a geometric (or, as he has put it, *economic*) Brownian motion

$$S_t = S_0 e^{\mu t} e^{\sigma W_t - \frac{\sigma^2}{2} t}. \tag{2}$$

That is, Samuelson assumed that the *logarithms* of the prices S_t (rather then the prices themselves) are governed by a linear model of type (1), so that

$$\ln \frac{S_t}{S_0} = \left(\mu - \frac{\sigma^2}{2}\right) t + \sigma W_t. \tag{3}$$

By Itô's formula (§ 3d) we immediately obtain that

$$dS_t = S_t(\mu\, dt + \sigma\, dW_t). \tag{4}$$

If we rewrite this relation in the following (not very rigorous) form

$$\frac{dS_t}{S_t} = \mu\, dt + \sigma\, dW_t, \tag{5}$$

then one can reasonably consider its discrete-time approximation (with time step Δ; here we set $\Delta S_t = S_t - S_{t-\Delta}$):

$$\frac{\Delta S_t}{S_{t-\Delta}} \approx \mu\, \Delta + \sigma\, \Delta W_t,$$

which is very much similar to the expression

$$\frac{\Delta S_n}{S_{n-1}} = \rho_n \tag{6}$$

with $\mathscr{F}_n$-measurable ρ_n, which we have already discussed (see, e.g., Chapter II, § 2a.6) in the description of the evolution of stock prices proceeding in discrete time.

It is also instructive to compare (5) with the expression for the (concomitant) bank account $B = (B_t)_{t\geqslant 0}$ with (fixed) interest rate $r > 0$, which is governed by the equation

$$dB_t = rB_t\, dt. \tag{7}$$

If the financial market under consideration is formed by a bank account $B = (B_t)_{t \geqslant 0}$ and stock of price $S = (S_t)_{t \geqslant 0}$ governed by equations (7) and (4), respectively, then we shall say that we have the ***standard diffusion*** (B, S)***-model*** or a standard diffusion (B, S)-market.

This standard diffusion model was considered in the calculations of option prices by F. Black and M. Scholes [44] and R. Merton [346]. The famous *Black–Scholes formula* for the rational (fair) price of European call options was invented just in the discussion of this model. (We devote Chapter VIII to these issues.)

It is fairly obvious that the standard model is based on assumptions that are not perfectly practical. In fact, it is implied that the interest rate r of the bank account is constant (while it actually fluctuates); the coefficients σ of volatility and μ of growth must also be constant (in fact, they change with time). In the deduction of the Black–Scholes formula one also assumes (see Chapter VIII, §§ 1b, c below) that the (B, S)-market is 'frictionless' (no transaction costs, no dividend payments, no delays in obtaining information or taking decisions), one can withdraw (or deposit) any amount from the bank account, and buy or sell any number of shares.

All this suggests that the standard diffusion (B, S)-market is, of course, a simplification. Nevertheless, it remains one of the most popular models.

The following words of F. Black concerning the 'simplicity' of this model ([43], 1988) are noteworthy in this connection:

> "*Yet that weakness is also its greatest strength. People like the model because they can easily understand its assumptions. The model is often good as a first approximation, and if you can see the holes in the assumptions you can use the model in more sophisticated ways.*"

2. One of these refinements, immediately suggesting itself, is to consider models with constants r, μ, and σ replaced by deterministic or random (adapted to $(\mathscr{F}_t)_{t \geqslant 0}$) functions $r(t)$, $\mu(t)$, and $\sigma(t)$:

$$dB_t = r(t) B_t \, dt, \tag{8}$$

$$dS_t = S_t (\mu(t) \, dt + \sigma(t) \, dW_t). \tag{9}$$

Of course, in our assessments of more complicated models we must be primarily backed by facts that have been established experimentally and cannot be explained, 'captured' by the standard (B, S)-model. On the other hand, these more refined models should not become complicated to a point where it is 'no more possible to calculate anything'.

In this connection we should mention the so-called '*smile effect*'. This is just one fact that the standard (B, S)-model fails to explain; and this failure has brought on various versions of its generalizations and refinements.

The smile effect is as follows.

Assume that the price of a stock is described by equation (4) with $S_0 = 1$ (for simplicity) and let $\mathbb{C} = \mathbb{C}(\sigma, T, K)$ be the rational price of the standard European call option with payoff $f_T = \max(S_T - K, 0)$.

The Black–Scholes formula for $\mathbb{C}$ (see formula (9) in Chapter I, § 1b or, for a more thorough discussion, Chapter VIII, § § 1b, c) explicitly describes the dependence of this price on the volatility σ, the exercise time T, and the strike price K.

However, we can consider the *actual* market prices of such options with given T and K and compare these with the values of $\mathbb{C}$.

Let $\widehat{\mathbb{C}}(T, K)$ be this (actual) price asked by financial markets.

We now define $\widehat{\sigma} = \widehat{\sigma}(T, K)$ as the solution of the equation

$$\mathbb{C}(\sigma, T, K) = \widehat{\mathbb{C}}(T, K),$$

where $\mathbb{C}(\sigma, T, K)$ is the function given by the Black–Scholes formula.

It is precisely this characteristic, $\widehat{\sigma}(T, K)$, which is called the *implied* volatility, that shows out certain 'holes'(using the expression of F. Black) in the initial standard model. For these are experimentally established facts that

(a) the value of $\widehat{\sigma}(T, K)$ changes with T for fixed K;

(b) the value of $\widehat{\sigma}(T, K)$ also changes with K for fixed T as a (downwards) convex function (this explains the name 'smile effect').

To take (a) into account R. Merton proposed ([346], 1973) to regard μ and σ in the standard model as *functions of time* $\mu = \mu(t)$ and $\sigma = \sigma(t)$. Schemes of that kind are indeed used in financial analysis, in particular in calculations involving American options.

The smile effect (b) appears more delicate; various modifications of the standard model ('diffusion with jumps', 'stochastic volatility', and others models) have been developed for its explanation.

The most transparent of them is the model suggested by B. Dupire [121], [122], in which

$$dS_t = S_t\big(\mu(t)\,dt + \sigma(S_t, t)\,dW_t\big), \tag{10}$$

where $\sigma = \sigma(S, t)$ is a function of the state S and time t.

In already mentioned papers [121] and [122] Dupire showed also that accepting the idea of an 'arbitrage-free complete market' one can estimate the unknown volatility on the basis of the *actual* (observed at time $t \leqslant T$ for the states $S_t = s$) prices $\mathbb{C}_{s,t}(K, T)$ of standard European call options with exercise time T and strike price K.

3. Conceptually, the process of the refinement of the standard (B, S)-model (4) and (7) has much in common with the development of the most simple discrete-time models.

On agreeing upon this we recall (see Chapter II, § 1d) that in the discrete-time case we started from the representation

$$S_n = S_0 e^{H_n} \tag{11}$$

with $H_n = h_1 + \dots + h_n$ and $h_n = \mu_n + \sigma_n \varepsilon_n$, where μ_n and σ_n were some (nonrandom) numbers and $\varepsilon_n \sim \mathscr{N}(0,1)$.

If the coefficients $\mu(t)$ and $\sigma(t)$ in (9) are deterministic, then we can write S_t as follows:

$$S_t = S_0 e^{H_t}, \tag{12}$$

where

$$H_t = \int_0^t \left(\mu(s) - \frac{\sigma^2(s)}{2}\right) ds + \int_0^t \sigma(s)\, dW_s. \tag{13}$$

Clearly, H_t has in this case the Gaussian distribution and (12) is a continual analog of the representation (11).

Further, in Chapter II, § 1d we considered the *AR*, *MA*, and *ARMA* models, in which the volatility σ_n was assumed to be constant ($\sigma_n \equiv \sigma = \text{Const}$), while the μ_n were defined by the 'past' variables $h_{n-1}, h_{n-2}, \dots$ and $\varepsilon_{n-1}, \varepsilon_{n-2}, \dots$.

Finally, in the *ARCH* and *GARCH* models we assumed that $h_n = \sigma_n \varepsilon_n$, while the volatility σ_n depended on the 'past' (see, e.g., (19) in Chapter II, § 1d).

We point out that all these models had a *single* source of randomness, *white (Gaussian) noise* $\varepsilon = (\varepsilon_n)$.

As regards the *stochastic volatility* model (see Chapter II, § 1d.7), it is assumed there that we have *two* sources of randomness, two independent *white (Gaussian) noises* $\varepsilon = (\varepsilon_n)$ and $\delta = (\delta_n)$ such that $h_n = \sigma_n \varepsilon_n$, where $\sigma_n = e^{\frac{1}{2}\Delta_n}$ and

$$\Delta_n = \alpha_0 + \sum_{i=1}^{p} \alpha_i \Delta_{n-i} + c\delta_n \tag{14}$$

(see Chapter II, § 3c).

In the same way, in the continuous-time case we can consider various counterparts to *ARCH–GARCH* models or to models of stochastic volatility kind.

In the second case we have, e.g., models of the following form:

$$dS_t = S_t\big(\mu(t, S_t, \sigma_t)\, dt + \sigma_t\, dW_t^{\varepsilon}\big), \tag{15}$$

$$d\Delta_t = a(t, \Delta_t)\, dt + b(t, \Delta_t)\, dW_t^{\delta} \tag{16}$$

where $\Delta_t = \ln \sigma_t^2$, while $W^{\varepsilon} = (W_t^{\varepsilon})_{t \geqslant 0}$ and $W^{\delta} = (W_t^{\delta})_{t \geqslant 0}$ are two independent Wiener processes (cf. (37) and (38) in Chapter II, § 3a; models of this kind were considered in [235], [364], [432], and [477]).

Using the models (15), (16) in the situation when one can observe only the process $S = (S_t)_{t \geqslant 0}$ (and not the volatility $\sigma = (\sigma_t)_{t \geqslant 0}$), we arrive at an 'incomplete' market, where one can give no clear definition of, say, a *rational price* of an option, so that one must resort to a more complicated analysis of *upper* and *lower* prices (see Chapter V, § 1b).

Diffusion models of *ARCH–GARCH* type in the continuous-time case seem to be more promising in this respect since one can avail of the well-developed machinery of an 'arbitrage-free complete' market.

Dupire's model considered above (10) is one example of such a *Markovian* model.

Pursuing the idea of the 'dependence on the past' inherent in the *ARCH–GARCH* models it is reasonable to consider, e.g., the representations

$$S_t = S_0 e^{H_t} \tag{17}$$

such that the diffusion process $H = (H_t)_{t\geqslant 0}$ is a component of a *multivariate diffusion process* $(H_t, H_t^1, \dots, H_t^{n-1})_{t\geqslant 0}$ generated by a single Wiener process.

A corresponding example is provided by a stationary Gaussian process with rational spectral density, which can be regarded as a component of a multivariate Markov process satisfying a linear system of stochastic differential equations. (See Theorem 15.4 and the system of equations (15.64) in [303] for details.)

4. We continue the discussion of the diffusion models introduced above and describing the evolution of stock prices in Chapter VII, where we shall consider these models in the context of an 'arbitrage-free complete market'.

We pointed out in Chapter I that the concept of the *absence of arbitrage* is precisely the economists' concept of *efficient market*, to which we stick in our analysis. In the discrete-time case, the *First fundamental theorem of asset pricing* (see Chapter V, § 2b) gives one a martingale criterion of whether a particular (B, S)-market is arbitrage-free. This property imposes certain constraints on both bank account $B = (B_n)_{n\geqslant 0}$ and stock price $S = (S_n)_{n\geqslant 0}$.

In the continuous-time case it is also reasonable to stay within the realm of *arbitrage-free* (B, S)-markets. Again, this imposes certain constraints on both dynamics of stock prices and evolution of the bank account (see Chapter VII).

It should be noted that (compared with discrete time) the continuous-time case is more delicate. This is primarily due to the technical complexity of the corresponding machinery of stochastic calculus and the considerable gap between the potentials of continuous and discrete trading.

Another property shaping the structure of a (B, S)-market is its *completeness* (or lack of it), which means that we can always make up a portfolio with value reproducing our 'liabilities' (see details in Chapter V, § 1b).

Generally speaking, this desirable property of a market is rather an exception than a rule. It is remarkable, however, that we obtain this property of completeness under fairly general assumptions in the case of diffusion (B, S)-markets (see Chapter VII, § 4a).

§ 4c. Diffusion Models of the Term Structure of Prices in a Family of Bonds

1. We already discussed bonds as IOU's and their market prices $\mathsf{P}(t, T)$ in general, in the very beginning of this book (Chapter I, § 1a). We also introduced there several characteristics of bonds such as the *initial price* $\mathsf{P}(0,T)$, the *face value* $\mathsf{P}(T,T)$ (which we, for definiteness set to be equal to one), the *current interest rate*, the *yield* to maturity, and some other. We mentioned that the question on the structure of the prices $\mathsf{P}(t,T)$, $0 \leqslant t < T$, regarded as a family of stochastic objects is, in a certain sense, more complicated than the question on the probabilistic structure of stock prices.

One difficulty here is that if, for a *fixed* T we regard $\mathsf{P}(t,T)$ as a stochastic process for $0 \leqslant t \leqslant T$, then, first of all, this process must be *conditional* because $\mathsf{P}(T,T) = 1$.

A typical example is a Brownian bridge $X = (X_t)_{t \leqslant T}$ considered in § 3a and governed by the equation

$$dX_t = \frac{1 - X_t}{T - t}\, dt + dW_t$$

with $X_0 = \alpha$. It has the property that $X_t \to 1$ as $t \uparrow T$.

In a similar way to Bachelier's using a linear Brownian motion (see formula (1) in § 4b) to simulate stock prices, one could treat the process $X = (X_t)_{t \leqslant T}$ as a prospective model of the evolution of the bond prices $\mathsf{P}(t,T)$, $t \leqslant T$.

The complications here are the same as with shares: the variables X_t, $t > 0$, can in general assume any values in $\mathbb{R}$, while for a bond price, by its meaning, we have $0 < \mathsf{P}(t,T) \leqslant 1$.

Another complication that one meets in the construction of models of bond prices relates to the fact that, as a rule, there are bonds with *various* times of maturity T on the market and in investors' portfolios. Therefore, in sound models of the evolution of bond prices one should not be restricted to the consideration of some particular value of T; rather, they must be constructed to cover *some subset* $\mathbb{T} \subseteq \mathbb{R}_+$ of terms containing all possible values of the maturity times of the bonds traded on a market. There should also be no *opportunities for arbitrage* on the financial market, i.e., no one should be able to buy a bond and sell it profitably running no risks.

2. In our construction of *models of the term structure* of the prices $\mathsf{P}(t,T)$, $t \leqslant T$, of bonds with maturity date T (we shall call them T-bonds) we shall assume that $\mathbb{T} = \mathbb{R}_+$. In other words, we assume that there exists a (continual) family of bonds with prices $\mathsf{P}(t,T)$, $0 \leqslant t \leqslant T$, $T \in \mathbb{R}_+$.

We shall assume in addition that there are no bond interest payments (coupon payments), i.e., we consider only *zero-coupon* bonds.

3. In dealing with a *single* T-bond or a *family* of T-bonds, $T > 0$, one uses several characteristics of bonds that shall be helpful to us over the entire process of model building and analyzing.

Namely, let us represent the price $\mathsf{P}(t,T)$ by the following expression:

$$\mathsf{P}(t,T) = e^{-r(t,T)(T-t)}, \quad t \leqslant T, \tag{1}$$

$$\mathsf{P}(t,T) = e^{-\int_t^T f(t,s)\, ds}, \quad t \leqslant T, \tag{2}$$

with some nonnegative functions $r(t,T)$ and $f(t,s)$, $0 \leqslant t \leqslant s \leqslant T$.

Clearly,

$$r(t,T) = -\frac{\ln \mathsf{P}(t,T)}{T-t}, \qquad t < T, \tag{3}$$

and (provided that $\mathsf{P}(t,T)$ is differentiable with respect to T, $T > t$)

$$f(t,T) = -\frac{\partial}{\partial T} \ln \mathsf{P}(t,T), \qquad t \leqslant T. \tag{4}$$

In the case of one zero-coupon T-bond, for $t \leqslant T$ we have defined its *yield* or *the yield* $\rho(T-t,T)$ *to the maturity time* by the formula

$$\mathsf{P}(t,T) = e^{-(T-t)\ln(1+\rho(T-t,t))} \tag{5}$$

(see formula (6) in Chapter I, § 1a). Comparing this with (1) we see that

$$r(t,T) = \ln\bigl(1 + \rho(T-t,t)\bigr). \tag{6}$$

The quantity $r(t,T)$ is also often called the *yield* of the T-bond at time $t < T$ (due for the remaining time $T-t$), and the function $t \rightsquigarrow r(t,T)$ $(t \leqslant T)$ is called the *yield curve of the* T*-bond*.

The quantities $r(t,T)$ are especially useful when one considers *various* T-bonds with time of maturity $T > t$. In this case, the function $T \rightsquigarrow r(t,T)$ is also called the *yield curve of the family of* T*-bonds* (at time t).

To avoid treating the cases of a single bond and a bond family separately we shall regard $r(t,T)$ as a function of *two variables*, t and T, assuming that $0 \leqslant t \leqslant T$ and $T > 0$. We shall call it the *yield* again.

The quantities $f(t,T)$ are usually called the *forward rates* for the contract sold at time t.

A key role in the subsequent analysis is played by the *instantaneous rate* $r(t)$ at time t that can be defined in terms of the forward interest rate by the equality

$$r(t) = f(t,t). \tag{7}$$

We point out also the following relation between $r(t,T)$ and $f(t,T)$:

$$f(t,T) = r(t,T) + (T-t)\frac{\partial r(t,T)}{\partial T}, \tag{8}$$

which is an immediate consequence of (4) and (1).

Hence, if, e.g., the derivative $\dfrac{\partial r(t,T)}{\partial T}$ is finite, then setting $T = t$ we obtain

$$r(t) = f(t,t) = r(t,t) \quad \Bigl(= \lim_{T \downarrow t} r(t,T)\Bigr). \tag{9}$$

4. We consider now the issue of the dynamics of the bond prices $\mathsf{P}(t,T)$. The two main approaches here are the *indirect* and the *direct* ones. (Cf. Chapter I, § 1b.5.)

Taking the first approach, one represents $\mathsf{P}(t,T)$ as the composite

$$\mathsf{P}(t,T) = F\big(t, r(t); T\big), \tag{10}$$

where $r = (r(t))_{t\geqslant 0}$ is some *instantaneous rate*.

In such models the entire term structure of prices is determined by a *single factor*, $r = (r(t))_{t\geqslant 0}$. For this reason, they are called *single-factor* models.

An important (and treatable by analytic means) subclass of such models is described by functions

$$F\big(t, r(t); T\big) = e^{\alpha(t,T) - r(t)\beta(t,T)}. \tag{11}$$

These models are said to be *affine*, or, sometimes, *exponential affine* ([117], [119]) since $\ln F(t,x;T) = \alpha(t,T) - x\beta(t,T)$ is a linear function of x with some coefficients $\alpha(t,T)$ and $\beta(t,T)$.

Another well-known approach was used in [219]. The corresponding models is called the *HJM-model* after its authors (D. Heath, R. Jarrow, and A. Morton). The idea of this approach is to seek the prices $\mathsf{P}(t,T)$ as solutions of certain stochastic differential equations of type

$$d\mathsf{P}(t,T) = \mathsf{P}(t,T)\big(A(t,T)\,dt + B(t,T)\,dW_t\big), \tag{12}$$

where $A(T,T) = B(T,T) = 0$ and $P(T,T) = 1$ (cf. formula (4) in § 4b). Alternatively, one can consider equations

$$df(t,T) = a(t,T)\,dt + b(t,T)\,dW_t \tag{13}$$

for the forward interest rates $f(t,T)$.

This is an appropriate place to recall that, throughout, we assume that we have some filtered probability space (stochastic basis)

$$\big(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geqslant 0}, \mathsf{P}\big).$$

All the functions in question ($\mathsf{P}(t,T)$, $f(t,T)$, $A(t,T)$, ...) are assumed to be $\mathscr{F}_t$-measurable for $t \leqslant T$. As usual, $W = (W_t, \mathscr{F}_t)_{t\geqslant 0}$ is a standard Wiener process; moreover, we assume that the conditions ensuring the existence of the stochastic integrals in (12) and (13) and the existence of a solution to (12) are satisfied.

By relation (2) between the prices $\mathsf{P}(t,T)$ and the forward interest rates $f(t,T)$, using Itô's formula and stochastic Fubini's theorem (see Lemma 12.5 in [303] or [395]), we can obtain the following relation between the coefficients of the above equations (see [219]):

$$a(t,T) = \frac{\partial B(t,T)}{\partial T}\,B(t,T) - \frac{\partial A(t,T)}{\partial T}\,, \tag{14}$$

$$b(t,T) = -\frac{\partial B(t,T)}{\partial T}, \tag{15}$$

or

$$A(t,T) = r(t) - \int_t^T a(t,s)\,ds + \frac{1}{2}\left(\int_t^T b(t,s)\,ds\right)^2, \tag{16}$$

$$B(t,T) = -\int_t^T b(t,s)\,ds. \tag{17}$$

By (13) we also obtain

$$dr(t) = \left(\frac{\partial f}{\partial T}(t,t) + a(t,t)\right) dt + b(t,t)\,dW_t. \tag{18}$$

Remark 2. In several papers (e.g., in [38], [359], and [421]), in place of the forward interest rate $f(t,T)$, the authors consider its modification

$$\widetilde{r}(t,x) = f(t,t+x).$$

They justify this by certain analytic simplification. For instance, after this *change of variables* equality (2) acquires a slightly more simple form:

$$\mathsf{P}(t,T) = \exp\left\{-\int_0^{T-t} \widetilde{r}(t,x)\,dx\right\}. \tag{19}$$

5. We discuss the structure of bond prices $\mathsf{P}(t,T)$, forward rates $f(t,T)$, and instantaneous $r(t)$ in greater detail, in the context of a 'complete arbitrage-free' T-bond market, in Chapter VII. Here we point out only that, as with a (B,S)-market (see § 4b.4), our condition of an *arbitrage-free* T-bond market imposes certain constraints on the structure of coefficients in (12) and (13) and on the coefficients $\alpha(t,T)$, $\beta(t,T)$, and the interest rates in the affine models (11), thus distinguishing some natural classes of *arbitrage-free* models of a bond market.

5. Semimartingale Models

§ 5a. Semimartingales and Stochastic Integrals

1. The abundance of the above-mentioned models designed for the description of the evolution of such financial indexes as, say, the price of a share, poses a natural problem of distinguishing a sufficiently general class of stochastic processes that on the one hand includes many of the processes occurring in these models and, on the other hand, yields to analytic means.

From many viewpoints, one such class is that of *semimartingales*, i.e., of stochastic processes $X = (X_t)_{t \geqslant 0}$ representable (not necessarily in a unique way) as sums

$$X_t = X_0 + A_t + M_t, \tag{1}$$

where $A = (A_t, \mathscr{F}_t)_{t \geqslant 0}$ is a process of bounded variation ($A \in \mathscr{V}$) and $M = (M_t, \mathscr{F}_t)_{t \geqslant 0}$ is a local martingale ($M \in \mathscr{M}_{\text{loc}}$) both defined on some *filtered probability space* (*stochastic basis*)

$$\big(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t \geqslant 0}, \mathsf{P}\big)$$

satisfying the *usual* conditions, i.e., the σ-algebra $\mathscr{F}$ is P-complete and the $\mathscr{F}_t$, $t \geqslant 0$, must contain all the sets in $\mathscr{F}$ of P-probability zero (cf. § 3b.3), and be right continuous ($\mathscr{F}_t = \mathscr{F}_{t+}$, $t \geqslant 0$); see, e.g., [250].

We also assume that $X = (X_t)_{t \geqslant 0}$ is a process *adapted* to $(\mathscr{F}_t)_{t \geqslant 0}$ and its trajectories $t \rightsquigarrow X_t(\omega)$, $\omega \in \Omega$, are right-continuous with limits from the left In French literature a process of this type is called *un processus càdlàg*—*C*ontinu *À D*roite avec des *L*imites *À G*auche.

2. We turn to semimartingales for several reasons. First, this class is fairly wide: it contains discrete-time processes $X = (X_n)_{n \geqslant 0}$ (for one can associate with such process the continuous-time process $X^* = (X^*)_{t \geqslant 0}$ with $X^*_t = X_{[t]}$, which clearly belongs to $\mathscr{V}$), diffusion processes, diffusion processes with jumps, point processes, processes with independent increments (with several exceptions; see [250; Chapter II, § 4c]), and many other processes.

The class of semimartingales is *stable* with respect to many transformations: absolutely continuous changes of measure, time changes, localization, changes of filtration, and so on (see a discussion in [250; Chapter I, § 4c]).

Second, there exists a well-developed machinery of stochastic calculus of semimartingales, based on the concepts of *Markov times*, *martingales*, *super-* and *submartingales*, *local martingales*, *optional* and *predictable σ-algebras*, and so on.

In a certain sense, the crucial factor of the success of stochastic calculus of semimartingales is the fact that it is possible to define *stochastic integrals* with respect to semimartingales.

Remark 1. One can find an absorbing concise exposition of the central ideas and concepts of the stochastic calculus of semimartingales and stochastic integrals with respect to semimartingales in the appendix to [139] written by P.-A. Meyer, a pioneer of modern stochastic calculus.

An important ingredient of the concept of stochastic bases $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geqslant 0}, \mathsf{P})$, which are definition domains for semimartingales, is a *flow* of σ-algebras $(\mathscr{F}_t)_{t\geqslant 0}$. As already mentioned in Chapter I, § 2a, this is also a key concept in financial theory: this is the flow of 'information' available on a financial market and underlying the traders' decisions.

It should be noted that, of course, one encounters processes that are not semimartingales in some 'natural' models (from the viewpoint of financial theory).

A typical example is a fractional Brownian motion $B^{\mathbb{H}} = (B_t^{\mathbb{H}})_{t\geqslant 0}$ (see § 2c) with an arbitrary Hurst parameter $0 < \mathbb{H} < 1$ (except for the case of $\mathbb{H} = 1/2$ corresponding to a usual Brownian motion).

3. We proceed now to stochastic integration with respect to semimartingales, which nicely describes the growth of capital in self-financing strategies.

Let $X = (X_t)_{t\geqslant 0}$ be a semimartingale on a stochastic basis $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geqslant 0}, \mathsf{P})$ and let $\mathscr{E}$ be the class of *simple* functions, i.e., of functions

$$f(t,\omega) = Y_0(\omega) I_{\{0\}}(t) + \sum_i Y_i(\omega) I_{(r_i,s_i]}(t) \tag{2}$$

that are linear combinations of finitely many *elementary* functions

$$f_i(t,\omega) = Y_i(\omega) I_{(r_i,s_i]}(t) \tag{3}$$

with $\mathscr{F}_{r_i}$-measurable random variables $Y_i(\omega)$; cf. § 3c.

As with Wiener processes (Brownian motions), a 'natural' way to define the *stochastic integral* $I_t(f)$ of a simple function (2) with respect to a semimartingale X (this integral is also denoted by $(f \cdot X)_t$, $\int_0^t f(s,\omega)\,dX_s$, or $\int_{(0,t]} f(s,\omega)\,dX_s$) is to set

$$I_t(f) = \sum_i Y_i(\omega)[X_{s_i\wedge t} - X_{r_i\wedge t}], \tag{4}$$

where $a \wedge b = \min(a,b)$.

Remark 2. The stochastic integral

$$I_t(f_i) = Y_i(\omega)[X_{s_i \wedge t} - X_{r_i \wedge t}]$$

has a perfectly transparent interpretation in financial theory: if $X = (X_t)_{t \geqslant 0}$ is, say, the price of a share, and if you are buying $Y_i(\omega)$ shares at time r_i (at the price X_{r_i}), then $Y_i(\omega)[X_{s_i} - X_{r_i}]$ is precisely your profit (which can also be negative) from selling these shares at time s_i, when their price is X_{s_i}.

We emphasize that, we need not assume that X is a semimartingale in this definition of the stochastic integrals $I(f) = (I_t(f))_{t \geqslant 0}$ of simple functions; expression (4) makes sense for *any* process $X = (X_t)_{t \geqslant 0}$.

However, our semimartingale assumption becomes decisive when we are going to *extend* the definition of the integral from simple functions $f = f(t, \omega)$ to broader function classes (with preservation of its 'natural' properties).

If $X = (X_t)_{t \geqslant 0}$ is a Brownian motion, then, in accordance with § 3c, we can define the stochastic integral $I_t(f)$ for each (measurable) function $f = (f(s, \omega))_{s \leqslant t}$, provided that the $f(s, \omega)$ are $\mathscr{F}_s$-measurable and

$$\int_0^t f^2(s, \omega)\, ds < \infty \qquad (\mathsf{P}\text{-a.s.}). \tag{5}$$

The key factor here is that we can *approximate* such functions by simple functions $(f_n, n \geqslant 1)$ such that the corresponding integrals $(I_t(f_n), n \geqslant 1)$ converge (in probability, at any rate). We have denoted the corresponding limit by $I_t(f)$ and have called it the stochastic integral of f with respect to the Brownian motion (over the interval $(0, t]$).

In replacing the Brownian motion by an *arbitrary* semimartingale, we base the construction of the stochastic integral $I_t(f)$ again on the approximation of f by simple functions $(f_n, n \geqslant 1)$ with well-defined integrals $I_t(f_n)$ and the subsequent passage to the limit as $n \to \infty$.

However, the problem of approximation is now more complicated, and we need to impose certain constraints on f *adapted* to the properties of the semimartingale X.

4. As an illustration, we present several results that seem appropriate here and have been obtained in the case of $X = M$, where $M = (M_t, \mathscr{F}_t)_{t \geqslant 0}$ is a *square integrable martingale* in the class $\mathscr{H}^2$ $(M \in \mathscr{H}^2)$, i.e., a martingale with

$$\sup_{t \geqslant 0} \mathsf{E} M_t^2 < \infty. \tag{6}$$

Let $\langle M \rangle = (\langle M \rangle_t, \mathscr{F}_t)_{t \geqslant 0}$ be the quadratic characteristic of martingales in $\mathscr{H}^2$ (or in $\mathscr{H}^2_{\mathrm{loc}}$), i.e., by definition, a predictable (see Definition 2 below) nondecreasing process such that the process $M^2 - \langle M \rangle = (M_t^2 - \langle M \rangle_t, \mathscr{F}_t)_{t \geqslant 0}$ is a martingale (see § 5b and cf. Chapter II, § 1b).

The following results are well-known (see, e.g., [303; Chapter 5]).

A) If the process $\langle M\rangle$ is *absolutely* continuous (P-a.s.) with respect to t, then the set $\mathscr{E}$ of simple functions is dense in the space L_1^2 of measurable functions $f = f(t,\omega)$ such that

$$\mathsf{E}\int_0^\infty f^2(t,\omega)\,d\langle M\rangle_t < \infty. \tag{7}$$

In other words, for each f in this space there exists a sequence of simple functions $f_n = (f_n(t,\omega))_{t\geqslant 0}$, $\omega\in\Omega$, $n\geqslant 1$, such that

$$\mathsf{E}\int_0^\infty |f(t,\omega) - f_n(t,\omega)|^2\,d\langle M\rangle_t \to 0. \tag{8}$$

Note that if $B = (B_t)_{t\geqslant 0}$ is a standard Brownian motion, then its quadratic characteristic $\langle B\rangle_t$ is identically equal to t for $t\geqslant 0$.

B) If the process $\langle M\rangle$ is *continuous* (P-a.s.), then the set $\mathscr{E}$ of simple functions is dense in the space L_2^2 of measurable functions $f = (f(t,\omega))_{t\geqslant 0}$ such that (7) holds and the variables $f(\tau(\omega),\omega)$ are $\mathscr{F}_\tau$-measurable for each (finite) Markov time $\tau = \tau(\omega)$.

C) In general, if we *do not* additionally ask for the regularity of the trajectories of $\langle M\rangle$, then the set $\mathscr{E}$ of simple functions is dense in the space L_3^2 of measurable functions $f = (f(t,\omega),\mathscr{F}_t)_{t\geqslant 0}$ satisfying (7) and *predictable* in the sense that we explain next.

First, let $X = (X_n,\mathscr{F}_n)_{n\geqslant 1}$ be a stochastic sequence defined on our stochastic basis. Then, in accordance with the standard definition (see Chapter II, § 1b), by the *predictability* of X we mean that the variables X_n are $\mathscr{F}_{n-1}$-measurable for all $n\geqslant 1$.

In the continuous-time case, the following definition of the predictability (with respect to the stochastic basis $(\Omega,\mathscr{F},(\mathscr{F}_t)_{t\geqslant 0},\mathsf{P})$) proved to be the most suitable one.

Let $\mathscr{P}$ the smallest σ-algebra in the space $\mathbb{R}_+\times\Omega$ such that if a (measurable) function $Y = (Y(t,\omega))_{t\geqslant 0,\omega\in\Omega}$ is $\mathscr{F}_t$-measurable for each $t\geqslant 0$ and its t-trajectories (for each fixed $\omega\in\Omega$) are *left-continuous*, then the map $(t,\omega)\rightsquigarrow Y(t,\omega)$ generated by this function is $\mathscr{P}$-measurable.

DEFINITION 1. We call the σ-algebra $\mathscr{P}$ in $\mathbb{R}_+\times\Omega$ the *σ-algebra of predictable sets.*

DEFINITION 2. We say that a stochastic process $X = (X_t(\omega))_{t\geqslant 0}$ defined on a stochastic basis $(\Omega,\mathscr{F},(\mathscr{F}_t)_{t\geqslant 0},\mathsf{P})$ is *predictable* if the map $(t,\omega)\rightsquigarrow X(t,\omega)$ $(= X_t(\omega))$ is $\mathscr{P}$-measurable.

5. The above results on the approximation of functions f enable one, by analogy with the case of a Brownian motion, to define the stochastic integral

$$I_\infty(f) = \int_0^\infty f(s,\omega)\, dM_s \tag{9}$$

for each $M \in \mathscr{H}^2$ by means of an isometry.

Then we can define the integrals $I_t(f)$ for $t > 0$ by the formula

$$I_t(f) = \int_0^\infty I(s \leqslant t) f(s,\omega)\, dM_s. \tag{10}$$

It should be pointed out that if we impose no regularity conditions (as in A) or B)) on $\langle M\rangle$, then the above discussion shows that for $M \in \mathscr{H}^2$ the stochastic integrals $I(f) = (I_t(f))_{t>0}$ are well defined for *each predictable bounded* function f.

The next step in the extension of the definition of the stochastic integral $I_t(f)$ (or $(f \cdot X)_t$; this notation stresses the role of the process X with respect to which the integration is carried out) is to *predictable locally bounded* functions f and *locally square integrable martingales* M (martingales in the class $\mathscr{H}^2_{\mathrm{loc}}$; if $\mathscr{K}$ is some class of processes, then we say that a process $Y = (Y_t(\omega))_{t\geqslant 0}$ is in the class $\mathscr{K}_{\mathrm{loc}}$ if there exists a sequence $(\tau_n)_{n\geqslant 1}$ of Markov times such that $\tau_n \uparrow \infty$ and the 'stopped' processes $Y^{\tau_n} = (Y_{t\wedge\tau_n})_{t\geqslant 0}$ are in $\mathscr{K}$ for each $n \geqslant 1$; cf. the definition in Chapter II, § 1c.)

If (τ_n) is a localizing sequence (for a locally bounded function f and a martingale $M \in \mathscr{H}^2_{\mathrm{loc}}$), then, in accordance with the above construction of stochastic integrals for bounded functions f and $M \in \mathscr{H}^2$, the integrals $f \cdot M^{\tau_n} = ((f \cdot M^{\tau_n})_t)_{t\geqslant 0}$ are well-defined. Moreover, it is easy to see that the integrals for distinct $n \geqslant 1$ are *compatible*, i.e., $(f \cdot M^{\tau_{n+1}})^{\tau_n} = f \cdot M^{\tau_n}$.

Hence there exists a (unique, up to stochastic indistinguishability) process $f \cdot M = ((f \cdot M)_t)_{t\geqslant 0}$ such that

$$(f \cdot M)^{\tau_n} = f \cdot M^{\tau_n}$$

for each $n \geqslant 1$.

The process $f \cdot M = (f \cdot M)_t, t > 0$ so defined belongs to the class $\mathscr{H}^2_{\mathrm{loc}}$ (see [250; Chapter I, § 4d] for details) and is called the *stochastic integral* of f with respect to M.

6. The final step in the construction of stochastic integrals $f \cdot X$ for locally bounded predictable functions $f = f(t,\omega)$ with respect to semimartingales $X = (X_t, \mathscr{F}_t)_{t\geqslant 0}$ is based on the following observation concerning the structure of semimartingales.

By definition, a semimartingale X is a process representable as (1), where $A = (A_t, \mathscr{F}_t)_{t\geqslant 0}$ is some process of bounded variation, i.e., $\int_0^t |dA_s(\omega)| < \infty$, for $t > 0$ and $\omega \in \Omega$ and $M = (M_t, \mathscr{F}_t)$ is a local martingale.

By an important result in general theory of martingales, *each local martingale M has a* (not unique, in general) *decomposition*

$$M_t = M_0 + M'_t + M''_t, \qquad t \geqslant 0, \tag{11}$$

where $M' = (M'_t, \mathscr{F}_t)_{t\geqslant 0}$ and $M'' = (M''_t, \mathscr{F})_{t\geqslant 0}$ are local martingales, $M'_0 = M''_0 = 0$, and M'' has bounded variation (we write this as $M'' \in \mathscr{V}$), while $M' \in \mathscr{H}^2_{\mathrm{loc}}$ (see [250; Chapter I, Proposition 4.17]).

Hence each semimartingale $X = (X_t, \mathscr{F}_t)_{t\geqslant 0}$ can be represented as a sum

$$X_t = X_0 + A'_t + M'_t, \tag{12}$$

where $A' = A + M''$ and $M' \in \mathscr{H}^2_{\mathrm{loc}}$.

For locally bounded functions f we have the well-defined Lebesgue–Stieltjes integrals (for each $\omega \in \Omega$)

$$(f \cdot A')_t = \int_{(0,t]} f(s,\omega)\, dA'_s, \tag{13}$$

and if, in addition, f is *predictable*, then the stochastic integrals $(f \cdot M')_t$ are also well defined. Thus, we can define the integrals $(f \cdot X)_t$ in a natural way, by setting

$$(f \cdot X)_t = (f \cdot A')_t + (f \cdot M')_t. \tag{14}$$

To show that this definition is consistent we must, of course, prove that such an integral is independent (up to stochastic equivalence) of the particular representation (12), i.e., if, in addition, $X_t = X_0 + \overline{A}_t + \overline{M}_t$ with $\overline{A} \in \mathscr{V}$ and $\overline{M} \in \mathscr{H}^2_{\mathrm{loc}}$, then

$$f \cdot A' + f \cdot M' = f \cdot \overline{A} + f \cdot \overline{M}. \tag{15}$$

This is obvious for elementary functions f and, by linearity, holds also for simple functions ($f \in \mathscr{E}$). If f is predictable and bounded, then it can be approximated by simple functions f_n convergent to f pointwise. Using this fact and localization we obtain the required property (15).

Remark 3. As regards the details of the proofs of the above results and several other constructions of stochastic integrals with respect to semimartingales, including vector-valued ones, see, e.g., [250], [248], or [303], and also Chapter VII, § 1a below.

7. We now dwell on several properties of stochastic integrals of *localy bounded* functions f with respect to semimartingales (see properties (4.33)–(4.37) in [250; Chapter I]):

a) $f \cdot X$ is a *semimartingale*;

b) the map $f \rightsquigarrow f \cdot X$ is *linear*;

c) if X is a local martingale ($X \in \mathscr{M}_{\rm loc}$), then $f \cdot X$ is also a *local* martingale;
d) for $X \in \mathscr{V}$ the stochastic integral coincides with the Stieltjes integral;
e) $(f \cdot X)_0 = 0$ and $f \cdot X = f \cdot (X - X_0)$;
f) $\Delta(f \cdot X) = f \Delta X$, where $\Delta X_t = X_t - X_{t-}$;
g) if g is a predictable locally bounded function, then $g \cdot (f \cdot X) = (gf) \cdot X$.

We point out also the following result (similar to Lebesgue's theorem on dominated convergence) concerning passages to the limit under the sign of stochastic integral:

h) if $g_n = g_n(t, \omega)$ are predictable processes convergent pointwise to $g = g(t, \omega)$ and if $|g_n(t, \omega)| \leqslant G(t, \omega)$, where $G = (G(t, \omega))_{t \geqslant 0}$ is a locally bounded predictable process, then $g_n \cdot X$ converges to $g \cdot X$ in measure uniformly on each finite interval, i.e.,

$$\sup_{s \leqslant t} |g_n \cdot X - g \cdot X| \xrightarrow{\mathsf{P}} 0. \tag{16}$$

8. In connection with the above construction of stochastic integrals $f \cdot X$ of predictable functions f with respect to semimartingales X there naturally arises the problem of their calculation using more simple procedures: e.g., based on the idea of *Riemann integration*.

Here the case of left-continuous functions $f = (f(t, \omega))_{t \geqslant 0}$ is of interest.

To state the corresponding result we give the following definition ([250; Chapter I, § 4d]).

For each $n \geqslant 1$ let

$$T^{(n)} = \{\tau^{(n)}(m),\ m \geqslant 1\}$$

be a family of Markov times $\tau^{(n)}(m)$ such that $\tau^{(n)}(m) \leqslant \tau^{(n)}(m+1)$ on the set $\{\tau^{(n)}(m) < \infty\}$.

We shall call $T^{(n)}$, $n \geqslant 1$, *Riemann* sequences if

$$\sup_m \left[\tau^{(n)}(m+1) \wedge t - \tau^{(n)}(m) \wedge t\right] \to 0 \tag{17}$$

for all $t \in \mathbb{R}_+$ and $\omega \in \Omega$.

We associate now with stochastic integrals $(f \cdot X)_t$ their $T^{(n)}$-*Riemann approximants*

$$T^{(n)}(f \cdot X)_t = \sum_m f\left(\tau^{(n)}(m), \omega\right)\left[X_{\tau^{(n)}(m+1) \wedge t} - X_{\tau^{(n)}(m) \wedge t}\right]. \tag{18}$$

It turns out that if a process $f = (f(t, \omega), \mathscr{F}_t)_{t \geqslant 0}$, has *left-continuous* trajectories, then these $T^{(n)}$-Riemann approximants $T^{(n)}(f \cdot X)$ converge to $f \cdot X$ in measure uniformly on each finite time interval $[0, t]$, $t > 0$, i.e.,

$$\sup_{u \leqslant t} |T^{(n)}(f \cdot X)_u - (f \cdot X)_u| \xrightarrow{\mathsf{P}} 0. \tag{19}$$

The proof is fairly easy. Let

$$f^{(n)}(t,\omega) = \sum_m f\big(\tau^{(n)}(m),\omega\big) I\big\{(t,\omega)\colon \tau^{(n)}(m) < t \leqslant \tau^{(n)}(m+1)\big\}. \tag{20}$$

Then the functions $f^{(n)}(t,\omega)$ are predictable and converge to f pointwise since f is left-continuous.

Let $K_t = \sup\limits_{s\leqslant t} |f_s|$. Clearly, the process $K = (K_t, \mathscr{F}_t)$ is left-continuous, locally bounded, and $|f^{(n)}| \leqslant K$.

Hence

$$\sup_{u\leqslant t} |(f^{(n)}\cdot X)_u - (f\cdot X)_u| \xrightarrow{\mathsf{P}} 0$$

by property (h), and to prove the required result (19) it suffices to observe that $T^{(n)}(f\cdot X) = f^{(n)}\cdot X$.

§ 5b. Doob–Meyer Decomposition. Compensators. Quadratic Variation

1. In the discrete-time case we have an important tool of 'martingale' analysis of (arbitrary!) stochastic sequences $H = (H_n, \mathscr{F}_n)$ with $\mathsf{E}|H_n| < \infty$ for $n \geqslant 0$: the *Doob decomposition*

$$H_n = H_0 + A_n + M_n \tag{1}$$

where $A = (A_n, \mathscr{F}_{n-1})$ is a predictable sequence, and $M = (M_n, \mathscr{F}_n)$ is a martingale (see formulas (1)–(5) Chapter II, in § 1b).

In the same way, in the continuous-time case, we have the *Doob–Meyer decomposition* (of submartingales), which plays a similar role and, together with the concept of a stochastic integral, underlies the stochastic calculus of semimartingales.

Let $H = (H_t, \mathscr{F}_t)_{t\geqslant 0}$ be a submartingale, i.e., a stochastic process such that the variables H_t are $\mathscr{F}_t$-measurable and integrable, $t \geqslant 0$, the trajectories of H are càdlàg (right-continuous and having limits from the left) trajectories, and (the *submartingale property*)

$$\mathsf{E}(H_t \,|\, \mathscr{F}_s) \geqslant H_s \quad (\mathsf{P}\text{-a.s.}), \quad s \leqslant t. \tag{2}$$

We shall say that an arbitrary stochastic process $Y = (Y_t)_{t\geqslant 0}$ belongs to the *Dirichlet class* (D) if

$$\sup_\tau \mathsf{E}\big\{|Y_\tau| I(|Y_\tau| \geqslant c)\big\} \to 0 \quad \text{as} \quad c \to \infty, \tag{3}$$

where we take the supremum over all finite Markov times. In other words, the family of random variables

$$\big\{Y_\tau \colon \tau \text{ is a finite Markov time}\big\}$$

must be *uniformly integrable*.

THEOREM 1 (the Doob–Meyer decomposition). *Each submartingale H in the class (D) has a (unique) decomposition*

$$H_t = H_0 + A_t + M_t, \tag{4}$$

where $A = (A_t, \mathscr{F}_t)$ is an increasing predictable process with $\mathsf{E}A_t < \infty$, $t > 0$, $A_0 = 0$, and $M = (M_t, \mathscr{F}_t)$ is a uniformly integrable martingale.

It is clear from (4) that each submartingale in the class (D) is a semimartingale such that, in addition, the corresponding process A (belonging to the class $\mathscr{V}$, see § 5a.6) is *predictable*. This justifies *distinguishing* a subclass of *special* semimartingales $X = (X_t, \mathscr{F}_t)_{t\geqslant 0}$ having a representation (4) with *predictable* process $A = (A_t, \mathscr{F}_t)$.

In the discrete-time case the Doob decomposition is *unique* (see Chapter II, § 1b). Likewise, any two representations (4) with *predictable* processes of bounded variation of a *special* semimartingale must coincide.

It is instructive to note that each special semimartingale is in fact a difference of two local submartingales (or, equivalently, of two local supermartingales).

Remark 1. The Doob–Meyer decomposition is a 'difficult' results in *martingale theory*. We do not present its proof here (see, e.g., [103], [248], or [303]); we outline instead a possible approach to the proof based on the Doob decomposition for discrete time.

Let $X = (X_t, \mathscr{F}_t)_{t\geqslant 0}$ be a submartingale and let $X^{(\Delta)} = \bigl(X_t^{(\Delta)}, \mathscr{F}_t^{(\Delta)}\bigr)_{t\geqslant 0}$ be its discrete Δ-approximation with

$$X_t^{(\Delta)} = X_{[t/\Delta]\Delta} \quad \text{and} \quad \mathscr{F}_t^{(\Delta)} = \mathscr{F}_{[t/\Delta]\Delta}.$$

By the Doob decomposition for discrete time,

$$X_t^{(\Delta)} = X_0 + A_t^{(\Delta)} + M_t^{(\Delta)},$$

where

$$\begin{aligned} A_t^{(\Delta)} = A_{[t/\Delta]\Delta}^{(\Delta)} &= \sum_{i=1}^{[t/\Delta]} \mathsf{E}\bigl(X_{i\Delta}^{(\Delta)} - X_{(i-1)\Delta}^{(\Delta)} \mid \mathscr{F}_{(i-1)\Delta}^{(\Delta)}\bigr) \\ &= \frac{1}{\Delta}\int_0^{[t/\Delta]\Delta} \mathsf{E}\bigl(X_{[s/\Delta]\Delta+\Delta} - X_{[s/\Delta]\Delta} \mid \mathscr{F}_{[s/\Delta]\Delta}\bigr)\, ds. \end{aligned}$$

Hence one could probably seek the nondecreasing predictable process $A = (A_t, \mathscr{F}_t)_{t\geqslant 0}$ participating in the Doob–Meyer decomposition in the following form:

$$A_t = \lim_{\Delta\downarrow 0} \frac{1}{\Delta}\int_0^t \mathsf{E}(X_{s+\Delta} - X_s \mid \mathscr{F}_s)\, ds.$$

Moreover, if this process is well-defined, then it remains only to prove that the *compensated* process $X - A = (X_t - A_t, \mathscr{F}_t)_{t\geqslant 0}$ is a uniformly integrable martingale.

It is now clear why one calls the process $A = (A_t, \mathscr{F}_t)_{t\geqslant 0}$ in the Doob–Meyer decomposition the *compensator* of the submartingale X.

2. The Doob–Meyer decomposition has several useful consequences (see [250; Chapter I, § 3b]). We point out two of them.

COROLLARY 1. *Each predictable local martingale of bounded variation* $X = (X_t, \mathscr{F}_t)_{t \geqslant 0}$ *with* $X_0 = 0$ *is stochastically indistinguishable from zero.*

COROLLARY 2. *Let* $A = (A_t, \mathscr{F}_t)_{t \geqslant 0}$ *be a process in the class* $\mathscr{A}_{\text{loc}}$, *i.e., locally integrable. Then there exists a unique (up to stochastic indistinguishability) predictable process* $\widetilde{A} = (\widetilde{A}_t, \mathscr{F}_t)_{t \geqslant 0}$ *(the compensator of* A*) such that* $A - \widetilde{A}$ *is a local martingale. If, in addition,* A *is a nondecreasing process, then* $\widetilde{A}$ *is also nondecreasing.*

3. Now, we consider the concepts of quadratic variation and quadratic covariance for semimartingales, which play important roles in stochastic calculus. (For example, these characteristics are explicitly involved in Itô's formula presented below, in § 5c.)

For the discrete-time case we introduced the quadratic variation of a martingale in Chapter II, § 1b. This definition is extended to the case of continuous time as follows.

DEFINITION 1. The *quadratic covariance* of two semimartingales, X and Y, is the process $[X, Y] = ([X, Y]_t, \mathscr{F}_t)_{t \geqslant 0}$ such that

$$[X, Y]_t = X_t Y_t - \int_0^t X_{s-}\, dY_s - \int_0^t Y_{s-}\, dX_s - X_0 Y_0. \tag{5}$$

(Note that the stochastic integrals in (5) are well defined because the left-continuous processes (X_{t-}) and (Y_{t-}) are locally bounded.)

DEFINITION 2. The *quadratic variation* of a semimartingale X is the process $[X, X] = ([X, X]_t, \mathscr{F}_t)_{t \geqslant 0}$ such that

$$[X, X]_t = X_t^2 - 2 \int_0^t X_{s-}\, dX_s - X_0^2. \tag{6}$$

One also writes $[X]$ in place of $[X, X]$ (cf. formula (10) in Chapter II, § 1b).

We note that, by definitions (5) and (6) we immediately obtain the *polarization* formula

$$[X, Y] = \frac{1}{4}\big([X + Y, X + Y] - [X - Y, X - Y]\big). \tag{7}$$

The names 'quadratic variation' and 'quadratic covariance' given to $[X, Y]$ and $[X, X]$ can be explained by the following arguments.

Let $T^{(n)}$, $n \geqslant 1$, be Riemann sequences (see § 5a.8) and let

$$S_t^{(n)}(X, Y) = \sum_m \big(X_{\tau^{(n)}(m+1)\wedge t} - X_{\tau^{(n)}(m)\wedge t}\big)\big(Y_{\tau^{(n)}(m+1)\wedge t} - Y_{\tau^{(n)}(m)\wedge t}\big). \tag{8}$$

Then

$$\sup_{u\leqslant t} \bigl|S_u^{(n)}(X,Y) - [X,Y]_u\bigr| \xrightarrow{\mathsf{P}} 0 \quad \text{as} \quad n\to\infty \tag{9}$$

for each $t>0$; in particular,

$$\sup_{u\leqslant t} \bigl|S_u^{(n)}(X,X) - [X,X]_u\bigr| \xrightarrow{\mathsf{P}} 0 \quad \text{as} \quad n\to\infty. \tag{10}$$

To prove (10) it suffices to observe that

$$S_u^{(n)}(X,X) = X_u^2 - X_0^2 - 2T^{(n)}(X_-\cdot X)_u \tag{11}$$

(see (6)) and that, by property (19) in § 5a,

$$T^{(n)}(X_-\cdot X)_u \to (X_-\cdot X)_u = \int_0^u X_{s-}\, dX_s.$$

Now, relation (9) is a consequence of (10) and polarization formula (7).

4. By (10) we obtain that $[X,X]$ is a nondecreasing process. Since it is cádlǵ, it follows that $[X,X]\in\mathscr{V}^+$, therefore (by (7)) the process $[X,Y]$ also has bounded variation ($[X,Y]\in\mathscr{V}$), provided that $Y\in\mathscr{V}^+$.

This, together with (5), proves the following result: *a product of two semimartingales is itself a semimartingale.*

Rewriting (5) as

$$X_tY_t = X_0Y_0 + \int_0^t X_{s-}\, dY_s + \int_0^t Y_{s-}\, dX_s + [X,Y]_t, \tag{12}$$

where $[X,Y]$ is the process defined by (7), we can regard this relation as the *formula of integration by parts for semimartingales.* Written in the differential form it is even more transparent:

$$d(XY) = X_-\, dY + Y_-\, dX + d[X,Y]. \tag{13}$$

Clearly, this is essentially a *semimartingale* generalization of the classical relation

$$\Delta(X_nY_n) = X_{n-1}\,\Delta Y_n + Y_{n-1}\,\Delta X_n + \Delta X_n\,\Delta Y_n \tag{14}$$

for sequences $X=(X_n)$ and $Y=(Y_n)$.

A list of properties of the quadratic variation and the quadratic covariance under various assumptions about X and Y can be found, e.g., in the book [250; Chapter I, § 4e].

We point out only several properties, where we always assume that $Y\in\mathscr{V}^+$:

a) if one of the semimartingales X and Y is continuous, then $[X,Y]=0$;

b) if X is a semimartingale and Y is a predictable process of bounded variation, then

$$[X,Y]=\Delta Y\cdot X \qquad \text{and} \qquad XY = Y\cdot X + X_-\cdot Y;$$

c) if Y is a predictable process and X is a local martingale, then $[X,Y]$ is a local martingale;

d) $\Delta[X,Y] = \Delta X\,\Delta Y$.

5. Besides the processes $[X,Y]$ and $X,X]$ just defined (which are often called 'square brackets') an important position in stochastic calculus is occupied by the processes $\langle X,Y\rangle$ and $\langle X,X\rangle$ ('angle brackets') that we define next.

Let $M = (M_t, \mathscr{F}_t)_{t\geqslant 0}$ be a square integrable martingale in the class $\mathscr{H}^2$. Then, by Doob's inequality (see, e.g., [303], and also (36) in § 3b for the case of a Brownian motion),

$$\mathsf{E}\sup_t M_t^2 \leqslant 4\sup_t \mathsf{E}M_t^2 < \infty. \tag{15}$$

Hence the process M^2, which is a submartingale (by Jensen's inequality), belongs to the class (D). In view of the Doob–Meyer decomposition, there exists a nondecreasing predictable integrable process, which we denote by $\langle M,M\rangle$ or $\langle M\rangle$, such that the difference $M^2 - \langle M,M\rangle$ is a square integrable martingale.

If $M \in \mathscr{H}^2_{\mathrm{loc}}$, then carrying out a suitable localization we can verify that there exists a nondecreasing predictable process (which we define again by $\langle M,M\rangle$ or $\langle M\rangle$) such that $M^2 - \langle M\rangle$ is a locally square integrable martingale.

For two martingales, M and N, in $\mathscr{H}^2_{\mathrm{loc}}$ their 'angle bracket' $\langle M,N\rangle$ is defined by the formula

$$\langle M,N\rangle = \frac{1}{4}\big(\langle M+N, M+N\rangle - \langle M-N, M-N\rangle\big). \tag{16}$$

It is straightforward that $\langle M,N\rangle$ is a predictable process of bounded variation and $MN - \langle M,N\rangle$ is a local martingale.

It follows from (5) that $MN - [M,N]$ is also a local martingale. Hence if the martingales M and N are in the class $\mathscr{H}^2_{\mathrm{loc}}$, then $[M,N] - \langle M,N\rangle$ is a local martingale.

In accordance with Corollary 2 in subsection 2, one calls the predictable process $\langle M,N\rangle$ the *compensator* of $[M,N]$ and one often writes

$$\langle M,N\rangle = \widetilde{[M,N]}. \tag{17}$$

Remark 2. In the discrete-time case, given two square integrable martingales, $M = (M_n, \mathscr{F}_n)$ and $N = (N_n, \mathscr{F}_n)$, the corresponding sequences $[M,N]$ and $\langle M,N\rangle$ are defined as follows:

$$[M,N]_n = \sum_{k=1}^n \Delta M_k\,\Delta N_k \tag{18}$$

and

$$\langle M,N\rangle_n = \sum_{k=1}^n \mathsf{E}(\Delta M_k\,\Delta N_k \mid \mathscr{F}_{k-1}), \tag{19}$$

where $\Delta M_k = M_k - M_{k-1}$ and $\Delta N_k = N_k - N_{k-1}$ (cf. Chapter II, § 1b.)

Relation (19) suggests the following (formal, but giving one a clear idea) representation for the quadratic covariance in the continuous-time case:

$$\langle M, N\rangle_t = \int_0^t \mathsf{E}(dM_s\, dN_s \mid \mathscr{F}_s). \tag{20}$$

(Cf. the formula for the A_t in Remark 3 at the end of subsection 1.)

6. We now discuss the definition of 'angle brackets' for semimartingales and their connection with 'square brackets'.

Let $X = (X_t, \mathscr{F}_t)_{t\geqslant 0}$ be a semimartingale with decomposition $X = X_0 + M + A$, where M is a local martingale ($M \in \mathscr{M}_{\text{loc}}$) and A is a process of bounded variation ($A \in \mathscr{V}$).

Besides the already used representation of the local martingale M as the sum $M = M_0 + M' + M''$ with $M'' \in \mathscr{V} \cap \mathscr{M}_{\text{loc}}$ and $M' \in \mathscr{H}^2_{\text{loc}} \cap \mathscr{M}_{\text{loc}}$ (see formula (11) in § 5a), each local martingale M can be represented (moreover, in a *unique* way) as a sum

$$M = M_0 + M + M^d, \tag{21}$$

where $M^c = (M_t^c, \mathscr{F}_t)_{t\geqslant 0}$ is a continuous local martingale and $M^d = (M_t^d, \mathscr{F}_t)_{t\geqslant 0}$ is a *purely discontinuous* local martingale. (We say that a local martingale X is purely discontinuous if $X_0 = 0$ and it is *orthogonal* to each continuous martingale Y, i.e., XY is a local martingale; see [250; Chapter I, § 4b] for detail).

Hence each semimartingale X can be represented as the sum

$$X = X_0 + M^c + M^d + A.$$

Remarkably, the *continuous* martingale component M^c of X is unambiguously defined (this is a consequence of the Doob–Meyer decomposition; see [250; Chapter I, 4.18 and 4.27] for a discussion), which explains why it is commonly denoted by X^c.

Clearly, $[X^c, X^c] = \langle X^c, X^c\rangle$; moreover, one can prove ([250; Chapter I, 4.52]) that

$$[X, X]_t = \langle X^c, X^c\rangle_t + \sum_{s\leqslant t} (\Delta X_s)^2 \tag{22}$$

and, for two semimartingales, X and Y,

$$[X, Y]_t = \langle X, Y\rangle_t + \sum_{s\leqslant t} \Delta X_s\, \Delta Y_s, \tag{23}$$

where we set by definition $\langle X, Y\rangle_t \equiv \langle X^c, X^c\rangle_t$.

Remark 3. For local martingales M we have

$$\sum_{s\leqslant t} (\Delta M_s)^2 < \infty \qquad (\mathsf{P}\text{-a. s.}) \tag{24}$$

for each $t > 0$ and

$$[M, M] \in \mathscr{A}_{\text{loc}} \tag{25}$$

(see, e.g., [250; Chapter I, § 4]). Since processes A of bounded variation satisfy the inequality $\sum_{s \leqslant t} |\Delta A_s| < \infty$ (P-a. s.) for each $t > 0$, it follows that

$$\sum_{s \leqslant t} (\Delta A_s)^2 < \infty \qquad \text{(P-a. s.)} \tag{26}$$

for the corresponding component of an arbitrary semimartingale $X = X_0 + M + A$. Hence

$$\sum_{s \leqslant t} (\Delta X_s)^2 < \infty \qquad \text{(P-a. s.)}$$

for each $t > 0$, and the right-hand sides of (22) and (23) are well defined and finite (P-a. s.).

§ 5c. Itô's Formula for Semimartingales. Generalizations

1. THEOREM (Itô's formula). *Let*

$$X = (X^1, \dots, X^d)$$

be a d-dimensional semimartingale and let $F = F(x_1, \dots, x_d)$ be a C^2-function in $\mathbb{R}^d$.

Then the process $F(X)$ is also a semimartingale and

$$\begin{aligned} F(X_t) = F(X_0) &+ \sum_{i \leqslant d} \big(D_i F(X_-)\big) \cdot X^i \\ &+ \frac{1}{2} \sum_{i,j \leqslant d} \big(D_{ij} F(X_-)\big) \cdot \langle X^{i,c}, X^{j,c} \rangle \\ &+ \sum_{s \leqslant t} \Big[F(X_s) - F(X_{s-}) - \sum_{i \leqslant d} D_i F(X_{s-}) \, \Delta X_s^i \Big], \end{aligned} \tag{1}$$

where $D_i F = \dfrac{\partial F}{\partial x_i}$, $D_{ij} F = \dfrac{\partial^2 F}{\partial x_i \partial x_j}$.

The proof can be found in many textbooks on stochastic calculus (e.g., [103], [248], or [250]).

2. We consider now two examples demonstrating the efficiency of Itô's formula in various issues of stochastic calculus.

EXAMPLE 1 (Doléans' equation and stochastic exponential). Let $X = (X_t, \mathscr{F}_t)_{t \geqslant 0}$ be a fixed semimartingale. We consider the problem of finding, in the class of càdlàg processes (processes with right-continuous trajectories having limits from the left) a solution $Y = (Y_t, \mathscr{F}_t)_{t \geqslant 0}$ to Doléans's equation

$$Y_t = 1 + \int_0^t Y_{s-}\, dX_s, \tag{2}$$

or, in the differential form,

$$dY = Y_-\, dX, \qquad Y_0 = 1. \tag{3}$$

By Itô's formula for the two processes

$$X_t^1 = X_t - X_0 - \frac{1}{2}\langle X^c, X^c\rangle_t$$

and

$$X_t^2 = \prod_{0<s\leqslant t} (1 + \Delta X_s) e^{-\Delta X_s}, \qquad X_0^2 = 1, \tag{4}$$

and the function $F(x_1, x_2) = e^{x_1} \cdot x_2$, we obtain that the process

$$\mathscr{E}(X)_t = F(X_t^1, X_t^2), \qquad t \geqslant 0,$$

i.e., the process

$$\mathscr{E}(X) = \big(\mathscr{E}(X)_t, \mathscr{F}_t\big)_{t\geqslant 0} \tag{5}$$

such that

$$\mathscr{E}(X)_t = e^{X_t - X_0 - \frac{1}{2}\langle X^c, X^c\rangle_t} \prod_{0<s\leqslant t} (1 + \Delta X_s) e^{-\Delta X_s}, \tag{6}$$

1) is a semimartingale;
2) satisfies Doléans's stochastic equation (2).

Moreover, the process $\mathscr{E}(X)$ (the Doléans *stochastic exponential*) is a unique (up to stochastic distinguishability) solution of (2) in the class of processes with càdlàg trajectories. (See the proof and a generalization to the case of complex-valued semimartingales in [250; Chapter I, § 4f].)

EXAMPLE 2 (Lévy's theorem; § 3b). Let $X = (X_t, \mathscr{F}_t)_{t\geqslant 0}$ be a continuous local martingale with $\langle X\rangle_t \equiv t$. Then X is a *Brownian motion.*

For the proof we consider the function $F(x) = e^{i\lambda x}$ and use Itô's formula (separately for the real and the imaginary parts). Then we obtain

$$e^{i\lambda X_t} = 1 + i\lambda \int_0^t e^{i\lambda X_s}\, dX_s - \frac{\lambda^2}{2}\int_0^t e^{i\lambda X_s}\, ds. \tag{7}$$

The integral $\int_0^t e^{i\lambda X_s}\, dX_s$ is a local martingale. We set $\tau_n = \inf\{t\colon |X_t| \geqslant n\}$. Let $\mathsf{E}(\,\cdot\,; A)$ be averaging over the set A. If $A \in \mathscr{F}_0$, then we see from (7) that

$$\mathsf{E}\bigl(e^{i\lambda X_{t\wedge\tau_n}}; A\bigr) = 1 - \frac{\lambda^2}{2}\mathsf{E}\biggl(\int_0^{t\wedge\tau_n} e^{i\lambda X_s}\, ds; A\biggr). \tag{8}$$

Hence, as $n \to \infty$, we obtain the following relation for $f_A(t) = \mathsf{E}(e^{i\lambda X_t}; A)$, $t \geqslant 0$:

$$f_A(t) = 1 - \frac{\lambda^2}{2}\int_0^t f_A(s)\, ds, \qquad t \geqslant 0; \tag{9}$$

moreover, it is clear that $|f_A(t)| \leqslant 1$ and $f_A(0) = \mathsf{P}(A)$.

The only solution of (9) satisfying these conditions is as follows:

$$f(A) = e^{-\frac{\lambda^2}{2}t}\,\mathsf{P}(A). \tag{10}$$

Hence

$$\mathsf{E}\bigl(e^{i\lambda X_t} \mid \mathscr{F}_0\bigr) = e^{-\frac{\lambda^2}{2}t}.$$

In a similar way we can show that for all s and t, $s \leqslant t$, we have

$$\mathsf{E}\bigl(e^{i\lambda(X_t - X_s)} \mid \mathscr{F}_s\bigr) = e^{-\frac{\lambda^2}{2}(t-s)}. \tag{11}$$

Consequently, $X = (X_t, \mathscr{F}_t)_{t\geqslant 0}$ is a process with independent Gaussian increments, $\mathsf{E}(X_t - X_s) = 0$, and $\mathsf{E}(X_t - X_s)^2 = t - s$. Hence, by our assumption of the continuity of the trajectories X is a Brownian motion (see the definition in § 1b).

3. Itô's formula, which is an important tool of stochastic calculus, can be generalized in several directions: to *nonsmooth* functions $F = F(x_1, \dots, x_d)$ and *semimartingales* $X = (X_1, \dots, X_d)$, to *smooth* functions $F = F(x_1, \dots, x_d)$ and *nonsemimartingales* $X = (X_1, \dots, X_d)$, and so on.

Following [166] and using some ideas from this paper we present here several results in this direction.

a) Let $d = 1$ and let X be a *standard Brownian motion* $B = (B_t)_{t\geqslant 0}$. Let $F = F(x)$ be an absolutely continuous function, i.e.,

$$F(x) = F(0) + \int_0^x f(y)\,dy; \tag{12}$$

here we assume that the (measurable) function $f = f(y)$ belongs to the class $L^2_{\mathrm{loc}}(\mathbb{R}^1)$, i.e.,

$$\int_{|x|\leqslant K} f^2(x)\,dx < \infty \tag{13}$$

for each $K > 0$.

Note that *neither* $f(B) = (f(B_t))_{t\geqslant 0}$ *nor* $F = (F(B_t))_{t\geqslant 0}$ are semimartingales in general, therefore we cannot use Itô's formula in the form (1).

However, it is proved in [166] that we have the following relation:

$$F(B_t) = F(0) + \int_0^t f(B_s)\,dB_s + \frac{1}{2}[f(B), B]_t, \tag{14}$$

where $[f(B), B]$ is the *quadratic covariance* of the processes $f(B)$ and B defined as follows:

$$[f(B), B]_t = \mathsf{P}\text{-}\lim_{n\to\infty} \sum_m \Big(f\big(B_{t^{(n)}(m+1)\wedge t}\big) - f\big(B_{t^{(n)}(m)\wedge t}\big)\Big) \times \Big(B_{t^{(n)}(m+1)\wedge t} - B_{t^{(n)}(m)\wedge t}\Big); \tag{15}$$

here the $T^{(n)} = \{t^{(n)}(m),\ m \geqslant 1\}$, $n \geqslant 1$, are Riemannian sequences of (deterministic) times $t^{(n)}$ defined in § 5a.8.

We emphasize that the process $f(B)$ is not a semimartingale in general, therefore, the mere existence of the limit (in probability P) in (15) is nontrivial. One of the results of [166] is the proof of the *existence* of this limit.

COROLLARY 1. *Let $F(x) \in C^2$. Then*

$$[f(B), B]_t = \int_0^t f'(B_s)\,ds$$

and (14) coincides with Itô's formula.

COROLLARY 2. *Let $F(x) = |x|$. Then $[f(B), B]_t = 2L_t(0)$, where*

$$L_t(0) = \lim_{\varepsilon\downarrow 0} \frac{1}{2\varepsilon} \int_0^t I(|B_s| \leqslant \varepsilon)\,ds \tag{16}$$

is the local time that the Brownian motion "spends" at zero over the period $[0, t]$. *Hence, from* (14) *we obtain Tanaka's formula*

$$|B_t| = \int_0^t \operatorname{Sgn} B_s \, dB_s + L_t(0) \tag{17}$$

(cf. the example in Chapter II, § 1b and § 3e in this chapter).

Note that the process $|B| = (|B_t|, \mathscr{F}_t)_{t \geqslant 0}$ is a submartingale, the stochastic integral is a martingale, and $L(0) = (L_t(0), \mathscr{F}_t)$ is a continuous (and therefore, predictable) nondecreasing process.

Hence we can regard (17) as an *explicit* formula for the Doob–Meyer decomposition of the semimartingale $|B|$.

b) Again, let $d = 1$, and let X be a *fractional Brownian motion* $B^{\mathbb{H}} = (B_t^{\mathbb{H}})_{t \geqslant 0}$ with Hurst parameter $\mathbb{H} \in \left(\frac{1}{2}, 1\right]$; see § 2c. As before, let $T^{(n)} = \{t^{(n)}(m), m \geqslant 1\}$, $n \geqslant 1$, be Riemann sequences of (deterministic) times $t^{(n)}(m)$.

Since $\mathsf{E}|\Delta B_t^{\mathbb{H}}|^2 = |\Delta t|^{2\mathbb{H}}$ (see (6) in § 2c), it follows that

$$\lim_{n \to \infty} \mathsf{E}\left[\sum_m \left|B^{\mathbb{H}}_{t^{(n)}(m+1) \wedge t} - B^{\mathbb{H}}_{t^{(n)}(m)}\right|^2\right] = 0,$$

and therefore, for $\mathbb{H} \in \left(\frac{1}{2}, 1\right]$ we have

$$\mathsf{P}\text{-}\lim_{n \to \infty} \sum_m \left|B^{\mathbb{H}}_{t^{(n)}(m+1) \wedge t} - B^{\mathbb{H}}_{t^{(n)}(m)}\right|^2 = 0 \tag{18}$$

for the limit in probability.

Remark. If $\mathbb{H} = \frac{1}{2}$, then the corresponding limit is equal to t, whereas it is $+\infty$ for $\mathbb{H} \in (0, \frac{1}{2})$. One usually calls processes with property (18) *zero-energy* processes (see Chapter IV, § 3a.6 and also, e.g., [166]).

Let $F = F(x)$ be a C^2-function and let $f(x) = F'(x)$.

For a Brownian motion $B = (B_t)_{t \geqslant 0}$ we have

$$dF(B_t) = F'(B_t) \, dB_t + \frac{1}{2} F''(B_t) \, (dB_t)^2, \tag{19}$$

which (on having agreed that $(dB_t)^2 = dt$—see § 3d for details) gives one the following *Itô's formula* written in the differential form:

$$dF(B_t) = F'(B_t) \, dB_t + \frac{1}{2} F''(B_t) \, dt. \tag{20}$$

Property (18) holds for a fractal Brownian motion $B^{\mathbb{H}} = (B_t^{\mathbb{H}})_{t \geqslant 0}$ with parameter $\frac{1}{2} < \mathbb{H} \leqslant 1$, which makes the following agreement look fairly natural:

$$(dB_t^{\mathbb{H}})^2 = 0. \tag{21}$$

(Cf. (12)–(14) in § 3d.)

If we replace B by $B^{\mathbb{H}}$ in the expansion (19), then, in view of (21), we arrive (only formally at this stage) at the representation (recall that $f = F'$)

$$dF(B_t^{\mathbb{H}}) = f(B_t^{\mathbb{H}})\,dB_t^{\mathbb{H}}, \tag{22}$$

which, in accordance with the standard conventions of stochastic calculus, must be interpreted in the integral form: for $s < t$ we have

$$F(B_t^{\mathbb{H}}) - F(B_s^{\mathbb{H}}) = \int_s^t f(B_u^{\mathbb{H}})\,dB_u^{\mathbb{H}} \quad (\mathsf{P}\text{-a.s.}). \tag{23}$$

We shall now prove this formula restricting ourselves for simplicity to the case of $F \in C^2$ and explaining also on the way how one should interpret the 'stochastic integral' in (23).

By Taylor's formula with remainder in the integral form

$$F(x) = F(y) + f(y)(x-y) + \int_y^x f'(u)(x-u)\,du.$$

Hence, proceeding as in [166] and [299], for each $T^{(n)}$-partitioning, $n \geqslant 1$, we obtain

$$\begin{aligned} F(B_t^{\mathbb{H}}) - F(B_0^{\mathbb{H}}) &= \sum_m \bigl[F\bigl(B^{\mathbb{H}}_{t\wedge t^{(n)}(m+1)}\bigr) - F\bigl(B^{\mathbb{H}}_{t\wedge t^{(n)}(m)}\bigr)\bigr] \\ &= \sum_m f\bigl(B^{\mathbb{H}}_{t\wedge t^{(n)}(m+1)}\bigr)\bigl(B^{\mathbb{H}}_{t\wedge t^{(n)}(m+1)} - B^{\mathbb{H}}_{t\wedge t^{(n)}(m)}\bigr) + R_t^{(n)}, \end{aligned} \tag{24}$$

where

$$R_t^{(n)} = \sum_m \int_{B^{\mathbb{H}}_{t\wedge t^{(n)}(m)}}^{B^{\mathbb{H}}_{t\wedge t^{(n)}(m+1)}} f'(u)\bigl(B^{\mathbb{H}}_{t\wedge t^{(n)}(m+1)} - u\bigr)\,du.$$

Clearly, $\mathsf{P}\Bigl(\sup\limits_{0\leqslant u\leqslant t} |f'(B_t^{\mathbb{H}})| < \infty\Bigr) = 1$ and, in view of (18),

$$|R_t^{(n)}| \leqslant \frac{1}{2} \sup_{0\leqslant u\leqslant t} |f'(B_u^{\mathbb{H}})| \cdot \sum_m \bigl|B^{\mathbb{H}}_{t\wedge t^{(n)}(m+1)} - B^{\mathbb{H}}_{t\wedge t^{(n)}(m)}\bigr|^2 \xrightarrow{\mathsf{P}} 0. \tag{25}$$

The left-hand side of (24) is independent of n and $R_t^{(n)} \xrightarrow{\mathsf{P}} 0$, therefore *there exists*

$$\lim_n \sum_m f\bigl(B^{\mathbb{H}}_{t\wedge t^{(n)}(m)}\bigr)\bigl(B^{\mathbb{H}}_{t\wedge t^{(n)}(m+1)} - B^{\mathbb{H}}_{t\wedge t^{(n)}(m)}\bigr) \tag{26}$$

(in the sense of convergence in measure), which we denote by

$$\int_0^t f(B_u^{\mathbb{H}})\,dB_u^{\mathbb{H}} \tag{27}$$

and call the *stochastic integral with respect to the fractional Brownian motion* $B^{\mathbb{H}} = (B_u^{\mathbb{H}})_{t\geqslant 0}$ ($\mathbb{H} \in (\frac{1}{2}, 1]$, $f \in C^1$).

Simultaneously, these arguments also prove required formula (23), which can be regarded as an analogue of *Itô's formula* for a fractional Brownian motion.

EXAMPLE 3. If $F(x) = x^2$, then (23) shows that for a fractional Brownian motion $B^{\mathbb{H}}$ with $\frac{1}{2} < \mathbb{H} \leqslant 1$ we have

$$d(B_t^{\mathbb{H}})^2 = 2B_t^{\mathbb{H}}\, dB_t^{\mathbb{H}}. \tag{28}$$

Recall that for a Brownian motion $B = B^{1/2}$,

$$d(B_t)^2 = 2B_t\, dB_t + dt. \tag{29}$$

EXAMPLE 4. If $F(x) = e^x$, then

$$d\big(e^{B_t^{\mathbb{H}}}\big) = e^{B_t^{\mathbb{H}}}\, dB_t^{\mathbb{H}} + \frac{1}{2} e^{B_t^{\mathbb{H}}}\, dt. \tag{30}$$

For a Brownian motion $B = B^{1/2}$ we have

$$d\big(e^{B_t}\big) = e^{B_t}\, dB_t + \frac{1}{2} e^{B_t}\, dt. \tag{31}$$

Our considerations could be extended also to the case of several fractional Brownian motions. For instance, assume that

$$(X_0, X_1, \dots, X_d) \equiv (B^{\mathbb{H}_0}, B^{\mathbb{H}_1}, \dots, B^{\mathbb{H}_d}),$$

where $\mathbb{H}_0 = \frac{1}{2}$ (so that $B^{\mathbb{H}_0}$ is a standard Brownian motion) and $\frac{1}{2} < \mathbb{H}_1 < \mathbb{H}_2 < \cdots < \mathbb{H}_d \leqslant 1$. If $F = F(t, x_0, x_1, \dots, x_d) \in C^{1,2,\dots,2}$, then (cf. (11) in § 3d)

$$dF = \frac{\partial F}{\partial t}\, dt + \left[\frac{\partial F}{\partial x_0}\, dX_0 + \frac{1}{2}\, \frac{\partial^2 F}{\partial x_0^2}\, (dX_0)^2\right] + \sum_{i=1}^{d} \frac{\partial F}{\partial x_i}\, dX_i. \tag{32}$$

EXAMPLE 5. We consider the process

$$S_t = e^{(\mu - \frac{\sigma_0^2}{2})t + \sigma_0 B_t + \sigma_1 B_t^{\mathbb{H}}}, \tag{33}$$

where $B = B^{1/2}$ is a standard Brownian motion and $B^{\mathbb{H}}$ is a fractional Brownian motion with $\frac{1}{2} < \mathbb{H} \leqslant 1$. Then $S = (S_t)_{t \geqslant 0}$ can be regarded as the solution of the stochastic differential equation

$$dS_t = S_t\big(\mu\, dt + \sigma_0\, dB_t + \sigma_1\, dB_t^{\mathbb{H}}\big) \tag{34}$$

with $S_0 = 1$. (Cf. (2) and (4) in § 4b.)

Chapter IV. Statistical Analysis of Financial Data

1. Empirical Data. Probabilistic and Statistical Models of Their Description. Statistics of 'Ticks'

§ 1a. Structural Changes in Financial Data Gathering and Analysis

1. Looking back to the changes in the empirical analysis of time-related financial data, one discovers the following features.

In the 1970s and earlier, one mostly operated with data averaged over *large* time intervals: a year, a quarter, a month, a week. Among the most widespread probabilistic and statistical models (developed for the description of the behavior of the logarithms of financial indices) at that time (and nowadays) there were models of random walk type (see Chapter I, § 2a), autoregressive, and moving average models, their combinations, and so on (see Chapter II, § 1d). It should be noted that most of these models were *linear*.

In the 1980s, in connection with the analysis of *daily* data, one saw the need to invoke *nonlinear* models; the *ARCH* and *GARCH* models and their various modifications (see Chapter II, § 1d) are the best known examples of these.

The analysis of *interday* data has become feasible in the 1990s. This is primarily a result of the general progress in computer technology and telecommunications accompanied by a sharp improvement (as compared with the 'paper-based' technology of data storage and processing) in the methods of the collection, registration, storage, and analysis of statistical information, arriving in a virtually *incessant* flow.

2. Besides routine news coming even through daily newspapers or TV (for example, the exchange rates, several indexes, the opening and the closing trading prices of major stocks and commodities, and so on), several information agencies (Reuters, Telerate, Knight Ridder, Bloomberg, to mention just a few) deliver on-line megabytes of most diversified financial information to their customers: for instance, one can learn about the recent bid and ask currency prices as well as the name and the location of the bidder.

Below is an example of a message from Reuters concerning the *exchange rates* (against the US dollar) that is displayed on its customers' monitors immediately after 7 h 27 GMT (see [204]):

0727	DEM	RABO	RABOBANK	UTR	1.6290/00	DEM	1.6365	1.6270
0727	FRF	BUEX	UECIC	PAR	5.5620/30	FRF	5.5835	5.5588
0726	NLG	RABO	RABOBANK	UTR	1.8233/38	NLG	1.8309	1.8220

Here 0727 and 0726 is the time of quotation announcements; DEM, FRF, and NLG are the abbreviations for particular currencies (the German mark, the French franc, and the Dutch florin); RABO and BUEX are RABOBANK in Utrecht (UTR) and UECIC bank and Paris (PAR), respectively; 1.6290 is the bid price; 00 following this bid price indicates the ask price, 1.6300; 1.6365 and 1.6270 are the high and the low prices of the last 24 hours. On the third line, 0726 means that RABOBANK announced (or maintained) its florin quotations (1.8233/38) at 7 h 26 and nobody (including RABOBANK itself) has announced new quotations since.

Choosing $t_0 = 7$ h 27 as time zero we shall see that, say, the *ask price of the dollar against the German mark*

$$S_t^a = \left(\frac{\text{DEM}}{\text{USD}}\right)_t^a, \qquad t \geqslant t_0,$$

behaves as in Fig. 28. (Cf. Fig. 6 in Chapter I, § 1b.)

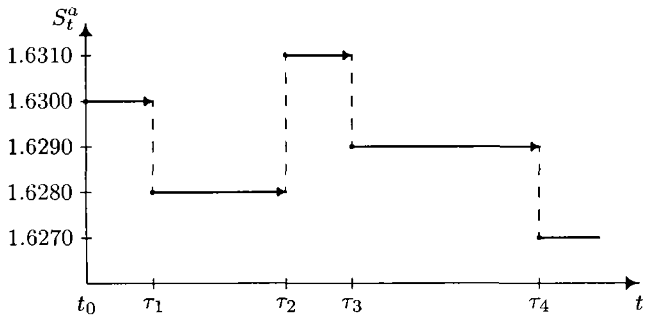


FIGURE 28. Behavior of the cross rate $S_t^a = \left(\frac{\text{DEM}}{\text{USD}}\right)_t^a$, $t \geqslant t_0$

That is, the price S_t^a keeps its value for some time (on the interval $[0, \tau_1)$); then, at the instant τ_1, it drops (one speaks about a 'tick' indicating that some bank announced another quotation $S_{\tau_1}^a$ at time τ_1), and so on.

This raises two questions:

(I) what can be said on the statistics of the lengths of the 'intertick' intervals $(\tau_{k+1} - \tau_k)$;

(II) what can be said on the statistics of the changes in price (absolute changes $S_{\tau_{k+1}}^a - S_{\tau_k}^a$ or relative changes $S_{\tau_{k+1}}^a / S_{\tau_k}^a$)?

Extracting this information from the available data is the *primary task of the statistical analysis* of the exchange rates or data on other financial indexes, which often exhibit patterns of dynamics similar to the above one.

Clearly, the function of such an analysis is the construction of *probabilistic and statistical models* of such processes as ask prices (S_t^a) or bid prices (S_t^b). This is, in the long run, important also for understanding the evolution of financial indexes, the pricing mechanisms, and for predictions of the price development in the future.

It should be noted that registering and processing statistical data, storing it in a form allowing an easy recovery is a difficult task, impossible without advanced technology. Still, it is equally clear that an access to the results of statistical analysis packed in a ready-to-use form gives one indisputable advantages in the securities markets as regards building an 'efficient portfolio', making rational investment decisions, the assessment of investment projects, securities, and so on.

3. The almost incessant registration and processing of statistical data, which have been made possible by the modern technology, reveals a *high-frequency* pattern of the behavior of financial indexes, which evolve in a fairly chaotic manner. This pattern is not discernible after discretization (with respect to time or the phase variables). Consequently, the appearance in financial mathematics of problems dealing with 'high-frequency' is the result of the new opportunities offered by almost *continuous* gathering of statistical data. It is this advanced technology of data gathering that enabled one to discover, besides the presence of *high frequency components*, several other peculiarities of the dynamics of financial indexes. Not plunging into detail here[a], we point out, for instance, the *nonlinear mode* of the formation of financial indexes and the *aftereffect*, the capacity of many indexes, prices, and so on, to *'remember' their past*.

4. To give one a notion of the frequency of ticks in currency *cross rates*, developing as in the above chart, and also an idea of the *bulk* of the available statistical information, we now cite the data of Olsen & Associates, an institute mentioned in the footnote on this page; see [90], [91], and also [204].

Over the period January 1, 1987–December 31, 1993, Reuters registered

$$8\,238\,532$$

ticks in the DEM/USD cross rate. Of these, 1 466 946 ticks occurred during one year, from October 1, 1992 to September 30, 1993. Over the same period, Reuters registered 570 814 ticks of the exchange rate JPY/USD (see also the table in § 1b).

[a]For that, see the proceedings [393] of the HFDF-1 conference organized by the Research Institute for Applied Economics Olsen & Associates (March 29–31, 1995, Zürich, Switzerland). The introductory talk "High Frequency Data in Financial Markets: Issues and Applications" by C. A. E. Goodhart and M. O'Hara provides a brilliant introduction into the range of problems arising in connection with 'high frequencies', describes new phenomena, peculiarities, and results, their interpretations, and suggests a program of a further research.

These are high-frequency data indeed: on a typical business day, there occurs on the average 4 500 ticks of the DEM/USD exchange rate and around 2 thousand ticks for JPY/USD. In July, 1994, the number of ticks of the DEM/USD rate was close to 9 000, that is, 15–20 ticks per minute. (On usual days, there occur on the average 3–4 ticks each minute.)

It should be noted that, of all the exchange rates, the rate DEM/USD is in general subject to most frequent changes. It is also worth noting that the quotations (bid and ask prices) cited by various agencies *are not the real transaction prices.* To our knowledge, the corresponding data or the statistics of the volumes of transactions are not easily accessible.

5. The above example relates to currency exchange rates; however, many other financial indexes behave in a similar way. We can refer to [127; p. 284] for a chart depicting the behavior of Siemens shares on the Frankfurt Stock Exchange on March 2, 1992, since the opening at 10 h 30. The pattern is the same as in Fig. 28: the price is flat for some time, then (at a random instant) it changes the value.

6. One can find extended statistics of stock prices 'ticks' in [217].

Various information relating to the ticks in the prices of all kinds of securities (including shares and bonds) can be obtained, for example, from

ISSM – the Institute for the Study of Securities Markets,
NYSE – the New York Stock Exchange.

Berkeley Options Data provides data on the ask and bid prices of options and on the instantaneous prices of CBOE (the Chicago Board of Options Exchange); in the Commodity Futures Trading Commission (CFTC) one can obtain data concerning the American futures market, while the current American stock prices can be, for example, obtained by e-mail, from the MIT Artificial Intelligence Laboratory (`http://www.stockmaster.com`).

§ 1b. Geography-Related Features of the Statistical Data on Exchange Rates

1. By contrast to, say, bourses, which can trade in stock, bonds, futures and are open only on 'trading days', the *currency exchange market*, the

FX-*market*

(Foreign Exchange or Forex) has several peculiarities that one would like to consider more closely.

It should be noted first of all that the FX-market is inherently *international*. It cannot be 'localized', has no premises (of the sort NYSE or CBOE have). Instead, this is an interweaving network of banks and exchange offices, equipped with modern communications, all over the world.

The FX-market operates *continuously round the clock*; it is more active during the five working days, and less active on week-ends or holidays (for example, on Easter Monday).

2. Judging by the periods of different *intensity* of currency trade during the day, one usually distinguishes the following *three geographic zones* (indicated is Greenwich Mean Time, GMT):

(1) the East-Asian, with center at Tokyo, 21:00–7:00;
(2) the European, with center at London, 7:00–13:00;
(3) the American, with center at New York, 13:00–20:00.

According to [D4], the *East-Asian zone* includes Australia, Hong Kong, India, Indonesia, Japan, South Korea, Malaysia, New Zealand, and Singapore.

The *European zone* covers Austria, Bahrain, Belgium, Germany, Denmark, Finland, France, Great Britain, Greece, Ireland, Italy, Israel, Jordan, Kuwait, Luxembourg, the Netherlands, Norway, Saudi Arabia, South Africa, Spain, Sweden, Switzerland, Turkey, and United Arab Emirates.

The *American zone* includes Argentina, Canada, Mexico, and the USA.

Sometimes one distinguishes four zones in place of the three in our scheme, where the fourth, *Pacific*, zone is included in the East Asian one.

To the already mentioned main centers of trade, Tokyo, London, and New York, one could add also Sydney, Hong Kong, Singapore (the East-Asian zone), Frankfurt-on-Main, Zurich (the European zone), and Toronto (the American zone).

Taking the American dollar (USD) as the basis, the bulk of the trade proceeds between the dollar and the following 'most important' currencies: the German mark (DEM), the Japanese yen (JPY), the British pound (GBP), and the Swiss franc (CHF) (we use the abbreviations adopted by the *International Organization for Standartization* (IOS), code 4217).

These 'basic' currencies (including, of course, the American dollar) are traded all over the globe. Other currencies are mostly traded in their geographic zones.

3. To gain some impression of the 'intensity' of currency exchange, that is, of the frequency of 'ticks' in the cross rates on the FX-market (a global market, as already mentioned) one can consider the following graph (borrowed from [181]) of the 'intensity' of changes in the DEM/USD exchange rate.

The *average* number of changes, 'ticks' occurring every 5 minutes (Fig. 29) or 20 minutes (Fig. 30) is plotted along the vertical axis and time is plotted along the horizontal axis.

In these charts one can distinguish *cycles*, which are clearly related to the rotation of Earth, and *inhomogeneity* in the number of changes ('ticks'). Three peaks of *activity*, related to the three geographic zones, are discernible.

The European and American peaks are fairly similar. The peak of activity in Europe occurs immediately after lunch, when the business day begins in America. The minimum of activity fall precisely at Tokyo lunch time, when it is night in Europe and America.

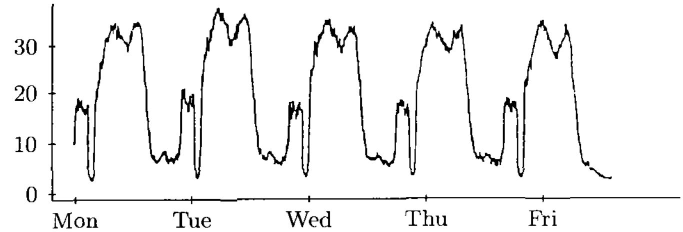


FIGURE 29. Intensity of the DEM/USD exchange rate from Monday through Friday, over 5-minute intervals

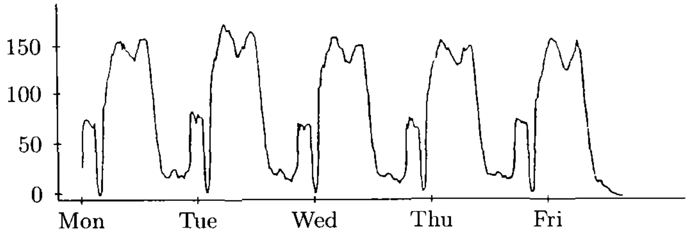


FIGURE 30. The same as in Fig. 29, over 20-minute intervals

We supplement Fig. 29 and 30 with the following chart from [427], depicting the *interday* intensity (the average number of 'ticks' occurring at each of the 24 hours) of the DEM/USD cross rate over the period October 5, 1992–September 26, 1993 (the data of Reuters).

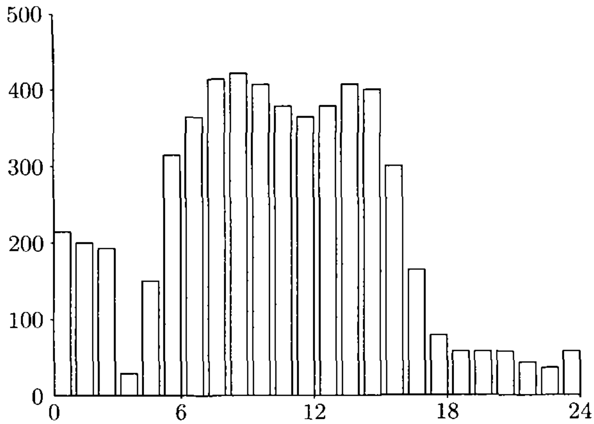


FIGURE 31. Average hourly number of 'ticks' during one day (24 hours) for the DEM/USD cross rate. Zero corresponds to 0h00 GMT

4. The FX-market is the largest financial market. According to Bank for International Settlements in Basel, Switzerland (1993), the *daily* turnover on this market was approximately \$ 832 bn ($832 \cdot 10^9$) in 1992!

One of the largest databases covering the FX-market is to our knowledge the database of Olsen & Associates (see the footnote in § 1a.4). The table based on its data gives one an idea of transactions on the FX-market in January 01, 1987–December 31, 1993.

Exchange rate	Total number of 'ticks' registered in the database	Average daily number of 'ticks' (52 weeks in a year, 5 business days in a week)
DEM/USD	8.238.532	4500
JPY/USD	4.230.041	2300
GBP/USD	3.469.421	1900
CHF/USD	3.452.559	1900
FRF/USD	2.098.679	1150
JPY/DEM	190.967	630
FRF/DEM	132.089	440
ITL/DEM	118.114	390
GBP/DEM	96.537	320
NLG/DEM	20.355	70

5. In the above description of the dynamics of exchange rates we were concerned only with the temporal aspect: the frequency (intensity) of changes as a function of time. Looking back at the classification in § 1a.2 we see that this discussion relates to issue (I), the statistics of 'intertick' intervals. However, we have said nothing so far about the character of these *changes* in the *values* of prices that occur at the instants of ticks. We turn to this question (issue (II)) in § 1d, while in the next section we shall consider the probabilistic and statistical models coming to mind in a natural way in connection with the trajectory behavior of the prices (S_t^a) and (S_t^a) (see Fig. 28).

§ 1c. Description of Financial Indexes as Stochastic Processes with Discrete Intervention of Chance

1. As already said, the monitor of, say, a Reuters customer interested the FX-market shows at each instant t two quotations: the ask price S_t^a and the bid price S_t^b of, say, the American dollar (USD) in German marks (DEM).

The difference $S_t^a - S_t^b$, the *spread*, is an important characteristic of the market situation. It is well known that the spread is positively correlated with the *volatilities* of the prices (which are understood *here* as the usual standard deviations). Thus, a rise in volatility (which increases risks, due to the decreasing accuracy of the predictions of the price development) prompts traders to widen the spread as a compensation for greater risks.

2. We now represent the prices S_t^a and S_t^b in the following form:

$$S_t^a = S_0^a e^{H_t^a}, \qquad S_t^b = S_0^b e^{H_t^b}. \tag{1}$$

We also set

$$H_t = \frac{1}{2}(H_t^a + H_t^b), \tag{2}$$

$$S_t = S_0 e^{H_t}, \quad \text{where} \quad S_0 = \sqrt{S_0^a \cdot S_0^b}\,. \tag{3}$$

Then

$$H_t = \ln \sqrt{S_t^a \cdot S_t^b} \tag{4}$$

is the logarithm of the geometric mean and

$$S_t = \sqrt{S_t^a \cdot S_t^b}\,. \tag{5}$$

It is the so-defined prices $S = (S_t)$ and their logarithms $H = (H_t)$ that one usually deals with in the analysis of exchange rates, so that the two *really existing* process (S_t^a) and (S_t^b) are reduced to a single one, (S_t).

The evolution of $S = (S_t)_{t\geqslant 0}$ and $H = (H_t)_{t\geqslant 0}$, where $H_t = \ln \dfrac{S_t}{S_0}$, in real ('physical') time t can be fairly adequately (see Fig. 28 in § 1a) described by stochastic processes with *discrete intervention of chance*:

$$S_t = S_0 + \sum_{k\geqslant 1} s_{\tau_k} I(\tau_k \leqslant t) \tag{6}$$

and

$$H_t = \sum_{k\geqslant 1} h_{\tau_k} I(\tau_k \leqslant t), \tag{7}$$

where $0 \equiv \tau_0 < \tau_1 < \tau_2 < \cdots$ are the successive instants of 'ticks', that is, the instants of price changes, and

$$s_{\tau_k} = \Delta S_{\tau_k}, \qquad h_{\tau_k} = \Delta H_{\tau_k}$$

are $\mathscr{F}_{\tau_k}$-measurable variables (here we set $\Delta S_{\tau_k} = S_{\tau_k} - S_{\tau_k-} = S_{\tau_k} - S_{\tau_{k-1}}$; the same holds for ΔH_{τ_k}).

Clearly,

$$h_{\tau_k} = \Delta H_{\tau_k} = \ln \frac{S_{\tau_k}}{S_{\tau_{k-1}}} = \ln\left(1 + \frac{\Delta S_{\tau_k}}{S_{\tau_{k-1}}}\right) = \ln\left(1 + \frac{s_{\tau_k}}{S_{\tau_{k-1}}}\right).$$

We already considered processes of types (6) and (7) in Chapter II, § 1f. We recall several concepts relating to these important processes, which we shall use in the present analysis of the evolution of currency cross rates and other financial indexes.

For the simplicity of notation we set

$$\xi_k \equiv s_{\tau_k} = \Delta S_{\tau_k}.$$

The probability distribution of the process $S = (S_t)_{t \geqslant 0}$ (which we denote by $\mathrm{Law}(S)$ or $\mathrm{Law}(S_t, t \geqslant 0)$) is completely defined by the *joint* distribution $\mathrm{Law}(\tau, \xi)$ of the *sequence*

$$(\tau, \xi) = (\tau_k, \xi_k)_{k \geqslant 1},$$

of 'ticks' τ_k and their 'marks' ξ_k. Such a sequence is commonly called a *marked point process* or a *multivariate point process* (see, for instance, [250; Chapter III, § 1c]).

The name *point process* is usually reserved for the sequence $\tau = (\tau_k)$ alone (see Chapter II, § 1f and, in greater detail, [250; Chapter II, § 3c]), which is in our case the sequence of the *instants of price jumps* $\tau_0 \equiv 0 < \tau_1 < \tau_2 < \cdots$.

Setting

$$N_t = \sum_{k \geqslant 1} I(\tau_k \leqslant t) \tag{8}$$

we obtain

$$\tau_k = \inf\{t \colon N_t = k\} \tag{9}$$

(here, as usual, $\tau_k = \infty$ if $\inf\{t \colon N_t = k\} = \varnothing$).

The process $N = (N_t)_{t \geqslant 0}$ is called a *counting* process, and formulas (8) and (9) establish clearly a one-to-one correspondence $N \Longleftrightarrow \tau$. As regards distributions, the *distribution* $\mathrm{Law}(\tau, \xi)$ is completely defined by the conditional distributions

$$\mathrm{Law}\left(\tau_{k+1} \;\middle|\; \begin{matrix} \tau_0, \ldots, \tau_k \\ \xi_0, \ldots, \xi_k \end{matrix}\right) \tag{10}$$

and

$$\mathrm{Law}\left(\xi_{k+1} \;\middle|\; \begin{matrix} \tau_0, \ldots, \tau_k, \tau_{k+1} \\ \xi_0, \ldots, \xi_k \end{matrix}\right), \tag{11}$$

where $\tau_0 = 0$ and $\xi_0 = S_0$.

3. So far, nothing has been said about τ_k and ξ_k as 'random' variables defined on some probability space. This question is worth a slightly more thorough discussion.

To be able to use the well-developed machinery of stochastic calculus in our analysis of stochastic processes (for instance, multivariate point processes that we consider now) and also to take account of the flow of information determining the prices, assume that from the very beginning we have some *filtered probability space* (stochastic basis)

$$(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geqslant 0}, \mathsf{P}).$$

Here $(\mathscr{F}_t)_{t\geqslant 0}$ is a flow (filtration) of σ-algebras $\mathscr{F}_t$ that are the 'carriers of market-related information available on the time intervals $[0,t]$'.

Once we have got this filtered probability space, it comes naturally to regard the τ_k as random variables ($\tau_k = \tau_k(\omega)$) that are *Markov times* with respect to $(\mathscr{F}_t)$:

$$\{\tau \leqslant t\} \in \mathscr{F}_t, \qquad t \geqslant 0. \tag{12}$$

In the same way we regard the ξ_k as $\mathscr{F}_{\tau_k}$-measurable random variables ($\xi_k = \xi_k(\omega)$), where $\mathscr{F}_{\tau_k}$ is the σ-algebra of events observable on the interval $[0,\tau_k]$, that is, of events $A \in \mathscr{F}$ such that

$$A \cap \{\tau_k \leqslant t\} \in \mathscr{F}_t \tag{13}$$

for each $t \geqslant 0$.

§ 1d. On the Statistics of 'Ticks'

1. We now discuss what is known about the *unconditional* probability distributions $\mathrm{Law}(\tau_1, \tau_2, \dots)$.

For $k \geqslant 1$ we set

$$\Delta_k = \tau_k - \tau_{k-1},$$

where $\tau_0 = 0$. Clearly, the knowledge of $\mathrm{Law}(\Delta_1, \Delta_2, \dots)$ is equivalent to the knowledge of $\mathrm{Law}(\tau_1, \tau_2, \dots)$, so that we can restrict ourselves to the analysis of the distribution of the time intervals between 'ticks'.

Starting from the *conjecture* that the variables $\Delta_1, \Delta_2, \dots$ are identically distributed and independent (this assumption enables one to justify, on the basis of the *Law of large numbers*, the standard statistical constructions of estimators for parameters, distributions, and so on) one can gain a clear impression of the character of their probability distribution from bar charts for the empirical density $\widehat{p}(\Delta)$ constructed on the basis of the available statistical data.

In [145] one can find the following results of such an analysis of 1 472 240 intervals between 'ticks' in the DEM/USD cross rate (the data supplied by Olsen & Associates [221]).

Up to a constant, we have

$$\widehat{p}(\Delta) \sim \begin{cases} \Delta^{-(1+\lambda_1)}, & 23\text{ s} \leqslant \Delta < 3\text{ m}, \\ \Delta^{-(1+\lambda_2)}, & 3\text{ m} \leqslant \Delta < 3\text{ h}, \end{cases} \tag{1}$$

where

$$\lambda_1 \approx 0.13 \quad \text{and} \quad \lambda_2 \approx 0.61. \tag{2}$$

(This is a result of the analysis of the logarithms $\ln \widehat{p}(\Delta)$ as functions of $\ln \Delta$; the estimates for λ_1 and λ_2 are obtained by the least squares method.)

It is worth recalling in connection with (1) that there is one distribution, well-known in the mathematical statistics, whose density decreases as a *power function*: the *Pareto distribution* with density

$$f_{\alpha b}(x) = \begin{cases} \dfrac{\alpha b^\alpha}{x^{\alpha+1}}, & x \geqslant b, \\ 0, & x < b \end{cases} \tag{3}$$

(here $\alpha > 0$, $b > 0$; see Table 2 in Chapter III, § 1a).

Note that quite often, in particular, in the finances literature, the authors mean by distributions of Pareto *type* (or simply *Pareto distributions*) ones whose density decreases as a power function at infinity.

Following this trend we can say that, as shows (1), we have a *Pareto distribution with exponent* $\alpha = \lambda_1$ on the interval [23 s, 3 m), and a *Pareto distribution with exponent* $\alpha = \lambda_2$ on the interval [3 m, 3 h).

It should be noted that, in the description of the probabilistic properties of a particular index, one can rarely expect (as it is often visible from the statistical analysis of the above kind) to make do with a single 'standard' distribution dependent on few unknown parameters. The reason probably lies with the fact that traders on the market have different goals, constraints, and react to risks in different ways.

2. In general, there are no a priori reasons to assume that the variables $\Delta_1, \Delta_2, \dots$ are *independent*. Moreover, as shown by empirical analysis, the time when the next 'tick' would occur essentially depends on the 'intensity', 'frequency' of 'ticks' in the past. The problem of an adequate description of the conditional distributions $\operatorname{Law}(\Delta_k \mid \Delta_1, \dots, \Delta_{k-1})$ is therefore timely.

In this connection we present, following [143], an interesting model serving for such a description. This model is *ARDM* (Autoregressive Conditional Duration Model), and it is related to the *ARCH* family.

Let $\psi_k = \psi_k(\Delta_1, \dots, \Delta_{k-1})$ be (sufficient) statistics such that

$$\mathsf{P}(\Delta_k \leqslant x \mid \Delta_1, \dots, \Delta_{k-1}) = \mathsf{P}(\Delta_k \leqslant x \mid \psi_k). \tag{4}$$

The simplest conditional distribution for Δ_k under condition ψ_k that comes to mind is the exponential distribution with density

$$p(\Delta \mid \psi_k) = \frac{1}{\psi_k} e^{-\frac{\Delta}{\psi_k}}, \qquad \Delta \geqslant 0, \tag{5}$$

where the (random) parameters (ψ_k) are defined recursively; namely,

$$\psi_k = \alpha_0 + \alpha_1 \Delta_{k-1} + \beta_1 \psi_{k-1} \tag{6}$$

with $\Delta_0 = 0$, $\psi_0 = 0$, $\alpha_0 > 0$, $\alpha_1 \geqslant 0$, and $\beta_1 \geqslant 0$.

Clearly, relations (4)–(6) define the conditional distributions

$$\mathrm{Law}(\Delta_k \mid \Delta_1, \dots, \Delta_{k-1})$$

completely and the resulting variables $\Delta_1, \Delta_2, \dots$ are, in general, *dependent*.

Note the following formula for the conditional expectations in this model:

$$\mathsf{E}(\Delta_k \mid \Delta_{k-1}, \dots, \Delta_1) = \psi_k. \tag{7}$$

This *first-order autoregressive model* (6) can obviously be generalized to higher orders (see Chapter II, § 1d and [143]).

2. Statistics of One-Dimensional Distributions

§ 2a. Discretizing Statistical Data

1. On gaining some ideas of the character of the 'intensity' (frequency) of 'ticks' (at any rate, in the example of currency cross rates) and of the one-dimensional distributions of the 'intertick' intervals $(\tau_k - \tau_{k-1})$, it now seems appropriate to consider the statistics of price 'changes', that is, of the sequence $(S_{\tau_k} - S_{\tau_{k-1}})_{k\geqslant 1}$ or of related variables, for instance, $h_{\tau_k} = \ln \dfrac{S_{\tau_k}}{S_{\tau_{k-1}}}$.

We point out straight away that 'daily', 'weekly', 'monthly', ... data are different from the 'interday' ones. The former can be regarded as data received at equal, 'regular', time intervals Δ (for instance, Δ can be one day or one week), while in the analysis of 'interday' statistics one must consider figures arriving in a '*nonregular way*', at random instants $\tau_1, \tau_2, \dots$ with different intervals $\Delta_1, \Delta_2, \dots$ between them (here $\Delta_k = \tau_k - \tau_{k-1}$).

This 'lack of regularity' brings forward certain difficulties in the application of the already developed methods of the statistical data analysis. For that reason, one usually carries out some preliminary data processing ('discretizing' the data, 'rejecting' abnormal observations, 'smoothing', separating trend components, and so on).

We now dwell on the methods of *discretization.*

We fix a 'reasonable' interval Δ of (real, physical) time. It should not be very small: in the case of the statistics of exchange rates we must ensure that this interval is *representative*, that is, it contains for sure a large number of 'ticks' (in other words, Δ must be considerably larger than the average time between two 'ticks'). Otherwise, there will be too many 'blank spaces' in our 'discretized data'.

For the 'exchange rates' of the 'basic' currencies it is recommended in [204] to take Δ not shorter than 10 minutes, which (besides the already mentioned 'representativeness') will also enable one to avoid uncertainties arising for small values of Δ, when the amplitude of spreads and the range of changes in bid and ask prices are *comparable*.

The *simplest* discretization method works as follows: having chosen Δ (say, 10 m, 20 m, 24 h, ...), we replace the piecewise constant continuous-time process $S = (S_t)_{t \geqslant 0}$, $t \geqslant 0$, by the sequence $S^\Delta = (S_{t_k})$ with discrete time $t_k = k\Delta$, $k = 0, 1, \dots$.

Using another widespread method one, first, replaces the piecewise constant process

$$S_t = S_0 + \sum_{k \geqslant 1} \xi_k I(\tau_k \leqslant t) \tag{1}$$

by its *continuous* modification $\widetilde{S} = (\widetilde{S}_t)$ obtained by linear interpolation between the values of (S_{τ_k}):

$$\widetilde{S}_t = S_{\tau_k} \frac{\tau_{k+1} - t}{\tau_{k+1} - \tau_k} + S_{\tau_{k+1}} \frac{t - \tau_k}{\tau_{k+1} - \tau_k}, \qquad \tau_k < t \leqslant \tau_{k+1}. \tag{2}$$

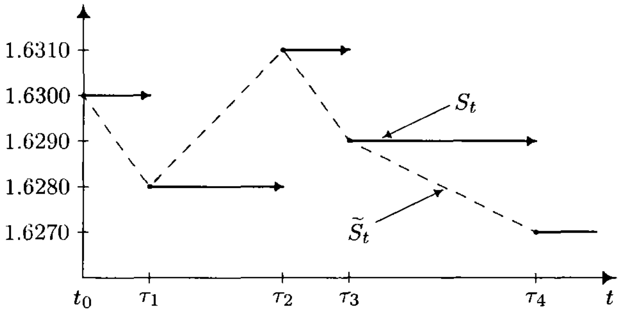


FIGURE 32. Piecewise-constant process (S_t) and its continuous modification $(\widetilde{S}_t)$

After that, we discretize this modification $\widetilde{S} = (\widetilde{S}_t)$ using the simplest method; that is, we construct the sequence $\widetilde{S}^\Delta = (\widetilde{S}_{t_k})$, where $t_k = k\Delta$, $k = 0, 1, \dots$, and Δ is the time interval that is of importance to a particular investor or trader (one month, one day, 20 minutes, 5 minutes, ...).

2. Beside discretization with respect to time, statistical data can also be *quantified, rounded off* with respect to the phase variable. Usually, this is carried out as follows.

We choose some $\gamma > 0$ and, instead of the original process $S = (S_t)_{t \geqslant 0}$, introduce a new one, $S(\gamma) = (S_t(\gamma))_{t \geqslant 0}$, with variables

$$S_t(\gamma) = \gamma \left[\frac{S_t}{\gamma} \right]. \tag{3}$$

For example, if $\gamma = 1$ and $S_t = 10.54$, then $S_t(1) = 10$; while if $\gamma = 3$, then $S_t(3) = 9$. Hence it is clear that (3) corresponds to rounding off with error not larger than γ.

If γ-*quantization* is carried out first, and Δ-*discretization* next, then we obtain from (S_t) another sequence, $S^{\Delta}(\gamma)$ or $\widetilde{S}^{\Delta}(\gamma)$.

Since $S_t(\gamma) \to S_t$ as $\gamma \to 0$, there arises the question how one can *consistently* choose the values of Δ and γ to achieve that the amount of information contained in the variables $S_{t_k}(\gamma)$ with $t_k = k\Delta$, $k = 0, 1, \dots$, is 'almost the same as the information in (S_t)'. As an initial approach it seems reasonable (following a suggestion of J. Jacod) to find out conditions on the convergence rates $\Delta \to 0$ and $\gamma \to 0$ ensuring the convergence of the finite-dimensional distributions of the processes $S^{\Delta}(\gamma)$ and $\widetilde{S}^{\Delta}(\gamma)$ to the corresponding distributions for S.

§ 2b. One-Dimensional Distributions of the Logarithms of Relative Price Changes. Deviation from the Gaussian Property and Leptokurtosis of Empirical Densities

1. We now consider the process of some cross rate (DEM/USD, say). Let

$$S = (S_t)_{t \geqslant 0},$$

and let $\widetilde{S} = (\widetilde{S}_t)_{t \geqslant 0}$ be the continuous modification of $S = (S_t)_{t \geqslant 0}$ obtained by linear interpolation.

Further, let $\widetilde{S}^{\Delta} = (\widetilde{S}_{t_k})_{k \geqslant 0}$, where $t_k = k\Delta$ be the 'Δ-discretization' of $\widetilde{S} = (\widetilde{S}_t)_{t \geqslant 0}$.

We have repeatedly mentioned (see, for instance, Chapter I, § 2a.4) that, in the analysis of price changes, it is not their amplitudes $\Delta\widetilde{S}_{t_k} \equiv \widetilde{S}_{t_k} - \widetilde{S}_{t_{k-1}}$ *themselves* that are of real economic significance, but, rather, the *relative* changes $\dfrac{\Delta\widetilde{S}_{t_k}}{\widetilde{S}_{t_{k-1}}} = \dfrac{\widetilde{S}_{t_k}}{\widetilde{S}_{t_{k-1}}} - 1$. It is understandable for that reason that one is usually not so much interested in the distribution of $\widetilde{S}_{t_k}$ as in the distribution of $\widetilde{H}_{t_k} = \ln\left(\dfrac{\widetilde{S}_{t_k}}{\widetilde{S}_0}\right)$.

We now set

$$\widetilde{h}_{t_k}^{(\Delta)} = \Delta\widetilde{H}_{t_k} \quad (= \widetilde{H}_{t_k} - \widetilde{H}_{t_{k-1}}), \tag{1}$$

where $t_k = k\Delta$, $k \geqslant 0$, and $\widetilde{H}_0 = 0$.

Bearing in mind our construction of $\widetilde{S}^{\Delta}$ from S and the notation $\left(h_t = \Delta H_t \right.$ and $\left. H_t = \ln \dfrac{S_t}{S_0}\right)$ introduced in § 2a, we obtain

$$\widetilde{h}_{t_k}^{(\Delta)} = \sum_{\{i:\, t_{k-1} < \tau_i \leqslant t_k\}} h_{\tau_i} + \widetilde{r}_{t_k}^{(\Delta)}, \tag{2}$$

where the term $\widetilde{r}_{t_k}^{(\Delta)}$ reflects effects related to the end-points of the partitioning intervals and can be called the 'remainder' because it is small compared with the sum.

Indeed, if, e.g., Δ is 1 hour, then (see the table in § 1b.4) the average number of 'ticks' of the DEM/USD exchange rate is approximately 187 (= 4500 : 24). Hence the value of the sum in (2) is determined by the 187 values of the h_{τ_i}. At the same time, $|\widetilde{r}_{t_k}^{(\Delta)}|$ is knowingly not larger than the sum of the absolute values of the four increments h_τ corresponding to the ticks occurring immediately before and after the instants t_{k-1} and t_k.

It must be pointed out that the sum in (2) is the sum of a *random number of random variables*, and it may have a fairly complicated distribution even if the distributions of its terms and of the random number of terms are relatively simple. This is a sort of a technical explanation why, as we show below, we cannot assume that the variables $\widetilde{h}_{t_k}^{(\Delta)}$ have a Gaussian distribution. True, we shall also see that the conjecture of the Gaussian distribution of $\widetilde{h}_{t_k}^{(\Delta)}$ becomes ever more likely with the decrease of Δ, which leads to the increase in the number of summands in (2). This is where the influence of the *Central limit theorem* about sums of large numbers of variables becomes apparent.

Remark. The notation $\widetilde{h}_{t_k}^{(\Delta)}$ is fairly unwieldy, although it is indicative of the construction of these variables. In what follows, we shall also denote them by $\widetilde{h}_k$ for simplicity (describing their construction explicitly and indicating the chosen value of Δ).

2. Thus, assume that we are given some value of $\Delta > 0$. In analyzing the joint distributions $\mathrm{Law}(\widetilde{h}_1, \widetilde{h}_2, \dots)$ of the sequence of 'discretized' variables $\widetilde{h}_1, \widetilde{h}_2, \dots$ it is reasonable to start with the *one-dimensional* distributions, making the assumption that these variables are *identically* distributed. (The suitability of this homogeneity conjecture as a first approximation is fully supported by the statistical analysis of many financial indexes; at any rate, for not excessively big time intervals. See also § 3c below for a description of another construction of the variables $\widetilde{h}_i$ that takes into account the geography-related peculiarities of the interday cycle.)

As already mentioned, there occurred 8.238.532 'ticks' in the DEM/USD exchange rate from January 01, 1987 through December 31, 1993 according to Olsen & Associates. One can get an idea of the estimates obtained for several *characteristics* of the one-dimensional distribution of the variables $\widetilde{h}_i = \widetilde{h}_{t_i}^{(\Delta)}$ on the basis of these statistical data from the following table borrowed from [204]:

Δ	N	Mean $\overline{h}_N$	Variance $\widehat{m}_2$	Skewness $\widehat{S}_N$	Kurtosis $\widehat{K}_N$
10 m	368.000	$-2.73 \cdot 10^{-7}$	$2.62{\cdot}10^{-7}$	0.17	35.11
1 h	61.200	$-1.63 \cdot 10^{-6}$	$1.45{\cdot}10^{-6}$	0.26	23.55
6 h	10.200	$-9.84 \cdot 10^{-6}$	$9.20{\cdot}10^{-6}$	0.24	9.44
24 h	2.100	$-4.00 \cdot 10^{-5}$	$3.81{\cdot}10^{-5}$	0.08	3.33

In this table, N is the total number of the discretization nodes $t_1, t_2, \dots, t_N$, where $t_k = k\Delta$,

$$\overline{h}_N = \frac{1}{N}\sum_{i=1}^{N} \widetilde{h}_{t_i}, \qquad \widehat{m}_k = \frac{1}{N}\sum_{i=1}^{N} \bigl(\widetilde{h}_{t_i} - \overline{h}_N\bigr)^k,$$

while

$$\widehat{S}_N = \frac{\widehat{m}_3}{\widehat{m}_2^{3/2}}$$

is the empirical skewness, and

$$\widehat{K}_N = \frac{\widehat{m}_4}{(\widehat{m}_2)^2} - 3$$

is the empirical kurtosis.

The theoretical skewness S_N of the normal distribution is zero. The fact that the empirical coefficient $\widehat{S}_N$ is positive means that the empirical (and maybe, also the actual) distribution density is *asymmetric*, the left-hand side of its graph is steeper than its right-hand side.

It is clear from the table that the modulus of the mean value is considerably less than the standard deviation, so that we can set it to be equal to zero in practice.

The most serious argument *against* the conjecture of normality is, of course, the *excessively large* kurtosis, growing, as we see, with the decrease of Δ. Since this coefficient is defined in terms of the fourth moment, this also suggests that the distribution of the $\widetilde{h}_k = \widetilde{h}_{t_k}^{(\Delta)}$ must have 'heavy tails', i.e., the corresponding density $p^{(\Delta)}(x)$ decreases relatively slowly as $|x| \to \infty$ (as compared with the normal density).

3. It is not the shape of the bar-charts of statistical densities alone that speaks for the deviation of the variables $\widetilde{h}_k$ from the normal (Gaussian) property (which is observable *not only* in the case of exchange rates, but also for other financial indexes, for instance, stock prices). It can be discovered also by standard statistical tests checking for deviations from normality, such as, for example,

(1) the *quantile method*,
(2) the χ^2-*test*,
(3) the *rank tests*.

We recall the essential features of these methods.

The quantile method can be most readily illustrated by the $Q\widehat{Q}$-plot (see Fig. 33), where the quantiles of the corresponding normal distribution $\mathcal{N}(\mu, \sigma^2)$ with parameters μ and σ^2 estimated on the basis of statistical data are plotted along the horizontal axis and the quantiles of the empirical distribution for the $\widetilde{h}_k$ are plotted along the vertical axis. (The *quantile* Q_p of order p, $0 < p < 1$, of the distribution

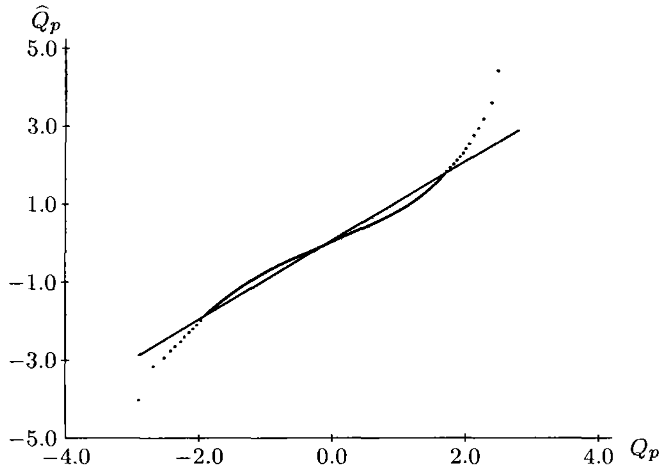


FIGURE 33. $Q\widehat{Q}$-quantile analysis of the DEM/USD exchange rate with $\Delta = 20$ minutes (according to Reuters; from October 5, 1992 through September 6, 1993; [427]). The quantiles $\widehat{Q}_p$ of the empirical distribution for the variables $\widetilde{h}_k = \widetilde{h}_{t_k}^{(\Delta)}$, $t_k = k\Delta$, $k = 1, 2, \dots$, are plotted along the vertical axis, and the quantiles Q_p of the normal distribution are plotted along the horisontal axis

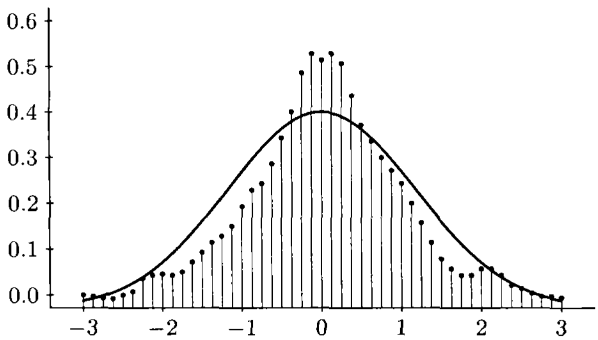


FIGURE 34. A typical graph of the empirical density (for the variables $\widetilde{h}_k = \widetilde{h}_{t_k}^{(\Delta)}$, $k = 1, 2, \dots$) and the graph of the corresponding theoretical (normal) density

of a random variable ξ is, by definition, the value of x such that $\mathsf{P}(\xi \leqslant x) \geqslant p$ and $\mathsf{P}(\xi \geqslant x) \geqslant 1 - p$.)

In the case of a good agreement between the empirical and the theoretical distributions the set $(Q_p, \widehat{Q}_p)$ must be clustered close to the bisector. However, this is not the case for the statistical data under consideration (exchange rates, stock prices, and so on). The $(Q_p, \widehat{Q}_p)$-graph in Fig. 33 describes the relations between the (normal) theoretical density and the empirical density, which we depict in Fig. 34.

4. The use of K. Pearson's χ^2-test in the role of a test of goodness-of-fit is based on the statistics

$$\widehat{\chi}^2 = \sum_{i=1}^{k} \frac{(\nu_i - np_i)^2}{np_i},$$

where the ν_i are the numbers of the observations falling within certain intervals I_i $\left(\sum_{i=1}^{k} I_i = \mathbb{R}\right)$ corresponding to a particular 'grouping' of the data, and the p_i are the probabilities of hitting these intervals calculated for the theoretical distribution under test.

In accordance with Pearson's criterion, the conjecture

$\mathscr{H}_0$: *the empirical data agree with the theoretical model*

is rejected with significance level α if $\widehat{\chi}^2 > \chi^2_{k-1,1-\alpha}$, where $\chi^2_{k-1,1-\alpha}$ is the $Q_{1-\alpha}$-quantile (that is, the quantile of order $1-\alpha$) of the χ^2-distribution with $k-1$ degrees of freedom. Recall that the χ^2-distribution with n degrees of freedom is the distribution of the random variable

$$\chi_n^2 = \xi_1^2 + \cdots + \xi_n^2,$$

where $\xi_1, \ldots, \xi_n$ are independent standard normally distributed ($\mathscr{N}(0,1)$) random variables. The density $f_n(x)$ of this distribution is

$$f_n(x) = \begin{cases} \dfrac{1}{2^{n/2}\Gamma(n/2)} x^{n/2-1} e^{-x/2}, & x > 0, \\ 0, & x \leqslant 0. \end{cases} \tag{3}$$

In [127] one can find the results of statistical testing for conjecture $\mathscr{H}_0$ with significance level $\alpha = 1\%$ in the case of the stock of ten major German companies and banks (BASF, BMW, Daimler Benz, Deutsche Bank, Dresdner Bank, Höchst, Preussag, Siemens, Thyssen, VW). This involves the data over a period of three years (October 2, 1989–September 30, 1992). The calculation of $\widehat{\chi}^2$ and $\chi^2_{k-1,1-\alpha}$ (with $k = 22$, on the basis of $n = 745$ observations) shows that conjecture $\mathscr{H}_0$ must be *rejected* for *all* these ten companies. For instance, the values of $\widehat{\chi}^2$ (with $p_i = 1/k$, $k = 22$) for BASF and Deutsche Bank are equal to 104.02 and 88.02, respectively, while the critical value of $\chi^2_{k-1,1-\alpha}$ for $k = 22$ and $\alpha = 0.01$ is 38.93. Hence $\widehat{\chi}^2$ is considerably larger than $\chi^2_{k-1,1-\alpha}$.

§ 2c. One-Dimensional Distributions of the Logarithms of Relative Price Changes. 'Heavy Tails' and Their Statistics

1. In view of deviations from the normal property and the 'heavy tails' of empirical densities the researchers have come to a common opinion that, for instance, for the 'right-hand tails' (that is, as $x \to +\infty$) we must have

$$\mathsf{P}(\widetilde{h}_{t_k}^{(\Delta)} > x) \sim x^{-\alpha} L(x), \tag{1}$$

where the 'tail index' $\alpha = \alpha(\Delta)$ is positive and $L = L(x)$ is a slowly varying function $\left(\dfrac{L(xy)}{L(x)} \to 1 \text{ as } x \to \infty \text{ for each } y > 0\right)$. A similar conclusion holds for 'left-hand tails'.

We note that discussions of '*heavy tails*' and *kurtosis* in the finances literature can be found in [46], [361], [419], and some other papers, and even in several publications dating back to the 1960s (see, e.g., [150], [317]).

There, it was pointed out that the kurtosis and the 'heavy tails' of a distribution density could occur, for instance, in the case of *mixtures* of normal distributions. (On this subject, see Chapter III, § 1d, where we explain how one can obtain, for example, hyperbolic distributions by 'mixing' normal distributions with different variances.)

In several papers (see, e.g., [46] and [390]), in connection with the search of suitable distributions for the $\widetilde{h}_{t_k}^{(\Delta)}$ the authors propose to use Student's t-distribution with density

$$f_n(x) = \frac{1}{\sqrt{\pi n}} \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}, \tag{2}$$

where n is an integer parameter, the 'number of the degrees of freedom'. Clearly, this is a distribution of Pareto type with 'heavy tails'.

2. After B. Mandelbrot (see, for instance, [318]–[324]) and E. Fama ([150]), it became a fashion in the finances literature to consider models of financial indexes based on *stable* distributions (see Chapter III, § 1a for the detail). Such a distribution has a *stability exponent* α in the interval $(0,2]$. If $\alpha = 2$, then the stable distribution is normal; for $0 < \alpha < 2$ this is a Pareto-type distribution satisfying (1), and the 'tail index' α is just the stability exponent.

This explains why the conjecture of a stable distribution with $0 < \alpha < 2$ arises naturally in our search for the distributions of the $\widetilde{h}_k = \widetilde{h}_{t_k}^{(\Delta)}$: such a distribution is marked by both '*heavy tails*' and strong *kurtosis*, which are noticeable in the statistical data. Another argument in favor of these distributions is the characteristic property of self-similarity (see Chapter III, § 2b): if X and Y are independent random variables having a stable distribution with *stability exponent* α, then their sum also has a stable distribution with the same exponent (equivalently, the *convolution* of the distributions of these variables is a distribution of the same type).

From the economic standpoint, this is a perfectly natural property of the *preservation* of the type of distribution under *time aggregation*. The fact that stable distributions have this property is an additional evidence in favor of their use.

However, operating stable distributions involves considerable difficulties, brought on by the following factors.

If X is a random variable with stable distribution of index α, $0 < \alpha < 2$, then $\mathsf{E}|X| < \infty$ only for $\alpha > 1$. More generally, $\mathsf{E}|X|^p < \infty$ if and only if $p < \alpha$.

Hence the tails of a stable distribution with $0 < \alpha < 2$ are so 'heavy' that the second moment is infinite. This leads to considerable theoretical complications (for instance, in the analysis of the quality of various estimators or tests based on variance); on the other hand this is difficult to explain from the economic standpoint or to substantiate by facts since one usually has only limited stock of suitable statistical data.

We must point out in this connection that the estimate of the actual value of the 'tail index' α is, in general, a fairly delicate task.

On the one hand, to get a good estimate of α one must have access to the results of sufficiently *many* observations, in order to collect a stock of 'extremal' values, which alone are suitable for the assessment of 'tail effects' and the 'tail index'. Unfortunately, on the other hand, the *large number* of 'nonextremal' observations inevitably contributes to *biased* estimates of the actual value of α.

3. As seen from the properties of stable distributions, using them for the description of the distributions of financial indexes we cannot possibly satisfy the following three conditions: *stability of the type of distribution under convolution, heavy tails with index* $0 < \alpha < 2$, and *the finiteness of the second moment* (*and, therefore, of the variance*).

Clearly, Pareto-type distributions with 'tail index' $\alpha > 3$ have *finite variance*. Although they are not preserved by convolutions, they nevertheless have an important property of the stability of the *order of decrease* of the distribution density under convolution.

More precisely, this means the following. If X and Y have the same Pareto-type distribution with 'tail index' α and are independent, then their sum $X + Y$ also have a Pareto-type distribution with the same 'tail index' α. In this sense, we can say that Pareto-type distributions satisfy the required property of the stability of the 'tail index' α under convolutions.

It is already clear from the above why we pay so much attention to this index α, which determines the behavior of the distributions of the variables $\widehat{h}_{t_k}^{(\Delta)}$ at infinity. We can give also an 'econo-financial' explanation to this interest towards α. The 'tail index' indicates, in particular, the role of 'speculators' on the market. If α is large, then extraordinary oscillations of prices are rare and the market is 'well-behaved'. Accepting this interpretation, a market characterized by a large value of α can be considered as 'efficient', so that the value of α is a measure of this 'efficiency'. (For a discussion devoted to this subject see, for example, [204]).

4. We now consider the estimates for the 'stability index' of stable distributions and, in general, for the 'tail indices' of Pareto-type distributions.

It should be noted from the onset that authors are not unanimous on the issue of the 'real' value of the 'tail index' α for particular exchange rates, stock prices, or other indexes. The reason, as already mentioned, lies in the complexity of the task of the construction of efficient estimators $\widehat{\alpha}_N$ for α (here N is the number of observations). Stating the problem of estimation for this parameter requires on its own right an accurate specification of all details of cooking the statistical 'stock' $\{\widehat{h}_{t_k}^{(\Delta)}\}$, a careful choice of the value of Δ, and so on.

In the finances literature, one often uses the following 'efficient' estimators $\widehat{\alpha}_N$ for α, which were suggested in [152] and [153]:

$$\widehat{\alpha}_N = 0.827\,\frac{\widehat{Q}_f - \widehat{Q}_{1-f}}{\widehat{Q}_{0.72} - \widehat{Q}_{0.28}}\,, \qquad 0.95 \leqslant f \leqslant 0.97, \tag{3}$$

where $\widehat{Q}_f$ is the quantile of order f constructed for a sample of size N under the assumption that the observable variables have a symmetric stable distribution.

If it were true that the distribution $\mathrm{Law}(\widetilde{h}_k)$ with $\widetilde{h}_k = \widetilde{h}_{t_k}^{(\Delta)}$ belongs to the class of stable distributions (with stability exponent α), then one could anticipate that the $\widehat{\alpha}_N$ stabilize (and converge to some value $\alpha < 2$) with the increase of the sample size N.

However, there is no unanimity on this point, as already mentioned. Some authors say that their estimators for certain financial indexes duly stabilize (see, for instance, [88] and [474]). On the other hand, one can find in many papers results of statistical analysis showing that the $\widehat{\alpha}_N$ do not merely show a tendency to growth, but approach values that are greater of equal to 2 (see, e.g., [27] and [207] as regards shares in the American market, or [127] as regards the stock of major German companies and banks). All this calls for a cautious approach towards the 'stability' conjecture, though, of course, there is no contradiction with the conjecture that 'tails' can be described in terms of Pareto-type distributions.

5. Now, following [204], we present certain results concerning the values of the 'tail index' α for the currency cross rates *under the assumption* that the $\widetilde{h}_k = \widetilde{h}_{t_k}^{(\Delta)}$ have a Pareto-type distribution satisfying (1).

The following values of the 'tail index' $\alpha = \alpha(\Delta)$ were obtained in [204] (see the table on the next page).

We now make several comments as regards this table.

In subsection 6 below we describe estimates of α based on the data of Olsen & Associates (cf. § 1b). For Δ equal to 6 hours the estimation accuracy is low, due to the insufficient amount of observations.

rate \ Δ	10 m	30 m	1 h	6 h
DEM/USD	3.11 ± 0.33	3.35 ± 0.29	3.50 ± 0.57	4.48 ± 1.64
JPY/USD	3.53 ± 0.21	3.55 ± 0.47	3.62 ± 0.46	3.86 ± 1.81
GBP/USD	3.44 ± 0.22	3.52 ± 0.46	4.01 ± 1.09	6.93 ± 10.79
CHF/USD	3.64 ± 0.41	3.74 ± 0.82	3.84 ± 0.77	4.39 ± 4.64
FRF/USD	3.34 ± 0.22	3.29 ± 0.47	3.40 ± 0.69	4.61 ± 1.21
FRF/DEM	3.11 ± 0.41	2.55 ± 0.23	2.43 ± 0.23	3.54 ± 1.42
NLG/DEM	3.05 ± 0.27	2.44 ± 0.08	2.19 ± 0.12	3.37 ± 1.43
ITL/DEM	3.31 ± 0.51	2.93 ± 1.17	2.54 ± 0.49	2.86 ± 0.98
GBP/DEM	3.68 ± 0.35	3.63 ± 0.42	4.18 ± 1.67	3.22 ± 0.79
JPY/DEM	3.69 ± 0.41	4.18 ± 0.90	4.13 ± 1.05	4.71 ± 1.61

Analyzing the figures in the table (which rest upon a *large database* and must be reliable for this reason), we can arrive at an important conclusion that the exchange rates of the basic currencies against the US dollar in the FX-market have (for $\Delta = 10$ minutes) a Pareto-type distribution with 'tail index' $\alpha \approx 3.5$, which increases with the increase of Δ. Hence it is now very likely that the variance of $\widetilde{h}_k = \widetilde{h}_{t_k}^{(\Delta)}$ must be *finite* (a desirable property!), although we cannot say the same about the *fourth* moment responsible for the leptokurtosis of distributions.

In another paper [91] by Olsen & Associates one can also find data on the rates XAU/USD and XAG/USD (XAU is gold and XAG is silver). For $\Delta = 10$ minutes the corresponding estimates for α are 4.32 ± 0.56 and 4.04 ± 1.71, respectively; for $\Delta = 30$ minutes these estimates are 3.88 ± 1.04 and 3.92 ± 0.73.

6. In this subsection we merely outline the construction of estimators $\widehat{\alpha}$ for the 'tail index' α used for the above table borrowed from [204] (we do not dwell on the 'bootstrap' and 'jackknife' methods significant for the evaluation of the bias and the standard deviation of these estimators).

We consider the *Pareto distribution* with density

$$f_{\alpha b}(x) = \frac{\alpha b^\alpha}{x^{\alpha+1}}, \qquad x \geqslant b, \tag{4}$$

where $f_{\alpha b}(x) = 0$ for $x < b$.

For $x \geqslant b$ we have

$$\ln f_{\alpha b}(x) = \ln \alpha + \alpha \ln b - (\alpha + 1) \ln x. \tag{5}$$

Hence the maximum likelihood estimator $\widehat{\alpha}_N$ constructed on the basis of N independent observations $(X_1, \dots, X_N)$ so as to satisfy the condition

$$\max_{\alpha} \prod_{k=1}^{N} f_{\alpha b}(X_k) = \prod_{k=1}^{N} f_{\widehat{\alpha}_N b}(X_k), \tag{6}$$

can be defined by the formula

$$\frac{1}{\widehat{\alpha}_N} = \frac{1}{N} \sum_{i=1}^{N} \ln\left(\frac{X_i}{b}\right). \tag{7}$$

Since

$$\alpha \int_{\beta}^{\infty} (\ln x) x^{-(\alpha+1)}\, dx = \beta^{\alpha}\left(\frac{1}{\alpha} + \ln \beta\right)$$

for $\alpha > 0$ and $\beta > 0$, it follows that

$$\mathsf{E} \ln \frac{X_i}{b} = \alpha b^{\alpha} \int_{b}^{\infty} \left(\ln \frac{x}{b}\right) x^{-(\alpha+1)}\, dx = \frac{1}{\alpha}.$$

Hence

$$\mathsf{E}\left(\frac{1}{\widehat{\alpha}_N}\right) = \frac{1}{\alpha},$$

that is, $1/\widehat{\alpha}_N$ is an unbiased estimator for $1/\alpha$, so that the estimator $\widehat{\alpha}_N$ for α has fairly good properties (of course, *provided* that the actual distribution is—*exactly*—a Pareto distribution with known 'starting point' b, rather than a Pareto-*type* distribution that has no well-defined 'starting point').

There arises a natural idea (see [223]) to use (7) nevertheless for the estimate of α in Pareto-*type* distributions, replacing the unknown 'starting point' b by some its suitable estimate.

For instance, we can proceed as follows. We choose a sufficiently large number M (although it should not be too large in comparison with N) and construct an estimator for α using a modification of (7) with b replaced by M and with the sum taken over $i \leqslant N$ such that $X_i \geqslant M$.

To this end we set

$$\widehat{\gamma}_{N,M} = \frac{\sum_{\{i \leqslant N:\, X_i \geqslant M\}} \ln \frac{X_i}{M}}{\sum_{\{i \leqslant N:\, X_i \geqslant M\}} I_M(X_i)}, \tag{8}$$

where

$$I_M(x) = \begin{cases} 1 & \text{if } x \geqslant M, \\ 0 & \text{if } x < M. \end{cases}$$

Setting

$$\nu_{N,M} = \sum_{\{i\leqslant N:\ X_i\geqslant M\}} I_M(X_i),$$

we can rewrite the formula for $\widehat{\gamma}_{N,M}$ as follows:

$$\widehat{\gamma}_{N,M} = \frac{1}{\nu_{N,M}} \sum_{\{i\leqslant N:\ X_i\geqslant M\}} \ln \frac{X_i}{M}, \tag{9}$$

and we can consider an estimator $\widehat{\alpha}_{N,M}$ for α such that

$$\frac{1}{\widehat{\alpha}_{N,M}} = \widehat{\gamma}_{N,M}. \tag{10}$$

We could also proceed otherwise. Namely, we order the sample $(X_1, X_2, \dots, X_N)$ to obtain another sample, $(X_1^*, X_2^*, \dots, X_N^*)$, with $X_1^* \geqslant X_2^* \geqslant \cdots \geqslant X_N^*$. We fix again some integer $M \ll N$ as the 'starting point' b and set

$$\gamma_{N,M}^* = \frac{1}{M} \sum_{1\leqslant i\leqslant M} \ln \frac{X_i^*}{X_M^*}. \tag{11}$$

Then we can consider the estimator $\alpha_{N,M}^*$ for α such that

$$\frac{1}{\alpha_{N,M}^*} = \gamma_{N,M}^*. \tag{12}$$

The estimator $\alpha_{N,M}^*$ so obtained was first proposed by B. M. Hill [223], and it is usually called the *Hill estimator*.

Obviously, the 'good' properties of this estimator depend on a *right choice* of M, the number of maximal order statistics generating the statistics $\alpha_{N,M}^*$. However, it is also clear that one can hardly expect to make a 'universal' choice of M, suitable for a wide class of slowly changing functions $L = L(x)$ describing the behavior of, say, the right-hand 'tail' in accordance with the formula

$$\mathsf{P}(X_i > x) \sim x^{-\alpha} L(x). \tag{13}$$

Usually one studies the properties of the above estimators $\alpha_{N,M}^*$ and $\widehat{\alpha}_{N,M}$ for a particular subclass of functions $L = L(x)$. For instance, we can assume that $L = L(x)$ belongs to the subclass

$$L_\gamma = \bigl\{L = L(x)\colon L(x) = 1 + cx^{-\gamma} + o(x^{-\gamma}),\ c > 0\bigr\}$$

with $\gamma > 0$. It is shown in [223] under this assumption that if $M \to \infty$ as $N \to \infty$ so that

$$\frac{M}{N^{\frac{2\gamma}{2\gamma+\alpha}}} \to 0,$$

then

$$\operatorname{Law}\bigl(\sqrt{M}\,(\alpha_{N,M}^* - \alpha)\bigr) \to \mathscr{N}(0, \alpha^2),$$

i.e., the estimators $(\alpha_{N,M}^*)$ are asymptotically normal.

§ 2d. One-Dimensional Distributions of the Logarithms of Relative Price Changes. Structure of the 'Central' Parts of Distributions

1. And yet, how can one, bearing in mind the properties of the distribution $\mathrm{Law}(\widetilde{h}_k)$, combine large kurtosis and 'tail index' $\alpha > 2$ (as in the case of currency exchange rates)?

Apparently, one can hardly expect to carry this out by means of a *single 'standard'* distribution. Taking into account the fact that the market is full of traders and investors with various interests and time horizons, a more attractive option is the one of invoking *several* 'standard' distributions, each valid in a particular domain of the range of the variables $\widetilde{h}_k$.

Many authors and, first of all, Mandelbrot are insistent in their advertising the use of stable distributions (and some their modifications) as extremely suitable for the ' central' part of this range. (See, for instance, the monograph [352] containing plenty of statistical data, the theory of stable distributions and their generalizations, and results of statistical analysis.)

In what follows we shall discuss the results of [330] relating to the use of stable laws in the description of the S&P500 Index in the corresponding 'central' zone. (To describe the 'tails' the authors of [330] propose to use the normal distribution; the underlying idea is that the shortage of suitable statistical data rules out reliable conclusions about the behavior of the 'tails'. See also [464].)

As regards other financial indexes we refer to [127], where one can find a minute statistical analysis of the financial characteristics of ten major German corporations and the conclusion that the *hyperbolic* distribution is extremely well suited for the 'central' zone.

In Chapter III, § 1d we gave a detailed description of the class of hyperbolic distributions, which, together with the class of stable distributions, provides one with a rather rich armory of theoretical distributions. As both *hyperbolic* and *stable* distributions can be described by *four* parameters, the hopes of reaching a good agreement between 'theory' and 'experiment' using their combinations seem well based.

2. We consider now the results of the statistical analysis of the data relating to the S&P500 Index that was carried out in [330].

The authors considered the six-year evolution of this index on NYSE (New York Stock Exchange) (January 1984 through December 1989). All in all, 1 447 514 ticks were registered (the data of the Chicago Mercantile Exchange). On the average, ticks occurred with one-minute intervals during 1984–85 and with 15-second intervals during 1986–87.

Since the exchange operates only at opening hours, the construction of the process describing the evolution of the S&P500 Index took into account only the 'trading time' t, and the closing prices of each day were adjusted to match the 'opening prices' of the next day.

Let $S = (S_t)$ be the resulting process. We shall now consider the changes of this process over some intervals of time Δ:

$$\Delta S_{t_k} = S_{t_k} - S_{t_{k-1}}, \tag{1}$$

where $t_k = k\Delta$. The interval Δ will vary from 1 minute to 10^3 minutes (In [330], Δ assumes the following values: $1, 3, 10, 32, 100, 316$, and 1000 minutes; the number of ticks corresponding to $\Delta = 1$ m is 493 545 and it is 562 for $\Delta = 1000$ m.)

Using the notation of § 2a, one can observe that

$$S_{t_k} \approx S_{t_{k-1}} e^{\widetilde{h}_{t_k}^{(\Delta)}} \approx S_{t_{k-1}} \bigl(1 + \widetilde{h}_{t_k}^{(\Delta)}\bigr)$$

and therefore, $\Delta S_{t_k} \approx S_{t_{k-1}} \widetilde{h}_{t_k}^{(\Delta)}$, since $\widetilde{S}_{t_k} \approx S_{t_k}$ and the increments $\Delta S_{t_k} \equiv S_{t_k} - S_{t_{k-1}}$ are small.

This (approximate) equality shows that for independent increments $(S_{t_k} - S_{t_{k-1}})$ the distribution $\mathsf{P}\bigl(\widetilde{h}_{t_k}^{(\Delta)} \leqslant x\bigr)$ and the conditional distribution $\mathsf{P}(\Delta S_{t_k} \leqslant x \mid S_{t_{k-1}} = y)$ have roughly the same behavior.

Considering the empirical densities $\widehat{p}^{(\Delta)}(x)$ of the variables ΔS_{t_k}, $t_k = k\Delta$, which are assumed to be identically distributed, the authors of [330] plot the graphs of the $\log_{10} \widehat{p}^{(\Delta)}(x)$. Schematically, they are as in Fig. 35.

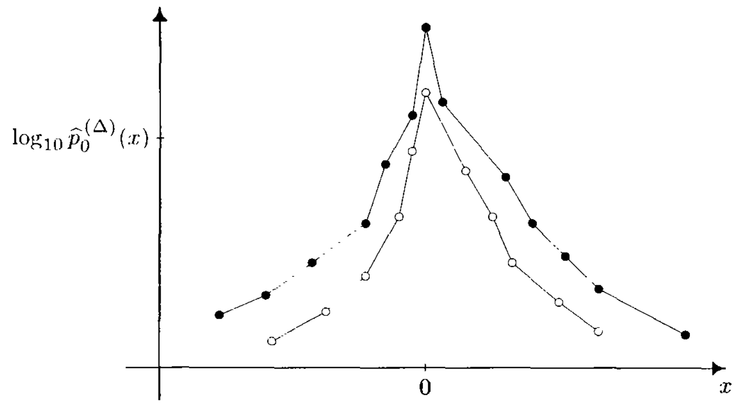


FIGURE 35. Sketches of the graphs of $\log_{10} \widehat{p}^{(\Delta)}(x)$ for two distinct values of Δ

A mere visual inspection of the many graphs of $\log_{10} \widehat{p}_{t_k}^{(\Delta)}(x)$ for various values of Δ shows that the distribution densities are fairly symmetric and are 'melting' as Δ *grows*. They decrease as $x \to \pm\infty$, but not as fast as it should be for a Gaussian distribution.

The unimodality and the symmetry that are visible here, together with the character of the decrease at infinity of the empirical densities indicate that it might be reasonable to direct attention to *stable symmetric* distributions. We recall that the characteristic function $\varphi(\theta) = \mathsf{E}e^{i\theta X}$ of a stable random variable X with symmetric distribution is as follows (see formula (14) in Chapter III, § 1a):

$$\varphi(\theta) = e^{-\sigma^\alpha|\theta|^\alpha}, \tag{2}$$

where $\sigma \geqslant 0$ and $0 < \alpha \leqslant 2$. Hence, once one have accepted the 'stability' conjecture one should first of all find an estimate for the parameter α.

Stable distributions have the Pareto type. In the symmetric case (see (7) and (8) in Chapter III, § 1a), if $0 < \alpha < 2$, then we have

$$\mathsf{P}(|X| > x) \sim \widetilde{c}_\alpha x^{-\alpha} \quad \text{as} \quad x \to \infty$$

with some constant $\widetilde{c}_\alpha$, and we could find the value of α using the techniques of the construction of estimators described in § 2c.

However, as rightly pointed out in [330], this method for the estimation of α is not very reliable due to the insufficient amount of observations; it requires many 'extremal' values. For that reason, the approach chosen in [330] is a different one: it takes into account, by contrast, only the results of observations fitting into the 'central' zone. The main idea of this approach is as follows.

Assume that the characteristic function

$$\varphi^{(\Delta)}(\theta) = \mathsf{E}e^{i\theta\Delta S_{t_k}}$$

has the following form:

$$\varphi^{(\Delta)}(\theta) = e^{-\gamma\Delta|\theta|^\alpha}. \tag{3}$$

Then the density $p^{(\Delta)}(x)$ of the distribution $\mathsf{P}(\Delta S_{t_k} \leqslant x)$ can be represented, by the *inversion formula*, as the integral

$$p^{(\Delta)}(x) = \frac{1}{\pi}\int_0^\infty e^{-\gamma\Delta|\theta|^\alpha}\cos\theta x\,d\theta.$$

For $x = 0$ we obtain

$$p^{(\Delta)}(0) = \frac{1}{\pi}\int_0^\infty e^{-\gamma\Delta|\theta|^\alpha}\,d\theta = \frac{\Gamma(1/\alpha)}{\pi\alpha(\gamma\theta\Delta)^{1/\alpha}}. \tag{4}$$

Hence

$$p^{(n\Delta)}(0) = n^{-1/\alpha}p^{(\Delta)}(0). \tag{5}$$

Of course, we could obtain the same without resorting to the representation (3), by applying directly to the definition of a stable law, which requires that

$$\mathrm{Law}(\Delta S_{t_1} + \Delta S_{t_2} + \cdots + \Delta S_{t_n}) = \mathrm{Law}\big(C_n(S_{t_1} - S_{t_0})\big), \qquad t_0 = 0, \tag{6}$$

and by using the equality

$$C_n = n^{1/\alpha}$$

(see Chapter III, § 1a).

Indeed, since $\Delta S_{t_1} + \Delta S_{t_2} + \cdots + \Delta S_{t_n} = S_{t_n} - S_0$, it follows that

$$\mathrm{Law}(S_{t_n} - S_0) = \mathrm{Law}\big(n^{1/\alpha}(S_{t_1} - S_0)\big).$$

Hence

$$p^{(n\Delta)}(x) = n^{-1/\alpha} p^{(\Delta)}(xn^{-1/\alpha}), \tag{7}$$

and setting $x = 0$ we obtain (5).

Relation (5) enables one to find an estimate $\widehat{\alpha}$ of the 'stability index' α by considering the empirical densities $\widehat{p}^{(n\Delta)}(0)$ with Δ equal to 1 minute and $n = 1, 3, 10, 32, 100, 316, 1000$, then passing to the logarithms, and using the least squares method. (This choice of $n = 1, 3, 10, \dots$ is explained by the fact that the corresponding values of $\log_{10} n$ are approximately equidistant: $\log_{10} 3 = 0.477$, $\log_{10} 10 = 1$, $\log_{10} 32 = 1.505$, ...)

The value of this estimate of α obtained in [330] is

$$\widehat{\alpha} = 1.40 \pm 0.05. \tag{8}$$

We point out at the onset that there is no contradiction whatsoever between this result and the estimate $\widehat{\alpha} \approx 3.5$ of the 'tail index' α in § 2c.5. The point is that these estimates are obtained under *different* assumptions about the character of distributions. In one case we assume that the distribution has 'stable' type, while in the other—that it is of Pareto type. Moreover (and this can be important), the object of the first research is the *currency cross rates* and the object of the other is the *S&P500 Index*. Generally speaking, there are no solid grounds to assume that the behavior of their distributions must be similar, for the defining factors in these two cases are distinct (the state of the *world* economy in the case of exchange rates and the state of the American *national* economy in the case of the S&P500 Index).

This idea of the different behavior of the exchange rates and financial indexes of S&P500 or DJIA kind is also substantiated by the results of the $\mathcal{R}/\mathcal{S}$-analysis described in § 4b below.

We note also that the estimate (8) is based on the 'central' values, while the estimate $\widehat{\alpha} \approx 3.5$ was found on the basis of the 'marginal' values. Hence this discrepancy is just another argument in favor of the above-mentioned thesis that the financial indexes must be described by different 'standard' distributions in different domains of their values, and that a single 'universal' distribution must be hard to find.

3. The conclusion that the empirical densities $\widehat{p}^{(n\Delta)}(x)$ can be well approximated by stable symmetric densities $p^{(n\Delta)}(x)$ in the 'central' domain can be substantiated also by the following arguments, based on the *self-similarity* property.

We consider a sample

$$\bigl(\Delta S_{t_1^{(n)}},\dots,\Delta S_{t_k^{(n)}}\bigr),\qquad t_i^{(n)}-t_{i-1}^{(n)}=n\Delta,$$

of size k, with time step $n\Delta$, where Δ is 1 minute. If we proceed now to the sample

$$\bigl(n^{-1/\alpha}\Delta S_{t_1^{(n)}},\dots,n^{-1/\alpha}\Delta S_{t_k^{(n)}}\bigr),$$

then this new sample must have the same distribution as

$$(\Delta S_{t_1},\dots,\Delta S_{t_k}),\qquad t_i-t_{i-1}=\Delta.$$

Hence, in accordance with (7), the corresponding estimates of the one-dimensional densities of the (identically distributed) variables $n^{-1/\alpha}\Delta S_{t_i^{(n)}}$ and ΔS_{t_i} must be 'very similar'.

A chart in [330] obtained by a 'superposition' (see Chapter III, § 2c.6 for information about this method) of the so transformed empirical densities for the values of Δ equal to 1, 3, 10, 32, 100, 316, and 1000 minutes strongly supports the stable distribution conjecture (with $\alpha\approx 1.40$).

One can find an estimate of the coefficient γ in (3) on the basis of the empirical density $\widehat{p}^{(\Delta)}(0)$ and the estimate $\widehat{\alpha}=1.40$, by using (4). The corresponding value is $\widehat{\gamma}=0.00375$.

3. Statistics of Volatility, Correlation Dependence, and Aftereffect in Prices

§ 3a. Volatility. Definition and Examples

1. Arguably, no concept in financial mathematics if as loosely interpreted and as widely discussed as '*volatility*'. A synonym to '*changeability*',[b] '*volatility*' has many definitions, and is used to denote various measures of changeability.

If $S_n = S_0 e^{H_n}$ with $H_0 = 0$ and $\Delta H_n = \sigma\varepsilon_n$, $n \geqslant 1$, where (ε_n) is Gaussian white noise ($\varepsilon_n \sim \mathcal{N}(0,1)$), then one means by volatility the natural measure of uncertainty and changeability, the *standard deviation* σ.

We recall that if $\xi \sim \mathcal{N}(\mu, \sigma^2)$ for a random variable ξ, then

$$\mathsf{P}(|\xi - \mu| \leqslant \sigma) \approx 0.68 \tag{1}$$

and

$$\mathsf{P}(|\xi - \mu| \leqslant 1.65\,\sigma) \approx 0.90. \tag{2}$$

Hence one can expect in approximately 90 % cases that the result of an observation of ξ deviates from the mean value μ by $1.65\,\sigma$ at most.

In employing the scheme '$S_n = S_{n-1}e^{h_n}$' one usually deals with small values of the h_n, so that

$$S_n \approx S_{n-1}(1 + h_n).$$

Hence, if $h_n = \sigma\varepsilon_n$, then, given the 'today' value of some price S_{n-1} we can say that its 'tomorrow' value S_n will in 90 % cases lie in the interval

$$[S_{n-1}(1 - 1.65\sigma), S_{n-1}(1 + 1.65\sigma)],$$

[b] "Random House Webster's concise dictionary" (Random House, New York, 1993) gives the following explanation of the adjective *volatile*: "1. evaporating rapidly. 2. tending or threatening to erupt in violence; explosive. 3. *changeable*; unstable."

so that it is only in 5 % cases that S_n is larger than $S_{n-1}(1+1.65\sigma)$ and it is smaller than $S_{n-1}(1-1.65\sigma)$ in 5 % cases.

Remark 1. This explains why, in some handbooks of finance (see, for instance, [404]), one measures volatility in terms of $\nu = 1.65\,\sigma$ in place of the standard deviation σ.

2. We have already seen before that the model '$h_n = \sigma\varepsilon_n$, $n \geqslant 1$' is fairly distant from real life. It would be more realistic to use conditional Gaussian models of the kind '$h_n = \sigma_n\varepsilon_n$, $n \geqslant 1$', with a *random* sequence $\sigma = (\sigma_n)_{n\geqslant 1}$ of $\mathscr{F}_{n-1}$-measurable variables σ_n and with $\mathscr{F}_n$-measurable ε_n, where $(\mathscr{F}_n)$ is the flow of 'information' (on the price values, say; see Chapter I, § 2a for greater detail).

There is an established tradition of calling $\sigma = (\sigma_n)_{n\geqslant 1}$ (in the above model) the sequence of 'volatilities'. The random nature of the latter can be reflected by saying that 'the volatility is volatile on its own'.

Note that

$$\mathsf{E}(h_n^2 \mid \mathscr{F}_{n-1}) = \sigma_n^2, \tag{3}$$

and the sequence $H = (H_n, \mathscr{F}_n)_{n\geqslant 1}$ of the variables $H_n = h_1 + \dots + h_n$, where $\mathsf{E}|h_n|^2 < \infty$ for $n \geqslant 1$, is a square integrable martingale with quadratic characteristic

$$\langle H\rangle_n = \sum_{k=1}^{n} \mathsf{E}(h_k^2 \mid \mathscr{F}_{k-1}), \qquad n \geqslant 1. \tag{4}$$

In view of (3),

$$\langle H\rangle_n = \sum_{k=1}^{n} \sigma_k^2; \tag{5}$$

therefore it is natural to call the quadratic characteristic

$$\langle H\rangle = (\langle H\rangle_n, \mathscr{F}_n)_{n\geqslant 1}$$

the *volatility* of the sequence H.

We note that

$$\mathsf{E}H_n^2 = \mathsf{E}\langle H\rangle_n. \tag{6}$$

3. For $ARCH(p)$ models we have

$$\sigma_n^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i h_{n-i}^2 \tag{7}$$

(see Chapter II, § 3a).

Hence the *estimation* problem for the volatilities σ_n reduces to a problem of *parametric* estimation for $\alpha_0, \alpha_1, \dots, \alpha_p$.

There exist also other, e.g., *nonparametric*, methods for obtaining estimates of volatility. For instance, if $h_n = \mu_n + \sigma_n \varepsilon_n$ for $n \geqslant 1$, where $\mu = (\mu_n)$ and $\sigma = (\sigma_n)$ are stationary sequences, then a standard estimator for σ_n is

$$\widehat{\sigma}_n = \sqrt{\frac{1}{n-1} \sum_{k=1}^{n} (h_k - \overline{h}_n)^2}\,, \tag{8}$$

where $\overline{h}_n = \dfrac{1}{n} \sum\limits_{k=1}^{n} h_k$.

It is worth noting that one can also regard the empiric volatility $\widehat{\sigma} = (\widehat{\sigma}_n)_{n \geqslant 1}$ as an index of financial statistics and to analyze it using the same methods and tools as in the case of the prices $S = (S_n)_{n \geqslant 1}$.

To this end we consider the variables

$$\widehat{r}_n = \ln \frac{\widehat{\sigma}_n}{\widehat{\sigma}_{n-1}}\,, \qquad n \geqslant 2. \tag{9}$$

Many authors and numerous observations (see, for example, [386; Chapter 10]) demonstrate that the values of the 'logarithmic returns' $\widehat{r} = (\widehat{r}_n)_{n \geqslant 2}$ oscillate rather swiftly, which indicates that the variables $\widehat{r}_n$ and $\widehat{r}_{n+1}$, $n \geqslant 2$, are negatively correlated. If we consider the example of the S&P500 Index and apply $\mathcal{R}/\mathcal{S}$-analysis to the corresponding values of $\widehat{r} = (\widehat{r}_n)_{n \geqslant 2}$ (see Chapter III, § 2a and § 4 of the present chapter), then the results will fully confirm this phenomenon of negative correlation (see [386; Chapter 10]). Moreover, we can assume in the first approximation that the $\widehat{r}_n$ are Gaussian variables, so that this negative correlation (coupled with the property of self-similarity standing out in observations) can be treated as an argument in favor of the thesis that this sequence is a *fractional noise* with Hurst parameter $\mathbb{H} < 1/2$. (According to [386], $\mathbb{H} \approx 0.31$ for the S&P500 Index.)

4. The well-known paper [44] (1973) of Black and Scholes made a significantly contribution to the understanding of the importance of the concept of volatility. This paper contains a formula for the fair (rational) price $\mathbb{C}_T$ of a standard call option (see Chapter I, § 1b). By this formula, the value of $\mathbb{C}_T$ is *independent* of μ (surprisingly, at the first glance), but *depends* on the value of the volatility σ participating in the formula describing the evolution of stock prices $S = (S_t)_{t \geqslant 0}$:

$$S_t = S_0 e^{H_t}, \qquad H_t = \sigma W_t + \left(\mu - \frac{\sigma^2}{2}\right) t, \tag{10}$$

where $W = (W_t)_{t \geqslant 0}$ is a standard Wiener process.

Of course, the assumption made in this model that the volatility σ in (10) is, first, a *constant* and, second, a *known* constant, is rather fantastic. Clearly, practical applications of the Black–Scholes formula require one to have at least a rough idea of the possible value of the volatility; this is necessary not only to find out the fair option prices, but also to assess the risks resulting from some or other decisions in models with prices described by formulas (9) and (10) in Chapter I, § 1b.

In this connection, we must discuss another (empirical) approach to the concept of 'volatility' that uses the Black–Scholes formula and the *real* option prices on securities markets.

For this definition, let $\mathbb{C}_t = \mathbb{C}_t(\sigma; T)$ be the (theoretical) value of the ask price at time $t < T$ of a standard European call-option with $f_T = (S_T - K)^+$ and with maturity time T.

The price $\mathbb{C}_t$ is theoretical. What we know in practice is the price $\widehat{\mathbb{C}}_t$ that was *actually* announced at the instant t, and we can ask for the root of the equation

$$\widehat{\mathbb{C}}_t = \mathbb{C}_t(\sigma; T). \tag{11}$$

This value of σ, denoted by $\widetilde{\sigma}_t$, is called the '*implied* volatility'; it is considered to be a good estimate for the 'actual' volatility.

It should be noted that, as regards its behavior, the 'implied volatility' is similar to the 'empirical volatility' defined (in the continuous-time case) by formulae of type (8). Its negative correlation and fractal structure are rather clearly visible (see, e.g., [386; Chapter 10]).

5. We now discuss another approach to the definition of volatility, based on the consideration of the *variation-related* characteristics of the process $H = (H_t)_{t\geqslant 0}$ defining the prices $S = (S_t)_{t\geqslant 0}$ by the formula $S_t = S_0 e^{H_t}$. The results of many statistical observations and economic arguments support the thesis that the processes $H = (H_t)_{t\geqslant 0}$ have a property of *self-similarity*, which means, in particular, that the distributions of the variables $H_{t+\Delta} - H_t$ with *distinct* values of $\Delta > 0$ are similar in certain respects (see Chapter III, § 2).

We recall that if $H = B_{\mathbb{H}}$ is a fractional Brownian motion, then

$$\mathsf{E}|H_{t+\Delta} - H_t| = \sqrt{\frac{2}{\pi}}\,\Delta^{\mathbb{H}} \tag{12}$$

for all $\Delta > 0$ and $t \geqslant 0$ and

$$\mathsf{E}|H_{t+\Delta} - H_t|^2 = \Delta^{2\mathbb{H}}. \tag{13}$$

For a strictly α-stable Lévy motion with $0 < \alpha \leqslant 2$ we have

$$\mathsf{E}|H_{t+\Delta} - H_t| = \mathsf{E}|H_\Delta| = \Delta^{1/\alpha}\mathsf{E}|H_1|. \tag{14}$$

Hence, setting $\mathbb{H} = 1/\alpha < 1$ we obtain

$$\mathsf{E}|H_{t+\Delta} - H_t| = \Delta^{\mathbb{H}}\mathsf{E}|H_1|, \tag{15}$$

which resembles formula (12) for a fractional Brownian motion.

All these formulas, together with arguments based on the Law of large numbers, suggest that it would be reasonable to introduce certain *variation-related* characteristics and to use them for testing the statistical hypothesis that the process $H = (H_t)_{t\geqslant 0}$ generating the prices $S = (S_t)_{t\geqslant 0}$ is a self-similar process of the same kind as a fractional Brownian motion or an α-stable Lévy motion.

It should be also pointed out that, from the standpoint of statistical analysis, different kinds of investors are interested in different *time intervals* and have different *time horizons.*

For instance, the *short-term* investors are eager to know the values of the prices $S = (S_t)_{t\geqslant 0}$ at times $t_k = k\Delta$, $k \geqslant 0$, with *small* $\Delta > 0$ (several minutes or even seconds). Such data are of little interest to *long-term* investors; what they value most are data on the price movements over *large* time intervals (months and even years), information on cycles (periodic or aperiodic) and their duration, information on trend phenomena, and so on.

Bearing this in mind, we shall explicitly indicate in what follows the chosen time interval Δ (taken as a unit of time, the 'characteristic' time measure of an investor) and also the interval $(a, b]$ on which we study the evolution and the 'changeability' of the financial index in question.

6. On can get a satisfactory understanding of the changeability of a process $H = (H_t)_{t\geqslant 0}$ on the time interval $(a, b]$ from the Δ-variation

$$\mathrm{Var}_{(a,b]}(H;\Delta) = \sum |H_{t_k} - H_{t_{k-1}}|, \tag{16}$$

where the sum is taken over all k such that $a \leqslant t_{k-1} < t_k \leqslant b$, and $t_k = k\Delta$.

Clearly, if a particular trajectory of $H = (H_t)_{a\leqslant t\leqslant b}$ is 'sufficiently regular' and $\Delta > 0$ is small, then the value of $\mathrm{Var}_{(a,b]}(H;\Delta)$ is close to the variation

$$\mathrm{Var}_{(a,b]}(H) \equiv \int_a^b |dH_s|, \tag{17}$$

which is by definition the supremum

$$\sup \sum |H_{t_k} - H_{t_{k-1}}| \tag{18}$$

taken over all finite partitionings $(t_0, \dots, t_n)$ of the interval $(a, b]$ such that $a = t_0 < t_1 < \cdots < t_n \leqslant b$.

In the statistical analysis of the processes $H = (H_t)_{t\geqslant 0}$ with presumably homogeneous increments, it is reasonable to consider, in place of the Δ-variations $\mathrm{Var}_{(a,b]}(H;\Delta)$, the normalized quantities

$$\nu_{(a,b]}(H;\Delta) = \frac{\mathrm{Var}_{(a,b]}(H;\Delta)}{\left[\frac{b-a}{\Delta}\right]}, \tag{19}$$

which we shall call the (empirical) Δ-*volatilities* on $(a, b]$.

It is often useful to consider the following Δ-volatility of order $\delta > 0$:

$$\nu_{(a,b]}^{(\delta)}(H;\Delta) = \frac{\operatorname{Var}_{(a,b]}^{(\delta)}(H;\Delta)}{\left[\frac{b-a}{\Delta}\right]}, \tag{20}$$

where

$$\operatorname{Var}_{(a,b]}^{(\delta)}(H;\Delta) = \sum |H_{t_k} - H_{t_{k-1}}|^{\delta} \tag{21}$$

and the summation proceeds as in (16).

We note that for a fractional Brownian motion $H = B_{\mathbb{H}}$ we have

$$\operatorname{Var}_{(a,b]}^{(2)}(H;\Delta) \xrightarrow{\mathsf{P}} \begin{cases} \infty, & 0 < \mathbb{H} < \frac{1}{2}, \\ (b-a), & \mathbb{H} = \frac{1}{2}, \\ 0, & \frac{1}{2} < \mathbb{H} \leqslant 1, \end{cases} \tag{22}$$

as $\Delta \to 0$, where "$\xrightarrow{\mathsf{P}}$" is convergence in probability.

If H is a strictly α-stable Lévy motion with $0 < \alpha < 2$, then

$$\operatorname{Var}_{(a,b]}^{(2)}(H;\Delta) \xrightarrow{\mathsf{P}} 0 \tag{23}$$

as $\Delta \to 0$.

Remark 2. One usually calls stochastic processes $H = (H_t)_{t \geqslant 0}$ with property (23) *zero-energy* processes (see, i.g., [166]). Thus, it follows from (22) and (23) that both fractional Brownian motion with $1/2 < \mathbb{H} \leqslant 1$ and strictly α-stable Lévy processes with $\mathbb{H} = 1/\alpha > 1/2$ are zero-energy processes.

7. The statistical analysis of volatility by means of $\mathcal{R}/\mathcal{S}$-analysis discussed below (see § 4) enables one to discover several remarkable and unexpected properties, which provide one with tools for the verification of some or other conjectures concerning the space-time structure of the processes $H = (H_t)_{t \geqslant 0}$ (for models with continuous time) and $H = (H_n)_{n \geqslant 0}$ (in the discrete-time case). For instance, one must definitely *discard* the conjecture of the independence of the variables h_n, $n \geqslant 1$ (generating the sequence $H = (H_n)_{n \geqslant 0}$), for many financial indexes. (In the continuous-time case one must accordingly discard the conjecture that $H = (H_t)_{t \geqslant 0}$ is a process with independent increments.)

Simultaneously, the analysis of Δ-volatility and $\mathcal{R}/\mathcal{S}$-statistics support the thesis that the variables h_n, $n \geqslant 1$, are in fact characterized by a rather *strong aftereffect*, and this allows one to cherish hopes of a 'nontrivial' prediction of the price development.

The fractal structure in volatilities can be exposed for many financial indexes (stock and bond prices, DJIA, the S&P500 Index, and so on). It is most clearly visible in currency exchange rates. We discuss this subject in the next section.

§ 3b. Periodicity and Fractal Structure of Volatility in Exchange Rates

1. In § 1b we presented the statistics (see Fig. 29 and 30) of the number of ticks occurring during a day or a week. They unambiguously indicate

the interday inhomogeneity

and

the presence of daily cycles (periodicity).

Representing the process $H = (H_t)_{t \geqslant 0}$ with $H_t = \ln \frac{S_t}{S_0}$ by the formula

$$H_t = \sum h_{\tau_k} I(\tau_k \leqslant t) \tag{1}$$

(cf. formula (7) in § 1c), we can say that Fig. 29 and Fig. 30 depict only the part of the development due to the 'time-related' components of H, the instants of ticks τ_k, but give no insight in the structure of the 'phase' component, the sequence (h_{τ_k}) or the sequence of $\widetilde{h}_k = \widetilde{h}_{t_k}^{(\Delta)}$ (see the notation in § 2b).

The above-introduced concept of Δ-volatility, based on the Δ-variation $\mathrm{Var}_{(a,b]}(H;\Delta)$, enables one to gain a distinct notion of the 'intensity' of change in the processes H and $\widetilde{H}$ both with respect to the time and the phase variables.

To this end, we consider the Δ-volatility $\nu_{(a,b]}(H;\Delta)$ on the interval $(a,b]$.

We note on the onset that if $a = (k-1)\Delta$ and $b = k\Delta$, then

$$\nu_{((k-1)\Delta,k\Delta]}(H;\Delta) = |\widetilde{H}_{k\Delta} - \widetilde{H}_{(k-1)\Delta}| = |\widetilde{h}_k| \tag{2}$$

(see the notation in § 2b).

We choose as the object of our study the DEM/USD exchange rate, so that $S_t = (\mathrm{DEM/USD})_t$ and $H_t = \ln \frac{S_t}{S_0}$.

We set Δ to be equal to 1 hour and

$$t = 1, 2, \ldots, 24 \quad \text{(hours)}$$

in the case of the analysis of the '24-hour cycle', and we set

$$t = 1, 2, \ldots, 168 \quad \text{(hours)}$$

in the case of the 'week cycle' (the clock is set going on Monday, 0:00 GMT, so that $t = 168$ corresponds to the end of the week).

The impressive database of Olsen & Associates enables one to obtain quite reliable estimates $\widehat{\nu}_{((k-1)\Delta,k\Delta]}(H;\Delta) = \widehat{|\widetilde{h}_k|}$ for the values of $\nu_{((k-1)\Delta,k\Delta]}(H;\Delta) \equiv |\widetilde{h}_k|$ for each day of the week.

To this end let time zero be 0:00 GMT of the first Monday covered by the database. If $\Delta = 1$ h, then setting $k = 1, 2, \dots, 24$, we obtain the intervals $(0,1], (1,2], \dots, (23,24]$ corresponding to the intervals (of GMT)

$$(0{:}00, 1{:}00],\ (1{:}00, 2{:}00],\ \dots,\ (23{:}00, 24{:}00].$$

As an estimate for $\widehat{\nu}_{((k-1)\Delta, k\Delta]}(H; \Delta)$ we take the arithmetic mean of the quantities $|\widetilde{H}^{(j)}_{k\Delta} - \widetilde{H}^{(j)}_{(k-1)\Delta}|$ calculated for all Mondays in the database (indexed by the integer j). In a similar way we obtain estimates of the $\widehat{\nu}_{((k-1)\Delta, k\Delta]}(H; \Delta)$ for Tuesdays ($k = 25, \dots, 48$), ..., and for Sundays ($k = 145, \dots, 168$).

The following charts (Fig. 36 and Fig. 37) from [427] are good illustrations of the *interday inhomogeneity* and of the *daily cycles* visible all over the week in the behavior of the Δ-volatility $\nu_{((k-1)\Delta, k\Delta]}(H; \Delta) = |\widetilde{h}_k|$ calculated for the one-hour intervals $((k-1)\Delta, k\Delta]$, $k = 1, 2, \dots$.

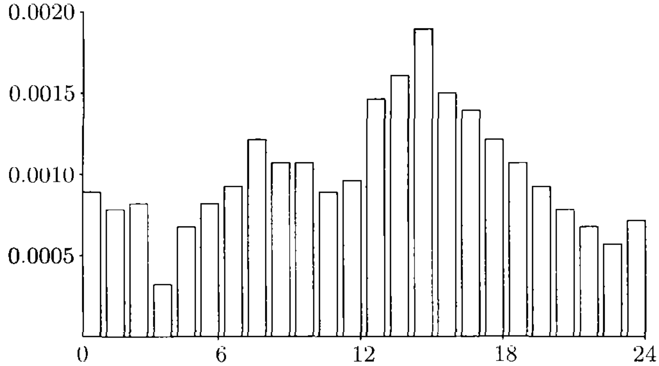


FIGURE 36. Δ-volatility of the DEM/USD cross rate during one day ($\Delta = 1$ hour), according to Reuters (05.10.1992–26.09.1993)

The above-mentioned daily periodicity of the Δ-volatility is also revealed by the analysis of its correlation properties. We devote the next section to this issue and to the discussion of the practical recommendations following from the statistical analysis of Δ-volatility.

2. We discuss now the properties of the Δ-volatility $\nu_t(\Delta) \equiv \nu_{(0,t]}(H; \Delta)$ regarded as a function of Δ for fixed t. We shall denote by $\widehat{\nu}_t(\Delta)$ its estimator $\widehat{\nu}_{(0,t]}(H; \Delta)$.

Assume that t is sufficiently large, for instance, $t = T =$ one year. We shall now evaluate $\widehat{\nu}_T(\Delta)$ for various Δ. No so long ago (see, i.g., [204], [362], [386], or [427]), the following remarkable property of the FX-market (and some other markets) was discovered: the behavior of the Δ-volatility is *highly regular*; namely,

$$\widehat{\nu}_T(\Delta) \approx C_T \Delta^{\mathbb{H}} \tag{3}$$

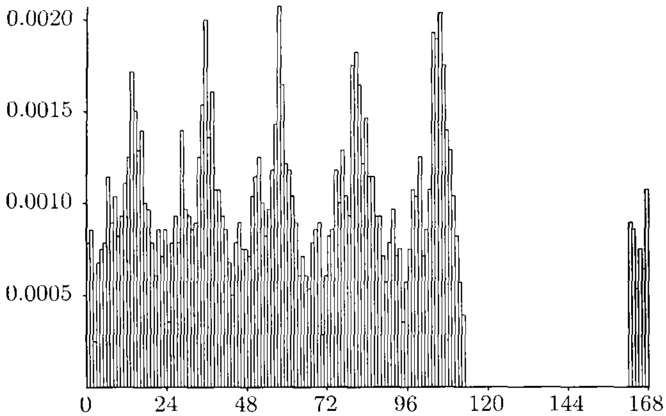


FIGURE 37. Δ-volatility of the DEM/USD cross rate during one week ($\Delta = 1$ hour), according to Reuters (05.10.1992–26.09.1993). The intervals $(0, 1], \ldots, (167, 168]$ correspond to the intervals $(0{:}00, 1{:}00], \ldots, (23{:}00, 24{:}00]$ of GMT

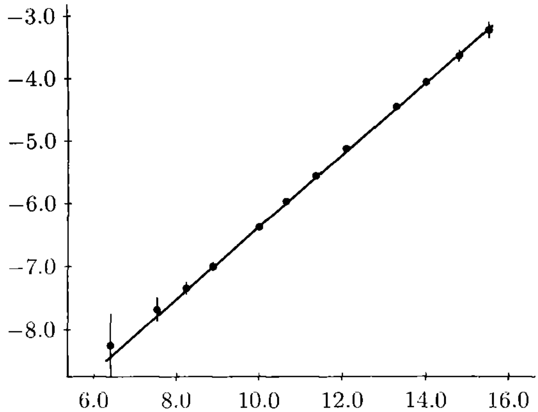


FIGURE 38. On the fractal structure of the Δ-volatility $\widehat{\nu}_T(\Delta)$. The values of $\ln \widehat{\nu}_T(\Delta)$ as a function of $\ln \Delta$ are plotted along the vertical axis

with certain constant C_T dependent on the currencies in question and with $\mathbb{H} \approx 0.585$ for the basic currencies.

To give (3) a more precise form we consider now the statistical data concerning $\log \widehat{\nu}_T(\Delta)$ as a function of $\ln \Delta$ with Δ ranging over a wide interval, from 10 minutes (= 600 seconds) to 2 months ($= 2 \times 30 \times 24 \times 60 \times 60 = 5\,184\,000$ seconds).

The chart in Fig. 38, which is constructed using the least squares method, shows that the empirical data cluster nicely along the straight line with slope $\mathbb{H} \cong 0.585$. Hence we can conclude from (3) that, for t large, the volatility $\nu_t(\Delta)$ regarded as a function of Δ has fractal structure with Hurst exponent $\mathbb{H} \cong 0.585$.

As shown in § 3a, we have $\mathsf{E}|H_\Delta| = \sqrt{2/\pi}\Delta^{1/2}$ for a Brownian motion $H = (H_t)_{t\geqslant 0}$, $\mathsf{E}|H_\Delta| = \sqrt{2/\pi}\Delta^{\mathbb{H}}$ for a fractional Brownian motion $H = (H_t)_{t\geqslant 0}$ with exponent $\mathbb{H}$, and $\mathsf{E}|H_\Delta| = \mathsf{E}|H_1|\Delta^{\mathbb{H}}$ with $\mathbb{H} = 1/\alpha < 1$ for a strictly α-stable Lévy process with $\alpha > 1$.

Thus, the experimentally obtained value $\mathbb{H} = 0.585 > 1/2$ supports the conjecture that the process $H = (H_t)$, $t \geqslant 0$, can be satisfactory described either by a fractional Brownian motion or by an α-stable Lévy process with $\alpha = \dfrac{1}{\mathbb{H}} \approx \dfrac{1}{0.585} \approx 1.7$.

Remark. As regards the estimates for $\mathbb{H}$ in the case of a fractional Brownian motion, see Chapter III, § 2c.6.

3. We now turn to Fig. 36. In it, the periods of maximum and minimum activity are clearly visible: 4:00 GMT (the minimum) corresponds to lunch time in Tokyo, Sydney, Singapore, and Hong Kong, when the life in the FX-market comes to a standstill. (This is nighttime in Europe and America). We have already pointed out that the maximum activity ($\approx$ 15:00) corresponds to time after lunch in Europe and the beginning of the business day in America.

The daily activity patterns during the five working days (Monday through Friday) are rather similar. Activity fades significantly on week-ends. On Saturday and most part of Sunday it is almost nonexistent. On Sunday evening, when the East Asian market begins its business day, activity starts to grow.

§ 3c. Correlation Properties

1. Again, we consider the DEM/USD exchange rate, which (as already pointed out in § 1a.4) is featured by high intensity of ticks (on the average, 3–4 ticks per minute on usual days and 15–20 ticks per minute on days of higher activity, as in July, 1994).

The above-described phenomena of periodicity in the occurrences of ticks and in Δ-volatility are visible also in the *correlation* analysis of the absolute values of the changes $|\Delta H|$. We present the corresponding results below, in subsection 3, while we start with the correlation analysis of the values of ΔH themselves.

2. Let $S_t = (\mathrm{DEM/USD})_t$ and let $H_t = \ln \dfrac{S_t}{S_0}$. We denote the results of an appropriate linear interpolation (see § 2b) by $\widetilde{S}_t$ and $\widetilde{H}_t = \ln \dfrac{\widetilde{S}_t}{S_0}$, respectively.

We choose a time interval Δ; let

$$\widetilde{h}_k = \widetilde{H}_{t_k} - \widetilde{H}_{t_{k-1}}$$

with $t_k = k\Delta$. (In § 2b we also used the notation $h_{t_k}^{(\Delta)}$; in accordance with § 3b, $|\widetilde{h}_k| = \nu_{(t_{k-1},t_k]}(H;\Delta)$.)

Let Δ be 1 minute and let $k = 1, 2, \dots, 60$. Then $\widetilde{h}_1, \widetilde{h}_2, \dots, \widetilde{h}_{60}$ is the sequence of consecutive (one-minute) increments of H over the period of *one* hour. We can assume that the sequence $\widetilde{h}_1, \widetilde{h}_2, \dots, \widetilde{h}_{60}$ is stationary (homogeneous) on this interval.

Traditionally, as a measure of the correlation dependence of stationary sequences $\widetilde{h} = (\widetilde{h}_1, \widetilde{h}_2, \dots)$, one takes their correlation function

$$\rho(k) = \frac{\mathsf{E}\,\widetilde{h}_n \widetilde{h}_{n+k} - \mathsf{E}\,\widetilde{h}_n \cdot \mathsf{E}\,\widetilde{h}_{n+k}}{\sqrt{\mathsf{D}\widetilde{h}_n \cdot \mathsf{D}\widetilde{h}_{n+k}}}, \tag{1}$$

(the *autocorrelation* function in the theory of stochastic processes).

The corresponding statistical analysis has been carried out by Olsen & Associates for the data in their (rather representative) database covering the period from January 05, 1987 through January 05, 1993 (see [204]). Using its results one can plot the following graph of the (empirical) autocorrelation function $\widehat{\rho}(k)$:

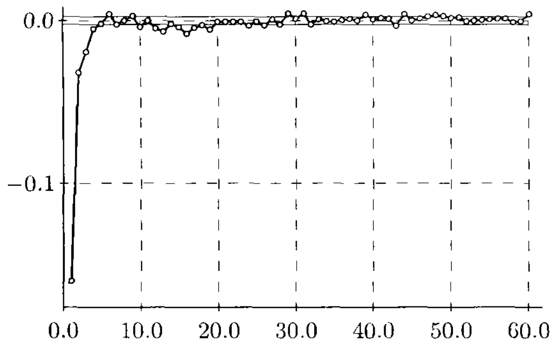


FIGURE 39. Empirical autocorrelation function $\widehat{\rho}(k)$ for the sequence of increments $\widetilde{h}_n = \widetilde{H}_{t_n} - \widetilde{H}_{t_{n-1}}$ corresponding to the DEM/USD cross rate, with $t_n = n\Delta$ and $\Delta = 1$ minute

In Fig. 39 one clearly sees a *negative* correlation on the interval of approximately 4 minutes ($\widehat{\rho}(1) < 0$, $\widehat{\rho}(2) < 0$, $\widehat{\rho}(3) < 0$, $\widehat{\rho}(4) \leqslant 0$), while most of the values of the $\widehat{\rho}(k)$ with $4 < k < 60$ are close to zero.

Bearing this observation in mind, we can assume that the variables $\widetilde{h}_n$ and $\widetilde{h}_m$, are virtually uncorrelated for $|n - m| > 4$.

We note that the *phenomenon of negative correlation* on small intervals ($|n - m| \leqslant 4$ minutes) was mentioned for the first time in [189] and [191]; it has been noticed for many financial indexes (see, for instance, [145] and [192]).

There exist various explanations in the literature of this negative correlation of the increments $\Delta\widetilde{H}$ on small intervals of time. For instance, the discussion in [204] essentially reduces to the observation that the traders in the FX-market are not uniform, their interests can 'point to different directions', and they can interpret available information in different ways. Traders often widen or narrow the spread when they are given instructions to 'get the market out of balance'. Moreover, many banks overstate their spreads systematically (see [192] in this connection).

A possible 'mathematical' explanation of the phenomenon of negative values of $\mathsf{Cov}(\widetilde{h}_n, \widetilde{h}_{n+k}) = \mathsf{E}\widetilde{h}_n\widetilde{h}_{n+k} - \mathsf{E}\widetilde{h}_n\mathsf{E}\widetilde{h}_{n+k}$ for small k can be, e.g., as follows (cf. [481]).

Let $\widetilde{H}_n = \widetilde{h}_1 + \cdots + \widetilde{h}_n$ with $\widetilde{h}_n = \mu_n + \sigma_n\varepsilon_n$, where the σ_n are $\mathscr{F}_{n-1}$-measurable and (ε_n) is a sequence of independent identically distributed random variables. We can also assume that the μ_n are $\mathscr{F}_{n-1}$-measurable variables. Judging by a large amount of statistical data, the 'mean values' μ_n are much smaller than σ_n (see, e.g., the table in § 2b.2) and can be set to be equal to zero for all practical purposes.

The values of $\widetilde{H}_n = \ln \dfrac{\widetilde{S}_n}{S_0}$ are in practice *not always known precisely*; it would be more realistic to assume that what we know are the values of $\widetilde{\widetilde{H}}_n = \widetilde{H}_n + \delta_n$, where (δ_n) is white noise, the noise component related to inaccuracies in our knowledge about the actual values of prices, rather to these values themselves.

We assume that (δ_n) is a sequence of independent random variables with $\mathsf{E}\delta_n = 0$ and $\mathsf{E}\delta_n^2 = C > 0$. Then, considering the sequence $\widetilde{\widetilde{h}} = (\widetilde{\widetilde{h}}_n)$ of the variables $\widetilde{\widetilde{h}}_n = \Delta\widetilde{\widetilde{H}}_n = \widetilde{h}_n + (\delta_n - \delta_{n-1})$ we obtain

$$\mathsf{E}\,\widetilde{\widetilde{h}}_n = 0, \qquad \mathsf{E}\,\widetilde{\widetilde{h}}{}_n^2 = \mathsf{E}\sigma_n^2 + 2C$$

and

$$\mathsf{E}\,\widetilde{\widetilde{h}}_n\,\widetilde{\widetilde{h}}_{n+1} = \mathsf{E}(\delta_n - \delta_{n-1})(\delta_{n+1} - \delta_n) = -C,$$
$$\mathsf{E}\,\widetilde{\widetilde{h}}_n\,\widetilde{\widetilde{h}}_{n+k} = 0, \qquad k > 1.$$

Hence the covariance function

$$\mathsf{Cov}(\widetilde{\widetilde{h}}_n, \widetilde{\widetilde{h}}_{n+k}) = \mathsf{E}\widetilde{\widetilde{h}}_n\widetilde{\widetilde{h}}_{n+k} - \mathsf{E}\widetilde{\widetilde{h}}_n \cdot \mathsf{E}\widetilde{\widetilde{h}}_{n+k}$$

(provided that $\mathsf{E}\sigma_n^2 = \mathsf{E}\sigma_1^2$, $n \geqslant 1$) can be described by the formula

$$\mathsf{Cov}(\widetilde{\widetilde{h}}_n, \widetilde{\widetilde{h}}_{n+k}) = \begin{cases} \mathsf{E}\sigma_1^2 + 2C, & k = 0, \\ -C, & k = 1, \\ 0, & k > 1. \end{cases}$$

3. To reveal cycles in volatilities by means of correlation analysis we proceed as follows.

We fix the interval Δ equal to 20 minutes. Let $t_0 = 0$ correspond to Monday, 0:00 GMT, let $t_1 = \Delta = 20$ m, $t_2 = 2\Delta = 40$ m, $t_3 = 3\Delta = 1$ h, ..., $t_{504} = 504\,\Delta = 1$ week, ..., $t_{2016} = 2016\,\Delta = 4$ weeks ($= 1$ month).

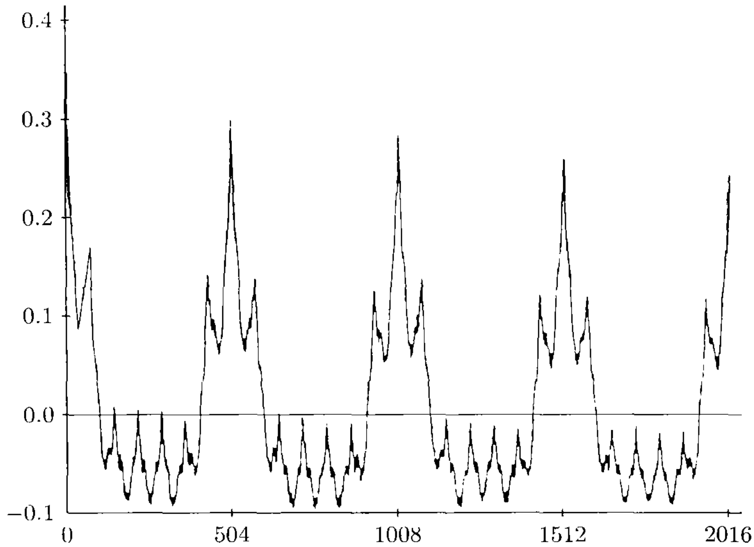


FIGURE 40. Empirical autocorrelation function $\widehat{R}(k)$ for the sequence $|\widetilde{h}_n| = |\widetilde{H}_{t_n} - \widetilde{H}_{t_{n-1}}|$ corresponding to the DEM/USD cross rate (the data of Reuters; October 10, 1992–September 26, 1993; [90], [204]). The value $k = 504$ corresponds to 1 week and $k = 2016$ to 4 weeks

We set $\widetilde{h}_n = \widetilde{H}_{t_n} - \widetilde{H}_{t_{n-1}}$; let

$$R(k) = \frac{\mathsf{E}|\widetilde{h}_n||\widetilde{h}_{n+k}| - \mathsf{E}|\widetilde{h}_n| \cdot \mathsf{E}|\widetilde{h}_{n+k}|}{\sqrt{\mathsf{D}|\widetilde{h}_n| \cdot \mathsf{D}|\widetilde{h}_{n+k}|}} \tag{2}$$

be the autocorrelation function of the sequence $|\widetilde{h}| = (|\widetilde{h}_1|, |\widetilde{h}_2|, \dots)$.

The graph of the corresponding empirical autocorrelation function $\widehat{R}(k)$ for $k = 0, 1, \dots, 2016$ (that is, over the period of *four* weeks) is plotted in Fig. 40. One clearly sees in it a *periodic* component in the autocorrelation function of the Δ-volatility of the sequence $|\widetilde{h}| = \big(|\widetilde{h}_n|\big)_{n\geqslant 1}$ with $|\widetilde{h}_n| = |\widetilde{H}_{t_n} - \widetilde{H}_{t_{n-1}}|$, $\Delta = t_n - t_{n-1}$.

As is known, to demonstrate the full strength of the correlation methods one requires that the sequence in question be stationary. We see, however, that Δ-volatility does not have this property. Hence there arises the natural desire to 'flatten' it in one or another way, making it a stationary, homogeneous sequence.

This procedure of 'flattening' the volatility is called '*devolatization*'. We discuss it in the next section, where we are paying most attention to the concept of 'change of time', well-known in the theory of random processes, and the idea of *operational* 'θ-time', which Olsen & Associates use methodically (see [90], [204], and [362]) in their analysis of the data relating to the FX-market.

§ 3d. 'Devolatization'. Operational Time

1. We start from the following example, which is a good illustration of the main stages of '*devolatization*', a procedure of 'flattening' the volatilities.

Let $H_t = \int_0^t \sigma(u)\,dB_u$, where $B = (B_t)_{t\geqslant 0}$ is a standard Brownian motion and $\sigma = (\sigma(t))_{t\geqslant 0}$ is some *deterministic* function (the 'activity') characterizing the 'contribution' of the dB_u, $u \leqslant t$, to the value of H_t. We note that

$$h_n \equiv H_n - H_{n-1} = \int_{n-1}^{n} \sigma(u)\,dB_u \overset{\mathrm{d}}{=} \sigma_n\varepsilon_n \tag{1}$$

for each $n \geqslant 1$, where $\varepsilon_n \sim \mathscr{N}(0,1)$, $\sigma_n^2 = \int_{n-1}^{n} \sigma^2(u)\,du$, and the symbol '$\overset{\mathrm{d}}{=}$' means that variables coincide in distribution.

Thus, if we can register the values of the process $H = (H_t)_{t\geqslant 0}$ only at discrete instants $n = 1, 2, \dots$, then the observed sequence $h_n \equiv H_n - H_{n-1}$ has a perfectly simple structure of a *Gaussian* sequence $(\sigma_n\varepsilon_n)_{n\geqslant 1}$ of independent random variables with zero means and, in general, inhomogeneous variances (volatilities) σ_n^2.

In the discussion that follows we present a method for transforming the data so that the inhomogeneous variables σ_n^2, $n \geqslant 1$, become 'flattened'.

We set

$$\tau(t) = \int_0^t \sigma^2(u)\,du \tag{2}$$

and

$$\tau^*(\theta) = \inf\left\{t\colon \int_0^t \sigma^2(u)\,du = \theta\right\} \quad (= \inf\{t\colon \tau(t) = \theta\}), \tag{3}$$

where $\theta \geqslant 0$.

We shall assume that $\sigma(t) > 0$ for each $t > 0$, $\int_0^t \sigma^2(u)\,du < \infty$ (this property ensures that the stochastic integral $\int_0^t \sigma(u)\,dB_u$ with respect to the Brownian motion $B = (B_u)_{u\geqslant 0}$ is well defined; see Chapter III, § 3c), and let $\int_0^t \sigma^2(u)\,du \uparrow \infty$ as $t \to \infty$.

Alongside *physical* time $t \geqslant 0$, we shall also consider new, *operational* 'time' θ defined by the formula

$$\theta = \tau(t). \tag{4}$$

The 'return' from this operational time θ to physical time is described by the inverse transfomation

$$t = \tau^*(\theta). \tag{5}$$

We note that

$$\int_0^{\tau^*(\theta)} \sigma^2(u)\, du = \theta \tag{6}$$

by (3), i.e., $\tau(\tau^*(\theta)) = \theta$, so that $\tau^*(\theta) = \tau^{-1}(\theta)$ and $\tau^*(\tau(t)) = t$.

We now consider the function $\theta = \tau(t)$ performing this transformation of physical time into operational time.

Since

$$\theta_2 - \theta_1 = \int_{t_1}^{t_2} \sigma^2(u)\, du, \tag{7}$$

we see that if the activity $\sigma^2(u)$ is small, then this transformation 'compresses' the physical time (as in Fig. 41).

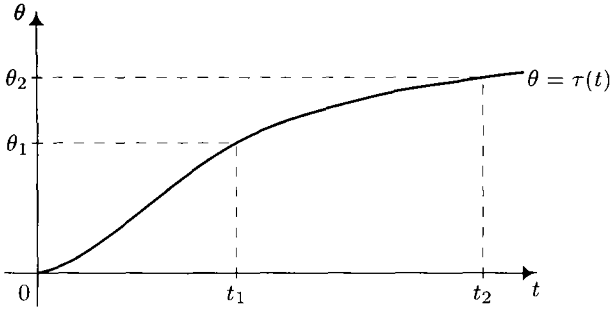


FIGURE 41. 'Compression' of a (large) time interval $[t_1, t_2]$ characterised by small activity into a (short) interval $[\theta_1, \theta_2]$ of θ-time

On the other hand, if the activity $\sigma^2(u)$ is large, then the process goes in the opposite direction: short intervals (t_1, t_2) of physical time (see Fig. 42) correspond to larger intervals (θ_1, θ_2) of operational time; time is being 'stretched'.

Now, we construct another process,

$$H^*_\theta = H_{\tau^*(\theta)}, \tag{8}$$

which proceeds in operational time. Clearly, we can 'return' from H^* to the old process by the formula

$$H_t = H^*_{\tau(t)}, \tag{9}$$

because $\tau^*(\tau(t)) = t$.

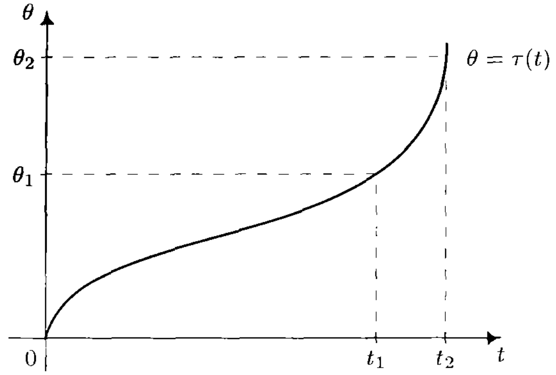


FIGURE 42. 'Stretching' a (short) interval $[t_1, t_2]$ characterized by large activity into a (long) interval $[\theta_1, \theta_2]$ of θ-time

Note that if $\theta_1 < \theta_2$, then

$$\begin{aligned} H^*_{\theta_2} - H^*_{\theta_1} &= H_{\tau^*(\theta_2)} - H_{\tau^*(\theta_1)} = \int_{\tau^*(\theta_1)}^{\tau^*(\theta_2)} \sigma(u)\, dB_u \\ &= \int_0^\infty I\bigl(\tau^*(\theta_1) < u \leqslant \tau^*(\theta_2)\bigr)\, \sigma(u)\, dB_u. \end{aligned}$$

Hence H^* is a process with *independent increments*, $H^*_0 = 0$, and $\mathsf{E} H^*_\theta = 0$. Moreover, by the properties of stochastic integrals (see Chapter III, § 3c) we obtain

$$\begin{aligned} \mathsf{E}\bigl|H^*_{\theta_2} - H^*_{\theta_1}\bigr|^2 &= \int_0^\infty I\bigl(\tau^*(\theta_1) < u \leqslant \tau^*(\theta_2)\bigr)\, \sigma^2(u)\, du \\ &= \int_{\tau^*(\theta_1)}^{\tau^*(\theta_2)} \sigma^2(u)\, du = \theta_2 - \theta_1 \end{aligned} \tag{10}$$

(the last equality is a consequence of (6)).

Since H^* is also a Gaussian process and has independent increments, zero mean, property (10), and continuous trajectories, it is just a *standard Brownian motion* (see Chapter III, § 3a), so that

$$H^*_\theta = \int_0^\theta \sigma^*(u)\, dH^*_u, \tag{11}$$

where $\sigma^*(u) \equiv 1$.

A comparison with the representation $H_t = \int_0^t \sigma(u)\, dB_u$, where $\sigma(u) \not\equiv 1$ in general, shows that the transition to operational time has 'flattened' the characteristic of activity $\sigma \equiv \sigma(u)$: when measured with respect to new 'time' θ, the level of activity is flat ($\sigma^*(u) \equiv 1$).

We have assumed above that $\sigma(u)$ is deterministic. However, $H^*_\theta = H_{\tau^*(\theta)}$ will be a Wiener process even if the change of time is *random*, defined by formula (3) with $\sigma(u) = \sigma(u;\omega)$, provided that $\int_0^t \sigma^2(u;\omega)\,du < \infty$ with probability one and $\int_0^t \sigma^2(u;\omega)\,du \uparrow \infty$ as $t \to \infty$. However, there exists a fundamental distinction between the cases of a *deterministic* function $\sigma = \sigma(u)$ and a *random* function $\sigma = \sigma(u;\omega)$: in the first case we can *calculate* the change $t \rightsquigarrow \theta = \tau(t)$ *in advance*, also for the 'future', which is impossible in the second case, because the 'random' changes of time corresponding to distinct realizations $\sigma = \sigma(u;\omega)$, $u \geqslant 0$, are distinct.

2. We consider now a sequence $h = (h_n)_{n\geqslant 1}$, where $h_n = \sigma_n \varepsilon_n$ with nonflat level of activity σ_n, $n \geqslant 1$. We shall treat n as *physical* ('old') time.

We introduce the sequence of times

$$\tau^*(\theta) = \min\left\{m \geqslant 1 \colon \sum_{k=1}^{m} \sigma_k^2 \geqslant \theta\right\},$$

with positive integer θ that we treat as *operational* ('new') time.

Also, let

$$h^*_\theta = \sum_{\tau^*(\theta-1)<k\leqslant\tau^*(\theta)} h_k$$

for $\theta = 1, 2, \dots$, where $\tau^*(0) = 0$.

We note that $\mathsf{E}h^*_0 = 0$ and the corresponding variances are

$$\mathsf{D}h^*_\theta = \mathsf{D}\Bigg[\sum_{\tau^*(\theta-1)<k\leqslant\tau^*(\theta)} h_k\Bigg] = \sum_{\tau^*(\theta-1)<k\leqslant\tau^*(\theta)} \sigma_k^2 \approx 1,$$

because the values of the σ_k^2 are usually fairly small (see the table in § 2b.2).

Thus, we can say that the transition to new 'time' θ transforms the inhomogeneous sequence $h = (h_n)_{n\geqslant 1}$ into an (almost) homogeneous sequence $h^* = (h^*_\theta)_{\theta\geqslant 1}$.

If the σ_n are random variables, that is $\sigma_n = \sigma_n(\omega)$, and our aim is to calculate the change of time for all instants (including the future), then we can implement the above-discuss idea of 'devolatization' by replacing the $\sigma_n^2(\omega)$ by their mean values $\mathsf{E}\sigma_n^2(\omega)$ or, in practice, by *estimates* of these mean values.

We see from the representation $h_n = \sigma_n \varepsilon_n$ that if the σ_n are $\mathscr{F}_{n-1}$-measurable, then $\mathsf{E}h_n^2 = \mathsf{E}\sigma_n^2$. Hence if time n of GMT falls, for example, on Monday, then we can take an estimate for $\mathsf{E}\sigma_n^2$ equal to the arithmetic mean of the values of h_k^2 over all k such that $((k-1)\Delta, k\Delta]$ corresponds to the same time interval of some other Monday covered by the database.

The change of time (2) is defined in terms of $\sigma(u)$ squared, which, of course, is not the only possible way to change time $t \rightsquigarrow \theta = \tau(t)$. For example, we could use the values of $|\sigma(u)|$ in place of $\sigma^2(u)$.

3. This is just the change of time used by Olsen & Associates ([90], [360]–[362]). They say that this kind of 'devolatization' captures *periodicity* better and the behavior pattern of the autocorrelation function of the 'devolatized' sequence $(|\widetilde{h}^*|)$ for the DEM/USD exchange rate is more 'smooth'.

Referring to [90] for the details, we shall present only the result of their statistical analysis of the properties of the autocorrelation function of the sequence $|\widetilde{h}^*|$.

As in § 3c.3, we assume that Δ is equal to 20 minutes, $t_n = n\Delta$, and $\widetilde{h}_n = \widetilde{H}_{t_n} - \widetilde{H}_{t_{n-1}}$.

In Fig. 42 (§ 3c) we plotted the graph of the empirical estimate $\widehat{R}(k)$ of the autocorrelation function

$$R(k) = \frac{\mathsf{E}|\widetilde{h}_n||\widetilde{h}_{n+k}| - \mathsf{E}|\widetilde{h}_n| \cdot \mathsf{E}|\widetilde{h}_{n+k}|}{\sqrt{\mathsf{D}(|\widetilde{h}_n|) \cdot \mathsf{D}(|\widetilde{h}_{n+k}|)}}, \tag{12}$$

in which the periodic structure is clearly visible.

In [90], after 'devolatization' and the transition to new, operational 'time' θ the authors plot a graph (see Fig. 43) of the *empirical* correlation function $\widehat{R}^*(\theta)$, $\theta \geqslant 0$, for the sequence $|h^*| = (|h^*_\theta|)_{\theta \geqslant 1}$. This graph is rather interesting for further analysis. In the same paper one can find a graph (see Fig. 44) of the transition $\tau^*(\theta)\colon \theta \rightsquigarrow t$ from operational time θ to real time t (just in our case of the DEM/USD cross rate; new time is normalized so that a week of physical time corresponds to a week of operational time).

It is clear from Fig. 44 that the function $t = \tau^*(\theta)$ is approximately linear during the five business days, while on the week-end, when the transactions in the FX-market fade, large intervals of physical time correspond to small intervals of operational time. (Only the latter are in fact interesting for business trading).

4. It should be noted that the above method of 'devolatizing' in order to eliminate the periodic component is not the only one used in the analysis of the FX-market. For instance, we can point out the papers [7], [13], and [306], where most diversified techniques, linear and nonlinear regression analysis, Fourier transformation, neural networks are used to find similar patterns in financial time series. We also point out the paper [297] by I. L. Legostaeva and this author, which relates to the same range of problems. In it (in connection with the study of the Wolf numbers characterizing solar activity), to analyze the 'trend' component $a = (a_k)$ of a sequence $\xi = (\xi_k)$ with an additive 'white noise' $\eta = (\eta_k)$ ($\xi_k = a_k + \eta_k$), the authors use a *minimax approach* suitable for the study of a much broader class of 'trend' sequences $a = (a_k)_{k \geqslant 1}$ than the standard polynomials classes of regression analysis. (See also [45], [338], and [416].)

5. For conclusion we present a graph of the periodic component of the 'activity' (see § 3b) corresponding to the CHF/USD cross rate, which is isolated by means of 'devolatization' (see [90]; cf. also Fig. 37 in § 3b).

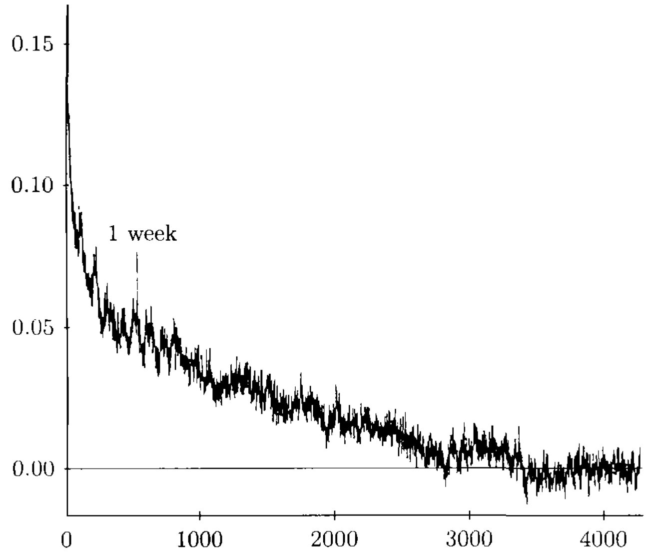


FIGURE 43. Empirical autocorrelation function $\widehat{R}^*(\theta)$ for the sequence $|\widetilde{h}^*| = (|\widetilde{h}^*_\theta|)_{\theta \geqslant 1}$ of the absolute values of 'devolatized' variables considered with respect to *operational* 'time' θ, with interval $\Delta\theta = 20$ m; the case of the DEM/USD cross rate ([90])

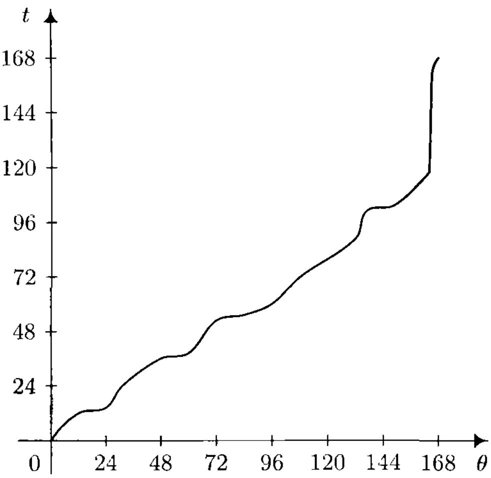


FIGURE 44. Graph of the transition $t = \tau^*(\theta)$ from operational time θ (plotted along the horizontal axis) to physical time (along the vertical axis). Time is measured in hours; 168 hours form 1 week ([90])

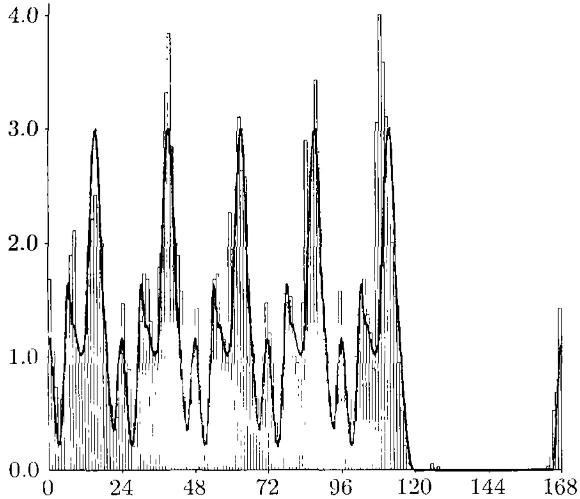


FIGURE 45. The thick curve is the periodic component of the 'activity' corresponding to the CHF/USD cross rate (over the period of 168 hours= 1 week)

In Fig. 45 we clearly see the geography-related structure of the periodic component (with 24-hour cycle during the week) arising from the differences in business time between three major FX-markets: East Asia, Europe, and America. In [D4] one can find an interesting representation of this component as the sum of three periodic components corresponding to these three markets. This can be helpful if one wants to take the factor of periodicity more consistently into account in the predictions of the evolution of exchange rates.

§ 3e. 'Cluster' Phenomenon and Aftereffect in Prices

1. We assumed in our initial scheme that the exchange rates (prices) $S = (S_t)_{t\geqslant 0}$ and their logarithms $H_t = \ln \dfrac{S_t}{S_0}$ could be described by stochastic processes with discrete intervention of chance:

$$S_t = S_0 + \sum_{n\geqslant 1} s_{\tau_n} I(\tau_n \leqslant t) \tag{1}$$

and

$$H_t = \sum_{n\geqslant 1} h_{\tau_n} I(\tau_n \leqslant t). \tag{2}$$

After that, we proceeded to their continuous modifications $\widetilde{S} = (\widetilde{S}_t)_{t\geqslant 0}$, $\widetilde{H} = (\widetilde{H}_t)_{t\geqslant 0}$ and, finally, to the variables $\widetilde{h}_n = \widetilde{H}_{t_n} - \widetilde{H}_{t_{n-1}}$, where $t_n - t_{n-1} \equiv \Delta$

(= const). It was for these variables $\widetilde{h}_n$ and for Δ equal to 1 minute that we discovered the *negative* autocovariance $\widetilde{\rho}(k) = \mathsf{E}\widetilde{h}_n\widetilde{h}_{n+k} - \mathsf{E}\widetilde{h}_n\,\mathsf{E}\widetilde{h}_{n+k}$ for small values of k $(k = 1, 2, 3, 4)$). For k large the autocovariance is close to zero, therefore we can assume for all practical purposes that the variables $\widetilde{h}_n$ and $\widetilde{h}_{n+k}$ are *uncorrelated* for such k.

Of course we are *far from saying* that they are *independent*; our analysis of the empirical autocovariance function $\widehat{R}(k)$ in § 3c (for the DEM/USD cross rate, as usual) showed that this was not the case.

The next step ('devolatization') enables us to flatten the level of 'activity' by means of the transition to new, operational time that takes into account the different intensity of changes in the values of the process $\widetilde{H} = (\widetilde{H}_t)_{t\geqslant 0}$ on different intervals.

As is clear from the statistical analysis of the sequence $(|\widetilde{h}_n|)_{n\geqslant 1}$ considered with respect to operational time θ, the autocorrelation function $\widehat{R}^*(\theta)$

1) is *rather large for small* θ;
2) *decreases fairly slowly with the growth of* θ.

It is maintained in [90] that for periods of about one month the behavior of $\widehat{R}^*(\theta)$ can be fairly well described by a *power* function, rather than an exponential; that is, we have

$$\widehat{R}^*(\theta) \sim k\theta^{-\alpha} \quad \text{as} \quad \theta \to \infty, \tag{3}$$

rather than

$$\widehat{R}^*(\theta) \sim k \exp(-\theta^{\beta}) \quad \text{as} \quad \theta \to \infty, \tag{4}$$

as could be expected and which holds in many popular models of financial mathematics (for instance, *ARCH* and *GARCH*; see [193] and [202] for greater detail).

This property of the slow decrease of the empirical autocorrelation function $\widehat{R}^*(\theta)$ has important practical consequences. It means that we have indeed a strong aftereffect in prices; 'prices remember their past', so to say. Hence one may hope to be able to predict price movements. To this end, one must, of course, learn to produce sequences $h = (h_n)_{n\geqslant 1}$ that have at least correlation properties similar to the ones observed in practice. (See [89], [360], and Chapter II, § 3b in this connection).

2. The fact that the autocorrelation is fairly *strong* for small θ is a convincing explanation for the *cluster property* of the 'activity' measured in terms of the volatility of $|\widetilde{h}_n|$.

This property, known since B. Mandelbrot's paper [322] (1963), essentially means the kind of behavior when *large values* of volatility are usually followed also by *large values* and *small values* are followed by *small* ones.

That is, if the variation $|\widetilde{h}_n| = |\widetilde{H}_{t_n} - \widetilde{H}_{t_{n-1}}|$ is large, then (with probability sufficiently close to one) the next value, $|\widetilde{h}_{n+1}|$, will also be large, while if $|\widetilde{h}_n|$ is small, then (with probability close to one) the next value will also be small. This property is clearly visible in Fig. 46; it can be observed in practice for many financial indexes.

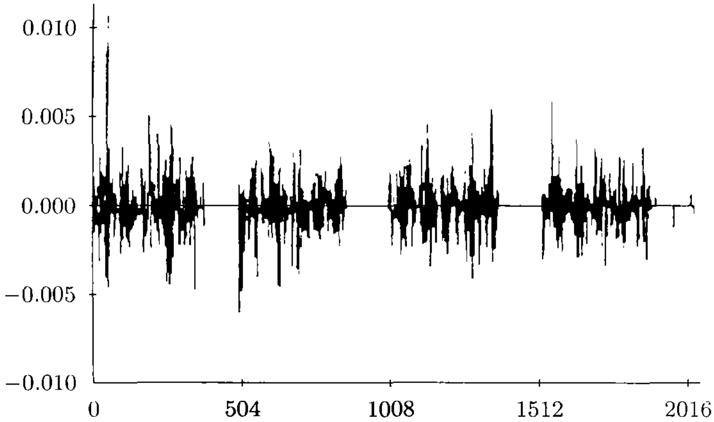


FIGURE 46. Cluster property of the variables $\widetilde{h}_k = \widetilde{h}^{(\Delta)}_{t_k}$ for the DEM/USD cross rate (the data of Reuters; October 5, 1992–November 2, 1992; [427]). The interval Δ is 20 minutes, the mark 504 corresponds to one week, and 2016 corresponds to four weeks. The clusters of large and small values of $|\widetilde{h}_k|$ are clearly visible.

We note that this cluster property is also well captured by $\mathcal{R}/\mathcal{S}$-analysis, which we discuss in the next section.

4. Statistical $\mathcal{R}/\mathcal{S}$-Analysis

§ 4a. Sources and Methods of $\mathcal{R}/\mathcal{S}$-Analysis

1. In Chapter III, § 2a we described the phenomena of long memory and self-similarity discovered by H. Hurst [236] (1951) in the statistical data concerning Nile's annual run-offs. This discovery brought him to the creation of the so-called $\mathcal{R}/\mathcal{S}$-analysis.

This method is not sufficiently well-known to practical statisticians. However, Hurst's method deserves greater attention, for it is robust and enables one to detect such phenomena in statistical data as the *cluster property*, *persistence* (in following a trend), *strong aftereffect*, *long memory*, *fast alternation* of successive values (antipersistence), the *fractal property*, the *existence of periodic or aperiodic cycles*. It can also *distinguish* between 'stochastic' and 'chaotic' noises and so on.

Besides Hurst's keystone paper [236], a fundamental role in the development of $\mathcal{R}/\mathcal{S}$-analysis, its methods and applications belongs to B. Mandelbrot and his colleagues ([314], [316]–[319], [321]–[325], [327]–[329]), and also to the works of E. Peters (which include two monographs [385] and [386]), containing large stocks of (mostly descriptive) information about the applications of $\mathcal{R}/\mathcal{S}$-analysis in financial markets.

2. Let $S = (S_n)_{n\geqslant 0}$ be some financial index and let $h_n = \ln \dfrac{S_n}{S_{n-1}}$, $n \geqslant 1$.

The main idea of the use of $\mathcal{R}/\mathcal{S}$-analysis in the study of the properties of $h = (h_n)_{n\geqslant 1}$ is as follows.

Let $H_n = h_1 + \cdots + h_n$, $n \geqslant 1$, and let

$$\mathcal{R}_n = \max_{k\leqslant n}\left(H_k - \frac{k}{n}H_n\right) - \min_{k\leqslant n}\left(H_k - \frac{k}{n}H_n\right) \tag{1}$$

(cf. Chapter III, § 2a).

The quantity $\overline{h}_n \equiv \dfrac{H_n}{n}$ is the *empirical mean* of the sample $(h_1, h_2, \dots, h_n)$, therefore $H_k - \dfrac{k}{n} H_n = \sum\limits_{i=1}^{k} (h_i - \overline{h}_n)$ is the deviation of H_k from the empirical mean value $\dfrac{k}{n} H_n$, and the quantity $\mathcal{R}_n$ itself characterizes the range of the sequences $(h_1, \dots, h_n)$ and $(H_1, \dots, H_n)$ relative to their empirical means.

Also let

$$\mathcal{S}_n^2 = \frac{1}{n} \sum_{k=1}^{n} h_k^2 - \left(\frac{1}{n} \sum_{k=1}^{n} h_k \right)^2 = \frac{1}{n} \sum_{k=1}^{n} (h_k - \overline{h}_n)^2 \tag{2}$$

be the *empirical variance* and let

$$\mathcal{Q}_n \equiv \frac{\mathcal{R}_n}{\mathcal{S}_n} \tag{3}$$

be the normalized, 'adjusted range' (in the terminology of [157]) of the cumulative sums H_k, $k \leqslant n$.

It is clear from (1)–(3) that $\mathcal{Q}_n$ has the important property of *invariance* under linear transformations $h_k \to c\,(h_k + m)$, $k \geqslant 1$. This valuable property makes this statistics *nonparametric* (at any rate, it is independent of the first two moments of the distributions of h_k, $k \geqslant 1$).

3. In the case when $h_1, h_2, \dots$ is a sequence of independent identically distributed random variables with $\mathsf{E} h_n = 0$ and $\mathsf{D} h_n = 1$, W. Feller [157] discovered that

$$\mathsf{E}\mathcal{R}_n \sim \sqrt{\frac{\pi}{2}}\, n^{1/2} \qquad (= 1.2533\ldots \times n^{1/2}) \tag{4}$$

and

$$\mathsf{D}\mathcal{R}_n \sim \left(\frac{\pi^2}{6} - \frac{\pi}{2} \right) n \qquad (= 0.07414\ldots \times n) \tag{5}$$

for large n.

This result can be readily understood using the Donsker–Prokhorov invariance principle (see, e.g., [39] or [250]) asserting that the asymptotic distribution for $\mathcal{R}_n / \sqrt{n}$ is the distribution of the range

$$R^* = \sup_{t \leqslant 1} B_t^0 - \inf_{t \leqslant 1} B_t^0 \tag{6}$$

of a *Brownian bridge* $B^0 = (B_t^0)_{t \leqslant 1}$, i.e., of the process

$$B_t^0 \equiv B_t - t B_1, \tag{7}$$

where $B = (B_t)_{t \geqslant 0}$ is a standard Brownian motion (see Chapter III, § 3a).

Indeed, let us consider the variable

$$\frac{\mathcal{R}_n}{\sqrt{n}} = \max_{k\leqslant n}\left[\frac{H_k}{\sqrt{n}} - \frac{k}{n}\frac{H_n}{\sqrt{n}}\right] - \min_{k\leqslant n}\left[\frac{H_k}{\sqrt{n}} - \frac{k}{n}\frac{H_n}{\sqrt{n}}\right]. \tag{8}$$

By the very meaning of *invariance* (i.e., the independence of the particular form of the distribution of the h_k), in looking for the limit distribution for $\mathcal{R}_n/\sqrt{n}$ as $n \to \infty$ we can assume that the h_k have the standard normal distribution $\mathscr{N}(0,1)$. If $B = (B_t)_{t\geqslant 0}$ is a standard Brownian motion, then the probability distributions for the collections $\{H_k/\sqrt{n},\ k = 1,\dots,n\}$ and $\{B_{k/n},\ k = 1,\dots,n\}$ are the same, so that

$$\begin{aligned}\frac{\mathcal{R}_n}{\sqrt{n}} &\overset{\mathrm{d}}{=} \max_{\{t\colon t=\frac{1}{n},\dots,\frac{n-1}{n},1\}}\bigl[B_t - tB_1\bigr] - \min_{\{t\colon t=\frac{1}{n},\dots,\frac{n-1}{n},1\}}\bigl[B_t - tB_1\bigr]\\ &\overset{\mathrm{d}}{=} \max_{\{t\colon t=\frac{1}{n},\dots,\frac{n-1}{n},1\}} B_t^0 - \min_{\{t\colon t=\frac{1}{n},\dots,\frac{n-1}{n},1\}} B_t^0,\end{aligned} \tag{9}$$

where the symbol '$\overset{\mathrm{d}}{=}$' means coincidence in distribution.

Hence it is clear that, as $n \to \infty$, the distribution of $\mathcal{R}_n/\sqrt{n}$ (weakly) converges to that of R^*. (Note that the functional $(\max - \min)(\,\cdot\,)$ is continuous in the space of right-continuous functions having limits from the left. See, for example, [39; Chapter 6], [250; Chapter VI], and [304] about this property and the convergence of measures in such function spaces.)

We know the following explicit formula for the density $f^*(x)$ of the distribution function $F^*(x) = \mathsf{P}(R^* \leqslant x)$ (see formula (4.3) in [157]):

$$\begin{aligned}f^* = xe''(x) + \sum_{k=2}^{\infty}\bigl\{2k(k-1)\bigl[e'\bigl((k-1)x\bigr) - e'(kx)\bigr]\\ + (k-1)^2 xe''\bigl((k-1)x\bigr) + k^2 xe''(kx)\bigr\},\end{aligned} \tag{10}$$

where $e(x) = e^{-2x^2}$.

Using this formula it is easy to see that

$$\mathsf{E}R^* = \sqrt{\frac{\pi}{2}} \quad\text{and}\quad \mathsf{D}R^* = \left(\frac{\pi^2}{6} - \frac{\pi}{2}\right). \tag{11}$$

We note by the way that $\mathsf{E}\sup\limits_{t\leqslant 1}|B_t| = \sqrt{\pi/2}$; see Chapter III, § 3b. Hence the *mean values* of the statistics R^* and $\sup\limits_{t\leqslant 1}|B_t|$ are the same.

4. Assuming that the variables $h_1, h_2, \dots$ are independent and identically distributed, with $\mathsf{E} h_i = 0$ and $\mathsf{D} h_i = 1$, we see that $\mathcal{S}_n^2 \to 1$ with probability one as $n \to \infty$ (the Strong law of large numbers). Hence the limit distribution for $\mathcal{Q}_n/\sqrt{n}$ as $n \to \infty$ is also the distribution of R^*.

The distribution of $\mathcal{Q}_n$ is independent of the mean value and the variance of h_k, $k \leqslant n$. This 'nonparametric' property brings us to the following criterion, which enables one to reject (with one or another level of reliability) the following hypothesis $\mathscr{H}_0$ lying in the foundation of the classical concept of an *efficient* market (see Chapter I, § § 2a, e): the prices in question are governed by the *random walk* model.

The main idea of this criterion, which is based on the $\mathcal{R}/\mathcal{S}$-statistics, is as follows (G. Hurst, [236]; [329], [386]).

If hypothesis $\mathscr{H}_0$ is valid, then for sufficiently large n the value of $\mathcal{R}_n/\mathcal{S}_n$ must be close to $\mathsf{E}_0 \dfrac{\mathcal{R}_n}{\mathcal{S}_n} \sim \sqrt{\dfrac{\pi}{2}\, n}$, therefore

$$\ln \frac{\mathcal{R}_n}{\mathcal{S}_n} \approx \ln \sqrt{\frac{\pi}{2}} + \frac{1}{2} \ln n. \tag{12}$$

(Of course, one can—and must—assign a precise probabilistic meaning to this relation and to relations (13) and (14) below by referring to limit theorems. We shall not do this here, taking the standpoint of a 'sensible' interpretation, which is common in practical statistics.)

Thus, choosing a logarithmic scale along both axes we see that the values of $\ln \dfrac{\mathcal{R}_n}{\mathcal{S}_n}$ must 'cluster' (provided $\mathscr{H}_0$ holds) along the straight line $a + \frac{1}{2} \ln n$ with $a = \ln \sqrt{\pi/2}$ (see Fig. 47).

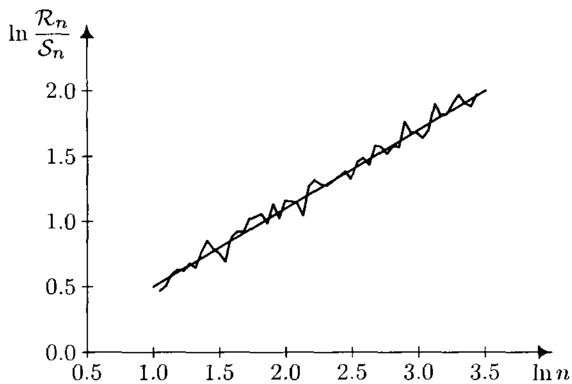


FIGURE 47. To $\mathcal{R}/\mathcal{S}$-analysis (the case of the validity of hypothesis $\mathscr{H}_0$)

This makes clear the method of $\mathcal{R}/\mathcal{S}$-analysis: given statistical data, we plot the points $\left(\ln n, \ln \dfrac{\mathcal{R}_n}{\mathcal{S}_n}\right)$ (i.e., we use logarithmic scales); then, using the least squares method, we draw a line $\widehat{a}_n + \widehat{b}_n \ln n$. If the value of $\widehat{b}_n$ 'significantly' deviates from 1/2, then hypothesis $\mathscr{H}_0$ must be rejected. (Of course, one must master the standard methods of statistical analysis so as to be able to calculate the significance of the deviation of $\widehat{b}_n$ from 1/2 under the hypothesis $\mathscr{H}_0$. This is not easy because the distribution of the statistics $\mathcal{R}_n/\mathcal{S}_n$ with finite n is difficult to find explicitly; we shall touch upon this subject below, in subsection 6.)

The principal significance of Hurst's research lies in his discovery (by means of $\mathcal{R}/\mathcal{S}$-analysis) that, in place of the expected (for Nile and other rivers) property

$$\frac{\mathcal{R}_n}{\mathcal{S}_n} \sim cn^{1/2} \tag{13}$$

one actually has

$$\frac{\mathcal{R}_n}{\mathcal{S}_n} \sim cn^{\mathbb{H}} \tag{14}$$

with $\mathbb{H}$ considerably larger than 1/2.

5. If an empirical study shows (somewhat unexpectedly) that (14) holds, then one should ask about models governing sequences $h = (h_n)_{n\geqslant 1}$ with this property.

We must also explain why $\mathbb{H} > 1/2$ in many cases. (As we shall see below, one possible explanation is that $h = (h_n)_{n\geqslant 1}$ is a sequence with *long memory* and positive correlation.)

In this connection, we discuss now several ideas, which are based, in particular, on numerical computer calculations and several results of $\mathcal{R}/\mathcal{S}$-analysis. Doing that we shall mainly follow [316] and [319].

As a rule, if a sequence $h = (h_n)_{n\geqslant 1}$ is characterized by 'weak dependence' (as are Markov, autoregressive, and some other sequences) then $\mathbb{H}$ (the *Hurst paremeter*) is close to 1/2. One usually says in this case that the system $h = (h_n)_{n\geqslant 1}$ has a 'short $\mathcal{R}/\mathcal{S}$-memory'.

If $h_n = B_{\mathbb{H}}(n) - B_{\mathbb{H}}(n-1)$, $n \geqslant 1$, where $B_{\mathbb{H}} = (B_{\mathbb{H}}(t))_{t\geqslant 0}$ is a fractional Brownian motion (Chapter III, § 2b), then the variables $\dfrac{\mathcal{R}_n}{\mathcal{S}_n n^{\mathbb{H}}}$ have a nontrivial limit distribution as $n \to \infty$, which can be indicated as in (14). Hence if a statistical study shows that $0 < \mathbb{H} \leqslant 1$ and $\mathbb{H} \neq 1/2$, then a *fractional Gaussian noise* is one possible explanation.

We recall (see Chapter III, § 2c) that in the case of such a noise, the corresponding correlation is positive if $\mathbb{H} > 1/2$ and negative if $\mathbb{H} < 1/2$. This explain why one talks about *persistence* in trend-following, a 'long memory', a 'strong aftereffect' in the first case: if the values have increased, then the odds are that they will increase further.

As pointed out in [386], the (hitherto widespread in the literature) opinion that $\mathbb{H}$ must be always larger than 1/2 in financial time series is ungrounded. The case $\mathbb{H} < 1/2$ occurs, for instance, for the returns of volatility (see Chapter III, § 2d.5 and § 4b.3 in this chapter), which are featured by strong alternation ('antipersistence').

6. We have already mentioned that in using $\mathcal{R}/\mathcal{S}$-analysis to determine the values of $\mathbb{H}$ in the conjectural relation (14) we must, of course, determine the degree of agreement between the empirical data and the model corresponding to the chosen value of $\mathbb{H}$.

In other words, there arises a standard question on the reliability of statistical inference, and one has to turn to the 'goodness-of-fit', 'significance', or other tests developed in mathematical statistics.

It should be pointed out that the complexity of the statistics $\dfrac{\mathcal{R}_n}{\mathcal{S}_n}$ leaves no hope for satisfactory formulas describing its probability distributions for various n even if we accept the null hypothesis $\mathscr{H}_0$. (Nevertheless, the question of the behavior of the mean value $\mathsf{E}_0 \dfrac{\mathcal{R}_n}{\mathcal{S}_n}$, where E_0 is averaging under condition $\mathscr{H}_0$, has been considered in [8].)

This explains the widespread use of the Monte Carlo method in $\mathcal{R}/\mathcal{S}$-analysis (see, for instance, [317], [329], [385], [386]), in particular, to assess the quality of the estimation of the unknown value of $\mathbb{H}$.

7. When adjusting theoretical models to real statistical data it is reasonable to start with simple schemes allowing an easy analytic study. For instance, describing the probabilistic structure of the sequences $h = (h_n)_{n\geqslant 1}$ one can well assume that this is fractional Gaussian noise with $0 < \mathbb{H} < 1$. If $\mathbb{H} = 1/2$, then we obtain usual Gaussian white noise, which lies at the core of many linear (AR, MA, $ARMA$) and nonlinear ($ARCH$, $GARCH$) models.

$\mathcal{R}/\mathcal{S}$-analysis produces good results in models where $h = (h_n)_{n\geqslant 1}$ is a sequence of a fractional Gaussian noise kind (see [317], [329], [385], [386]). In dealing with other models it is reasonable to consider, besides $\mathcal{Q}_n \equiv \dfrac{\mathcal{R}_n}{\mathcal{S}_n}$, the statistics

$$\mathcal{V}_n = \frac{\mathcal{Q}_n}{\sqrt{n}}. \tag{15}$$

A mere visual examination of the latter often brings important statistical conclusions.

This method is based on the thesis that, for 'white' noise ($\mathbb{H} = 1/2$) the statistics $\mathcal{V}_n$ must stabilize for large n (that is, $\mathcal{V}_n \to c$ for some constant c, where we understand the convergence in a suitable probabilistic sense).

If $h = (h_n)_{n\geqslant 1}$ is fractional Gaussian noise with $\mathbb{H} > 1/2$, then the values of $\mathcal{V}_n$ must *increase* (with n); by contrast, $\mathcal{V}_n$ must *decrease* for $\mathbb{H} < 1/2$.

Bearing this in mind we consider now the simplest, autoregressive model of order one ($AR(1)$), in which

$$h_n = \alpha_0 + \alpha_1 h_{n-1} + \varepsilon_n, \qquad n \geqslant 1. \tag{16}$$

The behavior of the sequence in this model is completely determined by the 'noise' terms ε_n and the initial value h_0.

The corresponding (to these $h_n, n \geqslant 1$) variables $\mathcal{V} = (\mathcal{V}_n)_{n\geqslant 1}$ increase with n. However, this does not mean that we have here fractional Gaussian noise with $\mathbb{H} > 1/2$ for the mere reason that this growth can be a consequence of the *linear dependence* in (16).

Hence, to understand the 'stochastic' nature of the sequence $\varepsilon = (\varepsilon_n)_{n\geqslant 1}$ in place of the variables $h = (h_n)_{n\geqslant 1}$ one would rather consider the linearly adjusted variables $h^0 = (h_n^0)_{n\geqslant 1}$, where $h_n^0 = h_n - (a_0 + a_1 h_{n-1})$ and a_0, a_1 are some estimates of the (unknown, in general) parameters, α_0 and α_1.

Producing a sequence $h = (h_n)_{n\geqslant 1}$ in accordance with (16), where $\varepsilon = (\varepsilon_n)_{n\geqslant 1}$ is white Gaussian noise, we can see that, indeed, the behavior of the corresponding variables $\mathcal{V}_n^\circ = \mathcal{V}_n(h^\circ)$ is just as it should be for fractional Gaussian noise with $\mathbb{H} = 1/2$. The following picture is a *crude* illustration to these phenomena:

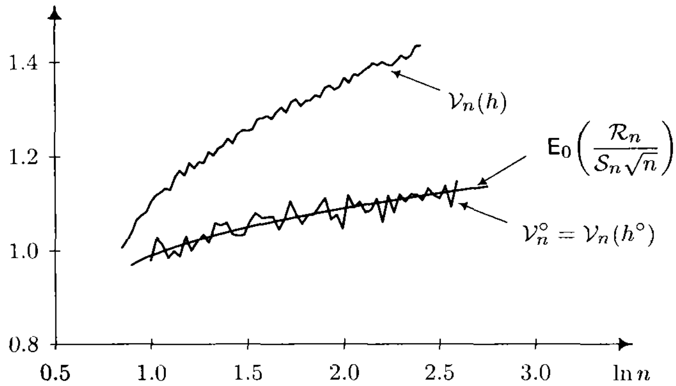


FIGURE 48. Statistics $\mathcal{V}_n(h)$ and $\mathcal{V}_n^\circ = \mathcal{V}_n(h^\circ)$ for the $AR(1)$ model. The expectation $\mathsf{E}_0\left(\frac{\mathcal{R}_n}{\mathcal{S}_n\sqrt{n}}\right)$ is evaluated under condition $\mathscr{H}_0$

If we consider the *linear* models $MA(1)$ or $ARMA(1,1)$, then the general pattern (see [386; Chapter 5]) is the same as in Fig. 48.

For the nonlinear $ARCH$ and $GARCH$ models realizations of $\mathcal{V}_n(h)$ and $\mathcal{V}_n(h^\circ)$ show another pattern of behavior relative to $\mathsf{E}_0\left(\frac{\mathcal{R}_n}{\mathcal{S}_n\sqrt{n}}\right)$, $n \geqslant 1$ (see Fig. 49). First, the graphs of $\mathcal{V}_n(h)$ and $\mathcal{V}_n(h^\circ)$ are fairly similar, which can be interpreted as the absence of linear dependence between the h_n. Second, if n is not very large, then the graphs of $\mathcal{V}_n(h)$ and $\mathcal{V}_n(h^\circ)$ lie slightly above the graph of $\mathsf{E}_0\left(\frac{\mathcal{R}_n}{\mathcal{S}_n\sqrt{n}}\right)$ corresponding to white noise, which can be interpreted as a 'slight persistence' in the model governing the variables h_n, $n \geqslant 1$. Third, the 'antipersistence' effect comes to the forefront with the growth of n.

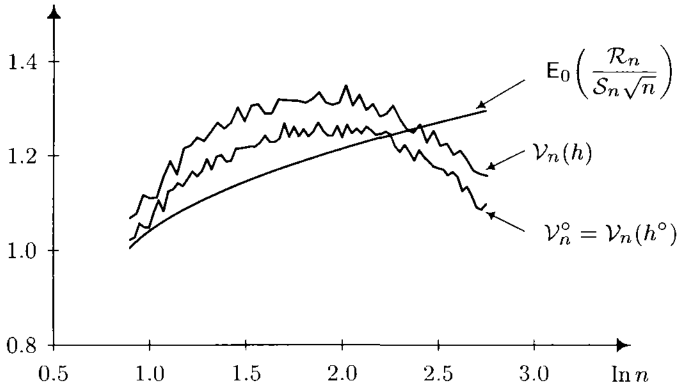


FIGURE 49. Statistics $\mathcal{V}_n(h)$ and $\mathcal{V}_n^\circ = \mathcal{V}_n(h^\circ)$ for the $ARCH(1)$ model

Remark. The terms 'persistence, 'antipersistence', and the other used in this section are adequate in the case of models of fractional Gaussian noise kind. However, the $ARCH$, $GARCH$, and kindred models (Chapter II, § § 3a, b) *are neither* fractional nor self-similar. Hence one must carry out a more thorough investigation to explain the phenomena of 'antipersistence' type in these models. This is all the more important as both linear and nonlinear models of these types are very popular in the analysis of financial time series and one needs to know what particular *local* or *global* features of empirical data related to their behavior in time can be 'captured' by these models.

8. As pointed out in Chapter I, § 2a, the underlying idea of M. Kendall's analysis of stock and commodity prices [269] is to discover *cycles* and *trends* in the behavior of these prices.

Market analysts and, in particular, 'technicians' (Chapter I, § 2e) do their research first of all on the premise that there do exist certain cycles and trends in the market, that market dynamics has rhythmic nature.

This explains why, in the analysis of financial series, one pays so much attention to finding *similar* segments of realizations in order to use these similarities in the predictions of the price development.

Statistical $\mathcal{R}/\mathcal{S}$-analysis proved to be very efficient not only in discovering the above-mentioned phenomena of 'aftereffect', 'long memory', 'persistence', and 'antipersistence', but also in the search of *periodic* or *aperiodic cycles* (see, e.g., [317], [319], [329], [385], [386].)

A classical example of a system where aperiodic cyclicity is clearly visible is *solar activity*.

As is known, there exists a convenient indicator of this activity, the *Wolf numbers* related to the number of 'black spots' on the sun surface. The monthly data for approximately 150 years are available and a mere visual inspection reveals the existence of an 11-year cycle.

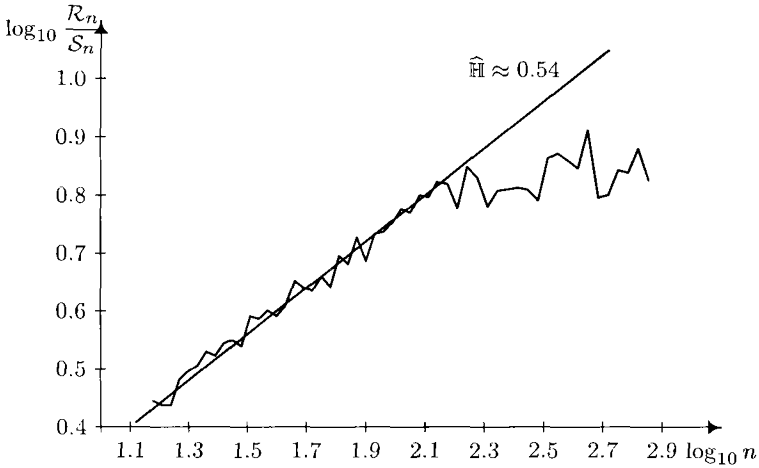


FIGURE 50. $\mathcal{R}/\mathcal{S}$-analysis of the Wolf numbers ($n = 1$ month, $2.12 = \log_{10}(12 \cdot 11)$, $h_n = x_n - x_{n-1}$, x_n are the monthly Wolf numbers)

The results of the $\mathcal{R}/\mathcal{S}$-analysis of the Wolf numbers (see [385; p. 78]) are roughly as shown in Fig. 50.

Estimating the parameter $\mathbb{H}$ we obtain $\widehat{\mathbb{H}} = 0.54$, which points to the tendency of keeping the current level of activity ('persistence'). It is also clearly visible in Fig. 50 that the behavior pattern of $\ln\left(\dfrac{\mathcal{R}_n}{\mathcal{S}_n}\right)$ changes drastically close to 11 years: the values of $\dfrac{\mathcal{R}_n}{\mathcal{S}_n}$ stabilize, which can be explained by a 11-year cycle of solar activity. Indeed, if there are periodic or 'aperiodic' cycles, then the range calculated for the second, third, and the following cycles cannot be much larger than its value obtained for the first cycle. In the same way, the empirical variance usually stabilizes. All this shows that $\mathcal{R}/\mathcal{S}$-analysis is good in discovering cycles in such phenomena as solar activity.

In conclusion we point out that the analysis of statistical dynamical systems usually deals with two kinds of noise: the 'endogenous noise' of the system, which in fact determines its statistical nature (for example, the stochastic nature of solar activity) and the 'outside noise', which is usually an additive noise resulting from measurement errors (occurring, say, in counting 'black spots', when one can take a cluster of small spots for a single big one).

All this taken into account, we must emphasize that $\mathcal{R}/\mathcal{S}$-analysis is robust as concerns 'outside noises'; this additional feature makes it rather efficient in the study of the 'intrinsic' stochastic nature of statistical dynamical systems.

§ 4b. $\mathcal{R}/\mathcal{S}$-Analysis of Some Financial Time Series

1. On gaining a general impression of the $\mathcal{R}/\mathcal{S}$-analysis of fractal, linear, and nonlinear models and its use for the detection of cycles, the idea to apply it to *concrete* financial time series (DJIA, the S&P500 Index, stock and bond prices, currency cross rates) comes naturally.

We have repeatedly pointed out (and we do it also now) that it is extremely important for the analysis of financial data to indicate the time intervals Δ at which investors, traders, ... 'read off' information. We shall indicate this interval Δ explicitly, and, keeping the notation of § 2b, we shall denote by $\widetilde{h}_i$ the variable $h_{i\Delta} = \ln \dfrac{S_{i\Delta}}{S_{(i-1)\Delta}}$, where S_t is the value at time t of the financial index under consideration.

Remark. One can find many data on the applications of $\mathcal{R}/\mathcal{S}$-analysis to time series, including financial ones, in [317], [323], [325], [327], [329] and in the later books [385] and [386]. The brief outline of these results that is presented below follows mainly [385] and [386].

2. DJIA (see Chapter I, § 1b.6; the statistical data are published in *The Wall Street Journal* since 1888). We set Δ to be 1, 5, or 20 days. In the following table we sum up the results of the $\mathcal{R}/\mathcal{S}$-analysis:

Δ	Number of obseravtions N	Estimate $\widehat{\mathbb{H}}_N$	Cycles (in days)
1 day	12 500	0.59	1 250
5 days	2 600	0.61	1 040
20 days	650	0.72	1 040

As regards the behavior of the statistics $\mathcal{V}_n$, one can see that the corresponding values first increase with n. This growth ceases for $n = 52$ (in the case of Δ equal to 20 days; that is, in 1040 days), which indicates the presence of a cyclic component in the data.

For the daily data (Δ is 1 day) the statistics $\mathcal{V}_n$ grows till n is approximately 1250. Thereupon, its behavior stabilizes, which indicates (cf. Fig. 49 in § 4a) a formation of a cycle (with approximately 4-year period; this is usually associated with the 4-year period between presidential elections in the USA).

3. The S&P500 Index (see Chapter I, § 1b.6; § 2d.2 in this chapter) develops in accordance with a 'tick' pattern. There exists a rather comprehensive database for this index. The analysis of monthly data (Δ = 1 month) from January, 1950 through July, 1988 shows ([385; Chapter 8]), that, as in the case of DJIA, there exists an approximately 4-year cycle.

A more refined $\mathcal{R}/\mathcal{S}$-analysis with $\Delta = 1$, 5, and 30 minutes (based on the data from 1989 through 1992; [386; Chapter 9]) delivers the following estimates $\widehat{\mathbb{H}}$ of the Hurst parameter:

$$\widehat{\mathbb{H}} = 0.603,\ 0.590,\ \text{and } 0.653, \tag{1}$$

which indicates a certain 'persistence' in the dynamics of S&P500.

It is worth noting that the transition from the variables $\widetilde{h}_n = \ln \dfrac{\widetilde{S}_n}{\widetilde{S}_{n-1}}$, $\widetilde{S}_n = S_{n\Delta}$, to the 'linearly adjusted' ones,

$$h_n^\circ = \widetilde{h}_n - (a_0 + a_1 \widetilde{h}_{n-1}),$$

results in smaller estimates of $\mathbb{H}$ (cf. (1)). Now,

$$\widehat{\mathbb{H}} = 0.551,\ 0.546,\ \text{and } 0.594,$$

which is close to the mean values $\mathsf{E}_0\widehat{\mathbb{H}}$ (equal to 0.538, 0.540, and 0.563, respectively) calculated under condition $\mathscr{H}_0$ on the basis of the same observations.

All this apparently means that for short intervals of time the behavior of the S&P500 Index can be described in the first approximation by traditional linear models (even by simple ones, such as the $AR(1)$ model).

This can serve an explanation for the wide-spread belief that high-frequency interday data have autoregressive character and for the fact that 'daily' traders make decisions by reacting to the last 'tick' rather then taking into account the 'long memory' and the past prices. The pattern changes, however, with the growth of the time interval Δ determining 'time sharp-sightedness' of the traders. In particular, the $\mathcal{R}/\mathcal{S}$-analysis carried out for Δ equal to, e.g., 1 month clearly reveals a fractal structure; namely, the value $\widehat{\mathbb{H}} \approx 0.78$ calculated on the 48-month basis is fairly large (the data from January, 1963 through December, 1989; [385; Chapter 8]).

We have already mentioned that large values of the Hurst parameter indicate the presence of 'persistence', which can bring about trends and cycles.

In our case, a mere visual analysis of the values of the $\mathcal{V}_n$ clearly indicates (cf. Fig. 49 in § 4a) the presence of a four-year cycle. As in the case of DJIA, this is usually explained by successive economic cycles, related maybe to the presidential elections in the USA.

It is appropriate to recall now (see § 3a) that the statistical $\mathcal{R}/\mathcal{S}$-analysis of the daily values of the variable $\widehat{r}_n = \ln \dfrac{\widehat{\sigma}_n}{\widehat{\sigma}_{n-1}}$, where the empirical dispersion $\widehat{\sigma}_n$ is calculated by formula (8) in § 3a on the basis of the S&P500 Index (from January 1, 1928 through December 31, 1989; [386, Chapter 10]) suggests an approximate value of the Hurst parameter of about 0.31, which is much less than 1/2 and indicates the phenomenon of 'antipersistence'. One visual consequence of this is the fast alternation of the values of $\widehat{r}_n$. In other words, if for an arbitrary n the value of $\widehat{\sigma}_n$ is *larger* than that of $\widehat{\sigma}_{n-1}$, then the value of $\widehat{\sigma}_{n+1}$ is very likely to be *smaller* than $\widehat{\sigma}_n$.

4. $\mathcal{R}/\mathcal{S}$-analysis of stock prices enables one not merely to reveal their fractal structure and discover cycles, but also to compare them from the viewpoint of the *riskiness* of the corresponding stock. According to the data in [385; Chapter 8], the values of the Hurst parameter $\mathbb{H} = \mathbb{H}(\,\cdot\,)$ and the lengths $\mathbb{C} = \mathbb{C}(\,\cdot\,)$ (measured in months) of the cycles for the S&P500 Index and the stock of several corporations included in this index are as follows:

$$
\begin{array}{ll}
\mathbb{H}(\text{S\&P500}) = 0.78, & \mathbb{C}(\text{S\&P500}) = 46; \\
\mathbb{H}(\text{IBM}) = 0.72, & \mathbb{C}(\text{IBM}) = 18; \\
\mathbb{H}(\text{Apple Computer}) = 0.75, & \mathbb{C}(\text{Apple Computer}) = 18; \\
\mathbb{H}(\text{Consolidated Edison}) = 0.68, & \mathbb{C}(\text{Consolidated Edison}) = 0.90.
\end{array}
$$

As we see, the S&P500 Index on its own has Hurst parameter larger than the corporations it covers. We also see that the value of this parameter for, say, Apple Computer is rather large ($\mathbb{H} = 0.75$), considerably larger than the parameter characteristic for, say, Consolidated Edison.

We note further that if $\mathbb{H} = 1$, then a fractional Brownian motion has the representation $B_1(t) = t\xi$, where ξ is a normally distributed random variable with expectation zero and variance one. All the 'randomness' of $B_1 = (B_1(t))_{t\geqslant 0}$ has its source in ξ, so that this is the least 'noisy' of all fractional Brownian motions with $0 < \mathbb{H} \leqslant 1$. Clearly, the noise component of the process $B_{\mathbb{H}}$ 'decreases' as $\mathbb{H} \uparrow 1$. In the financial context this means that models based on these processes become less risky. This observation of the decrease of the noise component can be put in the form of a precise statement: the process $B_{\mathbb{H}}$ converges weakly to B_1 as $\mathbb{H} \uparrow 1$. We note also that the correlation functions $\rho_{\mathbb{H}}(n) = \mathsf{E} h_k h_{n+k}$ are positive for $\mathbb{H} > 1/2$ ('persistence', the property of keeping the trends of dynamics) and $\rho_{\mathbb{H}}(n) \to 1$ for all n as $\mathbb{H} \uparrow 1$.

As noted in [385], such an interpretation seems more promising in cases where the processes in question have a large Hurst parameter $\mathbb{H}$, but do not have variance, which lies at the core of the concept of riskiness.

One reasonable explanation of the fact that the value of $\mathbb{H}$ corresponding to the S&P500 Index is large, so that securities based on this index are less risky than corporate stock, is diversification (see Chapter I, § 2b), which reduces the noise factor.

Remark 1. We emphasize that by 'lesser risks' here we mean 'less noise', 'stronger persistence', showing itself in the tendency to preserve the development trend. It should be pointed out, however, that systems with large $\mathbb{H}$ are prone to *sharp changes of the direction* of evolution. A long series of ups can be followed by a long series of downs.

Remark 2. Turning back to the values of $\mathbb{H}$ and $\mathbb{C}$ for the stock of various corporations, we should also mention the following observation made in [385]: usually, a high level of innovations in a company results in large values of $\mathbb{H}$ and short cycles, while a low level of innovations corresponds to small values of $\mathbb{H}$ and long cycles.

5. Bonds. The $\mathcal{R}/\mathcal{S}$-analysis of the monthly data concerning the American 30-year Treasury Bonds (T-Bonds) over the period from January, 1950 through December, 1989 (see [385; Chapter 8]) also discovers a rather high value of the fractality parameter $\widehat{\mathbb{H}} \approx 0.68$, with cycles of approximately 5 years.

6. Currency cross rates. There exist considerable differences between currency cross rates and such financial indexes as DJIA, the S&P500 Index, or stock and bond prices.

The point is that a purchase or a sale of stock has a direct relation to *investments* in the corresponding industry. The currency is different: it is bought or sold to *facilitate* consumption, the expansion of production, and so on. Moreover, large volumes of exchange influence at least two countries, are determined to a considerable degree by their political and economical situations, and depend strongly on the policy of the central banks (who make interventions in the money markets, change the interest rates, and so on).

Of course, these factors influence the statistical properties of cross rates and their dynamics.

One important characteristic of the changeability of cross rates is the Δ-volatility defined in terms of the increments $|H_{k\Delta} - H_{(k-1)\Delta}|$, where $H_t = \ln \dfrac{S_t}{S_0}$ for the cross rate S_t (see, for example, formula (5) in § 1c).

We must emphasize in this connection that the above-discussed statistics $\mathcal{R}_n$ and $\dfrac{\mathcal{R}_n}{\mathcal{S}_n}$ are *also*, in effect, characteristics of the changeability, 'range' of the process $H = (H_t)_{t \geqslant 0}$. It is therefore not surprising that many properties of exchange rates already described in the previous sections can be also detected by $\mathcal{R}/\mathcal{S}$-analysis.

By contrast to such financial indexes as DJIA or the S&P500, the fractal structure (at any rate, for small values of $\Delta > 0$) and the tendency towards its preservation in time are clearly visible in the evolution of the exchange rates.

This is revealed by a mere analysis of the behavior of the statistics $\ln \dfrac{\mathcal{R}_n}{\mathcal{S}_n}$ as functions of $\log n$ by the least squares method. This analysis shows that the values of $\ln \dfrac{R_n}{S_n}$, $n \geqslant 1$, distinctly cluster along the line $c + \widehat{\mathbb{H}} \ln n$, with $\widehat{\mathbb{H}}$ considerably larger than $1/2$ for most currencies. For instance, $\widehat{\mathbb{H}}$(JRY/USD) ≈ 0.64, $\widehat{\mathbb{H}}$(DEM/USD) ≈ 0.64, $\widehat{\mathbb{H}}$(GBP/USD) ≈ 0.61.

All this means that currency exchange rates have fractal structures with rather large Hurst parameters. It is worth recalling in this connection that, in the case of a fractional Brownian motion, $\mathsf{E}H_t^2$ grows as $|t|^{2\mathbb{H}}$. Hence for $\mathbb{H} > 1/2$ the dispersion of the values of $|H_t|$ is larger than for a standard Brownian motion, which indicates that the riskiness of exchange grows with time. Arguably, this can explain why high-intensity short-term trading is more popular in the currencies market than long-term operations.

Part 2

THEORY

Chapter V. Theory of Arbitrage in Stochastic Financial Models. Discrete Time

1. Investment Portfolio on a (B, S)-Market

§ 1a. Strategies Satisfying Balance Conditions

1. We assume that the securities market under consideration operates in the conditions of 'uncertainty' that can be described in the probabilistic framework in terms of a *filtered probability space*

$$(\Omega, \mathscr{F}, (\mathscr{F}_n)_{n\geqslant 0}, \mathsf{P}).$$

We interpret the flow $\mathbb{F} = (\mathscr{F}_n)_{n\geqslant 0}$ of σ-algebras as the 'flow of information' $\mathscr{F}_n$ accessible to all participants up to the instant n, $n \geqslant 0$.

We shall consider a (B, S)-market formed (by definition) by $d+1$ assets:

a bank account B (a *'risk-free' asset*)

and

stock ('risk' assets) $S = (S^1, \dots, S^d)$.

We assume that the evolution of the bank account can be described by a positive stochastic sequence

$$B = (B_n)_{n\geqslant 0},$$

where the variables B_n are $\mathscr{F}_{n-1}$-measurable for each $n \geqslant 1$.

The dynamics of the (value) of the ith risk asset S^i can also be described by a positive stochastic sequence

$$S^i = (S^i_n)_{n\geqslant 0},$$

where the S^i_n are $\mathscr{F}_n$-measurable variables for each $n \geqslant 0$.

From these definitions one can clearly see the crucial difference between a bank account and stock. Namely, the $\mathscr{F}_{n-1}$-measurability of B_n means that the state of the bank account at time n is already *clear* (provided that one has all information) *at time* $n-1$: the variable B_n is *predictable* (in this sense).

The situation with stock prices is entirely different: the variables S_n^i are $\mathscr{F}_n$-measurable, which means that their actual values *are known only after one obtains* all 'information' $\mathscr{F}_n$ arriving at time n.

This explains why one says that a bank account is 'risk-free' while stocks are 'risk' assets.

Setting

$$r_n = \frac{\Delta B_n}{B_{n-1}}, \qquad \rho_n^i = \frac{\Delta S_n^i}{S_{n-1}^i}, \tag{1}$$

we can write

$$\Delta B_n = r_n B_{n-1}, \tag{2}$$

$$\Delta S_n^i = \rho_n^i S_{n-1}^i, \tag{3}$$

where the (interest rates) r_n are $\mathscr{F}_{n-1}$-measurable and the ρ_n^i are $\mathscr{F}_n$-measurable.

Thus, for $n \geqslant 1$ we have

$$B_n = B_0 \prod_{1 \leqslant k \leqslant n} (1 + r_k) \tag{4}$$

and

$$S_n^i = S_0^i \prod_{1 \leqslant k \leqslant n} (1 + \rho_k^i). \tag{5}$$

In accordance with our terminology in Chapter II, § 1a, we say that the representations (4) and (5) are of 'simple interest' kind.

2. Imagine an investor on the (B, S)-market that can

a) *deposit money into the bank account and borrow from it*;

b) *sell and buy stock.*

We shall assume that a transfer of money from one asset into another can be done with no *transaction* costs and the assets are '*infinitely divisible*', i.e., the investor can buy or sell any portion of stock and withdraw or deposit any amount from the bank account.

We now give several definitions relating to the financial position of an investor in such a (B, S)-market and his actions.

DEFINITION 1. A (predictable) stochastic sequence

$$\pi = (\beta, \gamma)$$

where $\beta = (\beta_n(\omega))_{n\geqslant 0}$ and $\gamma = (\gamma_n^1(\omega), \dots, \gamma_n^d(\omega))_{n\geqslant 0}$ with $\mathscr{F}_{n-1}$-measurable $\beta_n(\omega)$ and $\gamma_n^i(\omega)$ for all $n \geqslant 0$ and $i = 1, \dots, d$ (we set $\mathscr{F}_{-1} = \mathscr{F}_0$) is called an investment *portfolio* on the (B, S)-market.

If $d = 1$, then we shall write γ_n and S_n in place of γ_n^1 and S_n^1.

We must now emphasize several important points.

The variables $\beta_n(\omega)$ and $\gamma_n^i(\omega)$ can be positive, equal to zero, or even *negative*, which means that the investor, in conformity with a) and b), can *borrow* from the bank account or *sell* stock *short*.

The assumption of $\mathscr{F}_{n-1}$-measurability means that the variables $\beta_n(\omega)$ and $\gamma_n^i(\omega)$ describing the financial position of the investor at time n (the amount he has on the bank account, the stock in his possession) are determined by the information available at time $n-1$, not n (the 'tomorrow' position is completely defined by the 'today' situation).

Time $n = 0$ plays a special role here (as in the entire theory of stochastic processes based on the concept of filtered probability spaces). This is reflected by the fact that the predictability at the instant $n = 0$ (formally, this must be equivalent to '$\mathscr{F}_{-1}$-measurability') is the same as $\mathscr{F}_0$-measurability. (The agreement $\mathscr{F}_{-1} = \mathscr{F}_0$ in the definition is convenient for a uniform approach to all instants $n \geqslant 0$.)

To emphasize the dynamics of an investment portfolio one often uses the term 'investment strategy' instead. We shall also use it sometimes.

We have assumed above that $n \geqslant 0$. Of course, all the definitions will be the same if we are bounded by some finite 'time horizon' N. In this case we shall assume that $0 \leqslant n \leqslant N$ in place of $n \geqslant 0$.

DEFINITION 2. The *value of an investment portfolio* π is the stochastic sequence

$$X^\pi = (X_n^\pi)_{n\geqslant 0},$$

where

$$X_n^\pi = \beta_n B_n + \sum_{i=1}^d \gamma_n^i S_n^i. \tag{6}$$

To avoid lengthy formulas we shall use 'coordinate-free' notation in what follows, denoting the scalar product

$$(\gamma_n, S_n) \equiv \sum_{i=1}^d \gamma_n^i S_n^i$$

of *vectors* $\gamma_n = (\gamma_n^1, \dots, \gamma_n^d)$ and $S_n = (S_n^1, \dots, S_n^d)$ by $\gamma_n S_n$.

Thus, let

$$X_n^\pi = \beta_n B_n + \gamma_n S_n. \tag{7}$$

For two arbitrary sequences $a = (a_n)_{n\geqslant 0}$ and $b = (b_n)_{n\geqslant 0}$ we have

$$\Delta(a_n b_n) = a_n \Delta b_n + b_{n-1}\Delta a_n. \tag{8}$$

Applying this formula (of 'discrete differentiation') to the right-hand side of (7) we see that

$$\Delta X_n^\pi = [\beta_n \Delta B_n + \gamma_n \Delta S_n] + [B_{n-1}\Delta\beta_n + S_{n-1}\Delta\gamma_n]. \tag{9}$$

This shows that changes $(\Delta X_n^\pi \equiv X_n^\pi - X_{n-1}^\pi)$ in the value of a portfolio are, in general, sums of two components: changes of the state of a unit bank account and of stock prices $(\beta_n \Delta B_n + \gamma_n \Delta S_n)$ and changes in the portfolio composition $(B_{n-1}\Delta\beta_n + S_{n-1}\Delta\gamma_n)$. It is reasonable to assume that *real* changes of this value are always due to the increments ΔB_n and ΔS_n (and not to $\Delta\beta_n$ and $\Delta\gamma_n$).

Thus, we conclude that the capital gains on the investment portfolio π are described by the sequence $G^\pi = (G_n^\pi)_{n\geqslant 0}$, where $G_0^\pi = 0$ and

$$G_n^\pi = \sum_{k=1}^n (\beta_k \Delta B_k + \gamma_k \Delta S_k). \tag{10}$$

Hence the value of the portfolio at time n is

$$X_n^\pi = X_0^\pi + G_n^\pi, \tag{11}$$

which brings us in a natural way to the following definition.

DEFINITION 3. We say that an investment portfolio π is *self-financing* if its value $X^\pi = (X_n^\pi)_{n\geqslant 0}$ can be represented as the following sum:

$$X_n^\pi = X_0^\pi + \sum_{k=1}^n (\beta_k \Delta B_k + \gamma_k \Delta S_k), \qquad n \geqslant 1. \tag{12}$$

That is, self-financing here is equivalent to the following condition describing 'admissible' portfolios π:

$$B_{n-1}\Delta\beta_n + S_{n-1}\Delta\gamma_n = 0, \qquad n \geqslant 1. \tag{13}$$

The message is perfectly clear: the change of the amount on the bank account can be only due to the change of the *package of shares* and vice versa.

We denote the class of self-financing strategies π by SF.

We note also that we have shown the equivalence of relations

$$(6) + (13) \iff (6) + (12).$$

3. It is perfectly clear that in the operations with the portfolio π one would better reduce the number of different assets in it or, at least, simplify their structure. In accordance with one possible (and commonly used here) approach, 'if we *a priori* have $B_n > 0$, $n \geqslant 1$, then we can set $B_n \equiv 1$'. This is a consequence of the following observation.

Along with the (B, S)-market we can consider a new market $(\widetilde{B}, \widetilde{S})$, where

$$\widetilde{B} = (\widetilde{B}_n)_{n\geqslant 0} \quad \text{with} \quad \widetilde{B}_n \equiv 1$$

and

$$\widetilde{S} = (\widetilde{S}_n)_{n\geqslant 0} \quad \text{with} \quad \widetilde{S}_n = \frac{S_n}{B_n}.$$

Then the value $\widetilde{X}^\pi = (\widetilde{X}^\pi_n)_{n\geqslant 0}$ of the portfolio $\pi = (\beta, \gamma)$ is as follows:

$$\widetilde{X}^\pi = \beta_n \widetilde{B}_n + \gamma_n \widetilde{S}_n = \beta_n + \gamma_n \widetilde{S}_n = \frac{1}{B_n}(\beta_n B_n + \gamma_n S_n) = \frac{X^\pi_n}{B_n}. \tag{14}$$

In addition, if π is self-financing in the (B, S)-market, then it has this property also on the $(\widetilde{B}, \widetilde{S})$-market, for (see (13))

$$\widetilde{B}_{n-1}\Delta\beta_n + \widetilde{S}_{n-1}\Delta\gamma_n = \frac{1}{B_{n-1}}(B_{n-1}\Delta\beta_n + S_{n-1}\Delta\gamma_n) = 0. \tag{15}$$

Since $\Delta\widetilde{B}_k \equiv 0$, it follows by (12) that

$$\widetilde{X}^\pi_n = \widetilde{X}^\pi_0 + \sum_{k=1}^n \gamma_k \Delta\widetilde{S}_k \tag{16}$$

for $\pi \in SF$, or, more explicitly,

$$\widetilde{X}^\pi_n = \widetilde{X}^\pi_0 + \sum_{k=1}^n \left[\sum_{i=1}^d \gamma^i_k \Delta\widetilde{S}^i_k\right], \qquad \widetilde{S}^i_k = \frac{S^i_k}{B_k}. \tag{17}$$

Thus, from (14) and (16) we see that the discounted value $\dfrac{X^\pi}{B} = \left(\dfrac{X^\pi_n}{B_n}\right)_{n\geqslant 0}$ of $\pi \in SF$ satisfies the relation

$$\Delta\left(\frac{X^\pi_n}{B_n}\right) = \gamma_n \Delta\left(\frac{S_n}{B_n}\right), \tag{18}$$

which, for all its simplicity, plays a key role in many calculations below based on the concept of 'arbitrage-free market'.

It follows from the above that

$$(6) + (12) \implies (6) + (18).$$

It is easy to verify that the reverse implication also holds. Thus, we have the following relation, which can be useful in the construction of self-financing strategies and in the inspection of whether a particular strategy is self-financing:

$$(6) + (13) \iff (6) + (12) \iff (6) + (18).$$

We note also that (18) is in a certain sense more consistent from the financial point of view than the equality

$$\Delta X_n^\pi = \beta_n \Delta B_n + \gamma_n \Delta S_n, \tag{19}$$

following from (12). Indeed, *comparing* prices one is more interested in their relative values, than in the absolute ones. (We have already mentioned this in Chapter I, § 2a.4.) This explains why we consider the discounted variables $\widetilde{B} = \dfrac{B}{B}$ $(\equiv 1)$ and $\widetilde{S} = \left(\dfrac{S}{B}\right)$ in place of B and S.

Our choice of the bank account as a *discount* factor is a mere convention; we could take any of the assets $S^1, \dots, S^d$ or their combination instead. However, the fact that the bank account is a 'predictable' asset simplifies the analysis to a certain extent. For instance, this 'predictability' means that the distribution $\operatorname{Law}\left(\dfrac{X_n^\pi}{B_n} \;\Big|\; \mathscr{F}_{n-1}\right)$ can be determined from the conditional distribution $\operatorname{Law}(X_n^\pi \mid \mathscr{F}_{n-1})$, because the 'condition' $\mathscr{F}_{n-1}$ means that the variable B_n is known. Moreover, the bank account plays the role of a 'reference point' in economics, a 'standard' in the valuation of other securities, which display the same behavior 'on the average'.

4. The above-discussed evolution of the capital X^π in a (B, S)-market relates to the case when there are no 'inflows or outflows of funds' and the 'transaction costs' are negligible. Of course, we can think of other schemes, where the change of capital ΔX_n^π does not proceed in accordance with (19), but has a more complicated form, where *shareholder dividends*, *consumption*, *transaction costs*, etc. are taken into account. We consider now several examples of this kind.

The case of 'dividends'. Let $D = (D_n = (D_n^1, \dots, D_n^d))_{n \geqslant 0}$ be a d-dimensional sequence of $\mathscr{F}_n$-measurable variables D_n^i, $D_0^i = 0$, and assume that $\delta_n^i \equiv \Delta D_n^i \geqslant 0$. We shall treat D_n^i as the *dividend income* generated by one share S_n^i. Then one has a right to assume that the change of the portfolio value $X^\pi = (X_n^\pi)_{n \geqslant 0}$ can be described by the formula

$$\Delta X_n^\pi = \beta_n \Delta B_n + \gamma_n (\Delta S_n + \Delta D_n), \tag{20}$$

while X_n^π is itself the following combination of the bank account, stock prices, and dividends:

$$X_n^\pi = \beta_n B_n + \gamma_n (S_n + \Delta D_n). \tag{21}$$

It is easy to obtain by (20) and (21) the following, suitable for this case, modification of condition (13) of self-financing:

$$B_{n-1}\Delta\beta_n + S_{n-1}\Delta\gamma_n - \delta_{n-1}\gamma_{n-1} = 0. \tag{22}$$

Moreover,

$$(21) + (22) \iff (21) + (20), \tag{23}$$

and (18) can be generalized as follows:

$$\Delta\left(\frac{X_n^\pi}{B_n}\right) = \gamma_n \left[\Delta\left(\frac{S_n}{B_n}\right) + \frac{\Delta D_n}{B_n}\right]. \tag{24}$$

The case of 'consumption and investment'. We assume that $C = (C_n)_{n\geqslant 0}$ and $I = (I_n)_{n\geqslant 0}$ are nonnegative nondecreasing ($\Delta C_n \geqslant 0$, $\Delta I_n \geqslant 0$) processes with $C_0 = 0$, $I_0 = 0$, and with $\mathscr{F}_n$-measurable C_n and I_n.

Assume that the evolution of the portfolio value $X^\pi = (X_n^\pi)_{n\geqslant 0}$ is described by the formula

$$\Delta X_n^\pi = \beta_n \Delta B_n + \gamma_n \Delta S_n + \Delta I_n - \Delta C_n. \tag{25}$$

This explains why we call this the case with 'consumption and investment'. The process $C = (C_n)_{n\geqslant 0}$ is called the 'consumption process' and $I = (I_n)_{n\geqslant 0}$ is called the 'investment process'.

Clearly, if we write $X_n^\pi = \beta_n B_n + \gamma_n S_n$, then (25) brings us to the following 'admissibility' constraint on the components of π:

$$B_{n-1}\Delta\beta_n + S_{n-1}\Delta\gamma_n = \Delta I_n - \Delta C_n. \tag{26}$$

Relation (18) has now the following generalization:

$$\Delta\left(\frac{X_n^\pi}{B_n}\right) = \gamma_n \Delta\left(\frac{S_n}{B_n}\right) + \frac{\Delta(I_n - C_n)}{B_{n-1}}. \tag{27}$$

Combining both cases (with 'dividends' and with 'consumption and investment') we see that for an 'admissible' (i.e., satisfying (26)) portfolio π we have

$$\Delta\left(\frac{X_n^\pi}{B_n}\right) = \gamma_n \left[\Delta\left(\frac{S_n}{B_n}\right) + \frac{\Delta D_n}{B_n}\right] + \frac{\Delta(I_n - C_n)}{B_{n-1}}. \tag{28}$$

The case of 'operating expense' (of stock trading). Relation (13) considered above has a clear financial meaning of a *budgetary, balance* restriction. It shows that if, e.g., $\Delta\gamma_n > 0$, then buying stock of (unit) price S_{n-1} we must withdraw from the bank account the amount $B_{n-1}\Delta\beta_n$ equal to $S_{n-1}\Delta\gamma_n$. On the other hand, if $\Delta\gamma_n < 0$ (i.e., we sell stock), then the amount put into the bank account is $B_{n-1}\Delta\beta_n = -S_{n-1}\Delta\gamma_n > 0$.

Imagine now that each stock transaction involves some operating costs. Then, purchasing $\Delta\gamma_n > 0$ shares of share price S_{n-1}, it is no longer sufficient, due to transaction costs, that we withdraw $S_{n-1}\Delta\gamma_n$ from the bank account; we must withdraw instead an amount $(1+\lambda)S_{n-1}\Delta\gamma_n$ with $\lambda > 0$.

On the other hand, selling stock ($\Delta\gamma_n < 0$), we cannot deposit all the amount of $-S_{n-1}\Delta\gamma_n$ into our bank account, but only a smaller sum, say, $-(1-\mu)S_{n-1}\Delta\gamma_n$ with $\mu > 0$.

It is now clear that, with non-zero transaction costs, we must replace balance condition (13) imposed on the portfolio π by the following condition of *admissibility*:

$$B_{n-1}\Delta\beta_n + (1+\lambda)S_{n-1}\Delta\gamma_n I(\Delta\gamma_n > 0) + (1-\mu)S_{n-1}\Delta\gamma_n I(\Delta\gamma_n < 0) = 0, \tag{29}$$

which can be rewritten as

$$B_{n-1}\Delta\beta_n + S_{n-1}\Delta\gamma_n + \lambda S_{n-1}(\Delta\gamma_n)^+ + \mu S_{n-1}(\Delta\gamma_n)^- = 0, \tag{30}$$

where $(\Delta\gamma_n)^+ = \max(0,\Delta\gamma_n)$ and $(\Delta\gamma_n)^- = -\min(0,\Delta\gamma_n)$.

If $\lambda = \mu$, then (30) takes the form

$$B_{n-1}\Delta\beta_n + S_{n-1}\big[\Delta\gamma_n + \lambda|\Delta\gamma_n|\big] = 0. \tag{31}$$

To find a counterpart to (18) we observe that, in view of (30),

$$\begin{aligned}
\Delta\left(\frac{X_n^\pi}{B_n}\right) &= \frac{X_n^\pi}{B_n} - \frac{X_{n-1}^\pi}{B_{n-1}} \\
&= \frac{\beta_n B_n + \gamma_n S_n}{B_n} - \frac{\beta_{n-1}B_{n-1} + \gamma_{n-1}S_{n-1}}{B_{n-1}} \\
&= \Delta\beta_n + \gamma_n\Delta\left(\frac{S_n}{B_n}\right) + \frac{S_{n-1}}{B_{n-1}}\Delta\gamma_n \\
&= \frac{B_{n-1}\Delta\beta_n + S_{n-1}\Delta\gamma_n}{B_{n-1}} + \gamma_n\Delta\left(\frac{S_n}{B_n}\right) \\
&= -\frac{S_{n-1}}{B_{n-1}}\Big(\mu(\Delta\gamma_n)^- + \lambda(\Delta\gamma_n)^+\Big) + \gamma_n\Delta\left(\frac{S_n}{B_n}\right).
\end{aligned}$$

Hence, if the transaction costs specified by the parameters λ and μ are non-zero and the balance requirement (29) holds, then the evolution of the discounted portfolio value $\dfrac{X_n^\pi}{B_n}$ with $X_n^\pi = \beta_n B_n + \gamma_n S_n$ can be described by the relation

$$\Delta\left(\frac{X_n^\pi}{B_n}\right) = \gamma_n\Delta\left(\frac{S_n}{B_n}\right) - \frac{S_{n-1}}{B_{n-1}}\big[\lambda(\Delta\gamma_n)^+ + \mu(\Delta\gamma_n)^-\big]. \tag{32}$$

If $\lambda = \mu$, then

$$\Delta\left(\frac{X_n^\pi}{B_n}\right) = \gamma_n \Delta\left(\frac{S_n}{B_n}\right) - \lambda \frac{S_{n-1}}{B_{n-1}} |\Delta\gamma_n|. \tag{33}$$

We can also regard this problem of transaction costs from another, equivalent, standpoint.

Let $\widehat{S}_n = \gamma_n S_n$ and $\widehat{B}_n = \beta_n B_n$ be the funds invested in stock and deposited in the bank account, respectively, for the portfolio π at time n.

Assume that

$$\begin{aligned} \Delta\widehat{S}_n &= \rho\widehat{S}_{n-1} + \Delta L_n - \Delta M_n, \\ \Delta\widehat{B}_n &= r\widehat{B}_{n-1} - (1+\lambda)\Delta L_n + (1-\mu)\Delta M_n, \end{aligned} \tag{34}$$

where L_n is the *cumulative* transfer of funds from the bank account into stock by time n, which has required (due to transaction costs) a withdrawal of the (larger) amount $(1+\lambda)L_n$. In a similar way, let M_n be the cumulative amount obtained for the stock sold, and let the smaller amount $(1-\mu)\Delta M_n$ (again, due to the transaction costs) be deposited into the account.

It is clear from (34) that we have the following relation between the total capital assets $X_n^\pi = \widehat{B}_n + \widehat{S}_n$, the strategy π, and the transfers (L, M), where $L = (L_n)$, $M = (M_n)$, and $L_0 = M_0 = 0$:

$$\begin{aligned} \Delta X_n^\pi &= r\widehat{B}_{n-1} + \rho\widehat{S}_{n-1} - (\lambda\Delta L_n + \mu\Delta M_n) \\ &= r\beta_{n-1}B_{n-1} + \rho\gamma_{n-1}S_{n-1} - (\lambda\Delta L_n + \mu\Delta M_n) \\ &= \beta_{n-1}\Delta B_n + \gamma_{n-1}\Delta S_n - (\lambda\Delta L_n + \mu\Delta M_n). \end{aligned}$$

However,

$$\Delta X_n^\pi = \beta_{n-1}\Delta B_n + \gamma_{n-1}\Delta S_n + B_{n-1}\Delta\beta_n + S_{n-1}\Delta\gamma_n.$$

Thus, we see that

$$B_{n-1}\Delta\beta_n + S_{n-1}\Delta\gamma_n + \lambda\Delta L_n + \mu\Delta M_n = 0. \tag{35}$$

Setting here

$$\Delta L_n = S_{n-1}(\Delta\gamma_n)^+, \qquad \Delta M_n = S_{n-1}(\Delta\gamma_n)^-,$$

we obtain that (35) is equivalent to (30).

The general case. We can combine these four cases into one, the case where there are dividends, consumption, investment, and transaction costs.

Namely, using the above notation we shall assume that the evolution of the portfolio value

$$X_n^\pi = \beta_n B_n + \gamma_n (S_n + \Delta D_n) \tag{36}$$

can be described by the formula

$$\Delta X_n^\pi = \beta_n \Delta B_n + \gamma_n(\Delta S_n + \Delta D_n) + \Delta I_n \\ - \Delta C_n - S_{n-1}\big[\lambda(\Delta\gamma_n)^+ + \mu(\Delta\gamma_n)^-\big]. \tag{37}$$

By (36) and (37) we obtain the following balance condition:

$$B_{n-1}\Delta\beta_n + S_{n-1}\Delta\gamma_n \\ = -\big[\Delta C_n - \Delta I_n + \delta_{n-1}\gamma_{n-1} + \mu S_{n-1}(\Delta\gamma_n)^- + \lambda S_{n-1}(\Delta\gamma_n)^+\big], \tag{38}$$

which describes admissible π.

Since

$$\Delta\left(\frac{X_n^\pi}{B_n}\right) = \frac{B_{n-1}\Delta\beta_n + S_{n-1}\Delta\gamma_n}{B_{n-1}} + \frac{\gamma_n\delta_n}{B_n} - \frac{\gamma_{n-1}\delta_{n-1}}{B_{n-1}} + \gamma_n\Delta\left(\frac{S_n}{B_n}\right),$$

it follows in view of (38) that

$$\Delta\left(\frac{X_n^\pi}{B_n}\right) = \gamma_n\left[\Delta\left(\frac{S_n}{B_n}\right) + \frac{\Delta D_n}{B_n}\right] + \frac{\Delta(I_n - C_n)}{B_{n-1}} \\ - \frac{S_{n-1}}{B_{n-1}}\big[\lambda(\Delta\gamma_n)^+ + \mu(\Delta\gamma_n)^-\big] \tag{39}$$

for an admissible portfolio. If $\lambda = \mu$, then

$$\Delta\left(\frac{X_n^\pi}{B_n}\right) = \gamma_n\left[\Delta\left(\frac{S_n}{B_n}\right) + \frac{\Delta D_n}{B_n}\right] + \frac{\Delta(I_n - C_n)}{B_{n-1}} - \frac{\lambda S_{n-1}|\Delta\gamma_n|}{B_{n-1}}. \tag{40}$$

5. We now turn to formulas (18) and (27) and assume that the sequence $\dfrac{S}{B} = \left(\dfrac{S_n}{B_n}\right)_{n\geqslant 1}$ is a (d-dimensional) *martingale* with respect to the basic flow $\mathbb{F} = (\mathscr{F}_n)_{n\geqslant 0}$ and the probability measure P.

By (18),

$$\frac{X_n^\pi}{B_n} = \frac{X_0^\pi}{B_0} + \sum_{k=1}^{n} \gamma_k \Delta\left(\frac{S_k}{B_k}\right), \tag{41}$$

so that (since the γ_k, $k \geqslant 1$, are predictable) the sequence

$$\left(\frac{X_n^\pi}{B_n} - \frac{X_0^\pi}{B_0}\right)_{n\geqslant 0} \tag{42}$$

is a *martingale transform* and, therefore (see Chapter II, § 1c), a *local martingale.*

We shall assume that X_0^π and B_0 are constants. Then for each portfolio π such that

$$\inf_n \frac{X_n^\pi}{B_n} \geqslant C > -\infty, \tag{43}$$

where C is some constant, we see that

$$\sum_{k=1}^{n} \gamma_k \Delta\left(\frac{S_k}{B_k}\right) \geqslant C - \frac{X_0^\pi}{B_0} \tag{44}$$

for each $n \geqslant 1$, i.e., the local martingale

$$\left(\sum_{k=1}^{n} \gamma_k \Delta\left(\frac{S_k}{B_k}\right)\right)_{n \geqslant 1}$$

is bounded below and, by a lemma in Chapter II, § 1c is a *martingale.*

Thus, if the discounted value of π satisfies (43) (for brevity we shall write $\pi \in \Pi_C$), then this discounted portfolio value $\left(\dfrac{X_n^\pi}{B_n}\right)_{n \geqslant 0}$ is a martingale and, in particular,

$$\mathsf{E}\,\frac{X_n^\pi}{B_n} = \frac{X_0^\pi}{B_0} \tag{45}$$

for each $n \geqslant 0$.

We can make the following observation concerning this property.

Assume that for some (B, S)-market the sequence $\dfrac{S}{B} = \left(\dfrac{S_n}{B_n}\right)_{n \geqslant 0}$ is a martingale. (Using a term from Chapter I, § 2a, this market is *efficient*). Then there is no strategy $\pi \in \Pi_C$ (i.e., satisfying condition (43) with some constant C) that would enable an investor (on an *efficient* market) to attain a *larger* mean discounted portfolio value $\mathsf{E}\,\dfrac{X_n^\pi}{B_n}$ than the initial discounted value $\dfrac{X_0^\pi}{B_0}$.

An economic explanation is that the strategies $\pi \in \Pi_C$ are not sufficiently *risky*: an investor has no opportunity of 'running into large debts' (which would mean that he has an unlimited credit).

In the same way, if $\pi \in \Pi^C$, i.e.,

$$\sup_n \frac{X_n^\pi}{B_n} \leqslant C < \infty \quad (\mathsf{P}\text{-a.s.}) \tag{46}$$

for some C, then the same lemma (in Chapter II, § 1e) shows that the sequence $\left(\dfrac{X_n^\pi}{B_n}\right)_{n \geqslant 0}$ is a martingale again. and therefore (45) holds (for an *efficient* market).

It is interesting in many problems of financial mathematics whether property (45) holds after a replacement of the deterministic instant n by a Markov time $\tau = \tau(\omega)$. This is important, e.g., for American options (see Chapter VI), where a stopping time, the time of taking a decision of, say, exercising the option, is already included in the notion of a strategy.

It can be immediately observed that if a stochastic sequence $Y = (Y_n, \mathscr{F}_n)$ is a martingale, then $\mathsf{E}(Y_n \mid \mathscr{F}_{n-1}) = Y_{n-1}$ and $\mathsf{E}|Y_n| < \infty$, so that $\mathsf{E}Y_n = \mathsf{E}Y_0$, while the property

$$\mathsf{E}Y_\tau = \mathsf{E}Y_0, \tag{47}$$

does not necessarily hold for a random stopping time $\tau = \tau(\omega) < \infty$, $\omega \in \Omega$ (cf. Chapter III, § 3a, where we consider the case of a Brownian motion). It holds if, e.g., τ is a bounded stopping time ($\tau(\omega) \leqslant N < \infty$, $\omega \in \Omega$). As regards *various sufficient conditions*, we can refer to § 3a in this chapter and to Chapter VII, § 2 in [439], where one can find several criterions of property (47). (In a few words, these criterions indicate that one should not allow 'very large' τ and/or $\mathsf{E}|Y_\tau|$ if (47) is to be satisfied.) We present an example where (47) fails in § 2b below.

6. In the preceding discussion we were operating an investment portfolio with no restrictions on the possible values of the β_n and $\gamma_n = (\gamma_n^1, \dots, \gamma_n^d)$; we set $\beta_n \in \mathbb{R}$ and $\gamma_n^i \in \mathbb{R}$.

In practice, however, there can be *various sorts of constraints.* For instance, the condition $\beta_n \geqslant 0$ means that we cannot borrow from the bank account at time n. If $\gamma_n \geqslant 0$, then short-selling is not allowed.

If we impose the condition

$$\frac{\gamma_n S_n}{X_n^\pi} \leqslant \alpha,$$

where α is a fixed constant, then the proportion of the 'risky' component $\gamma_n S_n$ in the total capital X_n^π should not be larger than α.

Other examples of constraints on π are conditions of the types

$$\mathsf{P}(X_N^\pi \in A) \geqslant 1 - \varepsilon$$

(for some fixed N, $\varepsilon > 0$, and some set A) or

$$\mathsf{P}(X_N^\pi \geqslant f_N) = 1$$

for some $\mathscr{F}_N$-measurable functions $f_N = f_N(\omega)$.

We consider the last case thoroughly in the next section, in connection with *hedging*.

§ 1b. Notion of 'Hedging'. Upper and Lower Prices. Complete and Incomplete Markets.

1. We shall assume that transactions in our (B, S)-market are made only at the instants $n = 0, 1, \dots, N$.

Let $f_N = f_N(\omega)$ be a non-negative $\mathscr{F}_N$-measurable function (*performance*) treated as an 'obligation', 'terminal' pay-off.

DEFINITION 1. An investment portfolio $\pi = (\beta, \gamma)$ with $\beta = (\beta_n)$, $\gamma = (\gamma_n)$, $n = 0, 1, \dots, N$, is called an *upper* (x, f_N)*-hedge* (a *lower* (x, f_N)*-hedge*) for $x \geqslant 0$ if $X_0^\pi \equiv x$ and $X_N^\pi \geqslant f_N$ (P-a.s.) (respectively, $X_N^\pi \leqslant f_N$ (P-a.s.)).

We say that an (x, f_N)-hedge π is *perfect* if $X_0^\pi \equiv x$, $x \geqslant 0$, and $X_N^\pi = f_N$ (P-a.s.).

The concept of *hedge* plays an extremely important role in financial mathematics and practice. This is an instrument of *protection* enabling one to have a guaranteed level of capital and *insuring* transactions on securities markets.

A definition below will help us formalize actions aimed at securing a certain level of capital.

2. Let
$$H^*(x, f_N; \mathsf{P}) = \{\pi : X_0^\pi = x,\ X_N^\pi \geqslant f_N \ (\mathsf{P}\text{-a.s.})\}$$
be the class of upper (x, f_N)-hedges and let
$$H_*(x, f_N; \mathsf{P}) = \{\pi : X_0^\pi = x,\ X_N^\pi \leqslant f_N \ (\mathsf{P}\text{-a.s.})\}$$
be the class of lower (x, f_N)-hedges.

DEFINITION 2. Let f_N be a pay-off function. Then we call the quantity
$$\mathbb{C}^*(f_N; \mathsf{P}) = \inf\{x \geqslant 0 : H^*(x, f_N; \mathsf{P}) \neq \varnothing\}$$
the *upper price* (of hedging against f_N).

The quantity
$$\mathbb{C}_*(f_N; \mathsf{P}) = \sup\{x \geqslant 0 : H_*(x, f_N; \mathsf{P}) \neq \varnothing\}$$
is called the *lower price* (of hedging against f_N).

Remark 1. As usual, we put $\mathbb{C}^*(f_N; \mathsf{P}) = \infty$ if $H^*(x, f_N; \mathsf{P}) = \varnothing$ for all $x \geqslant 0$. The set $H_*(0, f_N; \mathsf{P})$ is nonempty (it suffices to consider $\beta_n \equiv 0$ and $\gamma_n \equiv 0$). If $H_*(x, f_N; \mathsf{P}) \neq \varnothing$ for all $x \geqslant 0$, then $\mathbb{C}_*(f_N; \mathsf{P}) = \infty$.

Remark 2. We mention no 'balance' or other constraints on the investment portfolio in the above definitions. One standard constraint of this kind is the condition of '*self-financing*' (see (13) in the previous section). Of course, one must specify the requirements imposed on '*admissible*' strategies π in one's considerations of particular cases.

Remark 3. It is possible that the classes $H^*(x, f_N; \mathsf{P})$ and $H_*(x, f_N; \mathsf{P})$ are empty (for some x at any rate). In this case, one reasonable method for comparing the quality of different portfolios is to use the mean quadratic deviation

$$\mathsf{E}[X_N^\pi - f_N]^2$$

(see Chapter IV, § 1d)

3. We now discuss the concepts of 'upper' and 'lower' prices.

If you are *selling* some contract (with pay-off function f_N), then you surely wish to sell it at a high price. However, you must be aware that the buyer wants to buy a *secure* contract at a low price. In view of these opposite interests you must determine the *smallest acceptable* price, which, for one thing, would enable you to meet the terms of the contract (i.e., to pay the amount of f_N at time N), but would give you no opportunity for arbitrage, no riskless gains (no *free lunch*), for the buyer has no reason to accept that.

On the other hand, *purchasing* a contract you are certainly willing to buy it cheap, but you should expect no opportunities for arbitrage, no riskless revenues, for the seller has no reasons to agree either.

We claim now that the 'upper' and 'lower' prices $\mathbb{C}^* = \mathbb{C}^*(f_N; \mathsf{P})$ and $\mathbb{C}_* = \mathbb{C}_*(f_N; \mathsf{P})$ introduced above have the following property: the intervals $[0, \mathbb{C}_*)$ and $(\mathbb{C}^*, \infty)$ are the (maximal) sets of price values that give a buyer or a seller, respectively, *opportunities for arbitrage.*

Assume that the price x of a contract is larger than $\mathbb{C}^*$ and that it is sold. Then the seller can get a *free lunch* acting as follows.

He deducts from the total sum x an amount y such that $\mathbb{C}^* < y < x$ and uses this money to build a portfolio $\pi^*(y)$ such that $X_0^{\pi^*(y)} = y$ and $X_N^{\pi^*(y)} \geqslant f_N$ at time N. The existence of such a portfolio follows by the definition of $\mathbb{C}^*$ and the inequality $y > \mathbb{C}^*$. (We can describe the same action otherwise: the seller *invests* the amount y in the (B, S)-market in accordance with the strategy $\pi^*(y) = (\beta_n^*(y), \gamma_n^*(y))_{0 \leqslant n \leqslant N}$.)

The value of this portfolio $\pi^*(y)$ at time N is $X_N^{\pi^*(y)}$, and the total gains from the two transactions (selling the contract and buying the portfolio $\pi^*(y)$) are

$$(x - f_N) + (X_N^{\pi^*(y)} - y) = (x - y) + (X_N^{\pi^*(y)} - f_N) \geqslant x - y > 0.$$

Here $x + X_N^{\pi^*(y)}$ are the returns from the transactions (at time $n = 0$ and $n = N$) and $f_N + y$ is the amount payable at time $n = N$ and, at time $n = 0$, for the purchase of the portfolio $\pi^*(y)$.

Thus, $x - y$ are the net riskless (i.e., arbitrage) gains of the seller.

We consider now the opportunities for arbitrage existing for the buyer.

Assume that the buyer buys a contract stipulating the payment of the amount f_N at a price $x < \mathbb{C}_*$. To get a *free lunch* the buyer can choose y such that $x < y < \mathbb{C}_*$.

By the definition of $\mathbb{C}_*$ there exists a portfolio $\pi_*(y)$ of initial (i.e., corresponding to $n = 0$) value y such that its value at the instant $n = N$ is $X_N^{\pi_*(y)} \leqslant f_N$. Bearing this in mind the buyer (who has already paid x for getting the promised amount f_N at time N) proceeds as follows.

At time $n = 0$ he invests the amount $(-y)$ in accordance with the strategy $\overline{\pi}(-y) = -\pi_*(y)$, where $\pi_*(y) = (\beta_{*n}(y), \gamma_{*n}(y))_{0 \leqslant n \leqslant N}$.

Thus,

$$\overline{\pi}(-y) = \bigl(-\beta_{*n}(y), -\gamma_{*n}(y)\bigr)_{0 \leqslant n \leqslant N}$$

so that $(-y)$ is distributed among bonds and stock in accordance with the formula

$$-y = -\beta_{*0}(y)B_0 - \gamma_{*0}(y)S_0.$$

The value of $\overline{\pi}(-y)$ at time N is

$$X_N^{\overline{\pi}(-y)} = X_N^{-\pi_*(y)} = -X_N^{\pi_*(y)},$$

so that the total gains form the two transactions (buying the contract and investing $(-y)$) is

$$(f_N - x) + \bigl(X_N^{\overline{\pi}(-y)} - (-y)\bigr) = \bigl(f_N - X_N^{\pi_*(y)}\bigr) + (y - x) \geqslant y - x > 0,$$

which, as we see, are the net (arbitrage) gains of the buyer purchasing the contract at the price $x < \mathbb{C}_*$.

In the above discussion we considered the investment of a negative (!) amount $(-y)$. What is the actual meaning of this transaction?

In fact, we already encountered such short selling—e.g., in the discussion of options in Chapter I, § 1c.4. In the current case of the investment of $(-y)$ this means merely that you find a speculator (cf. Chapter VII, § 1a) who agrees to pay you the amount y (at time $n = 0$) in return to your promise of the payment of $X_N^{\pi_*(y)}$ at time $n = N$; the latter amount can turn in effect to be more or less than y due to the random nature of prices.

Remark. These arguments are (implicitly) based on the assumption that the market is sufficiently *liquid*: there exists a full spectrum of investors, traders, speculators, ..., with different interests, views, and expectations of the behavior of prices. All this indicates, incidentally, that for a rigorous discussion one must precisely specify the requirements and constraints on the admissible transactions. The classes of admissible strategies considered below (see, e.g., Chapter VI, § 1a) provide examples of such constraints that, in particular, rule out arbitrage.

Thus, we have two intervals, $[0, \mathbb{C}_*)$ and $(\mathbb{C}^*, \infty)$, of prices giving opportunities for arbitrage. At the same time, if $x \in [\mathbb{C}_*, \mathbb{C}^*]$, then the buyer and the seller have no such opportunities.

This justifies the name 'interval of *acceptable, mutually acceptable* prices' given to $[\mathbb{C}_*, \mathbb{C}^*]$.

We emphasize again that a transaction at a *mutually acceptable* price $x \in [\mathbb{C}_*, \mathbb{C}^*]$ gives no riskless gains to either side. Both can gain or lose due to the random behavior of prices. Hence, as already mentioned, a gain, and especially a big gain, must be regarded as

'*a compensation for the risk*'

We summarize our discussion of prices acceptable to buyer and seller in the following figure:

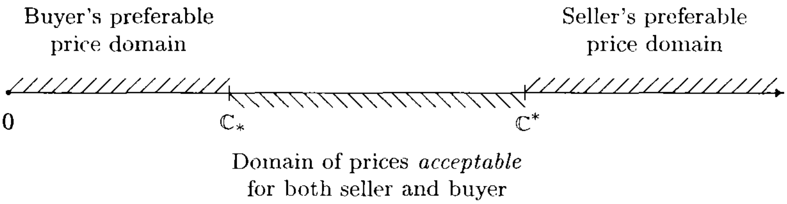


FIGURE 51. Domains of prices preferable and acceptable for buyer and seller ($\mathbb{C}_* = \mathbb{C}_*(f_N; \mathsf{P})$, $\mathbb{C}^* = \mathbb{C}^*(f_N; \mathsf{P})$)

4. We consider now more thoroughly the case where, for fixed x and a pay-off f_N, there exists a *perfect* (x, f_N)-hedge π, i.e., a strategy such that $X_0^\pi = x$ and $X_N^\pi = f_N$ (P-a.s.).

The equality $X_N^\pi = f_N$ means that the hedge π *replicates* the contingent claim f_N.

It is desirable for many reasons that each obligation be replicable (for some value of x). If this is the case, then the classes $H^*(x, f_N; \mathsf{P})$ and $H_*(x, f_N; \mathsf{P})$ clearly coincide and the upper and lower prices are also *the same*:

$$\mathbb{C}^*(f_N; \mathsf{P}) = \mathbb{C}_*(f_N; \mathsf{P}).$$

In other words, the *interval* of acceptable prices reduces in this case to a *single* price

$$C(f_N; \mathsf{P}) \quad (= \mathbb{C}^*(f_N; \mathsf{P}) = \mathbb{C}_*(f_N; \mathsf{P})),$$

the *rational* (*fair*) price of the contingent claim f_N acceptable for both buyer and seller. As explained above, a deviation from this price will inevitably give one of them *riskless* gains!

This case, in view of its importance, deserves special terminology.

DEFINITION 3. A (B, S)-market of securities is said to be *N-perfect* or *perfect with respect to time* N, if *each* $\mathscr{F}_N$-measurable (finite) pay-off function f_N can be replicated, i.e., for some x there exists a perfect (x, f_N)-hedge π, a portfolio such **that**

$$X_N^\pi = f_N \quad (\mathsf{P}\text{-a.s.}).$$

Otherwise the market is said to be *N-imperfect* or *imperfect with respect to* N.

In general, the condition that a (B, S)-market be perfect is a fairly strong one and it imposes rather serious constraints on the structure of the market. Incidentally, it is not necessary in many cases that a perfect hedge exist for *all* $\mathscr{F}_N$-measurable functions f_N; it suffices to consider, for instance, only *bounded* functions or functions in a certain subclass described by some or other conditions of integrability and measurability. (See, nevertheless, the theorem in § 4f.)

DEFINITION 4. We say that a (B, S)-market of securities is *N-complete* or *complete* with respect to time N if each *bounded* $\mathscr{F}_N$-measurable pay-off is replicable.

The question of whether and when a market is perfect or complete is of a major interest for financial mathematics and engineering; to answer it means to show that it is in principle possible (or impossible) to make up a portfolio π of value X_N^π *replicating* the pay-off f_N. (In the case where f_N cannot be precisely replicated one may, as already mentioned, pose the problem of finding a portfolio delivering $\inf \mathsf{E}[X_N^\pi - f_N]^2$, where the infimum is taken over all 'admissible' portfolios π: see Chapter VI, § 1d in this connection.)

It seems difficult to answer this question in a satisfactory way in the general case, making no additional assumptions about the *structure of the market* and the *probability space* underlying this market. However, in the case of so-called *arbitrage-free* markets (see the definitions in § 2a below) this problem of 'completeness' has an exhaustive solution in terms of the uniqueness of the so-called *martingale* measures (see Theorem A in § 2b and Theorem B in § 4a).

In the next section we consider a single-step model ($N = 1$) of a fairly simple (B, S)-market—an example where one can clearly see how *martingale* (*risk neutral*) measures enter in a natural way the calculations of the prices $\mathbb{C}_*(f_N; \mathsf{P})$ and $\mathbb{C}^*(f_N; \mathsf{P})$, and one can get acquainted with the *general pricing priciples* based essentially on the use of these measures.

§ 1c. Upper and Lower Prices in a Single-Step Model

1. We consider a simple 'single-step' model of a (B, S)-market formed by a bank account $B = (B_n)$ and some stock of price $S = (S_n)$, where $n = 0, 1$. We assume that the constants B_0 and S_0 are positive and (see § 1a)

$$\begin{aligned} B_1 &= B_0(1 + r), \\ S_1 &= S_0(1 + \rho), \end{aligned} \tag{1}$$

where the interest rate r is a constant ($r > -1$) and the rate ρ is a random variable ($\rho > -1$).

Since ρ is the source of all 'randomness' in this model, we can content ourselves with the consideration of a probability distribution $\mathsf{P} = \mathsf{P}(d\rho)$ on the set $\{\rho\colon -1 < \rho < \infty\}$ with Borel subsets.

2. We shall assume now that the support of $\mathsf{P}(d\rho)$ lies in the interval $[a,b]$, where $-1 < a < b < \infty$. (If $\mathsf{P}(d\rho)$ is concentrated at the *two* points $\{a\}$ and $\{b\}$, then (1) is the single-step *CRR*-model introduced in Chapter II, § 1e.)

Let $f = f(S_1)$ be the pay-off function and, for the simplicity of notation, let $\mathbb{C}^*(\mathsf{P}) = \mathbb{C}^*(f;\mathsf{P})$ and $\mathbb{C}_*(\mathsf{P}) = \mathbb{C}_*(f;\mathsf{P})$. Since $B_0 > 0$, we can assume without loss of generality that $B_0 = 1$.

In our single-step model the portfolio π can be described by a pair of numbers β and γ that must be specified at time $n = 0$.

In accordance with the definitions in the preceding section,

$$\mathbb{C}^*(\mathsf{P}) = \inf_{(\beta,\gamma)\in H^*(\mathsf{P})} (\beta + \gamma S_0) \tag{2}$$

and

$$\mathbb{C}_*(\mathsf{P}) = \sup_{(\beta,\gamma)\in H_*(\mathsf{P})} (\beta + \gamma S_0), \tag{3}$$

where

$$H^*(\mathsf{P}) = \big\{(\beta,\gamma)\colon \beta B_1 + \gamma S_1 \geqslant f(S_1),\ \mathsf{P}\text{-a.s.}\big\} \tag{4}$$

and

$$H_*(\mathsf{P}) = \big\{(\beta,\gamma)\colon \beta B_1 + \gamma S_1 \leqslant f(S_1),\ \mathsf{P}\text{-a.s.}\big\}. \tag{5}$$

We consider now the constraint

$$\beta(1+r) + \gamma S_0(1+\rho) \geqslant f\big(S_0(1+\rho)\big) \quad (\mathsf{P}\text{-a.s.}) \tag{6}$$

involved in the definition of $H^*(\mathsf{P})$ and introduce (however artificial it may seem at first glance) the class $\mathscr{P}(\mathsf{P}) = \{\widetilde{\mathsf{P}} = \widetilde{\mathsf{P}}(d\rho)\}$ of distributions on $[a,b]$ such that

$$\widetilde{\mathsf{P}} \sim \mathsf{P}$$

(i.e., the measures P and $\widetilde{\mathsf{P}}$ must be absolutely continuous with respect to each other: $\widetilde{\mathsf{P}} \ll \mathsf{P}$ and $\mathsf{P} \ll \widetilde{\mathsf{P}}$) and

$$\int_a^b \rho\, \widetilde{\mathsf{P}}(d\rho) = r. \tag{7}$$

We assume that $\mathscr{P}(\mathsf{P}) \neq \varnothing$. (This holds surely in the *CRR*-model considered below, in § 1d and in detail in § 3f.)

We point out the following important property underlying many calculations in financial mathematics:

> *if a measure $\widetilde{\mathsf{P}}$ is equivalent to P, then 'P-a.s.' in inequality* (6) *can be replaced by '$\widetilde{\mathsf{P}}$-a.s.'.*

Consequently, if $(\beta, \gamma) \in H^*(\mathsf{P})$, then

$$\beta(1+r) + \gamma S_0(1+\rho) \geqslant f(S_0(1+\rho)) \quad (\widetilde{\mathsf{P}}\text{-a.s.}), \tag{8}$$

therefore, integrating both parts of this inequality with respect to $\widetilde{\mathsf{P}} \in \mathscr{P}(\mathsf{P})$ and taking (7) into account we see that

$$\beta + \gamma S_0 \geqslant \mathsf{E}_{\widetilde{\mathsf{P}}} \frac{f(S_0(1+\rho))}{1+r}. \tag{9}$$

From this inequality we can obviously derive the following lower *bound* for $\mathbb{C}^*(\mathsf{P})$:

$$\begin{aligned} \mathbb{C}^*(\mathsf{P}) &= \inf_{(\beta,\gamma)\in H^*(\mathsf{P})} (\beta + \gamma S_0) \\ &\geqslant \sup_{\widetilde{\mathsf{P}}\in\mathscr{P}(\mathsf{P})} \mathsf{E}_{\widetilde{\mathsf{P}}} \frac{f(S_0(1+\rho))}{1+r} \quad (= x^*). \end{aligned} \tag{10}$$

In a similar way we obtain an upper bound for $\mathbb{C}_*(\mathsf{P})$:

$$\mathbb{C}_*(\mathsf{P}) \leqslant \inf_{\widetilde{\mathsf{P}}\in\mathscr{P}(\mathsf{P})} \mathsf{E}_{\widetilde{\mathsf{P}}} \frac{f(S_0(1+\rho))}{1+r} \quad (= x_*). \tag{11}$$

Thus,

$$\mathbb{C}_*(\mathsf{P}) \leqslant x_* \leqslant x^* \leqslant \mathbb{C}^*(\mathsf{P}), \tag{12}$$

which (provided that $\mathscr{P}(\mathsf{P}) \neq \varnothing$) proves the inequality $\mathbb{C}_*(\mathsf{P}) \leqslant \mathbb{C}^*(\mathsf{P})$, not all that obvious from the definitions of $\mathbb{C}_*(\mathsf{P})$ and $\mathbb{C}^*(\mathsf{P})$.

We proceed now to property (7), which can be written as

$$\mathsf{E}_{\widetilde{\mathsf{P}}} \frac{1+\rho}{1+r} = 1. \tag{13}$$

By (1), this is equivalent to

$$\mathsf{E}_{\widetilde{\mathsf{P}}} \frac{S_1}{B_1} = \frac{S_0}{B_0}. \tag{14}$$

We set $\mathscr{F}_0 = \{\varnothing, \mathbb{R}\}$ and $\mathscr{F}_1 = \sigma(\rho)$; then we see that

$$\mathsf{E}_{\widetilde{\mathsf{P}}}\left(\frac{S_1}{B_1} \,\Big|\, \mathscr{F}_0\right) = \frac{S_0}{B_0}.$$

i.e., the sequence $\left(\dfrac{S_n}{B_n}\right)_{n=0,1}$ is a *martingale* with respect to the measure $\widetilde{\mathsf{P}}$.

It is because of just this property (which holds also in a more general case) that measures $\widetilde{\mathsf{P}} \in \mathscr{P}(\mathsf{P})$ are usually called *martingale measures* in financial mathematics.

We note that their appearance in the calculation of, say, upper and lower prices, can at first seem forcible, because there is already a probability measure on the original basis and, presumably, all the pricing should be based on this measure alone. In effect, this is the case because

$$\mathsf{E}_{\widetilde{\mathsf{P}}} \frac{f(S_0(1+\rho))}{1+r} = \mathsf{E}_{\mathsf{P}}\, z(\rho) \frac{f(S_0(1+\rho))}{1+r}, \tag{15}$$

where $z(\rho) = \dfrac{d\widetilde{\mathsf{P}}}{d\mathsf{P}}$ is the Radon–Nikodym derivative of $\widetilde{\mathsf{P}}$ with respect to P.

However, the introduction of martingale measures has a deeper meaning because their existence is in direct connection with the *absence of arbitrage* on our (B,S)-market.

We shall discuss this issue in § 2 "Arbitrage-free market", while here we consider the question of *equality* signs in (10) and (11).

3. Assume that the function $f_x = f(S_0(1+x))$ is (downwards) convex and continuous on $[a,b]$. (We recall that each convex function on $[a,b]$ is continuous on the open interval (a,b) and can make jumps only at the *end-points* of this interval.)

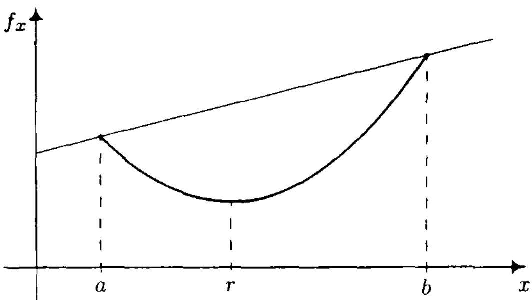


FIGURE 52. Pay-off function $f_x = f(S_0(1+x))$

We plot a straight line $y = y(x)$ through the points (a, f_a) and (b, f_b). If it has an equation

$$y(x) = \mu S_0(1+x) + \nu, \tag{16}$$

then, clearly,

$$\mu = \frac{f_b - f_a}{S_0(b-a)} \quad \text{and} \quad \nu = \frac{(1+b)f_a - (1+a)f_b}{(b-a)}. \tag{17}$$

Let us introduce the strategy $\pi^* = (\beta^*, \gamma^*)$ with

$$\beta^* = \frac{\nu}{1+r} \quad \text{and} \quad \gamma^* = \mu. \tag{18}$$

Since $(1+r)\beta^* + S_0(1+\rho)\gamma^* = \nu + \mu S_0(1+\rho) \geqslant f(\rho)$ for all $\rho \in [a,b]$, it follows that $\pi^* \in H^*(\mathsf{P})$, and therefore

$$\mathbb{C}^*(\mathsf{P}) = \inf_{(\beta,\gamma)\in H^*(\mathsf{P})} (\beta + \gamma S_0) \leqslant \beta^* + \gamma^* S_0 = \frac{\nu}{1+r} + \mu S_0. \tag{19}$$

We now make the following assumption about the set $\mathscr{P}(\mathsf{P})$ of martingale measures:

(A^*): there exists a sequence $(\widetilde{\mathsf{P}}_n)_{n\geqslant 1}$ of measures in $\mathscr{P}(\mathsf{P})$ that converges weakly to a measure P^* concentrated at the *two* points a and b.

If (A^*) holds, then we see from the equality

$$\mathsf{E}_{\widetilde{\mathsf{P}}_n} \frac{1+\rho}{1+r} = 1$$

that

$$\mathsf{E}_{\mathsf{P}^*} \frac{1+\rho}{1+r} = 1.$$

Hence the probabilities $p^* = \mathsf{P}^*\{b\}$ and $q^* = \mathsf{P}^*\{a\}$ satisfy the conditions

$$p^* + q^* = 1$$

and

$$bp^* + aq^* = r.$$

Consequently,

$$p^* = \frac{r-a}{b-a} \quad \text{and} \quad q^* = \frac{b-r}{b-a}. \tag{20}$$

Further, using our assumption of the weak convergence of $\widetilde{\mathsf{P}}_n$ to P^* again and in view of (19), we obtain

$$\begin{aligned}
\sup_{\widetilde{\mathsf{P}}\in\mathscr{P}(\mathsf{P})} \mathsf{E}_{\widetilde{\mathsf{P}}} \frac{f_\rho}{1+r} &\geqslant \lim_n \mathsf{E}_{\widetilde{\mathsf{P}}_n} \frac{f_\rho}{1+r} = \mathsf{E}_{\mathsf{P}^*} \frac{f_\rho}{1+r} \\
&= p^* \frac{f_b}{1+r} + q^* \frac{f_a}{1+r} = \frac{r-a}{b-a}\,\frac{f_b}{1+r} + \frac{b-r}{b-a}\,\frac{f_a}{1+r} \\
&= \frac{\nu}{1+r} + \mu S_0 = \beta^* + \gamma^* S_0 \\
&\geqslant \inf_{(\beta,\gamma)\in H^*(\mathsf{P})} (\beta + \gamma S_0) = \mathbb{C}^*(\mathsf{P}).
\end{aligned} \tag{21}$$

Combined with the reverse inequality (10) this means that

$$\mathbb{C}^*(\mathsf{P}) = \inf_{(\beta,\gamma)\in H^*(\mathsf{P})} (\beta + \gamma S_0) = \sup_{\widetilde{\mathsf{P}}\in\mathscr{P}(\mathsf{P})} \mathsf{E}_{\widetilde{\mathsf{P}}} \frac{f(S_0(1+\rho))}{1+r}. \tag{22}$$

Concerning the above discussion it should be noted that our assumption that $f = f(S_0(1+\rho))$ is a continuous convex function of $\rho \in [a,b]$ is fairly standard and should encounter no serious objections.

Assumption (A^*) about the existence of martingale measures weakly convergent to a measure concentrated at a and b is more controversial. However, it is not that 'dreadful' provided that there exist non-vanishing P-masses close to a and b (i.e., for each $\varepsilon > 0$ we have $\mathsf{P}[a, a+\varepsilon] > 0$ and $\mathsf{P}[b-\varepsilon, b] > 0$). If there exists *at least one* martingale measure $\widetilde{\mathsf{P}} \sim \mathsf{P}$ with the same properties (i.e., such that $\widetilde{\mathsf{P}}[a, a+\varepsilon] > 0$, $\widetilde{\mathsf{P}}[b-\varepsilon, b] > 0$ for each $\varepsilon > 0$, and $\int_a^b \rho\,\widetilde{\mathsf{P}}(d\rho) = r$), then we can construct the required sequence $\{\widetilde{\mathsf{P}}_n\}$ by 'pumping' the measure $\widetilde{\mathsf{P}}$ into ever smaller ($\varepsilon \downarrow 0$) neighborhoods $[a, a+\varepsilon]$ and $[b-\varepsilon, b]$ of the points a and b preserving at the same time the equivalence $\widetilde{\mathsf{P}}_n \sim \widetilde{\mathsf{P}}$.

In the next section we consider the *CRR* (Cox–Ross–Rubinstein) model in which the original measure P is concentrated at a and b and the construction of P^* proceeds without complications. (Actually, we already constructed this measure in (20).)

We now formulate the result on $\mathbb{C}^*$ obtained in this way.

THEOREM 1. *Assume that the pay-off function $f(S_0(1+\rho))$ is convex and continuous in ρ on $[a,b]$ and assume that the weak compactness condition (A^*) holds. Then the upper price can be expressed by the formula*

$$\mathbb{C}^*(\mathsf{P}) = \sup_{\widetilde{\mathsf{P}} \in \mathscr{P}(\mathsf{P})} \mathsf{E}_{\widetilde{\mathsf{P}}} \frac{f(S_0(1+\rho))}{1+r}. \tag{23}$$

Moreover, the supremum is attained at the measure P^* and

$$\mathbb{C}^*(\mathsf{P}) = \frac{r-a}{b-a}\,\frac{f_b}{1+r} + \frac{b-r}{b-a}\,\frac{f_a}{1+r}, \tag{24}$$

where $f_\rho = f(S_0(1+\rho))$.

4. We consider now the lower price $\mathbb{C}_*$.

By (3) and (11),

$$\mathbb{C}_*(\mathsf{P}) = \sup_{(\beta,\gamma) \in H_*(\mathsf{P})} (\beta + \gamma S_0) \leqslant \inf_{\widetilde{\mathsf{P}} \in \mathscr{P}(\mathsf{P})} \mathsf{E}_{\widetilde{\mathsf{P}}} \frac{f_\rho}{1+r}, \tag{25}$$

where $f_\rho = f(S_0(1+\rho))$, $\rho \in [a,b]$.

If f_ρ is (downwards) convex on $[a,b]$, then for each $r \in (a,b)$ there exists $\lambda = \lambda(r)$ such that

$$f\big(S_0(1+\rho)\big) \geqslant f\big(S_0(1+r)\big) + (\rho - r)\lambda(r) \tag{26}$$

for all $\rho \in [a,b]$, where

$$y(r) = f\big(S_0(1+r)\big) + (\rho - r)\lambda(r)$$

is the 'support line' (a line through r such that the graph of f_ρ lies above it).

Let $\widetilde{\mathsf{P}} \in \mathscr{P}(\mathsf{P})$. Then by (26) we obtain

$$\inf_{\widetilde{\mathsf{P}} \in \mathscr{P}(\mathsf{P})} \mathsf{E}_{\widetilde{\mathsf{P}}} \frac{f(S_0(1+\rho))}{1+r} \geqslant \frac{f(S_0(1+r))}{1+r}. \tag{27}$$

We set

$$\beta_* = \frac{f(S_0(1+r))}{1+r} - \lambda(r)$$

and

$$\gamma_* = \frac{\lambda(r)}{S_0}.$$

Then (26) takes the following form: for each $\rho \in [a, b]$ we have

$$f\big(S_0(1+\rho)\big) \geqslant \beta_*(1+r) + \gamma_* S_0(1+\rho),$$

which means that $\pi_* = (\beta_*, \gamma_*) \in H_*(\mathsf{P})$.

Following the pattern of the proof of Theorem 1 we assume now that

(A_*): there exists a sequence $\{\overline{\mathsf{P}}_n\}_{n\geqslant 1}$ of measures in $\mathscr{P}(\mathsf{P})$ converging weakly to a measure P_* concentrated at a *single* point r.

Then, assuming that f_ρ is a continuous function, we obtain

$$\begin{aligned}\inf_{\widetilde{\mathsf{P}} \in \mathscr{P}(\mathsf{P})} \mathsf{E}_{\widetilde{\mathsf{P}}} \frac{f_\rho}{1+r} &\leqslant \lim_n \mathsf{E}_{\overline{\mathsf{P}}_n} \frac{f_\rho}{1+r} = \mathsf{E}_{\mathsf{P}_*} \frac{f_\rho}{1+\rho} = \frac{f(S_0(1+r))}{1+r} \\ &= \beta_* + S_0\gamma_* \leqslant \sup_{(\beta,\gamma)\in H_*(\mathsf{P})} (\beta + S_0\gamma) = \mathbb{C}_*(\mathsf{P}),\end{aligned}$$

which, combined with (26), proves the following result.

THEOREM 2. *Let f_ρ be a continuous, downwards convex function on $[a, b]$ and assume that weak compactness condition (A_*) holds. Then the lower price can be expressed by the formula*

$$\mathbb{C}_*(\mathsf{P}) = \inf_{\widetilde{\mathsf{P}} \in \mathscr{P}(\mathsf{P})} \mathsf{E}_{\widetilde{\mathsf{P}}} \frac{f(S_0(1+\rho))}{1+r}. \tag{28}$$

Moreover, the infimum is attained at the measure P_* and

$$\mathbb{C}_*(\mathsf{P}) = \frac{f_r}{1+r}. \tag{29}$$

Remark. Let, for instance, $\mathsf{P}(d\rho) = \dfrac{d\rho}{b-a}$ be the uniform measure on $[a, b]$. Then conditions (A^*) and (A_*) are satisfied, so that the upper and the lower prices can be defined by (24) and (29).

5. In the above 'probabilistic' analysis of upper and lower prices we took it for granted that all the *uncertainty* in stock prices can be described in probabilistic terms. This was reflected in the assumption that ρ is a *random* variable with *probability distribution* $\mathsf{P}(d\rho)$.

However, we could treat ρ merely as a *chaotic* variable ranging in the interval $[a,b]$. In that case, in place of $H^*(\mathsf{P})$ and $H_*(\mathsf{P})$ it is reasonable to introduce the classes

$$\widehat{H}^* = \bigl\{(\beta,\gamma)\colon \beta(1+r)+\gamma S_0(1+\rho) \geqslant f(S_0(1+\rho)) \text{ for each } \rho\in[a,b]\bigr\}$$

and

$$\widehat{H}_* = \bigl\{(\beta,\gamma)\colon \beta(1+r)+\gamma S_0(1+\rho) \leqslant f(S_0(1+\rho)) \text{for each } \rho\in[a,b]\bigr\},$$

(i.e., to replace the condition 'P-a.s.' by the requirement: '*for each* $\rho\in[a,b]$'). Clearly,

$$\widehat{H}^* \subseteq H^*(\mathsf{P}) \quad\text{and}\quad \widehat{H}_* \subseteq H_*(\mathsf{P})$$

for each probability measure P.

We note that even if no probability measure has been fixed a priori, there are no indications against the (maybe, forcible) introduction of a distribution $\widehat{\mathsf{P}} = \widehat{\mathsf{P}}(d\rho)$ on $[a,b]$ such that (cf. (7))

$$\int_a^b \rho\,\widehat{\mathsf{P}}(d\rho) = r.$$

The class $\widehat{\mathscr{P}}$ of such measures is surely non-empty; it contains the above-discussed measures P^*, concentrated at a and b (see (20)), and P_*, concentrated at a single point r.

By analogy with (2) and (3) we now set

$$\widehat{\mathbb{C}}^* = \inf_{(\beta,\gamma)\in\widehat{H}^*} (\beta+\gamma S_0) \quad (\geqslant \mathbb{C}^*(\mathsf{P}))$$

and

$$\widehat{\mathbb{C}}_* = \sup_{(\beta,\gamma)\in\widehat{H}_*} (\beta+\gamma S_0) \quad (\leqslant \mathbb{C}_*(\mathsf{P})).$$

The above arguments (see (8), (9), and (10)) show that for each probability measure P we have

$$\begin{aligned}\widehat{\mathbb{C}}^* = \inf_{(\beta,\gamma)\in\widehat{H}^*} (\beta+\gamma S_0) &\geqslant \sup_{\widehat{\mathsf{P}}\in\widehat{\mathscr{P}}} \mathsf{E}_{\widehat{\mathsf{P}}} \frac{f(S_0(1+\rho))}{1+r}\\ &\geqslant \sup_{\widetilde{\mathsf{P}}\in\mathscr{P}(\mathsf{P})} \mathsf{E}_{\widetilde{\mathsf{P}}} \frac{f(S_0(1+\rho))}{1+r},\end{aligned} \tag{30}$$

and

$$\widehat{\mathbb{C}}_* = \sup_{(\beta,\gamma)\in\widehat{H}_*} (\beta+\gamma S_0) \leqslant \inf_{\widehat{\mathsf{P}}\in\widehat{\mathscr{P}}} \mathsf{E}_{\widehat{\mathsf{P}}} \frac{f(S_0(1+\rho))}{1+r}$$
$$\leqslant \inf_{\widetilde{\mathsf{P}}\in\mathscr{P}(\mathsf{P})} \mathsf{E}_{\widetilde{\mathsf{P}}} \frac{f(S_0(1+\rho))}{1+r}. \tag{31}$$

We note next that the class $\widehat{\mathscr{P}}$ contains also both 'two-point' measure P^* and 'single-point' measure P_* considered above, therefore (cf. (21))

$$\sup_{\widehat{\mathsf{P}}\in\widehat{\mathscr{P}}} \mathsf{E}_{\widehat{\mathsf{P}}} \frac{f(S_0(1+\rho))}{1+r} \geqslant \mathsf{E}_{\mathsf{P}^*} \frac{f(S_0(1+\rho))}{1+r} = p^* \frac{f_b}{1+r} + q^* \frac{f_a}{1+r} \tag{32}$$

and

$$\inf_{\widehat{\mathsf{P}}\in\widehat{\mathscr{P}}} \mathsf{E}_{\widehat{\mathsf{P}}} \frac{f(S_0(1+\rho))}{1+r} \leqslant \mathsf{E}_{\mathsf{P}_*} \frac{f(S_0(1+\rho))}{1+r} = \frac{f_r}{1+r}. \tag{33}$$

The strategy π^* discussed in the proof of Theorem 1 belongs clearly to the class $\widehat{H}^*$, and (provided that f_ρ is a (downwards) convex function on $[a,b]$)

$$\widehat{\mathbb{C}}^* = \inf_{(\beta,\gamma)\in\widehat{H}^*} (\beta+\gamma S_0) \leqslant \beta^* + \gamma^* S_0 = p^* \frac{f_b}{1+r} + q^* \frac{f_a}{1+r},$$

which together with (30) and (32) shows (cf. (23)) that

$$\widehat{\mathbb{C}}^* = \sup_{\widehat{\mathsf{P}}\in\widehat{\mathscr{P}}} \mathsf{E}_{\widehat{\mathsf{P}}} \frac{f(S_0(1+\rho))}{1+r}; \tag{34}$$

moreover (cf. (24)),

$$\widehat{\mathbb{C}}^* = p^* \frac{f_b}{1+r} + q^* \frac{f_a}{1+r}, \tag{35}$$

where $f_\rho = f(S_0(1+\rho))$.

In a similar way (cf. (28) and (29)),

$$\widehat{\mathbb{C}}_* = \inf_{\widehat{\mathsf{P}}\in\widehat{\mathscr{P}}} \mathsf{E}_{\widehat{\mathsf{P}}} \frac{f(S_0(1+\rho))}{1+r}; \tag{36}$$

moreover,

$$\widehat{\mathbb{C}}_* = \frac{f_r}{1+r}. \tag{37}$$

Thus, we have established the following result.

THEOREM 3. *If the variable ρ involved in* (1) *and ranging in* $[a, b]$ *is 'chaotic', then for each downwards convex function* $f = f(S_0(1+\rho))$ *the upper and lower prices* $\widehat{\mathbb{C}}^*$ *and* $\widehat{\mathbb{C}}_*$ *can be defined by formulas* (34), (35) *and* (36), (37), *respectively.*

Remark. Comparing Theorems 1, 2, and 3 we see that if the original probability measure P is sufficiently 'fuzzy' so that it has masses close to the points a, r, and b, then the class $\mathscr{P}(\mathsf{P})$ of martingale measures is as rich as $\widehat{\mathscr{P}}$ and we have equality signs in the inequalities $\widehat{\mathbb{C}}_* \leqslant \mathbb{C}_*(\mathsf{P})$ and $\mathbb{C}^*(\mathsf{P}) \leqslant \widehat{\mathbb{C}}^*$.

§ 1d. *CRR* Model: an Example of a Complete Market

1. We consider again a 'single-step' model of a (B, S)-market, in which we assume that

$$\begin{aligned} B_1 &= B_0(1+r), \\ S_1 &= S_0(1+\rho), \end{aligned} \tag{1}$$

where ρ is a random variable taking just *two values*, a and b, such that

$$-1 < a < r < b. \tag{2}$$

This simple (B, S)-market is called the *single-step CRR-model* (after J. C. Cox, S. A. Ross, and M. Rubinstein who considered it in [82]).

We assume that the initial distribution P of the random variable ρ is as follows:

$$p = \mathsf{P}\{b\} > 0, \qquad q = \mathsf{P}\{a\} > 0.$$

Then the *unique* (martingale) measure P^* equivalent to P and satisfying property (7) in the preceding section can be defined as follows:

$$\mathsf{P}^*\{b\} = p^*, \qquad \mathsf{P}^*\{a\} = q^*,$$

where (see (20) in § 1c)

$$p^* = \frac{r-a}{b-a}, \qquad q^* = \frac{b-r}{b-a}, \tag{3}$$

and that (see (11) and (12) in § 1c)

$$\mathbb{C}_*(\mathsf{P}) \leqslant \mathsf{E}_{\mathsf{P}^*} \frac{f_\rho}{1+r} \leqslant \mathbb{C}^*(\mathsf{P}). \tag{4}$$

In fact, the prices $\mathbb{C}_*(\mathsf{P})$ and $\mathbb{C}^*(\mathsf{P})$ are the same in our case for *each* pay-off function $f = f(S_0(1+\rho))$, so that their common value is given by the formula

$$\mathbb{C}(\mathsf{P}) = \mathsf{E}_{\mathsf{P}^*} \frac{f_\rho}{1+r} = p^* \frac{f_b}{1+r} + q^* \frac{f_a}{1+r}. \tag{5}$$

Actually, all we need for the proof of the equality $\mathbb{C}_*(\mathsf{P}) = \mathbb{C}^*(\mathsf{P})$ is already contained in the discussion in § 1c.

For let us consider the above-introduced strategy $\pi^* = (\beta^*, \gamma^*)$ with parameters

$$\beta^* = \frac{\nu}{1+r} \quad \text{and} \quad \gamma^* = \mu,$$

where ν and μ are as in formula (17) in § 1c.

For both $\rho = a$ and $\rho = b$ we have

$$\beta^*(1+r) + \gamma^* S_0(1+\rho) = f\big(S_0(1+\rho)\big),$$

therefore the pay-off function here can be *replicated*, so that $\pi^* \in H^*(\mathsf{P}) \cap H_*(\mathsf{P})$. Hence

$$\begin{aligned} \mathbb{C}_*(\mathsf{P}) &= \sup_{(\beta,\gamma)\in \widehat{H}_*(\mathsf{P})} (\beta + \gamma S_0) \geqslant \beta^* + \gamma^* S_0 \\ &\geqslant \inf_{(\beta,\gamma)\in \widehat{H}^*(\mathsf{P})} (\beta + \gamma S_0) = \mathbb{C}^*(\mathsf{P}). \end{aligned}$$

Combined with (4), this proved the required equality of $\mathbb{C}_*(\mathsf{P})$ and $\mathbb{C}^*(\mathsf{P})$ and formula (5) for their common value $\mathbb{C}(\mathsf{P})$.

2. We point out again the following important points revealed by our exposition here and in the previous section. They are valid also in cases when the martingale methods are used in a more general context of financial calculations relating to a fixed pay-off function:

(I) *if the class of martingale measures is non-empty*:

$$\mathscr{P}(\mathsf{P}) \neq \varnothing, \tag{6}$$

then

$$\mathbb{C}_*(\mathsf{P}) \leqslant \mathbb{C}^*(\mathsf{P})$$

(by formula (12), § 1c);

(II) *if*

$$H^*(\mathsf{P}) \cap H_*(\mathsf{P}) \neq \varnothing, \tag{7}$$

i.e., *the pay-off function is replicable, then*

$$\mathbb{C}_*(\mathsf{P}) \geqslant \mathbb{C}^*(\mathsf{P});$$

(III) *if both* (6) *and* (7) *hold, then the upper and the lower prices* $\mathbb{C}_*(\mathsf{P})$ *and* $\mathbb{C}^*(\mathsf{P})$ *are the same.*

We show in the next section that the non-emptiness of the class of martingale measures $\mathscr{P}(\mathsf{P})$ is directly related to the *absence of arbitrage.*

On the other hand, the non-emptiness of $H^*(\mathsf{P}) \cap H_*(\mathsf{P})$ (for an arbitrary pay-off function f) is related to the *uniqueness* of the martingale measure, i.e., to the question whether the set $\mathscr{P}(\mathsf{P})$ reduces to a single (martingale) measure $\widetilde{\mathsf{P}}$ equivalent to P ($\widetilde{\mathsf{P}} \sim \mathsf{P}$).

2. Arbitrage-Free Market

§ 2a. 'Arbitrage' and 'Absence of Arbitrage'

1. In a few words, the 'absence of arbitrage' on a market means that this market if 'fair', 'rational', and one can make no 'riskless' profits there. (Cf. the concept of 'efficient market' in Chapter I, § 2a, which is also based on a certain notion of what a 'rationally organized market' must be and where one simply postulates that prices on such a market must have the martingale property.)

For formal definitions we shall assume (as in § 1a) that we have a filtered probability space

$$(\Omega, \mathscr{F}, (\mathscr{F}_n)_{n\geqslant 0}, \mathsf{P}),$$

and a (B, S)-market on this space formed by $d + 1$ assets:

$$\text{a } \textit{bank account } B = (B_n)_{n\geqslant 0}$$

with $\mathscr{F}_{n-1}$-measurable B_n, $B_n > 0$, and

$$\text{a } d\textit{-dimensional risk asset } S = (S^1, \dots, S^d),$$

where $S^i = (S^i_n)_{n\geqslant 0}$, and the S^i_n are positive and $\mathscr{F}_n$-measurable.

Let $X^\pi = (X^\pi_n)_{n\geqslant 0}$ be the value

$$X^\pi_n = \beta_n B_n + \gamma_n S_n \quad \left(= \beta_n B_n + \sum_{i=1}^{d} \gamma^i_n S^i_n \right),$$

of the strategy $\pi = (\beta, \gamma)$ with predictable $\beta = (\beta_n)_{n\geqslant 0}$ and $\gamma = (\gamma^1, \dots, \gamma^d)$, $\gamma^i = (\gamma^i_n)_{n\geqslant 0}$.

If π is a *self-financing* strategy ($\pi \in SF$), then (see (12) in § 1a)

$$X_n^\pi = X_0^\pi + \sum_{k=1}^{n} (\beta_k \Delta B_k + \gamma_k \Delta S_k), \qquad n \geqslant 1, \tag{1}$$

and the discounted value of the portfolio $\widetilde{X}_n^\pi = \left(\dfrac{X_n^\pi}{B_n}\right)_{n\geqslant 0}$ satisfies the relations

$$\Delta\left(\frac{X_n^\pi}{B_n}\right) = \gamma_n \Delta\left(\frac{S_n}{B_n}\right), \tag{2}$$

which are of key importance for all the analysis that follows.

2. We fix some $N \geqslant 1$; we are interested in the value X_N^π of one or another strategy $\pi \in SF$ at this 'terminal' instant.

DEFINITION 1. We say that a self-financing strategy π brings about an *opportunity for arbitrage* (at time N) if, for starting capital

$$X_0^\pi = 0, \tag{3}$$

we have

$$X_N^\pi \geqslant 0 \quad (\mathsf{P}\text{-a.s.}) \tag{4}$$

and $X_N^\pi > 0$ with *positive* P-probability, i.e.,

$$\mathsf{P}(X_N^\pi > 0) > 0 \tag{5}$$

or, equivalently,

$$\mathsf{E} X_N^\pi > 0. \tag{6}$$

Let SF_{arb} be the class *of self-financing* strategies with opportunities for *arbitrage.* If $\pi \in SF_{\text{arb}}$ and $X_0^\pi = 0$, then

$$\mathsf{P}(X_N^\pi \geqslant 0) = 1 \implies \mathsf{P}(X_N^\pi > 0) > 0.$$

DEFINITION 2. We say that there exist *no opportunities for arbitrage* on a (B, S)-market or that the market is *arbitrage-free* if $SF_{\text{arb}} = \varnothing$. In other words, if the starting capital X_0^π of a strategy π is zero, then

$$\mathsf{P}(X_N^\pi \geqslant 0) = 1 \implies \mathsf{P}(X_N^\pi = 0) = 1.$$

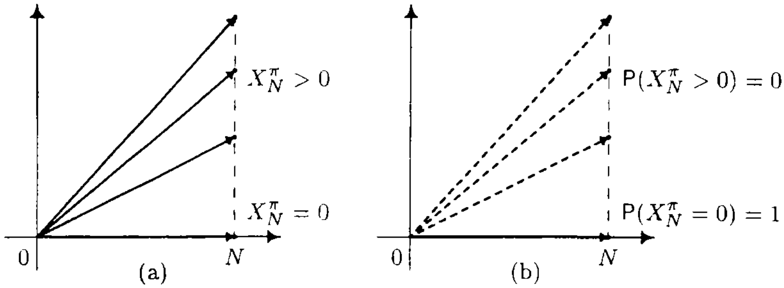


FIGURE 53. To the definition of an arbitrage-free market

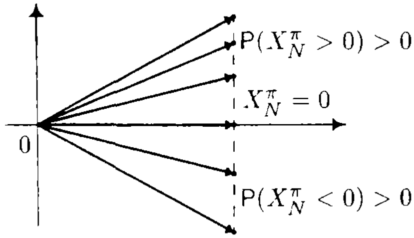


FIGURE 54. Typical pattern of transitions on an *arbitrage-free* market

Geometrically, this means that if π is arbitrage-free (and $X_N^\pi \geqslant 0$), then the chart (Fig. 53a) depicting transitions from $X_0^\pi = 0$ to X_N^π must actually be 'degenerate' as in Fig. 53b, where the dash lines correspond to transitions $X_0^\pi = 0 \to X_N^\pi$ of *probability zero*.

In general, if $X_0^\pi = \beta_0 B_0 + \gamma_0 S_0 = 0$, then the chart of transitions in an arbitrage-free market must be as in Fig. 54: if $\mathsf{P}(X_N^\pi = 0) < 1$, then both gains ($\mathsf{P}(X_N^\pi > 0) > 0$) and losses ($\mathsf{P}(X_N^\pi < 0) > 0$) must be possible. This can also be reformulated as follows: a strategy π (with $X_0^\pi = 0$) on an arbitrage-free market must be either trivial (i.e., $\mathsf{P}(X_N^\pi \neq 0) = 0$) or *risky* (i.e., we have *both* $\mathsf{P}(X_N^\pi > 0) > 0$ and $\mathsf{P}(X_N^\pi < 0) > 0$).

Along with the above definition of an *arbitrage-free* market, several other definitions are also used in finances (see, for instance, [251]). We present here two examples.

DEFINITION 3. a) A (B,S)-market is said to be *arbitrage-free in the weak sense* if for each self-financing strategy π satisfying the relations $X_0^\pi = 0$ and $X_n^\pi \geqslant 0$ (P-a.s.) for all $n \leqslant N$ we have $X_N^\pi = 0$ (P-a.s.).

b) A (B,S)-market is said to be *arbitrage-free in the strong sense* if for each self-financing strategy π the relations $X_0^\pi = 0$ and $X_N^\pi \geqslant 0$ (P-a.s.) mean that $X_n^\pi = 0$ (P-a.s.) for all $n \leqslant N$.

Remark. Note that in the above definitions we consider events of the form $\{X_N^\pi > 0\}$, $\{X_N^\pi \geqslant 0\}$, or $\{X_N^\pi = 0\}$, which are clearly the same as $\{\widetilde{X}_N^\pi > 0\}$, $\{\widetilde{X}_N^\pi \geqslant 0\}$, or $\{\widetilde{X}_N^\pi = 0\}$, respectively, where $\widetilde{X}_N^\pi = \dfrac{X_N^\pi}{B_N}$ (provided that $B_N > 0$). This explains why, in the discussion of the 'presence' or 'absence' of arbitrage on a (B,S)-market, one can restrict oneself to $(\widetilde{B},\widetilde{S})$-markets with $\widetilde{B}_n \equiv 1$ and $\widetilde{S}_n = \dfrac{S_n}{B_n}$. In other words, if we assume that $B_n > 0$, then we can also assume without loss of generality that $B_n \equiv 1$.

§ 2b. Martingale Criterion of the Absence of Arbitrage. First Fundamental Theorem

1. In our case of discrete time $n = 0, 1, \dots, N$ we have the following remarkable result, which, due to its importance, is called the *First fundamental asset pricing theorem.*

THEOREM A. *Assume that a*

$$(B,S)\text{-market}$$

on a filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_n), \mathsf{P})$ is formed by a bank account $B = (B_n)$, $B_n > 0$, and finitely many assets $S = (S^1, \dots, S^d)$, $S^i = (S_n^i)$.

Assume also that this market operates at the instants $n = 0, 1, \dots, N$, $\mathscr{F}_0 = \{\varnothing, \Omega\}$, and $\mathscr{F}_N = \mathscr{F}$.

Then this (B,S)-market is arbitrage-free if and only if there exists (at least one) measure $\widetilde{\mathsf{P}}$ (a 'martingale' measure) equivalent to the measure P such that the d-dimensional discounted sequence

$$\frac{S}{B} = \left(\frac{S_n}{B_n}\right)$$

is a $\widetilde{\mathsf{P}}$-martingale, i.e.,

$$\mathsf{E}_{\widetilde{\mathsf{P}}}\left|\frac{S_n^i}{B_n}\right| < \infty \tag{1}$$

for all $i = 1, \dots, d$ and $n = 0, 1, \dots, N$ and

$$\mathsf{E}_{\widetilde{\mathsf{P}}}\left(\frac{S_n^i}{B_n} \,\Big|\, \mathscr{F}_{n-1}\right) = \frac{S_{n-1}^i}{B_{n-1}} \qquad (\widetilde{\mathsf{P}}\text{-a.s.}) \tag{2}$$

for $n = 1, \dots, N$.

We split the proof of this result into several steps: we prove the sufficiency in § 2c and the necessity in § 2d. In § 2e we present another proof, of a slightly more general version of this result. Right now, we make several observations concerning the meaningfulness of the above criterion.

We have already mentioned that the assumption of the absence of arbitrage has a clear economic meaning; this is a desirable property of a market to be 'rational', 'efficient', 'fair'. The value of this theorem (which is due to J. M. Harrison and D. M. Kreps [214] and J. M. Harrison and S. R. Pliska [215] in the case of finite Ω and to R. C. Dalang, A. Morton, and W. Willinger [92] for arbitrary Ω) is that it shows a way to *analytic* calculations relevant to transactions of financial assets in these 'arbitrage-free' markets. (This is why it is called the *First fundamental asset pricing theorem.*) We have, in essence, already demonstrated this before, in our calculations of upper and lower prices (§ § 1b, c). We shall consistently use this criterion below; e.g., in our considerations of forward and futures prices or rational option prices (Chapter VI).

This theorem is also very important conceptually, for it demonstrates that the fairly vague *concept* of *efficient, rational* market (Chapter I, § 2a), put forward as some justification of the postulate of the martingale property of prices, becomes rigorous in the disguise of the *concept* of *arbitrage-free market*: a market is 'rationally organized' if the investors get no opportunities for *riskless* profits.

2. In operations with sequences $X = (X_n)$ that are martingales it is important to indicate not only the measure P, but also the flow of σ-algebras $(\mathscr{F}_n)$ in terms of which we state the martingale properties:

$$\text{the } X_n \text{ are } \mathscr{F}_n\text{-measurable,}$$

$$\mathsf{E}|X_n| < \infty,$$

$$\mathsf{E}(X_{n+1} \mid \mathscr{F}_n) = X_n \quad (\mathsf{P}\text{-a.s.}).$$

To emphasize this, we say that the martingale in question is a P-martingale or a $(\mathsf{P}, (\mathscr{F}_n))$-martingale and write $X = (X_n, \mathscr{F}_n, \mathsf{P})$.

Note that if X is a $(\mathsf{P}, (\mathscr{F}_n))$-martingale, then X is also a $(\mathsf{P}, (\mathscr{G}_n))$-martingale with respect to each 'smaller' flow $(\mathscr{G}_n)$ (such that $\mathscr{G}_n \subseteq \mathscr{F}_n$), provided that the X_n are $\mathscr{G}_n$-measurable. Indeed, by the 'telescopic' property of conditional expectations we see that

$$\mathsf{E}(X_{n+1} \mid \mathscr{G}_n) = \mathsf{E}\big(\mathsf{E}(X_{n+1} \mid \mathscr{F}_n) \mid \mathscr{G}_n\big) = \mathsf{E}(X_n \mid \mathscr{G}_n) = X_n \quad (\mathsf{P}\text{-a.s.}),$$

which means the martingale property.

Clearly, if X is a $(\mathsf{P}, (\mathscr{F}_n))$-martingale, then there exists the 'minimal' flow $(\mathscr{G}_n)$ such that X is a $(\mathscr{G}_n)$-martingale: the 'natural' flow generated by X, i.e., $\mathscr{G}_n = \sigma(X_0, X_1, \dots, X_n)$.

We can recall in this connection that we defined a 'weakly efficient' market (in Chapter I, § 2a) as a market where the 'information flow $(\mathscr{F}_n)$' was generated by the *past values of the prices* off all the assets 'traded' in this market, so that $(\mathscr{F}_n)$ is just the 'minimal' flow on such a market.

3. One might well wonder if the theorem still holds for $d = \infty$ or $N = \infty$.

The following example, which is due to W. Schachermayer ([424]), shows that if $d = \infty$ (and $N = 1$), then there exists an 'arbitrage-free' market *without* a 'martingale' measure, so that the 'necessity' part of the above theorem fails in general for $d = \infty$.

EXAMPLE 1. Let $\Omega = \{1, 2, \dots\}$, let $\mathscr{F}_0 = \{\varnothing, \Omega\}$, let $\mathscr{F} = \mathscr{F}_1$ be the σ-algebra generated by all finite subsets of Ω, and let $\mathsf{P} = \sum_{k=1}^{\infty} 2^{-k}\delta_k$, i.e., $\mathsf{P}\{k\} = 2^{-k}$.

We define the sequence of prices $S = (S_n^i)$ for $i = 1, 2, \dots$ and $n = 0, 1$ as follows:

$$\Delta S_1^i(\omega) = \begin{cases} 1, & \omega = i, \\ -1, & \omega = i + 1, \\ 0 & \text{otherwise.} \end{cases}$$

Clearly, the corresponding (B, S)-market with $B_0 = B_1 = 1$ is arbitrage-free. However, marginal measures do not necessarily exist. In fact, the value $X_1^\pi(\omega)$ of an arbitrary portfolio can be represented as the sum

$$X_1^\pi = c_0 + \sum_{i=1}^{\infty} c_i S_1^i = X_0^\pi + \sum_{i=1}^{\infty} c_i \Delta S_1^i,$$

where $X_0^\pi = c_0 + \sum_{i=1}^{\infty} c_i$ (here we assume that $\sum |c_i| < \infty$). If $X_0^\pi = 0$ (i.e., $c_0 + \sum_{i=1}^{\infty} c_i = 0$), then by the condition $X_1^\pi \geqslant 0$ we obtain

$$X_1^\pi(1) = c_1 \geqslant 0, \ X_1^\pi(2) = c_2 - c_1 \geqslant 0, \ \dots, \ X_1^\pi(k) = c_k - c_{k-1} \geqslant 0, \ \dots.$$

Hence all the c_i are equal to zero, so that $X_1^\pi = 0$ (P-a.s.). However, a martingale measure *cannot* exist.

For assume that there *exists* a measure $\widetilde{\mathsf{P}} \sim \mathsf{P}$ such that S is a martingale with respect to it. Then for each $i = 1, 2, \dots$ we have

$$\mathsf{E}_{\widetilde{\mathsf{P}}} \Delta S_1^i = 0,$$

i.e., $\widetilde{\mathsf{P}}\{i\} = \widetilde{\mathsf{P}}\{i + 1\}$ for $i = 1, 2, \dots$. Clearly, this is impossible for a probability measure.

The next counterexample relates to the question of whether the 'sufficiency' part of the theorem holds for $N = \infty$. It shows that the existence of a martingale measure does not ensure that there is no arbitrage: there can be opportunities for arbitrage described below. (Note that the prices S in this counterexample are not necessarily positive, which makes this example appear somewhat deficient.)

EXAMPLE 2. Let $\xi = (\xi_n)_{n\geqslant 0}$ be a sequence of independent identically distributed random variables on $(\Omega, \mathscr{F}, \mathsf{P})$ such that $\mathsf{P}(\xi_n = 1) = \mathsf{P}(\xi_n = -1) = \frac{1}{2}$.

We set $S_0 = 0$, $S_n = \xi_1 + \cdots + \xi_n$, $B_n \equiv 1$, and let

$$X_n^\pi = \sum_{1\leqslant k\leqslant n} \gamma_k \Delta S_k \quad \biggl(= \sum_{1\leqslant k\leqslant n} \gamma_k \xi_k\biggr),$$

where

$$\gamma_k = \begin{cases} 2^{k-1} & \text{if } \xi_1 = \cdots = \xi_{k-1} = -1, \\ 0 & \text{otherwise.} \end{cases}$$

As is well known, we can treat X_n^π as the gain of a gambler playing against a 'symmetric' adversary, when the outcomes are described by the variables ξ_k (he gains if $\xi_k = 1$ and loses if $\xi_k = -1$) and he *doubles the stake* after a loss.

Clearly, if $\xi_1 = \cdots = \xi_k = -1$ (i.e., the gambler loses all the time), then his gain is

$$X_k^\pi = -\sum_{i=1}^{k} 2^{i-1} = -(2^k - 1),$$

i.e., he is a *net loser.*

However, if he gains at the next instant $k+1$, i.e., if $\xi_{k+1} = 1$, then his gain becomes

$$X_{k+1}^\pi = X_k^\pi + 2^k = -(2^k - 1) + 2^k = 1.$$

Hence if the concept of 'strategy' embraces (in addition to a choice of a portfolio) also a (random) stopping *instant* τ, then the gain of our gambler can be positive. For let

$$\tau = \inf\bigl\{k : X_k^\pi = 1\bigr\}.$$

Since $\mathsf{P}(\tau = k) = \bigl(\frac{1}{2}\bigr)^k$, it follows that $\mathsf{P}(\tau < \infty) = 1$, and therefore $\mathsf{E} X_\tau^\pi = 1$ because $\mathsf{P}(X_\tau^\pi = 1) = 1$ (although the starting capital X_0^π was zero).

Thus, there *exists* an opportunity for arbitrage on our (B, S)-market with $B_k \equiv 1$. Namely, there exists a portfolio π such that $X_0^\pi = 0$, but the expectation $\mathsf{E} X_\tau^\pi = 1$ for some τ.

We note by the way that the strategy of *doubling the stake* after a loss used here implies that either the gambler is immensely rich or he can borrow indefinitely from somewhere; both variants are, of course, hardly probable.

For this reason, in the considerations of various issues of arbitrage theory, one should impose some sound, 'economically reasonable' constraints on the classes of admissible strategies. (See Chapter VII, § 1a on this subject.)

§ 2c. Martingale Criterion of the Absence of Arbitrage. Proof of Sufficiency

We claim that if there exists a *martingale* measure $\widetilde{\mathsf{P}}$ equivalent to the measure P such that the sequence

$$\frac{S}{B} = \left(\frac{S_n}{B_n}\right)_{0 \leqslant n \leqslant N}$$

is a $(\widetilde{\mathsf{P}}, (\mathscr{F}_n))$-martingale, then there can be no *arbitrage* on the (B, S)-market.

As already mentioned (see the end of § 2a), the condition $B_n > 0$, $n \geqslant 0$ (which we assume to hold), allows one to set $B_n \equiv 1$.

We use formula (2) in § 2a, in accordance to which,

$$X_n^\pi = X_0^\pi + G_n^\pi \quad \text{and} \quad G_n^\pi = \sum_{k=1}^{n} \gamma_k \Delta S_k, \tag{1}$$

where $S = (S_n)$ is a $\widetilde{\mathsf{P}}$-martingale.

To prove the required assertion we must show that if $\pi \in SF$ is a strategy such that $X_0^\pi = 0$ and $\mathsf{P}(X_N^\pi \geqslant 0) = 1$, i.e.,

$$G_N^\pi \equiv \sum_{k=1}^{N} \gamma_k \Delta S_k \geqslant 0 \tag{2}$$

(P-a.s., or, equivalently, $\widetilde{\mathsf{P}}$-a.s.), then $G_N^\pi = 0$ (P-a.s., or, equivalently, $\widetilde{\mathsf{P}}$-a.s.).

We use the theorem and the lemma in Chapter II, § 1c.

The sequence $(G_n^\pi)_{0 \leqslant n \leqslant N}$ is a martingale transform with respect to $\widetilde{\mathsf{P}}$ and, therefore, by the theorem a local martingale. Since $G_N^\pi \geqslant 0$, it follows by the lemma that $(G_n^\pi)_{0 \leqslant n \leqslant N}$ is in fact a $\widetilde{\mathsf{P}}$-martingale and therefore, $\mathsf{E}_{\widetilde{\mathsf{P}}} G_N^\pi = G_0^\pi = 0$. Hence $X_N^\pi = G_N^\pi = 0$ ($\widetilde{\mathsf{P}}$-a.s. and P-a.s.).

§ 2d. Martingale Criterion of the Absence of Arbitrage. Proof of Necessity (by Means of the Esscher Conditional Transformation)

1. We must now prove that the absence of arbitrage means the *existence* of a probability measure $\widetilde{\mathsf{P}} \sim \mathsf{P}$ in $(\Omega, \mathscr{F})$ such that the sequence $S = (S_n)_{0 \leqslant n \leqslant N}$ is a $\widetilde{\mathsf{P}}$-martingale.

There exist several proofs of this result and its generalization to the continuous-time case (see, e.g., [92], [100], [171], [215], [259], [443], and [455]). They all appeal in one or another way to concepts and results of functional analysis (the Hahn–Banach theorem, separation in finite-dimensional Euclidean spaces, Hilbert spaces, etc.).

At the same time, they all are 'existence proofs' and suggest no *explicit* constructions of martingale measures, to say nothing about an explicit description of *all* martingale measures $\widetilde{\mathsf{P}}$ equivalent to P.

It would be interesting for this reason to find a proof of the necessity part of the theorem where one explicitly constructs all martingale measures or at least some subclass of them. This becomes important once we take into account that in the calculation of upper and lower prices we must find the upper and lower bounds over the class of *all* measures $\widetilde{\mathsf{P}}$ equivalent to the original measure P (see § 1c).

It is in this way of an *explicit* construction of a martingale measure that we shall carry out the proof. We shall follow the ideas of L. C. G. Rogers [407] and use the construction of equivalent measures based on the *Esscher conditional transformations*.

2. To explain the main idea we consider first the single-step model ($N = 1$), where we assume for simplicity that $d = 1$, $B_0 = B_1 = 1$, and $\mathscr{F}_0 = \{\varnothing, \Omega\}$. We also assume that $\mathsf{P}(S_1 \neq S_0) > 0$. Otherwise we shall obtain an uninteresting *trivial* market and we can take the original measure P as a martingale measure.

Now, each portfolio π is a pair of *numbers* (β, γ). If $X_0^\pi = 0$, then only the pairs such that $\beta + \gamma S_0 = 0$ are admissible.

The assumption of the absence of arbitrage means that the following two conditions must be met on such a (non-trivial) market:

$$\mathsf{P}(\Delta S_1 > 0) > 0 \quad \text{and} \quad \mathsf{P}(\Delta S_1 < 0) > 0. \tag{1}$$

Hence Fig. 54 in § 2a takes now the following form:

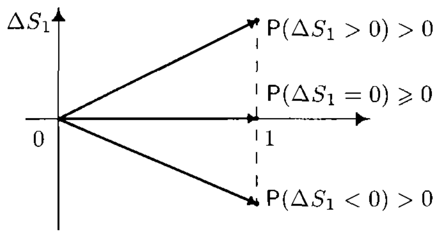


FIGURE 55. Typical arbitrage-free situation. Case $N = 1$

We must deduce from (1) that there exists a measure $\widetilde{\mathsf{P}} \sim \mathsf{P}$ such that

1) $\mathsf{E}_{\widetilde{\mathsf{P}}} |\Delta S_1| < \infty$,
2) $\mathsf{E}_{\widetilde{\mathsf{P}}} \Delta S_1 = 0$.

It can be useful to formulate the corresponding result with no mention of 'arbitrage', in the following, purely probabilistic form.

LEMMA 1. *Let X be a real-valued random variable with probability distribution* P *on* $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ *such that*

$$\mathsf{P}(X > 0) > 0 \quad \textit{and} \quad \mathsf{P}(X < 0) > 0. \tag{2}$$

Then there exists a measure $\widetilde{\mathsf{P}} \sim \mathsf{P}$ *such that*

$$\mathsf{E}_{\widetilde{\mathsf{P}}} e^{aX} < \infty \tag{3}$$

for each $a \in \mathbb{R}$; *in particular*,

$$\mathsf{E}_{\widetilde{\mathsf{P}}} |X| < \infty. \tag{4}$$

Moreover, $\widetilde{\mathsf{P}}$ *has the following property*:

$$\mathsf{E}_{\widetilde{\mathsf{P}}} X = 0. \tag{5}$$

Proof. Given the measure P, we construct first the probability measure

$$\mathsf{Q}(dx) = c\, e^{-x^2} \mathsf{P}(dx), \qquad x \in \mathbb{R},$$

where c is a normalizing coefficient, i.e.,

$$c^{-1} = \mathsf{E}_{\mathsf{P}} e^{-X^2}.$$

Let

$$\varphi(a) = \mathsf{E}_{\mathsf{Q}} e^{aX} \tag{6}$$

for $a \in \mathbb{R}$ and let

$$Z_a(x) = \frac{e^{ax}}{\varphi(a)}. \tag{7}$$

Clearly, $\mathsf{Q} \sim \mathsf{P}$ and it follows by the construction of Q that $\varphi(a) < \infty$ for each $a \in \mathbb{R}$ that $\varphi(a) > 0$.

It is equally clear that $Z_a(x) > 0$ and $\mathsf{E}_{\mathsf{Q}} Z_a(x) = 1$. Hence for *each* $a \in \mathbb{R}$ we can define the *probability* measure

$$\widetilde{\mathsf{P}}_a(dx) = Z_a(x) \mathsf{Q}(dx) \tag{8}$$

such that $\widetilde{\mathsf{P}}_a \sim \mathsf{Q} \sim \mathsf{P}$.

The function $\varphi = \varphi(a)$ defined for $a \in \mathbb{R}$ is *strictly convex* (downwards) since $\varphi''(a) > 0$.

Let

$$\varphi_* = \inf\{\varphi(a) \colon a \in \mathbb{R}\}.$$

Then two cases are possible:

1) there exists a_* such that $\varphi(a_*) = \varphi_*$

or

2) there exists no such a_*.

In the first case we clearly have $\varphi'(a_*) = 0$ and

$$\mathsf{E}_{\widetilde{\mathsf{P}}_{a_*}} X = \mathsf{E}_{\mathsf{Q}} \frac{X e^{a_* X}}{\varphi(a_*)} = \frac{\varphi'(a_*)}{\varphi(a_*)} = 0.$$

Thus, we can take $\widetilde{\mathsf{P}}_{a^*}$ as the required measure $\widetilde{\mathsf{P}}$ in case 1).

We now claim that assumption (2) rules out case 2).

Indeed, let $\{a_n\}$ be a sequence such that

$$\varphi_* < \varphi(a_n) \downarrow \varphi_*. \tag{9}$$

This sequence must approach $+\infty$ or $-\infty$ since otherwise we can choose a convergent subsequence and the minimum value is attained at a finite point, which contradicts assumption 2).

Let $u_n = \dfrac{a_n}{|a_n|}$ and let $u = \lim u_n$ $(= \pm 1)$.

By (2) we obtain

$$\mathsf{Q}(uX > 0) > 0.$$

Hence there exists $\delta > 0$ such that

$$\mathsf{Q}(uX > \delta) = \varepsilon > 0, \tag{10}$$

and we choose δ that is a continuity point of Q, i.e.,

$$\mathsf{Q}(uX = \delta) = 0.$$

Consequently,

$$\mathsf{Q}(a_n X > \delta |a_n|) = \mathsf{Q}(u_n X > \delta) \to \varepsilon \quad \text{as} \quad n \to \infty,$$

so that for n sufficiently large we have

$$\varphi(a_n) = \mathsf{E}_{\mathsf{Q}} e^{a_n X} \geqslant \mathsf{E}_{\mathsf{Q}} \left[e^{a_n X} I(a_n X > \delta |a_n|) \right] \geqslant \frac{1}{2} \varepsilon \cdot \exp(\delta |a_n|) \to \infty,$$

which contradicts (9), where $\varphi_* \leqslant 1$.

Remark. The above method of the construction of probability measures $\widetilde{\mathsf{P}}_a$, which is based on the Esscher transformation $x \leadsto \dfrac{e^{ax}}{\varphi(a)}$ defined by (7) is a known in the actuarial practice since F. Esscher [144] (1932). As regards the applications of this transformation in financial mathematics, see, for instance, [177] and [178], and as regards its applications in actuarial mathematics, see the book [52].

3. It is easy to see from the above proof of the lemma how one can generalize it to a vector-valued case, when one considers in place of X an ordered sequence $(X_0, X_1, \dots, X_N)$ of $\mathscr{F}_n$-measurable random variables X_n, $0 \leqslant n \leqslant N$, defined on a filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_n), \mathsf{P})$ with $\mathscr{F}_0 = \{\varnothing, \Omega\}$ and $\mathscr{F}_N = \mathscr{F}$.

LEMMA 2. *Assume that*

$$\mathsf{P}(X_n > 0 \,|\, \mathscr{F}_{n-1}) > 0 \quad \textit{and} \quad \mathsf{P}(X_n < 0 \,|\, \mathscr{F}_{n-1}) > 0 \tag{11}$$

for $1 \leqslant n \leqslant N$. *Then there exists a probability measure* $\widetilde{\mathsf{P}}$ *equivalent to* P *in the space* $(\Omega, \mathscr{F})$ *such that the sequence* $(X_0, X_1, \dots, X_N)$ *is a* $(\widetilde{\mathsf{P}}, (\mathscr{F}_n))$*-martingale difference.*

Proof. If necessary, we can proceed first from P to a new measure Q such that

$$\mathsf{Q}(d\omega) = c \exp\left\{ -\sum_{i=0}^{N} X_i^2(\omega) \right\} \mathsf{P}(d\omega) \tag{12}$$

and the generating function $\mathsf{E_Q} \exp\left\{ \sum\limits_{i=0}^{N} a_i X_i \right\}$ is finite.

We can construct the required measure $\widetilde{\mathsf{P}} = \widetilde{\mathsf{P}}(d\omega)$ as follows (cf. (8)).

Let

$$\varphi_n(a;\omega) = \mathsf{E}(e^{aX_n} \,|\, \mathscr{F}_{n-1})(\omega). \tag{13}$$

For each fixed ω these functions (in view of (11)) are strictly (downwards) convex in a. As in Lemma 1 we can show that there exists a *unique* (finite) value $a_n = a_n(\omega)$ such that the smallest value $\inf\limits_a \varphi_n(a;\omega)$ is attained at a_n.

Since here $\inf\limits_a = \inf\limits_{a \in \mathbb{Q}}$, where $\mathbb{Q}$ is the set of rationals, the function $\varphi_n(\omega) = \inf\limits_a \varphi_n(a;\omega)$ is $\mathscr{F}_{n-1}$-measurable, which means that $a_n(\omega)$ is also a $\mathscr{F}_{n-1}$-measurable function.

Indeed, if $[A, B]$ is a closed interval, then

$$\{\omega : a_n(\omega) \in [A, B]\} = \bigcap_m \bigcup_{a \in \mathbb{Q} \cap [A, B]} \left\{ \omega : \varphi_n(a;\omega) < \varphi_n(\omega) + \frac{1}{m} \right\} \in \mathscr{F}_{n-1},$$

and therefore $a_n(\omega)$ is $\mathscr{F}_{n-1}$-measurable.

Next, we define recursively a sequence $Z_0, Z_1(\omega), \dots, Z_N(\omega)$ by setting $Z_0 = 1$ and

$$Z_n(\omega) = Z_{n-1}(\omega) \frac{\exp\{a_n(\omega) X_n(\omega)\}}{\mathsf{E_Q}(\exp\{a_n X_n\} \,|\, \mathscr{F}_{n-1})(\omega)} \tag{14}$$

for $n \geqslant 1$. Clearly, the variables $Z_n(\omega)$ are $\mathscr{F}_n$-measurable and form a martingale:

$$\mathsf{E_Q}(Z_n \,|\, \mathscr{F}_{n-1}) = Z_{n-1} \quad (\mathsf{P}\text{-a.s.}).$$

We define the required measure $\widetilde{\mathsf{P}}$ by the formula

$$\widetilde{\mathsf{P}}(d\omega) = Z_N(\omega)\,\mathsf{P}(d\omega). \tag{15}$$

As in Lemma 1, it easily follows from this definition that $\mathsf{E}_{\widetilde{\mathsf{P}}}|X_n| < \infty$, $0 \leqslant n \leqslant N$, and

$$\mathsf{E}_{\widetilde{\mathsf{P}}}(X_n \,|\, \mathscr{F}_{n-1}) = 0, \qquad 1 \leqslant n \leqslant N. \tag{16}$$

By Lemma 1, $\mathsf{E}_{\widetilde{\mathsf{P}}} X_0 = 0$. Hence the sequence $(X_0, X_1, \dots, X_N)$ is a martingale difference with respect to $\widetilde{\mathsf{P}}$, which proves the lemma.

4. For $d = 1$, the *necessity* of the existence of a martingale measure $\widetilde{\mathsf{P}} \sim \mathsf{P}$ (in an arbitrage-free market) is a consequence of Lemma 2. For let $X_0 = S_0$, $X_1 = \Delta S_1, \dots, X_N = \Delta S_N$. Since there can be no arbitrage, we can assume without loss of generality that

$$\mathsf{P}(\Delta S_n > 0 \,|\, \mathscr{F}_{n-1}) > 0 \quad \text{and} \quad \mathsf{P}(\Delta S_n < 0 \,|\, \mathscr{F}_{n-1}) > 0 \tag{17}$$

for each $n = 1, \dots, N$.

Indeed, if $\mathsf{P}(\Delta S_n = 0) = 1$ for some n, then we can skip the instant n because no contribution to the value X_N^π of an arbitrary self-financing portfolio π can be made at this time.

On the other hand if there exists n such that

$$\mathsf{P}(\Delta S_n \geqslant 0) = 1$$

or $\mathsf{P}(\Delta S_n \leqslant 0) = 1$, then $\mathsf{P}(\Delta S_n = 0) = 1$ due to the absence of arbitrage. (Otherwise it is easy to construct a strategy π such that $X_N^\pi > 0$ with positive probability.) Again, the corresponding contribution to X_N^π is zero.

Thus, we can assume that (17) holds for all $n \leqslant N$, and the required *necessity* follows directly by Lemmas 1 and 2 as applied to $X_0 = S_0$ and $X_n = \Delta S_n$, $1 \leqslant n \leqslant N$.

5. We consider now the general case of $d \geqslant 1$. Conceptually, the proof is the same as for $d = 1$; it can be carried out using the following generalization of Lemma 2 to the vector-valued case.

LEMMA 3. *Let $(X_0, X_1, \dots, X_N)$ be a sequence of $\mathscr{F}_n$-measurable d-vectors*

$$X_n = \begin{pmatrix} X_n^1 \\ \vdots \\ X_n^d \end{pmatrix}, \qquad 0 \leqslant n \leqslant N,$$

defined on a filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_n)_{0 \leqslant n \leqslant N}, \mathsf{P})$ with $\mathscr{F}_0 = \{\varnothing, \Omega\}$, $\mathscr{F}_N = \mathscr{F}$.

Assume also that if

$$\gamma_n = \begin{pmatrix} \gamma_n^1 \\ \vdots \\ \gamma_n^d \end{pmatrix}$$

is a non-zero $\mathscr{F}_{n-1}$-measurable vector-valued variable with bounded component ($|\gamma_n^i(\omega)| \leqslant c < \infty$, $\omega \in \Omega$) such that

$$\mathsf{P}\big((\gamma_n, X_n) > 0 \,|\, \mathscr{F}_{n-1}\big) > 0 \quad (\mathsf{P}\text{-a.s.}),$$

then also

$$\mathsf{P}\big((\gamma_n, X_n) < 0 \,|\, \mathscr{F}_{n-1}\big) > 0 \quad (\mathsf{P}\text{-a.s.}),$$

where (γ_n, X_n) is the scalar product.

Then there exists a probability measure $\widetilde{\mathsf{P}}$ equivalent to P on $(\Omega, \mathscr{F})$ such that the sequence $(X_0, X_1, \dots, X_N)$ is a d-dimensional martingale difference with respect to $\widetilde{\mathsf{P}}$: $\mathsf{E}_{\widetilde{\mathsf{P}}}|X_n| < \infty$, $\mathsf{E}_{\widetilde{\mathsf{P}}} X_0 = 0$, and $\mathsf{E}_{\widetilde{\mathsf{P}}}(X_n \,|\, \mathscr{F}_{n-1}) = 0$, $1 \leqslant n \leqslant N$.

Proof. If the topological *support* (i.e., the smallest closed carrier) of the regular conditional probability $\mathsf{P}(X_n \in \cdot \,|\, \mathscr{F}_{n-1})(\omega)$ *does not lie* in a proper subspace of $\mathbb{R}^d$, then, as for $d = 1$, the functions

$$\varphi_n(a;\omega) = \mathsf{E}(e^{(a, X_n)} \,|\, \mathscr{F}_{n-1})(\omega), \qquad a \in \mathbb{R}^d,$$

are strictly convex and the smallest value $\inf \varphi_n(a;\omega)$ is attained at a unique point $a_n = a_n(\omega) \in \mathbb{R}^d$; moreover, the functions $a_n(\omega)$ are $\mathscr{F}_{n-1}$-measurable.

The case of a regular conditional distribution $\mathsf{P}(X_n \in \cdot \,|\, \mathscr{F}_{n-1})(\omega)$ concentrated at a proper subspace of $\mathbb{R}^d$ is slightly more delicate. As shown in [407], we can find in this case a unique $\mathscr{F}_{n-1}$-measurable variable $a_n = a_n(\omega)$ delivering the smallest value $\inf \varphi_n(a;\omega)$.

The required measure $\widetilde{\mathsf{P}}$ can be constructed as in (15) and (14) if we treat $a_n(\omega)X_n(\omega)$ (in (14)) as the scalar product of the vectors $a_n(\omega)$ and $X_n(\omega)$.

6. The above construction of a martingale measure based on the *Esscher conditional transformation* gives us only *one* particular measure, although the class of martingale measures equivalent to the original one can be more rich. We devote the next section to several approaches based of the *Girsanov transformation* that can be used in the construction of a family of measures $\widetilde{\mathsf{P}}$ equivalent to or absolutely continuous with respect to P such that the sequence of the discounted prices is a martingale with respect to these measures.

The Esscher transformation has long been in use in actuarial calculations (see, e.g. [52]). 'Esscher transformations' are less familiar under this name in *financial mathematics* (see nevertheless the already mentioned papers [177] and [178]), where several I. V. Girsanov's results on changes of measures known as 'Girsanov's theorem' play an important role.

In fact, these transformations have very much in common and we shall discuss this in detail in § 3 of this chapter. Here we only mention in connection with Lemma 1 that if X is a normally distributed random variable ($X \sim \mathscr{N}(0,1)$), then $\varphi(a) = \mathsf{E}_{\mathsf{P}} e^{aX} = e^{\frac{1}{2}a^2}$ and (see (7))

$$Z_a(x) = \frac{e^{ax}}{\varphi(a)} = e^{ax - \frac{1}{2}a^2}.$$

A reader acquainted with Girsanov's theorem will see at once that the 'Girsanov' exponential $e^{ax-\frac{1}{2}a^2}$ involved in this theorem (see § 3a) is just the Esscher transform (7).

§ 2e. Extended Version of the First Fundamental Theorem

1. Let $\mathscr{P}(\mathsf{P})$ and $\mathscr{P}_{\mathrm{loc}}(\mathsf{P})$ be the sets of all probability measures $\widetilde{\mathsf{P}} \sim \mathsf{P}$ such that the discounted prices

$$\frac{S}{B} = \left(\frac{S_n}{B_n}\right)_{0 \leqslant n \leqslant N}$$

are *martingales* and *local martingales*, respectively, relative to these measures.

Let $\mathscr{P}_b(\mathsf{P})$ be the set of measures $\widetilde{\mathsf{P}}$ in $\mathscr{P}(\mathsf{P})$ such that the Radon–Nikodym derivatives $\dfrac{d\widetilde{\mathsf{P}}}{d\mathsf{P}}$ are bounded above: $\dfrac{d\widetilde{\mathsf{P}}}{d\mathsf{P}}(\omega) \leqslant C(\widetilde{\mathsf{P}})$ (P-a.s.) for some constant $C(\widetilde{\mathsf{P}})$.

We can formulate Theorem A (the First fundamental theorem in § 2b) as follows: *the conditions*

(i) *a (B, S)-market is arbitrage-free*

and

(ii) *the set of martingale measures $\mathscr{P}(\mathsf{P})$ is not empty* ($\mathscr{P}(\mathsf{P}) \neq \varnothing$)

are equivalent.

Theorem A* below is a natural generalization of this version of the first fundamental theorem; it provides several equivalent characterizations of an arbitrage-free market and clears up the structure of the set of martingale measures. (We state and prove this theorem following [251].)

First of all, we introduce our notation.

Let $\mathsf{Q} = \mathsf{Q}(dx)$ be a probability measure in $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$ and let

$K(\mathsf{Q})$ be the *topological support* of Q (the smallest closed set carrying Q; [335; vol. 5]);

$L(\mathsf{Q})$ be the *closed convex hull* of $K(\mathsf{Q})$;

$H(\mathsf{Q})$ be the *smallest affine hyperplane* containing $K(\mathsf{Q})$ (clearly, $L(\mathsf{Q}) \subseteq H(\mathsf{Q})$);

$L^\circ(\mathsf{Q})$ be the *relative interior* of $L(\mathsf{Q})$ (in the topology of the hyperplane $H(\mathsf{Q})$).

For instance, if Q is concentrated at a *single* point a, then $H(\mathsf{Q})$ coincides with this point and $L(\mathsf{Q}) = L^\circ(\mathsf{Q}) = \{a\}$. Otherwise $H(\mathsf{Q})$ has dimension between 1 and d. If the dimension of $H(\mathsf{Q})$ is 1, then $L(\mathsf{Q})$ is a closed line segment and $L^\circ(\mathsf{Q})$ is an open interval.

Let $\mathsf{Q}_n(\omega, \cdot)$ and $\overline{\mathsf{Q}}_n(\omega, \cdot)$ be the regular conditional distributions

$$\mathsf{P}(\Delta S_n \in \cdot \mid \mathscr{F}_{n-1})(\omega) \quad \text{and} \quad \mathsf{P}\biggl(\Delta\Bigl(\frac{S_n}{B_n}\Bigr) \in \cdot \Bigm| \mathscr{F}_{n-1}\biggr)(\omega), \qquad 1 \leqslant n \leqslant N.$$

We note that since $B_n > 0$, $0 \leqslant n \leqslant N$, and the B_n are $\mathscr{F}_{n-1}$-measurable, the sets $K(\mathsf{Q})$, $L(\mathsf{Q})$, and $L^\circ(\mathsf{Q})$ are the same for $\mathsf{Q} = \mathsf{Q}_n$ and $\mathsf{Q} = \overline{\mathsf{Q}}_n$. Hence we shall assume without loss of generality in the statements and proofs below that $B_n \equiv 1$, $n \leqslant N$.

THEOREM A* (an extended version of the *First fundamental theorem*). *Assume that the conditions of Theorem* A *are satisfied. Then the following assertions are equivalent:*

a) *a (B,S)-market is arbitrage-free;*
a′) *a (B,S)-market is weakly arbitrage-free;*
a″) *a (B,S)-market is strongly arbitrage-free;*
b) $\mathscr{P}_b(\mathsf{P}) \neq \varnothing$;
c) $\mathscr{P}(\mathsf{P}) \neq \varnothing$;
d) $\mathscr{P}_{\mathrm{loc}}(\mathsf{P}) \neq \varnothing$;
e) $0 \in L^\circ(\overline{\mathsf{Q}}_n(\omega, \cdot))$ *for each* $n \in \{1, 2, \dots, N\}$ *and* P*-almost all* $\omega \in \Omega$.

First of all, we make several general observations concerning the above assertions.

As regards the definitions of arbitrage-free and weakly or strongly arbitrage-free markets, see § 2a.

It follows by the lemma in Chapter II, § 1c that, in fact, $\mathscr{P}(\mathsf{P}) = \mathscr{P}_{\mathrm{loc}}(\mathsf{P})$, and therefore

$$\text{c)} \iff \text{d)}.$$

Further, if the properties a), a′), a″, c), or e) hold with respect to some measure $\overline{\mathsf{P}}$ equivalent to P, then they also hold with respect to P. If b) holds with respect to a measure $\overline{\mathsf{P}} \sim \mathsf{P}$ (i.e., $\mathscr{P}_b(\overline{\mathsf{P}}) \neq \varnothing$) and the derivative $\dfrac{d\overline{\mathsf{P}}}{d\mathsf{P}}$ is bounded, then b) holds also for the original measure P.

We observe next that we can always find a measure $\overline{\mathsf{P}} \sim \mathsf{P}$ such that all the variables S_n, $n \leqslant N$, are integrable with respect to it and the derivative $\dfrac{d\overline{\mathsf{P}}}{d\mathsf{P}}$ is bounded. For instance, it suffices to set

$$d\overline{\mathsf{P}} = C \exp\biggl(-\sum_{i=1}^{d} \sum_{n=1}^{N} |S_n^i|\biggr) d\mathsf{P},$$

where C is a normalizing coefficient.

Hence we can assume throughout the proof of the theorem that $\mathsf{E}|S_n| < \infty$, $n \leqslant N$, for the original measure P.

Then, in view of the obvious implications

$$\text{a}'') \implies \text{a}) \implies \text{a}')$$

and

$$\text{b}) \implies \text{c}),$$

we must prove only that

$$\text{a}') \implies \text{e}) \implies \text{b})$$

and

$$\text{c}) \implies \text{a}'').$$

2. We introduce now several concepts that are necessary for the proof of these three implications and state two auxiliary results (Lemmas 1 and 2).

Let $\mathsf{Q} = \mathsf{Q}(dx)$ be a measure in $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$ such that

$$\int_{\mathbb{R}^d} |x| \, \mathsf{Q}(dx) < \infty. \tag{1}$$

(we shall set Q to be equal to the regular conditional distributions $\mathsf{Q}_n(\omega, dx)$ with $n \leqslant N$ in what follows, and (1) will hold (P-a.s.) in view of the assumption that $\mathsf{E}|S_n| < \infty$ for $n \leqslant N$.)

Let $x' = (x, v) \in E = \mathbb{R}^d \times (0, \infty)$ and let $\mathsf{Q}' = \mathsf{Q}'(dx')$ be the measure on the Borel σ-algebra $\mathscr{E}$ of E associated with $\mathsf{Q} = \mathsf{Q}(dx)$ in the following sense: Q is the 'first marginal' of Q', that is, $\mathsf{Q}(dx) = \mathsf{Q}'(dx, (0, \infty))$.

Let $Z(\mathsf{Q}')$ be the family of *positive* Borel functions $z = z(x, v)$ in E such that

$$\int_E |x| z(x, v) \, \mathsf{Q}'(dx; dv) < \infty, \tag{2}$$

$$\int_E z(x, v) \, \mathsf{Q}'(dx; dv) = 1, \tag{3}$$

and the functions $vz(x, v)$ are bounded.

Let $B(a, \varepsilon)$ be the closed ball in $\mathbb{R}^d$ with center at a and of radius ε. Let G be the collection of all families

$$g = \bigl(k, (a_i, \varepsilon_i, \alpha_i, \alpha_i')_{i=1,\dots,k}\bigr) \tag{4}$$

such that $k \in \{1, 2, \dots\}$, $a_i \in \mathbb{R}^d$, $\varepsilon_i > 0$, $\alpha_i > 0$, $\alpha_i' > 0$.

With each $g \in G$ we associate the positive function

$$z_g(x, v) = \frac{1}{\max(v, 1)} \sum_{i=1}^{k} \Bigl[\alpha_i I_{B(a_i, \varepsilon_i)}(x) + \alpha_i' I_{B^c(a_i, \varepsilon_i)}(x)\Bigr] \tag{5}$$

(on E), where $B^c = \mathbb{R}^d \setminus B$.

If $\mathsf{E}_{\mathsf{Q}'} z_g = 1$, then (1) shows that $z_g \in Z(\mathsf{Q}')$. Thus, we can define for such functions z_g the vectors

$$\varphi(g) = \int_E x z_g(x, v)\, \mathsf{Q}'(dx, dv) \tag{6}$$

(the barycenters).

We set

$$\Phi(\mathsf{Q}') = \{\varphi(g) \colon g \in G, \mathsf{E}_{\mathsf{Q}'} z_g = 1\}. \tag{7}$$

LEMMA 1. *We have the following relations:*

$$L(\mathsf{Q}) = \overline{\Phi}(\mathsf{Q}'), \tag{8}$$

$$L^\circ(\mathsf{Q}) \subseteq \Phi(\mathsf{Q}'). \tag{9}$$

If $0 \neq L^\circ(\mathsf{Q})$, *then there exists* γ *in* $\mathbb{R}^d$ *such that*

$$\mathsf{Q}\big(x \colon (\gamma, x) \geqslant 0\big) = 1 \quad \textit{and} \quad \mathsf{Q}\big(x \colon (\gamma, x) > 0\big) > 0. \tag{10}$$

Proof. First we shall prove that $L(\mathsf{Q}) \subseteq \overline{\Phi}(\mathsf{Q}')$.

Let $y \in K(\mathsf{Q})$. We claim that there exists a sequence y_n in $\Phi(\mathsf{Q}')$ such that $y_n \to y$. In other words, each point y in the topological support of Q is the limit of some sequence $y_n = \varphi(g_n)$, where $g_n \in G$, $n \geqslant 1$.

Let $A_n = B\big(y, \frac{1}{n}\big)$ be a ball of radius $1/n$ with center at y. We set

$$a_n = \mathsf{Q}''(A_n), \qquad a_n^c = \mathsf{Q}''(A_n^c)$$

and

$$b_n = \int_{A_n} x\, \mathsf{Q}''(dx), \qquad b_n' = \int_{A_n^c} x\, \mathsf{Q}''(dx),$$

where

$$\mathsf{Q}''(A) = \int_E I_A \frac{1}{\max(v, 1)}\, \mathsf{Q}'(dx, dv), \quad A \in \mathscr{B}(\mathbb{R}^d). \tag{11}$$

We choose δ_n such that $\delta_n a_n + \dfrac{a_n^c}{n} = 1$.

Since $y \in K(\mathsf{Q})$, it follows that $a_n > 0$. It is also clear that $\sup_n a_n' < \infty$. A fortiori, $\delta_n > 0$ (for large n at any rate).

Consequently, if $g_n = \big(1, \big(y, \frac{1}{n}, \delta_n, \frac{1}{n}\big)\big)$, then

$$\mathsf{E}_{\mathsf{Q}'} z_{g_n} = \delta_n a_n + \frac{a_n^c}{n} = 1 \tag{12}$$

for n large.

We set $y_n = \varphi(g_n)$. Then $y_n = \delta_n b_n + \dfrac{b_n^c}{n}$, where $\sup|b_n^c| < \infty$.

Since $\delta_n a_n \to 1$ and $b_n - y a_n \to 0$, it follows that $y_n \to y$, and therefore $K(\mathsf{Q})$ lies in the closure $\overline{\Phi}(\mathsf{Q}')$ of the set $\Phi(\mathsf{Q}')$:

$$K(\mathsf{Q}) \subseteq \overline{\Phi}(\mathsf{Q}'). \tag{13}$$

Now, note that $\Phi(\mathsf{Q}')$ is a convex set. Hence it follows from (13) that

$$L(\mathsf{Q}) \subseteq \overline{\Phi}(\mathsf{Q}'). \tag{14}$$

On the other hand each point $y \in \Phi(\mathsf{Q}')$ is the barycenter of a probability measure equivalent to Q (see (6), (7), and (11)), so that y lies in the closure of the convex hull $L(\mathsf{Q})$ of $K(\mathsf{Q})$.

Hence $\Phi(\mathsf{Q}') \subseteq L(\mathsf{Q})$, and therefore $\overline{\Phi}(\mathsf{Q}') \subseteq L(\mathsf{Q})$; together with (14) this shows that $L(\mathsf{Q}) = \overline{\Phi}(\mathsf{Q}')$.

We now use the fact that a convex set contains the interior of its closure. For the convex set $\Phi(\mathsf{Q}')$ (considered in the topology of $H(\mathsf{Q})$) this means that $L^\circ(\mathsf{Q}) \subseteq V(\mathsf{Q}')$.

Thus, we have proved (8) and (9).

We proceed now to the proof of the second assertion of the lemma.

Let $0 \notin L^\circ(\mathsf{Q})$. Then there are three possible cases, which can be treated using the standard separation machinery of convex analysis; see, e.g., [406].

The first case: $0 \notin H(\mathsf{Q})$. In this case let γ be a vector directed towards the affine hyperplane $H(\mathsf{Q})$ and orthogonal to it. Then $(\gamma, x) > 0$ for $x \in H(\mathsf{Q})$, which, of course, means that $\mathsf{Q}(x\colon (\gamma, x) > 0) = 1$.

The second case: $0 \in H(\mathsf{Q})$, but $0 \notin L(\mathsf{Q})$. Then there clearly exists a vector $\gamma \in H(\mathsf{Q})$ such that $(\gamma, x) > 0$ for all $x \in L(\mathsf{Q})$, so that $\mathsf{Q}(x\colon (\gamma, x) > 0) = 1$.

The third case: $0 \in H(\mathsf{Q})$, but $0 \in L(\mathsf{Q}) \setminus L^\circ(\mathsf{Q})$. Then both subsets $L(\mathsf{Q})$ and $K(\mathsf{Q})$ of the plane $H(\mathsf{Q})$ of some dimension q lie to one side of some $(q-1)$-dimensional hyperplane H' of $H(\mathsf{Q})$ containing 0. (If $q = 1$, then H' is reduced to the point $\{0\}$.) By definition, $H(\mathsf{Q})$ is the smallest affine plane containing $K(\mathsf{Q})$. Hence $K(\mathsf{Q})$ does not lie in H' and γ can be constructed as follows.

We consider an arbitrary non-zero vector γ in $H(\mathsf{Q})$ that is orthogonal to H and satisfies the inequality $(\gamma, x) \geqslant 0$ for each $x \in L(\mathsf{Q})$, so that $\mathsf{Q}(x\colon (\gamma, x) \geqslant 0) = 1$.

Next, we note that there exists $x \in K(\mathsf{Q})$ such that $(\gamma, x) > 0$. Bearing in mind that $K(\mathsf{Q})$ is the topological support of Q we see that $\mathsf{Q}(x\colon (\gamma, x) > 0) > 0$. This completes the proof of Lemma 1.

3. The next result that we shall require relates to the problem of the existence of a '*measurable selection*'.

LEMMA 2. *Let $(E, \mathscr{E}, \mu)$ be a probability space and let $(G, \mathscr{G})$ be a Polish space with Borel σ-algebra $\mathscr{G}$. Let A be a $\mathscr{E} \otimes \mathscr{G}$-measurable subset of $E \times G$. Then there exists a G-valued $\mathscr{E}/\mathscr{G}$-measurable function (a selector) $Y = Y(x)$, $x \in E$, such that $(x, Y(x)) \in A$ for μ-almost all x in the E-projection*

$$\pi(A) = \{x \colon (x, y) \in A \text{ for some } y \in G\}.$$

Remark 1. Here we mean by a *Polish* space a *separable topological space* that can be equipped with a metric consistent with the topology and making it a *complete* metric space.

In the literature devoted to measurable selection (see, e.g., [11]) one can find various versions of results on the existence of measurable selectors under various assumptions about the measurable spaces $(E, \mathscr{E})$ and $(G, \mathscr{G})$. The most easy way for us is to refer, e.g., to [102; Chapter III, Theorem 82] or [11; Appendix I, Theorem 1], where one can find the following result (in a slightly modified form).

PROPOSITION. *Let $(E, \mathscr{E})$ be an arbitrary measurable space and let $(G, \mathscr{G})$ be a Polish space. If $A \in \mathscr{E} \otimes \mathscr{G}$, then there exists a universally measurable function $\widehat{Y} = \widehat{Y}(x)$, $x \in E$, such that $(x, \widehat{Y}(x)) \in A$ for all $x \in \pi(A)$.*

(We recall that a G-valued function $\widehat{Y} = \widehat{Y}(x)$ in $(E, \mathscr{E})$ is said to be *universally measurable* if it is $\widehat{\mathscr{E}}/\mathscr{G}$-measurable, where $\widehat{\mathscr{E}} = \bigcap_\mu \mathscr{E}_\mu$ is the intersection of all the σ-algebras $\mathscr{E}_\mu$ that are the completions of $\mathscr{E}$ with respect to all finite measures μ in $(E, \mathscr{E})$.)

Thus, it follows from the above proposition that there *exists* a $\widehat{\mathscr{E}}/\mathscr{G}$-measurable function $\widehat{Y} = \widehat{Y}(x)$, $x \in E$, such that $(x, \widehat{Y}(x)) \in A$ for *all* $x \in \pi(A)$.

Since $\widehat{\mathscr{E}} = \bigcap_\mu \mathscr{E}_\mu$, the function $\widehat{Y}$ is, of course, $\mathscr{E}_\mu/\mathscr{G}$-measurable for each finite measure μ. Using the fact that $(G, \mathscr{G})$ is a Polish space and the σ-algebra $\mathscr{E}_\mu$ is the completion of $\mathscr{E}$ with respect to μ, it is easy to conclude (approximating $\widehat{Y}$ by simple functions) that there exists a $\mathscr{E}/\mathscr{G}$-measurable function Y such that $Y = \widehat{Y}$ (μ-a.s.).

Hence Lemma 2 is a consequence of the above proposition.

4. *Proof* a$'$) $\Longrightarrow$ e). Assume that the market is arbitrage-free, but e) fails. Then there exist $n \in \{1, 2, \dots, N\}$ and a $\mathscr{F}_{n-1}$-measurable set B with $\mathsf{P}(B) > 0$ such that 0 lies outside $L^\circ(\mathsf{Q}_n(\omega, \cdot\,))$ for almost all $\omega \in B$.

The set

$$A = \Big\{(\omega, \gamma) \in \Omega \times \mathbb{R}^d \colon \omega \in B,\ \mathsf{Q}_n\big(\omega, \{x \colon (\gamma, x) \geqslant 0\}\big) = 1, \\ \mathsf{Q}_n\big(\omega, \{x \colon (\gamma, x) > 0\}\big) > 0\Big\}$$

is $\mathscr{F}_{n-1} \otimes \mathscr{B}(\mathbb{R}^d)$-measurable and B is its projection $\pi(A)$ by the last assertion in Lemma 1.

By Lemma 2 on measurable selection there exists a $\mathscr{F}_{n-1}$-measurable vector $g = g(\omega)$ such that

$$\mathsf{Q}_n\big(\omega, \{x\colon (g(\omega), x) \geqslant 0\}\big) = 1, \qquad \mathsf{Q}_n\big(\omega, \{x\colon (g(\omega), x) > 0\}\big) > 0$$

for P-almost all $\omega \in B$.

Since B is a $\mathscr{F}_{n-1}$-measurable set, the function $\widetilde{g}(\omega) = g(\omega) I_B(\omega)$ is $\mathscr{F}_{n-1}$-measurable and, in accordance with (15), $(\widetilde{g}, \Delta S_n) \geqslant 0$ (P-a.s.) and $\mathsf{P}((\widetilde{g}, \Delta S_n) > 0) > 0$.

We set $\gamma_i = \widetilde{g} I_{i=n}$, $i = 1, \dots, N$, and consider a *self-financing* strategy $\pi = (\beta, \gamma)$ of value X^π satisfying the equalities $X_0^\pi = 0$ and $\Delta X_i^\pi = (\gamma_i, \Delta S_i)$. (We determine the parameters β_i—e.g., in the case of $B_i \equiv 1$—from the requirements that $X_i^\pi = (\gamma_i, S_i) + \beta_i$ must be equal to $\sum\limits_{j=1}^{i} (\gamma_j, \Delta S_j)$).

Clearly, for this strategy π we have $X_0^\pi = 0$, $X_i^\pi = 0$ for $i < n$, and $X_i^\pi = X_n^\pi \geqslant 0$ for all $i \geqslant n$; moreover, $\mathsf{P}(X_N^\pi > 0) > 0$, which contradicts condition a$'$).

5. *Proof* e) $\Longrightarrow$ b). We shall define the martingale measure $\widetilde{\mathsf{P}} \in \mathscr{P}_b(\mathsf{P})$ by the formula $d\widetilde{\mathsf{P}} = \widetilde{Z}\, d\mathsf{P}$, where

$$\widetilde{Z} = \prod_{n=1}^{N} \widetilde{z}_n \tag{16}$$

for some $\mathscr{F}_n$-measurable functions $\widetilde{z}_n$.

We consider the regular conditional probability

$$\mathsf{Q}_n(\omega, dx) = \mathsf{P}(\Delta S_n \in dx \mid \mathscr{F}_{n-1})(\omega),$$

which is also called the *transitional probability* (from $(\Omega, \mathscr{F}_{n-1})$ to $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$). In what follows, we construct from $\mathsf{Q}_n(\omega, dx)$, by means of a special procedure, transitional probabilities $\mathsf{Q}'_n(\omega, dx, dv)$ (from $(\Omega, \mathscr{F}_{n-1})$ to $(E, \mathscr{E})$ with $E = \mathbb{R}^d \times (0, \infty)$) such that

$$\mathsf{Q}'_n\big(\omega, dx, (0, \infty)\big) = \mathsf{Q}_n(\omega, dx). \tag{17}$$

We now return to our discussion in subsection 2 and set $\mathsf{Q}'(dx, dv)$ there to be equal to the measure $\mathsf{Q}'_n(\omega, dx, dv)$ ($n = 1, \dots, N$, $\omega \in \Omega$).

We also set (see (5)–(7))

$$A = \left\{(\omega, g) \in \Omega \times G\colon \int_E z_g(x, v)\, \mathsf{Q}'_n(\omega, dx, dv) = 1,\ \varphi(g) = 0\right\}.$$

Note that the above-introduced space G of elements g defined by (4) is a Polish space (with product topology). Let $\mathscr{G}$ be its Borel σ-algebra.

The set A is $\mathscr{F}_{n-1} \otimes \mathscr{G}$-measurable, and its projection $\pi(A)$ is equal to Ω by assumption e).

Hence, by Lemma 2 there exists a G-valued $\mathscr{F}_{n-1}$-measurable function $g_n = g_n(\omega)$ such that $(\omega, g_n(\omega)) \in A$ for P-almost all $\omega \in \Omega$.

We now consider the function

$$z_n(\omega, x, v) = z_{g_n(\omega)}(x, v) \tag{18}$$

in $\Omega \times E$. It is $\mathscr{F}_{n-1} \otimes \mathscr{E}$-measurable, and (P-a.s.) we have

$$\int_E z_n(\omega, x, v)\, \mathsf{Q}'(\omega, dx, dv) = 1 \tag{19}$$

and

$$\int_E x z_n(\omega, x, v)\, \mathsf{Q}'(\omega, dx, dv) = 0. \tag{20}$$

By (5),

$$\sup_v v z_n(\omega, x, v) < \infty$$

P-almost surely.

Using Lemma 2 again we see that there exists a $\mathscr{F}_{n-1}$-measurable positive function $V_{n-1} = V_{n-1}(\omega)$ such that

$$v z_n(\omega, x, v) \leqslant V_{n-1}(\omega) \tag{21}$$

for all (x, v) and P-almost all $\omega \in \Omega$.

We proceed now to a construction (by backward induction) of the sequence of measures $\mathsf{Q}'_n = \mathsf{Q}'_n(\omega, dx, dv)$ and the corresponding sequence of functions $z_n(\omega, x, v)$, $n = N, N-1, \dots$.

To this end we set $n = N$ and $V_N(\omega) = 1$, $\omega \in \Omega$. Let $\mathsf{Q}'_N = \mathsf{Q}'_N(\omega, dx, dv)$ be the regular $\mathscr{F}_{n-1}$-measurable conditional probability of the vector $(\Delta S_N, V_N)$. Clearly, the 'first marginal' of this measure is precisely $\mathsf{Q}_N = \mathsf{Q}_N(\omega, dx)$.

Let $z_N(\omega, x, v)$ be a function associated with Q'_N in accordance with the construction (18) and let $V_{N-1}(\omega)$ be defined by $z_N(\omega, x, v)$ in accordance with (21). Then we define Q'_{N-1} as the regular $\mathscr{F}_{N-2}$-measurable conditional distribution of the vector $(\Delta S_{N-1}, V_{N-1})$. In general, given Q'_n, we define $z_n(\omega, x, v)$ in accordance with (18) and $V_{n-1}(\omega)$ in accordance with (21).

Next we set Q'_{n-1} to be equal to the $\mathscr{F}_{n-2}$-measurable conditional distribution of $(\Delta S_{n-1}, V_{n-1})$ and so on.

We now set

$$\widetilde{z}_n(\omega) = z_n\big(\omega, \Delta S_n(\omega), V_n(\omega)\big) \tag{22}$$

and

$$\widetilde{Z}(\omega) = \prod_{n=1}^{N} \widetilde{z}_n(\omega). \tag{23}$$

Then, in view of (21) and (22),

$$\tilde{z}_n(\omega) \leqslant \frac{V_{n-1}(\omega)}{V_n(\omega)}. \tag{24}$$

By (23), (24) and bearing in mind that $V_N(\omega) = W_N(\omega) = 1$ we obtain

$$\widetilde{Z}(\omega) \leqslant V_0(\omega), \tag{25}$$

where $V_0(\omega) < \infty$ (P-a.s.). Since $\mathscr{F}_0 = \{\varnothing, \Omega\}$, the function $V_0(\omega)$ is constant (P-a.s.).

Further, from the definition of Q'_n and conditions (19) and (20) we see that (P-a.s.)

$$\mathsf{E}(\tilde{z}_n \mid \mathscr{F}_{n-1})(\omega) = 1 \tag{26}$$

and

$$\mathsf{E}(\Delta S_n \tilde{z}_n \mid \mathscr{F}_{n-1})(\omega) = 0. \tag{27}$$

By (26) we obtain that $\mathsf{E}\widetilde{Z}(\omega) = 1$, and therefore we can define a new probability measure $\widetilde{\mathsf{P}}$ by setting

$$\widetilde{\mathsf{P}}(d\omega) = \widetilde{Z}(\omega)\,\mathsf{P}(d\omega). \tag{28}$$

It is clear from (25) that $\widetilde{\mathsf{P}} \sim \mathsf{P}$.

Since

$$\widetilde{\mathsf{E}}|\Delta S_n| = \mathsf{E}\widetilde{Z}|\Delta S_n| \leqslant V_0 \mathsf{E}|\Delta S_n| < \infty$$

and

$$\widetilde{\mathsf{E}}(\Delta S_n \mid \mathscr{F}_{n-1}) = \mathsf{E}(\Delta S_n \tilde{z}_n \mid \mathscr{F}_{n-1}) = 0$$

by Bayes's formula (see formula (4) in § 3a) and (27), the sequence $S = (S_n)$ is a $\widetilde{\mathsf{P}}$-*martingale*.

By the above construction of the measure $\widetilde{\mathsf{P}}$ we obtain that $\widetilde{\mathsf{P}} \in \mathscr{P}_b(\mathsf{P})$. This proves the implication e) $\Longrightarrow$ b).

6. *Proof* e) $\Longrightarrow$ a$''$). Let $\widetilde{\mathsf{P}} \in \mathscr{P}(\mathsf{P}) \neq \varnothing$. Then $S = (S_n)_{n \leqslant N}$ is a $\widetilde{\mathsf{P}}$-martingale. Let π be a self-financing strategy such that $X_0^\pi = 0$ and $X_N^\pi \geqslant 0$.

Since $\Delta X_n^\pi = \gamma_n \, \Delta S_n$, the sequence $X^\pi = (X_n^\pi)_{n \leqslant N}$ is a martingale transform, and therefore X^π is a martingale by the lemma in Chapter II, § 1c. Hence $\mathsf{E}X_N^\pi = \mathsf{E}X_0^\pi = 0$, so that $X_N^\pi = 0$ (P-a.s.), which proves the required implication.

The proof of Theorem A* is complete.

Remark 2. If we do not assume that $B_n \equiv 1$, $n \leqslant N$, then in place of the regular conditional probabilities

$$\mathsf{Q}_n(\omega, dx) = \mathsf{P}(\Delta S_n \in dx \mid \mathscr{F}_{n-1})(\omega),$$

one must consider directly the regular modifications of the conditional probabilities

$$\overline{\mathsf{Q}}_n(\omega, dx) = \mathsf{P}\biggl(\Delta\biggl(\frac{S_n}{B_n}\biggr) \in dx \biggm| \mathscr{F}_{n-1}\biggr)(\omega).$$

The argument, in which one must bear in mind that the B_n, $n \leqslant N$, are positive and $\mathscr{F}_{n-1}$-measurable, remains essentially the same.

3. Construction of Martingale Measures by Means of an Absolutely Continuous Change of Measure

§ 3a. Main Definitions. Density Process

1. Let $(\Omega, \mathscr{F}, (\mathscr{F}_n)_{n\geqslant 1}, \mathsf{P})$ be a filtered probability space. Then we say that a measure $\widetilde{\mathsf{P}}$ in $(\Omega, \mathscr{F})$, is *absolutely continuous* with respect to P (we write $\widetilde{\mathsf{P}} \ll \mathsf{P}$) if $\widetilde{\mathsf{P}}(A) = 0$ for each $A \in \mathscr{F}$ such that $\mathsf{P}(A) = 0$.

We say that measures P and $\widetilde{\mathsf{P}}$ in the same measurable space $(\Omega, \mathscr{F})$ are *equivalent* (we write $\widetilde{\mathsf{P}} \sim \mathsf{P}$) if $\widetilde{\mathsf{P}} \ll \mathsf{P}$ and $\mathsf{P} \ll \widetilde{\mathsf{P}}$.

In many cases the condition of absolute continuity or equivalence is too restrictive or simply redundant: one can do with a weaker condition of *local* absolute continuity that can be explained as follows.

Let $\mathsf{P}_n = \mathsf{P} \,|\, \mathscr{F}_n$ be the restriction of the probability measure P to the σ-algebra $\mathscr{F}_n \subseteq \mathscr{F}$. (In other words, P_n is a measure in $(\Omega, \mathscr{F}_n)$ such that

$$\mathsf{P}_n(A) = \mathsf{P}(A)$$

for each $A \in \mathscr{F}_n$.) Then we say that a measure $\widetilde{\mathsf{P}}$ is *locally absolutely continuous* with respect to P (we write $\widetilde{\mathsf{P}} \overset{\text{loc}}{\ll} \mathsf{P}$) if

$$\widetilde{\mathsf{P}}_n \ll \mathsf{P}_n$$

for each $n \geqslant 1$.

Two measures, P and $\widetilde{\mathsf{P}}$, are said to be *locally equivalent* (we write $\widetilde{\mathsf{P}} \overset{\text{loc}}{\sim} \mathsf{P}$) if $\widetilde{\mathsf{P}} \overset{\text{loc}}{\ll} \mathsf{P}$ and $\mathsf{P} \overset{\text{loc}}{\ll} \widetilde{\mathsf{P}}$.

If, e.g., Ω is the space $\mathbb{R}^\infty$, the coordinate space of sequences $\omega = (x_1, x_2, \dots)$, $\mathscr{F}_n = \sigma(\omega\colon x_1, \dots, x_n)$ is the σ-algebra generated by the first n coordinates functions, $\mathscr{F} = \mathscr{B}(\mathbb{R}^\infty)$, and P, $\widetilde{\mathsf{P}}$ are probability measures in $(\Omega, \mathscr{F})$, then the relation $\widetilde{\mathsf{P}} \overset{\text{loc}}{\ll} \mathsf{P}$ means precisely that their corresponding *finite-dimensional* probability distributions are absolutely continuous.

Note that if we have $n \leqslant N < \infty$, then local absolute continuity is the same as absolute continuity. Thus, the 'local' concept can be of interest only for models with time parameter $n \in \mathbb{N} = \{1, 2, \dots\}$.

Besides the restrictions $\mathsf{P}_n = \mathsf{P} \,|\, \mathscr{F}_n$ of the measure P to the σ-algebras $\mathscr{F}_n$ we shall also consider the restrictions $\mathsf{P}_\tau = \mathsf{P} \,|\, \mathscr{F}_\tau$ of P to the σ-algebras $\mathscr{F}_\tau$ of sets $A \in \mathscr{F}$ such that $\{\tau(\omega) \leqslant n\} \cap A \in \mathscr{F}_n$ for each $n \geqslant 1$. We emphasize that, as usual, $\mathscr{F}_\infty = \mathscr{F}$ (and $\mathscr{F}_{\infty-} = \bigvee \mathscr{F}_n$); see [250; Chapter I, § 1a].

2. Let $\widetilde{\mathsf{P}} \overset{\text{loc}}{\ll} \mathsf{P}$. Then $\widetilde{\mathsf{P}}_n \ll \mathsf{P}_n$ for each $n \in \mathbb{N}$ and there exist (see, e.g., [439]) *Radon–Nikodym derivatives* denoted by

$$\frac{d\widetilde{\mathsf{P}}_n}{d\mathsf{P}_n} \quad \text{or} \quad \frac{d\widetilde{\mathsf{P}}_n}{d\mathsf{P}_n}(\omega)$$

and defined as $\mathscr{F}_n$-measurable functions $Z_n = Z_n(\omega)$ such that

$$\widetilde{\mathsf{P}}_n(A) = \int_A Z_n(\omega)\, \mathsf{P}_n(d\omega), \qquad A \in \mathscr{F}_n. \tag{1}$$

Remark. The Radon–Nikodym derivative $\dfrac{d\widetilde{\mathsf{P}}_n}{d\mathsf{P}_n}$ is defined only up to P_n-indistinguishability, i.e., if (1) holds for two functions, $Z_n(\omega)$ and $Z_n'(\omega)$, then

$$\mathsf{P}\big(Z_n(\omega) \neq Z_n'(\omega)\big) = 0.$$

Both $Z_n(\omega)$ and $Z_n'(\omega)$ are 'representatives' of the Radon–Nikodym derivative in this case. Saying that '$Z_n = \dfrac{d\widetilde{\mathsf{P}}_n}{d\mathsf{P}_n}$ is the Radon–Nikodym derivative' we mean that we have chosen *one representative* that we shall consider in what follows. It is easy to see that we can choose versions such that not only $\mathsf{P}\big(Z_n(\omega) \geqslant 0\big) = 1$, but also $Z_n(\omega) \geqslant 0$ for *all* $\omega \in \Omega$ and each $n \geqslant 1$. It is for this reason that one usually includes the *non-negativity* in the definition of the Radon–Nikodym derivative of a probability measure.

In what follows, we shall call the discrete-time process

$$Z = (Z_n)_{n \geqslant 1},$$

the *density process* (of the measures $\widetilde{\mathsf{P}}_n$ with respect to the P_n, $n \geqslant 1$, or of the measure $\widetilde{\mathsf{P}}$ with respect to the measure P).

For the convenience of references we combine in the next theorem several simple but useful and important properties of the density process.

THEOREM. *Assume that* $\widetilde{\mathsf{P}} \overset{\text{loc}}{\ll} \mathsf{P}$. *Then we have the following results.*

a) *The density process* $Z = (Z_n)$ *is a non-negative* $(\mathsf{P}, (\mathscr{F}_n))$*-martingale with* $\mathsf{E}Z_n = 1$.

b) *Let* $\mathscr{F} = \mathscr{F}_{\infty-}$, *where* $\mathscr{F}_{\infty-} = \bigvee \mathscr{F}_n$. *Then the following conditions are equivalent:*

(i) $\widetilde{\mathsf{P}} \ll \mathsf{P}$;

(ii) $Z = (Z_n)$ *is a uniformly integrable* $(\mathsf{P}, (\mathscr{F}_n))$*-martingale;*

(iii) $\widetilde{\mathsf{P}}\Bigl(\sup\limits_n Z_n < \infty\Bigr) = 1$.

c) *Let* $\tau = \inf\{n\colon Z_n = 0\}$ *be the first instant when the density process vanishes. Then it 'remains at the origin indefinitely', i.e.*

$$\mathsf{P}\bigl\{\omega\colon Z_n(\omega) \neq 0 \text{ for some } n \geqslant \tau(\omega)\bigr\} = 0.$$

d) *Let* τ *be some stopping time. Then the restrictions* $\widetilde{\mathsf{P}}_\tau = \widetilde{\mathsf{P}} \,|\, \mathscr{F}_\tau$ *and* $\mathsf{P}_\tau = \mathsf{P} \,|\, \mathscr{F}_\tau$ *to the* σ*-algebra* $\mathscr{F}_\tau$ (see Definition 2 in Chapter II, § 1f) *satisfy the relations* $\widetilde{\mathsf{P}}_\tau \ll \mathsf{P}_\tau$ *and*

$$Z_\tau = \frac{d\widetilde{\mathsf{P}}_\tau}{d\mathsf{P}_\tau}\,. \tag{2}$$

e) *We have the equality*

$$\widetilde{\mathsf{P}}\Bigl(\inf_n Z_n > 0\Bigr) = 1. \tag{3}$$

f) *If* $\mathsf{P}(Z_n > 0) = 1$ *for each* $n \geqslant 1$, *then* $\mathsf{P} \overset{\text{loc}}{\ll} \widetilde{\mathsf{P}}$ *and*

$$\widetilde{\mathsf{P}} \overset{\text{loc}}{\sim} \mathsf{P}.$$

Proof. a) By (1),

$$\widetilde{\mathsf{P}}_n(A) = \mathsf{E}(I_A Z_n) = \widetilde{\mathsf{P}}_{n+1}(A) = \mathsf{E}(I_A Z_{n+1})$$

for $A \in \mathscr{F}_n$, and therefore $\mathsf{E}I_A Z_n = \mathsf{E}I_A Z_{n+1}$. Hence $\mathsf{E}(Z_{n+1} \,|\, \mathscr{F}_n) = Z_n$ (P-a.s.) for each $n \geqslant 1$. It is also clear that $\mathsf{E}Z_n = \widetilde{\mathsf{P}}_n(\Omega) = 1$. Thus, Z is a $(\mathsf{P}, (\mathscr{F}_n))$-martingale.

b) (i) $\Longrightarrow$ (ii). We recall the following classical result of martingale theory ([109]; see also Chapter III, § 3b.4 in the continuous-time case).

DOOB'S CONVERGENCE THEOREM. *Let $X = (X_n)$ be a supermartingale with respect to the measure* P *and the flow* $(\mathscr{F}_n)$ *such that there exists an integrable random variable* Y *such that* $X_n \geqslant \mathsf{E}(Y \mid \mathscr{F}_n)$ (P-a.s.) *for each* $n \geqslant 1$.

Then X_n *converges* (P-a.s.) *to a finite limit* X_∞.

(The proof can be found in many handbooks, e.g., [439; Chapter VII, § 4].)

To prove the implication (i) $\Longrightarrow$ (iii) it suffices to observe that since $Z_n \geqslant 0$, it follows by this Doob's theorem that there exists with P-probability one a finite limit $\lim_n Z_n$. However, $\widetilde{\mathsf{P}} \ll \mathsf{P}$, therefore with $\widetilde{\mathsf{P}}$-probability one there also exists a finite limit, so that $\widetilde{\mathsf{P}}\left(\sup_n Z_n < \infty\right) = 1$.

(iii) $\Longrightarrow$ (ii). The *uniform integrability* of a family of random variables (ξ_n) means that

$$\lim_N \sup_n \mathsf{E}\big(|\xi_n| I(|\xi_n| > N)\big) = 0.$$

In our case (where $\xi_n = Z_n$), by (iii) we obtain

$$\mathsf{E}\big(Z_n I(Z_n > N)\big) = \widetilde{\mathsf{P}}(Z_n > N) \leqslant \widetilde{\mathsf{P}}\Big(\sup_n Z_n > N\Big) \to 0 \quad \text{as} \quad N \to \infty,$$

which proves (ii).

(ii) $\Longrightarrow$ (i). By Doob's theorem, $Z_n \to Z_\infty$ (P-a.s.). Hence the uniform integrability of the sequence (Z_n) ensures also the convergence in $L^1(\Omega, \mathscr{F}, \mathsf{P})$, i.e., $\mathsf{E}|Z_n - Z_\infty| \to 0$ as $n \to \infty$.

For sets $A \in \mathscr{F}_m$ and for $n \geqslant m$ we have

$$\widetilde{\mathsf{P}}(A) = \mathsf{E} I_A Z_m = \mathsf{E} I_A Z_n.$$

However, $\mathsf{E}|Z_n - Z_\infty| \to \infty$, and for each $A \in \mathscr{F}_m$ we obtain

$$\widetilde{\mathsf{P}}(A) = \mathsf{E} I_A Z_\infty.$$

Hence, using the standard techniques of monotonic classes ([439; § 2, Chapter II]) we conclude that the same equality holds in $\bigcup \mathscr{F}_n$ and in $\mathscr{F} = \sigma\big(\bigcup \mathscr{F}_n\big)$ $\big(\equiv \mathscr{F}_{-\infty} = \bigvee \mathscr{F}_n\big)$.

Thus, $\widetilde{\mathsf{P}} \ll \mathsf{P}$ and, moreover,

$$\frac{d\widetilde{\mathsf{P}}}{d\mathsf{P}} = Z_\infty,$$

where $Z_\infty = \lim Z_n$.

c) To prove this property we require another classical result of martingale theory ([109]; see Chapter III, § 3b.4 again for the continuous-time case).

DOOB'S OPTIONAL STOPPING THEOREM. *Let $X = (X_n)$ be a supermartingale such that*

$$X_n \geqslant \mathsf{E}(Y \mid \mathscr{F}_n), \qquad n \geqslant 1$$

for some integrable random variable Y.

Then the variables X_σ and X_τ are integrable for any two Markov times σ and τ, and

$$\mathsf{E}(X_\tau \mid \mathscr{F}_\sigma) \leqslant X_\sigma \quad (\mathsf{P}\text{-a.s.})$$

in the set $\{\sigma \leqslant \tau\}$.

(See the proof, e.g., in [109] or [439: Chapter VII, § 2].)

Remark. On $\{\omega\colon \tau(\omega) = \infty\}$ we set the value of $X_\tau(\omega)$ to be equal to $X_\infty(\omega)$, the limit of the $X_n(\omega)$, which exists by Doob's convergence theorem.

Besides $\tau = \inf\{n \geqslant 1\colon Z_n = 0\}$ we shall consider the times $\sigma_m = \inf\{n \geqslant 1\colon Z_n > 1/m\}$. It is easy to see that τ and σ_m are stopping times, i.e., the sets $\{\tau \leqslant n\}$ and $\{\sigma_m \leqslant n\}$ belong to $\mathscr{F}_n$ for each $n \geqslant 1$ and all $m \geqslant 1$. (We recall that, as always, $\tau(\omega) = \infty$ if $Z_n(\omega) > 0$ for all $n \geqslant 1$.)

By Doob's optional stopping theorem,

$$\mathsf{E}(Z_{\sigma_m} \mid \mathscr{F}_\tau) \leqslant Z_\tau = 0 \quad \text{on the set} \quad \{\omega\colon \tau(\omega) < \infty\}.$$

Hence $Z_{\sigma_m} I(\tau < \infty) = 0$, $m \geqslant 1$, so that $\sigma_m = \infty$ (P-a.s.), $m \geqslant 1$, which precisely means that

$$\mathsf{P}\bigl\{\omega\colon \exists\, n \geqslant \tau(\omega) \text{ with } Z_n(\omega) \neq 0\bigr\} = 0.$$

d) Let $A \in \mathscr{F}_\tau$. Then

$$\begin{aligned} \mathsf{E}\bigl[I_A \cdot I_{\{\tau<\infty\}} \cdot Z_\tau\bigr] &= \sum_{n\geqslant 1} \mathsf{E}\bigl[I_A \cdot I_{\{\tau=n\}} \cdot Z_\tau\bigr] = \sum_{n\geqslant 1} \mathsf{E}\bigl[I_A \cdot I_{\{\tau=n\}} \cdot Z_n\bigr] \\ &= \sum_{n\geqslant 1} \widetilde{\mathsf{P}}\bigl(A \cap I(\tau = n)\bigr) = \widetilde{\mathsf{P}}\bigl(A \cap \{\tau < \infty\}\bigr), \end{aligned}$$

which proves the required result.

e) We set

$$\tau_m = \inf\Bigl\{n\colon Z_n < \frac{1}{m}\Bigr\}.$$

Then

$$\widetilde{\mathsf{P}}(\tau_m < \infty) = \mathsf{E}\bigl(Z_{\tau_m} I(\tau_m < \infty)\bigr) \leqslant \frac{1}{m}$$

by d), and therefore

$$\widetilde{\mathsf{P}}\Bigl(\bigcap_m \{\tau_m < \infty\}\Bigr) = 0,$$

which is equivalent to the required equality $\widetilde{\mathsf{P}}\Bigl(\inf\limits_n Z_n > 0\Bigr) = 1$.

f) If $\widetilde{\mathsf{P}} \overset{\text{loc}}{\ll} \mathsf{P}$ and $\mathsf{P}(Z_n > 0) = 1$, $n \geqslant 1$, then also $\widetilde{\mathsf{P}}(Z_n > 0) = 1$.

For $A \in \mathscr{F}_n$ we set

$$Q_n(A) = \int_A Z_n^{-1}\, \widetilde{\mathsf{P}}(d\omega).$$

Then $\widetilde{\mathsf{P}}_n(d\omega) = Z_n\, \mathsf{P}_n(d\omega)$, and therefore

$$Q_n(A) = \int_A Z_n^{-1} Z_n\, \mathsf{P}(d\omega) = \mathsf{P}_n(A), \qquad n \geqslant 1.$$

Hence

$$\mathsf{P}_n(A) = \int_A Z_n^{-1}\, \widetilde{\mathsf{P}}_n(d\omega),$$

so that $\mathsf{P} \overset{\text{loc}}{\ll} \widetilde{\mathsf{P}}$.

Thus, we have proved all assertions a)–f) of the theorem.

3. The following 'technical' result is useful in the considerations of conditional expectations with respect to different measures. We shall use it repeatedly in what follows and call it the '*conversion lemma*'. Formula (4) is often called '(*generalized*) *Bayes' formula*' ([303; Chapter 7]).

LEMMA. *Let $\widetilde{\mathsf{P}}_n \ll \mathsf{P}_n$ and let Y be a bounded (or $\widetilde{\mathsf{P}}$-integrable) $\mathscr{F}_n$-measurable random variable. Then for each $m \leqslant n$ we have*

$$\widetilde{\mathsf{E}}(Y \mid \mathscr{F}_m) = \frac{1}{Z_m} \mathsf{E}(YZ_n \mid \mathscr{F}_m) \quad (\widetilde{\mathsf{P}}\text{-a.s.}). \tag{4}$$

Proof. For a start we observe that $\widetilde{\mathsf{P}}(Z_m > 0) = 1$ (see assertion e) in the above theorem). Further, we also have $Z_n(\omega) = 0$ (P-a.s.) in the set $\{\omega\colon Z_m(\omega) = 0\}$ for $n \geqslant m$. Taking this into account we shall assume that the right-hand side of (4) vanishes in this set.

By definition, $\widetilde{\mathsf{E}}(Y \mid \mathscr{F}_m)$ is a $\mathscr{F}_m$-measurable random variable such that

$$\widetilde{\mathsf{E}}\big[I_A \cdot \widetilde{\mathsf{E}}(Y \mid \mathscr{F}_m)\big] = \widetilde{\mathsf{E}}[I_A \cdot Y] \tag{5}$$

for each $A \in \mathscr{F}_m$, so that we need only verify that for the $\mathscr{F}_m$-measurable function on the right-hand side of (4) we have

$$\widetilde{\mathsf{E}}\left[I_A \cdot \frac{1}{Z_m} \mathsf{E}(YZ_n \mid \mathscr{F}_m)\right] = \widetilde{\mathsf{E}}[I_A \cdot Y]. \tag{6}$$

In fact, this follows from the following chain of equalities:

$$\begin{aligned}
\widetilde{\mathsf{E}}\left[I_A \cdot \frac{1}{Z_m} \mathsf{E}(YZ_n \mid \mathscr{F}_m)\right] &= \mathsf{E}\left[I_A \cdot \frac{1}{Z_m} \mathsf{E}(YZ_n \mid \mathscr{F}_m) \cdot Z_m\right] \\
&= \mathsf{E}\big[I_A \cdot \mathsf{E}(YZ_n \mid \mathscr{F}_m)\big] \\
&\overset{(\alpha)}{=} \mathsf{E}\big[I_A Y Z_n\big] \overset{(\beta)}{=} \mathsf{E}\big[I_A Y Z_m\big] = \widetilde{\mathsf{E}}\big[I_A Y\big],
\end{aligned}$$

where (α) holds by the definition of the conditional expectation $\mathsf{E}(YZ_n \mid \mathscr{F}_m)$ and (β) is a consequence of the $\mathscr{F}_m$-measurability of $I_A Y$ and the fact that $Z = (Z_n)$ is a martingale.

§ 3b. Discrete Version of Girsanov's Theorem. Conditionally Gaussian Case

1. It is reasonable to start our discussion of the construction of the probability measures $\widetilde{\mathsf{P}}$, that are (locally) absolutely continuous or equivalent to the original basic measure P involved in the definition of the filtered space

$$(\Omega, \mathscr{F}, (\mathscr{F}_n)_{n\geqslant 1}, \mathsf{P}),$$

with a *discrete* (with respect to time) *version* of a result established by I. V. Girsanov [183] for *processes of diffusion type*, which became the prototype for a number of results for martingales, local martingales, local measures, semimartingales, and so on. (See, e.g., [250; Chapter II].)

Let $n \geqslant 1$ be the time parameter and let $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots)$ be a sequence of $\mathscr{F}_n$-measurable random variables ε_n with distribution

$$\operatorname{Law}(\varepsilon_n \mid \mathscr{F}_{n-1}; \mathsf{P}) = \mathscr{N}(0,1). \tag{1}$$

In particular, this means that ε is a sequence of independent standard normally distributed random variables, $\varepsilon_n \sim \mathscr{N}(0,1)$.

Assume that, besides $\varepsilon = (\varepsilon_n)_{n\geqslant 1}$, we are given predictable sequences $\mu = (\mu_n)_{n\geqslant 1}$ and $\sigma = (\sigma_n)_{n\geqslant 1}$, i.e., sequences of $\mathscr{F}_{n-1}$-measurable variables μ_n and σ_n ($\mathscr{F}_0 = \{\varnothing, \Omega\}$). Moreover, we assume that $\sigma_n > 0$, which is based on the interpretation of this parameter as 'volatility' and the fact that we can simply leave observations with $\sigma_n = 0$ out of consideration.

We set $h = (h_n)_{n\geqslant 1}$, where

$$h_n = \mu_n + \sigma_n \varepsilon_n. \tag{2}$$

By (1) we obtain that the (regular) conditional distribution $\mathsf{P}(h_n \leqslant \cdot \mid \mathscr{F}_{n-1})$ can be defined as follows:

$$\mathsf{P}(h_n \leqslant x \mid \mathscr{F}_{n-1}) = \frac{1}{\sqrt{2\pi\sigma_n^2}} \int_{-\infty}^{x} e^{-\frac{(y-\mu_n)^2}{2\sigma_n^2}}\, dy, \tag{3}$$

or, symbolically,

$$\operatorname{Law}(h_n \mid \mathscr{F}_{n-1}; \mathsf{P}) = \mathscr{N}(\mu_n, \sigma_n^2), \tag{4}$$

which allows one to call $h = (h_n)$ a *conditionally Gaussian* sequence (with respect to P) with (conditional) expectation

$$\mathsf{E}(h_n \mid \mathscr{F}_{n-1}) = \mu_n \tag{5}$$

and variance

$$\mathsf{D}(h_n \mid \mathscr{F}_{n-1}) = \sigma_n^2. \tag{6}$$

We see from (4) that

$$\mathsf{E}(h_n - \mu_n \mid \mathscr{F}_{n-1}) = 0, \tag{5'}$$

$$\mathsf{D}(h_n - \mu_n \mid \mathscr{F}_{n-1}) = \sigma_n^2. \tag{6'}$$

Setting

$$H_n = \sum_{k=1}^{n} h_k, \quad A_n = \sum_{k=1}^{n} \mu_k, \quad \text{and} \quad M_n = \sum_{k=1}^{n} \sigma_k \varepsilon_k,$$

we can say that the variables H_n can be represented as follows in the *conditionally Gaussian case*:

$$H_n = A_n + M_n,$$

where $A = (A_n)$ is a *predictable sequence* and $M = (M_n)$ is a *conditionally Gaussian martingale with quadratic characteristic*

$$\langle M \rangle_n = \sum_{k=1}^{n} \sigma_k^2.$$

We set $W_n = \sum_{k=1}^{n} \varepsilon_k$, $\Delta W_n = \varepsilon_n$, $\Delta H_n = h_n$, and $\Delta = 1$. Then we can rewrite (2) in the *difference* form as follows:

$$\Delta H_n = \mu_n \Delta + \sigma_n \Delta W_n,$$

which one can regard as a discrete counterpart to the *stochastic differential*

$$dH_t = \mu_t \, dt + \sigma_t \, dW_t$$

of some Itô process $H = (H_t)$ generated by a Wiener process $W = (W_t)_{t \geqslant 0}$, with local drift $(\mu_t)_{t \geqslant 0}$ and local volatility $(\sigma_t)_{t \geqslant 0}$ (see, e.g., [303; Chapter 4] and Chapter III, § 3d).

In the present case of a conditionally Gaussian sequence (2) this discrete analog of Girsanov's theorem (which, as already mentioned, was proved by I. V. Girsanov in the *continuous-time* case) has a relation to the question of the *existence of a measure such that* $\widetilde{\mathsf{P}}$ *is absolutely continuous or equivalent to the measure* P *and the sequence* $h = (h_n)$ *is a (local) martingale difference with respect to* $\widetilde{\mathsf{P}}$. It is worthwhile to point out in this connection that the right-hand side of (2) contains two terms: the 'drift' μ_n and the 'discrete diffusion' $\sigma_n \varepsilon_n$, which is a martingale difference (with respect to P). The meaning of the above question lies in the existence of a measure $\widetilde{\mathsf{P}} \ll \mathsf{P}$ such that (h_n) has no 'drift' component with respect to $\widetilde{\mathsf{P}} \ll \mathsf{P}$ and reduces to 'discrete diffusion', i.e., is a (local) martingale difference.

2. Our construction of the measure $\widetilde{\mathsf{P}}$ is based of the sequence of (positive) random variables

$$Z_n = \exp\left\{-\sum_{k=1}^{n} \frac{\mu_k}{\sigma_k}\varepsilon_k - \frac{1}{2}\sum_{k=1}^{n}\left(\frac{\mu_k}{\sigma_k}\right)^2\right\}, \qquad n \geqslant 1. \tag{7}$$

LEMMA. 1) *The sequence* $Z = (Z_n)_{n\geqslant 1}$ *is a* $(\mathsf{P}, (\mathscr{F}_n))$*-martingale with* $\mathsf{E}Z_n = 1$, $n \geqslant 1$.

2) *Let* $\mathscr{F} = \bigvee \mathscr{F}_n$ *and assume that*

$$\mathsf{E}\exp\left(\frac{1}{2}\sum_{k=1}^{\infty}\left(\frac{\mu_k}{\sigma_k}\right)^2\right) < \infty \tag{8}$$

(the 'Novikov condition')

Then $Z = (Z_n)_{n\geqslant 1}$ *is a uniformly integrable martingale with limit* (P-a.s.) $Z_\infty = \lim Z_n$ *such that*

$$Z_\infty = \exp\left\{-\sum_{k=1}^{\infty} \frac{\mu_k}{\sigma_k}\varepsilon_k - \frac{1}{2}\sum_{k=1}^{\infty}\left(\frac{\mu_k}{\sigma_k}\right)^2\right\}, \tag{9}$$

and

$$Z_n = \mathsf{E}(Z_\infty \mid \mathscr{F}_n). \tag{10}$$

Proof. 1) This is obvious since for each $k \geqslant 1$ we obtain

$$\mathsf{E}\exp\left\{\frac{\mu_k}{\sigma_k}\varepsilon_k - \frac{1}{2}\left(\frac{\mu_k}{\sigma_k}\right)^2 \,\Big|\, \mathscr{F}_{k-1}\right\} = 1 \tag{11}$$

by the $\mathscr{F}_{k-1}$-measurability of the $\dfrac{\mu_k}{\sigma_k}$ and conditionally Gaussian property (1) (here $\mathscr{F}_0 = \{\varnothing, \Omega\}$).

2) The proof that the family (Z_n) is uniformly integrable if (8) holds is fairly complicated and can be found in the original paper [368] by A. A. Novikov as well as in many textbooks (see, e.g., [303; Chapter 7] or [402]).

However, this proof becomes relatively elementary once one imposes a stronger condition: *there exists* $\varepsilon > 0$ such that

$$\mathsf{E}\exp\left\{\left(\frac{1}{2}+\varepsilon\right)\sum_{k=1}^{\infty}\left(\frac{\mu_k}{\sigma_k}\right)^2\right\} < \infty. \tag{12}$$

Hence it would be reasonable to present this proof, which is what we do at the end of this section (see subsection 5).

3. We fix some $N \geqslant 1$ and shall consider the sequence (h_n) only for $n \leqslant N$. For simplicity we shall assume that $\mathscr{F} = \mathscr{F}_N$, so that $\mathsf{P}_N = \mathsf{P} \mid \mathscr{F}_N = \mathsf{P}$.

Since $Z_N > 0$ and $\mathsf{E} Z_N = 1$, we can define a probability measure $\widetilde{\mathsf{P}} = \widetilde{\mathsf{P}}(d\omega)$ in $(\Omega, \mathscr{F})$ by setting

$$\widetilde{\mathsf{P}}(d\omega) = Z_N(\omega)\, \mathsf{P}(d\omega). \tag{13}$$

We emphasize that we do not only have the relation $\widetilde{\mathsf{P}} \ll \mathsf{P}$, but also $\mathsf{P} \ll \widetilde{\mathsf{P}}$. Hence $\widetilde{\mathsf{P}} \sim \mathsf{P}$.

We consider now the properties of the sequence $(h_n)_{n \leqslant N}$ with respect to P.

By Bayes's formula (4) in § 3a, for each $\lambda \in \mathbb{R}$ and $n \leqslant N$ we obtain

$$\begin{aligned}
\widetilde{\mathsf{E}}\big(e^{i\lambda h_n} \mid \mathscr{F}_{n-1}\big) &= \mathsf{E}\Big(e^{\left(i\lambda\sigma_n - \frac{\mu_n}{\sigma_n}\right)\varepsilon_n + i\lambda\mu_n - \frac{1}{2}\left(\frac{\mu_n}{\sigma_n}\right)^2} \mid \mathscr{F}_{n-1}\Big) \\
&= \mathsf{E}\Big(e^{\left(i\lambda\sigma_n - \frac{\mu_n}{\sigma_n}\right)\varepsilon_n - \frac{1}{2}\left(i\lambda\sigma_n - \frac{\mu_n}{\sigma_n}\right)^2} \\
&\qquad\qquad \times e^{\frac{1}{2}\left(i\lambda\sigma_n - \frac{\mu_n}{\sigma_n}\right)^2 + i\lambda\mu_n - \frac{1}{2}\left(\frac{\mu_n}{\sigma_n}\right)^2} \mid \mathscr{F}_{n-1}\Big) \\
&= e^{-\frac{\lambda^2\sigma_n^2}{2}}
\end{aligned} \tag{14}$$

($\widetilde{\mathsf{P}}$-a.s.), where we use the equality

$$\mathsf{E} e^{\left(i\lambda\sigma_n - \frac{\mu_n}{\sigma_n}\right)\varepsilon_n - \frac{1}{2}\left(i\lambda\sigma_n - \frac{\mu_n}{\sigma_n}\right)^2} = 1$$

and the fact that the σ_n^2 are $\mathscr{F}_{n-1}$-measurable.

We obtain the equality

$$\widetilde{\mathsf{E}}\big(e^{i\lambda h_n} \mid \mathscr{F}_{n-1}\big) = e^{-\frac{\lambda^2\sigma_n^2}{2}} \quad (\widetilde{\mathsf{P}}\text{-a.s.}), \tag{15}$$

which means that the sequence $h = (h_n)$ remains *conditionally Gaussian* with respect to this new measure $\widetilde{\mathsf{P}}$, but has now the *trivial* 'drift' component:

$$\operatorname{Law}(h_n \mid \mathscr{F}_{n-1}; \widetilde{\mathsf{P}}) = \mathscr{N}(0, \sigma_n^2), \tag{16}$$

so that we obtain the following analogs of (5) and (6) in this case:

$$\widetilde{\mathsf{E}}(h_n \mid \mathscr{F}_{n-1}) = 0, \tag{17}$$

$$\widetilde{\mathsf{D}}(h_n \mid \mathscr{F}_{n-1}) = \sigma_n^2. \tag{18}$$

One can say that the transitions from P to the measure $\widetilde{\mathsf{P}}$ *eliminates* ('*kills*') the drift $\mu = (\mu_n)_{n \leqslant N}$ of the sequence $h = (h_n)_{n \leqslant N}$, *but preserves the conditional variance.*

We can also conclude from (16) that if $\widetilde{\varepsilon} = (\widetilde{\varepsilon}_n)_{n\leqslant N}$ is a sequence of $\mathscr{F}_n$-measurable variables with distribution

$$\mathrm{Law}(\widetilde{\varepsilon}_n \mid \mathscr{F}_{n-1}; \widetilde{\mathsf{P}}) = \mathscr{N}(0,1) \tag{19}$$

(one can always construct such a sequence, although it may be necessary to enlarge our initial probability space), then

$$\mathrm{Law}(h_n,\, n \leqslant N \mid \widetilde{\mathsf{P}}) = \mathrm{Law}(\sigma_n \widetilde{\varepsilon}_n,\, n \leqslant N \mid \widetilde{\mathsf{P}}). \tag{20}$$

Hence it is clear that the sequence $(h_n)_{n\leqslant N}$ 'behaves' as a local martingale difference $(\sigma_n\widetilde{\varepsilon}_n)_{n\leqslant N}$ with respect to $\widetilde{\mathsf{P}}$, while in terms of the original measure P a property similar to (20) can be expressed as follows:

$$\mathrm{Law}(h_n - \mu_n,\, n \leqslant N \mid \mathsf{P}) = \mathrm{Law}(\sigma_n \varepsilon_n,\, n \leqslant N \mid \mathsf{P}). \tag{21}$$

We now shift our standpoint.

We shall treat a measure $\widetilde{\mathsf{P}}$ and a sequence $h = (h_n)$ satisfying (20) as original ones. Then the transition from $\widetilde{\mathsf{P}}$ to a measure P (in accordance with (13)) gives us property (21), which can be interpreted as drift *appearing* in the local martingale difference $(\sigma_n\widetilde{\varepsilon}_n)_{n\leqslant N}$. This interpretation is more convenient when one is willing to state the corresponding general result on the transformations of local martingales under absolutely continuous changes of measure (see § 3d below).

Before summarizing the results just obtained we point out the following.

Assume that $\sigma_n^2(\omega)$ is independent of ω $(= \sigma_n^2)$. Using induction we obtain by (14) that

$$\begin{aligned}
\widetilde{\mathsf{E}}\big(e^{i\sum_{k=1}^N \lambda_k h_k}\big) &= \widetilde{\mathsf{E}}\big(e^{i\sum_{k=1}^N \lambda_k h_k}\,\widetilde{\mathsf{E}}(e^{i\lambda_N h_N} \mid \mathscr{F}_{N-1})\big) \\
&= e^{-\frac{\lambda_N^2\sigma_N^2}{2}}\,\widetilde{\mathsf{E}}\big(e^{i\sum_{k=1}^N \lambda_k h_k}\big) = \cdots = e^{-\frac{1}{2}\sum_{k=1}^N \lambda_k^2\sigma_k^2}.
\end{aligned}$$

Thus, with respect to the measure $\widetilde{\mathsf{P}}$ the sequence $(h_n)_{n\leqslant N}$ consists of *independent* normally distributed random variables $h_n \sim \mathscr{N}(0,\sigma_n^2)$, with expectations zero. (We could arrive to the same conclusion on the basis of (17)–(20).)

We now sum up the above results in the following theorem, which we could call the discrete analog of Girsanov's theorem.

THEOREM. *Let $h = (h_n)_{n\leqslant N}$ be a conditionally Gaussian sequence such that*

$$\mathrm{Law}(h_n \mid \mathscr{F}_{n-1}; \mathsf{P}) = \mathscr{N}(\mu_n, \sigma_n^2), \qquad n \leqslant N.$$

Let $\mathscr{F}_N = \mathscr{F}$ and let $\widetilde{\mathsf{P}}$ be the measure defined by formula (13) with density Z_N defined in (7).

Then

1) *the sequence* $h = (h_n)_{n\leqslant N}$ *is conditionally Gaussian with respect to* $\widetilde{\mathsf{P}}$:
$$\mathrm{Law}(h_n \mid \mathscr{F}_{n-1}; \widetilde{\mathsf{P}}) = \mathscr{N}(0, \sigma_n^2), \qquad n \leqslant N;$$
2) *if the* $\sigma_n^2 = \sigma_n^2(\omega)$, $n \leqslant N$, *are independent on* ω, *then* $h = (h_n)_{n\leqslant N}$ *is a sequence of independent Gaussian variables with respect to* $\widetilde{\mathsf{P}}$:
$$\mathrm{Law}(h_n \mid \widetilde{\mathsf{P}}) = \mathscr{N}(0, \sigma_n^2), \qquad n \leqslant N:$$
3) *if* $\mathscr{F} = \bigvee \mathscr{F}_n$ *and condition* (8) *is satisfied, then properties* 1) *and* 2) *hold with respect to the measure* $\widetilde{\mathsf{P}}(d\omega) = Z_\infty(\omega)\,\mathsf{P}(d\omega)$ *for all* $n \geqslant 1$, *where* $Z_\infty(\omega)$ *is as in* (9).

4. We note that the sequences (μ_n) and (σ_n) from the definition of h_n $(= \mu_n + \sigma_n\varepsilon_n)$ are directly involved in our construction of the measure in accordance with (13) and (7). For this reason and also to establish connections with our special choice of the values $a_n(\omega)$ in the discussion of the Esscher transformations (see § 2d), we consider now the sequence of processes $Z^{(b)} = (Z_n^{(b)})_{1\leqslant n\leqslant N}$ defined by the equalities
$$Z_n^{(b)}(\omega) = \exp\biggl\{-\sum_{k=1}^{n} b_k\varepsilon_k - \frac{1}{2}\sum_{k=1}^{n} b_k^2\biggr\}, \tag{22}$$
where the $b_k = b_k(\omega)$ are $\mathscr{F}_{k-1}$-measurable.

Since $\mathsf{E}Z_N^{(b)}(\omega) = 1$, we have a well-defined probability measure
$$\widetilde{\mathsf{P}}^{(b)}(d\omega) = Z_N^{(b)}(\omega)\,\mathsf{P}(d\omega) \tag{23}$$
in $\mathscr{F}_N = \mathscr{F}$.

The expectation with respect to this measure is
$$\widetilde{\mathsf{E}}^{(b)}(h_n \mid \mathscr{F}_{n-1}) = -b_n - \frac{\mu_n}{\sigma_n}. \tag{24}$$
It is now clear why one sets $b_n = -\dfrac{\mu_n}{\sigma_n}$, $n \leqslant N$, in Girsanov's theorem: it is under this choice that the sequence $h = (h_n)_{n\leqslant N}$ becomes a *local martingale difference*.

Further, if $X_n = \mu_n + \varepsilon_n$, then
$$\varphi_n(a;\omega) \equiv \mathsf{E}(e^{aX_n} \mid \mathscr{F}_{n-1}) = e^{\frac{1}{2}a^2 - a\mu_n}.$$
Hence
$$\inf_a \varphi_n(a;\omega) = \varphi_n\bigl(a_n(\omega);\omega\bigr)$$
where $a_n(\omega) = \mu_n$, $n \leqslant N$.

We used just these 'extremal' variables $a_n(\omega)$ in § 2d, in our construction (by means of the Esscher transformation) of a measure making the sequence (X_n) a martingale difference.

Thus, in the present case (of $\sigma_n \equiv 1$) both Girsanov and Esscher transformations bring us to the same measure $\widetilde{\mathsf{P}}$.

5. We now claim that if (12) holds, then the family $Z = (Z_n)_{n\geqslant 1}$ of the variables Z_n defined in (7) is uniformly integrable.

Let $\beta_n = -\dfrac{\mu_n}{\sigma_n}$. Then the condition (12) takes the following form: *there exists* $\delta > 0$ *such that*

$$\mathsf{E}\exp\biggl\{\Bigl(\frac{1}{2}+\delta\Bigr)\sum_{k=1}^{\infty}\beta_k^2\biggr\} < \infty. \tag{25}$$

In accordance with (7),

$$Z_n = \exp\biggl\{\sum_{k=1}^{n}\beta_k\varepsilon_k - \frac{1}{2}\sum_{k=1}^{n}\beta_k^2\biggr\}, \qquad n \geqslant 1. \tag{26}$$

Assume that $\varepsilon > 0$ and $p > 1$. We set

$$\psi_n^{(1)} = \exp\biggl\{(1+\varepsilon)\sum_{k=1}^{n}\beta_k\varepsilon_k - \frac{p(1+\varepsilon)^2}{2}\sum_{k=1}^{n}\beta_k^2\biggr\} \tag{27}$$

and

$$\psi_n^{(2)} = \exp\biggl\{\Bigl(\frac{p(1+\varepsilon)^2}{2} - \frac{1+\varepsilon}{2}\Bigr)\sum_{k=1}^{n}\beta_k^2\biggr\}. \tag{28}$$

To prove the uniform integrability of $(Z_n)_{n\geqslant 1}$ it suffices to show that

$$\sup_n \mathsf{E}Z_n^{1+\varepsilon} < \infty \tag{29}$$

for some $\varepsilon > 0$ (see, e.g., [439; Chapter II, § 6, Lemma 3]). Since

$$Z_n^{1+\varepsilon} = \psi_n^{(1)}\psi_n^{(2)},$$

it follows by Hölder's inequality that

$$\mathsf{E}Z_n^{1+\varepsilon} = \mathsf{E}\psi_n^{(1)}\psi_n^{(2)} \leqslant \bigl[\mathsf{E}\bigl(\psi_n^{(1)}\bigr)^p\bigr]^{1/p}\bigl[\mathsf{E}\bigl(\psi_n^{(2)}\bigr)^q\bigr]^{1/q} = \bigl[\mathsf{E}\bigl(\psi_n^{(2)}\bigr)^q\bigr]^{1/q} \tag{30}$$

$(1/p + 1/q = 1)$, where we use the equality

$$\mathsf{E}\bigl(\psi_n^{(1)}\bigr)^p = 1$$

(see (11)). We set $p = 1+\delta$ and $q = (1+\delta)/\delta$, and we choose $\delta > 0$ such that (25) holds. We now find $\varepsilon > 0$ such that

$$\varepsilon(1+\varepsilon) \leqslant \frac{\delta^2}{(1+\delta)(1+2\delta)}. \tag{31}$$

Then

$$(\psi_n^{(2)})^q \leqslant \exp\biggl\{\biggl(\frac{1}{2}+\frac{\varepsilon q(1+\varepsilon)(1+q)+1}{2(q-1)}\biggr)\sum_{k=1}^{n}\beta_k^2\biggr\}$$
$$\leqslant \exp\biggl\{\biggl(\frac{1}{2}+\delta\biggr)\sum_{k=1}^{n}\beta_k^2\biggr\} \leqslant \exp\biggl\{\biggl(\frac{1}{2}+\delta\biggr)\sum_{k=1}^{\infty}\beta_k^2\biggr\},$$

so that

$$\sup_n \mathsf{E} Z_n^{1+\varepsilon} \leqslant \Bigl[\sup_n \mathsf{E}(\psi_n^{(2)})^q\Bigr]^{1/q} \leqslant \biggl[\mathsf{E}\exp\biggl(\frac{1}{2}+\delta\biggr)\sum_{k=1}^{\infty}\beta_k^2\biggr]^{1/q} < \infty$$

by (25), which proves the required uniform integrability of (Z_n).

§ 3c. Martingale Property of the Prices in the Case of a Conditionally Gaussian and Logarithmically Conditionally Gaussian Distributions

1. Let $(\Omega, \mathscr{F}, (\mathscr{F}_n), \mathsf{P})$, $n \geqslant 0$, be the original filtered probability space. In our discussions of martingale measures turning the sequence of discounted prices $\dfrac{S}{B} = \left(\dfrac{S_n}{B_n}\right)$, where $B = (B_n)$ and $S = (S_n)$, into a martingale we consider first a somewhat *idealized* model of a (B, S)-market: we shall assume that $B_n \equiv 1$,

$$S_n = S_0 + H_n, \qquad n \geqslant 1, \tag{1}$$

where $H_n = \sum_{k=1}^{n} h_k$ and $S_0 = \text{Const}$. We shall also assume that $h = (h_n)$ is a conditionally Gaussian sequence: $h_n = \mu_n + \sigma_n \varepsilon_n$, where the μ_n and σ_n are $\mathscr{F}_{n-1}$-measurable, and $\varepsilon = (\varepsilon_n)$ is a sequence of independent, $\mathscr{N}(0,1)$-distributed, $\mathscr{F}_n$-measurable variables ε_n, $n \geqslant 1$ (see the preceding section for details).

Thus, we allow the prices S_n to take also *negative* values. (This is the 'idealization' we have mentioned.) Note, however, that L. Bachelier [12] considered precisely a model of this kind; see Chapter I, § 2a.

Let $\mu_n \equiv 0$. Then

$$S_n = S_0 + \sum_{k \leqslant n} \sigma_k \varepsilon_k. \tag{2}$$

Since the σ_k are $\mathscr{F}_{k-1}$-measurable and $\mathsf{E}(\varepsilon_k \mid \mathscr{F}_{k-1}) = 0$, it follows by (2) that the sequence of prices $S = (S_n)$ is a martingale transformation and, therefore, a local martingale (see the theorem in Chapter II, § 1c). Assuming additionally, that, e.g., $\mathsf{E}|\sigma_k \varepsilon_k| < \infty$, $k \geqslant 1$, we obtain that $S = (S_n)$ is a martingale with respect to the original measure P. (As regards general conditions ensuring that a local martingale is a martingale, see Chapter II, § 1c).

Assume now that the μ_n do not vanish identically for all $n \leqslant N$. In this case the discrete version of Girsanov's theorem (§ 3b) give us tools for a construction of measures (see formulas (8) and (6) in § 3b) such that the sequences $(S_n)_{n \leqslant N}$ are local martingales with respect to $\widetilde{\mathsf{P}}_N$ (or even are martingales if $\widetilde{\mathsf{E}}|\sigma_n \varepsilon_n| < \infty$ for all $n \leqslant N$).

2. We consider now a more down-to-earth situation where

$$S_n = S_0 e^{H_n} \tag{3}$$

in our (B, S)-market and $B_n \equiv 1$, $n \leqslant N$.

As mentioned in Chapter II, § 1a, the representation (3) of 'compound interest' kind is convenient for a statistical analysis but it is not perfectly convenient for the aims of stochastic analysis. The problem is that to verify the martingale property of the sequence $S = (S_n)$ one would better have a result saying: "in order that a sequence $S = (S_n)$ defined by (3) be a martingale it is sufficient that the sequence $H = (H_n)$ be a martingale". This is, however, not so in general, which explains why we turn to the representation

$$S_n = S_0 \mathscr{E}(\widehat{H})_n \tag{4}$$

('simple interest'), where (see Chapter II, § 1a)

$$\widehat{H}_n = H_n + \sum_{k \leqslant n} (e^{\Delta H_k} - \Delta H_k - 1) \tag{5}$$

and $\mathscr{E}(\widehat{H}) = (\mathscr{E}(\widehat{H})_n)_{n \geqslant 0}$ is the stochastic exponential constructed from $\widehat{H} = (\widehat{H}_n)$ by the formulas

$$\mathscr{E}(\widehat{H})_n = e^{\widehat{H}_n} \prod_{k \leqslant n} (1 + \Delta \widehat{H}_k) e^{-\Delta \widehat{H}_k}, \qquad n \geqslant 1, \tag{6}$$

and $\mathscr{E}(\widehat{H})_0 \equiv 1$.

Of course, we could simplify the right-hand sides in (5) and (6) and write these equations as

$$\widehat{H}_n = \sum_{k \leqslant n} (e^{\Delta H_k} - 1) \tag{7}$$

and

$$\mathscr{E}(\widehat{H})_n = \prod_{k \leqslant n} (1 + \Delta \widehat{H}_k). \tag{8}$$

Still, it is useful to point out again (see Chapter II, § 1a and Chapter III, § 5c) that bearing in mind a similar representation in the *continuous*-time case, we should regard as 'right' formulas ones close to (5) and (6) and not to (7) and (8). This has something to do with the problem of the convergence of the corresponding 'sums' $\sum_{s \leqslant t}$ and 'products' $\prod_{s \leqslant t}$, which, generally speaking, have *inifinitely* many terms and factors for each $t > 0$ in the continuous-time case.

The advantage of (4) over (3) lies in the following result.

PROPOSITION. *In order that a sequence $S = (S_n)$ defined by (4) be a martingale it is sufficient that the sequence $\widehat{H} = (\widehat{H}_n)_{n\geqslant 1}$ be a local martingale with $\Delta\widehat{H}_n \geqslant -1$ for $n \geqslant 1$.*

Indeed, from (6) or (8) we see that

$$\Delta\mathscr{E}(\widehat{H})_n = \mathscr{E}(\widehat{H})_{n-1}\Delta\widehat{H}_n \tag{9}$$

for $n \geqslant 1$. Assume that $(\widehat{H}_n)$ is a local martingale, and therefore is a martingale transformation (see the lemma in Chapter II, § 1c), which admits a representation

$$\widehat{H}_n = \sum_{k=1}^{n} a_k\Delta M_k \tag{10}$$

with $\mathscr{F}_{k-1}$-measurable a_k and some martingale $M = (M_n)$.

From (9) and (10) we see that

$$\Delta\mathscr{E}(\widehat{H})_n = a_n\mathscr{E}(\widehat{H})_{n-1}\Delta M_n,$$

i.e., $\mathscr{E}(\widehat{H})$ is a martingale transformation and, therefore, a local martingale.

If $\Delta\widehat{H}_n \geqslant -1$, then, surely, $\mathscr{E}(\widehat{H})_n \geqslant 0$. Hence, by the lemma in Chapter II, § 1c the local martingale $\mathscr{E}(\widehat{H})$ is actually a martingale.

The condition $\Delta\widehat{H}_n \geqslant -1$ holds in our case because

$$\Delta\widehat{H}_n = e^{\Delta H_n} - 1 \geqslant -1.$$

3. We shall assume that the $h_n = \Delta H_n$ are conditionally Gaussian variables with $h_n = \mu_n + \sigma_n\varepsilon_n$. Then it is natural to say that the sequence $S = (S_n)$ with $S_n = S_0e^{H_n}$ is *logarithmically conditionally Gaussian*, as in the title of § 3c.

First, we consider the question of the conditions ensuring that the sequence $S = (S_n)$ is a martingale with respect to the *original* measure P.

We have already seen that it is sufficient to this end that the sequence $\widehat{H} = (\widehat{H}_n)$ with $\Delta\widehat{H}_n = e^{\Delta H_n} - 1$ be a local martingale, i.e., $\mathsf{E}(|\Delta\widehat{H}_n| \mid \mathscr{F}_{n-1}) < \infty$ and $\mathsf{E}(\Delta\widehat{H}_n \mid \mathscr{F}_{n-1}) = 0$, or, equivalently,

$$\mathsf{E}(e^{\Delta H_n} \mid \mathscr{F}_{n-1}) = 1 \quad (\mathsf{P}\text{-a.s.}). \tag{11}$$

Since we assume that $\Delta H_n = \mu_n + \sigma_n\varepsilon_n$, we can rewrite condition (11) as follows:

$$\mathsf{E}(e^{\mu_n+\sigma_n\varepsilon_n} \mid \mathscr{F}_{n-1}) = 1, \tag{12}$$

which is equivalent to the relation

$$\mathsf{E}(e^{\sigma_n\varepsilon_n} \mid \mathscr{F}_{n-1}) = e^{-\mu_n}. \tag{13}$$

The left-hand side here is equal to $e^{\frac{1}{2}\sigma_n^2}$. Thus, we arrive at the condition

$$\mu_n + \frac{\sigma_n^2}{2} = 0 \quad (\mathsf{P}\text{-a.s.}), \qquad n \geqslant 1, \tag{14}$$

ensuring that the logarithmically conditionally Gaussian sequence

$$S_n = S_0 \exp\biggl\{\sum_{k=1}^{n} (\mu_k + \sigma_k \varepsilon_k)\biggr\}, \qquad n \geqslant 1,$$

is a martingale with respect to P. Of course, this is what one could expect because the sequence

$$\left(\exp\biggl\{\sum_{k=1}^{n}\Bigl(\sigma_k\varepsilon_k - \frac{\sigma_k^2}{2}\Bigr)\biggr\}\right)_{n\geqslant 1} \tag{15}$$

is a *martingale*, as already mentioned.

4. We now proceed to the case when (14) fails.

Assume that $n \leqslant N$. We shall construct the required measure $\widetilde{\mathsf{P}}$ on $\mathscr{F}_N = \mathscr{F}$ by means of the conditional Esscher transformation, in the following form:

$$\widetilde{\mathsf{P}}(d\omega) = Z_N(\omega)\, \mathsf{P}(d\omega)$$

with $Z_N(\omega) = \prod\limits_{1\leqslant n\leqslant N} z_n(\omega)$ and

$$z_n(\omega) = \frac{e^{a_n h_n}}{\mathsf{E}(e^{a_n h_n} \mid \mathscr{F}_{n-1})}, \tag{16}$$

where we shall choose the $\mathscr{F}_{k-1}$-measurable variables $a_k = a_k(\omega)$ (here $\mathscr{F}_0 = \{\varnothing, \Omega\}$) such that the sequence $(S_n)_{n\leqslant N}$ is a $(\widetilde{\mathsf{P}}, (\mathscr{F}_n))$-martingale.

In our case, when the prices are described by formulas (3), this means that, with necessity,

$$\mathsf{E}\bigl[e^{(a_n+1)h_n} \mid \mathscr{F}_{n-1}\bigr] = \mathsf{E}\bigl[e^{a_n h_n} \mid \mathscr{F}_{n-1}\bigr]. \tag{17}$$

Bearing in mind that $h_n = \mu_n + \sigma_n\varepsilon_n$, we see that equality (17) holds if

$$\mu_n + \frac{\sigma_n^2}{2} = -a_n\sigma_n^2, \tag{18}$$

i.e.,

$$a_n = -\frac{\mu_n}{\sigma_n^2} - \frac{1}{2}. \tag{19}$$

If (14) holds for all $n \leqslant N$, then $a_n = 0$ and $Z_N = 1$, i.e., $\widetilde{\mathsf{P}} = \mathsf{P}$.

Choosing a_n in accordance with (19) we obtain

$$\mathsf{E}(e^{a_n h_n} \mid \mathscr{F}_{n-1}) = \exp\left\{-\frac{\mu_n^2}{2\sigma_n^2} + \frac{\sigma_n^2}{8}\right\}.$$

Thus,

$$z_n = \frac{e^{a_n h_n}}{\mathsf{E}(e^{a_n h_n} \mid \mathscr{F}_{n-1})} = \exp\left\{-\left(\frac{\mu_n}{\sigma_n} + \frac{\sigma_n}{2}\right)\varepsilon_n - \frac{1}{2}\left(\frac{\mu_n}{\sigma_n} + \frac{\sigma_n}{2}\right)^2\right\}$$

and

$$Z_N = \exp\left\{-\sum_{n=1}^{N}\left[\left(\frac{\mu_n}{\sigma_n} + \frac{\sigma_n}{2}\right)\varepsilon_n + \frac{1}{2}\left(\frac{\mu_n}{\sigma_n} + \frac{\sigma_n}{2}\right)^2\right]\right\}. \tag{20}$$

Hence the sequence $S = (S_n)_{n \leqslant N}$ with

$$S_n = S_0 e^{H_n}, \quad H_n = h_1 + \cdots + h_n, \quad h_n = \mu_n + \sigma_n \varepsilon_n,$$

is a $\widetilde{\mathsf{P}}$-martingale with $\widetilde{\mathsf{E}} S_n = S_0$, and the density Z_N of the measure $\widetilde{\mathsf{P}}$ with respect to P is described in (20). If (14) holds for $n \leqslant N$, then $\widetilde{\mathsf{P}} = \mathsf{P}$ and $(S_n)_{n \leqslant N}$ is a $(\mathsf{P}, (\mathscr{F}_n))$-martingale, a martingale with respect to the original measure.

§ 3d. Discrete Version of Girsanov's Theorem. General Case

1. We have already mentioned that the discrete version of Girsanov's theorem for the conditionally Gaussian case had paved way to similar results on stochastic sequences $H = (H_n)$, where $h_n = \Delta H_n$ can have a more general structure than the one suggested by the representation $h_n = \mu_n + \sigma_n \varepsilon_n$.

To find generalizations of a 'right' form we shall analyze our proof of the implication

$$\widetilde{\mathsf{P}} \overset{\text{loc}}{\ll} \mathsf{P} \quad \textit{and} \quad \mathsf{E}(h_n \mid \mathscr{F}_{n-1}) = \mu_n \implies \widetilde{\mathsf{E}}(h_n \mid \mathscr{F}_{n-1}) = 0, \ n \geqslant 1$$

(established in the conditionally Gaussian case) once again. This proof in § 3b was based to a considerable extent on the conversion formula for conditional expectations ((4) in § 3a), which has the following form for $\widetilde{\mathsf{P}} \overset{\text{loc}}{\ll} \mathsf{P}$, $Y = H_n$ (where $\widetilde{\mathsf{E}}|H_n| < \infty$), and $m = n - 1$:

$$\widetilde{\mathsf{E}}(H_n \mid \mathscr{F}_{n-1}) = \frac{1}{Z_{n-1}} \mathsf{E}(H_n Z_n \mid \mathscr{F}_{n-1}) \quad (\widetilde{\mathsf{P}}\text{-a.s.}), \qquad n \geqslant 1. \tag{1}$$

Here $\widetilde{\mathsf{E}}$ is averaging with respect to $\widetilde{\mathsf{P}}$ and the right-hand side is set equal to zero if $Z_{n-1}(\omega) = 0$. We also set $\mathscr{F}_0 = \{\varnothing, \Omega\}$ and $Z_0(\omega) \equiv 1$ in what follows.

We claim that one can easily deduce from (1) the following equivalence:

$$\textit{if} \quad \widetilde{\mathsf{P}} \overset{\text{loc}}{\ll} \mathsf{P}, \quad \textit{then} \quad H \in \mathscr{M}(\widetilde{\mathsf{P}}) \iff HZ \in \mathscr{M}(\mathsf{P}), \tag{2}$$

where $\mathscr{M}(\mathsf{P})$ and $\mathscr{M}(\widetilde{\mathsf{P}})$ are the sets of martingales with respect to the measures P and $\widetilde{\mathsf{P}}$ (see Chapter II, § 1c).

Indeed, if $H \in \mathscr{M}(\widetilde{\mathsf{P}})$, then $\widetilde{\mathsf{E}}(H_n \mid \mathscr{F}_{n-1}) = H_{n-1}$ ($\widetilde{\mathsf{P}}$-a.s.), and it follows from (1) that $H_{n-1}Z_{n-1} = \mathsf{E}(H_n Z_n \mid \mathscr{F}_{n-1})$ ($\widetilde{\mathsf{P}}$-a.s.).

This equality holds not only $\widetilde{\mathsf{P}}$-a.s., but also P-a.s., for its both sides vanish on the set $\{Z_{n-1} = 0\}$ because we also have $Z_n = 0$ (P-a.s.) in this set. As regards the set $\{Z_{n-1} > 0\}$, the measures P and $\widetilde{\mathsf{P}}$ are equivalent there (in the sense of the equivalence $\mathsf{P}(\{Z_{n-1} > 0\} \cap A) = 0 \iff \widetilde{\mathsf{P}}(\{Z_{n-1} > 0\} \cap A) = 0$ for $A \in \mathscr{F}_{n-1}$), therefore the left-hand and the right-hand sides of this equality coincide also P-almost surely. Thus, we have proved the implication $\Longrightarrow$ in (2).

In a similar way, if $HZ \in \mathscr{M}(\mathsf{P})$, then $H_{n-1}Z_{n-1} = \mathsf{E}(H_n Z_n \mid \mathscr{F}_{n-1})$ (P and $\widetilde{\mathsf{P}}$-a.s.). Since $Z_{n-1} > 0$ $\widetilde{\mathsf{P}}$-a.s., it follows that

$$H_{n-1} = \frac{1}{Z_{n-1}} \mathsf{E}(H_n Z_n \mid \mathscr{F}_{n-1}) \quad (\widetilde{\mathsf{P}}\text{-a.s.}).$$

Hence it follows by (1) that $H \in \mathscr{M}(\widetilde{\mathsf{P}})$.

It is worth noting that the Bayes's formula (4) in § 3a (and, therefore, also formula (1)) is a consequence of the implication $\Longrightarrow$ in (2). For let Y be a $\mathscr{F}_n$-measurable variable such that $\widetilde{\mathsf{E}}|Y| < \infty$ and $\widetilde{\mathsf{P}} \overset{\text{loc}}{\ll} \mathsf{P}$. Then we consider the martingale $(H_m, \mathscr{F}_m, \widetilde{\mathsf{P}})_{m \leqslant n}$, where $H_m = \widetilde{\mathsf{E}}(Y \mid \mathscr{F}_m)$. By (2) we obtain

$$\mathsf{E}(Y Z_n \mid \mathscr{F}_m) = H_m Z_m \quad (\mathsf{P}\text{-a.s.}).$$

In particular, $\mathsf{E}(Y Z_n \mid \mathscr{F}_{n-1}) = H_{n-1} Z_{n-1}$ (P and $\widetilde{\mathsf{P}}$-a.s.).

Since $\widetilde{\mathsf{P}}(Z_{n-1} > 0) = 1$, it therefore follows that

$$\frac{1}{Z_{n-1}} \mathsf{E}(Y Z_n \mid \mathscr{F}_{n-1}) = H_{n-1} \quad (\widetilde{\mathsf{P}}\text{-a.s.}).$$

Together with the equality $H_{n-1} = \widetilde{\mathsf{E}}(Y \mid \mathscr{F}_{n-1})$ ($\widetilde{\mathsf{P}}$-a.s.), this proves the required formula (4) in the 'conversion lemma' in § 3a:

$$\widetilde{\mathsf{P}} \overset{\text{loc}}{\ll} \mathsf{P} \quad \textit{and} \quad \widetilde{\mathsf{E}}(Y \mid \mathscr{F}_{n-1}) = \frac{1}{Z_{n-1}} \mathsf{E}(Y Z_n \mid \mathscr{F}_{n-1}) \ (\widetilde{\mathsf{P}}\text{-a.s.}).$$

Hence the property (2) is a sort of a 'martingale' version of the conversion lemma for absolutely continuous changes of measures.

2. It can be useful to formulate (2), together with its 'local' version, as the following result (cf. [250; Chapter III, § 3b]).

LEMMA. *Assume that $\widetilde{\mathsf{P}} \overset{\text{loc}}{\ll} \mathsf{P}$ and let $Z = (Z_n)$ be the density process:*

$$Z_n = \frac{d\widetilde{\mathsf{P}}_n}{d\mathsf{P}_n}$$

with $\widetilde{\mathsf{P}}_n = \widetilde{\mathsf{P}} \mid \mathscr{F}_n$ and $\mathsf{P}_n = \mathsf{P} \mid \mathscr{F}_n$.

Let $H = (H_n, \mathscr{F}_n)$ be a stochastic sequence. Then

a) *H is a $\widetilde{\mathsf{P}}$-martingale ($H \in \mathscr{M}(\widetilde{\mathsf{P}})$) if and only if the sequence $HZ = (H_n Z_n, \mathscr{F}_n)$ is a P-martingale ($HZ \in \mathscr{M}(\mathsf{P})$):*

$$H \in \mathscr{M}(\widetilde{\mathsf{P}}) \iff HZ \in \mathscr{M}(\mathsf{P}).$$

b) *If, in addition, $\widetilde{\mathsf{P}} \overset{\text{loc}}{\sim} \mathsf{P}$, then H is a local $\widetilde{\mathsf{P}}$-martingale ($H \in \mathscr{M}_{\text{loc}}(\widetilde{\mathsf{P}})$) if and only if HZ is a local P-martingale ($HZ \in \mathscr{M}_{\text{loc}}(\mathsf{P})$):*

$$H \in \mathscr{M}_{\text{loc}}(\widetilde{\mathsf{P}}) \iff HZ \in \mathscr{M}_{\text{loc}}(\mathsf{P}). \tag{3}$$

Proof. a) We have already proved this using formula (1) (which is also of interest on its own right). However, this can also be proved directly, starting from the definition of a martingale.

We choose $m \leqslant n$ and $A \in \mathscr{F}_m$. Then $\widetilde{\mathsf{E}}(I_A H_n) = \mathsf{E}(I_A H_n Z_n)$, so that

$$\widetilde{\mathsf{E}}(I_A H_n) = \widetilde{\mathsf{E}}(I_A H_m) \iff \mathsf{E}(I_A H_n Z_n) = \mathsf{E}(I_A H_m Z_n).$$

However, $\mathsf{E}(I_A H_m Z_n) = \mathsf{E}(I_A H_m Z_m)$. Hence $H \in \mathscr{M}(\widetilde{\mathsf{P}}) \iff HZ \in \mathscr{M}(\mathsf{P})$.

b) We claim that $HZ \in \mathscr{M}_{\text{loc}}(\mathsf{P}) \implies H \in \mathscr{M}_{\text{loc}}(\widetilde{\mathsf{P}})$ (even if we assume only that $\widetilde{\mathsf{P}} \overset{\text{loc}}{\ll} \mathsf{P}$).

Let (τ_n) be a localizing sequence for HZ and let $\tau = \lim \tau_n$. Then $\mathsf{P}(\tau = \infty) = 1$ and $\widetilde{\mathsf{P}}(\tau < \infty) = \mathsf{E} Z_\tau I(\tau < \infty) = 0$ (since $\widetilde{\mathsf{P}} \overset{\text{loc}}{\ll} \mathsf{P}$ and by property d) in the theorem proved in § 3a). Hence $\widetilde{\mathsf{P}}(\tau = \infty) = 1$.

We note that

$$(H^{\tau_n} Z)_k = H_k^{\tau_n} Z_k = (H_k Z_k)^{\tau_n} + H_{\tau_n}\big(Z_k - Z_{\tau_n} I(k \geqslant \tau_n)\big).$$

Consequently, $H^{\tau_n} Z$ is a P-martingale and $H^{\tau_n} \in \mathscr{M}(\widetilde{\mathsf{P}})$ by assertion a). However, $\widetilde{\mathsf{P}}(\lim \tau_n = \infty) = 1$. Hence $H \in \mathscr{M}_{\text{loc}}(\widetilde{\mathsf{P}})$.

We claim now, that, conversely, if $\widetilde{\mathsf{P}} \overset{\text{loc}}{\sim} \mathsf{P}$, then $H \in \mathscr{M}_{\text{loc}}(\widetilde{\mathsf{P}}) \implies HZ \in \mathscr{M}_{\text{loc}}(\mathsf{P})$.

Let (σ_n) be a localizing sequence for $H \in \mathscr{M}_{\mathrm{loc}}(\widetilde{\mathsf{P}})$. Then $\widetilde{\mathsf{P}}(\lim \sigma_n = \infty) = 1$ and $H^{\sigma_n} \in \mathscr{M}(\widetilde{\mathsf{P}})$. Hence $H^{\sigma_n} Z \in \mathscr{M}(\mathsf{P})$ by a), and since

$$(HZ)_k^{\sigma_n} = H_k^{\sigma_n} Z_k - H_{\sigma_n}\big(Z_k - Z_{\sigma_n} I(k \geqslant \tau_n)\big)$$

(cf. the above formula for $(H^{\tau_n} Z)_k$), it follows that $(HZ)_k^{\sigma_n} \in \mathscr{M}(\mathsf{P})$. However, $\mathsf{P} \overset{\mathrm{loc}}{\ll} \widetilde{\mathsf{P}}$, so that $\mathsf{P}(\lim \sigma_n = \infty) = 1$. Hence $HZ \in \mathscr{M}_{\mathrm{loc}}(\mathsf{P})$.

The proof is complete.

3. Properties (1), (2), and (3) are of fundamental importance for the problem of the verification of the martingale property with respect to one or another measure $\widetilde{\mathsf{P}}$ for the sequences (H_n), $(\widehat{H}_n)$, (S_n), etc., because they reduce this task to the verification of the martingale property with respect to the original, basic measure P for the sequences $(H_n Z_n)$, $(\widehat{H}_n Z_n)$, $(S_n Z_n)$.

In fact, we have already used this in our proof of the discrete analog of Girsanov's theorem in the conditionally Gaussian case, when $H_n = h_1 + \cdots + h_n$ and $h_n = \mu_n + \sigma_n \varepsilon_n$, $n \leqslant N$, and the $\widetilde{\mathsf{P}}_N$ are the measures with densities

$$Z_N = \exp\left\{-\sum_{k=1}^{N} \frac{\mu_k}{\sigma_k}\varepsilon_k - \frac{1}{2}\sum_{k=1}^{N}\left(\frac{\mu_k}{\sigma_k}\right)^2\right\}, \tag{4}$$

explicitly constructed from (μ_k) and (σ_k).

However, the general case of the construction of the measures $\widetilde{\mathsf{P}}_N$ is considerably more complicated. The simplicity of the formulas for the Z_N in the conditionally Gaussian case is essentially the effect of the simple formulas for the H_n: $\Delta H_n \equiv h_n = \mu_n + \sigma_n \varepsilon_n$.

There exist various generalizations of Girsanov's theorem in the discrete-time case.

For a more clear perception of the results below as some generalizations of this theorem it seems reasonable to reformulate its already established version (i.e., the theorem in § 3b) in the conditionally Gaussian case.

We set

$$\alpha_n = \frac{Z_n}{Z_{n-1}} I(Z_{n-1} > 0). \tag{5}$$

Then

$$\alpha_n = \exp\left\{-\frac{\mu_n}{\sigma_n}\varepsilon_n - \frac{1}{2}\left(\frac{\mu_n}{\sigma_n}\right)^2\right\}, \tag{6}$$

and if $M_n = \sum_{k=1}^{n} \sigma_k \varepsilon_k$, then $M = (M_n) \in \mathscr{M}_{\mathrm{loc}}(\mathsf{P})$ and it is easy to show that

$$\mathsf{E}(\alpha_n \Delta M_n \mid \mathscr{F}_{n-1}) = -\mu_n. \tag{7}$$

Thus, leaving aside the questions of integrability for a while, in the conditionally Gaussian case we can state the theorem in § 3b as follows:

$$
\begin{aligned}
M \in \mathscr{M}_{\mathrm{loc}}(\mathsf{P}) &\Longleftrightarrow \mathsf{E}(\sigma_n\varepsilon_n \mid \mathscr{F}_{n-1}) = 0, \quad n \leqslant N \\
&\Longleftrightarrow \mathsf{E}(\mu_n + \sigma_n\varepsilon_n \mid \mathscr{F}_{n-1}) = \mu_n, \quad n \leqslant N \\
&\Longleftrightarrow \mathsf{E}(h_n \mid \mathscr{F}_{n-1}) = \mu_n, \quad n \leqslant N \\
&\Longrightarrow \widetilde{\mathsf{E}}(h_n \mid \mathscr{F}_{n-1}) = 0, \quad n \leqslant N \\
&\Longleftrightarrow \widetilde{\mathsf{E}}(\Delta M_n + \mu_n \mid \mathscr{F}_{n-1}) = 0, \quad n \leqslant N \\
&\Longleftrightarrow \widetilde{\mathsf{E}}\big(\Delta M_n - \mathsf{E}(\alpha_n \Delta M_n \mid \mathscr{F}_{n-1}) \mid \mathscr{F}_{n-1}\big) = 0, \quad n \leqslant N.
\end{aligned}
$$

In other words, the inclusion $M \in \mathscr{M}_{\mathrm{loc}}(\mathsf{P})$ means that the sequence $\widetilde{M} = (\widetilde{M}_n)_{n \leqslant N}$, where

$$
\widetilde{M}_n = M_n - \sum_{k=1}^{n} \mathsf{E}(\alpha_k \Delta M_k \mid \mathscr{F}_{k-1}), \tag{8}
$$

is a local martingale with respect to the measure $\widetilde{\mathsf{P}}(d\omega) = Z_N(\omega)\,\mathsf{P}(d\omega)$:

$$
M \in \mathscr{M}_{\mathrm{loc}}(\mathsf{P}) \implies \widetilde{M} \in \mathscr{M}_{\mathrm{loc}}(\widetilde{\mathsf{P}}). \tag{9}
$$

We emphasize an important point in the above analysis. In our discussion of the sequence $H = (H_n)$ with $\Delta H_n = \mu_n + \Delta M_n$ we were primarily interested in the martingale component of H. In effect, we 'traced' the change of the martingale component under a continuous change of measure. As we see, if we consider the measure $\widetilde{\mathsf{P}}$, then (M_n) is no longer a martingale: it has the representation

$$
M_n = \sum_{k=1}^{n} \mathsf{E}(\alpha_k \Delta M_k \mid \mathscr{F}_{k-1}) + \widetilde{M}_n,
$$

where $\widetilde{M} = (\widetilde{M}_n)$ is a $\widetilde{\mathsf{P}}$-martingale and $A = \big(A_n = \sum_{k=1}^{n} \mathsf{E}(\alpha_k \Delta M_k \mid \mathscr{F}_{k-1})\big)$ is some predictable drift. It is the appearance of this additional 'drift' term after an absolutely continuous change of measure that enables one to 'kill' the drift components of the original sequences $H = (H_n)$ by means of a transition to measures $\widetilde{\mathsf{P}}$ that are absolutely continuous (or locally absolutely continuous) with respect to P.

4. This interpretation of the above discrete version of Girsanov's theorem (in the conditionally Gaussian case) enables us to state the following general result on local martingales, where we *do not specify* that $M_n = \sum\limits_{k=1}^{n} \sigma_k \varepsilon_k$.

THEOREM 1. *Let* $M \in \mathscr{M}_{\text{loc}}(\mathsf{P})$ *with* $M_0 = 0$. *Assume that* $\widetilde{\mathsf{P}} \overset{\text{loc}}{\ll} \mathsf{P}$, *let* $Z_n = \dfrac{d\widetilde{\mathsf{P}}_n}{d\mathsf{P}_n}$, $n \geqslant 1$, *be the corresponding densities, and let* $\alpha_n = \dfrac{Z_n}{Z_{n-1}} I(Z_{n-1} > 0)$, *where* $Z_0 \equiv 1$. *Also, let*

$$\mathsf{E}\big(|\Delta M_n|\alpha_n \,|\, \mathscr{F}_{n-1}\big) < \infty \quad (\mathsf{P}\text{-a.s.}), \qquad n \geqslant 1. \tag{10}$$

Then the process $\widetilde{M} = (\widetilde{M}_n)$ *defined in* (8) *belongs to* $\mathscr{M}_{\text{loc}}(\widetilde{\mathsf{P}})$ (*i.e., is a local* $\widetilde{\mathsf{P}}$-*martingale*).

Proof. Again (as in the proof for the conditionally Gaussian case in § 3b), we use Bayes's formula (4) in § 3a:

$$\begin{aligned}
\widetilde{\mathsf{E}}(M_n \,|\, \mathscr{F}_{n-1}) &= \mathsf{E}(M_n\alpha_n \,|\, \mathscr{F}_{n-1}) \\
&= \mathsf{E}\big(\alpha_n(M_n - M_{n-1}) \,|\, \mathscr{F}_{n-1}\big) + \mathsf{E}\big(\alpha_n M_{n-1} \,|\, \mathscr{F}_{n-1}\big) \\
&= \mathsf{E}\big(\alpha_n \Delta M_n \,|\, \mathscr{F}_{n-1}\big) + M_{n-1}.
\end{aligned} \tag{11}$$

Hence, by assumption (10), we obtain

$$\widetilde{\mathsf{E}}\big(|M_n| \,|\, \mathscr{F}_{n-1}\big) \leqslant \mathsf{E}\big(|\alpha_n \Delta M_n| \,|\, \mathscr{F}_{n-1}\big) + |M_{n-1}| < \infty$$

(P and $\widetilde{\mathsf{P}}$-a.s.)

From (11) and (8) we immediately see that $\widetilde{\mathsf{E}}\big(|\widetilde{M}_n| \,|\, \mathscr{F}_{n-1}\big) < \infty$ and

$$\widetilde{\mathsf{E}}\big(\widetilde{M}_n \,|\, \mathscr{F}_{n-1}\big) = \widetilde{M}_{n-1}, \tag{12}$$

i.e., $\widetilde{M}$ is a local $\widetilde{\mathsf{P}}$-martingale ($\widetilde{M} \in \mathscr{M}_{\text{loc}}(\widetilde{\mathsf{P}})$) by the theorem in Chapter II, § 1c.1.

5. Now, assume that the basic sequence $H = (H_n)_{n \geqslant 1}$ has the representation

$$H_n = A_n + M_n, \tag{13}$$

where $A = (A_n)_{n \geqslant 1}$ is a predictable sequence (the A_n are $\mathscr{F}_{n-1}$-measurable for $n \geqslant 1$, where $\mathscr{F}_0 = \{\varnothing, \Omega\}$, and $A_0 = 0$) and $M = (M_n)_{n \geqslant 1} \in \mathscr{M}_{\text{loc}}(\mathsf{P})$.

Since $\mathsf{E}(|\Delta M_n| \,|\, \mathscr{F}_{n-1}) < \infty$ for a local martingale, it follows that

$$\mathsf{E}(|\Delta H_n| \,|\, \mathscr{F}_{n-1}) \leqslant |\Delta A_n| + \mathsf{E}(|\Delta M_n| \,|\, \mathscr{F}_{n-1}) < \infty,$$

so that the H_n, $n \geqslant 1$, have also the representations

$$H_n = \sum_{k=1}^{n} \mathsf{E}(\Delta H_k \,|\, \mathscr{F}_{k-1}) + \sum_{k=1}^{n} \big[\Delta H_k - \mathsf{E}(\Delta H_k \,|\, \mathscr{F}_{k-1})\big], \tag{14}$$

which we have called the *generalized Doob decomposition* (see Chapter II, § 1b) of the sequence $H = (H_n)_{n\geqslant 1}$.

As in the case of the usual Doob decomposition, the representation of the form (13) with predictable (A_n) is unique, and therefore

$$A_n = \sum_{k=1}^{n} \mathsf{E}(\Delta H_k \,|\, \mathscr{F}_{k-1}) \tag{15}$$

and

$$M_n = \sum_{k=1}^{n} \big[\Delta H_k - \mathsf{E}(\Delta H_k \,|\, \mathscr{F}_{k-1})\big]. \tag{16}$$

Theorem 1 above has the following simple generalization.

THEOREM 2. *Let $H = (H_n)_{n\geqslant 1}$ be a sequence with generalized Doob decomposition (14) and assume that (10) holds. Let $\widetilde{\mathsf{P}}$ be a measure such that $\widetilde{\mathsf{P}} \overset{\text{loc}}{\ll} \mathsf{P}$. Then the sequence $H = (H_n)_{n\geqslant 1}$ has a representation*

$$H_n = \widetilde{A}_n + \widetilde{M}_n, \tag{17}$$

or, equivalently,

$$H_n = \sum_{k=1}^{n} \widetilde{\mathsf{E}}(\Delta H_k \,|\, \mathscr{F}_{k-1}) + \sum_{k=1}^{n} \big[\Delta H_k - \widetilde{\mathsf{E}}(\Delta H_k \,|\, \mathscr{F}_{k-1})\big] \tag{18}$$

(a generalized Doob decomposition), *where*

$$\widetilde{A}_n = A_n + \sum_{k=1}^{n} \mathsf{E}(\alpha_k \Delta M_k \,|\, \mathscr{F}_{k-1}) \tag{19}$$

and the sequence $\widetilde{M}$ of the variables

$$\widetilde{M}_n = M_n - \sum_{k=1}^{n} \mathsf{E}(\alpha_k \Delta H_k \,|\, \mathscr{F}_{k-1}); \tag{20}$$

is a local $\widetilde{\mathsf{P}}$-martingale ($\widetilde{M} \in \mathscr{M}_{\text{loc}}(\widetilde{\mathsf{P}})$).

Proof. This is an immediate consequence of Theorem 1 applied to the sequence $M = (M_n)_{n\geqslant 1}$ with $M_n = H_n - A_n$.

6. Under the hypothesis of Theorem 2 we assume now that M and Z are (locally) *square* integrable martingales. Then they have a well-defined *predictable quadratic covariance* $\langle M, Z\rangle = (\langle M, Z\rangle_n)_{n\geqslant 0}$, where

$$\langle M, Z\rangle_n = \sum_{k=1}^{n} \mathsf{E}\big(\Delta M_k \Delta Z_k \mid \mathscr{F}_{k-1}\big). \tag{21}$$

One sometimes calls $\langle M, Z\rangle$ simply the 'angular brackets' of M and Z; as regards the corresponding definitions in the continuous-time case, see, e.g., [250; Chapter II, § § 4a, b] or [439; Chapter VII, § 1] and Chapter III, § 5b. We also recall that the *quadratic covariance* of the sequences $X = (X_n)_{n\geqslant 0}$ and $Y = (Y_n)_{n\geqslant 0}$ is the sequence $[X, Y]$ of variables

$$[X, Y]_n = \sum_{k=1}^{n} \Delta X_k \Delta Y_k. \tag{22}$$

By (21) and (22) we obtain that, in the case of (locally) square integrable martingales, the difference $[M, Z] - \langle M, Z\rangle$ is a local martingale (see [250; Chapter I, § 4e]).
We shall now assume that $\widetilde{\mathsf{P}} \overset{\text{loc}}{\sim} \mathsf{P}$. Then $Z_n > 0$ ($\widetilde{\mathsf{P}}$ and P-a.s.) and

$$\begin{aligned}\frac{\Delta\langle M, Z\rangle_n}{Z_{n-1}} &= \frac{\mathsf{E}[\Delta M_n \Delta Z_n \mid \mathscr{F}_{n-1}]}{Z_{n-1}} = \mathsf{E}\big[(\alpha_n - 1)\Delta M_n \mid \mathscr{F}_{n-1}\big] \\ &= \mathsf{E}\big[\alpha_n \Delta M_n \mid \mathscr{F}_{n-1}\big].\end{aligned} \tag{23}$$

Note that if $\langle M\rangle_n(\omega) = 0$ where $\langle M\rangle$ ($\equiv \langle M, M\rangle$) is the quadratic characteristic, then also $\langle M, Z\rangle_n(\omega) = 0$. Hence the left-hand side of (23) can be rewritten as follows (here $\langle M\rangle_n = \langle M, M\rangle_n$):

$$\frac{\Delta\langle M, Z\rangle_n}{Z_{n-1}} = -a_n \Delta\langle M\rangle_n, \tag{24}$$

where

$$a_n = -\frac{\Delta\langle M, Z\rangle_n}{\Delta\langle M\rangle_n Z_{n-1}}$$

and where we set $\dfrac{\Delta\langle M, Z\rangle_n}{\Delta\langle M\rangle_n}$ to be equal, say, to 1 if $\Delta\langle M\rangle_n = 0$.

Thus, (19) can be written as follows:

$$\widetilde{A}_n = A_n - \sum_{k=1}^{n} a_k \Delta\langle M\rangle_k. \tag{25}$$

Hence we can make an important conclusion as regards the structure of the original sequence H (with respect to P): *if this sequence is a local martingale with respect to a measure* $\widetilde{\mathsf{P}} \overset{\text{loc}}{\sim} \mathsf{P}$ (*i.e.*, $\widetilde{A} \equiv 0$), *then*

$$H_n = \sum_{k=1}^{n} a_k \Delta\langle M\rangle_k + M_n, \qquad n \geqslant 1, \tag{26}$$

or, in terms of increments,

$$\Delta H_n = a_n \Delta\langle M\rangle_n + \Delta M_n, \qquad n \geqslant 1. \tag{27}$$

7. So far we have based our arguments on the mere *existence* of the measure $\widetilde{\mathsf{P}}$, without specifying its structure or the structure of the sequence $\alpha = (\alpha_n)$ involved in the definition of the Radon–Nikodym derivatives

$$\frac{d\widetilde{\mathsf{P}}_n}{d\mathsf{P}_n} = \prod_{k=1}^{n} \alpha_k, \qquad n \geqslant 1. \tag{28}$$

By (23) and (24) we obtain

$$a_n \Delta\langle M\rangle_n = \mathsf{E}\big[(1-\alpha_n)\Delta M_n \,|\, \mathscr{F}_{n-1}\big]. \tag{29}$$

This relation can be regarded as an *equation* with respect to the ($\mathscr{F}_n$-measurable) variable α_n, and one can see that it has the following (*not necessarily unique*) solution:

$$\alpha_n = 1 - a_n \Delta M_n. \tag{30}$$

Of course, only those solutions α_n satisfying the relation $\mathsf{P}(\alpha_n > 0) = 1$, $n \geqslant 1$, are suitable for our aims. If this holds, then

$$\frac{d\widetilde{\mathsf{P}}_n}{d\mathsf{P}_n} = \prod_{k=1}^{n} (1 - a_k \Delta M_k) = \mathscr{E}\Big(-\sum_{k\leqslant\cdot} a_k \Delta M_k\Big)_n, \tag{31}$$

where $\mathscr{E} = (\mathscr{E}(R)_n)$ is the stochastic exponential (see Chapter II, § 1):

$$\mathscr{E}(R)_n = e^{R_n} \prod_{k\leqslant n} (1 + \Delta R_k) e^{-\Delta R_k} = \prod_{k\leqslant n} (1 + \Delta R_k). \tag{32}$$

Let $\widetilde{\mathsf{P}}$ be a probability measure such that its restrictions $\widetilde{\mathsf{P}}_n = \widetilde{\mathsf{P}} \,|\, \mathscr{F}_n$ can be recovered by formulas (31). Our original sequence $H = (H_n)$, which satisfies (27), becomes a *local martingale* with respect to this measure because $\Delta\widetilde{A}_n = a_n \Delta\langle M\rangle_n + \mathsf{E}(\alpha_n \Delta M_n \,|\, \mathscr{F}_{n-1}) = 0$ for $n \geqslant 1$ and $\widetilde{A}_0 = 0$.

We have already pointed out that this probability measure $\widetilde{\mathsf{P}}$, a *martingale measure*, is not unique in general. However, it has certain advantages. First, it can be *explicitly* constructed from the coefficients a_n. Second, it *has* certain properties of 'minimality', which justify its name of *minimal measure* (see [429] and Chapter VI, § 3d.6).

§ 3e. Integer-Valued Random Measures and Their Compensators. Transformation of Compensators under Absolutely Continuous Changes of Measures. 'Stochastic Integrals'

1. Let $H = (H_n)_{n \geqslant 0}$ be a stochastic sequence of random variables $H_n = H_n(\omega)$ defined on a filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_n)_{n \geqslant 0}, \mathsf{P})$. We shall assume that $H_0 = 0$ and $\mathscr{F}_0 = \{\varnothing, \Omega\}$.

There exist two ways to describe the probability distribution $\operatorname{Law}(H)$ of H: in terms of the *unconditional* distributions of the variables $H_1, H_2, \dots, H_n$,

$$\operatorname{Law}(H_1, H_2, \dots, H_n) \tag{1}$$

(or, equivalently, $\operatorname{Law}(\Delta H_1, \Delta H_2, \dots, \Delta H_n)$) or in terms of the (regular) *conditional* distributions of the variables ΔH_n:

$$\mathsf{P}(\Delta H_n \in \cdot \mid \mathscr{F}_{n-1}), \qquad n \geqslant 1. \tag{2}$$

The second way has certain advantages because, given conditional distributions, one can, of course, find the unconditional ones. (In addition, the conditional distributions exhibit the dependence of the ΔH_n on the 'past' in a more explicit form.) On the other hand, starting from unconditional distributions (1) one can recover *only* the conditional distributions $\mathsf{P}(\Delta H_n \in \cdot \mid \mathscr{F}^H_{n-1})$, where $\mathscr{F}^H_{n-1} = \sigma(\omega\colon H_1, \dots, H_{n-1})$, so that $\mathscr{F}^H_{n-1} \subseteq \mathscr{F}_{n-1}$ (this can also be a proper inclusion).

2. We consider a d-dimensional stochastic sequence

$$X = (X_n, \mathscr{F}_n)_{n \geqslant 0} \tag{3}$$

on a filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_n)_{n \geqslant 0}, \mathsf{P})$. Let $X_0 = 0$ and let $\mathscr{F}_0 = \{\varnothing, \Omega\}$.

We associate with X the sequence $\mu = (\mu_n(\cdot))_{n \geqslant 1}$ *of integer-valued random measures* defined as follows:

$$\mu_n(A; \omega) = I_A(\Delta X_n(\omega)), \qquad A \in \mathscr{B}(\mathbb{R}^d),$$

i.e.,

$$\mu_n(A; \omega) = \begin{cases} 1 & \text{if } \Delta X_n(\omega) \in A, \\ 0 & \text{if } \Delta X_n(\omega) \notin A. \end{cases}$$

Further, let $\nu = (\nu_n(\cdot))_{n \geqslant 1}$ be the sequence of *regular* conditional distributions $\nu_n(\cdot)$ of the variables ΔX_n with respect to the algebras $\mathscr{F}_{n-1}$, i.e., of the functions $\nu_n(A; \omega)$ (defined for $A \in \mathscr{B}(\mathbb{R}^d)$ and $\omega \in \Omega$) such that

1) $\nu_n(\cdot\,; \omega)$ is a probability measure on $(\mathbb{R}^d, \mathscr{B}(\mathbb{R}^d))$ for each $\omega \in \Omega$;
2) $\nu_n(A; \omega)$, regarded as a function of ω for fixed $A \in \mathscr{B}(\mathbb{R}^d)$, is some realization of the conditional probability $\mathsf{P}(\Delta X_n \in A \mid \mathscr{F}_{n-1})(\omega)$, i.e.,

$$\nu_n(A; \omega) = \mathsf{P}(\Delta X_n \in A \mid \mathscr{F}_{n-1})(\omega) \quad (\mathsf{P}\text{-a.s.}).$$

(The proof of the *existence* of such a realization of the conditional probability can be found, e.g., in [439; Chapter II, § 7].)

For regular conditional distributions the conditional expectations

$$\mathsf{E}[f(\Delta X_n) \mid \mathscr{F}_{n-1}](\omega)$$

can be calculated for a non-negative or bounded function f by means of *integration* over the regular conditional distributions $\nu_n(\,\cdot\,)$ *for fixed* ω:

$$\mathsf{E}\big[f(\Delta X_n) \mid \mathscr{F}_{n-1}\big](\omega) = \int_{\mathbb{R}^d} f(x)\,\nu_n(dx;\omega) \quad (\mathsf{P}\text{-a.s.}).$$

Thus, in the case under consideration we have

$$\nu_n(A;\,\cdot\,) = \mathsf{E}\big[\mu_n(A;\omega) \mid \mathscr{F}_{n-1}\big](\,\cdot\,),$$

so that for each $A \in \mathscr{B}(\mathbb{R}^d)$ the sequence

$$\big(\mu_n(A) - \nu_n(A)\big)_{n \geqslant 1}$$

with $\mu_n(A) = \mu_n(A;\omega)$ and $\nu_n(A) = \nu_n(A;\omega)$ is a *martingale difference* with respect to the measure P and the flow $(\mathscr{F}_n)$.

We set

$$\mu_{(0,n]}(A;\omega) = \sum_{k=1}^{n} \mu_k(A;\omega) \quad \text{and} \quad \nu_{(0,n]}(A;\omega) = \sum_{k=1}^{n} \nu_k(A;\omega).$$

Then, clearly, for each $A \in \mathscr{B}(\mathbb{R}^d)$ the sequence

$$\big(\mu_{(0,n]}(A;\omega) - \nu_{(0,n]}(A;\omega)\big)_{n \geqslant 1}$$

is a *martingale*. This property explains why one calls the (random) measure $\nu_{(0,n]}(\,\cdot\,)$ the *compensator* of the (random) measure $\mu_{(0,n]}(\,\cdot\,)$ and why one calls the sequence

$$\mu - \nu = \big(\mu_{(0,n]}(\,\cdot\,) - \nu_{(0,n]}(\,\cdot\,)\big)_{n \geqslant 1}$$

a *random martingale measure.*

Note that the representation

$$\mu = \nu + (\mu - \nu)$$

of the measure $\mu = (\mu_{(0,n]})_{n \geqslant 1}$ with predictable measure $\nu = (\nu_{(0,n]})_{n \geqslant 1}$ is a kind of the *Doob decomposition* (Chapter II, § 1b) into the sum of a predictable and a martingale components.

Remark. One can also associate with the sequence $X = (X_n, \mathscr{F}_n)_{n\geqslant 0}$ the integer-valued random measures of *jumps* $\mu^X = (\mu^X_{(0,n]}(\,\cdot\,))_{n\geqslant 1}$ where

$$\mu^X_{(0,n]}(A;\omega) = \sum_{k=1}^{n} \mu^X_k(A;\omega)$$

and

$$\mu^X_k(A;\omega) = I\bigl(\Delta X_k(\omega) \in A,\, \Delta X_k(\omega) \neq 0\bigr).$$

Clearly, if $A \in \mathscr{B}(\mathbb{R}^d \setminus \{0\})$, then $\mu_n(A;\omega) = \mu^X_n(A;\omega)$. All distinctions between these measures are concentrated at the 'no-jump' events $\{\omega\colon \Delta X_n(\omega) = 0\}$. If $\mathsf{P}\{\omega\colon \Delta X_n(\omega) = 0\} = 0$, then the measures μ and μ^X are essentially the same.

We note that these are the random jump measures μ^X, rather than μ, that play the *central role* in the *continuous-time* case, in the description of the properties of the jump components of stochastic processes in terms of integer-valued random measures. (See Chapter VII, § 3a of this book and [250; Chapter II, 1.16] for greater detail.)

3. In this subsection we consider the 'stochastic integrals'

$$w * \mu, \qquad w * \nu, \qquad w * (\mu - \nu)$$

with respect to the just introduced random measures μ, ν, and $\mu - \nu$.

Let $w = (w_k(\omega, x))_{k\geqslant 1}$ be a sequence of $\mathscr{F} \otimes \mathscr{B}(\mathbb{R}^d)$-measurable functions. We denote by $w * \mu$ the (dependent on ω) sequence of the sums of the Stieltjes integrals

$$(w * \mu)_n(\omega) = \sum_{k=1}^{n} \int_{\mathbb{R}^d} w_k(\omega, x)\, \mu_k(dx;\omega).$$

Our integer-valued random measures μ_k are very special in that they take only two values, 0 and 1. Hence

$$\int_{\mathbb{R}^d} w_k(\omega, x)\, \mu_k(dx;\omega) = w_k\bigl(\omega; \Delta X_k(\omega)\bigr).$$

Consequently, we have in fact

$$(w * \mu)_n(\omega) = \sum_{k=1}^{n} w_k\bigl(\omega; \Delta X_k(\omega)\bigr).$$

In a similar way we can define the 'stochastic' integrals $w * \nu$ and $w * (\mu - \nu)$ with respect to the measures ν and $\mu - \nu$ in terms of Stieltjes integrals. Note that we must impose the following condition of integrability on the $w_k(\omega, x)$ for the existence of these 'integrals':

$$\int_{\mathbb{R}^d} |w_k(\omega, x)|\, \nu_k(dx; \omega) < \infty$$

for all (or almost all) $\omega \in \Omega$ and $k \geqslant 1$.

It is easy to see in this case that

$$w * (\mu - \nu) = w * \mu - w * \nu.$$

(We must warn the reader against extending this property automatically to the case of *general* integer-valued random measures, e.g., the jump measures of stochastic processes with *continuous* time: the integral $w * (\mu - \nu)$ may be well defined while $w * \mu$ and $w * \nu$ are equal to $+\infty$, so that their difference has no definite value; see [250; Chapter III] for greater detail.)

Assuming additionally that the functions $w_k(\omega, x)$ are $\mathscr{F}_{k-1}$-measurable for each $x \in \mathbb{R}^d$ we obtain the predictable (i.e., $\mathscr{F}_{n-1}$-measurable) integrals $(w * \nu)_n$.

If, moreover,

$$\mathsf{E} \int_{\mathbb{R}^d} |w_k(\omega, x)|\, \nu_k(dx; \omega) < \infty \tag{4}$$

for each $k \geqslant 1$, then it is easy to see that the sequence $w*(\mu-\nu) = (w*(\mu-\nu)_n)_{n\geqslant 1}$ is a *martingale.*

Replacing (4) by the conditions

$$\mathsf{E} \int_{\mathbb{R}^d} |w_{k\wedge\tau_n}(\omega, x)|\, \nu_{k\wedge\tau_n}(dx; \omega) < \infty, \qquad k \geqslant 1, \quad n \geqslant 1, \tag{4$'$}$$

where (τ_n) is some localizing sequence of Markov times ($\tau_n \leqslant \tau_{n+1}$, $\tau_n \uparrow \infty$), we obtain a sequence $w * (\mu - \nu)$ that is a *local martingale.*

4. We now consider the Doob decomposition of the sequence $H = (H_n)_{n\geqslant 1}$, where we assume that $\mathsf{E}|h_n| < \infty$ for $h_n = \Delta H_n$, $n \geqslant 1$. We have (see Chapter II, § 1b)

$$H_n = A_n + M_n, \tag{5}$$

where

$$A_n = \sum_{k\leqslant n} \mathsf{E}(h_k \mid \mathscr{F}_{k-1}) \tag{6}$$

and

$$M_n = \sum_{k\leqslant n} \bigl[h_k - \mathsf{E}(h_k \mid \mathscr{F}_{k-1})\bigr]. \tag{7}$$

We can represent the variables A_n and M_n as follows in terms of the measures $\mu = (\mu_n)_{n\geqslant 1}$ introduced above and their compensators $\nu = (\nu_n)_{n\geqslant 1}$:

$$A_n = \sum_{k\leqslant n} \int_{\mathbb{R}} x\, \nu_k(dx;\omega), \tag{8}$$

$$M_n = \sum_{k\leqslant n} \int_{\mathbb{R}} x\, \bigl(\mu_k(dx;\omega) - \nu_k(dx;\omega)\bigr). \tag{9}$$

For brevity, one denotes the right-hand sides of (8) and (9) by

$$(x * \nu)_n \tag{10}$$

and

$$(x * (\mu - \nu))_n \tag{11}$$

respectively (see [250; Chapter II]).

Thus,

$$H_n = (x * \nu)_n + \bigl(x * (\mu - \nu)\bigr)_n, \tag{12}$$

or, in the *coordinate-free notation*,

$$H = x * \nu + x * (\mu - \nu). \tag{13}$$

Of course, $H = x * \mu$ in our case, so that (13) is in fact the equality

$$x * \mu = x * \nu + x * (\mu - \nu),$$

as obvious as is the Doob decomposition under the assumption $\mathsf{E}|h_n| < \infty$, $n \geqslant 1$.

In place of the condition that $\mathsf{E}|h_n| < \infty$ for $n \geqslant 1$ we shall now assume that (P-a.s.)

$$\mathsf{E}\bigl(|h_n| \,|\, \mathscr{F}_{n-1}\bigr) < \infty, \qquad n \geqslant 1. \tag{14}$$

Under this assumption, the sequences $A = (A_n)$ and $M = (M_n)$ are certainly well defined (by formulas (6) and (7)), and M is a local martingale since

$$\mathsf{E}(|\Delta M_n| \,|\, \mathscr{F}_{n-1}) < \infty \quad \text{and} \quad \mathsf{E}(\Delta M_n \,|\, \mathscr{F}_{n-1}) = 0.$$

Hence, *if* (14) *holds, then the sequence* $H = (H_n)$ *has a generalized Doob decomposition*

$$H = A + M \tag{15}$$

where $A = (A_n)$ and $M = (M_n)$ are as in (6) and (7).

Moreover, A is a predictable sequence and M is a local martingale. The representation (15) can be rewritten in terms of μ and ν, as equality (13).

Remark. We recall (see Chapter II, § 1b) that $\mathsf{E}(h_n \mid \mathscr{F}_{n-1})$ in (6) and (7) is the generalized conditional expectation defined as the difference

$$\mathsf{E}(h_n^+ \mid \mathscr{F}_{n-1}) - \mathsf{E}(h_n^- \mid \mathscr{F}_{n-1})$$

on the set $\{\omega\colon \mathsf{E}(|h_n| \mid \mathscr{F}_{n-1}) < \infty\}$ and in an arbitrary manner (e.g., as zero) on $\{\omega\colon \mathsf{E}(|h_n| \mid \mathscr{F}_{n-1}) = \infty\}$.

In the general case where (14) can fail, one can obtain an analog of (15) or (13) as follows (we already discussed that in Chapter II, § 1b).

Let $\varphi = \varphi(x)$ be a bounded truncation function, i.e., a function with compact support that is equal to x in a neighborhood of the origin. A typical example is the 'standard truncation function'

$$\varphi(x) = xI(|x| \leqslant 1). \tag{16}$$

Then

$$\begin{aligned} H_n = \sum_{k=1}^{n} h_k &= \sum_{k=1}^{n} \varphi(h_k) + \sum_{k=1}^{n} \big(h_k - \varphi(h_k)\big) \\ &= \sum_{k=1}^{n} \mathsf{E}[\varphi(h_k) \mid \mathscr{F}_{k-1}] \\ &\qquad + \sum_{k=1}^{n} \big[\varphi(h_k) - \mathsf{E}(\varphi(h_k) \mid \mathscr{F}_{k-1})\big] + \sum_{k=1}^{n} \big(h_k - \varphi(h_k)\big) \\ &= \sum_{k=1}^{n} \int \varphi(x)\, \nu_k(dx) + \sum_{k=1}^{n} \int \varphi(x)\big(\mu_k(dx) - \nu_k(dx)\big) \\ &\qquad + \sum_{k=1}^{n} \int \big(x - \varphi(x)\big)\, \mu_k(dx). \end{aligned} \tag{17}$$

Using the notation of (12) and (13), we obtain the following representation:

$$H_n = \big(\varphi(x) * \nu\big)_n + \big(\varphi(x) * (\mu - \nu)\big)_n + \big((x - \varphi(x)) * \mu\big)_n, \tag{18}$$

or, in the coordinate-free notation,

$$H = \varphi * \nu + \varphi * (\mu - \nu) + (x - \varphi) * \mu. \tag{19}$$

DEFINITION. We call (18) and (19) the *canonical representations* of the sequence $H = (H_n)_{n \geqslant 0}$, $H_0 = 0$, with truncation function $\varphi = \varphi(x)$.

With an eye to the continuous-time case it is useful to compare this definition and the *canonical representation of semimartingales* in [250; Chapter II, § 2c] and in Chapter VII, § 3a of the present book.

5. Let $H = (H_n)_{n\geqslant 1}$ be a sequence with generalized Doob decomposition

$$H_n = A_n + M_n$$

and assume that condition (10) in § 3d is satisfied. Then by Theorem 2 in the same § 3d we obtain the following representation with respect to an arbitrary measure $\widetilde{\mathsf{P}} \overset{\text{loc}}{\ll} \mathsf{P}$:

$$\begin{aligned} H_n &= \biggl[A_n + \sum_{k=1}^{n} \mathsf{E}(\alpha_k \Delta M_k \,|\, \mathscr{F}_{k-1})\biggr] + \biggl[M_n - \sum_{k=1}^{n} \mathsf{E}(\alpha_k \Delta M_k \,|\, \mathscr{F}_{k-1})\biggr] \\ &\equiv \widetilde{A}_n + \widetilde{M}_n, \end{aligned} \tag{20}$$

where $\widetilde{M} \in \mathscr{M}_{\text{loc}}(\widetilde{\mathsf{P}})$.

We now write down the canonical representations of H with respect to the measures P and $\widetilde{\mathsf{P}}$:

$$H = \varphi * \nu + \varphi * (\mu - \nu) + (x - \varphi) * \mu \qquad \text{(with respect to } \mathsf{P}) \tag{21}$$

and

$$H = \varphi * \widetilde{\nu} + \varphi * (\mu - \widetilde{\nu}) + (x - \varphi) * \mu \qquad \text{(with respect to } \widetilde{\mathsf{P}}), \tag{22}$$

where μ is the jump measure of the sequence H.

It is important for the stochastic calculus based on the canonical representations (21) and (22) that one knows how to *calculate* the compensators $\widetilde{\nu}$ for given compensators ν and the characteristics of the density process $Z = (Z_n)$. In particular, we are interested in formulas describing the *transformation of the 'drift' terms* $\varphi * \nu$ *and* $\varphi * \widetilde{\nu}$ *under a change of measure.*

We shall discuss this issue more closely under the assumption that $\widetilde{\mathsf{P}} \overset{\text{loc}}{\ll} \mathsf{P}$. Let

$$\nu_n(\,\cdot\,;\omega) = \mathsf{P}(h_n \in \cdot \,|\, \mathscr{F}_{n-1})(\omega)$$

and

$$\widetilde{\nu}_n(\,\cdot\,;\omega) = \widetilde{\mathsf{P}}(h_n \in \cdot \,|\, \mathscr{F}_{n-1})(\omega)$$

be the regular modifications of the corresponding conditional probabilities.

Bayes's formula (4) in § 3a assumes the following form for $Y = I_A(h_n)$, $A \in \mathscr{B}(\mathbb{R} \setminus \{0\})$, and $m = n-1$:

$$\widetilde{\mathsf{E}}(I_A(h_n) \,|\, \mathscr{F}_{n-1}) = \mathsf{E}\biggl(I_A(h_n) \frac{Z_n}{Z_{n-1}} \,\bigg|\, \mathscr{F}_{n-1}\biggr). \tag{23}$$

Hence the following conjecture looks plausible: for each $\omega \in \Omega$ the *conditional distributions* $\widetilde{\nu}_n(\,\cdot\,;\omega)$ *are absolutely continuous with respect to* $\nu_n(\,\cdot\,;\omega)$, i.e., there exists a $\mathscr{B}(\mathbb{R}\setminus\{0\})$-measurable (for each $\omega \in \Omega$) function $Y_n = Y_n(x,\omega)$ such that

$$\widetilde{\nu}_n(A;\omega) = \int_A Y_n(x,\omega)\,\nu_n(dx;\omega). \tag{24}$$

If it were so, then

$$\frac{d\widetilde{\nu}_n(\,\cdot\,;\omega)}{d\nu_n(\,\cdot\,;\omega)}(x) = Y_n(x,\omega), \tag{25}$$

i.e., the function $Y_n(x,\omega)$ would play the role of the *density* of one measure (more precisely, of one conditional distribution) with respect to the other.

We present now the proof of formula (24) (under the assumption $\widetilde{\mathsf{P}} \overset{\text{loc}}{\ll} \mathsf{P}$), which simultaneously delivers an 'explicit formula' for the density $Y_n = Y_n(x,\omega)$, $n \geqslant 1$.

We consider the conditional expectation on the right-hand side of (23). By definition, for each $B \in \mathscr{F}_{n-1}$ we have

$$\begin{aligned}
&\int_B \mathsf{E}\Big(I_A(h_n)\frac{Z_n}{Z_{n-1}}\;\Big|\;\mathscr{F}_{n-1}\Big)(\omega)\,(\mathsf{P}\,|\,\mathscr{F}_{n-1})(d\omega)\\
&\qquad= \int_B I_A(h_n)\frac{Z_n(\omega)}{Z_{n-1}(\omega)}(\mathsf{P}\,|\,\mathscr{F}_{n-1})(d\omega)\\
&\qquad= \int_B\Big[\int_A \frac{Z_n(\omega)}{Z_{n-1}(\omega)}\,\mu_n(dx;\omega)\Big](\mathsf{P}\,|\,\mathscr{F}_{n-1})(d\omega)\\
&\qquad= \int_{B\times A}\frac{Z_n(\omega)}{Z_{n-1}(\omega)}\,\mu_n(dx;\omega)(\mathsf{P}\,|\,\mathscr{F}_{n-1})(d\omega).
\end{aligned} \tag{26}$$

Let $M_n(dx,d\omega)$ be the 'skew product of measures'

$$\mu_n(dx;\omega)(\mathsf{P}\,|\,\mathscr{F}_{n-1})(d\omega)$$

on $\mathscr{B}(\mathbb{R}\setminus\{0\})\otimes\mathscr{F}_{n-1}$ and let $\mathsf{E}_{M_n}(\,\cdot\,|\,\mathscr{B}(\mathbb{R}\setminus\{0\})\otimes\mathscr{F}_{n-1})$ be the conditional expectation (with respect to the algebra $\mathscr{B}(\mathbb{R}\setminus\{0\})\otimes\mathscr{F}_{n-1}$ and measure $M_n = M_n(dx,d\omega)$) defined in a standard way (see, e.g., [439; Chapter II, § 7]), on the basis of the Radon–Nikodym theorem.

Then we derive from (26) by Fubini's theorem that

$$\begin{aligned}
&\int_B \mathsf{E}\Big(I_A(h_n)\frac{Z_n}{Z_{n-1}}\;\Big|\;\mathscr{F}_{n-1}\Big)(\omega)\,(\mathsf{P}\,|\,\mathscr{F}_{n-1})(d\omega)\\
&= \int_{B\times A}\frac{Z_n(\omega)}{Z_{n-1}(\omega)}\,M_n(dx;d\omega)\\
&= \int_{B\times A}\mathsf{E}_{M_n}\Big(\frac{Z_n}{Z_{n-1}}\;\Big|\;\mathscr{B}(\mathbb{R}\setminus\{0\})\otimes\mathscr{F}_{n-1}\Big)(x,\omega)\,M_n(dx;d\omega)\\
&= \int_B\Big[\int_A \mathsf{E}_{M_n}\Big(\frac{Z_n}{Z_{n-1}}\;\Big|\;\mathscr{B}(\mathbb{R}\setminus\{0\})\otimes\mathscr{F}_{n-1}\Big)(x,\omega)\,\mu_n(dx;\omega)\Big](\mathsf{P}\,|\,\mathscr{F}_{n-1})(d\omega)\\
&= \int_B\Big[\int_A \mathsf{E}_{M_n}\Big(\frac{Z_n}{Z_{n-1}}\;\Big|\;\mathscr{B}(\mathbb{R}\setminus\{0\})\otimes\mathscr{F}_{n-1}\Big)(x,\omega)\,\nu_n(dx;\omega)\Big](\mathsf{P}\,|\,\mathscr{F}_{n-1})(d\omega).
\end{aligned}$$

Since B is an arbitrary subset in $\mathscr{F}_{n-1}$, it follows that (P-a.s.)

$$\mathsf{E}\left(I_A(h_n)\frac{Z_n}{Z_{n-1}} \;\Big|\; \mathscr{F}_{n-1}\right)(\omega) = \int_A Y_n(x,\omega)\,\nu_n(dx;\omega), \tag{27}$$

where

$$Y_n(x,\omega) = \mathsf{E}_{M_n}\left(\frac{Z_n}{Z_{n-1}} \;\Big|\; \mathscr{B}(\mathbb{R}\setminus\{0\})\otimes\mathscr{F}_{n-1}\right)(x,\omega). \tag{28}$$

Comparing (23) with $\widetilde{\mathsf{E}}(I_A(h_n)\,|\,\mathscr{F}_{n-1}) = \int_A \widetilde{\nu}_n(dx;\omega)$ and formulas (27) and (28) we see that $\widetilde{\nu}_n(\,\cdot\,;\omega) \ll \nu_n(\,\cdot\,;\omega)$ for each $\omega\in\Omega$ and (25) holds.

This formula provides an answer to the above question: the '*drift*' *terms* $\varphi*\widetilde{\nu}$ *and* $\varphi*\nu$ *in* (22) *and* (21) *are connected by the relation*

$$\varphi*\widetilde{\nu} = \varphi*\nu + \varphi(Y-1)*\nu \tag{29}$$

(at any rate, if $(|\varphi(x)(Y-1)|*\nu)_n < \infty$, $n\geqslant 1$).

§ 3f. 'Predictable' Criteria of Arbitrage-Free (B,S)-Markets

1. By the *First fundamental asset pricing theorem* (§ 2b) a (B,S)-market formed by a bank account $B=(B_n)$ and d assets $S=(S^1,\dots,S^d)$, $S^i=(S^i_n)$, $0\leqslant n\leqslant N$, is arbitrage-free if and only if there exists a probability (martingale) measure $\widetilde{\mathsf{P}}$ equivalent to P on the initial filtered probability space $(\Omega,\mathscr{F},(\mathscr{F}_n)_{n\geqslant N},\mathsf{P})$ such that the d-dimensional sequence of discounted prices

$$\frac{S}{B} = \left(\frac{S_n}{B_n}\right)_{0\leqslant n\leqslant N}$$

is a $\widetilde{\mathsf{P}}$-martingale.

The description of the class $\mathscr{P}(\mathsf{P})$ of *all* such martingale measures $\widetilde{\mathsf{P}}\sim\mathsf{P}$ is also very interesting because we have already seen in § 1c that in looking for upper and lower prices one must consider the greatest and the least values over the class $\mathscr{P}(\mathsf{P})$.

Searching such measures it is reasonable to start with a slightly more general problem of the construction of martingale measures $\widetilde{\mathsf{P}}$, that are *locally* absolutely continuous with respect to P, leaving aside the question whether a measure so obtained satisfies incidentally the relation $\mathsf{P}\sim\widetilde{\mathsf{P}}$.

2. In the previous sections we exposed all the material pertaining to 'absolutely continuous changes of measure' that is necessary for a discussion of the problem of the construction of measures $\widetilde{\mathsf{P}} \overset{\text{loc}}{\ll} \mathsf{P}$.

We start with the case where $d=1$, $B_n\equiv 1$, $S=(S_n)$ is the only 'risk' asset, and

$$S_n = S_0 e^{H_n}. \tag{1}$$

We assume that we have a filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_n), \mathsf{P})$ and that the variables H_n are $\mathscr{F}_n$-measurable. As before, we shall set $h_n = \Delta H_n$ and $H_0 = 0$. Throughout, we shall assume that $\mathscr{F}_0 = \{\varnothing, \Omega\}$.

Using the notation and the results of § 3c, for $n \geqslant 1$ we set

$$\widehat{H}_n = \sum_{k=1}^{n} (e^{\Delta H_k} - 1) \tag{2}$$

and

$$\mathscr{E}(\widehat{H})_n = \prod_{k=1}^{n} (1 + \Delta \widehat{H}_k); \tag{3}$$

we also set $\widehat{H}_0 = 0$.

Note that

$$\Delta \mathscr{E}(\widehat{H})_n = \mathscr{E}(\widehat{H})_{n-1} \Delta \widehat{H}_n. \tag{4}$$

In our construction of a martingale measure $\widetilde{\mathsf{P}} \overset{\text{loc}}{\ll} \mathsf{P}$ such that $S = (S_n)$ is a $\widetilde{\mathsf{P}}$-martingale we can follow directly the above-described pattern: write down the canonical representation for $S = (S_n)$ (of type (13) in § 3e) and then find a measure $\widetilde{\mathsf{P}} \overset{\text{loc}}{\ll} \mathsf{P}$ 'killing' the drift term. We could proceed in that way; however, we have an additional property: prices are *positive*. This property enables us to consider the logarithms of the prices (i.e., the sequence $H = (H_n)$), which, as statistical analysis shows, have a more simple structure that the sequence $S = (S_n)$ of prices themselves.

Let $\widetilde{\mathsf{P}}$ be a measure such that $\widetilde{\mathsf{P}} \overset{\text{loc}}{\ll} \mathsf{P}$.

We set $Z_n = \dfrac{d\widetilde{\mathsf{P}}_n}{d\mathsf{P}_n}$, where $\mathsf{P}_n = \mathsf{P} \,|\, \mathscr{F}_n$ and $\widetilde{\mathsf{P}}_n = \widetilde{\mathsf{P}} \,|\, \mathscr{F}_n$. In what follows we set $\widetilde{\mathsf{P}}_0 = \mathsf{P}_0$, so that $Z_0 = 1$. Let $X = (X_n)_{n \geqslant 0}$ be some sequence of $\mathscr{F}_n$-measurable variables X_n, and let $X_0 = 0$.

By the lemma in § 3d,

if $\widetilde{\mathsf{P}} \overset{\text{loc}}{\ll} \mathsf{P}$, *then*

$$X \in \mathscr{M}(\widetilde{\mathsf{P}}) \iff XZ \in \mathscr{M}(\mathsf{P}), \tag{5}$$

and

if $\widetilde{\mathsf{P}} \overset{\text{loc}}{\sim} \mathsf{P}$, *then*

$$X \in \mathscr{M}_{\text{loc}}(\widetilde{\mathsf{P}}) \iff XZ \in \mathscr{M}_{\text{loc}}(\mathsf{P}). \tag{6}$$

Now let $X \in \mathscr{M}_{\text{loc}}(\widetilde{\mathsf{P}})$. By the theorem in Chapter II, § 1c the sequence X is a martingale transform, and therefore $\Delta X_n = a_n \Delta M_n$, where the a_n are $\mathscr{F}_{n-1}$-measurable and M is a martingale (with respect to $\widetilde{\mathsf{P}}$). Hence

$$\Delta \mathscr{E}(X)_n = \mathscr{E}(X)_{n-1} \Delta X_n = a_n \mathscr{E}(X)_{n-1} \Delta M_n,$$

and therefore $\mathscr{E}(X)$ is also a martingale transform, so that (again, by the theorem in Chapter II, § 1c) $\mathscr{E}(X) \in \mathscr{M}_{\text{loc}}(\widetilde{\mathsf{P}})$. Thus,

$$X \in \mathscr{M}_{\text{loc}}(\widetilde{\mathsf{P}}) \implies \mathscr{E}(X) \in \mathscr{M}_{\text{loc}}(\widetilde{\mathsf{P}}). \tag{7}$$

Assuming that $\mathscr{E}(X) \neq 0$ and considering

$$\Delta X_n = \Delta \mathscr{E}(X)_n / \mathscr{E}(X)_{n-1},$$

we can see in a similar way that

$$\mathscr{E}(X) \in \mathscr{M}_{\text{loc}}(\widetilde{\mathsf{P}}) \implies X \in \mathscr{M}_{\text{loc}}(\widetilde{\mathsf{P}}).$$

Hence, in view of (6), we obtain the following result.

LEMMA 1. *Let $\widetilde{\mathsf{P}} \overset{\text{loc}}{\sim} \mathsf{P}$ and assume that $\mathscr{E}(X) \neq 0$ ($\widetilde{\mathsf{P}}$-a.s.). Then*

$$\mathscr{E}(X) \in \mathscr{M}_{\text{loc}}(\widetilde{\mathsf{P}}) \iff X \in \mathscr{M}_{\text{loc}}(\widetilde{\mathsf{P}}) \iff XZ \in \mathscr{M}_{\text{loc}}(\mathsf{P}).$$

Assume now that $\mathscr{E}(X) \geqslant 0$ and $\mathscr{E}(X) \in \mathscr{M}(\widetilde{\mathsf{P}})$. Then, by the lemma in Chapter II, § 1c we obtain that $\mathscr{E}(X) \in \mathscr{M}(\widetilde{\mathsf{P}})$. Consequently, in view of (5), we have the following result.

LEMMA 2. *Let $\widetilde{\mathsf{P}} \overset{\text{loc}}{\ll} \mathsf{P}$ and assume that $\mathscr{E}(X) \geqslant 0$ ($\widetilde{\mathsf{P}}$-a.s.). Then*

$$\mathscr{E}(X) \in \mathscr{M}_{\text{loc}}(\widetilde{\mathsf{P}}) \iff \mathscr{E}(X) \in \mathscr{M}(\widetilde{\mathsf{P}}) \iff \mathscr{E}(X)Z \in \mathscr{M}(\mathsf{P}).$$

We point out the following implications (which must be interpreted componentwise) holding by (3):

$$\Delta X \neq -1 \implies \mathscr{E}(X) \neq 0,$$
$$\Delta X \geqslant -1 \implies \mathscr{E}(X) \geqslant 0,$$

and

$$\Delta X > -1 \iff \mathscr{E}(X) > 0.$$

Applying Lemmas 1 and 2 to the case of $X = \widehat{H}$, where $\widehat{H}$ is related to H by formula (2), we obtain the following result.

LEMMA 3. *Let* $S = (S_n)_{n \geqslant 0}$, *where*

$$S_n = S_0 e^{H_n}, \qquad H_0 = 0, \tag{8}$$

and let $\Delta \widehat{H}_n = e^{\Delta H_n} - 1$, $\widehat{H}_0 = 0$.

Then

$$S_n = S_0 \mathscr{E}(\widehat{H})_n \tag{9}$$

and

if $\widetilde{\mathsf{P}} \overset{\text{loc}}{\ll} \mathsf{P}$, *then*

$$S \in \mathscr{M}(\widetilde{\mathsf{P}}) \iff \mathscr{E}(\widehat{H})Z \in \mathscr{M}(\mathsf{P}); \tag{10}$$

if $\widetilde{\mathsf{P}} \overset{\text{loc}}{\sim} \mathsf{P}$, *then*

$$S \in \mathscr{M}_{\text{loc}}(\widetilde{\mathsf{P}}) \iff \widehat{H}Z \in \mathscr{M}_{\text{loc}}(\mathsf{P}). \tag{11}$$

3. These implications indicate the way in which one can seek measures $\widetilde{\mathsf{P}}$ such that $S \in \mathscr{M}(\widetilde{\mathsf{P}})$.

The sequence $Z = (Z_n)$ is a P-martingale. In accordance with (10) and (11), we must describe the non-negative P-martingales Z such that $\mathsf{E}Z_n \equiv 1$ and, in addition, either $\mathscr{E}(\widehat{H})Z \in \mathscr{M}(\mathsf{P})$ if we are looking for a measure $\widetilde{\mathsf{P}} \overset{\text{loc}}{\ll} \mathsf{P}$ or $\widehat{H}Z \in \mathscr{M}_{\text{loc}}(\mathsf{P})$ if we require that $\widetilde{\mathsf{P}} \overset{\text{loc}}{\sim} \mathsf{P}$.

The corresponding class of martingales $Z = (Z_n)$ in the conditionally Gaussian case is formed by the martingales

$$Z_n = \exp\left\{\sum_{k=1}^{n} b_k \varepsilon_k - \frac{1}{2}\sum_{k=1}^{n} b_k^2\right\} \tag{12}$$

(with $\mathscr{F}_{n-1}$-measurable b_k: see, for instance, formula (7) in § 3b) satisfying the difference equations

$$\Delta Z_n = Z_{n-1}\Delta N_n, \tag{13}$$

where the variables

$$\Delta N_n = e^{b_n \varepsilon_n - \frac{1}{2} b_n^2} - 1, \tag{14}$$

make up a generalized martingale difference, i.e.,

$$\mathsf{E}(|\Delta N_n| \mid \mathscr{F}_{n-1}) < \infty \quad \text{and} \quad \mathsf{E}(\Delta N_n \mid \mathscr{F}_{n-1}) = 0.$$

Hence a natural way to look for the density processes $Z = (Z_n)$ can be as follows.

We shall seek the required densities Z_n (necessary for the construction of the measures $\widetilde{\mathsf{P}}_n$ by $\widetilde{\mathsf{P}}_n(d\omega) = Z_n(\omega)\,\mathsf{P}_n(d\omega)$) in the form (13), i.e., we assume that

$$Z_n = \mathscr{E}(N)_n, \qquad Z_0 = 1, \tag{15}$$

where N are some (specified in what follows) *local* martingales with $N_0 = 0$, $\Delta N_n \geqslant -1$, and $\mathsf{E}\mathscr{E}(N)_n = 1$.

The question on the volume of the class of so obtained measures $\widetilde{\mathsf{P}} \overset{\text{loc}}{\ll} \mathsf{P}$ is far from simple. The point is that, first, it is not an easy problem to define whether (for a family of consistent 'finite-dimensional' distributions $\{\widetilde{\mathsf{P}}_n\}$) there exists a measure $\widetilde{\mathsf{P}}$ such that $\widetilde{\mathsf{P}} \mid \mathscr{F}_n = \widetilde{\mathsf{P}}_n$, $n \geqslant 1$. (See a counterexample in [439; Chapter II, § 3].)

Second, the question of the structure of all the martingales (or local martingales) on $(\Omega, \mathscr{F}, (\mathscr{F}_n)_{n\geqslant 1}, \mathsf{P})$ is not simple in principle. (See [250; Chapter III] on this issue.)

In what follows, we take a way that, while giving us no exhaustive answer to the question of the structure of all measures $\widetilde{\mathsf{P}}$ such that $\widetilde{\mathsf{P}} \overset{\text{loc}}{\ll} \mathsf{P}$ or $\widetilde{\mathsf{P}} \overset{\text{loc}}{\sim} \mathsf{P}$, is nevertheless fairly simple technically and brings one to a *broad class* of such measures. First of all, we make several observations.

It is clear from the above that, given a measure P, all measures $\widetilde{\mathsf{P}}$ such that $\widetilde{\mathsf{P}} \overset{\text{loc}}{\sim} \mathsf{P}$ are completely (as regards their finite-dimensional distributions $\{\widetilde{\mathsf{P}}_n\}$) described by their densities $Z = (Z_n)$. It follows from the assumption $\widetilde{\mathsf{P}} \overset{\text{loc}}{\sim} \mathsf{P}$ that $\mathsf{P}(Z_n > 0) = 1$, so that we can construct from $Z = (Z_n)$ a new sequence $N = (N_n)$ with $N_0 = 0$ and

$$\Delta N_n = \frac{\Delta Z_n}{Z_{n-1}}\,. \tag{16}$$

Clearly, $N \in \mathscr{M}_{\text{loc}}(\mathsf{P})$, and the sequences Z and N are in a one-to-one correspondence thanks to the equality $Z_n = \mathscr{E}(N)_n$.

Hence we can turn in our construction of measures $\widetilde{\mathsf{P}} \overset{\text{loc}}{\sim} \mathsf{P}$ from the densities $Z = (Z_n)$, where $Z_n = \dfrac{d\widetilde{\mathsf{P}}_n}{d\mathsf{P}_n}$, to the appropriate sequence $N = (N_n)$, which must satisfy the inequalities $\Delta N_n > -1$ to ensure that $Z_n > 0$ (P-a.s.) for $n \geqslant 1$.

Thus, we shall assume that $S_0 > 0$, $S_n = S_0\mathscr{E}(\widehat{H})_n$ for $n \geqslant 1$, and, in addition, $S_n > 0$ (which is equivalent to the relation $\Delta\widehat{H}_n > -1$).

Also let $Z = (Z_n)$, where $Z_n = \mathscr{E}(N)_n$ with $\Delta N_n > -1$, so that $Z_n > 0$ (P-a.s.).

Assume that there exists a measure $\widetilde{\mathsf{P}}$ such that its restrictions $\widetilde{\mathsf{P}}_n = \widetilde{\mathsf{P}} \mid \mathscr{F}_n$ satisfy the relations $\widetilde{\mathsf{P}}_n \sim \mathsf{P}_n$, $n \geqslant 1$, i.e., $\widetilde{\mathsf{P}} \overset{\text{loc}}{\sim} \mathsf{P}$.

Then, in view of (10),

$$S \in \mathscr{M}(\widetilde{\mathsf{P}}) \iff \mathscr{E}(\widehat{H})\mathscr{E}(N) \in \mathscr{M}(\mathsf{P}). \tag{17}$$

4. The following *Yor's formula* (M. Yor; see, e.g., [402]) can be immediately verified:

$$\mathscr{E}(\widehat{H})\mathscr{E}(N) = \mathscr{E}(\widehat{H} + N + [\widehat{H}, N]), \tag{18}$$

where

$$[\widehat{H}, N]_n = \sum_{k=1}^{n} \Delta \widehat{H}_k \Delta N_k.$$

By assumptions $\mathscr{E}(\widehat{H}) > 0$, $\mathscr{E}(N) > 0$, and equivalence (17) we obtain

$$\begin{aligned} S \in \mathscr{M}(\widetilde{\mathsf{P}}) &\iff \mathscr{E}(\widehat{H} + N + [\widehat{H}, N]) \in \mathscr{M}(\mathsf{P}) \\ &\iff \widehat{H} + N + [\widehat{H}, N] \in \mathscr{M}_{\mathrm{loc}}(\mathsf{P}). \end{aligned}$$

Hence if $N \in \mathscr{M}_{\mathrm{loc}}(\mathsf{P})$, $\Delta N > -1$, $Z_n = \mathscr{E}(N)_n$, and $d\widetilde{\mathsf{P}}_n = Z_n \, d\mathsf{P}$, then

$$S \in \mathscr{M}(\widetilde{\mathsf{P}}) \iff \widehat{H} + [\widehat{H}, N] \in \mathscr{M}_{\mathrm{loc}}(\mathsf{P}).$$

Since $\Delta(\widehat{H} + [\widehat{H}, N]) = \Delta\widehat{H}(1 + \Delta N)$, the inclusion $\widehat{H} + [\widehat{H}, N] \in \mathscr{M}_{\mathrm{loc}}(\mathsf{P})$ is equivalent to the condition that the sequence $\Delta\widehat{H}(1 + \Delta N) = \big(\Delta\widehat{H}_n(1 + \Delta N_n)\big)_n$ is a local P-martingale difference, or, the same (see the lemma in Chapter II, § 1c), that it is a generalized P-martingale difference and satisfies (P-a.s.) the relations

$$\mathsf{E}\big[\, |\Delta\widehat{H}_n(1 + \Delta N_n)| \,|\, \mathscr{F}_{n-1} \big] < \infty \tag{19}$$

and

$$\mathsf{E}\big[\, \Delta\widehat{H}_n(1 + \Delta N_n) \,|\, \mathscr{F}_{n-1} \big] = 0 \tag{20}$$

for all $n \geqslant 1$.

We note that conditions (19) and (20) are formulated in terms of the conditional expectation $\mathsf{E}(\,\cdot \mid \mathscr{F}_{n-1})$ (i.e., in 'predictable' terms).

Conditions (19) and (20) can be expressed in various forms. For instance, bearing in mind that $\Delta Z_n = Z_{n-1}\Delta N_n$, $\Delta N_n > -1$, and $\Delta\widehat{H}_n = e^{\Delta H_n} - 1$, we see that (20) is equivalent to the following condition on Z:

$$\mathsf{E}\left[e^{\Delta H_n} \frac{Z_n}{Z_{n-1}} \;\Big|\; \mathscr{F}_{n-1} \right] = 1. \tag{21}$$

Of course, we could derive this condition directly because if $\widetilde{\mathsf{P}} \overset{\mathrm{loc}}{\ll} \mathsf{P}$, then

$$\widetilde{\mathsf{E}}[e^{\Delta H_n} \mid \mathscr{F}_{n-1}] = \mathsf{E}\left[e^{\Delta H_n} \frac{Z_n}{Z_{n-1}} \;\Big|\; \mathscr{F}_{n-1} \right] \qquad (\mathsf{P}\text{-a.s.}) \tag{22}$$

by Bayes's formula ((4) in § 3a or (1) in § 3d), and therefore (20) is equivalent to the relation $\widetilde{\mathsf{E}}(S_n \mid \mathscr{F}_{n-1}) = S_{n-1}$ (P and $\widetilde{\mathsf{P}}$-a.s.). In a similar way, (19) $\iff$

$\widetilde{\mathsf{E}}(S_n \mid \mathscr{F}_{n-1}) < \infty$. Since the S_n are non-negative, *conditions* (19) *and* (20) *ensure that the sequence $S = (S_n)$ is a martingale with respect to the measure $\widetilde{\mathsf{P}}$ constructed from the sequence $N = (N_n)$.*

The above method can easily be extended to the multidimensional case, when $S = (S^1, \dots, S^d)$ and we consider the question of whether

$$\frac{S}{B} \in \mathscr{M}(\widetilde{\mathsf{P}})$$

with respect to some measure $\widetilde{\mathsf{P}} \overset{\text{loc}}{\sim} \mathsf{P}$.

This is the question we discuss next.

5. Our choice of the bank account $B = (B_n)$ as the normalizing factor has the advantage that the variables B_n are $\mathscr{F}_{n-1}$-measurable, which, as already mentioned, brings certain technical simplifications. However, there are no serious obstacles to considering *any other asset* in this role.

In this connection we discuss now the following situation.

Let $S^0 = (S^0_n)$ and $S^1 = (S^1_n)$ be two assets. We assume that

$$S^i_n = S^i_0 e^{H^i_n}, \qquad i = 0, 1,$$

where $H^i_0 = 0$, and let $\Delta Z_n = Z_{n-1}\Delta N_n$ and $\Delta \widehat{H}^i_n = e^{\Delta H^i_n} - 1$.

Setting S^1_0 and S^0_0 to be constants we consider the ratio

$$\frac{S^1}{S^0} = \left(\frac{S^1_n}{S^0_n}\right)_{n\geqslant 0},$$

and ask whether $\dfrac{S^1}{S^0} \in \mathscr{M}(\widetilde{\mathsf{P}})$. (To avoid the question of the existence of a measure $\widetilde{\mathsf{P}}$ such that $\widetilde{\mathsf{P}} \mid \mathscr{F}_n = \widetilde{\mathsf{P}}_n$, where $\widetilde{\mathsf{P}}_n(d\omega) = Z_n(\omega)\mathsf{P}_n(d\omega)$, we can assume that $n \leqslant N$ and $\mathscr{F} = \mathscr{F}_N$.)

Clearly

$$\frac{S^1}{S^0} Z = \frac{S^1_0}{S^0_0} Z_0 \cdot \mathscr{E}(\widehat{H}^1) \cdot \mathscr{E}^{-1}(\widehat{H}^0) \cdot \mathscr{E}(N) \tag{23}$$

(we use the coordinate-free notation). It is easy to verify directly that

$$\mathscr{E}^{-1}(\widehat{H}^0) = \mathscr{E}(-\widehat{H}^*), \tag{24}$$

where

$$\widehat{H}^*_n = \widehat{H}^0_n - \sum_{k=1}^{n} \frac{(\Delta \widehat{H}^0_k)^2}{1 + \Delta \widehat{H}^0_k} \quad \left(= \sum_{k=1}^{n} \frac{\Delta \widehat{H}^0_k}{1 + \Delta \widehat{H}^0_k} \right). \tag{25}$$

Hence $\dfrac{S^1}{S^0}Z$ is a product of *three* stochastic exponentials:

$$\frac{S^1}{S^0}Z = \frac{S_0^1}{S_0^0}Z_0 \cdot \mathscr{E}(\widehat{H}^1) \cdot \mathscr{E}(-\widehat{H}^*) \cdot \mathscr{E}(N), \tag{26}$$

so that using *Yor's formula* (18) we successively obtain

$$\begin{aligned} &\mathscr{E}(\widehat{H}^1) \cdot \mathscr{E}(-\widehat{H}^*) \cdot \mathscr{E}(N) \\ &\qquad = \mathscr{E}\biggl(\widehat{H}^1 - \widehat{H}^0 + N + \sum_{k\leqslant \cdot} \frac{(\Delta\widehat{H}_k^0 - \Delta\widehat{H}_k^1)(\Delta\widehat{H}_k^0 - \Delta N_k)}{1+\Delta\widehat{H}_k^0}\biggr). \end{aligned} \tag{27}$$

Hence one necessary and sufficient condition on N ensuring that $\dfrac{S^1}{S^0} \in \mathscr{M}(\widetilde{\mathsf{P}})$, is the following inclusion:

$$\widehat{H}^1 - \widehat{H}^0 + \sum_{k\leqslant \cdot} \frac{(\Delta\widehat{H}_k^0 - \Delta\widehat{H}_k^1)(\Delta\widehat{H}_k^0 - \Delta N_k)}{1+\Delta\widehat{H}_k^0} \in \mathscr{M}_{\mathrm{loc}}(\mathsf{P}), \tag{28}$$

or, equivalently,

$$\biggl(\frac{(\Delta\widehat{H}_k^1 - \Delta\widehat{H}_k^0)(1+\Delta N_k)}{1+\Delta\widehat{H}_k^0}\biggr)_{k\geqslant 1} \in \mathscr{M}_{\mathrm{loc}}(\mathsf{P}). \tag{29}$$

Since for d assets $S^1,\dots,S^d$ the question on the martingale property of the vector of discounted prices

$$\frac{S}{S^0} = \biggl(\frac{S^1}{S^0},\dots,\frac{S^d}{S^0}\biggr)$$

can be answered by component-wise analysis, we obtain from (29) the following general result.

THEOREM 1. *Let $(S^0, S^1,\dots,S^d)$ be $d+1$ assets defined on the filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_n)_{n\leqslant N}, \mathsf{P})$ with $\mathscr{F} = \mathscr{F}_N$ such that*

$$S_n^i = S_0^i e^{H_n^i}, \quad \textit{where} \quad H_0^i = 0,\ \ i = 1,\dots,d, \quad 1 \leqslant n \leqslant N,$$

or, equivalently,

$$\Delta S_n^i = S_{n-1}^i \Delta\widehat{H}_n^i, \tag{30}$$

where

$$\Delta\widehat{H}_n^i = e^{\Delta H_n^i} - 1, \qquad \widehat{H}_0^i = 0. \tag{31}$$

Assume also that the constants S_0^i are positive for all $i = 0, 1, \dots, n$.

Further, let $Z = (Z_n)_{0 \leqslant n \leqslant N}$ be a sequence of random variables such that

$$\Delta Z_n = Z_{n-1} \Delta N_n, \qquad Z_0 = 1, \tag{32}$$

where $\Delta N_n > -1$.

Then the ratio $\dfrac{S}{S^0}$ is a d-dimensional martingale with respect to a measure $\widetilde{\mathsf{P}}_N$ such that

$$\widetilde{\mathsf{P}}_N(d\omega) = Z_N(\omega)\,\mathsf{P}(d\omega) \tag{33}$$

if and only if for all i and n, $i = 1, \dots, d$ and $1 \leqslant n \leqslant N$, we have ($\mathsf{P}$-a.s.)

$$\mathsf{E}\left[\left|\frac{(\Delta \widehat{H}^i_n - \Delta \widehat{H}^0_n)(1 + \Delta N_n)}{1 + \Delta \widehat{H}^0_n}\right| \,\Big|\, \mathscr{F}_{n-1}\right] < \infty \tag{34}$$

and

$$\mathsf{E}\left[\frac{(\Delta \widehat{H}^i_n - \Delta \widehat{H}^0_n)(1 + \Delta N_n)}{1 + \Delta \widehat{H}^0_n} \,\Big|\, \mathscr{F}_{n-1}\right] = 0. \tag{35}$$

COROLLARY. *Let $S^0 = (S^0_n)$ be a 'risk-free' asset in the sense of the $\mathscr{F}_{n-1}$-measurability of the S^0_n (for example, $S^0 = B$ can be a bank account with fixed interest rate ($\Delta \widehat{H}^0_n \equiv r$)). Then (in view of the $\mathscr{F}_{n-1}$-measurability of the $\widehat{H}^0_n$) conditions (34) and (35) can be written in the following form:*

$$\mathsf{E}\big[|\Delta \widehat{H}^i_n (1 + \Delta N_n)| \,|\, \mathscr{F}_{n-1}\big] < \infty \tag{36}$$

and

$$\mathsf{E}\big[\,\Delta \widehat{H}^i_n (1 + \Delta N_n) \,|\, \mathscr{F}_{n-1}\big] = \Delta \widehat{H}^0_n \tag{37}$$

for $i = 1, \dots, d$ and n between 1 and N.

In particular, if $\Delta \widehat{H}^0_n \equiv 0$, then these conditions have the following form:

$$\mathsf{E}\big[|\Delta \widehat{H}^i_n (1 + \Delta N_n)| \,|\, \mathscr{F}_{n-1}\big] < \infty, \tag{38}$$

$$\mathsf{E}\big[\,\Delta \widehat{H}^i_n (1 + \Delta N_n) \,|\, \mathscr{F}_{n-1}\big] = 0, \tag{39}$$

i.e., are the same as earlier obtained conditions (19) and (20).

If, in addition, $\Delta N_n \equiv 0$, then the conditions in question reduce to

$$\mathsf{E}\big[|\Delta \widehat{H}^i_n| \,|\, \mathscr{F}_{n-1}\big] < \infty \tag{40}$$

and

$$\mathsf{E}\big[\,\Delta \widehat{H}^i_n \,|\, \mathscr{F}_{n-1}\big] = 0, \tag{41}$$

moreover, (41) is equivalent to the condition

$$\mathsf{E}[e^{\Delta H^i_n} \,|\, \mathscr{F}_{n-1}] = 1$$

(cf. (11) in § 3c), which is an obvious condition of the inclusion $S^i \in \mathscr{M}(\mathsf{P})$.

6. We consider now several examples illustrating the above criteria. To this end we observe first of all the following.

Assume that a (B, S)-market is formed by two assets: a bank account $B = (B_n)$ such that

$$\Delta B_n = r_n B_{n-1}$$

with $\mathscr{F}_{n-1}$-measurable r_n and stock $S = (S_n)$ such that

$$\Delta S_n = \rho_n S_{n-1}$$

with $\mathscr{F}_n$-measurable ρ_n. Then the conditions (36) and (37) can be rewritten as follows:

$$\mathsf{E}\big[|\rho_n(1 + \Delta N_n)| \,|\, \mathscr{F}_{n-1}\big] < \infty, \tag{42}$$

$$\mathsf{E}\big[\ \rho_n(1 + \Delta N_n)\ \,|\, \mathscr{F}_{n-1}\big] = r_n. \tag{43}$$

On the other hand, setting

$$B_n = B_{n-1}e^{r_n} \quad \text{and} \quad S_n = S_{n-1}e^{\rho_n}, \tag{44}$$

we can rewrite (36) and (37) as

$$\mathsf{E}\big[|(e^{\rho_n} - 1)(1 + \Delta N_n)| \,|\, \mathscr{F}_{n-1}\big] < \infty \tag{45}$$

and

$$\mathsf{E}\big[\ (e^{\rho_n} - 1)(1 + \Delta N_n)\ \,|\, \mathscr{F}_{n-1}\big] = e^{r_n} - 1. \tag{46}$$

EXAMPLE 1. We consider a single-step model (44) with $n = 0$ or 1, where we set $\mathscr{F}_0 = \{\varnothing, \Omega\}$ and $Z_0 = 1$. Then $1 + \Delta N_1 = Z_1$ and by (46),

$$\mathsf{E}e^{\rho_1} Z_1 = e^r.$$

We shall assume that $\Omega = \mathbb{R}$, $\rho_1(x) = x$, $Z_1 = Z(x)$, and let $F = F(x)$ be the probability distribution in Ω. Then the above condition is equivalent to the equality

$$\int_{-\infty}^{\infty} e^x Z_1(x)\, dF(x) = e^r. \tag{47}$$

Thus it is clear that finding *all* distributions $\widetilde{F} = \widetilde{F}(x)$ equivalent to $F = F(x)$ (in the sense of the equivalence of the corresponding Lebesgue–Stieltjes measures) is the same as describing *all positive* solutions $Z_1 = Z_1(x)$ of equation (47) satisfying the condition

$$\int_{-\infty}^{\infty} Z_1(x)\, dF(x) = 1. \tag{48}$$

For instance, if $F \sim \mathscr{N}(m,\sigma^2)$ with $\sigma^2 > 0$, then (45) and (46) take the following form:

$$\int_{-\infty}^{\infty} e^x Z_1(x)\varphi_{m,\sigma^2}(x)\,dx = e^r, \qquad \int_{-\infty}^{\infty} Z_1(x)\varphi_{m,\sigma^2}(x)\,dx = 1, \tag{49}$$

where $\varphi_{m,\sigma^2}(x) = \dfrac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}$ is the density of the normal distribution.

From (49) we see that if we seek a martingale measure *in the class of normal distributions* $\mathscr{N}(\widetilde{m},\widetilde{\sigma}^2)$ *with* $\widetilde{\sigma}^2 > 0$, i.e., if

$$Z_1(x) = \frac{\varphi_{\widetilde{m},\widetilde{\sigma}^2}(x)}{\varphi_{m,\sigma^2}(x)}, \tag{50}$$

then 'admissible' pairs $(\widetilde{m},\widetilde{\sigma}^2)$ must satisfy the condition

$$\int_{-\infty}^{\infty} e^x \varphi_{\widetilde{m},\widetilde{\sigma}^2}(x)\,dx = e^r,$$

which is equivalent to

$$\widetilde{m} + \frac{\widetilde{\sigma}^2}{2} = r. \tag{51}$$

In other words all the pairs $(\widetilde{m},\widetilde{\sigma}^2)$ with $\widetilde{\sigma}^2 > 0$ satisfying (51) are admissible. We have already encountered this condition with $r = 0$ in § 3c (see (14)).

Note that, besides the solutions $Z_1(x)$ of the form (50), the system (49) has other solutions, and their general form is probably unknown. This shows the complexity of the description problem for all martingale measures even in the case of the above simple 'single-step' scheme.

Our next example can be characterized as 'too simple' in this respect, for we have there a *unique*, easily calculated 'martingale' measure.

EXAMPLE 2. *The CRR-model.* Let

$$\begin{aligned} \Delta B_n &= rB_{n-1}, \\ \Delta S_n &= \rho_n S_{n-1}, \end{aligned} \tag{52}$$

where $n \leqslant N$ and B_0 and S_0 are positive constants.

It is assumed in the framework of this model that (ρ_n) is a sequence of *independent identically distributed* random variables taking two values, b and a, such that

$$-1 < a < r < b$$

and

$$\mathsf{P}_N(\rho_n = b) = p, \qquad \mathsf{P}_N(\rho_n = a) = q, \tag{53}$$

where $0 < p < 1$ and $p + q = 1$.

Since the variables ρ_n, $n \geqslant 1$, are the unique source of 'randomness' in the model, our space of elementary outcomes can be the space $\Omega = \{a, b\}^N$ of sequences $(x_1, \dots, x_N)$ with $x_i = a, b$, and the functions $\rho_n = \rho_n(x)$, $x = (x_1, \dots, x_N)$, can be defined coordinate-wise: $\rho_n(x) = x_n$.

We can define in a standard way a probability measure $\mathsf{P}_N = \mathsf{P}_N(x_1, \dots, x_N)$ such that $\rho_1, \dots, \rho_N$ are independent variables with respect to P_N and (53) holds; namely,

$$\mathsf{P}_N(x_1, \dots, x_N) = p^{\nu_b(x_1, \dots, x_N)} q^{N - \nu_b(x_1, \dots, x_N)},$$

where $\nu_b(x_1, \dots, x_N) = \sum_{i=1}^{N} I_b(x_i)$ is the number of the x_i equal to b.

We shall construct the measure $\widetilde{\mathsf{P}}_N \sim \mathsf{P}_N$ in several steps: we set $\mathsf{P}_n = \mathsf{P}_N \mid \mathscr{F}_n$, where $\mathscr{F}_n = \sigma(\rho_1, \dots, \rho_n)$, and define $\widetilde{\mathsf{P}}_1, \widetilde{\mathsf{P}}_2, \dots, \widetilde{\mathsf{P}}_N$ by the formulas

$$\widetilde{\mathsf{P}}_n(x_1, \dots, x_n) = Z_n(x_1, \dots, x_n)\, \mathsf{P}_n(x_1, \dots, x_n),$$

where we shall (successively) find the Z_n from condition (43), which, bearing in mind that $1 + \Delta N_n = \dfrac{Z_n}{Z_{n-1}}$, can be rewritten as

$$\mathsf{E}_{\mathsf{P}_n}\left[\rho_n \frac{Z_n}{Z_{n-1}} \;\Big|\; \mathscr{F}_{n-1}\right] = r. \tag{54}$$

For $n = 1$ (with $\mathscr{F}_0 = \{\varnothing, \Omega\}$) we obtain by (54) that

$$pbZ_1(b) + qaZ_1(a) = r, \tag{55}$$

which, together with the normalization condition

$$pZ_1(b) + qZ_1(a) = 1, \tag{56}$$

delivers *with necessity* the equalities

$$Z_1(b) = \frac{r-a}{b-a} \cdot \frac{1}{p} \quad \text{and} \quad Z_1(a) = \frac{b-r}{b-a} \cdot \frac{1}{q}. \tag{57}$$

We set

$$\widetilde{p} = \frac{r-a}{b-a} \quad \text{and} \quad \widetilde{q} = \frac{b-r}{b-a}.$$

Then

$$\begin{aligned} \widetilde{\mathsf{P}}_1(b) &= Z_1(b)\, \mathsf{P}_1(b) = \widetilde{p}, \\ \widetilde{\mathsf{P}}_1(a) &= Z_1(a)\, \mathsf{P}_1(a) = \widetilde{q}. \end{aligned} \tag{58}$$

To find $\widetilde{\mathsf{P}}_2$ (and, after that, $\widetilde{\mathsf{P}}_3, \dots, \widetilde{\mathsf{P}}_N$) we use (54) again. Based on the fact that ρ_1 and ρ_2 are independent with respect to P_2, we obtain by (52) that

$$bp\frac{Z_2(b,b)}{Z_1(b)} + aq\frac{Z_2(b,a)}{Z_1(b)} = r. \tag{59}$$

An additional condition on the values of $Z_2(b,b)$ and $Z_2(b,a)$ can be obtained from the martingale property

$$\mathsf{E}_{\mathsf{P}_2}[\,Z_2(\rho_1,\rho_2)\mid\rho_1 = b\,] = Z_1(b),$$

which brings us to the equality

$$p\frac{Z_2(b,b)}{Z_1(b)} + q\frac{Z_2(b,a)}{Z_1(b)} = 1. \tag{60}$$

Comparing (59) and (60) with (55) and (56) we see that

$$\frac{Z_2(b,b)}{Z_1(b)} = \frac{r-a}{b-a}\cdot\frac{1}{p} = \frac{\widetilde{p}}{p}, \qquad \frac{Z_2(b,a)}{Z_1(b)} = \frac{\widetilde{q}}{q}.$$

In a similar way,

$$\frac{Z_2(a,b)}{Z_1(a)} = \frac{\widetilde{p}}{p} \quad\text{and}\quad \frac{Z_2(a,a)}{Z_1(a)} = \frac{\widetilde{q}}{q}.$$

Hence

$$\widetilde{\mathsf{P}}_2(a,a) = Z_2(a,a)q^2 = Z_1(a)\frac{\widetilde{q}}{q}\cdot q^2 = \widetilde{q}^{\,2},$$
$$\widetilde{\mathsf{P}}_2(a,b) = \widetilde{q}\,\widetilde{p}, \qquad \widetilde{\mathsf{P}}_2(b,a) = \widetilde{p}\,\widetilde{q}, \qquad \widetilde{\mathsf{P}}_2(b,b) = \widetilde{p}^{\,2}.$$

It is now clear that the variables ρ_1 and ρ_2 are *identically distributed* and *independent* with respect to the measure $\widetilde{\mathsf{P}}_2$; moreover, $\widetilde{\mathsf{P}}_2(\rho_i = b) = \widetilde{p}$ and $\widetilde{\mathsf{P}}_2(\rho_i = a) = \widetilde{q}$, $i = 1,2$.

The next steps, the specification of $\widetilde{\mathsf{P}}_3, \dots, \widetilde{\mathsf{P}}_N$, proceed in a similar manner, which brings us to the following result.

THEOREM 2. *A martingale measure $\widetilde{\mathsf{P}}_N$ in the CRR model defined in* (52) *and* (53) *is unique and can be defined by the formula*

$$\widetilde{\mathsf{P}}_N(x_1,\dots,x_N) = \widetilde{p}^{\,\nu_b(x_1,\dots,x_N)}\widetilde{q}^{\,N-\nu_b(x_1,\dots,x_N)}, \tag{61}$$

where

$$\widetilde{p} = \frac{r-a}{b-a}, \qquad \widetilde{q} = \frac{b-r}{b-a}. \tag{62}$$

Remark. We note that it was not *a priori* obvious that the martingale measure $\widetilde{\mathsf{P}}_N$ was a *direct product* of 'one-dimensional' distributions:

$$\widetilde{\mathsf{P}}_N = \underbrace{\widetilde{\mathsf{P}}_1 \otimes \cdots \otimes \widetilde{\mathsf{P}}_1}_{N},$$

i.e., that the variables $\rho_1, \ldots, \rho_N$, which are independent and identically distributed with respect to the initial measure P_N, are also independent and identically distributed with respect to the martingale measure $\widetilde{\mathsf{P}}_N$.

4. Complete and Perfect Arbitrage-Free Markets

§ 4a. Martingale Criterion of a Complete Market. Statement of the Second Fundamental Theorem. Proof of Necessity

1. In accordance with the definitions in § 1b, a (B, S)-market defined on a filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_n), \mathsf{P})$, where $0 \leqslant n \leqslant N$, $\mathscr{F}_0 = \{\varnothing, \Omega\}$, and $\mathscr{F}_N = \mathscr{F}$, is said to be *complete* (*perfect*) or N-*complete* (N-*perfect*) if each $\mathscr{F}_N$-measurable *bounded* (*finite*) pay-off function $f_N = f_N(\omega)$ is replicable: there exist a self-financing portfolio π and an initial capital x such that $X_0^\pi = x$ and

$$X_N^\pi = f_N \quad (\mathsf{P}\text{-a.s.}).$$

Let $\mathscr{P}(\mathsf{P})$ be the collection of all martingale measures $\widetilde{\mathsf{P}} \sim \mathsf{P}$, such that the discounted prices $\dfrac{S}{B}$ are martingales. It is assumed (see § 1a, 2a) that $B = (B_n)_{0 \leqslant n \leqslant N}$ is a *risk-free* asset and $S = (S_n)_{0 \leqslant n \leqslant N}$ with $S_n = (S_n^1, \dots, S_n^d)$, $d < \infty$, is a multi-dimensional *risk* asset.

The asset B is usually regarded as a *bank account*; the assets S^i are called *stock*.

In what follows we assume that $B_n > 0$ for $n \geqslant 0$. Then we can set $B_n \equiv 1$, $n \geqslant 0$, without loss of generality.

The next result is so important that it may well be called the '*Second fundamental asset pricing theorem*'.

THEOREM B. *An arbitrage-free financial (B, S)-market (with $N < \infty$ and $d < \infty$) is complete if and only if the set $\mathscr{P}(\mathsf{P})$ of martingale measures contains a single element.*

Thus, while the absence of arbitrage means that

$$\mathscr{P}(\mathsf{P}) \neq \varnothing,$$

the completeness of an arbitrage-free market can (provisory) be written as

$$|\mathscr{P}(\mathsf{P})| = 1.$$

We now make several observations relating to the proof of this theorem.

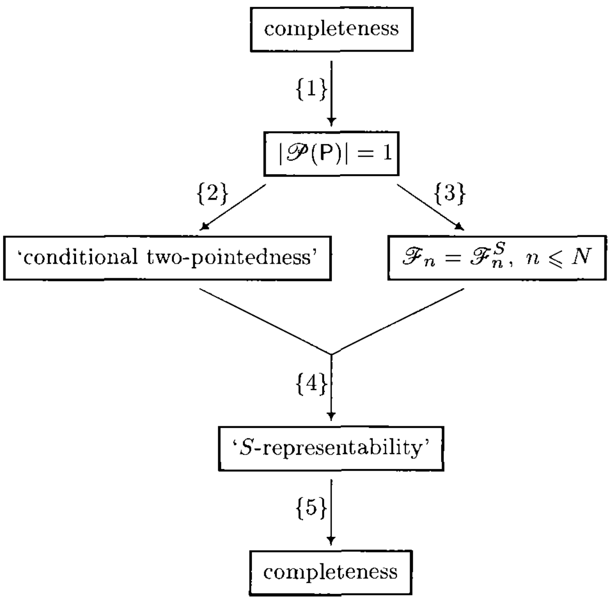
It is well known in stochastic calculus (see, e.g., [250; Chapter III]) that the *uniqueness* of a martingale measure is intimately connected with the issues of '*representatability*' of local martingales in terms of certain basic martingales. The corresponding results (especially in the continuous-time case) are considered to be technically difficult, because one must essentially use ideas and tools of the stochastic analysis of semimartingales and random measures in their proofs.

Incidentally, in the discrete-time case this range of issues pertaining to the 'representability' problem for local martingales and the completeness of a (B, S)-market can be discussed on a relatively elementary level. We start our discussion with the case of $d = 1$ (§ § 4a–4e). The general case $d \geqslant 1$ is considered in § 4f.

2. The idea of the proof of Theorem B in the case of $d = 1$ is to establish the following chain of implications (involving the concepts of 'conditional two-pointedness' and 'S-representability', which are introduced below, in § § 4b,e, and the equality $\mathscr{F}_n = \mathscr{F}_n^S$ meaning that the σ-algebra $\mathscr{F}_n$ coincides with the σ-algebra

$$\mathscr{F}_n^S = \sigma(S_1, \dots, S_n)$$

generated by the random variables $S_1, \dots, S_n$ up to sets of P-measure zero):



Implication $\{1\}$, i.e., the *necessity* in Theorem B, has a relatively easy proof that proceeds as follows.

Let $A \in \mathscr{F}_N$. We set $f_N = I_A(\omega)$. In view of the assumed completeness there exists a starting capital x and a self-financing strategy π such that if $X_0^\pi = x$, then $X_N^\pi = f_N$ (P-a.s.).

Since π is self-financing strategy, it follows that

$$X_n^\pi = X_0^\pi + \sum_{k=1}^n \gamma_k \Delta S_k.$$

Let P_i, $i = 1, 2$, be two martingale measures in the family $\mathscr{P}(\mathsf{P})$. Then $(X_n^\pi)_{n \leqslant N}$ is a martingale transform, and since $X_N^\pi = I_A$, the sequence $X^\pi = (X_n^\pi)_{n \leqslant N}$ is a martingale with respect to each martingale measure P_i, $i = 1, 2$, by the lemma in Chapter II, § 1c.

Hence

$$x = X_0^\pi = \mathsf{E}_{\mathsf{P}_i}(X_N^\pi \mid \mathscr{F}_0) = \mathsf{E}_{\mathsf{P}_i} I_A = \mathsf{P}_i(A)$$

for $i = 1, 2$, so that $\mathsf{P}_1(A) = \mathsf{P}_2(A)$, $A \in \mathscr{F}_N$.

Thus, the measures P_1 and P_2 are in fact the same, which proves that the set $\mathscr{P}(\mathsf{P})$ (non-empty due to the absence of arbitrage in our (B, S)-market) has at most one element ($|\mathscr{P}(\mathsf{P})| = 1$). This proves the necessity in Theorem B (implication $\{1\}$).

In the next section we consider the issues of 'representability', which are important for the discussion of implications $\{4\}$ and $\{5\}$.

§ 4b. Representability of Local Martingales. 'S-Representability'

From the standpoint of the 'general theory of martingales and stochastic calculus' (see [102], [103], [250], [304]) the assumption of 'completeness' is in fact equivalent to the so-called property of 'S-representability' of local martingales ([250; Chapter III]).

DEFINITION. Let $(\Omega, \mathscr{F}, (\mathscr{F}_n), \mathsf{P})$ be a filtered probability space with

a d-dimensional (basic) martingale $S = (S_n, \mathscr{F}_n, \mathsf{P})$

and

a (one-dimensional) local martingale $X = (X_n, \mathscr{F}_n, \mathsf{P})$.

Then we say that the local martingale X on $(\Omega, \mathscr{F}, (\mathscr{F}_n), \mathsf{P})$ admits an '*S-representation*' or a *representation in terms of the* P-*martingale* S if there exists a predictable sequence $\gamma = (\gamma_n)$, where $\gamma_n = (\gamma_n^1, \dots, \gamma_n^d)$, such that

$$X_n = X_0 + \sum_{k=1}^n \gamma_k \Delta S_k \quad \left(= X_0 + \sum_{k=1}^n \left(\sum_{j=1}^d \gamma_k^j \Delta S_k^j \right) \right) \tag{1}$$

P-a.s. for each $n \geqslant 1$, i.e., X is a martingale transform obtained from the P-martingale S by 'integration' of the predictable sequence γ; see Chapter II, § 1c.

The next result relates to implication {5} in the chain of implication on the previous page.

LEMMA. *Let (B, S) be an arbitrage-free market over a finite time horizon N, let $B_n \equiv 1$ for $n \leqslant N$, and let $\mathscr{P}(\mathsf{P})$ be the family of measures $\widetilde{\mathsf{P}}$ equivalent to P (on $(\Omega, \mathscr{F})$, where $\mathscr{F} = \mathscr{F}_N$), such that $S = (S_n)_{n\geqslant 0}$ is a $\widetilde{\mathsf{P}}$-martingale.*

Then this market is complete if and only if there exists a measure $\widetilde{\mathsf{P}} \in \mathscr{P}(\mathsf{P})$ such that each bounded martingale $X = (X_n, \mathscr{F}_n, \widetilde{\mathsf{P}})$ (with $|X_n(\omega)| \leqslant C$, $n \leqslant N$, $\omega \in \Omega$) on $(\Omega, \mathscr{F}, (\mathscr{F}_n), \widetilde{\mathsf{P}})$ admits an 'S-representation'.

Proof. (a) Assume that our (arbitrage-free) market is complete. We take an arbitrary measure in $\mathscr{P}(\mathsf{P})$ as a required measure $\widetilde{\mathsf{P}}$. Let $X = (X_n, \mathscr{F}_n, \widetilde{\mathsf{P}})_{n\leqslant N}$ be a martingale with $|X_n(\omega)| \leqslant C$, $n \leqslant N$, $\omega \in \Omega$.

We set $f_N = X_N$. The completeness assumption means that there exists a self-financing portfolio π and an initial capital x such that (P- and $\widetilde{\mathsf{P}}$-a.s.)

$$X_n^\pi = x + \sum_{k=1}^{n} \gamma_k \Delta S_k \tag{2}$$

and $X_N^\pi = f_N = X_N$. However, $|f_N| \leqslant C$, therefore $X^\pi = (X_n^\pi)_{n\leqslant N}$ is a $\widetilde{\mathsf{P}}$-martingale (the lemma in Chapter II, § 1c), so that the $\widetilde{\mathsf{P}}$-martingales X^π and X, which have the same terminal pay-off function f_N, actually coincide (P- and $\widetilde{\mathsf{P}}$-a.s.). Hence the martingale X admits an 'S-representation'.

(b) Now let $f_N = f_N(\omega)$ be a $\mathscr{F}_N$-measurable bounded function ($|f_N| \leqslant C < \infty$ P-a.s.). We claim that there exist a self-financing portfolio π and a starting capital x such that $X_N^\pi = f_N$ (P-a.s.).

By assumption, there exists a measure $\widetilde{\mathsf{P}} \in \mathscr{P}(\mathsf{P})$, such that each bounded $\widetilde{\mathsf{P}}$-martingale has an 'S-representation'.

We consider one such martingale

$$X = (X_n, \mathscr{F}_n, \widetilde{\mathsf{P}})_{n\leqslant N}, \quad \text{where} \quad X_n = \mathsf{E}_{\widetilde{\mathsf{P}}}(f_N \mid \mathscr{F}_n).$$

Since $|f_N| \leqslant C$, X is a bounded (Lévy) martingale and it has a representation (1) with some $\mathscr{F}_{k-1}$-measurable variables γ_k^j, $j = 1, \dots, d$, $k \leqslant N$.

For these variables we construct a portfolio $\pi^* = (\beta^*, \gamma^*)$ such that $\gamma^* = \gamma$ and $\beta_n^* = X_n - \sum_{j=1}^{d} \gamma_n^j S_n^j$.

By (1) we obtain that the β_n^* are $\mathscr{F}_{n-1}$-measurable. Moreover,

$$\sum_{j=1}^{d} S_{n-1}^j \Delta\gamma_n^{*j} + \Delta\beta_n^* = \sum_{j=1}^{d} S_{n-1}^j \Delta\gamma_n^j + \left(\Delta X_n - \Delta\Big(\sum_{j=1}^{d} \gamma_n^j S_n^j\Big)\right)$$
$$= \sum_{j=1}^{d} S_{n-1}^j \Delta\gamma_n^j + \sum_{j=1}^{d} \gamma_n^j \Delta S_n^j - \Delta\Big(\sum_{j=1}^{d} \gamma_n^j S_n^j\Big) = 0.$$

Hence π^* is a self-financing portfolio and

$$X_n^{\pi^*} = \beta_n^* + \sum_{j=1}^{d} \gamma_n^j S_n^j = X_n;$$

in particular, $X_N^{\pi^*} = X_N = f_N$ ($\widetilde{\mathsf{P}}$- and P-a.s.), i.e., our (B,S)-market is complete, which proves the lemma.

Remark. If we do not assume that $B_n \equiv 1$, $n \leqslant N$, then all our results are valid for the $\widetilde{\mathsf{P}}$-martingale $\dfrac{S}{B} = \left(\dfrac{S_n}{B_n}\right)_{n\leqslant N}$ in place of the $\widetilde{\mathsf{P}}$-martingale $S = (S_n)_{n\leqslant N}$.

§ 4c. Representability of Local Martingales ('μ-Representability' and '$(\mu-\nu)$-Representability')

1. The issue of 'S-representability' is, as shown in the preceding section, closely related to the 'completeness' of the corresponding market and the fact that the evolution of the capital X^π is described by (2).

As shown in § 4d, we have 'S-representability' in the *CRR*-model, so that the market is complete in this case. Generally speaking, completeness (and therefore also 'S-representability') is *an exception rather than a rule.* It is reasonable therefore to consider here another kind of representations of local martingales, which uses the concepts of random measures μ and martingale random measures $\mu-\nu$; see § 3e. It will be clear from what follows that 'μ-representations' and '$(\mu-\nu)$-representations' are significantly more widespread than 'S-representations'. Hence it often makes sense to find a 'μ' or '$(\mu-\nu)$-representation' first and attempt to transform them into an S-representation after that.

2. Let $S = (S^1, \dots, S^d)$ be a d-dimensional martingale on a filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_n), \mathsf{P})$ with $\mathscr{F}_0 = \{\varnothing, \Omega\}$ (with respect to the original measure P and the flow $(\mathscr{F}_n)$).

Let

$$\mathscr{F}_n^S = \sigma\big(S_k^j,\ k \leqslant n,\ j = 1, \dots, d\big),$$

$n \geqslant 1$, be the σ-algebra generated by the prices and let

$$X = (X_n, \mathscr{F}_n^S, \mathsf{P})$$

be a local martingale.

The increments $\Delta X_n = X_n - X_{n-1}$ are $\mathscr{F}_n^S$-measurable, therefore there exists a Borel function $f_n = f_n(x_1, \dots, x_n)$, $x_i \in \mathbb{R}^d$ such that

$$\Delta X_n(\omega) = f_n\big(\Delta S_1(\omega), \dots, \Delta S_n(\omega)\big), \qquad \omega \in \Omega.$$

(Since we have set $\mathscr{F}_0 = \{\varnothing, \Omega\}$, the vector S_0 is not random, and there exists a one-to-one correspondence between $(S_1, \dots, S_n)$ and $(\Delta S_1, \dots, \Delta S_n)$.)

For each $n \geqslant 1$ we now set

$$W_n(\omega, x) = f_n\big(\Delta S_1(\omega), \dots, \Delta S_{n-1}(\omega), x\big).$$

This function is clearly measurable in (ω, x), $\mathscr{B}(\mathbb{R}^d)$-measurable in x for each $\omega \in \Omega$ and $\mathscr{F}_{n-1}^S$-measurable for each $x \in \mathbb{R}^d$.

Let $\mu_n(A; \omega) = I(\Delta S_n(\omega) \in A)$, $A \in \mathscr{B}(\mathbb{R}^d)$, be the integer-valued random measure constructed from the increments $\Delta S_n(\omega)$, $n \geqslant 1$. Then

$$\Delta X_n(\omega) = \int_{\mathbb{R}^d} W_n(\omega, x)\, \mu_n(dx; \omega), \tag{1}$$

so that we obtain the so-called '*μ-representation*' of X:

$$X_n(\omega) = X_0(\omega) + \sum_{k=1}^{n} \int_{\mathbb{R}^d} W_k(\omega, x)\, \mu_k(dx; \omega), \tag{2}$$

or, in a more compact form,

$$X = X_0 + W * \mu \tag{3}$$

(see § 3e).

We now note that since X is a local martingale by assumption, it follows that $\mathsf{E}(|\Delta X_n| \,|\, \mathscr{F}_{n-1}^S) < \infty$ and $\mathsf{E}(\Delta X_n \,|\, \mathscr{F}_{n-1}^S) = 0$, $n \geqslant 1$. Hence if

$$\nu_n(A; \omega) = \mathsf{E}\big(\mu_n(A; \cdot\,) \,|\, \mathscr{F}_{n-1}^S\big)(\omega), \tag{4}$$

then we can see that

$$\int_{\mathbb{R}^d} W_k(\omega, x)\, \nu_k(dx; \omega) = \mathsf{E}(\Delta X_k \,|\, \mathscr{F}_{k-1}^S)(\omega) = 0.$$

Thus, besides (2) and (3), we obtain the so-called '*$(\mu-\nu)$-representation*'

$$X_n(\omega) = X_0(\omega) + \sum_{k=1}^{n} \int_{\mathbb{R}^d} W_k(\omega, x)\, \big(\mu_k(dx; \omega) - \nu_k(dx; \omega)\big), \tag{5}$$

or, in a more compact form,

$$X = X_0 + W * (\mu - \nu). \tag{6}$$

In may seem odd that, alongside the natural representation (3), we also consider the representation (6), which can be obtained from the first one in a trivial way, and honor it with a special name '$(\mu-\nu)$-representation', assigning to it certain importance by doing so. The reason is as follows.

First, in some more general situations (continuous time, more general flows of σ-algebras, etc.) similar representations of local martingales involve stochastic integrals just of the kind $W * (\mu - \nu)$ (see, e.g., [250; Chapter III, § 4.23, 4.24]). Second, the expressions of type $W * (\mu - \nu)$ (as compared with $W * \mu$) have one advantage: in general, there is no unique way to define the function W in these representations, while in expressions of the type $W * (\mu - \nu)$ these functions can often be chosen pretty simple.

We present the following example as an illustration.

Let $A \in \mathscr{B}(\mathbb{R}^d)$ and let $X^{(A)} = (X_n^{(A)}, \mathscr{F}_n^S, \mathsf{P})$ be a martingale with $X_0^{(A)} = 0$ and

$$\Delta X_n^{(A)}(\omega) = \mu_n(A;\omega) - \nu_n(A;\omega) = I\big(\Delta S_n(\omega) \in A\big) - \mathsf{E}\big(I(\Delta S_n) \in A \,|\, \mathscr{F}_{n-1}^S\big)(\omega).$$

If we set $W_n^{(A)}(\omega, x) = I_A(x)$, then

$$\Delta X_n^{(A)}(\omega) = \int_{\mathbb{R}^d} W_n^{(A)}(\omega, x)\big(\mu_n(dx;\omega) - \nu_n(dx;\omega)\big),$$

so that

$$X^{(A)} = W^{(A)} * (\mu - \nu).$$

On the other hand, setting

$$W_n(\omega, x) = I_A(x) - \mathsf{E}\big(I_A(\Delta S_n) \,|\, \mathscr{F}_{n-1}^S\big)(\omega)$$

we obtain

$$\int_{\mathbb{R}^d} W_n(\omega, x)\, \mu_n(dx;\omega) = \mu_n(A;\omega) - \nu_n(A;\omega) = \Delta X_n^{(A)}(\omega),$$

so that

$$X^{(A)} = W * \mu.$$

Clearly, the function $W^{(A)}$ is simpler as W.

These arguments show, in particular, that the integral $W * \mu$ does nor change after the replacement of the functions $W_n(\omega, x)$ by $W_n(\omega, x) + g_n'(x)$, where $g_n'(x)$ satisfies the equality

$$\int_{\mathbb{R}^d} g_n'(x)\, \mu_n(dx;\omega) = 0.$$

In turn, the integral $W * (\mu - \nu)$ does not change if we replace $W_n(\omega, x)$ by $W_n(\omega, x) + g_n''(\omega)$ because

$$\int_{\mathbb{R}^d} g_n''(x)\big(\mu_n(dx;\omega) - \nu_n(dx;\omega)\big) = 0.$$

In the next section we show how one can easily deduce an 'S-representation' from a '$(\mu-\nu)$-representation' in the CRR-model of Cox–Ross–Rubinstein.

§ 4d. 'S-Representability' in the Binomial CRR-Model

1. As shown in § 3f (Example 2 in subsection 5), there *exists* a martingale measure in the CRR-model (so that the corresponding market is arbitrage-free), which (assuming that discuss the coordinate probability space) is the *unique* martingale measure. By Theorem B, this is equivalent to the completeness of the corresponding market.

It would be interesting for that reason to trace down how the *uniqueness* of the martingale measure in this particular model delivers the 'S-representability' and, therefore, the completeness of the market (following by the lemma in § 4b).

First we recall some notation.

As explained in Chapter I, § 1e, the CRR-model of a (B, S) market defined on a filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_n), \mathsf{P})_{n\geqslant 0}$ is described by two sequences $B = (B_n)_{n\geqslant 0}$ and $S = (S_n)_{n\geqslant 0}$ such that

$$B_n = B_{n-1}(1 + r_n), \tag{1}$$

$$S_n = S_{n-1}(1 + \rho_n), \tag{2}$$

where the r_n are $\mathscr{F}_{n-1}$-measurable, the ρ_n are $\mathscr{F}_n$-measurable, and $B_0 > 0$ and $S_0 > 0$ are constants.

Since

$$\frac{S_n}{B_n} = \frac{S_{n-1}}{B_{n-1}} \cdot \frac{1+\rho_n}{1+r_n}, \tag{3}$$

it clearly follows that $\left(\dfrac{S_n}{B_n}\right)_{n\geqslant 0}$ is a martingale with respect to a measure $\widetilde{\mathsf{P}}$ if, first,

$$\widetilde{\mathsf{E}}\left|\frac{1+\rho_n}{1+r_n}\right| < \infty$$

(where $\widetilde{\mathsf{E}}$ is averaging with respect to $\widetilde{\mathsf{P}}$) and, second,

$$\widetilde{\mathsf{E}}\left(\frac{1+\rho_n}{1+r_n} \,\Big|\, \mathscr{F}_{n-1}\right) = 1. \tag{4}$$

In view of the $\mathscr{F}_{n-1}$-measurability of the variables r_n the condition (4) reduces to the equality

$$\widetilde{\mathsf{E}}(\rho_n \,|\, \mathscr{F}_{n-1}) = r_n. \tag{5}$$

2. In the Binomial *CRR*-model one has $r_n \equiv r$, where r is a constant and $(\rho_n)_{n\geqslant 1}$ is a sequence of *independent identically distributed* random variables taking two values, b and a, with positive probabilities

$$p = \mathsf{P}(\rho_n = b) \quad \text{and} \quad q = \mathsf{P}(\rho_n = a), \tag{6}$$

$p + q = 1$. (We also agree that $a < b$.) Finally, let $\mathscr{F}_n = \sigma(\rho_1, \dots, \rho_n)$ for $n \geqslant 1$ and let $\mathscr{F}_0 = \{\varnothing, \Omega\}$.

If we require that the sequence $\dfrac{S}{B} = \left(\dfrac{S_n}{B_n}\right)_{n\geqslant 0}$ be a martingale with respect to a measure $\widetilde{\mathsf{P}}$ such that $\widetilde{\mathsf{P}} \overset{\text{loc}}{\sim} \mathsf{P}$, then the quantities $\widetilde{p}_n = \widetilde{\mathsf{P}}(\rho_n = b)$ and $\widetilde{q}_n = \widetilde{\mathsf{P}}(\rho_n = a)$, in view of the equality $\widetilde{\mathsf{E}}\rho_n = r$, must satisfy the condition

$$b\widetilde{p}_n + a\widetilde{q}_n = r,$$

which, in view of the normalization $\widetilde{p}_n + \widetilde{q}_n = 1$, shows that

$$\widetilde{p}_n = \widetilde{p} \equiv \frac{r-a}{b-a}, \qquad \widetilde{q}_n = \widetilde{q} \equiv \frac{b-r}{b-a}. \tag{7}$$

In order that these values be positive one requires that $a < r < b$.

We shall also assume that $a > -1$. Then $S_n > 0$ for all $n \geqslant 1$ because $S_0 > 0$.

Let $X = (X_n, \mathscr{F}_n, \widetilde{\mathsf{P}})_{n\geqslant 0}$ be a martingale and let $\mathscr{F}_0 = \{\varnothing, \Omega\}$ and $\mathscr{F}_n = \sigma(\rho_1, \dots, \rho_n)$ for $n \geqslant 1$.

For $n \geqslant 1$ we set $\mu_n(A; \omega) = I(\rho_n(\omega) \in A)$ and $\widetilde{\nu}_n(A) = \widetilde{\mathsf{E}}\mu_n(A;\omega)$. Since ρ_n takes only two values, the measures $\mu_n(\,\cdot\,;\omega)$ and $\widetilde{\nu}_n(\,\cdot\,)$ are concentrated at a and b. Moreover,

$$\mu_n(\{a\};\omega) = I\big(\rho_n(\omega) = a\big), \qquad \widetilde{\nu}_n(\{a\}) = \widetilde{q}$$

and

$$\mu_n(\{b\};\omega) = I\big(\rho_n(\omega) = b\big), \qquad \widetilde{\nu}_n(\{b\}) = \widetilde{p}.$$

Let $g_n = g_n(x_1, \dots, x_n)$ be functions such that

$$X_n(\omega) = g_n\big(\rho_1(\omega), \dots, \rho_n(\omega)\big),$$

and therefore

$$\Delta X_n(\omega) = g_n\big(\rho_1(\omega), \dots, \rho_n(\omega)\big) - g_{n-1}\big(\rho_1(\omega), \dots, \rho_{n-1}(\omega)\big).$$

Since $\widetilde{\mathsf{E}}(\Delta X_n \mid \mathscr{F}_{n-1}) = 0$, it follows that

$$\widetilde{p} \cdot g_n\big(\rho_1(\omega), \dots, \rho_{n-1}(\omega), b\big) + \widetilde{q} \cdot g_n\big(\rho_1(\omega), \dots, \rho_{n-1}(\omega), a\big) \\ = g_{n-1}\big(\rho_1(\omega), \dots, \rho_{n-1}(\omega)\big)$$

or, equivalently,

$$\frac{g_n\big(\rho_1(\omega),\dots,\rho_{n-1}(\omega),b\big)-g_{n-1}\big(\rho_1(\omega),\dots,\rho_{n-1}(\omega)\big)}{\widetilde{q}}$$
$$=\frac{g_{n-1}\big(\rho_1(\omega),\dots,\rho_{n-1}(\omega)\big)-g_n\big(\rho_1(\omega),\dots,\rho_{n-1}(\omega),a\big)}{\widetilde{p}}. \tag{8}$$

In view of (7), this can be rewritten as follows:

$$\frac{g_n\big(\rho_1(\omega),\dots,\rho_{n-1}(\omega),b\big)-g_{n-1}\big(\rho_1(\omega),\dots,\rho_{n-1}(\omega)\big)}{b-r}$$
$$=\frac{g_n\big(\rho_1(\omega),\dots,\rho_{n-1}(\omega),a\big)-g_{n-1}\big(\rho_1(\omega),\dots,\rho_{n-1}(\omega)\big)}{a-r}. \tag{9}$$

We proceed now to 'μ-representations'. In accordance with (1) in § 4c,

$$\Delta X_n(\omega)=W_n\big(\omega,\rho_n(\omega)\big)=\int W_n(\omega,x)\,\mu_n(dx;\omega), \tag{10}$$

where

$$W_n(\omega,x)=g_n\big(\rho_1(\omega),\dots,\rho_{n-1}(\omega),x\big)-g_{n-1}\big(\rho_1(\omega),\dots,\rho_{n-1}(\omega)\big).$$

Setting

$$W_n'(\omega,x)=\frac{W_n(\omega,x)}{x-r}, \tag{11}$$

we see from (10) that

$$\Delta X_n(\omega)=\int (x-r)W_n'(\omega,x)\,\mu_n(dx;\omega). \tag{12}$$

Note that, in view of (9), the function $W_n'(\omega,x)$ is *independent* of x. Hence, denoting the right-hand side (or, equivalently, the left-hand side) of (9) by $\gamma_n'(\omega)$, we obtain

$$\Delta X_n(\omega)=\gamma_n'(\omega)\int (x-r)\,\mu_n(dx;\omega)=\gamma_n'(\omega)(\rho_n-r). \tag{13}$$

Thus, for $X=(X_n,\mathscr{F}_n,\widetilde{\mathsf{P}})$ we obtain the representation

$$X_n(\omega)=X_0(\omega)+\sum_{k=1}^{n}\gamma_k'(\omega)\big(\rho_k(\omega)-r\big). \tag{14}$$

Since

$$\Delta\left(\frac{S_n}{B_n}\right) = \frac{S_{n-1}}{B_{n-1}} \cdot \frac{\rho_n - r}{1+r},$$

it follows that

$$\rho_n - r = (1+r)\frac{B_{n-1}}{S_{n-1}}\Delta\left(\frac{S_n}{B_n}\right), \tag{15}$$

and therefore

$$X_n(\omega) = X_0(\omega) + \sum_{k=1}^{n} \gamma_k(\omega)\Delta\left(\frac{S_k(\omega)}{B_k}\right), \tag{16}$$

where

$$\gamma_k(\omega) = \gamma_k'(\omega)(1+r)\frac{B_{k-1}}{S_{k-1}} \tag{17}$$

are $\mathscr{F}_{k-1}$-measurable functions.

The sequence $\left(\frac{S_k}{B_k}\right)_{k\geqslant 0}$ is a martingale with respect to the measure $\widetilde{\mathsf{P}}$. Hence (16) is just the 'S-representation' of the P-martingale X with respect to the (basic) P-martingale $\left(\frac{S_k}{B_k}\right)_{k\geqslant 0}$.

Using the lemma in § 4b we see that the (B, S)-market described by the *CRR*-model is *complete* for each finite time horizon N.

3. *Remark.* It is worth noting that it was essential in our proof of the uniqueness of the martingale measure in the *CRR*-model in § 3f (Example 2 in subsection 5) that the original probability space was the coordinate one: $\Omega = \{x = (x_1, x_2, \dots)\}$, where $x_i = a$ or $x_i = b$; $\mathscr{F}_n = \sigma(x_1, \dots, x_n)$ for $n \geqslant 1$, and $\mathscr{F} = \bigvee \mathscr{F}_n$. It can be seen from what follows (see § 4f) that in an arbitrary filtered space $(\Omega, \mathscr{F}, (\mathscr{F}_n), \mathsf{P})$ the condition that a martingale measure be *unique* implies automatically that $\mathscr{F}_n = \sigma(S_1, \dots, S_n)$ for $n \geqslant 1$ (up to sets of P-measure zero).

§ 4e. Martingale Criterion of a Complete Market. Proof of Necessity for $d = 1$

1. In accordance with the diagram of implications in § 4a.2, to prove the necessity in Theorem B (i.e., the implication '$|\mathscr{P}(\mathsf{P})| = 1$' $\Longrightarrow$ 'completeness') we must verify implications $\{2\}$, $\{3\}$, and $\{4\}$ in this diagram. (We recall that we established implication $\{5\}$ by the lemma in § 4b and $d = 1$ by assumption.)

We start with the proof of implication $\{4\}$, where we assume that $B_n \equiv 1$ (and therefore $r_n \equiv 0$) for $n \geqslant 1$, which brings no loss of generality, as already mentioned.

To this end we point out that it was a key point of our proof of the 'S-representability' for the *CRR*-model that the probability distributions $\mathrm{Law}(\rho_n \mid \widetilde{\mathsf{P}})$, $n \geqslant 1$, were concentrated at two points (a and b, $a < b$).

In other words, it was important that these were 'two-point' distributions. The corresponding arguments prove to be valid also for more general models once the (regular) *conditional* distributions $\mathrm{Law}(\Delta S_n \mid \mathscr{F}_{n-1}; \widetilde{\mathsf{P}})$ or, equivalently, the *conditional* distributions $\mathrm{Law}(\rho_n \mid \mathscr{F}_{n-1}; \widetilde{\mathsf{P}})$ with $\rho_n = \dfrac{\Delta S_n}{S_{n-1}}$ are '*two-point*' and $\mathscr{F}_n = \sigma(S_1, \dots, S_n)$, $n \geqslant 1$.

Formally, the 'conditional two-pointedness' means that there exist two predictable sequences $a = (a_n)$ and $b = (b_n)$, of random variables $a_n = a_n(\omega)$ and $b = b_n(\omega)$, $n \geqslant 1$, such that

$$\widetilde{\mathsf{P}}(\rho_n = a_n \mid \mathscr{F}_{n-1})(\omega) + \widetilde{\mathsf{P}}(\rho_n = b_n \mid \mathscr{F}_{n-1})(\omega) = 1 \tag{1}$$

and $a_n(\omega) \leqslant 0$, $b_n(\omega) \geqslant 0$ for all $\omega \in \Omega$, $n \geqslant 1$. (If the values of $a_n(\omega)$ and $b_n(\omega)$ 'merge', then we clearly have $a_n(\omega) = b_n(\omega) = 0$; this is an uninteresting degenerate case corresponding to the situation when $\Delta S_n(\omega) = 0$. We can leave this case out of consideration from the very beginning and without loss of generality.)

Let $\widetilde{p}_n(\omega) = \widetilde{\mathsf{P}}(\rho_n = b_n \mid \mathscr{F}_{n-1})(\omega)$ and let $\widetilde{q}_n(\omega) = \widetilde{\mathsf{P}}(\rho_n = a_n \mid \mathscr{F}_{n-1})(\omega)$.

The martingale property of the sequence $S = (S_n, \mathscr{F}_n, \widetilde{\mathsf{P}})$ delivers the equalities $\widetilde{\mathsf{E}}(\rho_n \mid \mathscr{F}_{n-1}) = 0$, $n \geqslant 1$, which mean that

$$b_n(\omega)\widetilde{p}_n(\omega) + a_n(\omega)\widetilde{q}_n(\omega) = 0, \qquad n \geqslant 1. \tag{2}$$

By (1) and (2) we obtain (cf. formula (7) in § 4d)

$$\widetilde{p}_n(\omega) = \frac{-a_n(\omega)}{b_n(\omega) - a_n(\omega)} \quad \text{and} \quad \widetilde{q}_n(\omega) = \frac{b_n(\omega)}{b_n(\omega) - a_n(\omega)}. \tag{3}$$

(If $a_n(\omega) = b_n(\omega) = 0$, then we agree to set $\widetilde{p}_n(\omega) = \widetilde{q}_n(\omega) = \frac{1}{2}$.)

Let $X = (X_n, \mathscr{F}_n^S, \widetilde{\mathsf{P}})$ be a local martingale and let $g_n = g_n(x_1, \dots, x_n)$ be functions such that $X_n(\omega) = g_n(\rho_1(\omega), \dots, \rho_n(\omega))$. By analogy with formula (8) in § 4c we obtain

$$\begin{aligned}
&\frac{g_n\big(\rho_1(\omega), \dots, \rho_{n-1}(\omega), b_n(\omega)\big) - g_{n-1}\big(\rho_1(\omega), \dots, \rho_{n-1}(\omega)\big)}{\widetilde{q}_n(\omega)} \\
&\qquad = \frac{g_{n-1}\big(\rho_1(\omega), \dots, \rho_{n-1}(\omega)\big) - g_n\big(\rho_1(\omega), \dots, \rho_{n-1}(\omega), a_n(\omega)\big)}{\widetilde{p}_n(\omega)}.
\end{aligned} \tag{4}$$

Further, following the pattern of (9)–(17) (§ 4c) we see that X has an 'S-representation':

$$X_n = X_0 + \sum_{k=1}^{n} \gamma_k(\omega)\Delta S_k(\omega) \tag{5}$$

with $\mathscr{F}_{k-1}^S$-measurable functions $\gamma_k(\omega)$, $k \geqslant 1$.

Thus, implication {4} is established.

We consider now the proof of implication {2}, which says that the uniqueness of the martingale measure means the 'conditional two-pointedness' (for $d = 1$).

Taking into account that one can operate the conditional probabilities

$$\widetilde{\mathsf{P}}(\Delta S_n \in \cdot \mid \mathscr{F}_{n-1})(\omega)$$

like ordinary ones (for each fixed $\omega \in \Omega$), the required 'conditional two-pointedness' is equivalent to the following result.

I. Let $\mathsf{Q} = \mathsf{Q}(dx)$ be a probability distribution on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ such that

$$\int_{\mathbb{R}} |x|\, \mathsf{Q}(dx) < \infty, \qquad \int_{\mathbb{R}} x\, \mathsf{Q}(dx) = 0$$

('the martingale property'). Let $\mathscr{P}(\mathsf{Q})$ be the family of all measures $\widetilde{\mathsf{Q}} = \widetilde{\mathsf{Q}}(dx)$ equivalent to $\mathsf{Q} = \mathsf{Q}(dx)$ and having the properties

$$\int_{\mathbb{R}} |x|\, \widetilde{\mathsf{Q}}(dx) < \infty, \qquad \int_{\mathbb{R}} x\, \widetilde{\mathsf{Q}}(dx) = 0.$$

If the family $\mathscr{P}(\mathsf{Q})$ contains only the (original) measure Q, then, of necessity, this is a 'two-point' measure: there exist $a \leqslant 0$ and $b \geqslant 0$ such that

$$\mathsf{Q}(\{a\}) + \mathsf{Q}(\{b\}) = 1,$$

although these points can 'merge' at the origin ($a = b = 0$).

We can put this assertion in the following equivalent form:

II. Let $Z(\mathsf{Q})$ be the class of functions $z = z(x)$, $x \in \mathbb{R}$, such that

$$\mathsf{Q}\{x\colon 0 < z(x) < \infty\} = 1,$$
$$\int_{\mathbb{R}} |x| z(x)\, \mathsf{Q}(dx) < \infty, \qquad \int_{\mathbb{R}} x z(x)\, \mathsf{Q}(dx) = 0.$$

We assume that, for some measure Q, the class $Z(\mathsf{Q})$ contains only functions that are Q-indistinguishable from one ($\mathsf{Q}\{x\colon z(x) \neq 1\} = 0$). Then, of necessity, Q is concentrated at two points at most.

Finally, the same assertion can be reformulated as follows.

III. Let $\xi = \xi(x)$ be a random variable on a coordinate space with distribution $\mathsf{Q} = \mathsf{Q}(dx)$ on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$.

Assume that $\mathsf{E}|\xi| < \infty$, $\mathsf{E}\xi = 0$, and let Q be a measure such that if $\widetilde{\mathsf{Q}} \sim \mathsf{Q}$, $\widetilde{\mathsf{E}}|\xi| < \infty$, and $\widetilde{\mathsf{E}}\xi = 0$, then $\widetilde{\mathsf{Q}} = \mathsf{Q}$.

Then the support of Q consists of at most two points, $a \leqslant 0$ and $b \geqslant 0$, that can stick together at the origin ($a = b = 0$).

To prove these (equivalent) assertions we observe that each probability distribution $\mathsf{Q} = \mathsf{Q}(dx)$ on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ can be represented as the mixture

$$c_1\mathsf{Q}_1 + c_2\mathsf{Q}_2 + c_3\mathsf{Q}_3$$

of three: a purely discrete one (Q_1), an absolutely continuous one (Q_2), and a singular one (Q_3), with non-negative constants c_1, c_2, and c_3 of unit sum.

The idea of the proof becomes already transparent in the 'purely discrete' case, when we assume that Q is concentrated at *three* points x_-, x_0, and x_+, $x_- \leqslant x_0 \leqslant x_+$, with non-zero mass p_-, p_0, or p_+ at each point.

The condition $\mathsf{E}\xi = 0$ means that

$$x_-p_- + x_0p_0 + x_+p_+ = 0. \tag{6}$$

If $x_0 = 0$, then (6) assumes the form $x_-p_- + x_+p_+ = 0$.

We set

$$\widetilde{p}_- = \frac{p_-}{2}, \qquad \widetilde{p}_0 = \frac{1}{2} + \frac{p_0}{2}, \qquad \widetilde{p}_+ = \frac{p_+}{2}, \tag{7}$$

which corresponds to 'pumping' parts of the masses p_- and p_+ at x_- and x_+ to the point $x_0 = 0$.

It is clear from (7) that the measure $\widetilde{\mathsf{Q}} = \{\widetilde{p}_-, \widetilde{p}_0, \widetilde{p}_+\}$ concentrated at the three points x_-, x_0, and x_+, is a probability measure, $\widetilde{\mathsf{Q}} \sim \mathsf{Q}$, and $\widetilde{\mathsf{E}}\xi = 0$; moreover, $\widetilde{\mathsf{Q}} \neq \mathsf{Q}$, which contradicts the uniqueness of Q.

Thus, Q cannot be concentrated at *three* points including $x_0 = 0$.

Now let $x_0 \neq 0$. Then the idea the construction of a measure $\widetilde{\mathsf{Q}} \sim \mathsf{Q}$ by 'mass pumping' from x_- and x_+ to x_0 can be realized, e.g., as follows.

We set

$$\widetilde{p}_- = p_- - \varepsilon_-, \qquad \widetilde{p}_0 = p_0 + (\varepsilon_- + \varepsilon_+), \qquad \widetilde{p}_+ = p_+ - \varepsilon.$$

For sufficiently small ε_- and ε_+, $\widetilde{\mathsf{Q}} = \{\widetilde{p}_-, \widetilde{p}_0, \widetilde{p}_+\}$ is a probability measure and we must show that we can choose positive coefficients ε_- and ε_+ such that $\widetilde{\mathsf{E}}\xi = 0$, i.e.,

$$x_-p_- + x_0p_0 + x_+p_+ = (x_-p_- + x_0p_0 + x_+p_+) - (\varepsilon_-x_- + (\varepsilon_- + \varepsilon_+)x_0 - \varepsilon_+x_+) = 0.$$

Since $\mathsf{E}\xi = x_-p_- + x_0p_0 + x_+p_+ = 0$, these positive coefficients ε_- and ε_+ must satisfy the equality

$$\frac{\varepsilon_+}{\varepsilon_-} = \frac{x_0 - x_-}{x_+ - x_0}.$$

We set $\lambda = \dfrac{x_0 - x_-}{x_+ - x_0}$ (>0). Then it is clear that choosing sufficiently small ε_- first and setting $\varepsilon_+ = \lambda\varepsilon_-$ after that we can achieve the inequalities $\widetilde{p}_- > 0$, $\widetilde{p}_0 > 0$, and $\widetilde{p}_+ > 0$.

Hence $\widetilde{\mathsf{Q}} = \{\widetilde{p}_-, \widetilde{p}_0, \widetilde{p}_+\}$ is a probability measure, $\widetilde{\mathsf{Q}} \sim \mathsf{Q}$, $\widetilde{\mathsf{Q}} \neq \mathsf{Q}$, and $\widetilde{\mathsf{E}}\xi = 0$, which is again in contradiction with the uniqueness of the martingale measure Q, so that all the *three* masses p_-, p_0, and p_+ of the distribution Q cannot be positive.

This construction is easy to transfer to the case when a purely discrete martingale measure Q is concentrated at a countable set $\{x_i, i = 0, \pm 1, \pm 2, \dots\}$, $\cdots < x_{-2} < x_{-1} < x_0 < x_1 < x_2 < \cdots$, with respective probabilities $\{p_i, i = 0, \pm 1, \pm 2, \dots\}$.

If the set $\{x_i, i = 0, \pm 1, \pm 2, \dots\}$ contains the origin, say, $x_0 = 0$, then we must set

$$\widetilde{p}_i = \frac{p_i}{2}, \qquad i \neq 0,$$

and

$$\widetilde{p}_0 = \frac{1 - \sum_{i \neq 0} p_i}{2}.$$

Then $\sum\limits_i \widetilde{p}_i = 1$ and $\widetilde{\mathsf{E}}\xi = \sum\limits_i x_i \widetilde{p}_i = \frac{1}{2} \sum\limits_i x_i p_i = 0$.

The measure $\widetilde{\mathsf{Q}} = \{\widetilde{p}_i, i = 0, \pm 1, \pm 2, \dots\}$ is a probability measure, $\widetilde{\mathsf{Q}} \sim \mathsf{Q}$, $\widetilde{\mathsf{Q}} \neq \mathsf{Q}$, and $\widetilde{\mathsf{E}}\xi = 0$, which is incompatible with our assumption that the martingale measure is unique.

Now let $\{x_i, i = 0, \pm 1, \pm 2, \dots\}$ be a set of non-negative points $x_i \neq 0$. We construct a new distribution $\widetilde{\mathsf{Q}} = \{\widetilde{p}_i, i = 0, \pm 1, \pm 2, \dots\}$, where $\widetilde{p}_i = p_i$ for $i = \pm 2, \pm 3, \dots$ and, as above,

$$\widetilde{p}_{-1} = p_{-1} - \varepsilon_{-1}, \qquad \widetilde{p}_{+1} = p_{+1} - \varepsilon_{+1}, \qquad \widetilde{p}_0 = p_0 + (\varepsilon_{-1} + \varepsilon_{+1}).$$

Then

$$\widetilde{\mathsf{E}}\xi = \mathsf{E}\xi - \varepsilon_{-1}x_{-1} + (\varepsilon_{-1} + \varepsilon_{+1})x_0 - \varepsilon_{+1}x_{+1} = \varepsilon_+(x_0 - x_{+1}) + \varepsilon_{-1}(x_0 - x_{-1}),$$

and the same choice of ε_{-1} and ε_{+1} as in the above case of three points (x_-, x_0, x_+) brings us to a new martingale measure $\widetilde{\mathsf{Q}}$ distinct from Q but equivalent to it, which contradicts the assumption of the uniqueness of the martingale measure Q.

In a similar way we can consider cases where Q has absolutely continuous or singular components.

2. We now turn to the proof of implication $\{3\}$. We claim that the uniqueness of the martingale measure means that the σ-algebras $\mathscr{F}_n$ are generated by the prices S:

$$\mathscr{F}_n = \mathscr{F}_n^S \equiv \sigma(S_0, \dots, S_n), \qquad n \leqslant N.$$

We shall proceed by induction. (Note that the σ-algebras $\mathscr{F}_0$ and $\mathscr{F}_0^S$ are the same since, by assumption, $\mathscr{F}_0 = \{\varnothing, \Omega\}$ and S_0 is a non-random variable.)

Let $(\Omega, \mathscr{F}, (\mathscr{F}_n), \mathsf{P})_{n \leqslant N}$ be a filtered probability space and let $S = (S_n, \mathscr{F}_n, \mathsf{P})_{n \leqslant N}$ be a sequence of (stock) prices with $S_n = (S_n^1, \dots, S_n^d)$. To avoid additional notation we shall assume that P is itself a martingale measure.

Assuming that $\mathscr{F}_{n-1} = \mathscr{F}^S_{n-1}$ we consider a set $A \in \mathscr{F}_n$. Let

$$z = 1 + \frac{1}{2}\big(I_A - \mathsf{E}(I_A \,|\, \mathscr{F}^S_n)\big). \tag{8}$$

Clearly, $\frac{1}{2} \leqslant z \leqslant \frac{3}{2}$ and $\mathsf{E}z = 1$. Hence the measure P' with $\mathsf{P}'(d\omega) = z(\omega)\,\mathsf{P}(d\omega)$ is a probability measure and $\mathsf{P}' \sim \mathsf{P}$. Let $z_i = \mathsf{E}(z \,|\, \mathscr{F}_i)$. By Bayes's formula (see (4) in § 3a),

$$\mathsf{E}'(\Delta S_i \,|\, \mathscr{F}_{i-1}) = \mathsf{E}\biggl(\frac{z_i}{z_{i-1}}\Delta S_i \,\Big|\, \mathscr{F}_{i-1}\biggr). \tag{9}$$

In view of our assumption that $\mathscr{F}_{n-1} = \mathscr{F}^S_{n-1}$, it follows from (8) that $\mathsf{E}(z \,|\, \mathscr{F}_{n-1}) = 1$. At the same time z is a $\mathscr{F}_n$-measurable function. Hence $\dfrac{z_i}{z_{i-1}} = 1$ for $i \neq n$, so that $\mathsf{E}'(\Delta S_i \,|\, \mathscr{F}_{i-1}) = 0$ for all $i \neq n$.

Since $\dfrac{z_n}{z_{n-1}} = z$, $\mathsf{E}(z \,|\, \mathscr{F}^S_{n-1}) = 1$, and ΔS_n are $\mathscr{F}^S_n$-measurable, it follows by (12) that

$$\begin{aligned}
\mathsf{E}'(\Delta S_n \,|\, \mathscr{F}_{n-1}) &= \mathsf{E}(z\Delta S_n \,|\, \mathscr{F}_{n-1}) \\
&= \mathsf{E}(z\Delta S_n \,|\, \mathscr{F}^S_{n-1}) = \mathsf{E}\big(\mathsf{E}(z\Delta S_n \,|\, \mathscr{F}^S_n) \,|\, \mathscr{F}^S_{n-1}\big) \\
&= \mathsf{E}\big(\Delta S_n \mathsf{E}(z \,|\, \mathscr{F}^S_n) \,|\, \mathscr{F}^S_{n-1}\big) = \mathsf{E}(\Delta S_n \,|\, \mathscr{F}^S_{n-1}) = 0,
\end{aligned}$$

where we also use the equality $\mathsf{E}(z \,|\, \mathscr{F}^S_n) = 1$ holding by (8).

Hence the sequence of prices $(S_n, \mathscr{F}_n)_{n \leqslant N}$ is a P'-martingale.

Thus, the assumption that P is a *unique* martingale measure brings us to the equality $z = 1$ (P-a.s.), so that for each $A \in \mathscr{F}_n$ we obtain

$$I_A = \mathsf{E}(I_A | \mathscr{F}^S_n) \quad (\mathsf{P}\text{-a.s.})$$

by (11). Hence, $\mathscr{F}_n = \mathscr{F}^S_n$ up to sets of P-measure zero.

Using induction on n we can verify these relations for all $n \leqslant N$, which proves implication (3).

3. Thus, the uniqueness of the martingale measure P ensures the conclusions of the implications {2} and {3}. Combined, they mean 'S-representability', which, in turn, delivers the completeness of the market (by the Lemma in § 4a). This done, we have established the sufficiency part of Theorem B (for $d = 1$).

Remark 1. It is worth noting that, as shows the above proof of Theorem B, a discrete-time complete arbitrage-free market (with $N < \infty$ and $d = 1$) is in fact *discrete also in the phase variable* in the following sense: the σ-algebra $\mathscr{F}_N$ is *purely atomic* (up to P-measure zero) and contains at most 2^N atoms. This is an immediate consequence of a 'conditional two-pointedness'. (In the case of arbitrary $d < \infty$ the number of atoms in $\mathscr{F}_N$ is at most $(d+1)^N$.)

Remark 2. The fact that for $N < \infty$ and $d < \infty$ the σ-algebra $\mathscr{F}_N$ corresponding to a complete arbitrage-free market has at most $(d+1)^N$ elements shows that *completeness* and *perfection* mean the same for such markets.

§ 4f. Extended Version of the Second Fundamental Theorem

1. The above proof of Theorem B relates to the case of $d = 1$. (We have explicitly used this assumption in our proof of implications {2} and {4}.) In the general case of $d \geqslant 1$, it seems appropriate to present an extended version of this theorem including the equivalence of the completeness and the uniqueness of a martingale measure and several other equivalent characterizations.

First, let us introduce our notation.

We set $\overline{S}_n = \dfrac{S_n}{B_n}$ (the discounted prices),

$$\mathsf{Q}_n(\,\cdot\,;\omega) = \mathsf{P}(\Delta S_n \in \cdot \mid \mathscr{F}_{n-1})(\omega), \quad \overline{\mathsf{Q}}_n(\,\cdot\,;\omega) = \mathsf{P}(\Delta \overline{S}_n \in \cdot \mid \mathscr{F}_{n-1})(\omega).$$

We recall that vectors $a_1, \dots, a_k$ with $a_i \in \mathbb{R}^d$, $2 \leqslant k \leqslant d+1$, are said to be *affinely independent* if there exists $i \in \{1, \dots, k\}$ such that the $k-1$ vectors $(a_j - a_i)$, $j = 1, \dots, k$, $j \neq i$, are *linearly independent*. If this property holds for some $i \in \{1, \dots, k\}$, then it also holds for each integer $i = 1, \dots, k$. Note that this property of affine independence of d-dimensional vectors $a_1, \dots, a_k$ is equivalent to the fact that the minimal affine plane containing $a_1, \dots, a_k$ has dimension $k-1$.

THEOREM B* (an extended version of the *Second fundamental theorem*). *Assume that we have an arbitrage-free (B, S)-market with $B = (B_n)_{0 \leqslant n \leqslant N}$ and $S = (S_n)_{0 \leqslant n \leqslant N}$, $S_n = (S_n^1, \dots, S_n^d)$, where the $B_n > 0$ are $\mathscr{F}_{n-1}$-measurable, the $S_n^i \geqslant 0$ are $\mathscr{F}_n$-measurable, $N < \infty$, and $d < \infty$. Then the following conditions are equivalent.*

a) *The market is complete.*
b) *The market is perfect.*
c) *The set of martingale measure $\mathscr{P}(\mathsf{P})$ contains a unique measure.*
d) *The set of local martingale measures $\mathscr{P}_{\mathrm{loc}}(\mathsf{P})$ contains a unique measure.*
e) *There exists a measure P' in $\mathscr{P}_{\mathrm{loc}}(\mathsf{P})$ such that each martingale $M = (M_n, \mathscr{F}_n, \mathsf{P}')_{n \leqslant N}$ has an 'S-representation'*

$$M_n = M_0 + \sum_{i=1}^{n} \gamma_i \Delta \overline{S}_i, \qquad n \leqslant N,$$

with $\mathscr{F}_{i-1}$-measurable γ_i.
f) *$\mathscr{F}_n = \sigma(S_1, \dots, S_n)$ up to sets of P-measure zero, and there exist $(d+1)$-predictable $\mathbb{R}^d$-valued processes $(a_{1,n}, \dots, a_{d+1,n})$, $1 \leqslant n \leqslant N$, such that their values are affinely independent (for all n and ω) and the supports of the measures $\mathsf{Q}_n(\,\cdot\,;\omega)$ lie (P-a.s.) in the set $\{a_{1,n}(\omega), \dots, a_{d+1,n}(\omega)\}$.*
g) *$\mathscr{F}_n = \sigma(\overline{S}_1, \dots, \overline{S}_n)$ up to sets of P-measure zero and there exist $(d+1)$-predictable $\mathbb{R}^d$-valued processes $(\overline{a}_{1,n}, \dots, \overline{a}_{d+1,n})$, $1 \leqslant n \leqslant N$, such that their values are affinely independent (for all n and ω) and the supports of the measures $\overline{\mathsf{Q}}_n(\,\cdot\,;\omega)$ lie (P-a.s.) in the set $\{\overline{a}_{1,n}(\omega), \dots, \overline{a}_{d+1,n}(\omega)\}$.*

If these conditions hold, then the σ-algebra $\mathscr{F}_N$ is purely atomic (with respect to P) and consists of at most $(d+1)^N$ atoms.

Proof. For $d = 1$ this was explained in the preceding sections. In the general case of $d \geqslant 1$ the corresponding proof can be found in [251]. We refer the reader there for the technical detail related to the fact that the prices $S = (S^1, \dots, S^d)$ are vector-valued; we concentrate ourselves on outlining the proof and pointing out the differences between the cases $d = 1$ and $d > 1$.

First, we note that the equivalence of f) and g) is a simple consequence of the relation

$$\overline{a}_{i,n} = \frac{a_{i,n}}{B_n} + S_{n-1}\left(\frac{1}{B_n} - \frac{1}{B_{n-1}}\right)$$

between the $\overline{a}_{i,n}$ and the $a_{i,n}$.

Further, we clearly have the implication b)$\Longrightarrow$a) and by Theorem A* (§ 2e), d)$\Longleftrightarrow$c).

Hence to prove the theorem we must prove that

$$\begin{aligned} \text{a)} &\Longrightarrow \text{d)},\\ \text{c)} &\Longrightarrow \text{g)},\\ \text{g)} &\Longrightarrow \text{b)},\\ \text{a)} + \text{g)} &\Longrightarrow \text{e)},\\ \text{e)} &\Longrightarrow \text{a)}. \end{aligned}$$

The implication a)$\Longrightarrow$d) can be established in the same way as for $d = 1$ (see § 4a.2), where, in place of the martingale measures P_i, $i = 1, 2$, one must consider local martingale measures.

As regards the implication c)$\Longrightarrow$g), we already proved the equality $\mathscr{F}_n = \sigma(\overline{S}_1, \dots, \overline{S}_n)$ in § 4e.2, in our proof of implication {3} from § 4a.2. (The corresponding proof is in fact valid for each $d \geqslant 1$.)

The most tedious part of the proof of the implication c)$\Longrightarrow$g) is to describe the structure of the supports of the measures $\overline{\mathsf{Q}}_n(\,\cdot\,;\omega)$. For $d = 1$ these supports were 'two-point'. In the general case of $d \geqslant 1$ these supports consist of at most $d + 1$ points (in $\mathbb{R}^d$). This part of the proof is exposed at length in [251] and we omit it here. (Conceptually, the proof is the same as for $d = 1$ and proceeds as follows. Assume that P is itself a martingale measure. If the support of $\overline{\mathsf{Q}}_n(\,\cdot\,;\omega)$ contains more than $d + 1$ points, then, using the idea of 'mass pumping' again, we can construct a new measure P' by the formula $\mathsf{P}'(d\omega) = z(\omega)\,\mathsf{P}(d\omega)$. For a suitable choice of the $\mathscr{F}_N$-measurable function $z(\omega)$, P' is a martingale measure, $\mathsf{P}' \sim \mathsf{P}$, and $\mathsf{P}' \neq \mathsf{P}$. This contradicts, however, our assumption of the *uniqueness* of a martingale measure. In a similar way we can also prove that the $\mathbb{R}^d$-valued variables $\overline{a}_{1,n}, \dots, \overline{a}_{d+1,n}$ are affinely independent.)

We consider now the implication g)$\Longrightarrow$b). Let f_N be a $\mathscr{F}_N$-measurable random variable and assume that the original measure P is itself a martingale measure. Then it follows from g) that, in fact, f_N is a random variable with finitely many possible values.

We claim that f_N can be represented as follows:

$$f_N = x + \sum_{k=1}^{N} \gamma_k \Delta \overline{S}_k. \tag{1}$$

The sequence $\overline{X}_n \equiv x + \sum_{i=1}^{n} \gamma_i \Delta \overline{S}_i$, $n \leqslant N$, is a P-martingale, therefore the relations $x = \mathsf{E} f_N$ and

$$\gamma_n \Delta \overline{S}_n = \mathsf{E}(f_N \mid \mathscr{F}_n) - \mathsf{E}(f_N \mid \mathscr{F}_{n-1}) \tag{2}$$

must be satisfied.

Hence, to obtain the representation (1) we can set $x = \mathsf{E} f_N$ and then show (as for $d = 1$) that using condition g) we can construct $\mathscr{F}_{n-1}$-measurable functions γ_n with required property (2). (See [251] for greater detail.)

We now turn to the implication a) + g) $\Longrightarrow$ e). By g) the σ-algebra $\mathscr{F}_N$ is purely atomic. Hence all $\mathscr{F}_N$-measurable random variables can take only finitely many values and are therefore bounded.

Let $\mathsf{P}' \in \mathscr{P}_{\text{loc}}(\mathsf{P})$ and let $M = (M_n, \mathscr{F}_n, \mathsf{P}')_{n \leqslant N}$ be a martingale. By (a) there exist $x \in \mathbb{R}$ and a predictable process $\gamma = (\gamma_n)$ such that $M_N = x + \sum_{i=1}^{N} \gamma_i \Delta \overline{S}_i$.

The sequence $M' = (M'_n, \mathscr{F}_n, \mathsf{P}')_{n \leqslant N}$ of variables $M'_n = x + \sum_{i=1}^{n} \gamma_i \Delta \overline{S}_i$ is a local P'-martingale and, therefore, a martingale because all its elements are bounded. Since $M_N = M'_N$, the martingales M and M' coincide (P'-a.s.), which gives us assertion e).

Finally, to prove the implication e)$\Longrightarrow$a) it suffices to make the following observation (cf. Lemma § 4a).

Assume that each martingale $M = (M_n, \mathscr{F}_n, \mathsf{P}')$ with $\mathsf{P}' \in \mathscr{P}_{\text{loc}}(\mathsf{P})$ admits an 'S-representation' $M_n = M_0 + \sum_{i=1}^{n} \gamma_i \Delta \overline{S}_i$.

Let f_N be a $\mathscr{F}_N$-measurable bounded function. We consider the martingale $M_n = \mathsf{E}'(f_N \mid \mathscr{F}_n)$, $n \leqslant N$, where E' is averaging with respect to P'. By assumption,

$$f_N = M_N = M_0 + \sum_{i=1}^{N} \gamma_i \Delta \overline{S}_i \quad (\mathsf{P}'\text{-a.s.}).$$

Hence we can represent f_N as a sum:

$$f_N = x + \sum_{i=1}^{N} \gamma_i \Delta \overline{S}_i \quad (\mathsf{P}\text{-a.s.}),$$

where $x = M_0$ and $\gamma = (\gamma_i)_{i \leqslant N}$ is a predictable sequence, which means precisely that the market is complete.

These arguments complete our discussion of the above-described implications required in Theorem B*.

2. We present now several examples illustrating both Theorem B* and Theorem A*.

EXAMPLE 1 ($d = 1$). In the framework of the *CRR*-model (see § 4d) with $B_n \equiv 1$, $n \leqslant N$, we assume that $(\rho_n)_{n \leqslant N}$ is a sequence of independent identically distributed random variables taking two values, a and b, $a < b$.

Since $\Delta S_n = S_{n-1}\rho_n$, it follows that $\Delta S_n = S_{n-1}a$ or $\Delta S_n = S_{n-1}b$. By Theorem A*, in order that the market be *arbitrage-free* the interval (a, b) must contain the origin. Hence $a < 0 < b$. To ensure that the prices S are positive we must also set $a > -1$.

In this case $\Delta S_n = S_{n-1}\rho_n$, so that ΔS_n can take two values: $S_{n-1}b$ ('upward price motion') and $S_{n-1}a$ ('downward motion'). Hence the supports of the conditional distributions $\mathsf{Q}_n(\,\cdot\,;\omega)$ consist of two points: $S_{n-1}(\omega)a$ and $S_{n-1}(\omega)b$, while the 'price tree' $(S_0, S_1, S_2, \dots)$ (see Fig. 56) has the 'homogeneous Markov structure': if $(S_0, S_1, \dots, S_{n-1})$ is some realization of the price process, then the next transition brings $S_n = S_{n-1}B$ with probability $p = \mathsf{P}(\rho_n = b)$ and $S_n = S_{n-1}A$ with probability $q = \mathsf{P}(\rho_n = a)$, where $B = 1 + b$ and $A = 1 + a$.

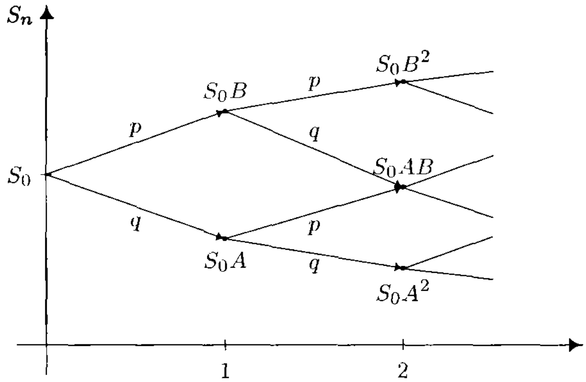


FIGURE 56. 'Price tree' $(S_0, S_1, S_2, \dots)$ in the *CRR*-model

If $-1 < a < 0 < b$, then there exists a unique martingale measure, so that the corresponding (B, S)-market is arbitrage-free and complete.

By Theorem B* we obtain that for $d = 1$ *all* complete arbitrage-free markets have fairly similar 'dyadic' branching structure of the prices.

Namely, for fixed 'history' $(S_0, S_1, \dots, S_{n-1})$ we have $S_n = S_{n-1}(1 + \rho_n)$, where the variables $\rho_n = \rho_n(S_0, S_1, \dots, S_{n-1})$ can take only *two* values, $a_n = a_n(S_0, S_1, \dots, S_{n-1})$ and $b_n = b_n(S_0, S_1, \dots, S_{n-1})$.

In the above model of Cox–Ross–Rubinstein the variables a_n and b_n are constant ($a_n = a$ and $b_n = b$). In general, they depend on the price history, but again, to obtain a complete arbitrage-free market with positive prices, we must set $-1 < a_n < 0 < b_n$.

EXAMPLE 2 ($d = 2$, $N = 1$). Let $B_0 = B_1 = 1$ and let $S = (S^1, S^2)$ be the prices of two kinds of stock, where $S_0^1 = S_0^2 = 2$. We consider a single-step model ($N = 1$); let

$$\Delta S_1 = \begin{pmatrix} \Delta S_1^1 \\ \Delta S_1^2 \end{pmatrix}$$

be the vector of price increments with $\Delta S_1^i = S_1^i - S_0^i = S_1^i - 2$, $i = 1, 2$.

In accordance with Theorem B*, to make the corresponding arbitrage-free market complete we need that the support of the measure $\mathsf{P}(\Delta S_1 \in \cdot)$ consist of three points in the plane,

$$\begin{pmatrix} a_1 \\ b_1 \end{pmatrix}, \qquad \begin{pmatrix} a_2 \\ b_2 \end{pmatrix}, \qquad \begin{pmatrix} a_3 \\ b_3 \end{pmatrix},$$

and the three corresponding vectors in $\mathbb{R}^2$ must be affinely independent. As already mentioned, this is equivalent to the linear independence of the vectors $\begin{pmatrix} a_1 - a_3 \\ b_1 - b_3 \end{pmatrix}$ and $\begin{pmatrix} a_2 - a_3 \\ b_2 - b_3 \end{pmatrix}$.

For instance, assume that the probability of each of the vectors

$$\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \qquad \begin{pmatrix} -1 \\ 1 \end{pmatrix}, \qquad \begin{pmatrix} 0 \\ -1 \end{pmatrix}$$

is equal to $\frac{1}{3}$. These vectors are affinely independent, and the measure assigning them probabilities $\frac{1}{4}$, $\frac{1}{4}$, and $\frac{1}{2}$, respectively, is martingale.

Chapter VI. Theory of Pricing in Stochastic Financial Models. Discrete Time

1. European Hedge Pricing on Arbitrage-Free Markets

§ 1a. Risks and Their Reduction

1. H. Markowitz's theory ([332], 1952), as expressed in his *mean-variance analysis* (see Chapter I, § 2b), suggests an approach to the pricing of investment risks and the reduction of their *nonsystematic* component that is based on the idea of *diversification* in the selection of an (optimal) investment portfolio.

Several other optimization problems arising in financial theory can, in view of the 'environmental uncertainties', be (as in H. Markowitz's case) ranked among the problems of *stochastic optimization*. In the first line, it should be mentioned that finance brings forward a series of nontraditional, nonstandard optimization problems relating to *hedging* (see Chapter V, § 1b about the notion of 'hedging'). They are nonstandard in the following sense: *optimal* hedging as a *control* must deliver certain properties with *probability one*, rather than, e.g., *on the average*, as is usual in stochastic optimization. (As regards problems involving the *mean square* criterion, see § 1d below.)

In what follows we put an accent on the discussion of hedging as a method of dynamical control of an investment portfolio. It should be noted that this method is crucial for pricing such (derivative) financial instruments as, e.g., options (see subsections 4 and 5). We can say more: it is in connection with the pricing of *option* contracts that the importance of hedging as a protection instrument has been understood and its basic methods have been developed.

2. We recall that we already encountered *hedging* in § 1b, Chapter 5, where in the simplest case of the *single-step* model we deduced formulas for both initial capital sufficient for the desired result and optimal (hedging) portfolio itself.

Explicit formulas for hedges are of considerable interest also in *several-step problems*, in which an investor seeks capital levels at a certain (fixed in advance) instant N that, with probability 1 (or, more generally, with certain positive probability), are not lower than the values at these instants of some fixed (random, in general) performance functional.

Problems of this kind have a direct relation to the pricing of *European options*; this connection is based on the following remarkably simple and seminal idea of F. Black and M. Scholes [44] and R. Merton [345]: in complete arbitrage-free markets

the dynamics of option prices must be replicated by the dynamics of the value of the optimal hedging strategy in the corresponding investment problem.

In the case of *American options*, besides the hedging of the writer, which realizes his 'control' strategy, we have one new 'optimization element'.

In fact, the buyer of a European option is *inert*: he is not engaged in financial activity and waits for the maturity date N of the option. American-type options are another matter. Here the buyer is an *acting* character in the trade: the contract allows him to choose the *instant* of exercising the option on his own (of course, within certain specified limits).

In selecting the corresponding hedging portfolio the writer must keep in mind the buyer's freedom to exercise the option at any time. Clearly, the *contradicting interests* of the writer and the buyer give rise to optimization problems of the *minimax* nature.

The present section (§ § 1a–1d) is devoted to *European hedging*. This name is inspired by an analogy with European options and means that we are discussing hedging against claims due at some moment that was *fixed* in advance. We consider American hedging in the next section (see, in particular, § 2c for the corresponding definitions).

3. As already mentioned, the buyer's control in American options reduces to the choice of the instant of exercising the contract, or, as it is often called, the *stopping time*.

Here, if, e.g., $f = (f_0, f_1, \dots, f_N)$ is a system of pay-off functions ($f_i = f_i(\omega)$, $i = 0, 1, \dots, N$) and the buyer chooses a stopping time $\tau = \tau(\omega)$, then his returns are $f_\tau = f_{\tau(\omega)}(\omega)$.

We represent now f_τ as follows:

$$f_\tau = f_0 + \sum_{k=1}^{\tau} \Delta f_k = f_0 + \sum_{k=1}^{N} I(k \leqslant \tau)\Delta f_k. \tag{1}$$

Note that the event $\{k \leqslant \tau\}$ belongs to $\mathscr{F}_{k-1}$. Hence (1) can be rewritten as follows:

$$f_\tau = f_0 + \sum_{k=1}^{N} \alpha_k \Delta f_k, \tag{2}$$

where $\alpha_k = I(k \leqslant \tau)$ is a $\mathscr{F}_{k-1}$-measurable random variable.

We can express this otherwise: the terms of an American option contract allow the buyer to exert only a predictable control $\alpha = (\alpha_k)_{k \leqslant N}$ of the form $\alpha_k = I(k \leqslant \tau)$.

In principle, one could imagine contracts allowing *other* forms of (predictable) control $\alpha = (\alpha_k)_{k \leqslant N}$ from the buyer's side. One possible example is a 'Passport option' (see, e.g., [6]) with pay-off function

$$f_N(\alpha) = \left[\sum_{k=1}^{N} \alpha_k \Delta S_k\right]^+, \tag{3}$$

where $|\alpha_k| \leqslant 1$ and $S = (S_k)_{k \leqslant N}$ is the stock price.

It is clear in this case that the writer of the contract must select a (hedging) portfolio $\pi = \pi(\alpha(\omega), \omega)$ such that for each admissible buyer's control $\alpha = \alpha(\omega)$ the value $X_N^\pi(\alpha(\omega), \omega)$ at the terminal date N is at least $f_N(\alpha(\omega))$ (P-a.s.).

4. Hedging (by an investor or the writer of an option) and the control exercised by the option buyer, say, are the two main 'optimization' components that should be reckoned with in derivatives pricing on *complete* markets.

For *incomplete* markets, we must add also the third component determined by 'natural factors'.

This means the following. There can exist only *one* martingale measure in complete arbitrage-free markets. However, in incomplete arbitrage-free financial markets '*Nature*' provides one with a whole *range* of martingale measures and, therefore, with *distinct* versions of the formalization of the market.

Which measures and formalizations are in fact operational remains unknown to the writer and the buyer of an option. Hence both writer and buyer, unless they have some additional arguments, must plan their strategy (of hedging, or the choice of the stopping time, etc.) bearing in mind the 'most unfavorable combination of natural factors possible'.

We shall express this formally by considering in many relations the least upper bounds over the class of all martingale measures that can describe potential 'environment'.

§ 1b. Main Hedge Pricing Formula. Complete Markets

1. We shall consider a complete arbitrage-free (B, S)-market with $N < \infty$ and $d < \infty$ (see the general scheme in Chapter V, § 2b). By assertion (f) of the extended version of the *Second fundamental theorem* (Chapter V, § 2e), such a *discrete-time* market is also *discrete with respect to the phase variable*, and all the $\mathscr{F}_N$-measurable random variables under consideration can take only finitely many values because the σ-algebra $\mathscr{F}_N$ consists of at most $(d+1)^N$ atoms. Thus, there can be no problems with integration in this case, and the concepts of complete market and perfect market are equivalent.

DEFINITION. The *price of perfect European hedging* of an ($\mathscr{F}_N$-measurable contingent claim f_N) is the quantity

$$\mathbb{C}(f_N; \mathsf{P}) = \inf\{x \colon \exists\, \pi \text{ with } X_0^\pi = x \text{ with } X_N^\pi = f_N \text{ (P-a.s.)}\} \tag{1}$$

(cf. Chapter V, § 1b).

Since the market in question is *arbitrage-free* and *complete* by assumption,

1) there *exists* a martingale measure $\widetilde{\mathsf{P}}$ equivalent to P such that the sequence $\left(\dfrac{S_n}{B_n}\right)_{n \leqslant N}$ is a martingale (the *First fundamental theorem*)

and

2) this measure is *unique*, and each contingent claim f_N can be replicated, i.e., there *exists* a ('perfect') hedge π such that $X_N^\pi = f_N$ (the *Second fundamental theorem*).

Hence if π is a *perfect* (x, f_N)-hedge, i.e., $X_0^\pi = x$ and $X_N^\pi = f_N$ (P-a.s.), then (see formula (18) in Chapter V, § 1a)

$$\frac{f_N}{B_N} = \frac{X_N^\pi}{B_N} = \frac{x}{B_0} + \sum_{k=1}^{n} \gamma_k \Delta\left(\frac{S_k}{B_k}\right), \tag{2}$$

and therefore

$$\widetilde{\mathsf{E}} \frac{f_N}{B_N} = \frac{x}{B_0},$$

i.e.,

$$x = B_0\, \widetilde{\mathsf{E}} \frac{f_N}{B_N}. \tag{3}$$

We note that the right-hand side of (3) is *independent* of the structure of the (x, f_N)-hedge π in question. In other words, if π' is another hedge, then the initial prices x and x' are the same.

Hence, we have the following result.

THEOREM 1 (*Main formula of the price of perfect European hedging in complete markets*). *The price $\mathbb{C}(f_N; \mathsf{P})$ of perfect hedging in a complete arbitrage-free market is described by the formula*

$$\boxed{\mathbb{C}(f_N; \mathsf{P}) = B_0\, \widetilde{\mathsf{E}} \frac{f_N}{B_N}.} \tag{4}$$

2. In hedging one must know not only the price $\mathbb{C}(f_N;\mathsf{P})$, but also the composition of the *portfolio* bringing about a perfect hedge. A standard method here is as follows (cf. Chapter V, § 4a).

We construct the martingale $M = (M_n, \mathscr{F}_n, \widetilde{\mathsf{P}})_{n\leqslant N}$ with $M_n = \widetilde{\mathsf{E}}\Big(\dfrac{f_N}{B_N}\;\Big|\;\mathscr{F}_n\Big)$. Since our market is *complete*, it follows by the *Second fundamental theorem* (or the lemma in Chapter V, § 4b) that M has an '$\dfrac{S}{B}$*-representation*'

$$M_n = M_0 + \sum_{k=1}^{n} \gamma_k \Delta\Big(\frac{S_k}{B_k}\Big) \tag{5}$$

with $\mathscr{F}_{k-1}$-measurable γ_k.

We set $\pi^* = (\beta^*, \gamma^*)$, where γ^* is equal to $\gamma = (\gamma_n)$ in (5) and $\beta_n^* = M_n - \dfrac{\gamma_n S_n}{B_n}$. It is easy to verify that this is a *self-financing* portfolio (see nevertheless the proof of the lemma in Chapter V, § 4b). Further, by construction,

$$\frac{X_0^{\pi^*}}{B_0} = M_0 \tag{6}$$

and

$$\Delta\Big(\frac{X_n^{\pi^*}}{B_n}\Big) = \gamma_n^* \Delta\Big(\frac{S_n}{B_n}\Big) = \gamma_n \Delta\Big(\frac{S_n}{B_n}\Big) = \Delta M_n.$$

Hence, for all $0 \leqslant n \leqslant N$ we have

$$\frac{X_n^{\pi^*}}{B_n} = M_n = \widetilde{\mathsf{E}}\Big(\frac{f_N}{B_N}\;\Big|\;\mathscr{F}_n\Big), \tag{7}$$

and, in particular,

$$X_N^{\pi^*} = f_N \quad (\widetilde{\mathsf{P}}\text{- and } \mathsf{P}\text{-a.s.}).$$

Thus, the portfolio π^* constructed on the basis of the '$\dfrac{S}{B}$-representation' is a perfect hedge (for f_N).

We can sum up the above results as follows.

THEOREM 2 ('Main formulas for a perfect hedge and its value'). *In an arbitrary arbitrage-free complete market there exists a self-financing perfect hedge* $\pi^* = (\beta^*, \gamma^*)$ *with initial capital*

$$X_0^{\pi^*} = \mathbb{C}(f_N;\mathsf{P}) \quad \Big(= B_0\,\widetilde{\mathsf{E}}\frac{f_N}{B_N}\Big),$$

that replicates f_N faithfully:

$$X_N^{\pi^*} = f_N \quad (\mathsf{P}\text{-a.s.}).$$

The dynamics of the capital $X_n^{\pi^*}$ is described by the formulas

$$X_n^{\pi^*} = B_n \widetilde{\mathsf{E}}\left(\frac{f_N}{B_N} \,\Big|\, \mathscr{F}_n \right), \qquad 0 \leqslant n \leqslant N;$$

the components $\gamma^* = (\gamma_n^*)$ are the same as in the '$\dfrac{S}{B}$-representation'

$$\widetilde{\mathsf{E}}\left(\frac{f_N}{B_N} \,\Big|\, \mathscr{F}_n \right) = \widetilde{\mathsf{E}}\frac{f_N}{B_N} + \sum_{k=1}^{n} \gamma_k^* \Delta\left(\frac{S_k}{B_k} \right), \qquad 1 \leqslant n \leqslant N,$$

and the components $\beta^* = (\beta_n^*)$ can be defined from the condition

$$X_n^{\pi^*} = \beta_n^* B_n + \gamma_n^* S_n.$$

3. We consider now the issue of hedge pricing in a somewhat more general framework, assuming that, in place of a single function f_N, we have a *sequence* $f_0, f_1, \dots, f_N$ of pay-off functions and f_i is $\mathscr{F}_i$-measurable, $0 \leqslant i \leqslant N$.

Let $\tau = \tau(\omega)$ be a *fixed* Markov time ranging in the set $\{0, 1, \dots, N\}$ and let f_τ be the terminal pay-off function constructed for this τ and $f_0, f_1, \dots, f_N$.

THEOREM 3. *If an arbitrage-free (B, S)-market is N-complete, then it is also τ-complete, that is, there exist a self-financing portfolio π and an initial capital x such that $X_0^\pi = x$ and $X_\tau^\pi = f_\tau$ (P-a.s.).*

Proof. This is simple: we construct a new pay-off function $f_N^* = f_{\tau \wedge N}$; then the perfect hedge π^* for f_N^* is simultaneously a perfect hedge for the initial pay-off function f_τ.

The corresponding price

$$\mathbb{C}(f_\tau; \mathsf{P}) = \min\bigl\{x \colon \exists\, \pi \text{ with } X_0^\pi = x \text{ and } X_\tau^\pi = f_\tau \ (\mathsf{P}\text{-a.s.})\bigr\}$$

can be evaluated by the formula

$$\mathbb{C}(f_\tau; \mathsf{P}) = B_0\, \widetilde{\mathsf{E}} \frac{f_\tau}{B_\tau}. \tag{8}$$

4. One can put the following question in connection with 'main formula' (4).

Assume that we consider a *complete and arbitrage-free* $(\widetilde{B}, S)$-market and let $\widetilde{\mathsf{P}}$ be a martingale measure for the discounted prices $\dfrac{S}{\widetilde{B}}$. It will be useful to formulate the last property in the following equivalent way: the *vector process* $\left(\dfrac{\widetilde{B}}{\widetilde{B}}, \dfrac{S^1}{\widetilde{B}}, \dots, \dfrac{S^d}{\widetilde{B}}\right)$ $\left(\textit{i.e., the process } \left(1, \dfrac{S^1}{\widetilde{B}}, \dots, \dfrac{S^d}{\widetilde{B}}\right)\right)$ *is a* $\widetilde{\mathsf{P}}$*-martingale.*

We assume now that there exists another (positive) discounting process $\overline{B} = (\overline{B}_n)_{n \leqslant N}$ and a measure $\overline{\mathsf{P}}$ equivalent to the initial measure P such that the discounted process

$$\left(\frac{\widetilde{B}}{\overline{B}}, \frac{S^1}{\overline{B}}, \dots, \frac{S^d}{\overline{B}}\right)$$

is a $\overline{\mathsf{P}}$*-martingale.*

One would expect the value of the price $\mathbb{C}(f_N; \mathsf{P})$ defined by (1) to be *independent* of the choice of the corresponding pairs $(\widetilde{B}, \widetilde{\mathsf{P}})$ and $(\overline{B}, \overline{\mathsf{P}})$.

In fact, this is just the case: we have the equalities

$$\widetilde{B}_0\, \widetilde{\mathsf{E}} \frac{f_N}{\widetilde{B}_N} = \overline{B}_0\, \overline{\mathsf{E}} \frac{f_N}{\overline{B}_N}, \tag{9}$$

moreover, even the 'price processes'

$$\left(\widetilde{B}_n \widetilde{\mathsf{E}}\left(\frac{f_N}{\widetilde{B}_N} \,\Big|\, \mathscr{F}_n\right)\right)_{n \leqslant N} \quad \text{and} \quad \left(\overline{B}_n \overline{\mathsf{E}}\left(\frac{f_N}{\overline{B}_N} \,\Big|\, \mathscr{F}_n\right)\right)_{n \leqslant N} \tag{10}$$

are the same.

We now discuss the factors underlying this coincidence. To this end we assume that $\overline{\mathsf{E}} \dfrac{\widetilde{B}_N}{\overline{B}_N} = 1$ (which is not a serious constraint). Then we can introduce a new measure $\widehat{\mathsf{P}}$ (on $\mathscr{F}_N$) by setting

$$d\widehat{\mathsf{P}} = \widehat{Z}_N \, d\widetilde{\mathsf{P}},$$

where $\widehat{Z}_n = \overline{Z}_n \dfrac{\widetilde{B}_n}{\overline{B}_n}$, $\overline{Z}_n = \dfrac{d\overline{\mathsf{P}}_n}{d\widetilde{\mathsf{P}}_n}$, $\overline{\mathsf{P}}_n = (\overline{\mathsf{P}} \,|\, \mathscr{F}_n)$, and $\widetilde{\mathsf{P}}_n = (\widetilde{\mathsf{P}} \,|\, \mathscr{F}_n)$, $n \leqslant N$.

The measure $\widehat{\mathsf{P}}$ is a probability measure, and by Bayes's formula (see the lemma in Chapter V, § 3a)

$$\begin{aligned}
\widehat{\mathsf{E}}\left(\frac{S_N}{\widetilde{B}_N} \,\Big|\, \mathscr{F}_n\right) &= \frac{1}{\widehat{Z}_n}\, \widetilde{\mathsf{E}}\left(\frac{S_N}{\widetilde{B}_N}\, \widehat{Z}_N \,\Big|\, \mathscr{F}_n\right) \\
&= \frac{1}{\overline{Z}_n \cdot \frac{\widetilde{B}_n}{\overline{B}_n}}\, \widetilde{\mathsf{E}}\left(\frac{S_N}{\overline{B}_N} \cdot \frac{\overline{B}_N}{\widetilde{B}_N} \cdot \widehat{Z}_N \,\Big|\, \mathscr{F}_n\right) \\
&= \frac{1}{\overline{Z}_n \cdot \frac{\widetilde{B}_n}{\overline{B}_n}}\, \widetilde{\mathsf{E}}\left(\frac{S_N}{\overline{B}_N} \cdot \overline{Z}_N \,\Big|\, \mathscr{F}_n\right) = \frac{1}{\frac{\widetilde{B}_n}{\overline{B}_n}}\, \overline{\mathsf{E}}\left(\frac{S_N}{\overline{B}_N} \,\Big|\, \mathscr{F}_n\right) = \frac{S_n}{\widetilde{B}_n}.
\end{aligned}$$

because $\overline{\mathsf{E}}\left(\frac{S_N}{\overline{B}_N} \mid \mathscr{F}_n\right) = \frac{S_n}{\overline{B}_n}$.

Hence the sequence $\left(\frac{S_n}{\widetilde{B}_n}\right)_{n\leqslant N}$ is a martingale both with respect to $\widetilde{\mathsf{P}}$ and with respect to $\widehat{\mathsf{P}}$.

However, if the market in question is *complete*, then the martingale measure must be *unique*, and therefore $\widehat{\mathsf{P}} = \widetilde{\mathsf{P}}$, that is, $\widehat{Z}_n = 1$ for $n \leqslant N$, which, by the definition of the $\widehat{Z}_n$, brings us to the equalities

$$\overline{Z}_n = \frac{d\overline{\mathsf{P}}_n}{d\widetilde{\mathsf{P}}_n} = \frac{\overline{B}_n}{\widetilde{B}_n}, \qquad n \leqslant N, \tag{11}$$

showing that

$$\overline{B}_n \overline{\mathsf{E}}\left(\frac{f_N}{\overline{B}_N} \mid \mathscr{F}_n\right) = \frac{\overline{B}_n}{\overline{Z}_n} \widetilde{\mathsf{E}}\left(\frac{f_N}{\overline{B}_N}\overline{Z}_N \mid \mathscr{F}_n\right) = \widetilde{B}_n \widetilde{\mathsf{E}}\left(\frac{f_N}{\widetilde{B}_N} \mid \mathscr{F}_n\right).$$

This proves (9) and (10), so that, indeed, the value of the price $\mathbb{C}(f_N; \mathsf{P})$ in a complete market is independent of our choice of a discounting process. In Chapter VII, § 1b we shall consider discounting in the continuous-time case (and in greater detail). At this point, however, we mention only that a successful choice of a discounting process can in many cases considerably reduce the analytic complexity of finding the prices $\mathbb{C}(f_N; \mathsf{P})$ and the corresponding perfect hedges. See, for instance, the calculations for 'Russian options' in § 5d and Chapter VIII, § 2c.

5. Thus, formula (4) fully solves on a complete arbitrage-free market the pricing problem for perfect hedges in the case of the discounting process B. If $\widetilde{\mathsf{P}}$ is the corresponding martingale measure $\left(\text{ i.e., } \frac{S}{B} \text{ is a martingale with respect to } \widetilde{\mathsf{P}}\right)$, then a transition to another discounting process $\overline{B}$ changes also the martingale measure: the new measure $\overline{\mathsf{P}}$, in accordance with (11), can be obtained by the formula

$$d\overline{\mathsf{P}} = \frac{\overline{B}_N}{B_N}\, d\widetilde{\mathsf{P}}. \tag{12}$$

On the other hand, if we are in an incomplete market, then there exist several martingale measures and it is no longer easy to understand what we *must call* the hedging price. If we have two distinct martingale measures, $\widetilde{\mathsf{P}}_1$ and $\widetilde{\mathsf{P}}_2$, and therefore can give distinct definitions of what it means for a market to be arbitrage-free, then the quantities $\widetilde{B}_0 \mathsf{E}_{\widetilde{\mathsf{P}}_1} \frac{f_N}{\widetilde{B}_N}$ and $\widetilde{B}_0 \mathsf{E}_{\widetilde{\mathsf{P}}_2} \frac{f_N}{\widetilde{B}_N}$, are, in general, also distinct (see § 1c below).

6. For an illustration of our above discussion of various discounting processes and the calculation of conditional expectations with respect to different measures we consider the following example.

Let f_N be the price of some contingent claim (expressed in USD). If we consider a complete arbitrage-free (dollar) market, then the price of a perfect hedge is $\widetilde{B}_0 \,\widetilde{\mathsf{E}} \dfrac{f_N}{\widetilde{B}_N}$, where $\widetilde{B} = (\widetilde{B}_n)_{n \leqslant N}$ is a (dollar) bank account.

Next, we consider a market with prices in German marks (DEM). Expressed in marks, the value of the same contingent claim becomes $f_N S_N$ (DEM), where

$$S_N = \left(\frac{\text{DEM}}{\text{USD}}\right)_N$$

is the cross rate at time N.

If $\overline{B} = (\overline{B}_n)_{n \leqslant N}$ is a DEM-bank account and the corresponding market is complete and arbitrage-free, then the price of a hedge against the contingent claim $f_N S_N$ (in DEM) is

$$\overline{B}_0 \,\overline{\mathsf{E}} \frac{S_N f_N}{\overline{B}_N},$$

which, converted into dollars, makes up

$$S_0^{-1} \overline{B}_0 \,\overline{\mathsf{E}} \frac{S_N f_N}{\overline{B}_N}.$$

What are the conditions ensuring that, as one would anticipate, this dollar price is equal to $\widetilde{B}_0 \,\widetilde{\mathsf{E}} \dfrac{f_N}{\widetilde{B}_N}$, or, in a more general form,

$$S_n^{-1} \overline{B}_n \,\overline{\mathsf{E}}\left(\frac{S_N f_N}{\overline{B}_N} \,\Big|\, \mathscr{F}_n \right) = \widetilde{B}_n \,\widetilde{\mathsf{E}}\left(\frac{f_N}{\widetilde{B}_N} \,\Big|\, \mathscr{F}_n \right)? \tag{13}$$

We assume that the DEM-market is arbitrage-free; then for the cross rate $S = (S_n)_{n \leqslant N}$, $S_n = \left(\dfrac{\text{DEM}}{\text{USD}}\right)_n$, the process $\left(\dfrac{S_n}{\overline{B}_n}\right)_{n \leqslant N}$ must be a $\overline{\mathsf{P}}$-martingale. Hence $\overline{\mathsf{E}}\left(\dfrac{S_N}{\overline{B}_N} \,\Big|\, \mathscr{F}_n \right) = \dfrac{S_n}{\overline{B}_n}$, and if $Z_n = \dfrac{d\overline{\mathsf{P}}_n}{d\widetilde{\mathsf{P}}_n}$, then, by Bayes's formula,

$$\frac{S_n}{\overline{B}_n} = \frac{1}{Z_n} \widetilde{\mathsf{E}}\left(\frac{S_N}{\overline{B}_N} Z_N \,\Big|\, \mathscr{F}_n \right) \qquad (\widetilde{\mathsf{P}}\text{-a.s.}). \tag{14}$$

Hence

$$\frac{S_n}{\widetilde{B}_n} \cdot \left(\frac{\widetilde{B}_n}{\overline{B}_n} Z_n \right) = \widetilde{\mathsf{E}}\left(\frac{S_N}{\widetilde{B}_N} \cdot \frac{\widetilde{B}_N}{\overline{B}_N} \cdot Z_N \,\Big|\, \mathscr{F}_n \right). \tag{15}$$

If the USD-market is also arbitrage-free, then $\left(\dfrac{S_n}{\widetilde{B}_n}\right)_{n\leqslant N}$ is a $\widetilde{\mathsf{P}}$-martingale, so that

$$\frac{S_n}{\widetilde{B}_n} = \widetilde{\mathsf{E}}\left(\frac{S_N}{\widetilde{B}_N} \,\Big|\, \mathscr{F}_n\right). \tag{16}$$

By (15), (16), and our assumption that the USD-market is complete, and therefore the measure $\widetilde{\mathsf{P}}$ is unique, we obtain

$$\frac{\widetilde{B}_N}{\overline{B}_N} \cdot \frac{d\overline{\mathsf{P}}}{d\widetilde{\mathsf{P}}} = 1 \tag{17}$$

(cf. (12)), and for all $n \leqslant N$,

$$\frac{\widetilde{B}_n}{\overline{B}_n} \cdot \frac{d\overline{\mathsf{P}}_n}{d\widetilde{\mathsf{P}}_n} = 1. \tag{18}$$

Since the process $Z = (Z_n, \mathscr{F}_n, \widetilde{\mathsf{P}})_{n\leqslant N}$ with $Z_n = \dfrac{d\overline{\mathsf{P}}_n}{d\widetilde{\mathsf{P}}_n}$ is a martingale, equality (18) means that $\left(\dfrac{\overline{B}_n}{\widetilde{B}_n}\right)_{n\leqslant N}$ is also a $\widetilde{\mathsf{P}}$-martingale. In fact one could foresee that this property would ensure the coincidence of the (dollar) prices of hedging against f_N in the USD and the DEM-market before any calculations: we could regard $\overline{B} = (\overline{B}_n)_{n\leqslant N}$ as one of the basic securities on the dollar market with dollar bank account $\widetilde{B} = (\widetilde{B}_n)_{n\leqslant N}$.

§ 1c. Main Hedge Pricing Formula. Incomplete Markets

1. As shown in the preceding section, the price $\mathbb{C}(f_N; \mathsf{P})$ of perfect hedging on a complete arbitrage-free market has the following expression:

$$\mathbb{C}(f_N; \mathsf{P}) = B_0 \widetilde{\mathsf{E}} \frac{f_N}{B_N}, \tag{1}$$

where $\widetilde{\mathsf{E}}$ is averaging with respect to the (unique) martingale measure $\widetilde{\mathsf{P}}$ such that $\dfrac{S}{B}$ is a $\widetilde{\mathsf{P}}$-martingale.

A similar question of hedging prices can be put, of course, also on an incomplete market. However, there does not necessarily exist a perfect self-financing hedge on such a market, therefore we must modify our definition of the hedging price and consider a somewhat wider class than *self-financing* strategies to which we could stick in the case of a complete market.

We recall that the value X^π of a self-financing strategy $\pi = (\beta, \gamma)$ on a complete market can in fact be defined in two ways: we either write

$$X_n^\pi = \beta_n B_n + \gamma_n S_n \tag{2}$$

with $B_{n-1}\Delta\beta_n + S_{n-1}\Delta\gamma_n = 0$ or

$$X_n^\pi = X_0^\pi + \sum_{k=1}^{n}(\beta_k\Delta B_k + \gamma_k\Delta S_k) \tag{3}$$

(see Chapter V, § 1a for greater detail).

Representation (3) is more convenient in a certain sense for it visualizes the *dynamics of the growth of capital*: X_0^π is the share of the initial capital in X_n^π, while

$$\Delta X_n^\pi = \beta_n\Delta B_n + \gamma_n\Delta S_n \tag{4}$$

is the increment.

In our case of hedging on an incomplete market it seems reasonable to consider, alongside the *portfolio* $\pi = (\beta,\gamma)$, also the *consumption process* $C = (C_n)_{n\geqslant 0}$, which is a non-negative non-decreasing process with $\mathscr{F}_n$-measurable C_n and $C_0 = 0$.

In fact, we already discussed this situation in Chapter V, § 1a, where we called it the *case with 'consumption'*. We assumed there that, in place of (4), the dynamics of the capital $X^{\pi,C}$ corresponding to a portfolio π and consumption C can be described by the relations

$$\Delta X_n^{\pi,C} = \beta_n\Delta B_n + \gamma_n\Delta S_n - \Delta C_n, \tag{5}$$

where $\Delta B_n + \gamma_n\Delta S_n$ is the share of the total increment that can be ascribed to the composition of the portfolio and is due to the 'market-related' changes ΔB_n and ΔS_n, while ΔC_n characterizes the expenses on consumption (e.g., the expenditures on the changes in the portfolio).

Thus, we shall now assume that the value $X^{\pi,C}$ of the *strategy* (π, C) can (by analogy with (3)) be calculated by the formulas

$$X_n^{\pi,C} = X_0^{\pi,C} + \sum_{k=1}^{n}(\beta_k\Delta B_k + \gamma_k\Delta S_k) - C_n, \qquad n\geqslant 1, \tag{6}$$

which is equivalent to

$$\Delta\left(\frac{X_n^{\pi,C}}{B_n}\right) = \gamma_n\Delta\left(\frac{S_n}{B_n}\right) - \frac{\Delta C_n}{B_{n-1}}, \qquad n\geqslant 1.$$

Remark 1. Setting $\beta_k' = \beta_k - \dfrac{\Delta C_k}{\Delta B_k}$ we see from (6) that

$$X_n^{\pi,C} = X_0^{\pi,C} + \sum_{k=1}^{n}(\beta_k'\Delta B_k + \gamma_k\Delta S_k).$$

This formula is very much similar to (3). However, while β_k in (3) is $\mathscr{F}_{k-1}$-measurable, β'_k is now only $\mathscr{F}_k$-measurable.

Remark 2. We have already mentioned that, generally speaking, perfect hedging is unattainable in incomplete markets, i.e., there does not necessarily exist hedging $\pi = (\beta, \gamma)$ such that $X_N^\pi = f_N$ (P-a.s.) At the same time, this does not rule out the possibility that modifying our definition of an admissible strategy we can attain the level of terminal capital *offsetting* (P-a.s.) the pay-off f_N. As will be clear from the proof of the theorem below, our calling upon 'consumption' enables us to find a strategy (π, C) such that $X_N^{\pi,C} = f_N$ (P-a.s.). This is one 'technical' argument in favor of considering consumption C alongside the portfolio π. On the other hand, our introduction of strategies with 'consumption', which must satisfy constraints of the form $\Delta C_n \geqslant 0$, has clear economic implications.

2. DEFINITION. The *upper price of European hedging* (against a $\mathscr{F}_N$-measurable pay-off f_N) is the quantity

$$\mathbb{C}^*(f_N; \mathsf{P}) = \inf\bigl\{x \colon \exists\,(\pi, C) \text{ with } X_0^{\pi,C} = x \text{ and } X_N^{\pi,C} \geqslant f_N \ (\mathsf{P}\text{-a.s.})\bigr\}. \tag{7}$$

Remark. Besides the upper price we can also consider the *lower price* of hedging (see the definition in § 1b). In our subsequent discussions we shall consider only the upper price and call it simply the *price* of hedging.

Let $\mathscr{P}(\mathsf{P})$ be the set of all martingale measures $\widetilde{\mathsf{P}}$ equivalent to P. We assume that $\mathscr{P}(\mathsf{P}) \neq \varnothing$.

The central result of pricing theory on *incomplete arbitrage-free* markets, which generalizes formula (1), is as follows.

THEOREM 1 (*Main formula of the price of European hedging on incomplete markets*). *Let f_N be a non-negative bounded $\mathscr{F}_N$-measurable function. Then, on an incomplete arbitrage-free market, the price $\mathbb{C}^*(f_N; \mathsf{P})$ can be calculated by the formula*

$$\boxed{\mathbb{C}^*(f_N; \mathsf{P}) = \sup_{\widetilde{\mathsf{P}} \in \mathscr{P}(\mathsf{P})} B_0 \mathsf{E}_{\widetilde{\mathsf{P}}} \frac{f_N}{B_N},} \tag{8}$$

where $\mathsf{E}_{\widetilde{\mathsf{P}}}$ is averaging with respect to measure $\widetilde{\mathsf{P}}$.

We have already encountered a special case of this result (Theorem 1 in Chapter V, § 1c; see also [93]) in the case of a single-step model.

The crucial point in the proof of (8) is the so-called *optional decomposition* (see § 2d below), which has a rather technical proof. The existence of this optional decomposition and formula (8) were established for the first time by N. El Karoui and M. Quenez [136] and D. O. Kramkov [281]; as regards generalizations and other proofs, see also [99], [163], and [164].

3. Proof of the theorem. Let (π, C) be an (x, f_N)-hedge, i.e., assume that $X_0^{\pi,C} = x$ and $X_N^{\pi,C} \geqslant f_N$ (P-a.s.).

Then (cf. formula (2) in § 1b)

$$0 \leqslant \frac{f_N}{B_N} \leqslant \frac{X_N^{\pi,C}}{B_N} = \frac{x}{B_0} + \sum_{k=1}^{N} \gamma_k \, \Delta\Big(\frac{S_k}{B_k}\Big) - \sum_{k=1}^{N} \frac{\Delta C_k}{B_{k-1}}$$
$$\leqslant \frac{x}{B_0} + \sum_{k=1}^{N} \gamma_k \, \Delta\Big(\frac{S_k}{B_k}\Big), \tag{9}$$

so that for *each* measure $\widetilde{\mathsf{P}} \in \mathscr{P}(\mathsf{P})$ we have

$$B_0 \, \mathsf{E}_{\widetilde{\mathsf{P}}} \frac{f_N}{B_N} \leqslant x, \tag{10}$$

because $\mathsf{E}_{\widetilde{\mathsf{P}}} \sum_{k=1}^{N} \gamma_k \, \Delta\Big(\dfrac{S_k}{B_k}\Big) = 0$, which is a consequence of assertion 2) of the lemma in Chapter II, § 1c and the inequality $\sum_{k=1}^{N} \gamma_k \, \Delta\Big(\dfrac{S_k}{B_k}\Big) \geqslant -\dfrac{x}{B_0}$ following by (9).

Hence

$$\sup_{\widetilde{\mathsf{P}} \in \mathscr{P}(\mathsf{P})} B_0 \, \mathsf{E}_{\widetilde{\mathsf{P}}} \frac{f_N}{B_N} \leqslant \mathbb{C}^*(f_N; \mathsf{P}). \tag{11}$$

To prove the reverse inequality we set

$$Y_n = \operatorname*{ess\,sup}_{\widetilde{\mathsf{P}} \in \mathscr{P}(\mathsf{P})} \mathsf{E}_{\widetilde{\mathsf{P}}} \Big(\frac{f_N}{B_N} \,\Big|\, \mathscr{F}_n\Big), \tag{12}$$

where, by definition, the essential supremum Y_n is the $\mathscr{F}_n$-measurable random variable that, on the one hand, satisfies the inequality

$$Y_n \geqslant \mathsf{E}_{\widetilde{\mathsf{P}}} \Big(\frac{f_N}{B_N} \,\Big|\, \mathscr{F}_n\Big) \qquad (\mathsf{P}\text{-a.s.}) \tag{13}$$

for each measure $\widetilde{\mathsf{P}} \in \mathscr{P}(\mathsf{P})$, and, on the other hand, has the following property ('minimality'): if $\overline{Y}_n$ is another variable dominating the right-hand side of (13), then $Y_n \leqslant \overline{Y}_n$ (P-a.s.).

As shown in § 2b, the sequence $Y = (Y_n, \mathscr{F}_n)_{n \leqslant N}$ is a supermartingale with respect to an *arbitrary* (!) measure $\mathsf{Q} \in \mathscr{P}(\mathsf{P})$, i.e.,

$$\mathsf{E}_{\mathsf{Q}}(Y_{n+1} \,|\, \mathscr{F}_n) \leqslant Y_n \quad (\mathsf{Q}\text{-a.s.}). \tag{14}$$

Recall that, as follows from the classical Doob decomposition (§ 1b, Chapter II), for each *particular* measure Q we can find a martingale $M^{\mathsf{Q}} = (M_n^{\mathsf{Q}}, \mathscr{F}_n, \mathsf{Q})_{0 \leqslant n \leqslant N}$, where $M_0^{\mathsf{Q}} = 0$, and a predictable non-decreasing process $A^{\mathsf{Q}} = (A_n^{\mathsf{Q}}, \mathscr{F}_{n-1}, \mathsf{Q})_{1 \leqslant n \leqslant N}$, where $A_0^{\mathsf{Q}} = 0$, such that

$$Y_n = Y_0 + M_n^{\mathsf{Q}} - A_n^{\mathsf{Q}}. \tag{15}$$

It is remarkable that if $Y = (Y_n, \mathscr{F}_n)$ is a supermartingale with respect to *each* measure Q in the family $\mathscr{P}(\mathsf{P})$, then Y has a *universal* (i.e., independent of Q) *decomposition*

$$Y_n = Y_0 + \overline{M}_n - \overline{C}_n, \tag{16}$$

where $\overline{M} = (\overline{M}_n, \mathscr{F}_n)$ is a martingale with respect to *each* measure $\mathsf{Q} \in \mathscr{P}(\mathsf{P})$ and $\overline{C} = (\overline{C}_n, \mathscr{F}_n)$ is a non-decreasing process with $\overline{C}_0 = 0$.

We point out that while the process A^{Q} in the Doob decomposition (15) is predictable (i.e., the A_n^{Q} are $\mathscr{F}_{n-1}$-measurable), the process $\overline{C} = (\overline{C}_n, \mathscr{F}_n)$ in (16) is only *optional* (i.e., the $\overline{C}_n$ are $\mathscr{F}_n$-measurable). This explains why (16) is called the *optional decomposition.*

In the special case of the supermartingale $Y = (Y_n, \mathscr{F}_n)$ defined by (12) we can further specify the structure of the martingale $\overline{M} = (\overline{M}_n, \mathscr{F}_n)$:

$$\overline{M}_n = \sum_{k=1}^{n} \overline{\gamma}_k \, \Delta\Big(\frac{S_k}{B_k}\Big), \tag{17}$$

where $\overline{\gamma} = (\overline{\gamma}_n, \mathscr{F}_{n-1})$ is a predictable process. (We emphasize that this is far from trivial and can be established simultaneously with the proof of the optional decomposition; see § 2d.)

For the processes $\overline{\gamma}$, $\overline{C}$, and Y_0 introduced by (16) and (17) we shall now construct the portfolio $\widetilde{\pi} = (\widetilde{\beta}, \widetilde{\gamma})$ and the consumption process $\widetilde{C}$ such that for the corresponding capital we have $X_0^{\widetilde{\pi}, \widetilde{C}} = B_0 \sup\limits_{\widetilde{\mathsf{P}} \in \mathscr{P}(\mathsf{P})} \mathsf{E}_{\widetilde{\mathsf{P}}} \dfrac{f_N}{B_N}$ and $X_N^{\widetilde{\pi}, \widetilde{C}} \geqslant f_N$. Of course, this means that

$$\mathbb{C}^*(f_N; \mathsf{P}) \leqslant X_0^{\widetilde{\pi}, \widetilde{C}} = B_0 \sup_{\widetilde{\mathsf{P}} \in \mathscr{P}(\mathsf{P})} \mathsf{E}_{\widetilde{\mathsf{P}}} \frac{f_N}{B_N},$$

which, together with (11), brings us to (8).

The required portfolio $\widetilde{\pi} = (\widetilde{\beta}, \widetilde{\gamma})$ and the consumption process $\widetilde{C}$ can be defined as follows:

$$\widetilde{\gamma}_n = \overline{\gamma}_n, \qquad \widetilde{\beta}_n = Y_n - \widetilde{\gamma}_n \frac{S_n}{B_n}, \tag{18}$$

$$\widetilde{C}_n = \sum_{k=1}^{n} B_{k-1} \, \Delta \overline{C}_k, \tag{19}$$

where $\overline{\gamma}$ and $\overline{C}$ are as in the optional decomposition of the supermartingale Y.

For $\widetilde{\pi}$ and $\widetilde{C}$ so defined the starting capital is

$$X_0^{\widetilde{\pi},\widetilde{C}} = \widetilde{\beta}_0 B_0 + \widetilde{\gamma}_0 S_0 = Y_0 B_0.$$

In the scheme with consumption we assume (see Chapter V, § 1a.4) that the increment in the capital can be described by the formula

$$\Delta X_n^{\widetilde{\pi},\widetilde{C}} = \widetilde{\beta}_n \,\Delta B_n + \widetilde{\gamma}_n \,\Delta S_n - \Delta \widetilde{C}_n, \tag{20}$$

which, as already mentioned, yields

$$\Delta\left(\frac{X_n^{\widetilde{\pi},\widetilde{C}}}{B_n}\right) = \widetilde{\gamma}_n \,\Delta\left(\frac{S_n}{B_n}\right) - \frac{\Delta \widetilde{C}_n}{B_{n-1}} \tag{21}$$

(cf. also (27) in Chapter V, § 1a).

By (16)–(19),

$$\Delta\left(\frac{X_n^{\widetilde{\pi},\widetilde{C}}}{B_n}\right) = \Delta Y_n, \tag{22}$$

and since $\dfrac{X_0^{\widetilde{\pi},\widetilde{C}}}{B_0} = Y_0$, it follows that

$$\frac{X_N^{\widetilde{\pi},\widetilde{C}}}{B_N} = Y_N = \frac{f_N}{B_N}\,. \tag{23}$$

Thus, $X_N^{\widetilde{\pi},\widetilde{C}} = f_N$, so that the proposed strategy $(\widetilde{\pi}, \widetilde{C})$ with starting capital

$$X_0^{\widetilde{\pi},\widetilde{C}} = B_0 Y_0 = B_0 \sup_{\widetilde{\mathsf{P}} \in \mathscr{P}(\mathsf{P})} \mathsf{E}_{\widetilde{\mathsf{P}}} \frac{f_N}{B_N}$$

gives one a *perfect hedge*: $X_N^{\widetilde{\pi},\widetilde{C}} = f_N$.

Hence

$$\mathbb{C}^*(f_N; \mathsf{P}) \leqslant B_0 \sup_{\widetilde{\mathsf{P}} \in \mathscr{P}(\mathsf{P})} \mathsf{E}_{\widetilde{\mathsf{P}}} \frac{f_N}{B_N}\,.$$

Together with (11) this proves required formula (8) (provided that we have the optional decomposition).

Thus, we have proved Theorem 1 and, incidentally, established the following result (cf. Theorem 2 in § 1b).

THEOREM 2 (*Main formulas for a perfect hedge, its value, and consumption*). *On an arbitrage-free market it is possible to find a self-financing hedge $\pi^* = (\beta^*, \gamma^*)$ and consumption C^* such that the value of this hedge, $X_n^{\pi^*} = \beta_n^* B_n + \gamma_n^* S_n$, changes in accordance with the balance condition $\Delta X_n^{\pi^*} = \beta_n^* \Delta B_n + \gamma_n^* \Delta S_n - \Delta C_n^*$ and satisfies the relations*

$$X_0^{\pi^*} = \mathbb{C}^*(f_N; \mathsf{P}) \quad \left(= \sup_{\widetilde{\mathsf{P}} \in \mathscr{P}(\mathsf{P})} B_0 \, \mathsf{E}_{\widetilde{\mathsf{P}}} \frac{f_N}{B_N} \right)$$

and

$$X_N^{\pi^*} = f_N \quad (\mathsf{P}\text{-a.s.}).$$

The value $X_n^{\pi^*}$ of this hedge can be calculated by the formula

$$X_n^{\pi^*} = B_n \operatorname*{ess\,sup}_{\widetilde{\mathsf{P}} \in \mathscr{P}(\mathsf{P})} \mathsf{E}_{\widetilde{\mathsf{P}}} \left(\frac{f_N}{B_N} \,\Big|\, \mathscr{F}_n \right),$$

the components $\gamma^* = (\gamma_n^*)$ and $C^* = (C_n^*)$ can be found from the optional decomposition

$$\operatorname*{ess\,sup}_{\widetilde{\mathsf{P}} \in \mathscr{P}(\mathsf{P})} \mathsf{E}_{\widetilde{\mathsf{P}}} \left(\frac{f_N}{B_N} \,\Big|\, \mathscr{F}_n \right) = \sup_{\widetilde{\mathsf{P}} \in \mathscr{P}(\mathsf{P})} \mathsf{E}_{\widetilde{\mathsf{P}}} \frac{f_N}{B_N} + \sum_{k=1}^{n} \gamma_k^* \, \Delta\left(\frac{S_k}{B_k} \right) - \sum_{k=1}^{n} \frac{\Delta C_k^*}{B_{k-1}},$$

and the components $\beta^* = (\beta_n^*)$ can be found from the condition $X_n^{\pi^*} = \beta_n^* B_n + \gamma_n^* S_n$.

§ 1d. Hedge Pricing on the Basis of the Mean Square Criterion

1. Let $f_N = f_N(\omega)$ be a $\mathscr{F}_N$-measurable pay-off. In a *complete arbitrage-free* market there exist a starting capital x and a strategy π such that a trader can *replicate* f_N faithfully (in the sense that $X_N^\pi(x) = f_N$ with probability one, where $X_N^\pi(x)$ is the value X_N of the strategy π such that $X_0^\pi = x$).

However, if a market (with or without arbitrage) is *incomplete*, then the situation becomes considerably more complex and one cannot hope to replicate f_N faithfully.

In § 1c we considered the calculations of the hedging price $\mathbb{C}^*(f_N; \mathsf{P})$ on an arbitrage-free market when the *hedging strategy* π was to ensure that

$$X_N^\pi(\mathbb{C}^*(f_N; \mathsf{P})) \geqslant f_N \quad (\mathsf{P}\text{-a.s.}).$$

In the present section we shall understand optimal hedging in a somewhat different sense, as the replication of f_N with 'maximum precision'.

Our choice of the measure of accuracy is, in a sense, a matter of convention; it depends on one's aims, the chances of finding a precise solution of the corresponding optimization problem, and so on.

In what follows, we measure the quality of replication by the mean square deviation

$$R_N(\pi, x) = \mathsf{E}[X_N^\pi(x) - f_N]^2, \tag{1}$$

which in many cases enables one to find the 'optimal' components x^* and π^* bringing about the minimum of $\mathsf{E}[X_N^\pi(x) - f_N]^2$:

$$\inf_{(\pi,x)} R_N(\pi; x) = R_N(\pi^*; x^*). \tag{2}$$

2. Let $(\Omega, \mathscr{F}, (\mathscr{F}_n)_{n\leqslant N}, \mathsf{P})$ be a fixed filtered probability space, let $\mathscr{F}_0 = \{\varnothing, \Omega\}$, and let $\mathscr{F}_N = \mathscr{F}$. Let $S = (S_n^1, \dots, S_n^d)_{n\leqslant N}$ be a sequence of prices of some d-dimensional asset and assume that $\mathsf{E}f_N^2 < \infty$.

If the sequence of prices is a *martingale* and, moreover, a square integrable martingale with respect to the original measure P, then the optimization problem (2) is easy to analyze in the class of strategies π satisfying the inequality $\mathsf{E}(X_N^\pi(x))^2 < \infty$. (We emphasize that we do not assume the *uniqueness* of the martingale measure P and, therefore, the completeness of the market.)

For a strategy $\pi = (\gamma^1, \dots, \gamma^d)$, where the $\gamma^i = (\gamma_n^i)_{n\leqslant N}$ are predictable variables, let

$$X_n^\pi(x) = x + \sum_{k=1}^n (\gamma_k, \Delta S_k) = x + \sum_{k=1}^n \biggl(\sum_{i=1}^d \gamma_k^i \, \Delta S_k^i\biggr) \tag{3}$$

be its value.

Since the sequence $(X_n^\pi(x))_{n\leqslant N}$ is a martingale, it follows that

$$\mathsf{E}X_N^\pi(x) = x. \tag{4}$$

Setting $\xi = X_N^\pi(x) - f_N$ we see in view of the obvious equality

$$\mathsf{E}\xi^2 = (\mathsf{E}\xi)^2 + \mathsf{E}(\xi - \mathsf{E}\xi)^2$$

that

$$R_N(\pi; x) = \bigl[\mathsf{E}(f_N - x)\bigr]^2 + \mathsf{E}\bigl[\bigl(X_N^\pi(x) - x\bigr) - \bigl(f_N - \mathsf{E}f_N\bigr)\bigr]^2. \tag{5}$$

We shall show below that for each pair (π, x) with $\mathsf{E}(X_N^\pi(x))^2 < \infty$ there exists a pair (π^*, x) such that $R_N(\pi; x) \geqslant R_N(\pi^*; x)$ and $X_N^{\pi^*}(x) - x$ is *independent* of x. Hence it follows by (5) that the smallest lower bound $\inf\limits_x \Bigl[\inf\limits_\pi R_N(\pi; x)\Bigr]$ is attained for

$$x^* = \mathsf{E}f_N. \tag{6}$$

For $i = 1, \dots, d$ we set

$$\gamma_n^{*i} = \frac{\mathsf{E}(f_N \, \Delta S_n^i | \mathscr{F}_{n-1})}{\mathsf{E}((\Delta S_n^i)^2 | \mathscr{F}_{n-1})} \tag{7}$$

(where we put $0/0 = 0$) and consider the martingale $L^* = (L_n^*)_{n\leqslant N}$ with

$$L_n^* = \mathsf{E}\biggl[f_N - \sum_{k=1}^{N}(\gamma_k^*, \Delta S_k) \biggm| \mathscr{F}_n\biggr] - x. \tag{8}$$

Clearly, f_N has the following decomposition:

$$f_N = x + \sum_{k=1}^{N}(\gamma_k^*, \Delta S_k) + L_N^*. \tag{9}$$

Using the definition (7) we can verify directly that

$$\mathsf{E}\bigl(\Delta L_n^* \cdot (\gamma_n, \Delta S_n) \mid \mathscr{F}_{n-1}\bigr) = 0. \tag{10}$$

Note that this property is equivalent to the 'orthogonality' of the square integrable martingales $(L_n^*)_{n\leqslant N}$ and $\Bigl(\sum\limits_{k=1}^{n}(\gamma_k, \Delta S_k)\Bigr)_{n\leqslant N}$ in the sense that their product is also a martingale. It is worth noting in this connection that (9) is called the *Kunita–Watanabe decomposition* in the 'general theory of martingales'.

By (9) and (10) we obtain that for each pair (π, x),

$$\begin{aligned}
R_N(\pi; x) &= \mathsf{E}\biggl[f_N - \biggl(x + \sum_{k=1}^{N}(\gamma_k, \Delta S_k)\biggr)\biggr]^2 \\
&= \mathsf{E}\biggl[\sum_{k=1}^{N}(\gamma_k^* - \gamma_k, \Delta S_k) + L_N^*\biggr]^2 \\
&= \mathsf{E}\biggl[\sum_{k=1}^{N}(\gamma_k^* - \gamma_k, \Delta S_k)\biggr]^2 + \mathsf{E}[L_N^*]^2 \geqslant \mathsf{E}[L_N^*]^2 \\
&= \mathsf{E}\biggl[f_N - \biggl(x + \sum_{k=1}^{N}(\gamma_k^*, \Delta S_k)\biggr)\biggr]^2 \\
&= R_N(\pi^*; x) \geqslant R_N(\pi^*; x^*).
\end{aligned} \tag{11}$$

Moreover, the first inequality turns to an equality for $\gamma = \gamma^*$.

Thus, we have proved the following result.

THEOREM. *Assume that the original measure* P *is a martingale measure. Then the optimal hedge* $\pi^* = (\gamma^{*1}, \dots, \gamma^{*d})$ *in the problem* (2) (*in the sense of the mean square criterion*) *is described by formulas* (7), $x^* = \mathsf{E} f_N$, *and*

$$R_N(\pi^*; x^*) = \mathsf{E}\biggl[f_N - \biggl(x^* + \sum_{k=1}^{N}\biggl(\sum_{i=1}^{d}\gamma_k^{*i}\,\Delta S_k^i\biggr)\biggr)\biggr]^2. \tag{12}$$

Remark. If P is *not* a martingale measure (with respect to the prices S), then the question on the existence of an *optimal* pair (x^*, π^*) and its search become fairly complicated. See, e.g., the papers of H. Föllmer, M. Schweizer, and D. Sondermann [167], [168], [425], and also [195], [194] in this connection.

We note also that the concept of a *minimal* martingale measure that we mentioned in passing at the end of § 3d, Chapter V has been developed precisely in connection with the above problem of hedging on the basis of the mean square test.

§ 1e. Forward Contracts and Futures Contracts

1. In the present section we shall show how to price forward contracts and futures contracts, important investment instruments used on financial markets alongside options.

In accordance with the definitions in Chapter I, § 1c, forwards and futures are sale contracts for some asset that must be delivered at a specified instant in the future at a specified price (the 'forward' or the 'futures' price, respectively).

There exists an essential distinction between futures and forwards, although both are sale contracts.

Forwards are in fact mere sale agreements between the two parties concerned, with no intermediaries.

Futures are also sale agreements, but they are concluded at an exchange and involve a clearing house through which the payments are made and which is the guarantor of the contract.

2. Assume that the market price of the asset in question can be described by a stochastic sequence $S = (S_k)_{k \leqslant N}$, where N is the maturity date of the contract, which we identify with the instant of delivery.

Clearly, if the deal is struck at time N, when the market price of the asset is S_N, then for any reasonable definition of the forward or the futures price it must be equal to S_N. This becomes another matter, of course, if the contract is sold at time $n < N$; the crucial question here is how we must understand the *fair price of the contract* (on an arbitrage-free market).

For a formalization, let us consider the scheme of a (B, S)-market described in Chapter V, § 1a, where $B = (B_n)$ is a bank account and $S = (S_n)$ is the traded asset. (If the asset in question is in fact *one* component of a d-dimensional vector of risk assets, then, in the arbitrage-free case, all our conclusion remain valid).

We shall consider the case with dividends discussed already in § 1a.4, when the value $X = (X_n^\pi)_{n \leqslant N}$ of the buyer's strategy $\pi = (\beta, \gamma)$ is described by the formula

$$X_n^\pi = \beta_n B_n + \gamma_n \Delta D_n, \tag{1}$$

and its changes are described as

$$\Delta X_n^\pi = \beta_n \, \Delta B_n + \gamma_n \, \Delta D_n, \tag{2}$$

where γ_n is the number of 'units' of the asset bought and $D = (D_n, \mathscr{F}_n)_{n \leqslant N}$, $D_0 = 0$, is the (not necessarily positive) process of overall dividends connected to the asset S.

We describe now the structure of dividends in the above cases of a forward or a futures contract and find the 'fair' prices of these contracts.

3. Assume that a forward contract is sold at time n and that both parties agree (on the basis of the 'information' $\mathscr{F}_n$) that the (forward) delivery price of the asset will be $\mathbb{F}_n(N)$.

Then, by the very mechanics of forward contracts, the overall dividends (positive or negative) can be represented as follows:

$$D_k = 0, \qquad n \leqslant k < N, \tag{3}$$

and

$$D_N = S_N - \mathbb{F}_n(N). \tag{4}$$

By (1) and (2) (see also (24) in Chapter V, § 1a) we obtain

$$\Delta\left(\frac{X_k^\pi}{B_k}\right) = \gamma_k \frac{\Delta D_k}{B_k}, \tag{5}$$

and therefore

$$\frac{X_N^\pi}{B_N} = \frac{X_n^\pi}{B_n} + \sum_{k=n+1}^{N} \gamma_k \frac{\Delta D_k}{B_k}, \qquad n < N. \tag{6}$$

Clearly, for a forward contract sold at time n we can set $\gamma_k = 0$ for $k \leqslant n$ and $\gamma_k = \gamma_{n+1}$ for all $k \geqslant n+1$, where γ_{n+1} is the number of units of the asset S requested in the contract.

By (6) we obtain

$$\frac{X_N^\pi}{B_N} = \frac{X_n^\pi}{B_n} + \gamma_{n+1} \frac{S_N - \mathbb{F}_n(N)}{B_N} \tag{7}$$

and we immediately arrive at the following conclusion.

Assume that the (B,S)-market under consideration is *arbitrage-free and complete.* Let $\widetilde{\mathsf{P}}$ be the unique martingale measure such that $\left(\dfrac{S_n}{B_n}\right)_{n \leqslant N}$ is a martingale with respect to it. Assume also that the $\mathscr{F}_n$-measurable prices $\mathbb{F}_n(N)$ satisfy the relation

$$\mathsf{E}_{\widetilde{\mathsf{P}}}\left(\frac{S_N - \mathbb{F}_n(N)}{B_N} \,\Big|\, \mathscr{F}_n\right) = 0, \qquad n \leqslant N, \tag{8}$$

i.e., let

$$\mathbb{F}_n(N) = \frac{\mathsf{E}_{\widetilde{\mathsf{P}}}\left(\dfrac{S_N}{B_N} \,\Big|\, \mathscr{F}_n\right)}{\mathsf{E}_{\widetilde{\mathsf{P}}}\left(\dfrac{1}{B_N} \,\Big|\, \mathscr{F}_n\right)}. \tag{9}$$

Then we see from (7) that

$$\mathsf{E}_{\widetilde{\mathsf{P}}} \frac{X_N^\pi}{B_N} = \mathsf{E}_{\widetilde{\mathsf{P}}} \frac{X_n^\pi}{B_n}, \tag{10}$$

so that the forward contract sold at time n at the price $\mathbb{F}_n(N)$ defined by (9) is arbitrage-free (i.e., if $X_n^\pi = 0$ and $\mathsf{P}(X_N^\pi \geqslant 0) = 1$, then $\mathsf{P}(X_N^\pi = 0) = 1$; see Definition 2 in Chapter V, § 2a). Of course, the value of $\mathbb{F}_n(N)$ (the *forward price*) can be regarded as the fair price of the forward contract.

Note that our assumption about an arbitrage-free (B, S)-market gives us the equality

$$\mathsf{E}_{\widetilde{\mathsf{P}}} \left(\frac{S_N}{B_N} \,\Big|\, \mathscr{F}_n \right) = \frac{S_n}{B_n}.$$

Thus, we see from (9) that the *arbitrage-free forward prices* $\mathbb{F}_n(N)$ *can be defined as follows*:

$$\boxed{\mathbb{F}_n(N) = \frac{S_n}{\mathsf{E}_{\widetilde{\mathsf{P}}} \left(\dfrac{B_n}{B_N} \,\Big|\, \mathscr{F}_n \right)}, \qquad n \leqslant N.} \tag{11}$$

4. We now proceed to futures contracts. Assume that we sell a contract of this kind at time n, at an $\mathscr{F}_n$-measurable price $\Phi_n(N)$ (the futures price). Immediately, the mechanism of settling through the clearing house comes into play. This can be roughly (skipping the issue of the margin account, the amounts deposited into it, and so on) described as follows in terms of (positive or negative) dividends.

If the market futures price at time $n+1$, $\Phi_{n+1}(N)$, turns out less than $\Phi_n(N)$, then the buyer puts the amount $\Phi_n(N) - \Phi_{n+1}(N)$ into the seller's account. If $\Phi_{n+1}(N) > \Phi_n(N)$, then, conversely, this is the seller who deposits the amount $\Phi_{n+1}(N) - \Phi_n(N)$ into the buyer's account.

We shall set $\delta_0 = \Phi_0(N)$ and

$$\delta_n = \Phi_n(N) - \Phi_{n-1}(N), \qquad n \geqslant 1.$$

Also, let

$$D_n = \delta_0 + \delta_1 + \cdots + \delta_n, \tag{12}$$

so that $\Delta D_n = \delta_n$, $n \geqslant 1$.

By (6) we see that (cf. (7))

$$\frac{X_N^\pi}{B_N} = \frac{X_n^\pi}{B_n} + \gamma_{n+1} \sum_{k=n+1}^{N} \frac{\Delta D_k}{B_k}. \tag{13}$$

Hence, similarly to forward contracts, we conclude that if $\widetilde{\mathsf{P}}$ is a unique martingale measure for the (B, S)-market in question, then the condition

$$\mathsf{E}_{\widetilde{\mathsf{P}}} \left(\sum_{k=n+1}^{N} \frac{\Delta D_k}{B_k} \,\Big|\, \mathscr{F}_n \right) = 0 \tag{14}$$

imposed on the prices $\Phi_n(N), \dots, \Phi_{n+1}(N)$ ensures that the futures contract sold at time n is arbitrate-free.

Assume that the sequence $D = (D_n)_{n \leqslant N}$ is a martingale with respect to $\widetilde{\mathsf{P}}$. Then (14) holds for each $n \geqslant 0$. In fact, since the predictable variables B_k are positive, we also have the converse result.

The condition that $D = (D_n)_{n \leqslant N}$ be a martingale means that

$$\mathsf{E}_{\widetilde{\mathsf{P}}}(D_N \mid \mathscr{F}_n) = D_n. \tag{15}$$

However, $D_n = \delta_0 + \cdots + \delta_n = \Phi_n(N)$ and $D_N = \Phi_N(N) = S_N$. Hence if the futures prices are

$$\boxed{\Phi_n(N) = \mathsf{E}_{\widetilde{\mathsf{P}}}(S_N \mid \mathscr{F}_n), \qquad n \leqslant N,} \tag{16}$$

then the corresponding futures contracts are arbitrage-free by (15).

Remark. Let $B = (B_n)_{n \leqslant N}$ be a deterministic sequence. Then, clearly,

$$\Phi_n(N) = \mathsf{E}_{\widetilde{\mathsf{P}}}\left(\frac{S_N}{B_N} \,\Big|\, \mathscr{F}_n\right) \cdot B_N = \frac{S_n}{B_n} \cdot B_N,$$

and comparing with (11) we obtain the well-known result that for a *deterministic* bank account $B = (B_n)$ *the forward and the futures prices are the same.*

2. American Hedge Pricing on Arbitrage-Free Markets

§ 2a. Optimal Stopping Problems. Supermartingale Characterization

1. The supermartingale characterization in § 1c of the sequence $Y = (Y_n)$ with respect to each measure in the family $\mathscr{P}(\mathsf{P})$ is not that striking if one treats the operation of taking the essential supremum in (12), § 1c as an optimization problem of finding the 'best' probability measure. For this interpretation, the above supermartingale property is a mere consequence of the well-known 'optimality principle' for the price process (the 'Bellman function') in a *stochastic optimization* problem.

A special case of such a problem is the optimal stopping problem for a stochastic sequence $f = (f_n)_{n \leqslant N}$, which seems a reasonable starting point in our discussion of 'supermartingale characterizations' in optimization problems. The emphasis on this case is also justified by our discussion of American options below (the buyer of such an option has a *right to choose* the date of exercising, so that the latter can be regarded as an 'optimization element'), and also because the necessity to have a '*sufficiently rich*' class of objects when considering ess sup is clearly visible there.

2. Let $f = (f_n, \mathscr{F}_n)_{0 \leqslant n \leqslant N}$ be a stochastic sequence on $(\Omega, \mathscr{F}, (\mathscr{F}_n)_{0 \leqslant n \leqslant N}, \mathsf{P})$, where $\mathscr{F}_0 = \{\varnothing, \Omega\}$ and $\mathscr{F}_N = \mathscr{F}$. We shall assume that $\mathsf{E}|f_n| < \infty$ for all $n \leqslant N < \infty$.

We are interested in the problems of finding

1) the functions (prices)

$$V_n^N = \sup_{\tau \in \mathfrak{M}_n^N} \mathsf{E} f_\tau, \tag{1}$$

where the supremum is taken over the class $\mathfrak{M}_n^N$ of all stopping times τ such that $n \leqslant \tau \leqslant N$, and

2) the *optimal stopping time* (which is well defined in our case).

We do not formulate here the optimal stopping problem in its most general form (see subsection 4 below), when N can be infinite (and $\mathfrak{M}_n^\infty$ is the class of all *finite* stopping times $\tau \geqslant n$); we restrict ourselves to *finite* 'time horizons'. The main reason is that this case is relatively easy to study, but on the other hand one must already use backward induction, which is one of the main tools in the search of both prices V_n^N and the corresponding optimal stopping times.

3. We introduce a sequence $\gamma^N = (\gamma_n^N)_{0\leqslant n\leqslant N}$ by the following (*ad hoc*) definition:

$$\begin{aligned} \gamma_N^N &= f_N, \\ \gamma_n^N &= \max(f_n, \mathsf{E}(\gamma_{n+1}^N \mid \mathscr{F}_n)). \end{aligned} \tag{2}$$

We also set

$$\tau_n^N = \min\{n \leqslant i \leqslant N : f_i = \gamma_i^N\}$$

for $0 \leqslant n \leqslant N$.

The following is one of the central results of the theory of optimal stopping problems for a finite time interval $0 \leqslant n \leqslant N$ (cf. [75: Chapter 3] and [441; Chapter 2]).

THEOREM 1. *The sequence $\gamma^N = (\gamma_n^N)_{n\leqslant N}$ defined by recursive relations* (2) *and the stopping times τ_n^N, $0 \leqslant n \leqslant N$, have the following properties:*

(a) $\tau_n^N \in \mathfrak{M}_n^N$;

(b) $\mathsf{E}(f_{\tau_n^N} \mid \mathscr{F}_n) = \gamma_n^N$;

(c) $\mathsf{E}(f_\tau \mid \mathscr{F}_n) \leqslant \mathsf{E}(f_{\tau_n^N} \mid \mathscr{F}_n) = \gamma_n^N$ *for each* $\tau \in \mathfrak{M}_n^N$;

(d) $\gamma_n^N = \operatorname*{ess\,sup}_{\tau\in\mathfrak{M}_n^N} \mathsf{E}(f_\tau \mid \mathscr{F}_n)$ *and, in particular,* $\gamma_0^N = \sup_{\tau\in\mathfrak{M}_0^N} \mathsf{E} f_\tau = \mathsf{E} f_{\tau_0^N}$;

(e) $V_n^N = \mathsf{E}\gamma_n^N$.

Proof. We drop for simplicity the superscript N in this proof and throughout subsection 3 and write γ_n, V_n, $\mathfrak{M}_n$, and τ_n in place of γ_n^N, V_n^N, $\mathfrak{M}_n^N$, and τ_n^N. Property (a) is a consequence of the definition of τ_n. Properties (b) and (c) are obvious for $n = N$. Now, we shall proceed by induction.

Assume that we have already established these properties for $n = N, N-1, \dots, k$. We claim that they also hold for $n = k-1$.

Let $\tau \in \mathfrak{M}_{k-1}$ and let $A \in \mathscr{F}_{k-1}$. We set $\overline{\tau} = \max(\tau, k)$. Clearly, $\overline{\tau} \in \mathfrak{M}_k$ and for $A \in \mathscr{F}_{k-1}$, bearing in mind that $\{\tau \geqslant k\} \in \mathscr{F}_{k-1}$, we obtain

$$\begin{aligned} \mathsf{E}[I_A f_\tau] &= \mathsf{E}[I_{A\cap\{\tau=k-1\}} f_\tau] + \mathsf{E}[I_{A\cap\{\tau\geqslant k\}} f_\tau] \\ &= \mathsf{E}[I_{A\cap\{\tau=k-1\}} f_{k-1}] + \mathsf{E}[I_{A\cap\{\tau\geqslant k\}} \mathsf{E}(f_\tau \mid \mathscr{F}_{k-1})] \\ &= \mathsf{E}[I_{A\cap\{\tau=k-1\}} f_{k-1}] + \mathsf{E}[I_{A\cap\{\tau\geqslant k\}} \mathsf{E}(\mathsf{E}(f_{\overline{\tau}} \mid \mathscr{F}_k) \mid \mathscr{F}_{k-1})] \\ &= \mathsf{E}[I_{A\cap\{\tau=k-1\}} f_{k-1}] + \mathsf{E}[I_{A\cap\{\tau\geqslant k\}} \mathsf{E}(\gamma_k \mid \mathscr{F}_{k-1})] \\ &\leqslant \mathsf{E}[I_A \gamma_{k-1}], \end{aligned} \tag{3}$$

where the last inequality is a consequence of (2).

Thus, $\mathsf{E}(f_\tau \mid \mathscr{F}_{k-1}) \leqslant \gamma_{k-1}$. We shall have proved required assertions (b) and (c) for $n = k-1$ once we have shown that

$$\mathsf{E}(f_{\tau_{k-1}} \mid \mathscr{F}_{k-1}) = \gamma_{k-1}. \tag{4}$$

To this end we consider the chain of inequalities in (3); we claim that we actually have the equality signs everywhere in (3) for $\tau = \tau_{k-1}$.

Indeed, by the definition of τ_{k-1} we have $\tau = \tau_k$ in the set $\{\tau_{k-1} \geqslant k\}$. Since $\mathsf{E}(f_{\tau_k} \mid \mathscr{F}_k) = \gamma_k$ by the inductive hypothesis, it follows that in (3) we have

$$\begin{aligned}\mathsf{E}\big[I_A f_{\tau_{k-1}}\big] &= \mathsf{E}\big[I_{A\cap\{\tau_{k-1}=k-1\}} f_{k-1}\big] \\ &\qquad + \mathsf{E}\big[I_{A\cap\{\tau_{k-1}\geqslant k\}} \mathsf{E}\big(\mathsf{E}(f_{\tau_k} \mid \mathscr{F}_k) \mid \mathscr{F}_{k-1}\big)\big] \\ &= \mathsf{E}\big[I_{A\cap\{\tau_{k-1}=k-1\}} f_{k-1}\big] + \mathsf{E}\big[I_{A\cap\{\tau_{k-1}\geqslant k\}} \mathsf{E}(\gamma_k \mid \mathscr{F}_k)\big] \\ &= \mathsf{E}\big[I_A \gamma_{k-1}\big],\end{aligned}$$

where the last equality follows from the definition of γ_{k-1} as $\max(f_{k-1}, \mathsf{E}(\gamma_k \mid \mathscr{F}_{k-1}))$, which means that $\gamma_{k-1} = f_{k-1}$ on $\{\tau_{k-1} = k-1\}$, while the inequality $f_{k-1} < \gamma_{k-1}$ in the set $\{\tau_{k-1} > k-1\}$ means that $\gamma_{k-1} = \mathsf{E}(\gamma_k \mid \mathscr{F}_{k-1})$ there.

Thus, we have proved (b) and (c), and therefore also assertion (d). Finally, since for each $\tau \in \mathfrak{M}_k$ we have

$$\mathsf{E} f_\tau \leqslant \mathsf{E} f_{\tau_k} = \mathsf{E}\gamma_k$$

by (c), it follows that $V_k = \sup_{\tau\in\mathfrak{M}_k} \mathsf{E} f_\tau = \mathsf{E} f_{\tau_k} = \mathsf{E}\gamma_k$, i.e., assertion (e) holds.

Remark 1. We used in the above proof the fact that if $\tau \in \mathfrak{M}_{k-1}$, then the stopping time $\widetilde{\tau} = \max(\tau, k)$ belongs to $\mathfrak{M}_k$. In our case the class $\mathfrak{M}_k$ has this property simply by definition. This means that the $\mathfrak{M}_k$, $k \leqslant N$, are 'sufficiently rich' classes in a certain sense. (See the end of subsection 1 in this connection.)

COROLLARY 1. *The sequence $\gamma = (\gamma_n)_{n\leqslant N}$ is a supermartingale. Moreover, γ is the smallest supermartingale majorizing the sequence $f = (f_n)_{n\leqslant N}$ in the following sense: if $\widetilde{\gamma} = (\widetilde{\gamma}_n)_{n\leqslant N}$ is also a supermartingale and $\widetilde{\gamma}_n \geqslant f_n$ for all $n \leqslant N$, then $\gamma_n \leqslant \widetilde{\gamma}_n$ (P-a.s.), $n \leqslant N$.*

Indeed, the fact that $\gamma = (\gamma_n)_{n\leqslant N}$ is a supermartingale majorizing $f = (f_n)_{n\leqslant N}$ follows from recursive relations (2).

Further, it is clear that $\widetilde{\gamma}_N \geqslant f_N$ and, for $n < N$,

$$\widetilde{\gamma}_n \geqslant \max\big(f_n, \mathsf{E}(\widetilde{\gamma}_{n+1} \mid \mathscr{F}_n)\big). \tag{5}$$

Since $\gamma_N = f_N$, it follows that $\widetilde{\gamma}_N \geqslant f_N$ and

$$\begin{aligned}\widetilde{\gamma}_{N-1} &\geqslant \max\big(f_{N-1}, \mathsf{E}(\widetilde{\gamma}_N \mid \mathscr{F}_{N-1})\big) \\ &\geqslant \max\big(f_{N-1}, \mathsf{E}(\gamma_N \mid \mathscr{F}_{N-1})\big) = \gamma_{N-1}.\end{aligned}$$

In a similar way we can show that $\widetilde{\gamma}_n \geqslant \gamma_n$ for each $n \leqslant N-1$.

COROLLARY 2. Corollary 1 can be formulated otherwise: *if $\gamma = (\gamma_n)_{n\leqslant N}$ is a solution of the recursive system of equations*

$$\gamma_n = \max\bigl(f_n, \mathsf{E}(\gamma_{n+1} \,|\, \mathscr{F}_n)\bigr), \qquad n < N, \tag{6}$$

with $\gamma_N = f_N$, then $\gamma_n \leqslant \widetilde{\gamma}_n$ for $n \leqslant N$ and each sequence $\widetilde{\gamma} = (\widetilde{\gamma}_n)_{n\leqslant N}$ such that $\widetilde{\gamma}_N \geqslant f_N$ and

$$\widetilde{\gamma}_n \geqslant \max\bigl(f_n, \mathsf{E}(\widetilde{\gamma}_{n+1} \,|\, \mathscr{F}_n)\bigr), \qquad n < N. \tag{7}$$

We claim that among all such solutions $\widetilde{\gamma} = (\widetilde{\gamma}_n)_{n\leqslant N}$ we can find the *smallest* solution $\gamma = (\gamma_n)_{n\leqslant N}$ with $\gamma_N = f_N$ satisfying the system of equations (6).

We set $\overline{\gamma}_N = f_N$, and let

$$\overline{\gamma}_n = \max\bigl(f_n, \mathsf{E}(\overline{\gamma}_{n+1} \,|\, \mathscr{F}_n)\bigr) \tag{8}$$

for $n < N$. Clearly, $\overline{\gamma}_n \geqslant f_n$ for all $n \leqslant N$ and $\overline{\gamma} = (\overline{\gamma}_n)_{n\leqslant N}$ is a supermartingale. Since $\gamma = (\gamma_n)_{n\leqslant N}$ has by assumption the property of minimality, it follows that

$$\max\bigl(f_n, \mathsf{E}(\overline{\gamma}_{n+1} \,|\, \mathscr{F}_n)\bigr) = \overline{\gamma}_n = \gamma_n. \tag{9}$$

Hence

$$\max\bigl(f_n, \mathsf{E}(\overline{\gamma}_{n+1} \,|\, \mathscr{F}_n)\bigr) = \overline{\gamma}_n \geqslant \gamma_n \geqslant \max\bigl(f_n, \mathsf{E}(\gamma_{n+1} \,|\, \mathscr{F}_n)\bigr)$$

for $n < N$ and, for $n = N$,

$$f_N = \overline{\gamma}_N \geqslant \gamma_N \geqslant f_N.$$

Consequently, $\gamma_N = \overline{\gamma}_N = f_N$ and, in view of (9), for all $n < N$ we have

$$\gamma_n = \overline{\gamma}_n.$$

Thus, the *smallest supermartingale* $\gamma = (\gamma_n)_{n\leqslant N}$, *majorizing the sequence* $f = (f_n)_{n\leqslant N}$ satisfies the equations (6) with $\gamma_N = f_N$.

COROLLARY 3. *The variable*

$$\tau_0^N = \min\{0 \leqslant i \leqslant N : f_i = \gamma_i^N\}$$

is an optimal stopping time in the class $\mathfrak{M}_0^N$, i.e.,

$$\sup_{\tau\in\mathfrak{M}_0^N} \mathsf{E} f_\tau = \mathsf{E} f_{\tau_0^N} \;(= \gamma_0^N).$$

4. We now consider the possible generalizations of Theorem 1 to the case when $N = \infty$. More precisely, we shall assume that $\mathfrak{M}_n^* \equiv \mathfrak{M}_n^\infty$ is the class of all *finite* Markov times $\tau = \tau(\omega)$ such that $\tau(\omega) \geqslant n$, $\omega \in \Omega$. We denote the class $\mathfrak{M}_0^\infty$ by $\mathfrak{M}^*$.

Further, let $f = (f_n, \mathscr{F}_n)_{n\geqslant 0}$ be a stochastic sequence defined on $(\Omega, \mathscr{F}, (\mathscr{F}_n)_{n\geqslant 0}, \mathsf{P})$, and let

$$V_n^* = \sup_{\tau\in\mathfrak{M}_n^*} \mathsf{E} f_\tau, \tag{10}$$

$$\gamma_n^* = \operatorname*{ess\,sup}_{\tau\in\mathfrak{M}_n^*} \mathsf{E}(f_\tau \mid \mathscr{F}_n), \tag{11}$$

$$\tau_n^* = \inf\{k \geqslant n \colon f_k = \gamma_k\}. \tag{12}$$

Of course one would expect that $\gamma_n^* = \lim\limits_{N\to\infty} \gamma_n^N$ (under certain conditions) and that one can make the limit transition in (2), which yields the following equations for $\gamma^* = (\gamma_n^*)$:

$$\gamma_n^* = \max\bigl(f_n, \mathsf{E}(\gamma_{n+1}^* \mid \mathscr{F}_n)\bigr). \tag{13}$$

In a similar way it would be natural if τ_n^*, defined in (12), were an optimal time in the class $\mathfrak{M}_n^*$ in the following sense:

$$V_n^* = \sup_{\tau\in\mathfrak{M}_n^*} \mathsf{E} f_\tau = \mathsf{E} f_{\tau_n^*} \tag{14}$$

and

$$V_n^* = \mathsf{E}\gamma_n^*. \tag{15}$$

As shown in the general theory of optimal stopping times (exposed, e.g., in [75] and [441]), these results hold, indeed, under certain conditions (but not always!).

Referring to the above-mentioned monographs [75] and [441] for details we present just one, fairly general, result in this direction.

THEOREM 2. *Let $f = (f_n, \mathscr{F}_n)_{n\geqslant 0}$ be a stochastic sequence with $\mathsf{E}\sup\limits_n f_n^- < \infty$. Then we have the following results.*

a) *The sequence $\gamma^* = (\gamma_n^*)_{n\geqslant 0}$, where*

$$\gamma_n^* = \operatorname*{ess\,sup}_{\tau\in\mathfrak{M}_n^*} \mathsf{E}(f_\tau \mid \mathscr{F}_n). \tag{16}$$

satisfies the recursive relations

$$\gamma_n^* = \max\bigl(f_n, \mathsf{E}(\gamma_{n+1}^* \mid \mathscr{F}_n)\bigr) \tag{17}$$

and is therefore a supermartingale majorizing the sequence $f = (f_n)$.

b) *The sequence* $\gamma^* = (\gamma_n^*)_{n\geqslant 0}$ *is the smallest supermartingale majorizing* $f = (f_n)_{n\geqslant 0}$.

c) *Let* $\tau_n^* = \inf\{k \geqslant n \colon f_k = \gamma_k^*\}$. *If* $\mathsf{E}\sup_n |f_n| < \infty$ *and* $\mathsf{P}(\tau_n^* < \infty) = 1$, *then* τ_n^* *is an optimal stopping time:*

$$V_n^* \equiv \sup_{\tau\in\mathfrak{M}_n^*} \mathsf{E} f_\tau = \mathsf{E} f_{\tau_n^*}, \tag{18}$$

$$\gamma_n^* \equiv \operatorname*{ess\,sup}_{\tau\in\mathfrak{M}_n^*} \mathsf{E}(f_\tau \mid \mathscr{F}_n) = \mathsf{E}(f_{\tau_n^*} \mid \mathscr{F}_n). \tag{19}$$

d) *For each* $n \geqslant 0$,

$$\gamma_n^N \uparrow \gamma_n^* \tag{20}$$

(P-*a.s.*) *as* $N \to \infty$.

5. We have already mentioned that for a finite time horizon ($N < \infty$) the optimal stopping problem can be solved by backward induction, by evaluating the variables $\gamma_N^N, \gamma_{N-1}^N, \dots, \gamma_0^N$ recursively. This is possible since $\gamma_N^N = f_N$ and the γ_n^N satisfy recursive relations (13).

On the other hand, if the time horizon is infinite ($N = \infty$), then the problem of finding the sequence $\gamma = (\gamma_n)_{n\geqslant 0}$ becomes more delicate: in place of an explicit formula describing the situation at time N we must use additional characteristics and properties of the prices V_n, $n \geqslant 0$. For example, one can sometimes seek the required solution of the system (17) using the observation that it must be the *smallest* of all solutions.

The techniques of solution of optimal stopping problems are better developed in the Markovian case.

For a description we assume that we have a *homogeneous Markov process* $X = (x_n, \mathscr{F}_n, \mathsf{P}_x)$ with discrete time $n = 0, 1, \dots$, with *phase* space $(E, \mathscr{B})$, and with probability measures P_x on $\mathscr{F} = \bigvee \mathscr{F}_n$ corresponding to each initial state $x \in E$ (see [126] and [441] for greater detail).

Let T be the one-step *transition operator* (i.e., $Tf(x) = \mathsf{E}_x f(x_1)$ for a measurable function $f = f(x_1)$ such that $\mathsf{E}_x|f(x_1)| < \infty$ for $x \in \mathsf{E}$, where E_x is averaging with respect to the measure P_x, $x \in E$).

Also, let $g = g(x)$ be some $\mathscr{B}$-measurable function of $x \in E$.

We set f_n (see the above discussion) to be equal to $g(x_n)$, $n \geqslant 0$, and assume that $\mathsf{E}_x[\sup g^-(x_n)] < \infty$, $x \in E$. We also set

$$s(x) = \sup_{\tau\in\mathfrak{M}^*} \mathsf{E}_x g(x_\tau), \tag{21}$$

where $\mathfrak{M}^* = \{\tau \colon \tau(\omega) < \infty,\ \omega \in \Omega\}$ is the class of all finite Markov times.

The optimal stopping problem for the Markov process X consists in finding the function $s(x)$ and optimal stopping times τ^* (satisfying the equality $s(x) = \mathsf{E}_x g(x_{\tau^*})$, $x \in E$) or ε-optimal times τ_ε^* (satisfying the relation $s(x) - \varepsilon \leqslant \mathsf{E}_x g(X_{\tau_\varepsilon^*})$, $x \in E$), provided that such times exist.

The advantages of the Markovian case are obvious: the above-considered conditional expectations $\mathsf{E}_x(\cdot \mid \mathscr{F}_n)$ depend on the 'history' $\mathscr{F}_n$ only through the value of the process at time n, i.e., through x_n. In particular, γ_n and γ_n^N are functions of x_n alone.

We present now a Markovian version of Theorems 1 and 2, which we shall use in what follows (in Chapter VI), e.g., in our analysis of American options.

THEOREM 3. *Let $g = g(x)$ be a $\mathscr{B}(\mathbb{R})$-measurable function with $\mathsf{E}_x g^-(x_k) < \infty$, $x \in E$, $k \leqslant N$, and let*

$$s_N(x) = \sup_{\tau \in \mathfrak{M}_0^N} \mathsf{E}_x g(x_\tau), \tag{22}$$

where $\mathfrak{M}_0^N = \{\tau\colon 0 \leqslant \tau \leqslant N\}$, $N \geqslant 0$.

Let

$$Qg(x) = \max\bigl(g(x), Tg(x)\bigr) \tag{23}$$

and let

$$\tau_0^N = \min\bigl\{0 \leqslant m \leqslant N\colon s_{N-m}(x_m) = g(x_m)\bigr\}. \tag{24}$$

Then

(a)

$$s_N(x) = Q^N g(x); \tag{25}$$

(b)

$$s_N(x) = \max\bigl(g(x), Ts_{N-1}(x)\bigr), \tag{26}$$

where $s_0(x) = g(x)$;

(c) *the Markov time τ_0^N is optimal in the class $\mathfrak{M}_0^N$:*

$$\mathsf{E}_x g\bigl(x_{\tau_0^N}\bigr) = s_N(x), \qquad x \in E; \tag{27}$$

(d) *the sequence $\gamma^N = (\gamma_m^N, \mathscr{F}_m)_{m \leqslant N}$ with $\gamma_m^N = s_{N-m}(x_m)$ is a supermartingale for each $N \geqslant 0$.*

Proof. This result and its generalizations can be found in Chapter 2 of the monograph [441] devoted to the Markovian approach to optimal stopping problems. Of course, Theorem 3 is also a consequence of Theorem 1 above, with the only exception: one must in any case study the structure of the operators $Qg(x)$ and their iterations $Q^n g(x)$. (See § 2.2 in [441] for a fuller account.)

It follows from this theorem that the optimal stopping times in the class $\mathfrak{M}_0^N = \{\tau\colon 0 \leqslant \tau \leqslant N\}$ have the following structure for fixed $N < \infty$.

Let

$$D_n^N = \{x\colon s_{N-n}(x) = g(x)\}, \qquad 0 \leqslant n \leqslant N, \tag{28}$$

and let

$$C_n^N = E \setminus D_n^N. \tag{29}$$

In accordance with Theorem 3 the optimal time τ_0^N can be expressed as follows:

$$\tau_0^N = \min\{0 \leqslant n \leqslant N : x_n \in D_n^N\}. \tag{30}$$

In other words, $D_0^N, D_1^N, \dots, D_N^N = E$ is a sequence of *stopping domains* while $C_0^N, C_1^N, \dots, C_N^N = \varnothing$ is a sequence of *domains of continued observations*.

Note that

$$D_0^N \subseteq D_1^N \subseteq \cdots \subseteq D_N^N = E$$

and

$$C_0^N \supseteq C_1^N \supseteq \cdots \supseteq C_N^N = \varnothing.$$

Hence no observations are carried out for $x_0 \in D_0^N$ and $\tau_0^N = 0$ in that case, whereas for $x_0 \in C_0^N$ one performs an observation and if $x_1 \in D_1^N$, then the observations are terminated. If $x_1 \in C_1^N$, then the next observation is performed, and so on. All observations stop at the terminal instant N (i.e., $D_N^N = E$).

Remark 2. It follows from Theorem 1 that the qualitative picture changes little if, in place of the price $s_N(x)$ defined by (21), we consider the prices

$$s_N(x) = \sup_{\tau \in \mathfrak{M}_0^N} \mathsf{E}_x \left[\beta^\tau g(x_\tau) - \sum_{k=1}^{\tau} c(x_{k-1})\right] \tag{21$'$}$$

with *discount* and *observation fees* (here $0 < \beta \leqslant 1$ and $c(x) \geqslant 0$ for $x \in E$; if $\tau = 0$, then the expression in $[\,\cdot\,]$ is set to be equal to $g(x)$).

Formula (25) holds good with

$$\mathsf{Q}g(x) = \max\bigl(g(x), \beta Tg(x) - c(x)\bigr), \tag{23$'$}$$

and recursive equation (26) takes the following form:

$$s_N(x) = \max\bigl(g(x), \beta T s_{N-1}(x) - c(x)\bigr), \tag{26$'$}$$

where $s_0(x) = g(x)$. See [441; Chapter 2] for greater detail.

EXAMPLE 1. Let $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots)$ be Bernoulli variables,

$$\mathsf{P}(\varepsilon_i = 1) = \mathsf{P}(\varepsilon_i = -1) = \frac{1}{2}.$$

Let $x_n = x + (\varepsilon_1 + \cdots + \varepsilon_n)$, where $x \in E = \{0, \pm 1, \dots\}$, and let

$$s_N(x) = \sup_{\tau \in \mathfrak{M}_0^N} \mathsf{E}_x \beta^\tau x_\tau.$$

If $\beta = 1$, then $\mathsf{Q}g(x) = g(x)$ for $g(x) = x$, $x \in E$, and we can take $\tau_0^N \equiv 0$ as an optimal stopping time.

On the other hand, if $0 < \beta < 1$, then for $g(x) = x$ we see that $\mathsf{Q}^n g(x) = x$ for $x = 0, 1, 2, \dots$ and $\mathsf{Q}^n g(x) = \beta^n x$ for $x = -1, -2, \dots$. Hence $s_N(x) = \max(x, \beta^N x)$ in this case. The optimal time is now

$$\tau_0^N = \min\{0 \leqslant n \leqslant N \colon x_n \in \{0, 1, 2, \dots\}\}.$$

(If $x_n \in \{-1, -2, \dots\}$ for all $0 \leqslant n \leqslant N$, then we set τ_0^N to be equal to N.)

Note that for $0 < \beta < 1$ we have

$$s_N(x) \uparrow x^+ = \max(0, x) \quad \text{as} \quad N \to \infty.$$

6. We consider now the optimal stopping problem for infinite horizon ($N = \infty$). We set

$$s(x) = \sup_{\tau \in \mathfrak{M}^*} \mathsf{E}_x g(x_\tau), \tag{31}$$

where $\mathfrak{M}^* = \{\tau \colon 0 \leqslant \tau(\omega) < \infty\}$ is the class of all finite Markov times.

To state the corresponding result on the structure of the *price* $s = s(x)$, $x \in E$, and the optimal (or ε-optimal) stopping times we recall the following definition.

DEFINITION (see, e.g., [441]). We call a function $f = f(x)$ such that $\mathsf{E}_x|f(x_1)| < \infty$, $x \in E$, and

$$f(x) \geqslant Tf(x) \tag{32}$$

an *excessive function* for the homogeneous Markov process $X = (x_n, \mathscr{F}_n, \mathsf{P}_x)_{n \geqslant 0}$, $x \in E$.

If, in addition, $f(x) \geqslant g(x)$, then we call $f = f(x)$ an *excessive majorant* of the function $g = g(x)$.

Clearly, if $f = f(x)$ is an excessive majorant of $g = g(x)$, then

$$f(x) \geqslant \max\big(g(x), Tf(x)\big). \tag{33}$$

The following theorem reveals the role of excessive majorants in optimal stopping problems for homogeneous Markov processes

$$X = (x_n, \mathscr{F}_n, \mathsf{P}_x)_{n \geqslant 0}, \qquad x \in E.$$

THEOREM 4. *Let $g = g(x)$ be a function such that $\mathsf{E}_x\Big[\sup_n g^-(x_n)\Big] < \infty$, $x \in E$. Then*

(a) *the price $s = s(x)$ is the smallest excessive majorant of $g = g(x)$;*

(b) $s(x) = \lim_{n\to\infty} Q^n g(x)$ $\Big(= \lim_{n\to\infty} s_n(x)\Big)$ *and* $s(x)$ *satisfies the equation*

$$s(x) = \max\bigl(g(x), Ts(x)\bigr) \tag{34}$$

(cf. (33));

(c) *if* $\mathsf{E}_x\Big[\sup_n |g(x_n)|\Big] < \infty$, $x \in E$, *then the time*

$$\tau_\varepsilon^* = \inf\{n \geqslant 0\colon s(x_n) \leqslant g(x_n) + \varepsilon\} \tag{35}$$

is ε*-optimal for each* $\varepsilon > 0$, *i.e.*,

$$s(x) - \varepsilon \leqslant \mathsf{E}_x g(x_{\tilde{\tau}_\varepsilon}), \qquad x \in E; \tag{36}$$

(d) *let* $\mathsf{E}_x\Big[\sup_n |g(x_n)|\Big] < \infty$ *and*

$$\tau^* = \inf\{n \geqslant 0\colon s(x_n) = g(x_n)\},$$

i.e., $\tau^* = \tau_0^*$; *if* $\mathsf{P}_x(\tau^* < \infty) = 1$, $x \in E$, *then* τ^* *is an optimal time*:

$$s(x) = \mathsf{E}_x g(x_{\tau^*}), \qquad x \in E; \tag{37}$$

(e) *if* E *is a finite set, then* τ^* *is an optimal time.*

As regards the proof of this theorem and its applications, see [441; Chapter 2]. In subsection 5 we discuss its applications to pricing of American options.

Remark 3. By analogy with (21′) we could consider also the optimal stopping problem in the case with discounting ($0 < \beta \leqslant 1$) and observation fee $c(x) \geqslant 0$.

We set

$$s(x) = \sup \mathsf{E}_x \left[\beta^\tau g(x_\tau) - \sum_{k=0}^{\tau-1} \beta^k c(x_k)\right] \tag{31′}$$

where we take the supremum over the class

$$\mathfrak{M}^*_{(\beta,c)} = \left\{\tau \in \mathfrak{M}^*\colon \mathsf{E}_x \sum_{k=0}^{\tau-1} \beta^k c(x_k) < \infty,\ x \in E\right\},$$

and assume that $g(x) \geqslant 0$.

Under these assumptions the price $s(x)$ is the smallest (β, c)-excessive majorant of $g(x)$ (see [441; Chapter 2]), i.e., the smallest function $f(x)$ such that $f(x) \geqslant g(x)$ and

$$f(x) \geqslant \beta T f(x) - c(x). \tag{33′}$$

Moreover,

$$s(x) = \max\bigl(g(x), \beta T s(x) - c(x)\bigr) \tag{34′}$$

and

$$s(x) = \lim_{n\to\infty} Q_{(\beta,c)} g(x),$$

where

$$Q_{(\beta,c)} g(x) = \max\bigl(g(x), \beta T g(x) - c(x)\bigr).$$

EXAMPLE 2. Let $x_n = x + (\varepsilon_1 + \cdots + \varepsilon_n)$, where $x \in E = \{0, \pm 1, \dots\}$ and $\varepsilon = (\varepsilon_n)$ is a Bernoulli sequence from Example 1. For $x \in E$ we set

$$s(x) = \sup \mathsf{E}_x(|x_\tau| - c\tau), \tag{38}$$

where we consider the supremum over all stopping times τ such that $\mathsf{E}_x \tau < \infty$. For such τ we have

$$\mathsf{E}_x x_\tau^2 = x^2 + \mathsf{E}_x \tau, \tag{39}$$

so that

$$\mathsf{E}_x(|x_\tau| - c\tau) = cx^2 + \mathsf{E}_x(|x_\tau| - c|x_\tau|^2). \tag{40}$$

Hence

$$s(x) = cx^2 + \sup \mathsf{E}_x g(x_\tau),$$

where $g(x) = |x| - c|x|^2$ and the supremum is taken over τ such that $\mathsf{E}_x \tau < \infty$.

Since $g(x)$ attains its maximum for $x = \pm \dfrac{1}{2c}$, it follows in the case when $\dfrac{1}{2c}$ is an integer that

$$s(x) \leqslant cx^2 + \frac{1}{4c}\,. \tag{41}$$

We now set $\tau_c = \inf\left\{n\colon |x_n| = \dfrac{1}{2c}\right\}$. If $|x| \leqslant \dfrac{1}{2c}$, then, of course, $|x_{\tau_c}| \leqslant \dfrac{1}{2c}$ and therefore

$$\frac{1}{(2c)^2} \geqslant \mathsf{E}_x x^2_{(\tau_c \wedge N)} = x^2 + \mathsf{E}_x(\tau_c \wedge N).$$

Passing to the limit as $N \to \infty$ (and using the monotonic convergence theorem) we see that $\mathsf{E}_x \tau_c \leqslant \dfrac{1}{(2c)^2} < \infty$. Hence τ_c satisfies (39), which shows that if $|x| \leqslant \dfrac{1}{2c}$, then we actually have $\mathsf{E}_x \tau_c = \dfrac{1}{(2c)^2} - x^2$.

Since

$$\mathsf{E}_x(|x_{\tau_c}| - c\tau_c) = \frac{1}{2c} - c\left(\frac{1}{(2c)^2} - x^2\right) = cx^2 + \frac{1}{4c}$$

for $|x| \leqslant \dfrac{1}{2c}$, it follows by (41) that τ_c is an optimal stopping time $\left(\text{for } |x| \leqslant \dfrac{1}{2c}\right)$.

§ 2b. Complete and Incomplete Markets. Supermartingale Characterization of Hedging Prices

1. We now return to the proof of formula (8) for hedge prices on incomplete markets exposed in § 1c.

As mentioned there, this proof is based on the following two facts: the *supermartingale property* of the sequence $Y = (Y_n)_{n \leqslant N}$ with respect to each measure in the family $\mathscr{P}(\mathsf{P})$ and the *optional decomposition* for $Y = (Y_n)_{n \leqslant N}$.

In the present section we consider the supermartingale property not only for the sequence $Y = (Y_n)_{n \leqslant N}$ defined by formulas (12) in § 1c, but also for a more general sequence defined by formula (1) below, which makes it possible to study American hedging (see the remark in § 1a).

The optional decomposition for $Y = (Y_n)_{n \leqslant N}$ is discussed in § 2d.

2. Let $(\Omega, \mathscr{F}, (\mathscr{F}_n)_{n \leqslant N}, \mathsf{P})$ be our underlying probability space and let (B, S) be a market formed by a bank account $B = (B_n)_{n \leqslant N}$ (for which we assume that $B_n \equiv 1$) and a d-dimensional stock $S = (S^1, \dots, S^d)$, $S^i = (S^i_n)_{n \leqslant N}$. Let $\mathscr{F}_0 = \{\varnothing, \Omega\}$ and let $\mathscr{F}_N = \mathscr{F}$.

Let $\mathscr{P}(\mathsf{P})$ be a non-empty set of martingale measures equivalent to P and let $f = (f_0, f_1, \dots, f_N)$ be a sequence of $\mathscr{F}_n$-measurable nonnegative functions f_n, $n \leqslant N$, such that $\mathsf{E}_{\widetilde{\mathsf{P}}} f_k < \infty$, $\widetilde{\mathsf{P}} \in \mathscr{P}(\mathsf{P})$, $0 \leqslant k \leqslant N$.

We set

$$Y_n = \operatorname*{ess\,sup}_{\widetilde{\mathsf{P}} \in \mathscr{P}(\mathsf{P}),\, \tau \in \mathfrak{M}_n^N} \mathsf{E}_{\widetilde{\mathsf{P}}}(f_\tau \mid \mathscr{F}_n). \tag{1}$$

THEOREM. *The sequence $Y = (Y_n)_{n \leqslant N}$ is a supermartingale with respect to each measure in $\mathscr{P}(\mathsf{P})$.*

Proof. Basically, this can be carried out along the same lines as the proof (in the preceding section) of the fact that the sequence $\gamma = (\gamma_n)_{n \leqslant N}$ is a supermartingale.

We can proceed as follows.

We choose some ('basic') measure in the set $\mathscr{P}(\mathsf{P})$. To avoid extra notation, let P be this measure (P is therefore assumed to be a martingale measure). We shall verify that Y is a supermartingale with respect to P.

If $\widetilde{\mathsf{P}} \in \mathscr{P}(\mathsf{P})$, then we set

$$\widetilde{Z}_N = \frac{d\widetilde{\mathsf{P}}}{d\mathsf{P}}, \quad \widetilde{Z}_n = \frac{d\widetilde{\mathsf{P}}_n}{d\mathsf{P}_n}, \quad \text{where} \quad \widetilde{\mathsf{P}}_n = \widetilde{\mathsf{P}} \mid \mathscr{F}_n, \quad \text{and} \quad \mathsf{P}_n = \mathsf{P} \mid \mathscr{F}_n.$$

For $n = 0$ we set $\widetilde{Z}_0 = 1$.

Let

$$\widetilde{\rho}_n = \frac{\widetilde{Z}_n}{\widetilde{Z}_{n-1}}. \tag{2}$$

Since $\widetilde{\mathsf{P}} \sim \mathsf{P}$, it follows that

$$\mathsf{P}(\widetilde{Z}_{n-1} > 0) = \widetilde{\mathsf{P}}(\widetilde{Z}_{n-1} > 0) = 1$$

for each $n \leqslant N$.

Setting $\widetilde{m}_n = \widetilde{\rho}_n - 1$, $\widetilde{M}_n = \sum_{k=1}^{n} \widetilde{m}_k$, and $\widetilde{M}_0 = 0$ we have

$$\Delta \widetilde{Z}_n = \widetilde{Z}_{n-1}\, \Delta \widetilde{M}_n. \tag{3}$$

From (3) we can see that

$$\widetilde{Z}_n = \mathscr{E}(\widetilde{M})_n \equiv \prod_{k=1}^{n} (1 + \Delta \widetilde{M}_k) = \prod_{k=1}^{n} \widetilde{\rho}_k, \tag{4}$$

where $\mathscr{E}(\widetilde{M})$ is the stochastic exponential (see Chapter II, § 1a).

It follows from the above that having chosen P to be the 'basic' measure we can completely characterize $\widetilde{\mathsf{P}}$ and its restrictions $\widetilde{\mathsf{P}}_n$, $n \leqslant N$, by one of the sequences $(\widetilde{Z}_n)$, $(\widetilde{M}_n)$, or $(\widetilde{\rho}_n)$.

By *Bayes's formula* ((4) in Chapter V, § 3a), for each stopping time τ (with respect to $(\mathscr{F}_n)$) and $n \leqslant N$ we have

$$\begin{aligned} \mathsf{E}_{\widetilde{\mathsf{P}}}(f_\tau \mid \mathscr{F}_n) &= \frac{1}{\widetilde{Z}_n} \mathsf{E}_{\mathsf{P}}(f_\tau \widetilde{Z}_\tau \mid \mathscr{F}_n) \\ &= \mathsf{E}_{\mathsf{P}}(\widetilde{\rho}_{n+1} \cdots \widetilde{\rho}_\tau f_\tau \mid \mathscr{F}_n) \\ &= \mathsf{E}_{\mathsf{P}}(\overline{\rho}_1 \cdots \overline{\rho}_n \cdot \overline{\rho}_{n+1} \cdots \overline{\rho}_\tau f_\tau \mid \mathscr{F}_n) \\ &= \mathsf{E}_{\mathsf{P}}(f_\tau \overline{Z}_\tau \mid \mathscr{F}_n), \end{aligned} \tag{5}$$

where $\overline{\rho}_1 = \cdots = \overline{\rho}_n = 1$, $\overline{\rho}_k = \widetilde{\rho}_k$, $k > n$, and $\overline{Z}_k = \overline{\rho}_1 \cdots \overline{\rho}_k$.

It is interesting to note that defining a measure $\overline{\mathsf{P}}$ by the equality $d\overline{\mathsf{P}} = \overline{Z}_N \, d\mathsf{P}$ we obtain

$$\overline{\mathsf{P}}(A) = \begin{cases} \mathsf{P}(A), & A \in \mathscr{F}_k, \ k \leqslant n, \\ \widetilde{\mathsf{P}}(A), & A \in \mathscr{F}_k, \ k > n. \end{cases} \tag{6}$$

Clearly, $\overline{\mathsf{P}} \sim \mathsf{P}$.

In view of our notation we can rewrite the definition of (1) as follows:

$$Y_n = \operatorname{ess\,sup} \mathsf{E}_{\mathsf{P}}(f_\tau \overline{Z}_\tau \mid \mathscr{F}_n),$$

where ess sup is taken over the class $\mathfrak{M}_n^N$ of stopping times τ such that $n \leqslant \tau \leqslant N$, and over P-martingales in the class $\mathscr{Z}_n^N$ of positive martingales $\overline{Z} = (\overline{Z}_k)_{k \leqslant N}$ such that $\overline{Z}_0 = \cdots = \overline{Z}_n = 1$ or, equivalently, $\overline{Z}_0 = \overline{\rho}_1 = \cdots = \overline{\rho}_n = 1$.

The sets with $k \leqslant N$ obviously satisfy the relations

$$\mathfrak{M}_k^N \subseteq \mathfrak{M}_{k-1}^N \quad \text{and} \quad \mathscr{Z}_k^N \subseteq \mathscr{Z}_{k-1}^N,$$

which play an important role in the proof of the supermartingale property of the sequence $Y = (Y_n, \mathscr{F}_n)_{n \leqslant N}$.

By the definition of the essential supremum (see, e.g., [75; Chapter 1]), there exist sequences of times $\tau^{(i)}$ and martingales $\overline{Z}^{(i)}$ belonging to the classes $\mathfrak{M}_k^N$ and $\mathscr{Z}_k^N$, respectively, such that

$$\operatorname*{ess\,sup}_{\tau \in \mathfrak{M}_k^N, \, \overline{Z} \in \mathscr{Z}_k^N} \mathsf{E}(f_\tau \overline{Z}_\tau \mid \mathscr{F}_k) = \lim_i \uparrow \mathsf{E}\big(f_{\tau^{(i)}} \overline{Z}^{(i)}_{\tau^{(i)}} \mid \mathscr{F}_k\big), \tag{7}$$

where $\lim\limits_{i}\uparrow$ is the limit of an increasing sequence.

Hence, by the monotone convergence theorem

$$\begin{aligned}
\mathsf{E}_{\mathsf{P}}(Y_k \mid \mathscr{F}_{k-1}) &= \mathsf{E}_{\mathsf{P}}\Big(\operatorname*{ess\,sup}_{\tau \in \mathfrak{M}_k^N,\, \overline{Z} \in \mathscr{Z}_k^N} \mathsf{E}(f_\tau \overline{Z}_\tau \mid \mathscr{F}_k) \;\Big|\; \mathscr{F}_{k-1}\Big) \\
&= \mathsf{E}_{\mathsf{P}}\Big(\lim_{i}\uparrow \mathsf{E}_{\mathsf{P}}\big(f_{\tau^{(i)}} \overline{Z}^{(i)}_{\tau^{(i)}} \mid \mathscr{F}_k\big) \;\Big|\; \mathscr{F}_{k-1}\Big) \\
&= \lim_{i}\uparrow \mathsf{E}_{\mathsf{P}}\big(f_{\tau^{(i)}} \overline{Z}^{(i)}_{\tau^{(i)}} \mid \mathscr{F}_{k-1}\big) \\
&\leqslant \operatorname*{ess\,sup}_{\tau \in \mathfrak{M}_k^N,\, \overline{Z} \in \mathscr{Z}_k^N} \mathsf{E}_{\mathsf{P}}(f_\tau \overline{Z}_\tau \mid \mathscr{F}_{k-1}) \\
&\leqslant \operatorname*{ess\,sup}_{\tau \in \mathfrak{M}_{k-1}^N,\, \overline{Z} \in \mathscr{Z}_{k-1}^N} \mathsf{E}_{\mathsf{P}}(f_\tau \overline{Z}_\tau \mid \mathscr{F}_{k-1}) = Y_{k-1},
\end{aligned}$$

which is just the required supermartingale property.

Remark. We can extend Theorem 1 to the *control with stopping* case, with f_τ replaced by a functional $\sum\limits_{k=1}^{\tau} \alpha_k \, \Delta g_k + f_\tau$, where $g = (g_0, g_1, \dots, g_N)$ is a sequence of $\mathscr{F}_n$-measurable functions g_n and $\alpha = (\alpha_1, \dots, \alpha_N)$ is a control belonging to a sufficiently *rich* class of predictable sequences. (See § 1a.2 in this connection.)

§ 2c. Complete and Incomplete Markets. Main Formulas for Hedging Prices

1. By the main formula of the price $\mathbb{C}^*(f_N; \mathsf{P})$ of European hedging on an arbitrage-free (B, S)-market (§ 1c),

$$\mathbb{C}^*(f_N; \mathsf{P}) = \sup_{\widetilde{\mathsf{P}} \in \mathscr{P}(\mathsf{P})} B_0 \, \mathsf{E}_{\widetilde{\mathsf{P}}} \frac{f_N}{B_N}. \tag{1}$$

We proceed now to a more complex financial instrument, *American hedging* on a (B, S)-market. We shall assume here that *the set of martingale measures* $\mathscr{P}(\mathsf{P})$ *is non-empty.*

We have mentioned on several occasions that we must often (e.g., in the study of American options) consider in place of a single pay-off function f_N an entire system of functions $f = (f_n)_{n \leqslant N}$ interpreted as follows. If the buyer exercises the option at time n, then the corresponding amount (payable by the writer) is described by the $\mathscr{F}_n$-measurable function $f_n = f_n(\omega)$.

Of course, the writer must choose only strategies π of value $X^\pi = (X_n^\pi)_{n \leqslant N}$ satisfying the *hedging condition*

$$X_\tau^\pi \geqslant f_\tau \quad (\mathsf{P}\text{-a.s.}), \tag{2}$$

which guarantees his ability to meet the contract terms for *each* stopping time $\tau = \tau(\omega)$ that can be chosen by the buyer to exercise the contract.

2. For more precise formulations of the corresponding problems we introduce now several definitions.

Following § 2a we set

$$\mathfrak{M}_n^N = \{\tau = \tau(\omega) \colon n \leqslant \tau(\omega) \leqslant N,\ \omega \in \Omega\}.$$

If $\pi = (\beta = (\beta_n)_{n \leqslant N}, \gamma = (\gamma_n)_{n \leqslant N})$ is a securities portfolio of value

$$X_n^\pi = \beta_n B_n + \gamma_n S_n, \qquad n \leqslant N, \tag{3}$$

then we shall understand its *self-financing* property in the sense of the following balance condition:

$$\Delta X_n^\pi = \beta_n \Delta B_n + \gamma_n \Delta S_n - \Delta C_n \tag{4}$$

where $C = (C_n)_{n \geqslant 0}$ is a non-negative process, $C_0 = 0$, and C_n is $\mathscr{F}_n$-measurable. (Cf. the 'consumption' case in Chapter V, § 1a).

To point out the dependence of the capital X_n^π on the 'consumption' we shall denote it by $X_n^{\pi,C}$ (as in § 1c).

DEFINITION 1. By the *upper price of American hedging* (against a system of $\mathscr{F}_n$-measurable payment functions f_n, $n \leqslant N$) we mean the quantity

$$\widetilde{\mathbb{C}}(f_N; \mathsf{P}) = \inf\{x \colon \exists\,(\pi, C) \text{ with } X_0^{\pi,C} = x \text{ with } X_\tau^{\pi,C} \geqslant f_\tau \ (\mathsf{P}\text{-a.s.})\ \forall\,\tau \in \mathfrak{M}_0^N\}. \tag{5}$$

DEFINITION 2. We say that a strategy (π, C) is *perfect* if $X_n^{\pi,C} \geqslant f_n$ for each $n \leqslant N$ and $X_N^{\pi,C} = f_N$ (P-a.s.).

THEOREM 1 (*Main formula for American hedging price*). *Assume that $\mathscr{P}(\mathsf{P}) \neq \varnothing$ and let $f = (f_n)_{n \leqslant N}$ be a sequence of non-negative payment functions such that*

$$\sup_{\widetilde{\mathsf{P}} \in \mathscr{P}(\mathsf{P})} \mathsf{E}_{\widetilde{\mathsf{P}}} \frac{f_n}{B_n} < \infty, \qquad n \leqslant N. \tag{6}$$

Then the upper price of American hedging is

$$\boxed{\widetilde{\mathbb{C}}(f_N; \mathsf{P}) = \sup_{\widetilde{\mathsf{P}} \in \mathscr{P}(\mathsf{P}),\, \tau \in \mathfrak{M}_0^N} B_0\, \mathsf{E}_{\widetilde{\mathsf{P}}} \frac{f_\tau}{B_\tau}\,.} \tag{7}$$

Proof. We have already made necessary preparatory work for the proof of this formula.

As in the case of European hedging we prove first that

$$\sup_{\widetilde{\mathsf{P}} \in \mathscr{P}(\mathsf{P}),\, \tau \in \mathfrak{M}_0^N} B_0\, \mathsf{E}_{\widetilde{\mathsf{P}}} \frac{f_\tau}{B_\tau} \leqslant \widetilde{\mathbb{C}}(f_N; \mathsf{P}). \tag{8}$$

If the set of hedges is empty, then $\widetilde{\mathbb{C}}(f_N; \mathsf{P}) = \infty$ (by Definition 1). and (8) is obvious.

Now let π be a self-financing hedge (with consumption C) such that $X_0^{\pi,C} = x < \infty$.

By analogy with (9), we see from § 1c that

$$\frac{f_\tau}{B_\tau} \leqslant \frac{X_\tau^{\pi,C}}{B_\tau} = \frac{x}{B_0} + \sum_{k=1}^{\tau} \gamma_k \, \Delta\left(\frac{S_k}{B_k}\right) - \sum_{k=1}^{\tau} \frac{\Delta C_k}{B_{k-1}} \tag{9}$$

for each $\tau \in \mathfrak{M}_0^N$. In particular,

$$\sum_{k=1}^{N} \gamma_k \, \Delta\left(\frac{S_k}{B_k}\right) \geqslant -\frac{x}{B_0}\,.$$

Hence we obtain by assertion 2) of the lemma in Chapter II, § 1c that the sequence

$$\left(\sum_{k=1}^{N} \gamma_k \, \Delta\left(\frac{S_k}{B_k}\right)\right)_{n \leqslant N}$$

is a martingale with respect to each measure $\widetilde{\mathsf{P}} \in \mathscr{P}(\mathsf{P})$.

Applying Doob's stopping theorem (see Chapter V, § 3a) we see from (9) that

$$\sup_{\widetilde{\mathsf{P}} \in \mathscr{P}(\mathsf{P})} \mathsf{E}_{\widetilde{\mathsf{P}}} \frac{f_\tau}{B_\tau} \leqslant \frac{x}{B_0}\,, \tag{10}$$

and therefore (8) holds.

The proof of the reverse inequality to (8) is more complicated. It is sufficient to find a portfolio $\widetilde{\pi} = (\widetilde{\beta}, \widetilde{\gamma})$ and 'consumption' $\widetilde{C}$ such that the capital $X^{\widetilde{\pi},\widetilde{C}}$ satisfies the 'balance' conditions

$$\Delta X_n^{\widetilde{\pi},\widetilde{C}} = \widetilde{\beta}_n \, \Delta B_n + \widetilde{\gamma}_n \, \Delta S_n - \Delta \widetilde{C}_n, \qquad n \leqslant N, \tag{11}$$

the starting capital is

$$X_0^{\widetilde{\pi},\widetilde{C}} = \sup_{\widetilde{\mathsf{P}} \in \mathscr{P}(\mathsf{P}),\, \tau \in \mathfrak{M}_0^N} B_0 \, \mathsf{E}_{\widetilde{\mathsf{P}}} \frac{f_\tau}{B_\tau}\,, \tag{12}$$

and (P-a.s.)

$$X_\tau^{\widetilde{\pi},\widetilde{C}} \geqslant f_\tau, \quad \forall\, \tau \in \mathfrak{M}_0^N. \tag{13}$$

To this end we consider a sequence $\widetilde{Y} = (\widetilde{Y}_n)_{n\leqslant N}$ with

$$\widetilde{Y}_n = \operatorname*{ess\,sup}_{\widetilde{\mathsf{P}}\in\mathscr{P}(\mathsf{P}),\,\tau\in\mathfrak{M}_0^N} \mathsf{E}_{\widetilde{\mathsf{P}}}\left(\frac{f_\tau}{B_\tau}\,\Big|\,\mathscr{F}_n\right). \tag{14}$$

It follows by the theorem in § 2b that $\widetilde{Y} = (\widetilde{Y}_n)_{n\leqslant N}$ is a *supermartingale* with respect to each measure $\widetilde{\mathsf{P}} \in \mathscr{P}(\mathsf{P})$. On the other hand it follows from § 2d that the supermartingale $\widetilde{Y} = (\widetilde{Y}_n)_{n\leqslant N}$ has the *optional* decomposition (holding $\widetilde{\mathsf{P}}$-a.s. for each measure $\widetilde{\mathsf{P}} \in \mathscr{P}(\mathsf{P})$)

$$\widetilde{Y}_n = \widetilde{Y}_0 + \sum_{k=1}^n \widetilde{\gamma}_k\,\Delta\left(\frac{S_n}{B_n}\right) - \sum_{k=1}^n \frac{\Delta\widetilde{C}_k}{B_{k-1}} \tag{15}$$

with some predictable sequence $\widetilde{\gamma} = (\widetilde{\gamma}_n)_{n\leqslant N}$ and non-negative sequence $\widetilde{C} = (\widetilde{C}_n)_{n\leqslant N}$ such that $\widetilde{C}_0 = 0$ and the $\widetilde{C}_n$ are $\mathscr{F}_n$-measurable.

Finding $\widetilde{\gamma}$ and $\widetilde{C}$ from the decomposition (15) we define $\widetilde{\beta} = (\widetilde{\beta}_n)_{n\leqslant N}$ by setting

$$\widetilde{\beta}_n = \widetilde{Y}_n - \widetilde{\gamma}_n \frac{S_n}{B_n}\,. \tag{16}$$

The value of the strategy $(\widetilde{\pi}, \widetilde{C})$ is

$$X_n^{\widetilde{\pi},\widetilde{C}} = \widetilde{\beta}_n B_n + \widetilde{\gamma}_n S_n, \tag{17}$$

and 'balance' condition (11) is satisfied in view of (15). Since $X_n^{\widetilde{\pi},\widetilde{C}} = B_n\widetilde{Y}_n$, the capital $X_n^{\widetilde{\pi},\widetilde{C}}$ has the following representation by (14):

$$X_n^{\widetilde{\pi},\widetilde{C}} = \operatorname*{ess\,sup}_{\widetilde{\mathsf{P}}\in\mathscr{P}(\mathsf{P}),\,\tau\in\mathfrak{M}_n^N} B_n\mathsf{E}_{\widetilde{\mathsf{P}}}\left(\frac{f_\tau}{B_\tau}\,\Big|\,\mathscr{F}_n\right). \tag{18}$$

We conclude from this formula that

1) the initial capital of the strategy $(\widetilde{\pi}, \widetilde{C})$ is defined by (12);
2) $(\widetilde{\pi}, \widetilde{C})$ is a *hedging* strategy in the following sense: $X_n^{\widetilde{\pi},\widetilde{C}} \geqslant f_n$ for $n \leqslant N$, or, equivalently, (13) is satisfied;
3) $(\widetilde{\pi}, \widetilde{C})$ has the following *replication* property: $X_N^{\widetilde{\pi},\widetilde{C}} = f_N$ (P-a.s.).

The proof is complete and, on the way, we have established also the following result (cf. Theorem 2 in § 1c).

THEOREM 2 (*Main formulas for hedging, consumption, and capital*). *Under the assumptions of Theorem 1 there exists a hedge* $\widetilde{\pi} = (\widetilde{\beta}, \widetilde{\gamma})$ *and consumption* $\widetilde{C}$ *such that the dynamics of the corresponding capital* $X_n^{\widetilde{\pi},\widetilde{C}} = \widetilde{\beta}_n B_n + \widetilde{\gamma}_n S_n$ *satisfies 'balance' conditions* (11). *Moreover,* $X_0^{\widetilde{\pi},\widetilde{C}}$ *satisfies* (12), *the dynamics of* $X_n^{\widetilde{\pi},\widetilde{C}}$ *is determined by* (18), *the components of* $\widetilde{\gamma} = (\widetilde{\gamma}_n)$ *and consumption* $\widetilde{C} = (\widetilde{C}_n)$ *can be found from the optional decomposition* (15), *and* $\widetilde{\beta} = (\widetilde{\beta}_n)$ *can be found from* (16).

4. In connection with the assumption '$\mathscr{P}(\mathsf{P}) \neq \varnothing$' in Theorems 1 and 2 of this section and the assumption of the absence of arbitrage in Theorem 1 and 2 of § 1b we can make the following observation (taking the hedging of option contracts as an example).

The standard definition of the *absence of arbitrage* (see Definition 2 in Chapter V, § 2a) relates to some particular instant N, e.g., the maturity date of a European option.

On the other hand, in dealing with American options one must consider in place of a single instant N an entire class $\mathfrak{M}_0^N$ of stopping times τ. For that reason, in place of the assumption '$\mathscr{P}(\mathsf{P}) \neq \varnothing$' in Theorems 1 and 2, it would be more logical to assume that the market is *arbitrage-free in the strong sense* (see Definition 3 in Chapter V, § 2a).

By the *extended* version of the *First fundamental theorem* (Chapter V, § 1e) the 'arbitrage-free' state (whether in the strong sense or not) is equivalent in our case of discrete time ($n \leqslant N < \infty$) and finitely many stocks ($d < \infty$) to the condition $\mathscr{P}(\mathsf{P}) \neq \varnothing$.

The reasons why we have actually put this condition in the form $\mathscr{P}(\mathsf{P}) \neq \varnothing$ in the statement of the theorem are as follows. First, this property can often be verified (even in the continuous-time case); second, the term 'arbitrage-free in the strong sense' is not widely accepted and the question of the equivalence of different interpretations of the 'absence of arbitrage' and the condition $\mathscr{P}(\mathsf{P}) \neq \varnothing$ is not always that simple, especially, in the continuous-time case. (See Chapter VII, § § 1a, c in this connection.)

5. As regards the writer of an American option, he must first of all choose a strategy (π, C) enabling him to meet the terms of the contract. This imposes the following constraint on the capital $X^{\pi,C}$: the 'hedging condition' $X_\tau^{\widetilde{\pi},\widetilde{C}} \geqslant f_\tau$ must be satisfied (P-a.s.) for each $\tau \in \mathfrak{M}_0^N$.

Now, there arises the natural question on the particular instant $\tau = \tau(\omega)$ at which it would be *reasonable* for the buyer to exercise the contract.

We shall consider the case of a *complete arbitrage-free* market, which in the present context of discrete time ($n \leqslant N < \infty$) and finitely many stocks ($d < \infty$) is equivalent to the existence of a unique martingale measure $\widetilde{\mathsf{P}}$ (the *Second fundamental theorem*).

Before answering this question we reformulate and somewhat improve on Theorems 1 and 2 in the case under consideration.

THEOREM 3. *Let $\widetilde{\mathsf{P}}$ be a unique martingale measure ($|\mathscr{P}(\mathsf{P})| = 1$) and let $f = (f_n)_{n\leqslant N}$ be a sequence of non-negative pay-off functions with $\mathsf{E}_{\widetilde{\mathsf{P}}} \dfrac{f_n}{B_n} < \infty$, $n \leqslant N$. Then we have the following results.*

1) *The upper price is*

$$\boxed{\widetilde{\mathbb{C}}(f_N;\mathsf{P}) = \sup_{\tau\in\mathfrak{M}_0^N} B_0\,\mathsf{E}_{\widetilde{\mathsf{P}}}\frac{f_\tau}{B_\tau}\,.} \tag{19}$$

2) *There exists a self-financing strategy $(\widetilde{\pi},\widetilde{C})$ such that the corresponding capital $X^{\widetilde{\pi},\widetilde{C}}$ satisfies the 'balance' condition*

$$\Delta X_n^{\widetilde{\pi},\widetilde{C}} = \widetilde{\beta}_n\,\Delta B_n + \widetilde{\gamma}_n\,\Delta - \Delta\widetilde{C}_n; \tag{20}$$

$$X_0^{\widetilde{\pi},\widetilde{C}} = \sup_{\tau\in\mathfrak{M}_0^N} B_0\,\mathsf{E}_{\widetilde{\mathsf{P}}}\frac{f_\tau}{B_\tau} \quad (= \widetilde{\mathbb{C}}^*(f_N;\mathsf{P})); \tag{21}$$

$$X_\tau^{\widetilde{\pi},\widetilde{C}} \geqslant f_\tau, \quad \forall\,\tau\in\mathfrak{M}_0^N. \tag{22}$$

The dynamics of $X^{\widetilde{\pi},\widetilde{C}}$ is described by the formulas

$$X_n^{\widetilde{\pi},\widetilde{C}} = B_n \operatorname*{ess\,sup}_{\tau\in\mathfrak{M}_n^N} \mathsf{E}_{\widetilde{\mathsf{P}}}\biggl(\frac{f_\tau}{B_\tau}\biggm|\mathscr{F}_n\biggr). \tag{23}$$

3) *The components $\widetilde{\gamma} = (\widetilde{\gamma}_n)$ and $\widetilde{C} = (\widetilde{C}_n)$ can be determined from the Doob decomposition for the supermartingale $\widetilde{Y} = (\widetilde{Y}_n,\mathscr{F}_n,\widetilde{\mathsf{P}})_{n\leqslant N}$ with*

$$\widetilde{Y}_n = \operatorname*{ess\,sup}_{\tau\in\mathfrak{M}_n^N} \mathsf{E}_{\widetilde{\mathsf{P}}}\biggl(\frac{f_\tau}{B_\tau}\biggm|\mathscr{F}_n\biggr),$$

which has the following form:

$$\widetilde{Y}_n = \widetilde{Y}_0 + \sum_{k=1}^{n}\widetilde{\gamma}_k\,\Delta\biggl(\frac{S_k}{B_k}\biggr) - \widetilde{C}_n. \tag{24}$$

The components $\widetilde{\beta} = (\widetilde{\beta}_n)_{n\leqslant N}$ are defined by the relations

$$\widetilde{\beta}_n = \widetilde{Y}_n - \widetilde{\gamma}_n\frac{S_n}{B_n}\,. \tag{25}$$

4) *In the optimal stopping problem of finding*

$$\sup_{\tau\in\mathfrak{M}_0^N} \mathsf{E}_{\widetilde{\mathsf{P}}}\frac{f_\tau}{B_\tau}\,, \tag{26}$$

the stopping time

$$\widetilde{\tau} = \min\left\{0 \leqslant n \leqslant N : Y_n = \frac{f_n}{B_n}\right\} \tag{27}$$

is optimal, i.e.,

$$\sup_{\tau \in \mathfrak{M}_0^N} \mathsf{E}_{\widetilde{\mathsf{P}}} \frac{f_\tau}{B_\tau} = \mathsf{E}_{\widetilde{\mathsf{P}}} \frac{f_{\widetilde{\tau}}}{B_{\widetilde{\tau}}}. \tag{28}$$

Moreover,

$$X_{\widetilde{\tau}}^{\widetilde{\pi},\widetilde{C}} = f_{\widetilde{\tau}} \quad (\mathsf{P}\text{-a.s.}), \tag{29}$$

and the sequence $(\widetilde{Y}_n)_{n \leqslant N}$ is the smallest $\widetilde{\mathsf{P}}$-supermartingale majorizing $(f_n)_{n \leqslant N}$.

Proof. Assertions 1)–3) follow from Theorems 1 and 2. It should be noted only that, since the martingale measure is unique, we need not refer to optional decompositions; it suffices to use directly the *Doob decomposition*

$$\widetilde{Y}_n = \widetilde{M}_n - \widetilde{C}_n \tag{30}$$

of the supermartingale $\widetilde{Y} = (\widetilde{Y}_n, \mathscr{F}_n, \widetilde{\mathsf{P}})_{n \leqslant N}$ (see Chapter III, § 1b).

By the *extended* version of the *Second fundamental theorem* (Chapter V, § 4f) the martingale $\widetilde{M} = (\widetilde{M}_n, \mathscr{F}_n, \widetilde{\mathsf{P}})$ can be represented as follows:

$$\widetilde{M}_n = \widetilde{Y}_0 + \sum_{k=1}^{n} \widetilde{\gamma}_k \, \Delta\left(\frac{S_k}{B_k}\right). \tag{31}$$

Together with (30), this brings us to the required representation (24).

As regards assertion 4), this is a special case of Corollaries 1 and 3 to Theorem 1 in § 2a.

6. We now proceed directly to the question whether (judging on the basis of the information contained in the flow $(\mathscr{F}_n)$) it would be 'reasonable' of the buyer to choose the exercise time $\widetilde{\tau}$.

Both buyer and writer operate with the understanding that the option price $\widetilde{\mathbb{C}}(f_N; \mathsf{P})$ defined by (19) is *mutually acceptable.* (See Chapter V, § 1b.4.)

We consider now all strategies $(\overline{\pi}, \overline{C})$ with initial capital $X_0^{\overline{\pi},\overline{C}} = \widetilde{\mathbb{C}}(f_N; \mathsf{P})$ that bring about hedging, i.e., strategies such that $X_n^{\overline{\pi},\overline{C}} \geqslant f_n$ for $n \leqslant N$. We denote the class of these strategies by $\prod(\widetilde{\mathbb{C}}(f_N; \mathsf{P}))$.

This class contains a strategy $(\widetilde{\pi}, \widetilde{C})$ of the *minimum* value, such that

$$f_n \leqslant X_n^{\widetilde{\pi},\widetilde{C}} \leqslant X_n^{\overline{\pi},\overline{C}}, \qquad n \leqslant N, \tag{32}$$

for each $(\overline{\pi}, \overline{C}) \in \prod(\widetilde{\mathbb{C}}(f_N; \mathsf{P}))$. For it follows from the 'balance' conditions that $\overline{Y}_n = \dfrac{X_n^{\overline{\pi},\overline{C}}}{B_n}$, $n \leqslant N$, is a $\widetilde{\mathsf{P}}$-*supermartingale* majorizing $\dfrac{f_n}{B_n}$, $n \leqslant N$, while it follows from assertion 4) in Theorem 3 that the sequence $\widetilde{Y}_n$, $n \leqslant N$, is the *smallest* $\widetilde{\mathsf{P}}$-supermartingale majorizing $\dfrac{f_n}{B_n}$, $n \leqslant N$.

Hence $\widetilde{Y}_n \leqslant \overline{Y}_n$ for $n \leqslant N$. Together with the relation

$$\frac{f_n}{B_n} \leqslant \widetilde{Y}_n = \frac{X_n^{\widetilde{\pi},\widetilde{C}}}{B_n},$$

this proves inequalities (32).

These inequalities show that for each stopping time τ we have

$$f_\tau \leqslant X_\tau^{\widetilde{\pi},\widetilde{C}} \leqslant X_\tau^{\overline{\pi},\overline{C}}. \tag{33}$$

Clearly, the buyer must chose τ such that for no strategy $(\overline{\pi}, \overline{C}) \in \prod(\widetilde{\mathbb{C}}(f_N; \mathsf{P}))$ the writer would get profits $X_\tau^{\overline{\pi},\overline{C}} - f_\tau > 0$ with positive probability. In other words, the buyer may consider only the stopping times τ such that

$$f_\tau = X_\tau^{\overline{\pi},\overline{C}} \quad (\mathsf{P}\text{-a.s.}), \quad \forall\, (\overline{\pi}, \overline{C}) = \prod(\widetilde{\mathbb{C}}(f_N; \mathsf{P})). \tag{34}$$

All this justifies the following definition.

DEFINITION. Stopping times τ satisfying (34) are called *rational* exercise times.

THEOREM 4. *Each stopping time τ that solves the optimal stopping problem* (26) *(i.e., each stopping time satisfying* (28)*) is a rational exercise time.*

Proof. Let $(\overline{\pi}, \overline{C}) = \prod(\widetilde{\mathbb{C}}(f_N; \mathsf{P}))$. Then, in view of the $\widetilde{\mathsf{P}}$-supermartingale property of the sequence $\overline{Y}_n = \dfrac{X_n^{\overline{\pi},\overline{C}}}{B_n}$, $n \leqslant N$, we see that

$$\begin{aligned}\widetilde{\mathbb{C}}(f_N; \mathsf{P}) = X_0^{\overline{\pi},\overline{C}} &\geqslant B_0\, \mathsf{E}_{\widetilde{\mathsf{P}}} \frac{X_\tau^{\overline{\pi},\overline{C}}}{B_\tau} \geqslant B_0\, \mathsf{E}_{\widetilde{\mathsf{P}}} \frac{f_\tau}{B_\tau} \\ &= B_0 \sup_{\tau \in \mathfrak{M}_0^N} \mathsf{E}_{\widetilde{\mathsf{P}}} \frac{f_\tau}{B_\tau} = \widetilde{\mathbb{C}}(f_N; \mathsf{P}).\end{aligned}$$

Hence

$$\mathsf{E}_{\widetilde{\mathsf{P}}} \frac{X_\tau^{\overline{\pi},\overline{C}}}{B_\tau} = \mathsf{E}_{\widetilde{\mathsf{P}}} \frac{f_\tau}{B_\tau},$$

which, coupled with the property $X_\tau^{\overline{\pi},\overline{C}} \geqslant f_\tau$, proves that, in fact, $X_\tau^{\overline{\pi},\overline{C}} = f_\tau$ (P-a.s.), i.e., τ is a rational time.

Remark. It may be useful to repeat at this point that a solution of the optimal stopping problem (26) provides one with both value of the *rational price* $\widetilde{\mathbb{C}}(f_N; \mathsf{P})$ and *rational* exercise time. Usually, one cannot find $\widetilde{\mathbb{C}}(f_N; \mathsf{P})$ or $\widetilde{\tau}$ *separately*; they can be found only *in tandem*, by solving (26).

§ 2d. Optional Decomposition

1. For a filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_n)_{n\leqslant N}, \mathsf{P})$ with $\mathscr{F}_0 = \{\varnothing, \Omega\}$, let $X = (X_n)_{n\leqslant N}$ be a real-valued process and let $S = (S_n)_{n\leqslant N}$, where $S_n = (S_n^1, \dots, S_n^d)$, be a $\mathbb{R}^d$-valued process that are both adapted to $(\mathscr{F}_n)_{n\leqslant N}$: the X_n and S_n^i must be $\mathscr{F}_n$-measurable for $n \leqslant N$ and $i = 1, \dots, d$.

Let $\mathscr{P}(\mathsf{P})$ be the set of probability measures $\widetilde{\mathsf{P}}$ on $(\Omega, \mathscr{F})$ such that $\widetilde{\mathsf{P}} \sim \mathsf{P}$ and S is a $\widetilde{\mathsf{P}}$-*martingale*. We assume that $\mathscr{P}(\mathsf{P}) \neq \varnothing$.

As regards the process $X = (X_n)_{n\leqslant N}$, our main assumption is that it is a *supermartingale* with respect to each measure $\widetilde{\mathsf{P}} \in \mathscr{P}(\mathsf{P})$.

Considering X with respect to a particular measure $\widetilde{\mathsf{P}} \in \mathscr{P}(\mathsf{P})$ we see that, by the Doob decomposition (Chapter II, § 1b),

$$X_n = X_0 + M_n^{(\widetilde{\mathsf{P}})} - C_n^{(\widetilde{\mathsf{P}})}, \tag{1}$$

where $M^{(\widetilde{\mathsf{P}})} = (M_n^{(\widetilde{\mathsf{P}})}, \mathscr{F}_n, \widetilde{\mathsf{P}})$ is a martingale, $C^{(\widetilde{\mathsf{P}})} = (C_n^{(\widetilde{\mathsf{P}})}, \mathscr{F}_{n-1}, \widetilde{\mathsf{P}})$ is a non-decreasing predictable process, $M_0^{(\widetilde{\mathsf{P}})} = 0$, and $C_0^{(\widetilde{\mathsf{P}})} = 0$. The components $M^{(\widetilde{\mathsf{P}})}$ and $C^{(\widetilde{\mathsf{P}})}$ in (1) depend on our choice of the measure $\widetilde{\mathsf{P}}$, which we emphasize by our notation.

In the theorem following next we describe another decomposition of X, the so-called *optional* decomposition. It is remarkable for its *universal* nature: its components (see (2)) are the same for all $\widetilde{\mathsf{P}} \in \mathscr{P}(\mathsf{P})$.

THEOREM. *If a process X is a supermartingale with respect to each martingale measure $\widetilde{\mathsf{P}} \in \mathscr{P}(\mathsf{P})$, then it has an (optional) decomposition*

$$X_n = X_0 + \sum_{k=1}^{n} (\gamma_k, \Delta S_k) - C_n, \qquad n \leqslant N, \tag{2}$$

with predictable $\mathbb{R}^d$-valued process $\gamma = (\gamma_k)_{k\leqslant N}$ and non-decreasing process $C = (C_n)_{n\leqslant N}$ of $\mathscr{F}_n$-measurable variables C_n.

Before turning to the proof we note that there exists an essential difference between (1) and (2): the variables $C_n^{(\widetilde{\mathsf{P}})}$ in (1) are $\mathscr{F}_{n-1}$-measurable, whereas C_n in (2) is $\mathscr{F}_n$-measurable. It is just for this reason that (2) is called an *optional* decomposition.

Remark. In the present, discrete-time, case one says that a process $C = (C_n)_{n\leqslant N}$ is optional with respect to $(\mathscr{F}_n)_{n\leqslant N}$ when it is merely *consistent* with (or *adapted* to) the filtration $(\mathscr{F}_n)_{n\leqslant N}$, i.e., when C_n is $\mathscr{F}_n$-measurable, $n \geqslant 0$. See Chapter III, § 5a and [250; Chapter I, § 1c] for the concept of optional process.

First versions of the above theorem were established in [136] and [281] for the continuous-time case (as already mentioned in § 1c).

Before long, several other papers devoted to both discrete and continuous time were published ([99], [163]–[165]), in which, in particular, the assumptions of [136] and [281] were weakened and various versions of proofs were suggested.

The proof below follows the scheme of H. Föllmer and Yu. M. Kabanov [163], [164], which is based on the idea of treating the γ_k in (2) as *Lagrange multipliers* in certain *problems of optimization with constraints.* (We shall also use several results of Chapter V, § 2e in the proof).

2. In accordance with [163] and [164], we shall have established the decomposition (2) once we have shown that for each $n = 1, \dots, N$ the variables $\Delta X_n \equiv X_n - X_{n-1}$ can be represented as the differences

$$\Delta X_n = (\gamma_n, \Delta S_n) - c_n \tag{3}$$

where γ_n is a $\mathscr{F}_{n-1}$-measurable $\mathbb{R}^d$-valued variable and c_n is a non-negative $\mathscr{F}_n$-measurable variable.

To obtain such a representation of ΔX_n it is in fact sufficient to show that (under the above assumptions about X and S) there exists a $\mathscr{F}_{n-1}$-measurable $\mathbb{R}^d$-valued variable γ_n such that

$$\Delta X_n - (\gamma_n, \Delta S_n) \leqslant 0. \tag{4}$$

In this case, we can take $(\gamma_n, \Delta S_n) - \Delta X_n$ as the required variable c_n.

We note also that if $\widetilde{\mathsf{P}} \in \mathscr{P}(\mathsf{P})$, then

$$\mathsf{E}_{\widetilde{\mathsf{P}}}|\Delta S_n| < \infty, \qquad \mathsf{E}_{\widetilde{\mathsf{P}}}(\Delta S_n \mid \mathscr{F}_{n-1}) = 0 \tag{5}$$

and

$$\mathsf{E}_{\widetilde{\mathsf{P}}}|\Delta X_n| < \infty, \qquad \mathsf{E}_{\widetilde{\mathsf{P}}}(\Delta X_n \mid \mathscr{F}_{n-1}) \leqslant 0. \tag{6}$$

If P_n and $\widetilde{\mathsf{P}}_n$ are the restrictions of P and $\widetilde{\mathsf{P}}$ to $\mathscr{F}$ and $Z_n = \dfrac{d\widetilde{\mathsf{P}}_n}{d\mathsf{P}_n}$, then by Bayes's formula (Chapter V, § 3a),

$$\mathsf{E}_{\widetilde{\mathsf{P}}}(\Delta S_n \mid \mathscr{F}_{n-1}) = \mathsf{E}_{\widetilde{\mathsf{P}}}(z_n \Delta S_n \mid \mathscr{F}_{n-1}), \tag{7}$$

$$\mathsf{E}_{\widetilde{\mathsf{P}}}(\Delta X_n \mid \mathscr{F}_{n-1}) = \mathsf{E}_{\widetilde{\mathsf{P}}}(z_n \Delta X_n \mid \mathscr{F}_{n-1}), \tag{8}$$

where $z_n = \dfrac{Z_n}{Z_{n-1}}$.

Hence it is easy to see that (4) is a consequence of the following general result (where we set $\xi = \Delta X_n$ and $\eta = \Delta S_n$), which is also of independent, 'purely probabilistic', interest.

LEMMA. *Let* $(\Omega, \mathscr{F}, \mathsf{P})$ *be a probability space and let* ξ *and* $\eta = (\eta^1, \dots, \eta^d)$ *be a real-valued variable and an* $\mathbb{R}^d$*-valued variable in this space.*

Let $\mathscr{G}$ *be a* σ*-subalgebra of* $\mathscr{F}$ *and let* Z *be the set of all random variables* $z > 0$ *such that,* P-a.s.,

$$\mathsf{E}(z \mid \mathscr{G}) = 1, \qquad \mathsf{E}(|\xi| z \mid \mathscr{G}) < \infty, \qquad \mathsf{E}(|\eta| z \mid \mathscr{G}) < \infty \tag{9}$$

and

$$\mathsf{E}(z\eta \mid \mathscr{G}) = 0. \tag{10}$$

If $Z \neq \varnothing$ *and*

$$\mathsf{E}(z\xi \mid \mathscr{G}) \leqslant 0 \tag{11}$$

for all $z \in Z$, *then there exists a* $\mathscr{G}$*-measurable* $\mathbb{R}^d$*-valued vector* λ^* *such that*

$$\xi + (\lambda^*, \eta) \leqslant 0 \quad (\mathsf{P}\text{-a.s.}). \tag{12}$$

Proof. a) The idea of the proof is already transparent when $\mathscr{G}$ is a trivial σ-subalgebra, i.e., $\mathscr{G} = \{\varnothing, \Omega\}$. In this case the required vector λ^* is *nonrandom* and can be treated (as shown below) as a Lagrange multiplier in a certain optimization problem.

If $\mathscr{G}$ is a nontrivial σ-subalgebra, then we can carry out the same arguments for *each* ω and, again, obtain a vector λ^* (not uniquely defined, in general), which depends on ω. After that, the entire problem is reduced to the proof that λ^* can be chosen $\mathscr{G}$*-measurable.*

We recall that we encountered the same measurability problem in Chapter V, § 2e, in our proof of the *extended* version of the *First fundamental theorem* (in the discussion of the implications a$'$) $\Longrightarrow$ e) and e) $\Longrightarrow$ b)). We referred there to certain general results on the existence of a measurable selection (Lemma 2 in Chapter V, § 2e).

The same techniques are applicable in our context for the proof of the existence of a $\mathscr{G}$-measurable selection λ^*. Referring to [163] and [164] for detail we note that there is no such problem of measurability in the case of a discrete space Ω.

b) Thus, we shall assume that $\mathscr{G} = \{\varnothing, \Omega\}$.

Let $\mathsf{Q} = \mathsf{Q}(dx, dy)$ be the measure in $\mathbb{R} \times \mathbb{R}^d$ generated by the variables ξ and $\eta = (\eta^1, \dots, \eta^d)$, i.e.,

$$\mathsf{Q}(dx, dx) = \mathsf{P}\big(\xi \in dx, \eta \in dy\big).$$

Without loss of generality we can assume that the family of random variables $\eta^1, \dots, \eta^d$ is (P-a.s.) a *linearly independent system*, i.e., if $a^1\eta^1 + \cdots + a^d\eta^d = 0$ (P-a.s.) for some coefficients $a^1, \dots, a^d$, then $a^1 = \cdots = a^d = 0$. Indeed, $\eta^1, \dots, \eta^d$ enter (12) in a linear manner. If they were linearly dependent, then the problem could be reduced to another, with vector η of lower dimension.

As in Chapter V, § 2e, let $L^\circ(\mathsf{Q})$ be the relative interior of the closed convex hull $L(\mathsf{Q})$ of the topological support $K(\mathsf{Q})$ of the measure Q.

Let $x' = (x, y)$, $x \in \mathbb{R}^d$, $y \in \mathbb{R}^{d+1}$, and let $Z(\mathsf{Q})$ be the family of Borel functions $z = z(x') > 0$ such that $\mathsf{E}_{\mathsf{Q}} z = 1$ and $\mathsf{E}_{\mathsf{Q}}|x'|z < \infty$.

Let

$$\Phi(\mathsf{Q}) = \{\varphi(z) \colon \varphi(z) = \mathsf{E}_{\mathsf{Q}} x'z,\ z \in Z(\mathsf{Q})\}$$

be the family of barycenters (of the measures $d\mathsf{Q}' = z\,d\mathsf{Q}$).

By Lemma 1 in Chapter V, § 2e we obtain that $L^\circ(\mathsf{Q}) \subseteq \Phi(\mathsf{Q})$ (in fact, $L^\circ(\mathsf{Q}) = \Phi(\mathsf{Q})$) and if $0 \notin L^\circ(\mathsf{Q})$, then there exists $\gamma' \in \mathbb{R}^{d+1}$ such that

$$\mathsf{Q}\{x' \colon (\gamma', x') \geqslant 0\} = 1 \quad \text{and} \quad \mathsf{Q}\{x' \colon (\gamma', x') > 0\} > 0. \tag{13}$$

To prove the existence of a vector λ^* with property (12) we consider separately the following two cases:

(i) $0 \notin L^\circ(\mathsf{Q})$,

(ii) $0 \in L^\circ(\mathsf{Q})$.

Case (i). By (13), there exist numbers γ and $\gamma^1, \dots, \gamma^d$ such that (P-a.s.)

$$\gamma\xi + (\gamma^1\eta^1 + \dots + \gamma^d\eta^d) \geqslant 0 \tag{14}$$

and, with positive P-probability,

$$\gamma\xi + (\gamma^1\eta^1 + \dots + \gamma^d\eta^d) > 0. \tag{15}$$

We claim that $\gamma \neq 0$. For if $\gamma = 0$, then $\gamma^1\eta^1 + \dots + \gamma^d\eta^d \geqslant 0$ (P-a.s.). By hypothesis, there exists a martingale measure $\widetilde{\mathsf{P}} \sim \mathsf{P}$ such that

$$\mathsf{E}_{\widetilde{\mathsf{P}}}(\gamma^1\eta^1 + \dots + \gamma^d\eta^d) = 0,$$

therefore $\gamma^1\eta^1 + \dots + \gamma^d\eta^d = 0$ (both $\widetilde{\mathsf{P}}$- and P-a.s.).

Since $\eta^1, \dots, \eta^d$ are assumed to be linearly independent, this means that $\gamma^1 = \dots = \gamma^d = 0$, which, however, contradicts (15).

Hence $\gamma \neq 0$ and we see from relation (14) and our assumptions that $\mathsf{E}_{\widetilde{\mathsf{P}}}\xi \leqslant 0$, $\mathsf{E}_{\widetilde{\mathsf{P}}}\eta^i = 0$ that $\gamma < 0$.

Setting $\lambda^i = \dfrac{\gamma_i}{\gamma}$, $i = 1, \dots, d$, we obtain by (14) the following inequality:

$$\xi + (\lambda^1\eta^1 + \dots + \lambda^d\eta^d) \leqslant 0,$$

which proves (12) with $\lambda^* = (\lambda^1, \dots, \lambda^d)$ in case (i).

Case (ii) is slightly more complicated. It is in this case that we shall use the idea of [163] and [164] about Lagrange multipliers.

Let

$$\varphi_\xi(z) = \mathsf{E}_{\mathsf{Q}} xz, \qquad \varphi_\eta(z) = \mathsf{E}_{\mathsf{Q}} yz$$

be the components of the barycenter $\varphi(z)$.

We set

$$Z_0(\mathsf{Q}) = \{z \in Z(\mathsf{Q}) \colon \varphi_\eta(z) = 0\}$$

and

$$\Phi_0(\mathsf{Q}) = \{\varphi(z) = (\varphi_\xi(z), \varphi_\eta(z)) \colon z \in Z_0(\mathsf{Q})\}.$$

By hypothesis, $Z_0(\mathsf{Q}) \neq \varnothing$ and

$$z \in Z_0(\mathsf{Q}) \implies \varphi_\xi(z) \leqslant 0. \tag{16}$$

If $0 \in L^\circ(\mathsf{Q})$, then, as already observed, $L^\circ(\mathsf{Q}) \subseteq \Phi(\mathsf{Q})$, and by Chapter V, § 2e we obtain that there exists $z_0 \in Z_0(\mathsf{Q})$ such that $\varphi_\xi(z_0) = 0$. Hence, in case (ii),

$$\sup_{z \in Z_0(\mathsf{Q})} \varphi_\xi(z) = 0, \tag{17}$$

which we can interpret as follows: in this case the value of f^* in the *optimization problem*

$$\text{“}\textit{find } f^* \equiv \sup_{z \in Z_0(\mathsf{Q})} \varphi_\xi(z)\text{”} \tag{18}$$

is equal to zero.

Following [163] and [164] we now reformulate (18) as an *optimization problem with constraints*:

$$\text{“}\textit{find } f^* \equiv \sup_{z \in Z(\mathsf{Q})} \varphi_\xi(z) \textit{ under the additional constraint } \varphi_\eta(z) = 0\text{”}. \tag{19}$$

According to the principles of variational calculus, for some *non-zero* vector λ^* (the Lagrange multiplier) the problem (19) is equivalent to the following *optimization problem*:

$$\text{“}\textit{find } f^* \equiv \sup_{z \in Z(\mathsf{Q})} \bigl(\varphi_\xi(z) + \lambda^* \varphi_\eta(z)\bigr)\text{”}. \tag{20}$$

(For simplicity, we denote here and below the scalar product (a, b) of vectors a and b by ab.)

We shall now prove that the problems (19) and (20) are equivalent (this is interesting on its own, although it would suffice for our aims to show that, under the assumption (ii), we have

$$\sup_{z \in Z(\mathsf{Q})} \bigl(\varphi_\xi(z) + \lambda^* \varphi_\eta(z)\bigr) \leqslant 0; \tag{21}$$

at any rate, for some non-zero vector λ^*).

Let

$$A = \{(x, y) \in \mathbb{R} \times \mathbb{R}^d : x < \varphi_\xi(z) \text{ and } y = \varphi_\eta(z) \text{ for some } z \in Z(\mathsf{Q})\}.$$

This set is nonempty and convex. By assumption (ii), $(0,0) \notin A$, therefore we can *separate* the origin and A by a hyperplane, i.e., *there exists a nonzero vector* $\lambda = (\lambda_1, \lambda_2) \in \mathbb{R} \times \mathbb{R}^d$ *such that*

$$\lambda_1 x + \lambda_2 y \leqslant 0 \tag{22}$$

for all (x, y) *in the closure of* A. (See, e.g., [241; § 0.1]; note that we have in fact employed the idea of 'separation' also in case (i)).

Note that A contains all points $(x, 0)$ with negative x. Hence $\lambda_1 \geqslant 0$ in (22).

Further, if we assume that $\lambda_1 = 0$, then we see from (21) that

$$\mathsf{E}_{\mathsf{Q}}(\lambda_2 y)z = \lambda_2 \mathsf{E}_{\mathsf{Q}} yz = \lambda_2 \varphi_\eta(z) \leqslant 0 \tag{23}$$

for all $z \in Z(\mathsf{Q})$. Further, $z \in Z(\mathsf{Q})$, so that $\lambda_2 y \leqslant 0$ (Q-a.s.), i.e., $\lambda_2 \eta \leqslant 0$ (P-a.s.).

Since $\mathscr{P}(\mathsf{P}) \neq \varnothing$, it follows that $\mathsf{E}_{\widetilde{\mathsf{P}}} \lambda_2 \eta = 0$ for some measure $\widetilde{\mathsf{P}}$ in $\mathscr{P}(\mathsf{P})$. Combined with the property $\lambda_2 \eta \leqslant 0$ ($\widetilde{\mathsf{P}}$-a.s.) this shows that we have linear dependence ($\lambda_2 \eta = 0$), which is ruled out by the above assumption. Hence $\lambda_2 = 0$.

Consequently, if $\lambda_1 = 0$, then also $\lambda_2 = 0$, which contradicts the assumption that the vector (λ_1, λ_2) in (21) is not zero.

Thus, $\lambda_1 > 0$.

We now set $\lambda^* = \dfrac{\lambda_2}{\lambda_1}$. Then we see from (21) and the definition of A that for all $z \in Z(\mathsf{Q})$ and $\varepsilon > 0$ we have

$$\bigl(\varphi_\xi(z) - \varepsilon\bigr) + \lambda^* \varphi_\eta(z) \leqslant 0.$$

Passing to the limit as $\varepsilon \to 0$ we obtain

$$\varphi_\xi(z) + \lambda^* \varphi_\eta(z) \leqslant 0, \qquad z \in Z(\mathsf{Q}),$$

which is equivalent to the inequality

$$\sup_{z \in Z(\mathsf{Q})} \bigl(\varphi_\xi(z) + \lambda^* \varphi_\eta(z)\bigr) \leqslant 0. \tag{24}$$

We now observe that, obviously,

$$\sup_{z \in Z(\mathsf{Q})} \bigl(\varphi_\xi(z) + \lambda \varphi_\eta(z)\bigr) \geqslant \sup_{z \in Z_0(\mathsf{Q})} \bigl(\varphi_\xi(z) + \lambda \varphi_\eta(z)\bigr) = \sup_{z \in Z_0(\mathsf{Q})} \varphi_\xi(z) \tag{25}$$

for each $\lambda \in \mathbb{R}^d$.

If (ii) holds, then the right-hand side of (25) is equal to zero, while the left-hand side vanishes for $\lambda = \lambda^*$. Hence, under assumption (ii) we obtain

$$\sup_{z \in Z_0(\mathsf{Q})} \varphi_\xi(z) = 0 \iff \sup_{z \in Z(\mathsf{Q})} \bigl(\varphi_\xi(z) + \lambda^* \varphi_\eta(z)\bigr) = 0,$$

which establishes the equivalence of (19) and (20) and can be interpreted as follows: the Lagrange multiplier 'lifts' the constraint $\varphi_\eta(z) = 0$ in the optimization problem (19).

We now turn back to (24) or, equivalently, to the inequality

$$\varphi_\xi(z) + \lambda^* \varphi_\eta(z) \leqslant 0, \qquad z \in Z(\mathsf{Q}), \tag{26}$$

which can be rewritten as follows:

$$\mathsf{E}_{\mathsf{Q}} z(x + \lambda^* y) \leqslant 0, \qquad z \in Z(\mathsf{Q}).$$

The space $Z(\mathsf{Q})$ is 'sufficiently rich', therefore $x + \lambda^* y \leqslant 0$ (Q-a.s.), which proves required property (12) in the case of $\mathscr{G} = \{\varnothing, \Omega\}$.

In the general case, as already mentioned, one must consider in place of $\mathsf{Q}(dx, dy)$ *regular* conditional distributions

$$\mathsf{Q}(\omega; dx, dy) = \mathsf{P}(\xi \in dx, \eta \in dy \,|\, \mathscr{G})(\omega)$$

and prove the existence of $\lambda^* = \lambda^*(\omega)$ for each ω. At the final stage one must show that we can choose a $\mathscr{G}$-measurable version of the function $\lambda^* = \lambda^*(\omega)$, $\omega \in \Omega$, by using general results on measurable selections, such as Lemma 2 in Chapter V, § 2e. See the details in [163] and [164].

3. Scheme of Series of 'Large' Arbitrage-Free Markets and Asymptotic Arbitrage

§ 3a. One Model of 'Large' Financial Markets

1. We encountered already 'large' markets and the concept of asymptotic arbitrage in Chapter I, § 2d, in the discussion of the basic principles of *Arbitrage pricing theory* (APT) pioneered by S. Ross [412].

As in H. Markowitz's theory ([331]–[333], see also Chapter I, § 2b), based on the analysis of the *mean value* and *variance* of the capital corresponding to various investment portfolios, *asymptotic* arbitrage in Ross's theory is also defined in terms of these parameters.

In the present section we define asymptotic arbitrage in a somewhat different manner, which is more consistent with the concept of arbitrage considered in Chapter V, § 2a and more adequate to the martingale approach permeating our entire presentation.

2. Developing further our initial model of a (B, S)-market formed by a bank account $B = (B_k)_{k\leqslant n}$ and a d-dimensional stock $S = (S_k)_{k\leqslant n}$ (where $S_k = (S_k^1, \dots, S_k^d)$) both defined on some *filtered* probability space

$$(\Omega, \mathscr{F}, (\mathscr{F}_k)_{k\leqslant n}, \mathsf{P}),$$

we assume now that we have a *scheme of series* of n-markets

$$(B^n, S^n) = (B_k^n, S_k^n)_{k\leqslant k(n)}$$

with $S_k^n = (S_k^{n,1}, \dots, S_k^{n,d(n)})$, and each market is defined on a probability space

$$(\Omega^n, \mathscr{F}^n, (\mathscr{F}_k^n)_{k\leqslant k(n)}, \mathsf{P}^n) \tag{1}$$

'of its own'.

Here we assume that $n \geqslant 1$, $\mathscr{F}_0^n = \{\varnothing, \Omega^n\}$, $\mathscr{F}^n = \mathscr{F}_{k(n)}^n$ with $k(n) < \infty$ and $d(n) < \infty$.

We are mainly interested in the cases when $k(n) \to \infty$ and (or) $d(n) \to \infty$ as $n \to \infty$. It is in this sense that we shall speak about 'large' markets.

Remark. In our considerations of the scheme of series of probability spaces the index n is the *number of a series*, while k plays the role of a *time parameter.*

3. Let $X^{\pi(n)} = (X_k^{\pi(n)})_{k \leqslant k(n)}$ be a capital corresponding to some self-financing portfolio $\pi(n)$ in a (B^n, S^n)-market.

Recall that we agree to assume that the variables B_k^n are positive and $\mathscr{F}_{k-1}^n$-measurable. We explained at the end of § 2b in Chapter V that we can assume without loss of generality that $B_k^n \equiv 1$, which corresponds to a transition to discounted prices. In this case,

$$X_k^{\pi(n)} = X_0^{\pi(n)} + \sum_{l=1}^{k} (\gamma_l^n, \Delta S_l^n),$$

where $(\gamma_l^n, \Delta S_l^n) = \sum\limits_{i=1}^{d(n)} \gamma_l^{n,i} \Delta S_l^{n,i}$.

DEFINITION 1. We shall say that a sequence of strategies $\pi = (\pi(n))_{n \geqslant 1}$ realizes an *asymptotic arbitrage* in a *scheme of series* $(\mathbb{B}, \mathbb{S}) = \{(B^n, S^n),\ n \geqslant 1\}$ of n-markets (B^n, S^n) if

$$\lim_n X_0^{\pi(n)} = 0, \tag{2}$$

$$X_{k(n)}^{\pi(n)} \geqslant -c(n) \quad (\mathsf{P}^n\text{-a.s.}), \quad n \geqslant 1, \tag{3}$$

where $0 \leqslant c(n) \downarrow 0$ as $n \to \infty$ and

$$\lim_{\varepsilon \downarrow 0} \varlimsup_n \mathsf{P}^n \big(X_{k(n)}^{\pi(n)} \geqslant \varepsilon \big) > 0. \tag{4}$$

Using the above-introduced concepts and notation we can say (slightly broadening the arguments of Chapter V, § 2a) that the *asymptotic arbitrage in APT* considered in Chapter I, § 2d occurs if there exists a *subsequence* $(n') \subseteq (n)$ and a *sequence* of strategies $(\pi(n'))$ such that $X_0^{\pi(n')} = 0$ and $\mathsf{E} X_{k(n')}^{\pi(n')} \to \infty$, $\mathsf{D} X_{k(n')}^{\pi(n')} \to 0$ as $n' \to \infty$.

Definition (4) of asymptotic arbitrage is more convenient and preferable from the 'martingale' standpoint to that of *APT*, which can be explained as follows.

First, we can consider (4) as a natural generalization of our earlier definition of opportunities for arbitrage (Chapter V, § 2a), and the latter, as we know from the First fundamental theorem, relates arbitrage theory and the theory of martingales and stochastic calculus in a straightforward way.

Second, taking the definition (4) or a similar one (see [260], [261], [273]) we can find effective criteria of the absence of asymptotic arbitrage. They include criteria in terms of such well-known concepts of the theory of stochastic processes as *Hellinger integrals* and *Hellinger processes* (see [250; Chapter V]).

4. DEFINITION 2. We say that a $(\mathbb{B}, \mathbb{S})$-market that is a collection $\{(B^n, S^n),\ n \geqslant 1\}$ of n-markets is *locally arbitrage-free* if the market (B^n, S^n) is arbitrage-free (Chapter V, § 2a) for each $n \geqslant 1$.

The central question in what follows is to find conditions ensuring that there is *no* asymptotic arbitrage (in the sense of Definition 1 above) on a locally arbitrage-free $(\mathbb{B}, \mathbb{S})$-market.

The existence of asymptotic arbitrage as $n \to \infty$ can have various reasons: the growth in the number of shares $(d(n) \to \infty)$, the increase of the time interval $(k(n) \to \infty$; see Example 2 in Chapter V, § 2b), breaking down of the asymptotic equivalence of measures. It can also be brought about by a combination of these factors.

In connection with the above-mentioned asymptotic equivalence of measures and the related asymptotic analogues of *absolute continuity* and *singularity* of sequences of probability measures it should be noted that their precise definitions involve the concepts of *contiguity* and *complete asymptotic separability* (see [250; Chapter V], where one can find also criteria of these properties formulated in terms of Hellinger integrals and processes).

The importance of these concepts in the problem of asymptotic arbitrage on large markets has been pointed out for the first time by Yu. M. Kabanov and D. O. Kramkov [261] who have introduced the concepts of asymptotic arbitrage of the first and the second kinds. (In accordance with the nomenclature in [261] the asymptotic arbitrage of Definition 1 is of the *first kind*.) Later on, the theory of asymptotic arbitrage has been significantly developed by I. Klein and W. Schachermayer [273] and by Yu. M. Kabanov and D. O. Kramkov in [260].

§ 3b. Criteria of the Absence of Asymptotic Arbitrage

1. Let $X^{\pi(n)} = (X_k^{\pi(n)})_{k \leqslant k(n)}$ be the capital corresponding to some self-financing portfolio $\pi(n)$ in a (B^n, S^n)-market with $B_k^n \equiv 1$:

$$X_k^{\pi(n)} = X_0^{\pi(n)} + \sum_{l=1}^{k} (\gamma_l^n, \Delta S_l^n). \tag{1}$$

If Q^n is a measure on $(\Omega^n, \mathscr{F}^n)$ such that $\mathsf{Q}^n \ll \mathsf{P}^n$, then we obtain by Bayes's formula ((4) in Chapter V, § 3a) that (Q^n-a.s.)

$$\mathsf{E}_{\mathsf{Q}^n}\big(X_{k(n)}^{\pi(n)} \,|\, \mathscr{F}_k^n\big) = \frac{1}{Z_k^n} \mathsf{E}_{\mathsf{P}^n}\big(X_{k(n)}^{\pi(n)} Z_{k(n)}^n \,|\, \mathscr{F}_k^n\big), \tag{2}$$

where $Z_k^n = \dfrac{d\mathsf{Q}_k^n}{d\mathsf{P}_k^n}$, $\mathsf{Q}_k^n = \mathsf{Q}^n \,|\, \mathscr{F}_k^n$, and $\mathsf{P}_k^n = \mathsf{P}^n \,|\, \mathscr{F}_k^n$. (Here we assume that the expectation on the left-hand side of (2) is well defined.)

We assume that each of the n-markets (B^n, S^n), $n \geqslant 1$, is *arbitrage-free*, and therefore, in accordance with the *First fundamental theorem* (Chapter V, § § 2b, c), the family of martingale measures $\mathscr{P}(\mathsf{P}^n)$ is nonempty.

Let $\widetilde{\mathsf{P}}^n \in \mathscr{P}(\mathsf{P}^n)$ and let $\pi(n)$ be a strategy such that

$$\mathsf{E}_{\widetilde{\mathsf{P}}^n} \bigl(X_{k(n)}^{\pi(n)}\bigr)^- < \infty. \tag{3}$$

Then we obtain by the lemma in Chapter II, § 1c, that the sequence $X^{\pi(n)} = \bigl(X_k^{\pi(n)}\bigr)_{k \leqslant k(n)}$ is a $\widetilde{\mathsf{P}}^n$-martingale. Consequently, $\mathsf{E}_{\widetilde{\mathsf{P}}^n} \bigl|X_{k(n)}^{\pi(n)}\bigr| < \infty$ and

$$X_k^{\pi(n)} = \mathsf{E}_{\widetilde{\mathsf{P}}^n} \bigl(X_{k(n)}^{\pi(n)} \mid \mathscr{F}_k^n\bigr) \tag{4}$$

for $k \leqslant k(n)$. Clearly, $\mathsf{E}_{\mathsf{P}^n} \bigl|X_{k(n)}^{\pi(n)} Z_{k(n)}^n\bigr| = \mathsf{E}_{\widetilde{\mathsf{P}}^n} \bigl|X_{k(n)}^{\pi(n)}\bigr| < \infty$, and, by (2) and (4),

$$X_k^{\pi(n)} Z_k^n = \mathsf{E}_{\mathsf{P}^n} \bigl(X_{k(n)}^{\pi(n)} Z_{k(n)}^n \mid \mathscr{F}_k^n\bigr) \quad (\widetilde{\mathsf{P}}^n\text{- and } \mathsf{P}^n\text{-a.s.}). \tag{5}$$

Hence assumption (3) ensures in our case of discrete time $k \leqslant k(n) < \infty$ that

$$\bigl(X_n^{\pi(n)}, \mathscr{F}_k^n, \widetilde{\mathsf{P}}^n\bigr)_{k \leqslant k(n)} \quad \text{and} \quad \bigl(X_n^{\pi(n)} Z_k^n, \mathscr{F}_k^n, \mathsf{P}^n\bigr)_{k \leqslant k(n)}$$

are martingales. (Cf. the lemma in Chapter V, § 3d.)

In particular,

$$X_0^{\pi(n)} = \mathsf{E}_{\widetilde{\mathsf{P}}^n} X_{k(n)}^{\pi(n)}, \tag{6}$$

$$X_0^{\pi(n)} = \mathsf{E}_{\mathsf{P}^n} X_{k(n)}^{\pi(n)} Z_{k(n)}^n. \tag{7}$$

Each of these equivalent relations can be used in the search of conditions ensuring that the strategy $\pi(n)$ is *arbitrage-free* and the sequence of strategies $\pi = (\pi(n))_{n \geqslant 1}$ is *asymptotically arbitrage-free*. (Note that, in fact, we used just these relations in our proof of the sufficiency in the *First fundamental theorem*, in Chapter V, § 2c.)

Remark. It is not necessary for formulas (4)–(7) to hold that $\widetilde{\mathsf{P}}^n \sim \mathsf{P}^n$. It suffices that $\widetilde{\mathsf{P}}^n \ll \mathsf{P}^n$. We must nevertheless assume that $\mathsf{P}^n \ll \widetilde{\mathsf{P}}^n$ if we want to deduce from (7), say, the absence of arbitrage. For an explanation assume that $X_0^{\pi(n)} = 0$, $X_{k(n)}^{\pi(n)} \geqslant 0$ (P^n-a.s.), and $A = \bigl\{X_{k(n)}^{\pi(n)} > 0\bigr\}$. Then it is clear from (7) that we can say nothing on the probability $\mathsf{P}^n(A)$ of a set A if $Z_{k(n)}^n = 0$ on this set. This explains why we can deduce the equality $\mathsf{P}^n\bigl(X_{k(n)}^{\pi(n)} = 0\bigr) = 1$ from the assumption $\mathsf{P}^n\bigl(X_{k(n)}^{\pi(n)} \geqslant 0\bigr) = 1$ and the relation $0 = \mathsf{E}_{\mathsf{P}^n} X_{k(n)}^{\pi(n)} Z_{k(n)}^n$ only if $\mathsf{P}^n(Z_{k(n)}^n > 0) = 1$. (By assertion f) of the theorem in Chapter V, § 3a, this means that $\mathsf{P}^n \ll \widetilde{\mathsf{P}}^n$.)

Thus, the condition (in the definition) that the martingale measure $\widetilde{\mathsf{P}}^n$ be *equivalent* to the measure P^n ensures, in particular, that

$$0 < Z_{k(n)}^n < \infty \quad (\mathsf{P}^n\text{-a.s.}). \tag{8}$$

2. For simplicity, we start our search of criteria of the absence of asymptotic arbitrage from the *stationary case*, when we have a $(\mathbb{B},\mathbb{S})$-market, $(\mathbb{B},\mathbb{S}) = \{(B^n, S^n),\ n \geqslant 1\}$, of the following structure.

There exist a probability space $(\Omega, \mathscr{F}, (\mathscr{F}_k)_{k\geqslant 0}, \mathsf{P})$, $\mathscr{F} = \bigvee \mathscr{F}_k$ and a $(d+1)$-dimensional process $(B,S) = (B_k, S_k)_{k\geqslant 0}$ of $\mathscr{F}_{k-1}$-measurable variables B_k and $\mathscr{F}_k$-measurable variables $S_k = (S_k^1, \dots, S_k^d)$ such that each n-market has the structure $(B^n, S^n) = (B_k, S_k)_{k\leqslant k(n)}$ with $k(n) = n$.

Clearly (using the language of the scheme of series), we can assume that the market (B^n, S^n) is defined on its own probability space $(\Omega, \mathscr{F}^n, (\mathscr{F}_k^n)_{k\leqslant n}, \mathsf{P}^n)$, which has the properties $\mathscr{F}^n = \mathscr{F}_n$, $\mathscr{F}_k^n = \mathscr{F}_k$, $k \leqslant n$, and $\mathsf{P}^n = \mathsf{P} \,|\, \mathscr{F}^n$.

We can express this otherwise: in the 'stationary case' the number $d(n)$ of shares is independent of n $(d(n) \equiv d)$ and the market (B^{n+1}, S^{n+1}) is an 'extension' of the market (B^n, S^n) for each n.

Let $\widetilde{\mathbb{P}} = \{(\widetilde{\mathsf{P}}_k)_{k\geqslant 1}\}$ be the family of sequences $(\widetilde{\mathsf{P}}_k)_{k\geqslant 1}$ of martingale measures $\widetilde{\mathsf{P}}_k$ that have the property of compatibility: $\widetilde{\mathsf{P}}_{k+1} \,|\, \mathscr{F}_k = \widetilde{\mathsf{P}}_k$, $k \geqslant 1$.

Given such sequence of measures $(\widetilde{\mathsf{P}}_k)_{k\geqslant 1}$ we can consider the associated sequence $Z = (Z_k)_{k\geqslant 0}$ of Radon Nikodym derivatives $Z_k = \dfrac{d\widetilde{\mathsf{P}}_k}{d\mathsf{P}_k}$, $k \geqslant 1$, $Z_0 = 1$.

Let

$$\mathbb{Z}_k = \left\{ Z_k \colon Z_k = \frac{d\widetilde{\mathsf{P}}_k}{d\mathsf{P}_k},\ \widetilde{\mathsf{P}}_k \in \mathscr{P}(\mathsf{P}^k) \right\}$$

and let

$$\mathbb{Z}_\infty = \left\{ Z_\infty \colon Z_\infty = \varlimsup_k \frac{d\widetilde{\mathsf{P}}_k}{d\mathsf{P}_k},\ (\widetilde{\mathsf{P}}_k)_{k\geqslant 1} \in \widetilde{\mathbb{P}} \right\}.$$

Although we do not assume the existence of a measure $\widetilde{\mathsf{P}}$ on $(\Omega, \mathscr{F})$ such that $\widetilde{\mathsf{P}}_k = \widetilde{\mathsf{P}} \,|\, \mathscr{F}_k$, note that the sequence $(Z_k, \mathscr{F}_k)_{k\geqslant 0}$ is nevertheless a (positive) P-martingale thanks to the property $\widetilde{\mathsf{P}}_{k+1} \,|\, \mathscr{F}_k = \widetilde{\mathsf{P}}_k$. Hence, by Doob's convergence theorem (Chapter V, § 3a) there exists (P-a.s.) $\lim Z_k$ $(= Z_\infty)$, and moreover, $0 \leqslant \mathsf{E} Z_\infty \leqslant 1$.

THEOREM 1 (the *stationary case*). *If $(\mathbb{B},\mathbb{S}) = \{(B^n, S^n),\ n \geqslant 1\}$ is a locally arbitrage-free market, then the condition*

$$\lim_{\varepsilon\downarrow 0} \varlimsup_k \inf_{Z_k\in\mathbb{Z}_k} \mathsf{P}(Z_k < \varepsilon) = 0 \tag{9}$$

is necessary and sufficient, while the condition

$$\lim_{\varepsilon\downarrow 0} \inf_{Z_\infty\in\mathbb{Z}_\infty} \mathsf{P}(Z_\infty < \varepsilon) = 0 \tag{10}$$

is sufficient for the absence of asymptotic arbitrage.

Proof. The *sufficiency* of condition (9) is relatively easy to establish; we present its proof below. (The sufficiency of (10) is a consequence of the implication (10) $\Rightarrow$ (9).) The proof of the *necessity* is more complicated. It is based on several results on *contiguity* of probability measures and is presented in the next section, § 3c (subsection 9).

Let $(\widetilde{\mathsf{P}}_n)_{n\geqslant 1} \in \widetilde{\mathbb{P}}$, and let $\pi = (\pi(n))_{n\geqslant 1}$ be a sequence of strategies on a $(\mathbb{B}, \mathbb{S})$-market, $(\mathbb{B}, \mathbb{S}) = \{(B^n, S^n),\ n \geqslant 1\}$, satisfying condition (3) in § 3a. Then we have (3) and therefore also (6), which takes the following form in our 'stationary case':

$$X_0^{\pi(n)} = \mathsf{E} Z_n X_n^{\pi(n)}. \tag{11}$$

Hence, choosing $\varepsilon > 0$ we obtain

$$\begin{aligned}
\mathsf{E} Z_n X_n^{\pi(n)} &= \mathsf{E} Z_n X_n^{\pi(n)} \Big(I\big(-c(n) \leqslant X_n^{\pi(n)} < 0\big) \\
&\qquad + I\big(0 \leqslant X_n^{\pi(n)} < \varepsilon\big) + I\big(X_n^{\pi(n)} \geqslant \varepsilon\big)\Big) \\
&\geqslant -c(n) + \mathsf{E} Z_n X_n^{\pi(n)} I(Z_n \geqslant \varepsilon) I(X_n^{\pi(n)} \geqslant \varepsilon) \\
&\geqslant -c(n) + \varepsilon^2 \mathsf{P}(X_n^{\pi(n)} \geqslant \varepsilon,\ Z_n \geqslant \varepsilon) \\
&\geqslant -c(n) + \varepsilon^2 [\mathsf{P}(X_n^{\pi(n)} \geqslant \varepsilon) - \mathsf{P}(Z_n, \varepsilon)],
\end{aligned}$$

and therefore

$$X_0^{\pi(n)} + c(n) + \varepsilon^2 \mathsf{P}(Z_n < \varepsilon) \geqslant \varepsilon^2 \mathsf{P}(X_n^{\pi(n)} \geqslant \varepsilon). \tag{12}$$

If conditions (2) and (3) in § 7a are satisfied, then we see from (12) that

$$\lim_{\varepsilon\downarrow 0} \varlimsup_n \inf_{Z_n \in \mathbb{Z}_n} \mathsf{P}(Z_n < \varepsilon) \geqslant \lim_{\varepsilon\downarrow 0} \varlimsup_n \mathsf{P}(X_n^{\pi(n)} \geqslant \varepsilon) \tag{13}$$

(because $(\widetilde{\mathsf{P}}_n)_{n\geqslant 1}$ is an arbitrary sequence in $\widetilde{\mathbb{P}}$.

Hence it is clear that if (9) and, of course, (10) hold, then the sequence of strategies $\pi = (\pi(n))_{n\geqslant 1}$ with properties (2)–(4) in § 3a cannot realize asymptotic arbitrage.

COROLLARY. *Let $(\widetilde{\mathsf{P}}_n)_{n\geqslant 1}$ be some sequence in $\widetilde{\mathbb{P}}$ and let $Z_\infty = \varliminf \dfrac{d\widetilde{\mathsf{P}}_n}{d\mathsf{P}_n}$. Then the condition $\mathsf{P}(Z_\infty > 0) = 1$ ensures the absence of asymptotic arbitrage.*

3. We now proceed to the general case of arbitrage-free n-markets (B^n, S^n), $n \geqslant 1$, defined on filtered probability spaces

$$(\Omega^n, \mathscr{F}^n, (\mathscr{F}_k^n)_{k\leqslant k(n)}, \mathsf{P}^n)$$

'of their own', where $\mathscr{F}^n_{k(n)} = \mathscr{F}^n$.

If $\widetilde{\mathsf{P}}^n_{k(n)}$ is a martingale measure, $\widetilde{\mathsf{P}}^n_{k(n)} \sim \mathsf{P}^n_{k(n)}$, and $Z^n_{k(n)} = \dfrac{d\widetilde{\mathsf{P}}^n_{k(n)}}{d\mathsf{P}^n_{k(n)}}$, then in a similar way to (12) we obtain

$$X_0^{\pi(n)} + c(n) + \varepsilon^2\, \mathsf{P}^n\big(Z^n_{k(n)} < \varepsilon\big) \geqslant \varepsilon^2\, \mathsf{P}^n\big(X^{\pi(n)}_{k(n)} \geqslant \varepsilon\big). \tag{14}$$

We set

$$\mathbb{Z}^n_{k(n)} = \left\{ Z^n_{k(n)} \colon Z^n_{k(n)} = \frac{d\widetilde{\mathsf{P}}^n_{k(n)}}{d\mathsf{P}^n_{k(n)}},\ \widetilde{\mathsf{P}}^n_{k(n)} \in \mathscr{P}\big(\mathsf{P}^n_{k(n)}\big) \right\}.$$

THEOREM 2. *Let $(\mathbb{B}, \mathbb{S}) = \{(B^n, S^n),\ n \geqslant 1\}$ be a 'large' locally arbitrage-free market. Then the condition*

$$\lim_{\varepsilon \downarrow 0} \overline{\lim_{n}} \inf_{Z^n_{k(n)} \in \mathbb{Z}^n_{k(n)}} \mathsf{P}^n\big(Z^n_{k(n)} < \varepsilon\big) = 0 \tag{15}$$

is necessary and sufficient for the absence of asymptotic arbitrage.

Proof. The sufficiency of (15) follows from (14) as in the 'stationary' case. The proof of the necessity see in § 3c.9.

§ 3c. Asymptotic Arbitrage and Contiguity

1. It is clear from our previous discussion of *arbitrage theory* in this chapter that the issue of the *absolute continuity of probability measures* plays an important role there. As will be obvious from what we write below, a crucial role in the *theory of asymptotic arbitrage* is that of the concept of *contiguity* of probability measures, one of important concepts used in asymptotic problems of mathematical statistics.

To introduce this concept in a most natural way we consider the *stationary case* first (see § 3b.2 for the definition and notation).

Let $(\widetilde{\mathsf{P}}_n)_{n\geqslant 1}$ be a sequence of (compatible) martingale measures in $\widetilde{\mathbb{P}}$. Assume that there exists also a measure $\widetilde{\mathsf{P}}$ on $(\Omega, \mathscr{F})$ such that $\widetilde{\mathsf{P}} \,|\, \mathscr{F}_n = \widetilde{\mathsf{P}}_n$, $n \geqslant 1$.

We recall (see Chapter V, § 3a) that two measures, $\widetilde{\mathsf{P}}$ and P, are said to be *locally equivalent* ($\widetilde{\mathsf{P}} \overset{\text{loc}}{\sim} \mathsf{P}$) if $\widetilde{\mathsf{P}}_n \sim \mathsf{P}$, $n \geqslant 1$. It should be pointed out here that the relation $\widetilde{\mathsf{P}} \overset{\text{loc}}{\sim} \mathsf{P}$ *does not mean* in general that $\widetilde{\mathsf{P}} \ll \mathsf{P}$, $\mathsf{P} \ll \widetilde{\mathsf{P}}$, or $\widetilde{\mathsf{P}} \sim \mathsf{P}$.

It is clear from Theorem 1 in § 3b that if

$$\mathsf{P}(Z_\infty > 0) = 1, \tag{1}$$

where $Z_\infty = \lim Z_n$, $Z_n = \dfrac{d\widetilde{\mathsf{P}}_n}{d\mathsf{P}_n}$, then there is no asymptotic arbitrage.

By the theorem in Chapter V, § 3a condition (1) is equivalent to the relation $\mathsf{P} \ll \widetilde{\mathsf{P}}$ (under our assumption that $\widetilde{\mathsf{P}} \overset{\mathrm{loc}}{\sim} \mathsf{P}$). Thus, we can regard the relation $\mathsf{P} \ll \widetilde{\mathsf{P}}$ (or (1) if one likes it better) as just an additional (to $\widetilde{\mathsf{P}} \overset{\mathrm{loc}}{\sim} \mathsf{P}$) restriction on the behavior of the probability measures associated with n-markets that prevents asymptotic arbitrage as $n \to \infty$.

In the case when there *exists no measure* $\widetilde{\mathsf{P}}$ on the probability space $(\Omega, \mathscr{F}, (\mathscr{F}_n)_{n\geqslant 1})$ with $\mathscr{F} = \bigvee \mathscr{F}_n$ such that $\widetilde{\mathsf{P}}|\mathscr{F}_n = \widetilde{\mathsf{P}}_n$, $n \geqslant 1$, the following definition is helpful for finding a counterpart to the assertion

$$\mathsf{P}(Z_\infty > 0) = 1 \iff \mathsf{P} \ll \widetilde{\mathsf{P}}. \tag{2}$$

DEFINITION 1. Let Q^n and $\widetilde{\mathsf{Q}}^n$, $n \geqslant 1$, be probability measures on measurable spaces $(E^n, \mathscr{E}^n)$. Then we say that the sequence $(\widetilde{\mathsf{Q}}^n)_{n\geqslant 1}$ is *contiguous* with respect to $(\mathsf{Q}^n)_{n\geqslant 1}$ (our notation is $(\widetilde{\mathsf{Q}}^n) \lhd (\mathsf{Q}^n)$) if for all sequences of sets $A^n \in \mathscr{E}^n$ such that $\mathsf{Q}^n(A^n) \to 0$ as $n \to \infty$ we have $\widetilde{\mathsf{Q}}^n(A^n) \to 0$ as $n \to \infty$.

Remark 1. In the case when the sets $(E^n, \mathscr{E}^n)$ and the measures Q^n and $\widetilde{\mathsf{Q}}^n$ are independent on n $((E^n, \mathscr{E}^n) \equiv (E, \mathscr{E})$, $\mathsf{Q}^n \equiv \mathsf{Q}$, $\widetilde{\mathsf{Q}}^n \equiv \widetilde{\mathsf{Q}})$ contiguity $(\widetilde{\mathsf{Q}}^n) \lhd (\mathsf{Q}^n)$ becomes the standard property of the absolute continuity $\widetilde{\mathsf{Q}} \ll \mathsf{Q}$ of the measures Q and $\widetilde{\mathsf{Q}}$ on $(E, \mathscr{E})$.

THEOREM 1 (the *stationary case*). *Let* $(\widetilde{\mathsf{P}}_n)$ *be a sequence of martingale measures in* $\widetilde{\mathbb{P}}$. *Then*

$$\mathsf{P}(Z_\infty > 0) = 1 \iff (\mathsf{P}_n) \lhd (\widetilde{\mathsf{P}}_n). \tag{3}$$

The condition of contiguity $(\mathsf{P}_n) \lhd (\widetilde{\mathsf{P}}_n)$ *ensures the absence of asymptotic arbitrage.*

If $(\widetilde{\mathsf{P}}_n)$ *is a unique martingale sequence, then the condition* $(\mathsf{P}_n) \lhd (\widetilde{\mathsf{P}}_n)$ *is necessary and sufficient for the absence of asymptotic arbitrage.*

Proof. This is an immediate consequence of Theorem 1 in the previous section and Lemma 1 below, which contains several useful contiguity criteria. For a formulation we require additional definitions and notation.

2. Let Q and $\widetilde{\mathsf{Q}}$ be two probability measures on the measurable space $(E, \mathscr{E})$, $\overline{\mathsf{Q}} = \frac{1}{2}(\mathsf{Q} + \widetilde{\mathsf{Q}})$. We set $\mathfrak{z} = \dfrac{d\mathsf{Q}}{d\overline{\mathsf{Q}}}$, $\tilde{\mathfrak{z}} = \dfrac{d\widetilde{\mathsf{Q}}}{d\overline{\mathsf{Q}}}$, and $Z = \dfrac{\tilde{\mathfrak{z}}}{\mathfrak{z}}$. (Here we choose representatives $\mathfrak{z}$ and $\tilde{\mathfrak{z}}$ of the Radon–Nikodym derivatives such that $\mathfrak{z} + \tilde{\mathfrak{z}} \equiv 2$.)

We recall that, in accordance with the *Lebesgue decomposition* (see, e.g., [439; Chapter III, § 9]), we can represent $\widetilde{\mathsf{Q}}$ as follows:

$$\widetilde{\mathsf{Q}} = \widetilde{\mathsf{Q}}_1 + \widetilde{\mathsf{Q}}_2.$$

where

$$\widetilde{\mathsf{Q}}_1(A) = \mathsf{E}_{\mathsf{Q}} Z I_A \quad \text{and} \quad \widetilde{\mathsf{Q}}_2(A) = \widetilde{\mathsf{Q}}(A \cap (Z = \infty)).$$

Since $\mathsf{Q}(Z < \infty) = 1$, it follows that $\widetilde{\mathsf{Q}}_1 \ll \mathsf{Q}$ and $\widetilde{\mathsf{Q}}_2 \perp \mathsf{Q}$.

Hence Z is just the Radon-Nikodym derivative of the absolutely continuous component of $\widetilde{\mathsf{Q}}$ with respect to the measure Q. $\Bigl($It is in this sense that we shall use the notation $Z = \dfrac{d\widetilde{\mathsf{Q}}}{d\mathsf{Q}}$.$\Bigr)$

DEFINITION 2. Let $\alpha \in (0, 1)$ and let

$$H(\alpha; \mathsf{Q}, \widetilde{\mathsf{Q}}) = \mathsf{E}_{\overline{\mathsf{Q}}}\, \mathfrak{z}^{\alpha} \tilde{\mathfrak{z}}^{1-\alpha}, \tag{4}$$

$$H(\mathsf{Q}, \widetilde{\mathsf{Q}}) = H\Bigl(\frac{1}{2}; \mathsf{Q}, \widetilde{\mathsf{Q}}\Bigr), \tag{5}$$

$$\rho(\mathsf{Q}, \widetilde{\mathsf{Q}}) = \sqrt{1 - H(\mathsf{Q}, \widetilde{\mathsf{Q}})}\,. \tag{6}$$

We call the quantity $H(\alpha; \mathsf{Q}, \widetilde{\mathsf{Q}})$ the *Hellinger integral of order* α (of the measures Q and $\widetilde{\mathsf{Q}}$); we call the quantity $H(\mathsf{Q}, \widetilde{\mathsf{Q}})$ simply the *Hellinger integral*, and $\rho(\mathsf{Q}, \widetilde{\mathsf{Q}})$ is the *Hellinger distance* between Q and $\widetilde{\mathsf{Q}}$.

It can be proved (see, e.g., [250; Chapter IV, § 1a]) that $H(\alpha; \mathsf{Q}, \widetilde{\mathsf{Q}})$ is in fact *independent* of the dominating measure $\overline{\mathsf{Q}}$. This explains the widespread symbolic notation

$$H(\alpha; \mathsf{Q}, \widetilde{\mathsf{Q}}) = \int_E (d\mathsf{Q})^{\alpha} (d\widetilde{\mathsf{Q}})^{1-\alpha}, \tag{7}$$

$$H(\mathsf{Q}, \widetilde{\mathsf{Q}}) = \int_E \sqrt{d\mathsf{Q}\, d\widetilde{\mathsf{Q}}}, \tag{8}$$

$$\rho(\mathsf{Q}, \widetilde{\mathsf{Q}}) = \frac{1}{2} \int_E \Bigl(\sqrt{d\mathsf{Q}} - \sqrt{d\widetilde{\mathsf{Q}}}\Bigr)^2. \tag{9}$$

EXAMPLE. Let $\mathsf{Q} = \mathsf{Q}_1 \times \mathsf{Q}_2 \times \cdots$, $\widetilde{\mathsf{Q}} = \widetilde{\mathsf{Q}}_1 \times \widetilde{\mathsf{Q}}_2 \times \cdots$, where Q_k and $\widetilde{\mathsf{Q}}_k$ are the Gaussian measures on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ with densities

$$q_k(x) = \frac{1}{\sqrt{2\pi}\,\sigma_k}\, e^{-\frac{(x-\mu_k)^2}{2\sigma_k^2}}$$

and

$$\widetilde{q}_k(x) = \frac{1}{\sqrt{2\pi}\,\sigma_k}\, e^{-\frac{(x-\widetilde{\mu}_k)^2}{2\sigma_k^2}}\,.$$

Then

$$\begin{aligned}H(\alpha; \mathsf{Q}_k, \widetilde{\mathsf{Q}}_k) &= \int_{\mathbb{R}} (d\mathsf{Q}_k)^{\alpha} (d\widetilde{\mathsf{Q}}_k)^{1-\alpha} = \int_{\mathbb{R}} \left(\frac{d\widetilde{\mathsf{Q}}_k}{d\mathsf{Q}_k}\right)^{1-\alpha} d\mathsf{Q}_k \\ &= \int_{\mathbb{R}} q_k^{\alpha}(x) \widetilde{q}_k^{\,1-\alpha}(x)\, dx = \exp\left\{-\frac{\alpha(1-\alpha)}{2}\left(\frac{\mu_k - \widetilde{\mu}_k}{\sigma_k}\right)^2\right\},\end{aligned}$$

and therefore

$$H(\alpha; \mathsf{Q}, \widetilde{\mathsf{Q}}) = \exp\left\{-\frac{\alpha(1-\alpha)}{2}\sum_{k=1}^{\infty}\left(\frac{\mu_k - \widetilde{\mu}_k}{\sigma_k}\right)^2\right\}. \tag{10}$$

LEMMA 1. *Let $(E^n, \mathscr{E}^n)$ be measurable spaces endowed with probability measures Q^n and $\widetilde{\mathsf{Q}}^n$, $n \geqslant 1$. Then the following conditions are equivalent* $\left(\text{here } \mathfrak{z}^n = \dfrac{d\mathsf{Q}^n}{d\overline{\mathsf{Q}}^n},\ Z^n = \dfrac{d\widetilde{\mathsf{Q}}^n}{d\mathsf{Q}^n}, \text{ and } \overline{\mathsf{Q}}^n = \dfrac{1}{2}(\mathsf{Q}^n + \widetilde{\mathsf{Q}}^n)\right)$:

a) $(\widetilde{\mathsf{Q}}^n) \lhd (\mathsf{Q}^n)$;

b) $\lim\limits_{\varepsilon\downarrow 0} \overline{\lim\limits_{n}}\, \widetilde{\mathsf{Q}}^n(\mathfrak{z}^n < \varepsilon) = 0$;

c) $\lim\limits_{N\uparrow\infty} \overline{\lim\limits_{n}}\, \widetilde{\mathsf{Q}}^n(Z^n > N) = 0$;

d) $\lim\limits_{\alpha\downarrow 0} \underline{\lim\limits_{n}}\, H(\alpha; \mathsf{Q}^n, \widetilde{\mathsf{Q}}^n) = 1$.

Proof. This can be found in [250; Chapter V, Lemma 1.6].

The proof of equivalence (3) in Theorem 1 is an immediate consequence of the equivalence of a) and c) in Lemma 1 in the case of $\mathsf{Q}^n = \widetilde{\mathsf{P}}_n$ and $\widetilde{\mathsf{Q}}^n = \mathsf{P}_n$:

$$\begin{aligned}(\mathsf{P}_n) \lhd (\widetilde{\mathsf{P}}_n) &\iff \lim_{N\uparrow\infty} \overline{\lim_{n}}\ \mathsf{P}\left(\frac{d\mathsf{P}_n}{d\widetilde{\mathsf{P}}_n} > N\right) = 0 \\ &\iff \lim_{\varepsilon\downarrow 0} \overline{\lim_{n}}\ \mathsf{P}\left(\frac{d\widetilde{\mathsf{P}}_n}{d\mathsf{P}_n} < \varepsilon\right) = 0 \\ &\iff \lim_{\varepsilon\downarrow 0} \mathsf{P}(Z_\infty < \varepsilon) = 0 \iff \mathsf{P}(Z_\infty > 0) = 1,\end{aligned}$$

where Z_∞ is equal to $\lim \dfrac{d\widetilde{\mathsf{P}}_n}{d\mathsf{P}_n}$, which exists P-a.s.

The absence of asymptotic arbitrage under the assumption $(\mathsf{P}_n) \lhd (\widetilde{\mathsf{P}}_n)$ is a *consequence of the corollary to Theorem 1 in § 3b and property (3) just proved.*

3. The extension of Theorem 1 to the general (*nonstationary*) case proceeds without complications.

We shall stick to the pattern put forward in § 3b.3.

THEOREM 2. *Let* $(\widetilde{\mathsf{P}}^n_{k(n)})_{n\geqslant 1}$ *be a 'chain' of martingale measures on* (B^n, S^n)-*markets (here* $\widetilde{\mathsf{P}}^n_{k(n)} \sim \mathsf{P}^n_{k(n)}$*). Then the condition* $(\mathsf{P}^n_{k(n)}) \lhd (\widetilde{\mathsf{P}}^n_{k(n)})$ *ensures the absence of asymptotic arbitrage.*

Proof. Since conditions a) and c) in Lemma 1 are equivalent, it follows that

$$(\mathsf{P}^n_{k(n)}) \lhd (\widetilde{\mathsf{P}}^n_{k(n)}) \iff \lim_{\varepsilon\downarrow 0} \overline{\lim_{n}}\, \mathsf{P}^n(Z^n_{k(n)} < \varepsilon) = 0, \tag{11}$$

where $Z^n_{k(n)} = \dfrac{d\widetilde{\mathsf{P}}^n_{k(n)}}{d\mathsf{P}^n_{k(n)}}$ and $\mathsf{P}^n_{k(n)} = \mathsf{P}^n \mid \mathscr{F}^n_{k(n)}$.

The required absence of asymptotic arbitrage in the case of the contiguity $(\mathsf{P}^n_{k(n)}) \lhd (\widetilde{\mathsf{P}}^n_{k(n)})$ is a consequence of (11) and Theorem 2 in § 3b.

4. So far, we have formulated conditions of the absence of asymptotic arbitrage in terms of the asymptotic properties of the likelihood ratios $Z^n_{k(n)}$ (Theorems 1 and 2 in § 3b) or in terms of contiguity (Theorems 1 and 2 in the present section). Lemma 1 above suggests a necessary and sufficient condition of the contiguity $(\widetilde{\mathsf{Q}}^n) \lhd (\mathsf{Q}^n)$ in terms of the asymptotic properties of the *Hellinger integrals* of order $\alpha \in (0, 1)$:

$$(\widetilde{\mathsf{Q}}^n) \lhd (\mathsf{Q}^n) \iff \lim_{\alpha\downarrow 0} \underline{\lim_{n}}\, H(\alpha; \mathsf{Q}^n, \widetilde{\mathsf{Q}}^n) = 1. \tag{12}$$

It is often easy to analyze this integral and to establish in that way the absence of asymptotic arbitrage. (See examples in subsection 5.) A simplest particular case here is that of the direct product of measures.

Namely, assume that

$$E^n = E^n_1 \times \cdots \times E^n_{k(n)}, \qquad \mathscr{E}^n = \mathscr{E}^n_1 \times \cdots \times \mathscr{E}^n_{k(n)},$$
$$\mathsf{Q}^n = \mathsf{Q}^n_1 \times \cdots \times \mathsf{Q}^n_{k(n)}, \qquad \widetilde{\mathsf{Q}}^n = \widetilde{\mathsf{Q}}^n_1 \times \cdots \times \widetilde{\mathsf{Q}}^n_{k(n)},$$

where Q^n_k and $\widetilde{\mathsf{Q}}^n_k$ are probability measures in $(E^n_k, \mathscr{E}^n_k)$.

Clearly,

$$\begin{aligned} H(\alpha; \mathsf{Q}^n, \widetilde{\mathsf{Q}}^n) &= \prod_{k=1}^{k(n)} H(\alpha; \mathsf{Q}^n_k, \widetilde{\mathsf{Q}}^n_k) \\ &= \prod_{k=1}^{k(n)} \bigl[1 - \bigl(1 - H(\alpha; \mathsf{Q}^n_k, \widetilde{\mathsf{Q}}^n_k)\bigr)\bigr] \end{aligned} \tag{13}$$

and

$$\lim_{\alpha\downarrow 0}\varliminf_{n} H(\alpha;\mathsf{Q}^n,\widetilde{\mathsf{Q}}^n)=1 \iff \lim_{\alpha\downarrow 0}\varlimsup_{n}\sum_{k=1}^{k(n)}\bigl(1-H(\alpha;\mathsf{Q}^n_k,\widetilde{\mathsf{Q}}^n_k)\bigr)=0. \tag{14}$$

Thus, in the case of a direct product we have

$$(\widetilde{\mathsf{Q}}^n)\lhd(\mathsf{Q}^n) \iff \lim_{\alpha\downarrow 0}\varlimsup_{n}\sum_{k=1}^{k(n)}\bigl(1-H(\alpha;\mathsf{Q}^n_k,\widetilde{\mathsf{Q}}^n_k)\bigr)=0. \tag{15}$$

5. We now present several examples, which on the one hand show the efficiency of the criteria of the absence of asymptotic arbitrage based on Hellinger integrals of order α and, on the other hand, can clarify the arguments and conclusions of *APT*, the theory described in Chapter I, § 2d. Examples 1 and 2 are particular cases of examples in [260].

EXAMPLE 1 ('Large' stationary market with $d(n)=1$ and $k(n)=n$). We consider the probability space $(\Omega,\mathscr{F},\mathsf{P})$, where $\Omega=\{-1,1\}^\infty$ is the set of binary sequences $x=(x_1,x_2,\dots)$ $(x_i=\pm 1)$ and P is a measure such that $\mathsf{P}\{x\colon (x_1,\dots,x_n)\}=2^{-n}$. Let $\varepsilon_i(x)=x_i$, $i=1,2,\dots$. Then $\varepsilon=(\varepsilon_1,\varepsilon_2,\dots)$ is a sequence of Bernoulli random variables with $\mathsf{P}(\varepsilon_i=\pm 1)=\frac{1}{2}$.

We assume that each (B^n,S^n)-market defined on $(\Omega,\mathscr{F}^n,\mathsf{P}^n)$ with $\mathscr{F}^n=\mathscr{F}_n=\sigma(\varepsilon_1,\dots,\varepsilon_n)$ and $\mathsf{P}^n=\mathsf{P}|\mathscr{F}^n$ has the properties $B^n_k\equiv 1$ and $S^n=(S_1,\dots,S_n)$ with

$$S_k=S_{k-1}(1+\rho_k)\quad\text{and}\quad S_0=1, \tag{16}$$

where $\rho_k=\mu_k+\sigma_k\varepsilon_k$, $\sigma_k>0$, and $\max(-\sigma_k,\sigma_{k-1})<\mu_k<\sigma_k$ (cf. condition (2) in Chapter V, § 1d).

We can rewrite (16) as follows:

$$S_k=S_{k-1}\bigl(1+\sigma_k(\varepsilon_k-b_k)\bigr) \tag{17}$$

where $b_k=-(\mu_k/\sigma_k)$. (Note that $|b_k|<1$.)

It follows from (17) and Theorem 2 in Chapter V, § 3f that there *exists* a *unique* martingale measure, which is a direct product $\widetilde{\mathsf{P}}^n=\widetilde{\mathsf{P}}^n_1\times\widetilde{\mathsf{P}}^n_2\times\dots\times\widetilde{\mathsf{P}}^n_n$ and has the following properties: the variables $\varepsilon_1,\varepsilon_2,\dots,\varepsilon_n$ are independent with respect to this measure and

$$\widetilde{\mathsf{P}}^n(\varepsilon_k=1)=\frac{1}{2}(1+b_k),\qquad \widetilde{\mathsf{P}}^n(\varepsilon_k=-1)=\frac{1}{2}(1-b_k).$$

Since

$$H(\alpha;\widetilde{\mathsf{P}}^n,\mathsf{P}^n)=\prod_{k=1}^{n}\biggl[\frac{(1+b_k)^\alpha+(1-b_k)^\alpha}{2}\biggr], \tag{18}$$

it follows by (12) and (15) that

$$(\mathsf{P}^n) \lhd (\widetilde{\mathsf{P}}^n) \iff \lim_{\alpha\downarrow 0} \underline{\lim_{n}} \sum_{k=1}^{n} \prod_{k=1}^{n} \left[\frac{(1+b_k)^\alpha + (1-b_k)^\alpha}{2}\right] = 1$$

$$\iff \lim_{\alpha\downarrow 0} \overline{\lim_{n}} \sum_{k=1}^{n} \prod_{k=1}^{n} \left[1 - \frac{(1+b_k)^\alpha + (1-b_k)^\alpha}{2}\right] = 0.$$

Hence it is easy to conclude that

$$(\mathsf{P}^n) \lhd (\widetilde{\mathsf{P}}^n) \iff \sum_{k=1}^{\infty} b_k^2 < \infty.$$

Recalling that $b_k = -\dfrac{\mu_k}{\sigma_k}$ and using Theorem 1 we see that the condition $\sum_{k=1}^{\infty} \left(\dfrac{\mu_k}{\sigma_k}\right)^2 < \infty$ is *necessary* and *sufficient* for the absence of asymptotic arbitrage on our 'large' stationary market.

EXAMPLE 2 ('Large' market with $k(n) \equiv 1$ and $d(n) = n$). Consider a single-step model of (B^n, S^n)-markets, where $B^n = (B_k^n)$, $S^n = (S_k^0, S_k^1, \dots, S_k^{n-1})$ with $k = 0$ or 1, $B_0^n = B_1^n = 1$, and

$$S_1^i = S_0^i(1+\rho^i), \qquad S_0^i > 0, \tag{19}$$

where

$$\rho^0 = \mu_0 + \sigma_0\varepsilon_0, \tag{20}$$

$$\rho^i = \mu_i + \sigma_i(c_i\varepsilon_0 + \bar{c}_i\varepsilon_i), \quad i \geqslant 1. \tag{21}$$

We also assume that $\sigma_i > 0$, $\bar{c}_i > 0$, $c_i^2 + \bar{c}_i^2 = 1$, and $\varepsilon = (\varepsilon_0, \varepsilon_1, \dots)$ is a sequence of independent Bernoulli random variables taking the values ± 1 with probabilities $\frac{1}{2}$.

With an eye on the theories *CAPM* and *APT*, we recommend to interpret S_k^i, $i \geqslant 1$, as the price at time k of some stock that is traded on a 'large', 'global' market and S_k^0 as some general index of this market (for example, the S&P500 Index of the market of the 500 stocks covered by it; see Chapter I, § 1b.6).

Let $\beta_i = \dfrac{c_i\sigma_i}{\sigma_0}$, $i \geqslant 1$, and let

$$b_0 = -\frac{\mu_0}{\sigma_0} \quad \text{and} \quad b_i = \frac{\mu_0\beta_i - \mu_i}{\sigma_i\bar{c}_i}, \tag{22}$$

where $|b_0| \neq 1$ and $|b_i| \neq 1$, $i \geqslant 1$.

Using this notation, we obtain by (19)–(21) that

$$S_1^0 = S_0^0\bigl(1 + \sigma_0(\varepsilon_0 - b_0)\bigr), \tag{23}$$

and

$$S_1^i = S_0^i\bigl(1 + \sigma_i c_i(\varepsilon_0 - b_0) + \sigma_i \bar{c}_i(\varepsilon_i - b_i)\bigr) \tag{24}$$

for $i \geqslant 1$.

Sticking to the above-described scheme of series of (B^n, S^n)-markets we can assume that each of the markets is defined on a probability space $(\Omega, \mathscr{F}^n, \mathsf{P}^n)$, where $\mathscr{F}^n = \sigma(\varepsilon_0, \varepsilon_1, \dots, \varepsilon_{n-1})$, $\mathsf{P}^n = \mathsf{P} \mid \mathscr{F}^n$, and Ω and P are the same as in the above example.

It is easy to see from (23) and (24) that, in the framework of this scheme, for each $n \geqslant 1$ there *exists* (at least one) martingale measure. For we can consider a measure $\widetilde{\mathsf{P}}^n$ (having again the structure of a direct product) such that the variables $\varepsilon_0, \varepsilon_1, \dots, \varepsilon_{n-1}$ are independent with respect to $\widetilde{\mathsf{P}}^n$, namely,

$$\widetilde{\mathsf{P}}^n(\varepsilon_i = 1) = \frac{1}{2}(1 + b_i), \quad \text{and} \quad \widetilde{\mathsf{P}}^n(\varepsilon_i = -1) = \frac{1}{2}(1 - b_i).$$

It is straightforward that

$$H(\alpha; \widetilde{\mathsf{P}}^n, \mathsf{P}^n) = \prod_{i=0}^{n-1}\left[\frac{(1 + b_i)^\alpha + (1 - b_i)^\alpha}{2}\right]. \tag{25}$$

As in the previous example, we conclude that

$$(\mathsf{P}^n) \lhd (\widetilde{\mathsf{P}}^n) \iff \sum_{i=0}^{\infty} b_i^2 < \infty,$$

so that, by Theorem 2, the condition

$$\sum_{i=1}^{\infty}\left(\frac{\mu_0\beta_i - \mu_i}{\sigma_i \bar{c}_i}\right)^2 < \infty \tag{26}$$

$\left(\text{in addition to } \left|\dfrac{\mu_0}{\sigma_0}\right| < 1 \text{ and } \left|\dfrac{\mu_0\beta_i - \mu_i}{\sigma_i \bar{c}_i}\right| < 1,\ i \geqslant 1\right)$ ensures the absence of the asymptotic arbitrage. (Compare formula (26) with (4) in Chapter I, § 2c and (19) in Chapter I, § 2d).

Remark 2. We point out one crucial distinction between the above examples. In the first of them, where $k \leqslant n$, there exists a *unique* martingale measure $\widetilde{\mathsf{P}}^n$, which allows us to claim (on the basis of Theorem 1) that the condition $\sum\limits_{k=1}^{\infty}\left(\dfrac{\mu_k}{\sigma_k}\right)^2 < \infty$ is necessary and sufficient for the absence of asymptotic arbitrage.

On the other hand, in the second example, where n is the index of a series, the measure $\widetilde{\mathsf{P}}^n$ is not unique for $n \geqslant 1$. This explains why (26) is only a *sufficient* condition of the absence of asymptotic arbitrage. $\Big($As shown in [260], the condition that is both necessary and sufficient has the form $\varliminf_i [\min(1+b_i, 1-b_i)] > 0.\Big)$

EXAMPLE 3. We consider a *stationary logarithmically Gaussian* (Chapter V, § 3c) market $(\mathbb{B}, \mathbb{S}) = \{(B^n, S^n),\ n \geqslant 1\}$ with $B_k^n \equiv 1$ and $S^n = (S_0, S_1, \dots, S_n)$, where

$$S_k = S_0 e^{h_1 + \cdots + h_k}, \qquad S_0 > 0. \tag{27}$$

We assume that $h_k = \mu_k + \sigma_k \varepsilon_k$, $k \geqslant 1$, where $(\varepsilon_1, \varepsilon_2, \dots)$ is a sequence of independent, normally distributed $(\mathscr{N}(0,1))$ random variables defined on a probability space $(\Omega, \mathscr{F}, \mathsf{P})$ and $\sigma_k > 0$, $k \geqslant 1$.

Let $\mathscr{F}_n = \sigma(\varepsilon_1, \dots, \varepsilon_n)$ and let $\mathsf{P}_n = \mathsf{P}|\mathscr{F}_n$, $n \geqslant 1$. It was shown in Chapter V, § 3c that if

$$Z_n = \exp\biggl\{-\sum_{k=1}^{n}\Bigl(\frac{\mu_k}{\sigma_k} + \frac{\sigma_k}{2}\Bigr)\varepsilon_k + \frac{1}{2}\sum_{k=1}^{n}\Bigl(\frac{\mu_k}{\sigma_k} + \frac{\sigma_k}{2}\Bigr)^2\biggr\}, \tag{28}$$

then the sequence of prices $(S_k)_{k \leqslant n}$ is a martingale with respect to the measure $\widetilde{\mathsf{P}}_n$ such that $d\widetilde{\mathsf{P}}_n = Z_n\, d\mathsf{P}_n$; moreover, $\operatorname{Law}(h_k \mid \widetilde{\mathsf{P}}_n) = \mathscr{N}(\widetilde{\mu}_k, \sigma_k)$, where

$$\widetilde{\mu}_k = -\frac{\sigma_k^2}{2}, \qquad k \leqslant n.$$

It is now easy to see from formula (10) that

$$H(\alpha; \widetilde{\mathsf{P}}_n, \mathsf{P}_n) = \exp\biggl\{-\frac{\alpha(1-\alpha)}{2}\sum_{k=1}^{n}\Bigl(\frac{\mu_k}{\sigma_k} + \frac{\sigma_k}{2}\Bigr)^2\biggr\}. \tag{29}$$

By (12) we obtain

$$(\mathsf{P}^n) \vartriangleleft (\widetilde{\mathsf{P}}^n) \iff \sum_{k=1}^{\infty}\Bigl(\frac{\mu_k}{\sigma_k} + \frac{\sigma_k}{2}\Bigr)^2 < \infty,$$

so that it follows from Theorem 1 that the condition

$$\sum_{k=1}^{\infty}\Bigl(\frac{\mu_k}{\sigma_k} + \frac{\sigma_k}{2}\Bigr)^2 < \infty \tag{30}$$

ensures the absence of asymptotic arbitrage.

Remark 3. If $\dfrac{\mu_k}{\sigma_k} + \dfrac{\sigma_k}{2} = 0$, i.e.,

$$\mu_k = -\frac{\sigma_k^2}{2}, \tag{31}$$

then the initial probability measure P is a martingale measure for $S = (S_k)_{k\geqslant 0}$.

Note that condition (30) is also necessary and sufficient for the relation $(\widetilde{\mathsf{P}}^n) \lhd (\mathsf{P}^n)$. Hence this condition is necessary and sufficient in order that the sequence of measures (P^n) and $(\widetilde{\mathsf{P}}^n)$ be mutually contiguous; we denote this property by $(\mathsf{P}^n) \lhd\rhd (\widetilde{\mathsf{P}}^n)$.

6. We now discuss briefly the concept of *complete asymptotic separability*, a natural 'asymptotic' counterpart of the concept of singularity.

DEFINITION 3. Let Q^n and $\widetilde{\mathsf{Q}}^n$, $n \geqslant 1$, be sequences of probability measures on measurable spaces $(E^n, \mathscr{E}^n)$. Then we say that $(\widetilde{\mathsf{Q}}^n)_{n\geqslant 1}$ and $(\mathsf{Q}^n)_{n\geqslant 1}$ have the property of *complete asymptotic separability* (and we write $(\widetilde{\mathsf{Q}}^n) \vartriangle (\mathsf{Q}^n)$) if there exists a subsequence (n_k), $n_k \uparrow \infty$ as $k \uparrow \infty$, such that for each k there exist a subset $A^{n_k} \in \mathscr{E}^{n_k}$ such that $\mathsf{Q}^{n_k}(A^{n_k}) \to 1$ and $\widetilde{\mathsf{Q}}^{n_k}(A^{n_k}) \to 0$ as $k \uparrow \infty$.

LEMMA 2. *Let $(E^n, \mathscr{E}^n)$ be measurable spaces endowed with probability measures Q^n and $\widetilde{\mathsf{Q}}^n$, $n \geqslant 1$. Then the following properties are equivalent* $\left(\textit{here } \mathfrak{z}^n = \dfrac{d\mathsf{Q}^n}{d\overline{\mathsf{Q}}^n},\ Z^n = \dfrac{d\widetilde{\mathsf{Q}}^n}{d\mathsf{Q}^n}, \textit{ and } \overline{\mathsf{Q}}^n = \dfrac{1}{2}(\mathsf{Q}^n + \widetilde{\mathsf{Q}}^n)\right)$:

a) $(\widetilde{\mathsf{P}}^n) \vartriangle (\mathsf{P}^n)$;

b) $\varliminf_n \widetilde{\mathsf{Q}}^n(\mathfrak{z}^n \geqslant \varepsilon) = 0$ *for all* $\varepsilon > 0$;

c) $\varlimsup_n \widetilde{\mathsf{Q}}^n(Z^n \leqslant N) = 0$ *for all* $N > 0$;

d) $\lim_{\alpha\downarrow 0} \varliminf_n H(\alpha; \mathsf{Q}^n, \widetilde{\mathsf{Q}}^n) = 0$;

e) $\varliminf_n H(\alpha; \mathsf{Q}^n, \widetilde{\mathsf{Q}}^n) = 0$ *for all* $\alpha \in (0,1)$;

f) $\varliminf_n H(\alpha; \mathsf{Q}^n, \widetilde{\mathsf{Q}}^n) = 0$ *for some* $\alpha \in (0,1)$.

Proof. This can be found in [250; Chapter V, Lemma 1.9].

7. Our analysis in Examples 1–3 of the cases when there is no asymptotic arbitrage demonstrates the efficiency of criteria formulated in terms of the asymptotic properties of Hellinger integrals of order $\alpha > 0$.

For *filtered* probability spaces (as in Examples 1 and 3) it can be useful to consider also the so-called *Hellinger process*: we can also formulate criteria of absolute continuity, continuity, and other properties of probability measures with respect to one another in terms of such a process.

We now present what can be regarded as an introduction into the range of issues related to Hellinger processes by means of a *discrete-time* example. (See [250; Chapters IV and V] for greater detail.)

Let P and $\widetilde{\mathsf{P}}$ be probability measures on a filtered measurable space $(\Omega, \mathscr{F}, (\mathscr{F}_n)_{n\geqslant 0})$, where $\mathscr{F}_0 = \{\varnothing, \Omega\}$ and $\mathscr{F} = \bigvee \mathscr{F}_n$.

Let $\mathsf{P}_n = \mathsf{P} \,|\, \mathscr{F}_n$ and $\widetilde{\mathsf{P}}_n = \widetilde{\mathsf{P}} \,|\, \mathscr{F}_n$ be their restrictions to $\mathscr{F}_n$, $n \geqslant 1$, let $\mathsf{Q} = \frac{1}{2}(\mathsf{P} + \widetilde{\mathsf{P}})$, and let $\mathsf{Q}_n = \mathsf{Q} \,|\, \mathscr{F}_n$.

We set $\mathfrak{z}_n = \dfrac{d\mathsf{P}_n}{d\mathsf{Q}_n}$, $\tilde{\mathfrak{z}}_n = \dfrac{d\widetilde{\mathsf{P}}_n}{d\mathsf{Q}_n}$, $\beta_n = \dfrac{\mathfrak{z}_n}{\mathfrak{z}_{n-1}}$, and $\widetilde{\beta}_n = \dfrac{\tilde{\mathfrak{z}}_n}{\tilde{\mathfrak{z}}_{n-1}}$ (here we set $0/0 = 0$; recall that $\mathfrak{z}_n = 0$ if $\mathfrak{z}_{n-1} = 0$ and $\tilde{\mathfrak{z}}_n = 0$ if $\tilde{\mathfrak{z}}_{n-1} = 0$).

Using this notation we can express the Hellinger integral $H_n(\alpha) \equiv H(\alpha; \mathsf{P}_n, \widetilde{\mathsf{P}}_n)$ of order α as follows:

$$H_n(\alpha) = \mathsf{E}_{\mathsf{Q}}\, \mathfrak{z}_n^{\alpha}\, \tilde{\mathfrak{z}}_n^{1-\alpha}. \tag{32}$$

We consider now the process $Y(\alpha) = (Y_n(\alpha))_{n\geqslant 0}$ of the variables

$$Y_n(\alpha) = \mathfrak{z}_n^{\alpha}\, \tilde{\mathfrak{z}}_n^{1-\alpha}. \tag{33}$$

Let $f_\alpha(u, v) = u^\alpha v^{1-\alpha}$. This function is downwards convex (for $u \geqslant 0$ and $v \geqslant 0$), and therefore we have

$$\mathsf{E}_{\mathsf{Q}}(Y_n(\alpha) \,|\, \mathscr{F}_m) \leqslant Y_m(\alpha) \tag{34}$$

(Q-a.s.) for $m \leqslant n$ by Jensen's inequality.

Hence the sequence $Y(\alpha) = (Y_n(\alpha), \mathscr{F}_n, \mathsf{Q})$ is a (bounded) supermartingale, which, in view of the Doob decomposition (Chapter II, § 1b), can be represented as follows:

$$Y_n(\alpha) = M_n(\alpha) - A_n(\alpha), \tag{35}$$

where $M(\alpha) = (M_n(\alpha), \mathscr{F}_n, \mathsf{Q})$ is a martingale and $A(\alpha) = (A_n(\alpha), \mathscr{F}_{n-1}, \mathsf{Q})$ is a nondecreasing predictable process with $A_0(\alpha) = 0$.

The specific structure of $Y(\alpha)$ (see (33)) enables us to give the following representation of the predictable process $A(\alpha)$:

$$A_n(\alpha) = \sum_{k=1}^{n} Y_{k-1}(\alpha)\, \Delta h_k(\alpha) \tag{36}$$

where $h(\alpha) = (h_k(\alpha))_{k\geqslant 0}$, $h_0(\alpha) = 0$, is some nondecreasing predictable process.

In general, there is no canonical way to define this process. For instance, both processes

$$h_n(\alpha) = \sum_{k=1}^{n} \mathsf{E}_{\mathsf{Q}}\big(1 - \beta_k^{\alpha} \widetilde{\beta}_k^{1-\alpha} \,|\, \mathscr{F}_{k-1}\big) \tag{37}$$

and

$$h_n(\alpha) = \sum_{k=1}^{n} \mathsf{E}_{\mathsf{Q}}\big(\varphi_\alpha(\beta_k, \widetilde{\beta}_k) \,|\, \mathscr{F}_{k-1}\big), \tag{38}$$

where $\varphi_\alpha(u,v) = \alpha u + (1-\alpha)v - u^\alpha v^{1-\alpha}$, $0 < \alpha < 1$, satisfy the above requirements. This can be proved by a direct verification of the fact that the process $M(\alpha) = (M_n(\alpha), \mathscr{F}_n, \mathsf{Q})$ with

$$M_n(\alpha) = Y_n(\alpha) + \sum_{k=1}^{n} Y_{k-1}(\alpha)\,\Delta h_k(\alpha) \tag{39}$$

is a martingale for both of them. (See also [250; Chapter IV, § 1e].)

DEFINITION 4. By a *Hellinger process of order* $\alpha \in (0,1)$ we mean an arbitrary predictable process $h(\alpha) = (h_k(\alpha))_{k\geqslant 0}$, $h_0(\alpha) = 0$ such that the process $M(\alpha) = (M_k(\alpha), \mathscr{F}_k, \mathsf{Q})_{k\geqslant 0}$ defined by (39) is a martingale.

Remark 4. Assume that $\widetilde{\mathsf{P}} \overset{\text{loc}}{\ll} \mathsf{P}$, i.e., $\widetilde{\mathsf{P}}_n \ll \mathsf{P}_n$ for $n \geqslant 0$. Let $Z_n = \dfrac{d\widetilde{\mathsf{P}}_n}{d\mathsf{P}_n}$ and let $\rho_n = \dfrac{Z_n}{Z_{n-1}}$. Then the two processes $h_n(\alpha)$ defined by (37) and (38), respectively, have the following representations:

$$h_n(\alpha) = \sum_{k=1}^{n} \mathsf{E}_{\mathsf{P}}\big(1 - \rho_k^{1-\alpha} \,|\, \mathscr{F}_{k-1}\big) \tag{40}$$

and

$$h_n(\alpha) = \sum_{k=1}^{n} \mathsf{E}_{\mathsf{P}}\big(\varphi_\alpha(1, \rho_k) \,|\, \mathscr{F}_{k-1}\big). \tag{41}$$

Remark 5. We consider now the 'direct product scheme' by setting $\Omega = E_1 \times E_2 \times \cdots$, $\mathscr{F} = \mathscr{E}_1 \otimes \mathscr{E}_2 \otimes \cdots$, $\mathsf{P} = \mathsf{Q}_1 \times \mathsf{Q}_2 \times \cdots$, and $\widetilde{\mathsf{P}} = \widetilde{\mathsf{Q}}_1 \times \widetilde{\mathsf{Q}}_2 \times \cdots$, where Q_i and $\widetilde{\mathsf{Q}}_i$ are probability measures on $(E_i, \mathscr{E}_i)$.

In this case we set $\mathscr{F}_n = \mathscr{E}_1 \otimes \cdots \otimes \mathscr{E}_n$, $\mathsf{P}_n = \mathsf{Q}_1 \times \cdots \times \mathsf{Q}_n$, and $\widetilde{\mathsf{P}}_n = \widetilde{\mathsf{Q}}_1 \times \cdots \times \widetilde{\mathsf{Q}}_n$. Then the property $\widetilde{\mathsf{P}} \overset{\text{loc}}{\ll} \mathsf{P}$ is equivalent to the relation $\widetilde{\mathsf{Q}}_n \ll \mathsf{Q}_n$, $n \geqslant 1$, and we set $\rho_n = \dfrac{d\widetilde{\mathsf{Q}}_n}{d\mathsf{Q}_n}$.

Since $\mathsf{E}_{\mathsf{P}}\rho_n = 1$, the right-hand sides of (36) and (37) are the same. The corresponding Hellinger process $h(\alpha) = (h_n(\alpha))$ with

$$h_n(\alpha) = \sum_{k=1}^{n} \mathsf{E}_{\mathsf{P}}\big(1 - \rho_k^{1-\alpha}\big), \qquad n \geqslant 1,$$

is *deterministic* and

$$h_n(\alpha) = \sum_{k=1}^{n} \bigl(1 - H(\alpha; \mathsf{Q}_k, \widetilde{\mathsf{Q}}_k)\bigr). \tag{42}$$

If $H_n(\alpha) = H(\alpha; \mathsf{P}_n, \widetilde{\mathsf{P}}_n)$, then in our 'direct product case' we have

$$\begin{aligned} H_n(\alpha) &= H_{n-1}(\alpha) H(\alpha; \mathsf{Q}_n; \widetilde{\mathsf{Q}}_n) \\ &= H_{n-1}(\alpha)\bigl[1 - \bigl(1 - H(\alpha; \mathsf{Q}_n, \widetilde{\mathsf{Q}}_n)\bigr)\bigr]. \end{aligned}$$

Using the notation (42) we obtain

$$\Delta H_n(\alpha) = -H_{n-1}(\alpha)\, \Delta h_n(\alpha).$$

We already encountered difference equations of this kind in Chapter II (see formula (11) in § 1a). We expressed their solution in terms of *stochastic exponentials*:

$$H_n(\alpha) = H_0(\alpha) \mathscr{E}\bigl(-h(\alpha)\bigr)_n,$$

where

$$\begin{aligned} \mathscr{E}\bigl(-h(\alpha)\bigr)_n &= e^{-h_n(\alpha)} \prod_{k=1}^{n} \bigl(1 - \Delta h_k(\alpha)\bigr) e^{\Delta h_k(\alpha)} \\ &= \prod_{k=1}^{n} \bigl(1 - \Delta h_k(\alpha)\bigr) \quad \Bigl(= \prod_{k=1}^{n} H(\alpha; \mathsf{Q}_k, \widetilde{\mathsf{Q}}_k)\Bigr). \end{aligned}$$

This agrees completely with the expansion

$$H_n(\alpha) = H_0(\alpha) \prod_{k=1}^{n} H(\alpha; \mathsf{Q}_k, \widetilde{\mathsf{Q}}_k)$$

holding in the present case.

The next results (relating to the 'scheme of series') reveal the role of the stochastic exponential in the issues of the contiguity and complete asymptotic separability of sequences of probability measures $(\mathsf{P}^n)_{n\geqslant 1}$ and $(\widetilde{\mathsf{P}}^n)_{n\geqslant 1}$ on filtered measurable spaces $(\Omega^n, \mathscr{F}^n, (\mathscr{F}^n_k)_{k\leqslant k(n)})$, $n \geqslant 1$, with $\mathscr{F}^n = \mathscr{F}^n_{k(n)}$ and $\mathscr{F}^n_0 = \{\varnothing, \Omega^n\}$.

By analogy with (39) (but modifying our notation in an obvious way so as to fall in the scheme of series) we shall denote by

$$h^n_{k(n)}(\alpha) = \sum_{k=1}^{k(n)} \mathsf{E}_{\mathsf{Q}^n}\bigl(1 - (\beta^n_k)^{\alpha} (\widetilde{\beta}^n_k)^{1-\alpha} \,|\, \mathscr{F}^n_{k-1}\bigr) \tag{42$'$}$$

a Hellinger process of order $\alpha \in (0,1)$ corresponding to the measures P^n and $\widetilde{\mathsf{P}}^n$.

LEMMA 3. *The following conditions are equivalent:*

a) $(\widetilde{\mathsf{P}}^n) \lhd (\mathsf{P}^n)$;

b) $(\widetilde{\mathsf{P}}_0^n) \lhd (\mathsf{P}_0^n)$ *and*

$$\lim_{\alpha\downarrow 0} \overline{\lim_n}\, \widetilde{\mathsf{P}}^n\big(h_{k(n)}^n(\alpha) > \varepsilon\big) = 0 \tag{43}$$

for each $\varepsilon > 0$.

COROLLARY. *Consider a stationary case, when* P *and* $\widetilde{\mathsf{P}}$ *are probability measures on a filtered measurable space* $(\Omega, \mathscr{F}, (\mathscr{F}_k)_{k\geqslant 0})$. *Let* $\mathsf{P}_N = \mathsf{P}\,|\,\mathscr{F}_N$ *and* $\widetilde{\mathsf{P}}_N = \widetilde{\mathsf{P}}\,|\,\mathscr{F}_N$.

Then, for each $N \geqslant 1$ *we have the absolute continuity* $\widetilde{\mathsf{P}}_N \ll \mathsf{P}_N$ *if and only if*

$$\widetilde{\mathsf{P}}_0 \ll \mathsf{P}_0 \quad \textit{and} \quad h_N(\alpha) \xrightarrow{\widetilde{\mathsf{P}}} 0 \quad \textit{as} \quad \alpha \downarrow 0. \tag{44}$$

LEMMA 4. *If*

$$\overline{\lim_n}\, \widetilde{\mathsf{P}}^n\big(h_{k(n)}^n(\tfrac{1}{2}) > N\big) = 1 \tag{45}$$

for each $N \geqslant 0$, *then* $(\widetilde{\mathsf{P}}^n) \bigtriangleup (\mathsf{P}^n)$.

COROLLAY. *In the stationary case both conditions* $\widetilde{\mathsf{P}}_0 \perp \mathsf{P}_0$ *and*

$$\widetilde{\mathsf{P}}\big(h_N(\tfrac{1}{2}) = \infty\big) = 1 \tag{46}$$

are sufficient for the relation $\widetilde{\mathsf{P}}_N \perp \mathsf{P}_N$.

Criteria for the properties $\widetilde{\mathsf{P}} \ll \mathsf{P}$ and $\widetilde{\mathsf{P}} \perp \mathsf{P}$ take a particularly simple form if we assume additionally that $\widetilde{\mathsf{P}} \overset{\text{loc}}{\ll} \mathsf{P}$ (i.e., $\widetilde{\mathsf{P}}_n \ll \mathsf{P}_n$, $n \geqslant 0$). Namely,

$$\widetilde{\mathsf{P}} \ll \mathsf{P} \iff \widetilde{\mathsf{P}}\bigg\{\sum_{k=1}^{\infty} \mathsf{E}_{\mathsf{P}}\big[(1-\sqrt{\alpha_n}\,)^2 \,|\, \mathscr{F}_{n-1}\big] < \infty\bigg\} = 1,$$

$$\widetilde{\mathsf{P}} \perp \mathsf{P} \iff \widetilde{\mathsf{P}}\bigg\{\sum_{k=1}^{\infty} \mathsf{E}_{\mathsf{P}}\big[(1-\sqrt{\alpha_n}\,)^2 \,|\, \mathscr{F}_{n-1}\big] = \infty\bigg\} = 1,$$

where $\alpha_n = \dfrac{d\widetilde{\mathsf{P}}_n}{d\mathsf{P}_n}$.

The proofs of Lemmas 3 and 4 and the corollaries to them can be found in [250; Chapter V, § 2c and Chapter IV, § 2c].

8. We proceed now to the proof of the *necessity* of conditions (9) and (15) in Theorems 1 and 2 of § 3b.

It can be recommended to this end to use the following generalization of the concept of contiguity introduced in [260] in connection with asymptotic arbitrage on *incomplete* markets.

For each $n \geqslant 0$ let $(E^n, \mathscr{E}^n)$ be a measurable space with probability measure $\widetilde{\mathsf{Q}}^n$ and let $\mathbb{Q}^n = \{\mathsf{Q}^n\}$ be a *family* of probability measures Q^n in this space. (In what follows $\mathbb{Q}^n$ will always be a family $\mathscr{P}(\mathsf{P}^n)$ of *martingale* measures.)

We associate with a family $\mathbb{Q}^n = \{\mathsf{Q}^n\}$ of measures Q^n their *upper envelope* $\sup \mathsf{Q}^n$, the function of sets $A \in \mathscr{E}^n$ such that

$$(\sup \mathsf{Q}^n)(A) = \sup_{\mathsf{Q}^n \in \mathbb{Q}^n} \mathsf{Q}^n(A). \tag{47}$$

We shall also denote by $\operatorname{conv} \mathbb{Q}^n$ the convex hull of $\mathbb{Q}^n$.

DEFINITION 5. We say that a sequence of measures $(\widetilde{\mathsf{Q}}^n)_{n\geqslant 1}$ is *contiguous to the sequence of upper envelopes* $(\sup \mathsf{Q}^n)_{n\geqslant 1}$ (we write $(\widetilde{\mathsf{Q}}^n) \lhd (\sup \mathsf{Q}^n)$) if for each sequence of sets $A^n \in \mathscr{E}^n$, $n \geqslant 1$ such that $(\sup \mathsf{Q}^n)(A^n) \to 0$ we also have $\widetilde{\mathsf{Q}}^n(A^n) \to 0$.

Assume that

$$\mathfrak{z}^n(\mathsf{Q}) = \frac{d\mathsf{Q}}{d\overline{\mathsf{Q}}^n}, \qquad Z^n(\mathsf{Q}) = \frac{d\widetilde{\mathsf{Q}}^n}{d\mathsf{Q}}$$

for $\mathsf{Q} \in \operatorname{conv} \mathbb{Q}^n$, where $\overline{\mathsf{Q}}^n = \frac{1}{2}(\mathsf{Q} + \widetilde{\mathsf{Q}}^n)$.

The following result of [260] is a straightforward generalization of Lemma 1.

LEMMA 5. *The following conditions are equivalent:*

a) $(\widetilde{\mathsf{Q}}^n) \lhd (\sup \mathsf{Q}^n)$;

b) $\displaystyle \lim_{\varepsilon\downarrow 0} \varlimsup_n \inf_{\mathsf{Q}\in\operatorname{conv}\mathbb{Q}^n} \widetilde{\mathsf{Q}}^n\big(\mathfrak{z}^n(\mathsf{Q}) < \varepsilon\big) = 0;$

c) $\displaystyle \lim_{N\uparrow\infty} \varlimsup_n \inf_{\mathsf{Q}\in\operatorname{conv}\mathbb{Q}^n} \widetilde{\mathsf{Q}}^n\big(Z^n(\mathsf{Q}) > N\big) = 0;$

d) $\displaystyle \lim_{\alpha\downarrow 0} \varliminf_n \sup_{\mathsf{Q}\in\operatorname{conv}\mathbb{Q}^n} \mathsf{H}\big(\alpha; \mathsf{Q}, \widetilde{\mathsf{Q}}^n\big) = 1.$

9. Proceeding to the direct proof of the necessity of (9) and (15) in Theorems 1 and 2 in § 3b we shall set $\widetilde{\mathsf{Q}}^n = \mathsf{P}^n$ $(\equiv \mathsf{P}^n_{k(n)})$ and, as already mentioned, we take the family $\mathscr{P}(\mathsf{P}^n)$ of all martingale measures $\widetilde{\mathsf{P}}^n_{k(n)}$ as $\mathbb{Q}^n$, $n \geqslant 1$.

Clearly, $\mathscr{P}(\mathsf{P}^n)$ is also equal to $\operatorname{conv} \mathbb{Q}^n$ in this case, therefore condition c) in Lemma 5 takes the following form:

$$\lim_{N\uparrow\infty} \varlimsup_n \inf_{\widetilde{\mathsf{P}}^n_{k(n)}\in\mathscr{P}(\mathsf{P}^n_{k(n)})} \mathsf{P}^n_{k(n)}\left(\frac{d\mathsf{P}^n_{k(n)}}{d\widetilde{\mathsf{P}}^n_{k(n)}} > N\right) = 0,$$

which is equivalent to the following condition (since $\widetilde{\mathsf{P}}^n_{k(n)} \sim \mathsf{P}^n_{k(n)}$):

$$\lim_{\varepsilon \downarrow 0} \varlimsup_n \inf_{Z^n_{k(n)} \in \mathbb{Z}^n_{k(n)}} \mathsf{P}^n_{k(n)}\bigl(Z^n_{k(n)} < \varepsilon\bigr) = 0, \tag{48}$$

where

$$\mathbb{Z}^n_{k(n)} = \left\{ Z^n_{k(n)} : Z^n_{k(n)} = \frac{d\widetilde{\mathsf{P}}^n_{k(n)}}{d\mathsf{P}^n_{k(n)}},\ \widetilde{\mathsf{P}}^n_{k(n)} \in \mathscr{P}(\mathsf{P}^n_{k(n)}) \right\}.$$

Since (48) is precisely the same as (15) in § 3b, it follows by the equivalence of conditions a) and c) in Lemma 5 that the 'sufficiency' part of Theorem 2 in § 3b can be reformulated as follows: *the condition*

$$\bigl(\mathsf{P}^n_{k(n)}\bigr) \lhd \bigl(\sup \widetilde{\mathsf{P}}^n_{k(n)}\bigr) \tag{49}$$

means the absence of asymptotic arbitrage.

Hence (again, by the equivalence of a) and c) in Lemma 5) to prove the necessity in this theorem we must show that *the absence of asymptotic arbitrage means* (49).

We shall carry out a proof by contradiction.

Picking a subsequence if necessary we can assume that the sets $A^n \in \mathscr{F}^n_{k(n)}$ satisfy the relation

$$\bigl(\sup \widetilde{\mathsf{P}}^n_{k(n)}\bigr)(A^n) \to 0, \tag{50}$$

but $\mathsf{P}^n_{k(n)} \to \alpha > 0$.

We claim that *we have* asymptotic arbitrage in this case.

For a proof we consider the process

$$X^n_k = \operatorname*{ess\,sup}_{\widetilde{\mathsf{P}}^n_{k(n)} \in \mathscr{P}(\mathsf{P}^n_{k(n)})} \mathsf{E}_{\widetilde{\mathsf{P}}^n_{k(n)}} \bigl(I_{A^n} \mid \mathscr{F}^n_k\bigr), \qquad k \leqslant k(n). \tag{51}$$

By the theorem in § 2b, $X^n = (X^n_k)$ is a supermartingale with respect to each measure $\widetilde{\mathsf{P}}^n_{k(n)} \in \mathscr{P}(\mathsf{P}^n_{k(n)})$ and by the theorem in § 2d it has an *optional* decomposition

$$X^n_k = X^n_0 + \sum_{j=1}^k (\gamma^n_j, \Delta S^n_j) - C^n_k, \tag{52}$$

where $C^n_0 = 0$, the C^n_k are $\mathscr{F}^n_k$-measurable, and the γ^n_k are $\mathscr{F}^n_{k-1}$-measurable.

Based on (52), we define (for each $n \geqslant 1$) a strategy $\pi^n = (\beta^n, \gamma^n)$ with $\beta^n = (\beta^n_k)_{k \geqslant 0}$ and $\gamma^n = (\gamma^n_k)_{k \geqslant 0}$ such that its value $X^{\pi^n}_k$ is equal to $X^n_0 + \sum\limits_{j=1}^n (\gamma^n_j, \Delta S^n_j)$.

To this end it suffices to choose β^n_0 and γ^n_0 such that $\beta^n_0 + (\gamma^n_0, S^n_0) = X^n_0$ (for simplicity we always set $B^n_k \equiv 1$), to take the γ^n_k with $k \geqslant 1$ as in the decomposition (52), and to define β^n_k from the condition of self-financing.

For such strategies π^n we clearly have $X_k^{\pi^n} = X_k^n + C_k^n \geqslant 0$ for all $k \leqslant k(n)$, and as $n \to \infty$ we have

$$X_0^{\pi^n} = \sup_{\widetilde{\mathsf{P}}^n_{k(n)} \in \mathscr{P}(\mathsf{P}^n_{k(n)})} \mathsf{E}_{\widetilde{\mathsf{P}}^n_{k(n)}} I_{A^n} = \sup_{\widetilde{\mathsf{P}}^n_{k(n)} \in \mathscr{P}(\mathsf{P}^n_{k(n)})} \widetilde{\mathsf{P}}^n_{k(n)}(A^n) \to 0$$

by assumption (50).

Thus, conditions (2) and (3) in Definition 1 in § 3a hold for the strategy $\pi = (\pi^n)_{n\geqslant 1}$.

To complete the proof it remains to observe that the strategy $\pi = (\pi^n)_{n\geqslant 1}$ satisfies also condition (4) in the same Definition 1 because

$$\varlimsup_n \mathsf{P}^n(X^n_{k(n)} \geqslant 1) \geqslant \varlimsup_n \mathsf{P}^n(X^n_{k(n)} = 1) = \lim_n \mathsf{P}^n(A^n) = \alpha > 0.$$

This proves the necessity in both Theorems 2 and 1.

COROLLARY. To emphasize once again the importance of the concept of contiguity in the problems of asymptotic arbitrage in models of 'large' financial markets, we put Theorem 2 in the following (equivalent) form: *the condition* $(\mathsf{P}^n_{k(n)}) \lhd (\sup \widetilde{\mathsf{P}}^n_{k(n)})$ *is necessary and sufficient for the absence of asymptotic arbitrage on a 'large' locally arbitrage-free market* $(\mathbb{B}, \mathbb{S}) = \{(B^n, S^n),\ n \geqslant 1\}$.

Clearly, for a *complete* market this is the 'standard' condition of the contiguity

$$(\mathsf{P}^n_{k(n)}) \lhd (\widetilde{\mathsf{P}}^n_{k(n)}) \tag{53}$$

of the family $(\mathsf{P}^n_{k(n)})_{n\geqslant 1}$ of original probability measures $\mathsf{P}^n_{k(n)}$ to the family $(\widetilde{\mathsf{P}}^n_{k(n)})_{n\geqslant 1}$ of martingale measures $\widetilde{\mathsf{P}}^n_{k(n)}$ (which are unique for each $n \geqslant 1$).

Remark 6. In Theorems 1 and 2 we formulated sufficient conditions of the absence of asymptotic arbitrage in terms of the contiguity (53) to *some* 'chain' of martingale measures $(\widetilde{\mathsf{P}}^n_{k(n)})$.

It is worth noting that, as shown in [260], the converse result is in fact also true: if we have asymptotic arbitrage, then there *exists* a 'chain' of martingale measures $(\widetilde{\mathsf{P}}^n_{k(n)})_{n\geqslant 1}$ satisfying condition of contiguity (53).

§ 3d. Some Issues of Approximation and Convergence in the Scheme of Series of Arbitrage-Free Markets

1. In the models of 'large' financial markets discussed in § § 3a, b, c we always assume that there exists a *scheme of series* of n-markets (B^n, S^n), $n \geqslant 1$, each of which is arbitrage-free; after that we consider the question of the absence of *asymptotic arbitrage.* It should be noted that we make there no assumptions about the *existence* of a 'limit' market (B, S).

In the present section we discuss the case when, besides 'prelimit' n-markets (B^n, S^n) defined on the probability spaces $(\Omega^n, \mathscr{F}^n, \mathsf{P}^n)$, $n \geqslant 1$, we also have a 'limit' market (B, S) defined on $(\Omega, \mathscr{F}, \mathsf{P})$ and the (weak) convergence

$$\mathrm{Law}(B^n, S^n \,|\, \mathsf{P}^n) \to \mathrm{Law}(B, S \,|\, \mathsf{P}) \tag{1}$$

as $n \to \infty$.

We are mainly interested in understanding (under the assumption (1)) the issue of the convergence

$$\mathrm{Law}(B^n, S^n \,|\, \widetilde{\mathsf{P}}^n) \to \mathrm{Law}(B, S \,|\, \widetilde{\mathsf{P}}), \tag{2}$$

where $\widetilde{\mathsf{P}}^n$ and $\widetilde{\mathsf{P}}$ are some or other martingale measures in the classes $\mathscr{P}(\mathsf{P}^n)$ and $\mathscr{P}(\mathsf{P})$, respectively, and in finding appropriate measures $\widetilde{\mathsf{P}}^n$ and $\widetilde{\mathsf{P}}$ such that the convergence (2) can be ensured.

It seems timely to recall in this connection that we have already come across *various* methods of the construction of martingale measures, based, for example, on the Girsanov and the Esscher transformations. We also recall that the concept of *minimal* martingale measure discussed in Chapter V, § 3d, has been developed (in works of H. Föllmer and M. Schweizer, e.g., [167] and [429]) just in connection with the question on martingale measures in $\mathscr{P}(\mathsf{P}^n)$ 'eligible' to enter the construction of chains of measures $(\widetilde{\mathsf{P}}^n)_{n\geqslant 1}$ used for financial calculations. (It seems appropriate to point out here that, actually, one uses *martingale measures* $\widetilde{\mathsf{P}}^n$ and $\widetilde{\mathsf{P}}$ rather then the original—*physical*, as one puts it sometimes—measures P^n and P in, say, pricing of hedges or rational option pricing; see, e.g., the main pricing formula for European-type hedges on incomplete markets ((8) in § 1c) or formula (20) in § 4b.)

2. Proceeding to a discussion of the above questions we should recall the following two 'classical' models of (B, S)-markets distinguished by their simplicity and popularity in the finances literature:

the *Cox–Ross–Rubinstein model*

(or the binomial model; see Chapter II, § 1e) in the discrete-time case and

the *Black–Merton–Scholes model*

(or the standard diffusion model, based on geometric Brownian motion; see Chapter III, § 4b) in the continuous-time case.

As is well known, the first of them is (for a small time step $\Delta > 0$) a satisfactory approximation to the second, and pricing (of standard options, say) based on the first model yields results close to those obtained for the second model of a (B, S)-market, in which

$$B_t = B_0 e^{rt} \quad \text{and} \quad S_t = S_0 e^{(\mu - \frac{\sigma^2}{2})t + \sigma W_t}, \tag{3}$$

where $W = (W_t)_{t\geqslant 0}$ is a standard Wiener process.

In line with the *invariance principle* well known in the theory of limit theorems (see, e.g., [39] and [250]), a Wiener process can be a result of a limit transition in a great variety of random walk schemes. It is not surprising therefore that, for instance, in binomial models of (B^n, S^n)-markets (with discrete time step $\Delta = 1/n$) defined on some probability spaces $(\Omega^n, \mathscr{F}^n, \mathsf{P}^n)$ we have convergence to the (B, S)-model of Black–Merton–Scholes in the sense of (1) as $n \to \infty$, where P is a probability measure such that $W = (W_t)_{t \geqslant 0}$ is a Wiener process with respect to P.

Note that the existence of the convergence (1) for these two 'classical' models as well as for other models of financial markets is, as we see, only a part of the general *problem of the convergence* of (B^n, S^n)-markets to some 'limit' (B, S)-market. An equally important question is that of the convergence as $n \to \infty$ of the distributions

$$\operatorname{Law}(B^n, S^n \mid \widetilde{\mathsf{P}}^n) \to \operatorname{Law}(B, S \mid \widetilde{\mathsf{P}}), \tag{4}$$

where $\widetilde{\mathsf{P}}^n$ and $\widetilde{\mathsf{P}}$ are *martingale* (risk-neutral) measures for (B^n, S^n)- and (B, S)-models, respectively.

Here one must bear in mind the following aspects related to the *completeness* or *incompleteness* of the (arbitrage-free) markets in question.

If (B^n, S^n)-markets are complete, then (at any rate, if the *Second fundamental theorem* holds; see Chapter V, § 4a) each collection $\mathscr{P}(\mathsf{P}^n)$ of martingale measures contains a *unique* element, the question as to whether (4) holds under condition (1) is connected directly to the *contiguity* of the families of measures $(\widetilde{\mathsf{P}}^n)$ and (P^n), and it can be given a fairly complete answer in the framework of the *stochastic invariance principle*, which has been studied in detail, e.g., in [250; Chapter X, § 3].

However, the situation is more complicated if the arbitrage-free (B^n, S^n)-markets in question are *incomplete*.

In this case the sets of martingale measures $\mathscr{P}(\mathsf{P}^n)$ contain in general more than one element and there arises a tricky problem of the choice of a *chain* $(\widetilde{\mathsf{P}}^n)_{n \geqslant 1}$ of the corresponding martingale measures ensuring the convergence (4).

In accordance with the definition of weak convergence we can reformulate (4) as the limit relation

$$\mathsf{E}_{\widetilde{\mathsf{P}}^n} f(B^n, S^n) \to \mathsf{E}_{\widetilde{\mathsf{P}}} f(B, S) \tag{4'}$$

for continuous bounded functionals in the space of (càdlàg) trajectories of processes under consideration (which, in the above context, are usually assumed to be martingales).

Note that if $Z^n = \dfrac{d\widetilde{\mathsf{P}}^n}{d\mathsf{P}^n}$ and $Z = \dfrac{d\widetilde{\mathsf{P}}}{d\mathsf{P}}$, then

$$\mathsf{E}_{\widetilde{\mathsf{P}}^n} f(B^n, S^n) = \mathsf{E}_{\mathsf{P}^n} Z^n f(B^n, S^n)$$

and

$$\mathsf{E}_{\widetilde{\mathsf{P}}} f(B, S) = \mathsf{E}_{\mathsf{P}} Z f(B, S),$$

and therefore, clearly, (4′) is most closely connected with the convergence

$$\operatorname{Law}(B^n, S^n, Z^n \mid \mathsf{P}^n) \to \operatorname{Law}(B, S, Z \mid \mathsf{P}), \tag{5}$$

relating to the 'functional convergence' issues of the theory of limit theorems for stochastic process. See [39] and [250] for greater detail.

It is worth noting that the convergence (4′) follows from (5) if the family of random variables $\{Z^n f(B^n, S^n);\ n \geqslant 1\}$ is *uniformly integrable*, i.e.,

$$\lim_{N\to\infty} \varlimsup_{n} \mathsf{E}_{\mathsf{P}^n}\Big(|Z^n f(B^n, S^n)|\, I\big(|Z^n f(B^n, S^n)| > N\big)\Big) = 0.$$

3. As an example we consider the question whether properties (1) and (2) hold for the 'prelimit' models of Cox–Ross–Rubinstein and the 'limit' model of Black–Merton–Scholes, which are both arbitrage-free and complete.

Let $(\Omega^n, \mathscr{F}^n, \mathsf{P}^n)$ be a probability space, and assume that we have a binomial (B^n, S^n)-market with *piecewise-constant trajectories* (Chapter IV, § 2a) defined in this space in accordance with the 'simple return' pattern (Chapter II, § 1a): for $0 \leqslant t \leqslant 1$, $k = 1, \dots, n$, and $n \geqslant 1$ we have

$$B_t^n = B_0^n \prod_{k=1}^{[nt]} (1 + r_k^n) \tag{6}$$

and

$$S_t^n = S_0^n \prod_{k=1}^{[nt]} (1 + \rho_k^n), \tag{7}$$

where the *bank interest rates* are

$$r_k^n = \frac{r}{n}, \qquad r \geqslant 0, \tag{8}$$

and the *market stock returns* are

$$\rho_k^n = \frac{\mu}{n} + \xi_k^n, \qquad \mu \geqslant 0. \tag{9}$$

In the *homogeneous* Cox–Ross–Rubinstein models the variables $\xi_1^n, \dots, \xi_n^n$ are *independent* and *identically distributed*, with

$$\mathsf{P}^n\left(\xi_k^n = \frac{b}{\sqrt{n}}\right) = p \quad \text{and} \quad \mathsf{P}^n\left(\xi_k^n = -\frac{a}{\sqrt{n}}\right) = q, \tag{10}$$

where a, b, p, and q are some positive constants, $p + q = 1$.

By (9) and (10) we obtain

$$\mathsf{E}_{\mathsf{P}^n} \rho_k^n = \frac{\mu}{n} + \frac{1}{\sqrt{n}}(pb - qa), \tag{11}$$

$$\mathsf{E}_{\mathsf{P}^n} (\rho_k^n)^2 = \frac{1}{n}(pb^2 + qa^2) + O\left(\frac{1}{n^{3/2}}\right), \tag{12}$$

and

$$\mathsf{D}_{\mathsf{P}^n} \rho_k^n = \frac{pq}{n}(b + a)^2 + O\left(\frac{1}{n^{3/2}}\right). \tag{13}$$

For sufficiently large n we have $1 + \rho_k^n > 0$ and

$$\begin{aligned} S_t^n &= S_0^n \exp\left\{\sum_{k=1}^{[nt]} \ln(1 + \rho_k^n)\right\} \\ &= S_0^n \exp\left\{\sum_{k=1}^{[nt]} \left(\rho_k^n - \frac{(\rho_k^n)^2}{2}\right) + O\left(\frac{1}{\sqrt{n}}\right)\right\}. \end{aligned} \tag{14}$$

Assume that

$$pb - qa = 0. \tag{15}$$

Then, setting

$$\sigma^2 = pb^2 + qa^2 \tag{16}$$

and bearing in mind that $pq(b + a)^2 = \sigma^2$ and

$$\mathsf{E}_{\mathsf{P}^n} \rho_k^n = \frac{\mu}{n}, \quad \mathsf{E}_{\mathsf{P}^n} (\rho_k^n)^2 = \frac{\sigma^2}{n} + O\left(\frac{1}{n^{3/2}}\right), \quad \mathsf{D}_{\mathsf{P}^n} \rho_k^n = \frac{\sigma^2}{n}, \tag{17}$$

we obtain

$$\sum_{k=1}^{[nt]} \mathsf{E}_{\mathsf{P}^n} \left[\rho_k^n - \frac{(\rho_k^n)^2}{2}\right] \to \left(\mu - \frac{\sigma^2}{2}\right)t, \tag{18}$$

$$\sum_{k=1}^{[nt]} \mathsf{D}_{\mathsf{P}^n} \left[\rho_k^n - \frac{(\rho_k^n)^2}{2}\right] \to \sigma^2 t \tag{19}$$

as $n \to \infty$.

The following *Lindeberg condition* is clearly satisfied in the above case:

$$\lim_{n\to\infty} \sum_{k=1}^{n} \mathsf{E}_{\mathsf{P}^n} \Big[|\ln(1 + \rho_k^n)|^2 I(|\ln(1 + \rho_k^n)| > \varepsilon)\Big] = 0 \quad \text{for} \quad \varepsilon > 0. \tag{20}$$

Hence it follows by the *functional Central limit theorem* ([250; Chapter VII, Theorem 5.4]) that (as $S_0^n \to S_0$)

$$\mathrm{Law}\big(S_t^n;\ t \leqslant 1 \,|\, \mathsf{P}^n\big) \to \mathrm{Law}\big(S_t;\ t \leqslant 1 \,|\, \mathsf{P}\big), \tag{21}$$

where

$$S_t = S_0 \exp\left\{\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W_t\right\} \tag{22}$$

and $W = (W_t)_{t\leqslant 1}$ is a Wiener process (with respect to the measure P).

Thus, if $B_0^n \to B_0$, then

$$\mathrm{Law}\big(B_t^n,\, S_t^n;\ t \leqslant 1 \,|\, \mathsf{P}^n\big) \to \mathrm{Law}\big(B_t,\, S_t;\ t \leqslant 1 \,|\, \mathsf{P}\big), \tag{23}$$

i.e., we have convergence (1).

We turn now to an analogue of property (23) in the case when P^n and P are replaced by martingale measures $\widetilde{\mathsf{P}}^n$ and $\widetilde{\mathsf{P}}$.

If

$$\widetilde{\mathsf{P}}^n\left(\xi_k^n = \frac{b}{\sqrt{n}}\right) = \widetilde{p}^n, \qquad \widetilde{\mathsf{P}}^n\left(\xi_k^n = -\frac{a}{\sqrt{n}}\right) = \widetilde{q}^n$$

for $k = 1, \dots, n$, and the variables $\xi_1^n, \dots, \xi_n^n$ are independent with respect to $\widetilde{\mathsf{P}}^n$, then the martingale condition

$$\mathsf{E}_{\widetilde{\mathsf{P}}^n}\left(\frac{S_k^n}{B_k^n} \,\Big|\, \mathscr{F}_{k-1}^n\right) = \frac{S_{k-1}^n}{B_{k-1}^n}, \qquad 1 \leqslant k \leqslant n,$$

brings us to the relations

$$\mathsf{E}_{\widetilde{\mathsf{P}}^n}\, \rho_k^n = r_k^n, \qquad 1 \leqslant k \leqslant n,$$

which are already equivalent to the condition

$$b\widetilde{p}^n - a\widetilde{q}^n = \frac{r-\mu}{\sqrt{n}}\,. \tag{24}$$

Taking account of the equality $\widetilde{p}^n + \widetilde{q}^n = 1$, we find that

$$\begin{aligned} \widetilde{p}^n &= \frac{a}{a+b} + \frac{1}{\sqrt{n}}\frac{r-\mu}{a+b}\,, \\ \widetilde{q}^n &= \frac{b}{a+b} - \frac{1}{\sqrt{n}}\frac{r-\mu}{a+b}\,. \end{aligned} \tag{25}$$

(Cf. Example 2 in Chapter V, § 3f, where the *uniqueness* of the martingale measure $\widetilde{\mathsf{P}}^n$ and the *independence* of $\xi_1^n, \dots, \xi_k^n$ with respect to this measure were also established.)

Hence

$$\mathsf{E}_{\widetilde{\mathsf{P}}^n}\rho_k^n = \frac{r}{n},$$
$$\mathsf{E}_{\widetilde{\mathsf{P}}^n}(\rho_k^n)^2 = \frac{\widetilde{\sigma}^2}{n} + O\Big(\frac{1}{n^{3/2}}\Big),$$
$$\mathsf{D}_{\widetilde{\mathsf{P}}^n}\rho_k^n = \frac{\widetilde{\sigma}^2}{n} + O\Big(\frac{1}{n^{3/2}}\Big),$$

where $\widetilde{\sigma}^2 = ab$.

Note that the conditions $p + q = 1$ and $pb - qa = 0$ mean that $ab = pb^2 + qa^2$. In other words, $\sigma^2 = \widetilde{\sigma}^2$ and therefore

$$\sum_{k=1}^{[nt]} \mathsf{E}_{\widetilde{\mathsf{P}}^n}\left[\rho_k^n - \frac{(\rho_k^n)^2}{2}\right] \to \left(r - \frac{\sigma^2}{2}\right)t,$$
$$\sum_{k=1}^{[nt]} \mathsf{D}_{\widetilde{\mathsf{P}}^n}\left[\rho_k^n - \frac{(\rho_k^n)^2}{2}\right] \to \sigma^2 t.$$

The Lindeberg condition (20) withstands the replacement of the measure P^n by $\widetilde{\mathsf{P}}^n$. Hence using the functional Central limit theorem (as $S_0^n \to S_0$) we obtain

$$\operatorname{Law}(S_t^n;\ t \leqslant 1 \,|\, \widetilde{\mathsf{P}}^n) \to \operatorname{Law}(S_t;\ t \leqslant 1 \,|\, \widetilde{\mathsf{P}}), \tag{26}$$

where

$$S_t = S_0 \exp\left\{\left(r - \frac{\sigma^2}{2}\right)t + \sigma\widetilde{W}_t\right\},$$

$\widetilde{W} = (\widetilde{W}_t)_{t\leqslant 1}$ is a Wiener process with respect to $\widetilde{\mathsf{P}}$, which is a unique martingale measure existing by Girsanov's theorem (Chapter III, § 3e):

$$d\widetilde{\mathsf{P}} = \exp\left\{-\frac{\mu - r}{\sigma}W_1 - \frac{1}{2}\Big(\frac{\mu - r}{\sigma}\Big)^2\right\} d\mathsf{P}.$$

In addition, $\widetilde{W}_t = W_t + \dfrac{\mu - r}{\sigma}t$.

Thus, we have the following result.

THEOREM. *If the parameters $a > 0$, $b > 0$, $p > 0$, and $q > 0$ in the 'prelimit' Cox–Ross-Rubinstein models defined by (6)–(10) satisfy the conditions $pb - qa = 0$ and $p + q = 1$, then one has the convergence (21) and (26) to the 'limit' Black–Merton–Scholes models.*

4. We now modify slightly the 'prelimit' Cox–Ross–Rubinstein models, dropping the restrictive condition $pb - qa = 0$ so as to retain the right to use the functional limit theorem.

To this end we assume that the variables ρ_k^n are still defined by formulas (9) for odd $k = 1, 3, \dots$, whereas for even $k = 2, 4, \dots$ we have

$$\rho_k^n = \frac{\mu}{n} + \eta_k^n, \qquad \mu > 0,$$

where

$$\mathsf{P}^n\left(\eta_k^n = -\frac{b}{\sqrt{n}}\right) = p, \qquad \mathsf{P}^n\left(\eta_k^n = \frac{a}{\sqrt{n}}\right) = q.$$

Then for $k = 2, 4, \dots$ we obtain

$$\begin{aligned}
\mathsf{E}_{\mathsf{P}^n} \rho_k^n &= \frac{\mu}{n} - \frac{1}{\sqrt{n}}(pb - aq), \\
\mathsf{E}_{\mathsf{P}^n} (\rho_k^n)^2 &= \frac{1}{n}(pb^2 + qa^2) + O\left(\frac{1}{n^{3/2}}\right), \\
\mathsf{D}_{\mathsf{P}^n} \rho_k^n &= \frac{pq}{n}(a+b)^2 + O\left(\frac{1}{n^{3/2}}\right).
\end{aligned}$$

Hence, in view of (11)–(13) (for odd k),

$$\begin{aligned}
&\sum_{k=1}^{[nt]} \mathsf{E}_{\mathsf{P}^n}\left[\rho_k^n - \frac{(\rho_k^n)^2}{2}\right] \to \left(\mu - \frac{\sigma^2}{2}\right)t, \\
&\sum_{k=1}^{[nt]} \mathsf{D}_{\mathsf{P}^n}\left[\rho_k^n - \frac{(\rho_k^n)^2}{2}\right] \to \sigma^2 t,
\end{aligned}$$

where

$$\sigma^2 = pq(a+b)^2.$$

Thus, in our case of the *inhomogeneous* Cox–Ross–Rubinstein model we obtain the same result as in the *homogeneous* case; namely,

$$\mathrm{Law}\big(S_t^n;\ t \leqslant 1 \,|\, \mathsf{P}^n\big) \to \mathrm{Law}\big(S_t;\ t \leqslant 1 \,|\, \mathsf{P}\big), \tag{27}$$

where $S = (S_t)_{t \leqslant 1}$ is a process defined by (22) with $\sigma^2 = pq(a+b)^2$.

Now let $\widehat{\mathsf{P}}^n$ be martingale measures such that the variables $\rho_1^n, \dots, \rho_n^n$ are independent again and

$$\widehat{\mathsf{P}}^n\left(\xi_k^n = \frac{b}{\sqrt{n}}\right) = \widehat{p}_k^n, \qquad \widehat{\mathsf{P}}^n\left(\xi_k^n = -\frac{a}{\sqrt{n}}\right) = \widehat{q}_k^n$$

for *odd* k, where $\widehat{p}_k^n = \widetilde{p}^n$ and $\widehat{q}_k^n = \widetilde{q}^n$ and

$$\widehat{\mathsf{P}}^n\left(\eta_k^n = -\frac{b}{\sqrt{n}}\right) = \widehat{p}_k^n, \qquad \widehat{\mathsf{P}}^n\left(\eta_k^n = \frac{a}{\sqrt{n}}\right) = \widehat{q}_k^n$$

for *even* k with $\widehat{p}_k^n$ and $\widehat{q}_k^n$ defined (due to the martingale condition) by the formulas $\widehat{p}_k^n = \widehat{p}^n$ and $\widehat{q}_k^n = \widehat{q}^n$ with

$$\widehat{p}^n = \frac{a}{a+b} - \frac{1}{\sqrt{n}}\frac{r-\mu}{a+b},$$
$$\widehat{q}^n = \frac{b}{a+b} + \frac{1}{\sqrt{n}}\frac{r-\mu}{a+b}.$$

Then for *all* $k = 1, \dots, n$ we have

$$\mathsf{E}_{\widehat{\mathsf{P}}^n}\rho_k^n = \frac{r}{n},$$
$$\mathsf{E}_{\widehat{\mathsf{P}}^n}(\rho_k^n)^2 = \frac{ab}{n} + O\left(\frac{1}{n^{3/2}}\right),$$

and

$$\mathsf{D}_{\widehat{\mathsf{P}}^n}\rho_k^n = \frac{ab}{n} + O\left(\frac{1}{n^{3/2}}\right).$$

Consequently,

$$\sum_{k=1}^{[nt]} \mathsf{E}_{\widehat{\mathsf{P}}^n}\left[\rho_k^n - \frac{(\rho_k^n)^2}{2}\right] \to \left(r - \frac{\widehat{\sigma}^2}{2}\right)t,$$
$$\sum_{k=1}^{[nt]} \mathsf{D}_{\widehat{\mathsf{P}}^n}\left[\rho_k^n - \frac{(\rho_k^n)^2}{2}\right] \to \widehat{\sigma}^2 t,$$

where $\widehat{\sigma}^2 = ab$, and in view of the Lindeberg condition (which is still satisfied) we see that (as $S_0^n \to S_0$)

$$\operatorname{Law}\big(S_t^n;\ t \leqslant 1 \,|\, \widehat{\mathsf{P}}^n\big) \to \operatorname{Law}\big(S_t;\ t \leqslant 1 \,|\, \widehat{\mathsf{P}}\big), \tag{28}$$

where

$$S_t = S_0 \exp\left\{\left(r - \frac{\widehat{\sigma}^2}{2}\right)t + \widehat{\sigma}\widehat{W}_t\right\}$$

and $\widehat{W} = (\widehat{W}_t)_{t\leqslant 1}$ is a Wiener process with respect to the measure $\widehat{\mathsf{P}}$ such that

$$d\widehat{\mathsf{P}} = \exp\left\{-\frac{\mu - r}{\widehat{\sigma}}W_1 - \frac{1}{2}\left(\frac{\mu - r}{\widehat{\sigma}}\right)^2\right\}d\mathsf{P}.$$

We point out that, in general, $ab \neq pq(a+b)^2$, and therefore $\widehat{\sigma}^2 \neq \sigma^2$. Hence if the 'limiting' model has volatility σ^2 and the parameters $a > 0$, $b > 0$, $p > 0$, and $q > 0$ satisfy the equalities $p + q = 1$ and $\sigma^2 = pq(a+b)^2$, then we have the functional convergence (27); however the 'limit' volatility $\widehat{\sigma}^2$ in (28) can happen to be *distinct* from σ^2 (if we drop the 'restrictive' condition $pb - aq = 0$).

This example of an *inhomogeneous* Cox–Ross–Rubinstein model shows that in choosing such models as *approximations* to the Black–Merton–Scholes model with parameters (μ, σ^2) one must be careful in the selection of the parameters (p, q, a, b) of the 'prelimit' models, since even if there is convergence (27) with respect to the original probability measure, the corresponding convergence (to the Black–Merton–Scholes models with parameters (μ, σ^2)) with respect to martingale measures may well fail. This, in turn, means that the rational (hedging) prices $\mathbb{C}_1^n$ in the 'prelimit' models *do not necessarily converge* to the (anticipated) price $\mathbb{C}_1$ in the 'limiting' model.

Clearly, a similar situation can arise in the framework of other approximation schemes.

5. In connection with the above cases of the convergence of the processes $S^n = (S_t^n)_{t \leqslant 1}$ to $S = (S_t)_{t \leqslant 1}$ – both with respect to the original probability measures (P^n and P) and with respect to the martingale measures ($\widetilde{\mathsf{P}}^n$ and $\widetilde{\mathsf{P}}$)—it seems appropriate to present several general results in this direction.

Assuming for simplicity that $B_t^n \equiv 1$ and $B_t \equiv 1$ for $t \leqslant 1$ and $n \geqslant 1$, we observe first of all that the weak convergence of the laws $\mathrm{Law}(S^n \mid \mathsf{P}^n)$ does not imply the weak convergence of $\mathrm{Law}(S^n \mid \widetilde{\mathsf{P}}^n)$ even in the presence of the contiguity $(\widetilde{\mathsf{P}}^n) \lhd (\mathsf{P}^n)$, although the sequence $(\mathrm{Law}(S^n \mid \widetilde{\mathsf{P}}^n))_{n \geqslant 1}$ is nevertheless *tight* (see the monograph [250; Chapter X, § 3]), and therefore, in general, can have *several* limit measures (corresponding to different subsequences).

One standard trick ensuring the *uniqueness* of the limit probability measure is that, besides the condition of the weak convergence of the laws $\mathrm{Law}(S^n \mid \mathsf{P}^n)$ and the contiguity $(\widetilde{\mathsf{P}}^n) \lhd (\mathsf{P}^n)$, we assume the weak convergence of the *joint* distributions $\mathrm{Law}(S^n, Z^n \mid \mathsf{P}^n)$, where $Z^n = (Z_t^n)_{t \leqslant 1}$ and the $Z_t^n = \dfrac{d\widetilde{\mathsf{P}}_t^n}{d\mathsf{P}_t^n}$ are the densities of the $\widetilde{\mathsf{P}}_t^n$ with respect to the P_t^n (here $\widetilde{\mathsf{P}}_t^n$ and P_t^n are the restrictions of $\widetilde{\mathsf{P}}^n$ and P^n to the σ-algebra $\mathscr{F}_t^n$ from the stochastic basis $(\Omega^n, \mathscr{F}^n, (\mathscr{F}_t^n)_{t \leqslant 1}, \mathsf{P}^n)$ underlying the processes $S^n = (S_t^n)_{t \leqslant 1}$).

Then it follows from the generalized version of so-called *LeCam's third lemma* (see [250; Chapter X, Theorem 3.3]) that the *sequence of laws* $\mathrm{Law}(S^n, Z^n \mid \widetilde{\mathsf{P}}^n)$ *converges weakly to some probability measure that is absolutely continuous with respect to the measure that is the weak limit of* $\mathrm{Law}(S^n, Z^n \mid \mathsf{P}^n)$, $n \geqslant 1$. *If, in addition,* $\mathrm{Law}(S^n, Z^n \mid \mathsf{P}^n) \to \mathrm{Law}(S, Z \mid \mathsf{P})$, *then* $\mathrm{Law}(S^n, Z^n \mid \widetilde{\mathsf{P}}^n) \to \mathrm{Law}(S, Z \mid \widetilde{\mathsf{P}})$, *where* $d\widetilde{\mathsf{P}} = Z\, d\mathsf{P}$.

6. For illustrations of these results we turn again to the Cox–Ross–Rubinstein model (6)–(7) considered above, set for simplicity $r = 0$ and $B_0^n = 1$, and put this model in the form used already in the construction of the *minimal* martingale measure in Chapter V, § 3d.7 (see also [392]).

Let $H_k^n = \sum_{l=1}^{k} \rho_l^n$. If $pb - qa = 0$, then the sequence $M^n = (M_k^n)_{k \leqslant n}$, where $M_k^n = \sum_{l=1}^{k} \xi_l^n$, is a square integrable martingale with quadratic characteristic $\langle M^n \rangle = (\langle M^n \rangle_k)_{k \leqslant n}$, where $\langle M^n \rangle_k = \sum_{l=1}^{k} \mathsf{D}\xi_l^n = \sigma^2 k$, $\sigma^2 = pb^2 + qa^2$ $(= ab)$.

Setting $a_k^n = \mu/\sigma^2$ we can write the Doob decomposition (Chapter II, § 1b) of $H^n = (H_k^n)_{k \leqslant n}$ in the following form (in terms of the increments):

$$\Delta H_k^n = a_k^n \, \Delta \langle M^n \rangle_k + \Delta M_k^n. \tag{29}$$

Assume that $b\mu < \sigma^2$. Then the (minimal) measure $\widetilde{\mathsf{P}}^n$ such that

$$d\widetilde{\mathsf{P}}^n = \prod_{k=1}^{n} (1 - a_k^n \, \Delta M_k^n) \, d\mathsf{P}^n \tag{30}$$

(cf. formula (31) in Chapter V, § 3d) is a *probability* measure and, moreover, a (unique) *martingale* measure for the sequence $H^n = (H_k^n)_{k \leqslant n}$, as follows from Chapter V, § 3d.7 and can be verified directly.

We set

$$Z_t^n = \prod_{k=1}^{[nt]} (1 - a_k^n \, \Delta M_k^n) = \mathscr{E}\Big(- \sum_{k \leqslant \cdot} a_k^n \, \Delta M_k^n \Big)_{[nt]} \tag{31}$$

and represent S_t^n as follows:

$$\begin{aligned} S_t^n &= S_0^n \prod_{k=1}^{[nt]} (1 + \Delta H_k^n) = S_0^n \mathscr{E}(H^n)_{[nt]} \\ &= S_0^n \mathscr{E}\Big(\sum_{k \leqslant} a_k^n \, \Delta \langle M^n \rangle_k + M^n \Big)_{[nt]}. \end{aligned} \tag{32}$$

Let also $M = (M_t)_{t \leqslant 1}$ be a square integrable martingale (on some stochastic basis $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t \leqslant 1}, \mathsf{P})$) with quadratic characteristic $\langle M \rangle = (\langle M \rangle_t)_{t \leqslant 1}$, let $a = (a_t)_{t \leqslant 1}$ be a predictable process with $a^2 \cdot \langle M \rangle_1 < \infty$ $\Big($i.e., with $\int_0^1 a_t^2 \, d\langle M \rangle_t < \infty\Big)$, let

$$H_t = \int_0^t a_s \, d\langle M \rangle_s + M_t \tag{33}$$

and let

$$Z_t = \mathscr{E}\biggl(-\int_0^{\cdot} a_s\,dM_s\biggr)_t, \qquad S_t = S_0\mathscr{E}(H)_t. \tag{34}$$

The structure of relations (31)–(34) suggests the conditions that one must impose on these processes to obtain the weak convergence

$$\mathrm{Law}\bigl(S_t^n,\, Z_t^n;\ t \leqslant 1 \,|\, \mathsf{P}^n\bigr) \to \mathrm{Law}\bigl(S_t,\, Z_t;\ t \leqslant 1 \,|\, \mathsf{P}\bigr). \tag{35}$$

Thus, if $a^n = (a_t^n)_{t\leqslant 1}$, $M^n = (M_t^n)_{t\leqslant 1}$ and $\langle M^n\rangle = (\langle M^n\rangle_t)_{t\leqslant 1}$ are piecewise-constant processes constructed from $(a_k^n)_{k\leqslant n}$, $(M_k^n)_{k\leqslant n}$, and $(\langle M^n\rangle_k)_{k\leqslant n}$, then it is clearly sufficient for (35) that the distributions

$$\mathrm{Law}\biggl(M_t^n,\ \sum_{k=1}^{[nt]} a_k^n\,\Delta M_k^n,\ \sum_{k=1}^{[nt]} a_k^n\,\Delta\langle M^n\rangle_k;\ t \leqslant 1 \biggm| \mathsf{P}^n\biggr)$$

converge to

$$\mathrm{Law}\biggl(M_t,\ \int_0^t a_s\,dM_s,\ \int_0^t a_s\,d\langle M\rangle_s;\ t \leqslant 1 \biggm| \mathsf{P}\biggr)$$

and $S_0^n \to S_0$.

Various conditions ensuring such convergence of *martingales* and *stochastic integrals* can be found, e.g., in [250; Chapter IX] and [254]. In particular, it follows from theorems 2.6 and 2.11 in [254] that the required convergence occurs once

$$\mathrm{Law}\bigl(M_t^n,\, a_t^n;\ t \leqslant 1 \,|\, \mathsf{P}^n\bigr) \to \mathrm{Law}\bigl(M_t,\, a_t;\ t \leqslant 1 \,|\, \mathsf{P}\bigr)$$

and the jumps of martingales satisfy the following condition of *uniform smallness*:

$$\sup_n \mathsf{E}_{\mathsf{P}^n}\Bigl[\sup_{t\leqslant 1} |\Delta M_t^n|\Bigr] < \infty.$$

This clearly holds for the model (6)–(7) and (as seen from subsection 3) we can take as the limit $M = (M_t)_{t\leqslant 1}$ the process with $M_t = \sigma W_t$ and $\sigma^2 = pb^2 + qa^2$, where $W = (W_t)_{t\leqslant 1}$ is a standard Wiener process. Then $H_t = \mu t + \sigma W_t$ and $S_t = S_0\mathscr{E}(H)_t$. Since $d\mathscr{E}(H)_t = \mathscr{E}(H)_t\,dH_t$, it follows that $dS_t = S_t(\mu\,dt + \sigma\,dW_t)$. Hence, as one would expect, the process $S = (S_t)_{t\leqslant 1}$ is just a geometric Brownian motion:

$$S_t = S_0 \exp\biggl\{\Bigl(\mu - \frac{\sigma^2}{2}\Bigr)t + \sigma W_t\biggr\}.$$

We also see from (34) that

$$Z_t = \exp\biggl\{-\frac{\mu}{\sigma}W_t - \frac{1}{2}\Bigl(\frac{\mu}{\sigma}\Bigr)^2 t\biggr\}.$$

In the present case we have the contiguity $(\widetilde{\mathsf{P}}^n) \lhd (\mathsf{P}^n)$, which can be proved as in Example 1 in § 3c.5. Hence it follows from the above-cited generalized version of LeCam's third lemma that

$$\mathrm{Law}\big(S_t^n;\ t \leqslant 1 \,|\, \widetilde{\mathsf{P}}^n\big) \to \mathrm{Law}\big(S_t;\ t \leqslant 1 \,|\, \widetilde{\mathsf{P}}\big),$$

where $d\widetilde{\mathsf{P}} = Z_1 d\mathsf{P}$.

Since, by Girsanov's theorem (Chapter III, § 3e), the process $\widetilde{W} = (\widetilde{W}_t)_{t\leqslant 1}$ with $\widetilde{W}_t = W_t + \dfrac{\mu}{\sigma} t$ is Wiener with respect to the measure $\widetilde{\mathsf{P}}$, it follows that

$$\begin{aligned}\mathrm{Law}\big(\mu t + \sigma W_t;\ t \leqslant 1 \,|\, \widetilde{\mathsf{P}}\big) &= \mathrm{Law}\big(\sigma \widetilde{W}_t;\ t \leqslant 1 \,|\, \widetilde{\mathsf{P}}\big) \\ &= \mathrm{Law}\big(\sigma W_t;\ t \leqslant 1 \,|\, \mathsf{P}\big).\end{aligned}$$

Hence

$$\mathrm{Law}\big(S_t;\ t \leqslant 1 \,|\, \widetilde{\mathsf{P}}\big) = \mathrm{Law}\big(S_0 e^{-\frac{\sigma^2}{2} t + \sigma W_t};\ t \leqslant 1 \,|\, \mathsf{P}\big),$$

which has been already proved in subsection 3 by a direct application of the functional Central limit theorem (see (26)).

4. European Options on a Binomial (B, S)-Market

§ 4a. Problems of Option Pricing

1. Following a long-established tradition in finance and in accordance with the nomenclature in Chapter I, § 1a we distinguish two kinds of financial instruments and, in particular, securities,

basic (primary)

and

derivative (secondary).

We discussed basic securities (stock, bonds, currency) at length in the first chapters. We considered various *models* of their dynamics, and the results of *statistical analysis* revealing such phenomena in the behavior of financial data as the *cluster property*, *fractality*, *long memory*, and some other.

We have also paid much attention to the *theory* underlying derivatives pricing, which is based upon the concept of arbitrage-free, 'fair' financial market.

As pointed out in Chapter I, not only speculators may develop an interest in *derivatives*. Importantly, they play the role of *hedging* instruments, which *protect* one from financial risks incurred by uncertainties of the price development.

For instance, if the current price of a share in corporation A is $S_0 = 100$ and the investor anticipates its growth ($S_1 = 120$), then he can buy a share (at time $n = 0$) and then (at time $n = 1$) sell it pocketing the profit of $S_1 - S_0 = 120 - 100 = 20$.

Of course, the price can also drop ($S_0 = 100 \downarrow S_1 = 80$), and then selling the share will bring losses: $S_1 - S_0 = 80 - 100 = -20$.

Thus, we have two possible patterns of development:

$$
S_0 = 100 \quad \begin{array}{l} \nearrow \; S_1 = 120 \\ \searrow \; S_1 = 80 \end{array}
$$

and 'large' gains (= 20, i.e., 20% of the price $S_0 = 100$) are accompanied by a 'large' risk of losses ($= -20$, i.e., 20% of $S_0 = 100$).

Besides the above strategy (of buying and selling *basic* securities), an investor can also look at the *derivatives* market. For example, he can buy a call option with maturity time $n = 1$, which gives him a right (see Chapter I, § 1c) to buy a share at time $n = 1$ at a price of $K = 100$, say.

In this case, if $S_0 = 100$ and $S_1 = 120$, then the investor buys the share at the price $K = 100$ (fixed in advance) and immediately sells it (on the so-called *spot* market[a]) at the market price $S_1 = 120$, pocketing the profit of $(S_1 - K)^+ = 20$.

Of course, for buying this option that grants the right of a purchase at a fixed price ($K = 100$) one pays certain premium to the writer. Assume that this premium is $\mathbb{C}_1 = 10$. If the stock moves up ($S_0 = 100 \uparrow S_1 = 120$), then the buyer's *net profit* is 10. On the other hand, if the prices drop ($S_0 = 100 \downarrow S_1 = 80$), then the buyer does not exercise his option (there is no sense in the purchase of a share at the price $K = 100$ when one can buy it at the lower market price $S_1 = 80$), and takes *losses* in the amount of the premium (= 10) paid for the option.

Hence, the purchase of a call option (which is just one kind of derivatives) *reduces the risks* of an investor (he can now lose only 10 in place of 20 'units'), but, of course, his potential profits have also slimmed (10 'units' in place of 20).

Thus, we can say that the strategy of a speculator anticipating a rise of price (of a 'bull', using the term of Chapter I, § 1c) is better insured against possible losses if it is based on the purchase of a call option than when it is based on transactions involving stock directly.

We consider now a 'bear', a speculator who anticipates a *drop* of prices (the price of a share, in our case). In principle, it is not against the rules on many markets to sell stock one does not have at the moment. Assume that our 'bear' undertakes to sell stock at time $n = 1$ and the corresponding exercise price is 100 'units'. If the price S_1 drops to 80, in line with the 'bear's forecasts, then he will buy a share at this (market) price $S_1 = 80$ and take a profit of 20 'units'. However, if $S_1 = 120$, then his losses will be 20 'units'. Again, both huge profits and huge losses are possible.

Similarly to the 'bull', the 'bear' can turn to the *derivatives* market. For instance, he can buy a put option (for 10 'units', with $K = 100$), which gives him the right to sell the share (which he does not necessarily have at the moment) at the price $K = 100$. Later on, if $S_1 = 80$, then he buys a share at this market price and sells it for $K = 100$ 'units', as stays written in the contract. Allowing for the premium, his net profit is $20 - 10 = 10$ 'units' if prices actually drop ($S_0 = 100 \downarrow S_1 = 80$). On the other hand, if they rise, then the 'bear' takes losses

[a]The following terms are often used in the finances literature (see, e.g., [50]): deals (contracts) providing for a delivery or certain actions at some moment in the future (options, futures, forwards, etc.) are usually said to be *forward*, while the ones including instant delivery are said to be *spot*.

of 10 'units'. Hence, the purchase of a put option reduces *speculative risks*, but reduces accordingly also the possible *speculative gains*.

The above figures are fairly arbitrary; nevertheless, our illustration of the 'speculative' and 'protective' functions of derivatives is adequate on the whole.

2. One cardinal issue relating to options is the value of the *fair*, *rational* premium paid for the purchase of an option contract. This is important for the buyer and also for the writer, who sells the derivative and must use the premium to *provide* for his ability to meet the terms of the contract. Of course, the writer of options is also interested in the assessment of the overall profits or losses from their flotation on the market.

Note that a pricing theory for some or other *derivatives* must be built upon concrete models describing *basic* derivatives and general assumptions about the structure and the mechanisms of securities markets. The simplest in this respect is the (B, S)-market described by the binomial *CRR*-model of Cox–Ross–Rubinstein (see Chapter II, § 1e). For all its simplicity, in its analysis one can more easily understand general principles and find examples of pricing based on the concept of 'absence of arbitrage'. Our discussion will be built around options, both for their own sake and because many problems related to the markets of other derivatives can either be reformulated in the language of options or benefit by the well-developed techniques of option pricing, which is based on the plain and fruitful idea of hedging.

§ 4b. Rational Pricing and Hedging Strategies. General Pay-Off Functions

1. In the *CRR*-model of a (B, S)-market formed by two assets, a bank account $B = (B_n)$ and a stock $S = (S_n)$, one assumes that

$$\begin{aligned} \Delta B_n &= rB_{n-1}, \\ \Delta S_n &= \rho_n S_{n-1}, \end{aligned} \tag{1}$$

where (ρ_n) is a sequence of *independent random variables* taking two values, a and b, $a < b$, and r is the interest rate, $-1 < a < r < b$.

Moreover, we assume that the sequence $\rho = (\rho_n)$ of variables defined on the underlying filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_n), \mathsf{P})$ has the property

$$\mathsf{P}(\rho_n = b) = p \quad \text{and} \quad \mathsf{P}(\rho_n = a) = q,$$

where $p + q = 1$, $0 < p, q < 1$, and the ρ_n are $\mathscr{F}_n$-measurable for each n.

All the randomness in this model is due to the variables ρ_n, therefore, we can take as the space Ω of elementary outcomes either the space $\Omega_N = \{a, b\}^N$ of finite sequences $x = (x_1, x_2, \dots, x_N)$ such that $x_n = a$ or $x_n = b$ (if $n \leqslant N$) or the space $\Omega_\infty = \{a, b\}^\infty$ of infinite sequences $x = (x_1, x_2, \dots)$ with $x_n = a, b$ (if $n \in \{1, 2, \dots\}$). Then $\rho_n(x) = x_n$ and since both spaces Ω_N and Ω_∞ are discrete, probability measures P_N and P on the corresponding systems of Borel sets are completely defined by their finite-dimensional distributions $\mathsf{P}_n = \mathsf{P}_n(x_1,\dots,x_n)$, where $n \leqslant N$ or $n < \infty$.

If $\nu_b(x_1,\dots,x_n) = \sum_{i=1}^{n} I_b(x_i)$ is the number of the components x_i equal to b for $i \leqslant n$, then, obviously,

$$\mathsf{P}_n(x_1,\dots,x_n) = p^{\nu_b(x_1,\dots,x_n)} q^{n-\nu_b(x_1,\dots,x_n)}. \tag{2}$$

Putting this another way we can say that P_n is equal to $\underbrace{\mathsf{Q}\otimes\cdots\otimes\mathsf{Q}}_{n\,\text{times}}$, the direct product of the measures Q such that $\mathsf{Q}(\{b\}) = p$ and $\mathsf{Q}(\{a\}) = q$, where $p > 0$, $q > 0$, and $p+q = 1$. As shown in Chapter V, § 1d the *CRR*-model is *arbitrage-free* and *complete*, while, by the First and the Second fundamental theorems, for each $n \geqslant 1$ there *exists* a *unique* martingale measure $\widetilde{\mathsf{P}}_n \sim \mathsf{P}_n$, which has the following simple structure (cf. (2)):

$$\widetilde{\mathsf{P}}_n(x_1,\dots,x_n) = \widetilde{p}^{\,\nu_b(x_1,\dots,x_n)} \widetilde{q}^{\,n-\nu_b(x_1,\dots,x_n)}, \tag{3}$$

where

$$\widetilde{p} = \frac{r-a}{b-a}, \qquad \widetilde{q} = \frac{b-r}{b-a}. \tag{4}$$

We see from (3) that $\widetilde{\mathsf{P}}_n$, like P_n, has the structure of a *direct product*: $\widetilde{\mathsf{P}}_n = \underbrace{\widetilde{\mathsf{Q}}\otimes\cdots\otimes\widetilde{\mathsf{Q}}}_{n\,\text{times}}$, where $\widetilde{\mathsf{Q}}(\{b\}) = \widetilde{p}$ and $\widetilde{\mathsf{Q}}(\{a\}) = \widetilde{q}$.

2. We shall now consider *European options* of maturity $N < \infty$ with pay-off (contingent claim) f_N depending, in general, on all the variables $S_0, S_1, \dots, S_N$, or, equivalently, on S_0 and $\rho_1,\dots,\rho_N$. (See Chapter I, § 1c for various concepts related to options.)

We have already mentioned that both writer (issuer) and buyer of an option contract come across the central problem of the correct definition of the 'fair' ('rational') price of this contract.

In accordance with Chapter V, § 1b, if a market is *complete* and *arbitrage-free* (as is our binomial (B,S)-market), then as a *fair* price, one can reasonably regard the following quantity (the price of perfect European hedging):

$$\mathbb{C}(f_N;\mathsf{P}) = \inf\bigl\{x \geqslant 0 \colon \exists \pi \text{ such that } X_0^\pi = x \text{ and } X_N^\pi = f_N \ (\mathsf{P}\text{-a.s.})\bigr\}, \tag{5}$$

where $X^\pi = (X_n^\pi)_{0\leqslant n\leqslant N}$ is the value of the self-financing strategy $\pi = (\beta,\gamma)$. (See Chapter V, § 1b and § 1b in the present chapter for greater detail.)

Moreover, one can calculate $\mathbb{C}(f_N;\mathsf{P})$ by formula (4) in § 1b; namely,

$$\mathbb{C}(f_N;\mathsf{P}) = B_0 \widetilde{\mathsf{E}} \frac{f_N}{B_N}, \tag{6}$$

where $\widetilde{\mathsf{E}}$ is averaging with respect to the martingale measure $\widetilde{\mathsf{P}}_N$.

For the model (1) we have $B_N = B_0(1+r)^N$. Hence we obtain there

$$\mathbb{C}(f_N;\mathsf{P}) = \widetilde{\mathsf{E}}\frac{f_N}{(1+r)^N}, \tag{7}$$

and, in principle, this gives one a complete answer to the question on the rational price of an option contract with pay-off f_N.

Remarkably, the option writer in this model, on taking the premium $\mathbb{C}(f_N;\mathsf{P})$ from the buyer, can build a portfolio $\widetilde{\pi} = (\widetilde{\beta},\widetilde{\gamma})$ of value $X^{\widetilde{\pi}} = (X_n^{\widetilde{\pi}})_{n\leqslant N}$ *replicating* faithfully the pay-off f_N at the instant N. As mentioned, e.g., in § 1b, one standard way of finding this portfolio $\widetilde{\pi} = (\widetilde{\beta},\widetilde{\gamma})$ is as follows.

We consider the martingale $M = (M_n, \mathscr{F}_n, \widetilde{\mathsf{P}}_N)_{n\leqslant N}$, where

$$M_n = \widetilde{\mathsf{E}}\biggl(\frac{f_N}{B_N}\biggm| \mathscr{F}_n\biggr).$$

In view of the '$\dfrac{S}{B}$*-representation*', there exists a predictable sequence $\widetilde{\gamma} = (\widetilde{\gamma}_i)_{i\leqslant N}$ such that

$$M_n = M_0 + \sum_{k=1}^{n} \widetilde{\gamma}_k \Delta\biggl(\frac{S_k}{B_k}\biggr), \qquad n \leqslant N. \tag{8}$$

Setting $\widetilde{\beta}_k = M_k - \dfrac{\widetilde{\gamma}_k S_k}{B_k}$ we obtain (see Chapter V, § 4b and § 1b in the present chapter) a self-financing hedge $\widetilde{\pi} = (\widetilde{\beta},\widetilde{\gamma})$ of value

$$X_k^{\widetilde{\pi}} = \widetilde{\beta}_k B_k + \widetilde{\gamma}_k S_k = B_k \widetilde{\mathsf{E}}\biggl(\frac{f_N}{B_N}\biggm| \mathscr{F}_k\biggr)$$

such that

$$X_0^{\widetilde{\pi}} = \mathbb{C}(f_N;\mathsf{P}) \tag{9}$$

and we have the property of *perfect* hedging:

$$X_N^{\widetilde{\pi}} = f_N.$$

Since

$$\Delta\biggl(\frac{S_k}{B_k}\biggr) = \frac{S_{k-1}(\rho_k - r)}{B_k}, \tag{10}$$

it follows by (8) that

$$M_n = M_0 + \sum_{k=1}^{n} \widetilde{\alpha}_k^{(\rho)}(\rho_k - r) = M_0 + \sum_{k=1}^{n} \widetilde{\alpha}_k^{(\rho)} \Delta m_k^{(\rho)}, \tag{11}$$

where $\widetilde{\alpha}_k^{(\rho)}$ and $\widetilde{\gamma}_k$ are connected by the equality

$$\widetilde{\gamma}_k = \frac{\widetilde{\alpha}_k^{(\rho)} B_k}{S_{k-1}}, \tag{12}$$

and the sequence $m^{(\rho)} = (m_n^{(\rho)}, \mathscr{F}_n, \widetilde{\mathsf{P}}_N)_{n \leqslant N}$ of the variables

$$m_n^{(\rho)} = \sum_{k=1}^{n} (\rho_k - r) \tag{13}$$

is a martingale.

As will be clear from what follows, it makes sense to consider alongside the sequence $\rho = (\rho_n)$ also the sequence $\delta = (\delta_n)$ of the variables

$$\delta_n = \frac{\rho_n - a}{b - a}. \tag{14}$$

Clearly,

$$\rho_n = \begin{cases} b \\ a \end{cases} \iff \delta_n = \begin{cases} 1 \\ 0 \end{cases} \tag{15}$$

and $\mathscr{F}_n = \sigma(\rho_1, \ldots, \rho_n) = \sigma(\delta_1, \ldots, \delta_n)$.

Since

$$\delta_k - \widetilde{p} = \frac{\rho_k - r}{b - a}, \tag{16}$$

it follows that, in addition to (8) and (11), we have also the representation

$$M_n = M_0 + \sum_{k=1}^{n} \widetilde{\alpha}_k^{(\delta)} m_k^{(\delta)}, \tag{17}$$

where the sequence $m^{(\delta)} = (m_k^{(\delta)}, \mathscr{F}_k, \widetilde{\mathsf{P}}_N)$ of variables

$$m_l^{(\delta)} = \sum_{l=1}^{n} (\delta_n - \widetilde{p}) \tag{18}$$

is a martingale and

$$\widetilde{\alpha}_k^{(\delta)} = (b - a)\widetilde{\alpha}_k^{(\rho)}. \tag{19}$$

We sum up the above results as follows.

THEOREM 1. 1) *In the framework of the CRR-model* (1), *for each N and each $\mathscr{F}_N$-measurable pay-off f_N the fair price $\mathbb{C}(f_N; \mathsf{P})$ can be described by the formula*

$$\mathbb{C}(f_N; \mathsf{P}) = \widetilde{\mathsf{E}} \frac{f_N}{(1+r)^N}, \tag{20}$$

where $\widetilde{\mathsf{E}}$ is averaging with respect to the martingale measure $\widetilde{\mathsf{P}}_N$.

2) *There exists a perfect self-financing hedge $\widetilde{\pi} = (\widetilde{\beta}, \widetilde{\gamma})$ of value $X^{\widetilde{\pi}} = (X_n^{\widetilde{\pi}})_{n \leqslant N}$ such that*

$$X_0^{\widetilde{\pi}} = \mathbb{C}(f_N; \mathsf{P}), \qquad X_N^{\widetilde{\pi}} = f_N$$

and

$$X_n^{\widetilde{\pi}} = \widetilde{\mathsf{E}}\left(\frac{f_N}{(1+r)^N} \,\Big|\, \mathscr{F}_n \right). \tag{21}$$

3) *The components $\widetilde{\beta} = (\widetilde{\beta}_n)_{n \leqslant N}$ and $\widetilde{\gamma} = (\widetilde{\gamma}_n)_{n \leqslant N}$ of the hedge $\widetilde{\pi}$ satisfy the relation*

$$\widetilde{\beta}_n = M_n - \frac{\widetilde{\gamma}_n S_n}{B_n},$$

where $\widetilde{\gamma}_n$, $n \leqslant N$, can be determined from the '$\dfrac{S}{B}$-representation' (8) for the martingale $M = (M_n, \mathscr{F}_n, \widetilde{\mathsf{P}}_N)_{n \leqslant N}$ with

$$M_n = \widetilde{\mathsf{E}}\left(\frac{f_N}{B_N} \,\Big|\, \mathscr{F}_n \right).$$

3. As seen from the statement of this theorem, finding a perfect hedge $\widetilde{\pi} = (\widetilde{\beta}, \widetilde{\gamma})$ is intimately connected with the representation of the martingale $M = (M_n, \mathscr{F}_n, \widetilde{\mathsf{P}}_N)_{n \leqslant N}$ in one of the equivalent forms (8), (11), or (17). The next result concerns one interesting case of such a representation.

THEOREM 2. *Assume that a pay-off function is as follows:*

$$f_N = B_N\, g(\Delta_N), \tag{22}$$

where $g = g(\Delta_N)$ is a function of $\Delta_N = \delta_1 + \cdots + \delta_N$.

Then the coefficients in the representation (17) *are*

$$\widetilde{\alpha}_k^{(\delta)} = G_{N-k}(\Delta_{k-1}; \widetilde{p}), \qquad 1 \leqslant i \leqslant N, \tag{23}$$

where $\Delta_0 = 0$ and

$$G_n(x; \widetilde{p}) = \sum_{k=0}^{n} \bigl[g(x+k+1) - g(x+k)\bigr] C_n^k\, \widetilde{p}^{\,k}\, \widetilde{q}^{\,n-k}. \tag{24}$$

Proof. First of all, we note that $M_n = \widetilde{\mathsf{E}}(M_N \mid \mathscr{F}_n)$, where $M_N = \dfrac{f_N}{B_N}$.

Since $\Delta M_n = \widetilde{\alpha}_n^{(\delta)} \Delta m_n^{(\delta)}$, the value of $\widetilde{\alpha}_n^{(\delta)} = \widetilde{\alpha}_n^{(\delta)}(\delta_1, \dots, \delta_{n-1})$ can be expressed as follows:

$$\begin{aligned}\widetilde{\alpha}_n^{(\delta)} &= \frac{\widetilde{\mathsf{E}}(M_N \mid \delta_1, \dots, \delta_{n-1}, 1) - \widetilde{\mathsf{E}}(M_N \mid \delta_1, \dots, \delta_{n-1})}{1-\widetilde{p}} \\ &= \frac{\widetilde{\mathsf{E}}\big(g(\Delta_N) \mid \delta_1, \dots, \delta_{n-1}, 1\big) - \widetilde{\mathsf{E}}\big(g(\Delta_N) \mid \delta_1, \dots \delta_{n-1}\big)}{1-\widetilde{p}}. \end{aligned} \tag{25}$$

In the set $\{\omega\colon \Delta_{n-1} = x,\ \delta_n = 1\}$ we have

$$\widetilde{\mathsf{E}}\big(g(\Delta_N) \mid \mathscr{F}_n\big) = \widetilde{\mathsf{E}}\, g(x + 1 + \Delta_N - \Delta_n)$$

and

$$\begin{aligned}\widetilde{\mathsf{E}}\big(g(\Delta_N) \mid \mathscr{F}_{n-1}\big) &= \widetilde{\mathsf{E}}\, g(x + \Delta_N - \Delta_{n-1}) \\ &= \widetilde{p}\, \widetilde{\mathsf{E}}\, g(x + 1 + \Delta_N - \Delta_n) + (1-\widetilde{p})\, \widetilde{\mathsf{E}}\, g(x + \Delta_N - \Delta_n).\end{aligned}$$

Hence we obtain there

$$\begin{aligned}&\widetilde{\mathsf{E}}\big(g(\Delta_N) \mid \mathscr{F}_n\big) - \widetilde{\mathsf{E}}\big(g(\Delta_N) \mid \mathscr{F}_{n-1}\big) \\ &\quad = (1-\widetilde{p})\, \widetilde{\mathsf{E}}\big[g(x + 1 + \Delta_N - \Delta_n) - g(x + \Delta_N - \Delta_n)\big] \\ &\quad = (1-\widetilde{p}) \sum_{k=0}^{N-n} \big[g(x+1+k) - g(x+k)\big] C_{N-n}^k\, \widetilde{p}^k (1-\widetilde{p})^{N-n-k} \\ &\quad = (1-\widetilde{p}) G_{N-n}(x;\widetilde{p}),\end{aligned}$$

which, in view of (25), brings us to the required representation (23).

§ 4c. Rational Pricing and Hedging Strategies. Markovian Pay-Off Functions

1. We shall now assume that the pay-off function f_N has the 'Markovian' form $f_N = f(S_N)$, where $f = f(x)$ is a nonnegative function of $x \geqslant 0$.

Let

$$X_n^{\widetilde{\pi}} = \widetilde{\mathsf{E}}\big[f(S_N)(1+r)^{-(N-n)} \mid \mathscr{F}_n\big] \tag{1}$$

be the value of a perfect hedge $\widetilde{\pi}$ at time n; in particular,

$$\mathbb{C}(f_N;\mathsf{P}) = X_0^{\widetilde{\pi}} = \widetilde{\mathsf{E}}\big[f(S_N)(1+r)^{-N}\big]. \tag{2}$$

We set

$$F_n(x;p) = \sum_{k=0}^{n} f\big(x(1+b)^k(1+a)^{n-k}\big) C_n^k\, p^k (1-p)^{n-k}. \tag{3}$$

We have

$$\prod_{n<k\leqslant N} (1+\rho_k) = (1+b)^{\Delta_N-\Delta_n}(1+a)^{(N-n)-(\Delta_N-\Delta_n)} \tag{4}$$

with $\Delta_n = \delta_1 + \cdots + \delta_n$ for $\delta_k = \dfrac{\rho_k - a}{b-a}$, and therefore

$$\widetilde{\mathsf{E}} f\Big(x \prod_{n<k\leqslant N} (1+\rho_k)\Big) = F_{N-n}(x;\widetilde{p}) \tag{5}$$

with $\widetilde{p} = \dfrac{r-a}{b-a}$.

Taking finally into account the equality

$$S_N = S_n \prod_{n<k\leqslant N} (1+\rho_k), \tag{6}$$

we obtain by (5) the following result.

THEOREM 1. *The value $X^{\widetilde{\pi}} = (X_n^{\widetilde{\pi}})_{n\leqslant N}$ of a perfect hedge $\widetilde{\pi}$ in the CRR-model with Markovian pay-off function $f_N = f(S_N)$ can be described by the formulas*

$$X_n^{\widetilde{\pi}} = (1+r)^{-(N-n)} F_{N-n}(S_n;\widetilde{p}). \tag{7}$$

In particular, the rational option price is

$$\mathbb{C}(f_N;\mathsf{P}) \equiv X_0^{\widetilde{\pi}} = (1+r)^{-N} F_N(S_0;\widetilde{p}). \tag{8}$$

2. We consider now the question of the *structure* of the perfect hedge $\widetilde{\pi} = (\widetilde{\beta}, \widetilde{\gamma})$. We set

$$g(x) = \frac{f\big(S_0(1+b)^x(1+a)^{N-x}\big)}{B_N}. \tag{9}$$

Then, by Theorem 2 in the preceding section, the coefficients in the representation

$$M_N = M_0 + \sum_{k=1}^{N} \widetilde{\alpha}_k^{(\delta)} (\delta_k - \widetilde{p})$$

for

$$M_N = \frac{X_N^{\widetilde{\pi}}}{B_N} = \frac{f(S_N)}{B_N}$$

are predictable functions:

$$\widetilde{\alpha}_i^{(\delta)} = G_{N-i}(\Delta_{i-1}; \widetilde{p}), \tag{10}$$

where

$$G_{N-i}(x; \widetilde{p}) = \frac{1}{B_N} \sum_{k=0}^{N-i} C_{N-i}^k \widetilde{p}^k (1-\widetilde{p})^{N-i-k}$$
$$\times \left[f\left(S_0(1+a)^N \left(\frac{1+b}{1+a} \right)^{x+k+1} \right) - f\left(S_0(1+a)^N \left(\frac{1+b}{1+a} \right)^{x+k} \right) \right]. \tag{11}$$

Setting here $x = \Delta_{i-1}$ and taking account of the equality

$$S_{i-1} = S_0(1+a)^{i-1} \left(\frac{1+b}{1+a} \right)^{\Delta_{i-1}}$$

and notation (3), we obtain by (11) that

$$G_{N-i}(\Delta_i; \widetilde{p}) = \frac{1}{B_N} \big[F_{N-i}\big(S_{i-1}(1+b); \widetilde{p}\big) - F_{N-i}\big(S_{i-1}(1+a); \widetilde{p}\big) \big]. \tag{12}$$

Now, note that, by formulas (12) and (16) in § 4b,

$$\widetilde{\gamma}_i = \frac{\widetilde{\alpha}_i^{(\delta)} B_i}{S_{i-1}(b-a)} = \frac{G_{N-i}(\Delta_{i-1}; \widetilde{p}) B_i}{S_{i-1}(b-a)}. \tag{13}$$

By (12) and (13) we obtain

$$\widetilde{\gamma}_i = (1+r)^{-(N-i)} \cdot \frac{F_{N-i}(S_{i-1}(1+b); \widetilde{p}) - F_{N-i}(S_{i-1}(1+a); \widetilde{p})}{S_{i-1}(b-a)}. \tag{14}$$

As in § 4b, we set

$$\widetilde{\beta}_i = M_i - \frac{\widetilde{\gamma}_i S_i}{B_i}.$$

The strategy $\widetilde{\pi} = (\widetilde{\beta}, \widetilde{\gamma})$ is self-financing, therefore

$$\Delta \widetilde{\beta}_i \cdot B_{i-1} + \Delta \widetilde{\gamma}_i \cdot S_{i-1} = 0.$$

Hence

$$X_{i-1}^{\widetilde{\pi}} = \widetilde{\beta}_{i-1} B_{i-1} + \widetilde{\gamma}_{i-1} S_{i-1} = \widetilde{\beta}_i B_{i-1} + \widetilde{\gamma}_i S_{i-1},$$

so that

$$\widetilde{\beta}_i = \frac{X_{i-1}^{\widetilde{\pi}}}{B_{i-1}} - \frac{\widetilde{\gamma}_i S_{i-1}}{B_{i-1}}.$$

Bearing in mind (7) and (14) we obtain

$$\begin{aligned} \widetilde{\beta}_i &= \frac{F_{N-i+1}(S_{i-1};\widetilde{p})}{B_N} - \frac{G_{N-i}(\Delta_{i-1};\widetilde{p})(1+r)}{b-a} \\ &= \frac{1}{B_N}\Big\{F_{N-i+1}(S_{i-1};\widetilde{p}) \\ &\qquad - \frac{1+r}{b-a}\big[F_{N-i}(S_{i-1}(1+b);\widetilde{p}) - F_{N-i}\big(S_{i-1}(1+a);\widetilde{p}\big)\big]\Big\}. \end{aligned} \tag{15}$$

We now sum up the results so obtained.

THEOREM 2. *The components $\widetilde{\beta} = (\widetilde{\beta}_i)_{i\leqslant N}$ and $\widetilde{\gamma} = (\widetilde{\gamma}_i)_{i\leqslant N}$ of a perfect hedge $\widetilde{\pi} = (\widetilde{\beta}, \widetilde{\gamma})$ in the CRR-model with $f_N = f(S_N)$ can be defined by formulas* (14) *and* (15).

COROLLARY 1. *The predictable functions $\widetilde{\beta}_i$ and $\widetilde{\gamma}_i$ depend on the 'past' only through the variable S_{i-1}:*

$$\widetilde{\beta}_i = \widetilde{\beta}_i(S_{i-1}), \qquad \widetilde{\gamma}_i = \widetilde{\gamma}_i(S_{i-1}).$$

COROLLARY 2. *Let $f = f(x)$ be a nondecreasing nonnegative function. Then it follows from* (3) *and* (14) *that if $\widetilde{\pi} = (\widetilde{\beta}, \widetilde{\gamma})$ is a perfect hedge, then $\widetilde{\gamma}_i \geqslant 0$ for all $i \leqslant N$.*

Remark. One can interpret negative $\widetilde{\gamma}_i$ as borrowing stock (short-selling). Then Corollary 2 means that if $f(x)$ is non-decreasing, then no short-selling is necessary for perfect hedging.

§ 4d. Standard Call and Put Options

1. For a standard call option,

$$f(S_N) = (S_N - K)^+,$$

where N is the maturity time and K is the strike price. Of course, formulas for the rational price and perfect hedge obtained in the preceding section look more simple in this case.

By definition (3) in § 4c,

$$F_n(S_0;\widetilde{p}) = \sum_{k=0}^{n} C_n^k \widetilde{p}^k (1-\widetilde{p})^{n-k} \max\left\{0,\ S_0(1+a)^N\Big(\frac{1+b}{1+a}\Big)^k - K\right\}. \tag{1}$$

Let

$$K_0 = K_0\left(a, b, N; \frac{S_0}{K}\right)$$

be the smallest integer such that

$$S_0(1+a)^N\Big(\frac{1+b}{1+a}\Big)^{K_0} > K. \tag{2}$$

For $f_N = (S_N - K)^+$ we shall denote for brevity $\mathbb{C}(f_N; \mathsf{P})$ by $\mathbb{C}_N$ (or by $\mathbb{C}_N^{(K)}$ if we want to underline the dependence on K).

If $K_0 > N$, then $F_N(S_0;\widetilde{p}) = 0$, and therefore the rational price $\mathbb{C}_N$ is equal to zero (see (8), § 4c); this is understandable since we surely have $S_N < K$ in this case, and the purchase of an option brings no profits.

We shall assume for that reason that $K_0 \leqslant N$. Then

$$\begin{aligned}
\mathbb{C}_N &= (1+r)^{-N} F_N(S_0;\widetilde{p}) \\
&= S_0 \sum_{k=K_0}^{N} C_N^k \widetilde{p}^k (1-\widetilde{p})^{N-k}\Big(\frac{1+a}{1+r}\Big)^N\Big(\frac{1+b}{1+a}\Big)^k \\
&\quad - K(1+r)^{-N} \sum_{k=K_0}^{N} C_N^k \widetilde{p}^k (1-\widetilde{p})^{N-k}.
\end{aligned} \tag{3}$$

We set

$$p^* = \frac{1+b}{1+r}\,\widetilde{p}, \tag{4}$$

$$\mathbb{B}(j, N; p) = \sum_{k=j}^{N} C_N^k p^k (1-p)^{N-k}. \tag{5}$$

Using this notation, we can formulate the result so obtained (which is originally due to J. C. Cox, S. A. Ross, and M. Rubinstein [82]) as follows.

THEOREM. *The fair (rational) price of a standard European option with pay-off* $f(S_N) = (S_N - K)^+$ *is*

$$\mathbb{C}_N = S_0\mathbb{B}(K_0, N; p^*) - K(1+r)^{-N}\mathbb{B}(K_0, N; \widetilde{p}), \tag{6}$$

where

$$K_0 = 1 + \left[\ln\frac{K}{S_0(1+a)^N}\Big/\ln\frac{1+a}{1+b}\right]. \tag{7}$$

If $K_0 > N$, *then* $\mathbb{C}_N = 0$.

2. Since

$$(K - S_N)^+ = (S_N - K)^+ - S_N + K,$$

the rational (fair) price $\mathbb{P}_N$ of a put option can be defined by the formula

$$\begin{aligned}\mathbb{P}_N &= \widetilde{\mathsf{E}}(1+r)^{-N}(K - S_N)^+ \\ &= \mathbb{C}_N - \widetilde{\mathsf{E}}(1+r)^{-N}S_N + K(1+r)^{-N}.\end{aligned} \tag{8}$$

Here $\widetilde{\mathsf{E}}(1+r)^{-N}S_N = S_0$. Hence we have the following identity (the *call-put parity*):

$$\mathbb{P}_N = \mathbb{C}_N - S_0 + K(1+r)^{-N}. \tag{9}$$

3. Let $f = f(S_N)$ be a pay-off function and let $\mathbb{C}_N^{(f)} = B_0\,\widetilde{\mathsf{E}}\,\dfrac{f(S_N)}{B_N}$ be the corresponding rational (fair) price.

The following observation (see, e.g., [121], [122]) shows how one can use the rational prices $\mathbb{C}_N^{(K)}$ corresponding to the pay-off $(S_N - K)^+$, $K \geqslant 0$, in the search of the values of $\mathbb{C}_N^{(f)}$ for options with *other* types of pay-off functions f.

Assume that the derivative of the pay-off function $f = f(x)$, $x \geqslant 0$, can be expressed as an integral: $f'(x) = \int_0^x \mu(dy)$, where $\mu = \mu(dy)$ is a finite measure (not necessarily of constant sign) on $(\mathbb{R}_+, \mathscr{B}(\mathbb{R}_+))$. (If $f(y)$ has a second derivative in the usual sense, then $\mu(dy) = f''(y)\,dy$.) Then it is straightforward that

$$f(x) = f(0) + xf'(0) + \int_0^\infty (x - K)^+\,\mu(dK)$$

and therefore

$$f(S_N) = f(0) + S_N f'(0) + \int_0^\infty (S_N - K)^+\,\mu(dK) \quad (\mathsf{P}\text{-a.s.}).$$

We consider now the expectation with respect to the martingale measure $\widetilde{\mathsf{P}}_N$ and obtain

$$\widetilde{\mathsf{E}}\frac{f(S_N)}{B_N} = \frac{f(0)}{B_N} + \frac{S_0}{B_0}f'(0) + \int_0^\infty \widetilde{\mathsf{E}}\,\frac{(S_N - K)^+}{B_N}\,\mu(dK),$$

so that, by formula (6) in § 1b,

$$\mathbb{C}_N^{(f)} = (1+r)^{-N}f(0) + S_0 f'(0) + \int_0^\infty \mathbb{C}_N^{(K)}\,\mu(dK). \tag{10}$$

Note that if $f(x) = (x - K_*)^+$, $K_* > 0$, then $\mu(dK)$ is concentrated at the point K_* (i.e., $\mu_*(dK) = \delta_{\{K_*\}}(dx)$) and $\mathbb{C}_N^{(f)} = \mathbb{C}_N^{(K_*)}$, as one would expect.

4. Formulas (6) and (9) answer the question on the *rational price* of put and call options. It is also of considerable practical interest to the option writer to know how to find a perfect hedge $\widetilde{\pi} = (\widetilde{\beta}, \widetilde{\gamma})$; this can be carried out on the basis of formulas (15) and (14) in the preceding section. We do not analyze these formulas thoroughly here; we content ourselves with one simple example, the idea of which is borrowed from [162]. (See also [443] and a similar illustrative example at the beginning of this chapter.)

EXAMPLE. Consider two currencies, $\mathbb{A}$ and $\mathbb{B}$. Let S_n be the price of 100 units of $\mathbb{A}$ expressed in the units of $\mathbb{B}$ for $n = 0$ and 1. Let $S_0 = 150$ and assume that the price S_1 at time $n = 1$ is expected to be either 180 (the currency $\mathbb{A}$ rises) or 90 (the currency $\mathbb{A}$ falls).

We write

$$S_1 = S_0(1 + \rho_1), \tag{11}$$

and see that ρ_1 can take two values, $b = \frac{1}{5}$ and $a = -\frac{2}{5}$, which correspond to a rise or a drop in the cross rate of $\mathbb{A}$.

Let $B_0 = 1$ (in the units of $\mathbb{B}$) and let $r = 0$. Thus, we assume (for simplicity) that funds put into a bank account bring no profits and no interest on loans is taken.

Let $N = 1$ and let $f(S_1) = (S_1 - K)^+$, where $K = 150(\mathbb{B})$, i.e., $K = 150$ (units of $\mathbb{B}$). Thus, if the currency $\mathbb{A}$ rises, then a buyer of a call option obtains $180 - 150 = 30$ (units of $\mathbb{B}$), whereas if the exchange rate falls, then $f(S_1) = 0$.

So far, nothing has been said on the probabilities of the events $\rho_1 = b$ and $\rho_1 = a$. Assuming that $\mathbb{A}$ can rise or fall with probability $\frac{1}{2}$ we obtain that $\mathsf{E}f(S_1) = 30 \cdot \frac{1}{2} = 15$. A classical view, dating back to the times of J. Bernoulli and C. Huygens (see, e.g., [186; pp. 397–402]), is that $\mathsf{E}f(S_1) = 15$ (units of $\mathbb{B}$) could be a reasonable price of such an option.

It should be emphasized, however, that this quantity *depends* essentially on our assumption on the values of the probabilities $p = \mathsf{P}(\rho_1 = b)$ and $1 - p = \mathsf{P}(\rho_1 = a)$. If $p = \frac{1}{2}$, then, as we see, $\mathsf{E}f(S_1) = 15\,(\mathbb{B})$. However, if $p \neq \frac{1}{2}$, then we obtain another value of $\mathsf{E}f(S_1)$.

Taking into account that, in real life, one usually *has no conclusive evidence in favor of some or other values of* p, one understands that the classical approach to the calculation of rational prices is far from satisfactory.

Rational pricing theory exposed above works under the assumption that p, $0 < p < 1$, is *arbitrary*. The value that must enter the (classical scheme of) calculations is

$$\widetilde{p} = \frac{r - a}{b - a}.$$

In our example

$$\widetilde{p} = \frac{0 + \frac{2}{5}}{\frac{1}{5} + \frac{2}{5}} = \frac{2}{3}.$$

If $N = 1$, then the corresponding value $K_0 = K_0(a, b, 1; S_0/K)$ is 1 for $a = -\frac{2}{5}$, $b = \frac{1}{5}$, and $S_0 = K = 150$, therefore

$$\mathbb{C}_1 = S_0 \widetilde{p}(1+b) - k\widetilde{p} = S_0 \widetilde{p} b = 150 \cdot \frac{2}{3} \cdot \frac{1}{5} = 20$$

by (3).

Hence the buyer of an option must pay a premium $\mathbb{C}_1$ of 20 units of $\mathbb{B}$, which can now be regarded as the *starting capital* $X_0 = 20\,(\mathbb{B})$ of the writer (issuer), who invests it on the market.

We represent now X_0 in the standard form (for (B, S)-markets) $X_0 = \beta_0 B_0 + \gamma_0 S_0$. Setting $B_0 = 1$ and $S_0 = 150$ we can express the capital $X_0 = 20\,(\mathbb{B})$ as follows: $20 = 0 + \frac{2}{15} \cdot 150$. That is, $\beta_0 = 0$ and $\gamma_0 = \frac{2}{15}$, which can be deciphered as follows: the issuer puts 0 units of $\mathbb{B}$ into the bank account, while $\gamma_0 \cdot S_0 = \frac{2}{15} \cdot 150 = 20$ units $\mathbb{B}$ can be converted into the currency $\mathbb{A}$.

Assume that the issuer can also borrow money (from the bank account B, in the currency $\mathbb{B}$), which, of course, should be paid back in the future. Then we can represent the initial capital $X_0 = 20\,(\mathbb{B})$ as the sum $X_0 = -30 + \frac{1}{3} \cdot 150$, which corresponds to the portfolio $(\beta_0, \gamma_0) = (-30, \frac{1}{3})$ meaning that the issuer borrows 30 units of $\mathbb{B}$ and can now exchange $\frac{1}{3} \cdot 150 = 50$ units of $\mathbb{B}$ for 33.33 units of $\mathbb{A}$.

Assume that, as an investor on our (B, S)-market, the issuer chooses $(\beta_1, \gamma_1) = (\beta_0, \gamma_0)$. What does his portfolio bring at the instant $N = 1$?

By the assumption that $B_1 = B_0 = 1$, there will be $\beta_1 B_1 = -30$ units of $\mathbb{B}$ in the bank account.

If the currency $\mathbb{A}$ rises ($180\,\mathbb{B} = 100\,\mathbb{A}$), then 33.33 units of $\mathbb{A}$ will be worth 60 units of $\mathbb{B}$, of which 30 is the outstanding debt. On paying it back the issuer still has $60 - 30 = 30$ units of $\mathbb{B}$, which he will pay to the buyer of the option to meet the conditions of the contract.

On the other hand, if $\mathbb{A}$ falls, then 33.33 units of $\mathbb{A}$ will be worth 30 units of $\mathbb{B}$, which should be paid back to the bank. Nothing must be paid to the buyer (who has lost), so that the issuer 'comes clean'.

Our choice of the portfolio $(\beta_1, \gamma_1) = (-30, \frac{1}{3})$ may appear *ad hoc*. However, these are just the values suggested by the above theory.

In fact, by formula (14) in § 4c the 'optimal' value $\gamma_1 = \gamma_1(S_0)$ that is a component of a perfect hedge can be calculated as follows:

$$\begin{aligned}
\gamma_1(S_0) &= \frac{F_0(S_0(1+b); \widetilde{p}) - F_0(S_0(1+a); \widetilde{p})}{S_0(b-a)} \\
&= \frac{f(S_0(1+b)) - f(S_0(1+a))}{S_0(b-a)} = \frac{f(S_0(1+b))}{S_0(b-a)} \\
&= \frac{(S_0(1+b) - K)^+}{S_0(b-a)} = \frac{b}{b-a} = \frac{1/5}{1/5 + 2/5} = \frac{1}{3}.
\end{aligned}$$

The value $\beta_1 = \beta_0$ can be defined from the condition

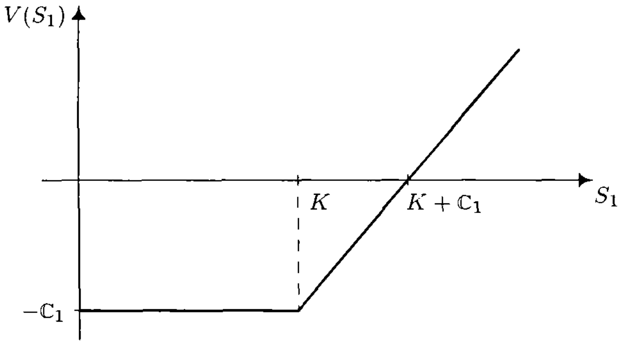
$$X_0 = \beta_0 + \gamma_0 S_0.$$

Since $X_0 = 20$, $\gamma_0 = \frac{1}{3}$, and $S_0 = 150$, it follows that $\beta_1 = \beta_0 = -30$, just as chosen above.

It is clear from these arguments that the buyer's net profit $V(S_1)$ (as a function of S_1 for fixed K) can be described by the formula

$$V(S_1) = (S_1 - K)^+ - \mathbb{C}_1.$$

The graph of this function is as follows:



Of course, the question on the writer's profits also seems appropriate in this context.

It is easy to see that there are none in the above example, for both cases of raising and falling currency $\mathbb{A}$. So, how can there be someone ready to float options and other kinds of derivatives on financial markets?

In fact, the situation is more complex because, first of all, one must take into account overheads, the broker's commission, taxes, and the like, which, of course, increases the size of the premium calculated above. For instance, the commission can be regarded as the writer's profit. Moreover, one must bear in mind that the writer has control (not necessarily for long) over the collected premiums and can use these funds for gaining some money for himself.

Some may also wonder at the variety of kinds of options and other derivatives traded in the market.

One possible explanation is that there are always some who expect currencies, stock prices, and the like to rise or fall. Hence there must exist someone who derives profit from this. This is what issuers are doing by floating call options (designed for 'bulls'), put options (designed for 'bears'), or their combinations with derivatives of other types.

§ 4e. Option-Based Strategies (Combinations and Spreads)

1. In practice one can encounter most diversified kinds of options and their combinations. We listed several kinds of options ('with aftereffect', 'Asian', etc.) in Chapter I, § 1c. Some of them, due to their peculiarity and intricacy, are called 'exotic options' (see, e.g., [414]).

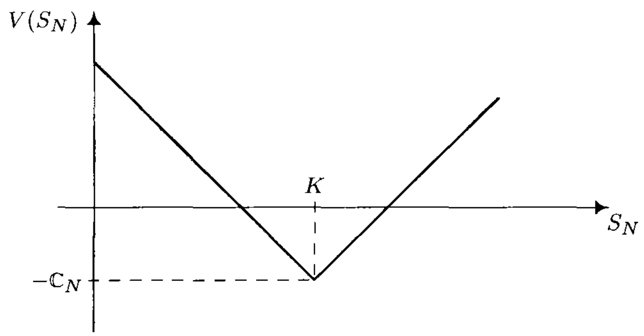
We now list (and characterize) several popular strategies based on *different kinds of options*. Usually, one classifies these strategies between *combinations* and *spreads*. The distinction is that combinations are made up from options of *different* kinds, while spreads include options of *one* kind. (See, e.g., [50] for a more thorough description, details of corresponding calculations, and a list of books touching upon *financial engineering*, which considerably relies upon option-based strategies.)

2. Combinations

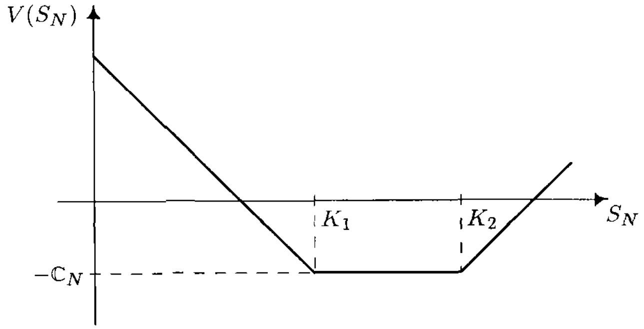
Straddle is a combination of call and put options for the same stock with the same strike price K and of the same maturity N. The buyer's gain-and-loss function $V(S_N)$ ($= f(S_N) - \mathbb{C}_N$) for such a combination is as follows:

$$V(S_N) = |S_N - K| - \mathbb{C}_N.$$

Its graph is as in the chart below.



Strangle is a combination of call and put options of the same maturity N, but of different strike prices K_1 and K_2. The typical graph of the buyer's gain-and-loss function $V(S_N)$ is as follows:



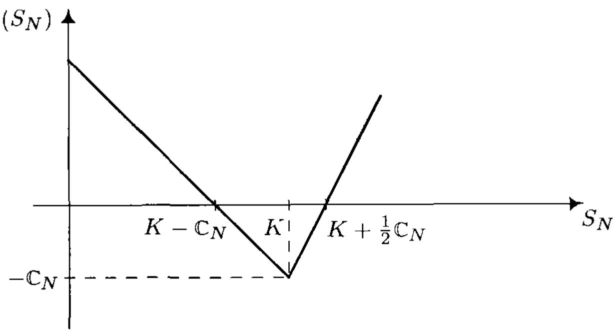
Analytically, $V(S_N)$ has the following expression:

$$V(S_N) = |S_N - K_2| I(S_N > K_2) + |S_N - K_1| I(S_N < K_1) - \mathbb{C}_N.$$

Strap is a combination of one put option and two call options of the same maturity N, but, in general, of different strike prices K_1 and K_2. If $K_1 = K_2 = K$, then

$$V(S_N) = 2|S_N - K| I(S_N > K) + |S_N - K| I(S_N < K) - \mathbb{C}_N.$$

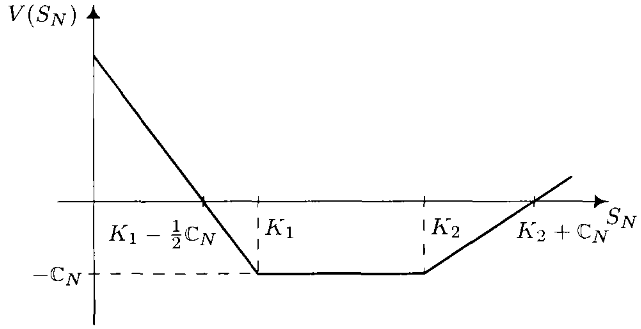
The graph depicting the behavior of $V(S_N)$ is now asymmetric:



Strip is a combination of one call option and two put options of the same maturity N, but, in general, of different strike prices K_1 and K_2. The gain-and loss function is

$$V(S_N) = |S_N - K_2| I(S_N > K_2) + 2|S_N - K_1| I(S_N < K_1) - \mathbb{C}_N;$$

and its graph has the following form:

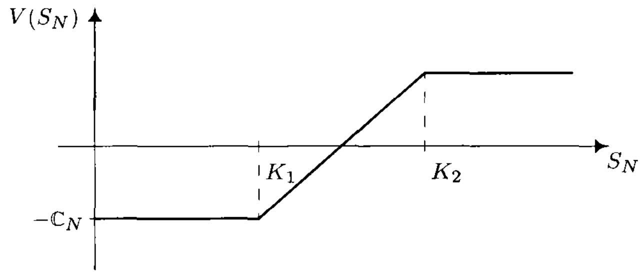


3. Spreads

Bull spread is the strategy of buying a call option with strike price K_1 and selling a call option with (higher) strike price $K_2 > K_1$. Here

$$V(S_N) = |K_2 - K_1| I(S_N \geqslant K_2) + |S_N - K_1| I(K_1 < S_N < K_2) - \mathbb{C}_N,$$

and the graph of this function is as follows:

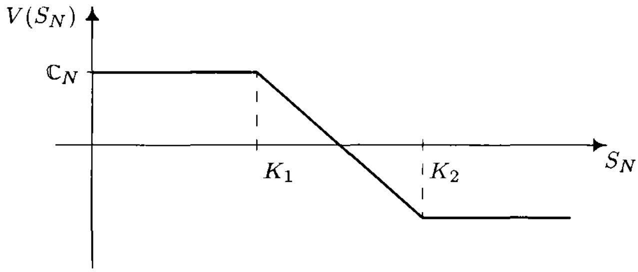


It is reasonable to buy a bull spread when an investor anticipates a rise in prices (of some stock, say), but wants to reduce potential losses. However, this combination restricts also the potential gains.

Bear spread is the strategy of selling a call option with strike price K_1 and buying a call option with strike price $K_2 > K_1$. For this combination

$$V(S_N) = -|K_2 - K_1| I(S_N \geqslant K_2) + |S_N - K_1| I(K_1 < S_N < K_2) + \mathbb{C}_N.$$

The corresponding graph is as follows:



Such a combination makes sense if the investor anticipates a drop in prices, but wants to cap losses due to possible rises of stock.

As regards other kinds of spreads, see [50], § 24.

4. On securities markets one comes across other *combinations* besides the above-mentioned ones, involving standard (call and put) options. For instance, there exists a strategy of buying *options* (a derivative security) and *stock* (the underlying security) at the same time. Investors choose such strategies in an attempt to insure against a drop in stock prices below certain level. If this occurs, then the investor who has bought a put option can sell his stock at the (higher) strike price, rather than at the (lower) spot price. (See [50], § 22.)

5. American Options on a Binomial (B, S)-Market

§ 5a. American Option Pricing

1. For American options the main pricing questions (both in the discrete and the continuous-time cases) take the following form:

(i) what is the rational (fair, mutually acceptable) price of option contracts with fixed collection of pay-off functions?
(ii) what is the rational time for exercising an option?
(iii) what optimal hedging strategy of an option writer ensures his ability to meet the conditions of the contract?

In the present section, concerned with American option pricing in the *discrete-time* case, we pay most attention to the first two questions, (i) and (ii). In principle, questions of type (iii), on particular hedging strategies, are answered by Theorems 2 and 3 in § 2c.

2. We shall stick to the *CRR*-model of a (B, S)-market described in § 4b. That is, we assume that $\Delta B_n = rB_{n-1}$ and $\Delta S_n = \rho_n S_{n-1}$; here $\rho = (\rho_n)$ is a sequence of independent identically distributed random variables such that $\mathsf{P}(\rho_n = b) = p$ and $\mathsf{P}(\rho_n = a) = q$, where $-1 < a < r < b$, $p + q = 1$, $0 < p < 1$.

An additional assumption enabling us to simplify considerably the analysis that follows is that

$$b = \lambda - 1 \quad \text{and} \quad a = \lambda^{-1} - 1 \tag{1}$$

for some $\lambda > 1$.

Thus, in place of *two* parameters, a and b, determining the evolution of the prices S_n, $n \geqslant 1$, we must fix a *single* parameter $\lambda > 1$, which defines a and b by formulas (1).

Clearly, in this case we have

$$S_n = S_0 \lambda^{\varepsilon_1 + \cdots + \varepsilon_n}, \tag{2}$$

where $\mathsf{P}(\varepsilon_i = 1) = \mathsf{P}(\rho_i = b) = p$ and $\mathsf{P}(\varepsilon_i = -1) = \mathsf{P}(\rho_i = a) = q$ (cf. Chapter II, § 1e).

Assuming also that S_0 belongs to the set $E = \{\lambda^k,\ k = 0, \pm 1, \dots\}$, we see that for each $n \geqslant 1$ the state S_n also belongs to E.

The sequence $S = (S_n)_{n\geqslant 0}$ described by relation (2) with $S_0 \in E$ is usually called a *geometric random walk* over the set of states $E = \{\lambda^k\, k = 0, \pm 1, \dots\}$ (cf. Chapter II, § 1e).

Let $x \in E$ and let $\mathsf{P}_x = \operatorname{Law}\bigl((S_n)_{n\geqslant 0} \,|\, \mathsf{P},\, S_0 = x\bigr)$ be the probability distribution of the sequence $(S_n)_{n\geqslant 0}$ with respect to P under the assumption that $S_0 = x$.

In accordance with the standard nomenclature of the theory of stochastic processes, we can say that the sequence $S = (S_n)_{n\geqslant 0}$ with family of probabilities P_x, $x \in E$, makes up a *homogeneous Markov random walk*, or a *homogeneous Markov process* (with discrete times).

Let T be the one-step *transition operator*, i.e., for a function $g = g(x)$ on E we set

$$Tg(x) = \mathsf{E}_x g(S_1), \qquad x \in E, \tag{3}$$

where E_x is averaging with respect to the measure P_x.

In our case (2) we have

$$Tg(x) = pg(\lambda x) + (1-p)g\Bigl(\frac{x}{\lambda}\Bigr). \tag{4}$$

3. The (B,S)-market described by the *CRR*-model is both arbitrage-free and complete, and the unique martingale measure $\widetilde{\mathsf{P}}$ has the following properties:

$$\widetilde{\mathsf{P}}(\varepsilon_i = 1) = \widetilde{\mathsf{P}}(\rho_i = b) = \frac{r-a}{b-a}, \qquad \widetilde{\mathsf{P}}(\varepsilon_i = -1) = \widetilde{\mathsf{P}}(\rho_i = a) = \frac{b-r}{b-a}.$$

(See, e.g., Chapter V, § 1d.)

The 'arbitrage-free martingale' ideology of Chapter V requires that all probabilistic calculations proceed with respect to the martingale measure $\widetilde{\mathsf{P}}$, rather than the original measure P. To avoid additional notation we shall assume that $\mathsf{P} = \widetilde{\mathsf{P}}$ from the very beginning, so that

$$p = \frac{r-a}{b-a} \quad \text{and} \quad q = \frac{b-r}{b-a}. \tag{5}$$

Bearing in mind (1) we see that

$$p = \frac{\alpha^{-1} - \lambda^{-1}}{\lambda - \lambda^{-1}}, \qquad q = \frac{\lambda - \alpha^{-1}}{\lambda - \lambda^{-1}}, \tag{6}$$

where $\alpha = (1+r)^{-1}$.

Let $f = (f_0, f_1, \dots)$ be a *system* of pay-off functions defined, as usual, on a filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_n)_{n\geqslant 0}, \mathsf{P})$ with $\mathscr{F}_0 = \{\varnothing, \Omega\}$.

In accordance with § 2a, let $\mathfrak{M}_n^N$ be the class of stopping times τ such that $n \leqslant \tau \leqslant N$. Let $\mathfrak{M}_n^\infty$ be the class of *finite* stopping times such that $\tau \geqslant n$.

The buyer of an American option chooses himself the instant τ of exercising it and obtains the amount of f_τ. If the contract is made at time $n = 0$ and its *expiry* date is $n = N$, then the buyer of an American option can choose any τ in the class $\mathfrak{M}_0^N$ as the time of exercising. Of course, the writer must allow for the *most unfavorable* buyer's choice of τ and (in incomplete markets) the 'Nature's choice' of a martingale measure among the possible ones. Thus, in accordance with § 1a, the writer of an American option must opt for a strategy bringing about American hedging.

Our (B, S)-market is complete, and the upper price of American hedging (see (5) in § 2c)

$$\widetilde{\mathbb{C}}_N(f; \mathsf{P}) = \inf\{y\colon \exists\, \pi \text{ with } X_0^\pi = y,\ X_\tau^\pi \geqslant f_\tau \ (\mathsf{P}\text{-a.s.}),\ \forall\, \tau \in \mathfrak{M}_0^N\}, \tag{7}$$

which can be reasonably taken as the price of the American option in question, can be calculated by the formula

$$\widetilde{\mathbb{C}}_N(f; \mathsf{P}) = \sup_{\tau\in\mathfrak{M}_0^N} B_0\, \mathsf{E}\, \frac{f_\tau}{B_\tau} \tag{8}$$

(see (19) in § 2c).

Recall that E here is averaging with respect to the (martingale) measure P.

4. We put the accent in the above discussion on the questions *how* and under *what* conditions the *seller* of an option can meet the terms of the contract.

According to the general theory of American hedges (subsection 2), the premium $\widetilde{\mathbb{C}}_N(f; \mathsf{P})$ for the option contract defined by (8) is the *smallest* price at which the writer can meet these terms.

The buyer is aware of this fact, and the price $\widetilde{\mathbb{C}}_N(f; \mathsf{P})$ is in this sense acceptable for both parties. Now, in accordance with the general theory, the writer can select a hedging portfolio $\overline{\pi}$ such that its value $X_\tau^{\overline{\pi}}$ is at least f_τ for each $\tau \in \mathfrak{M}_0^N$.

We now discuss the question *how* the *buyer*, who agrees to pay the premium $\widetilde{\mathbb{C}}_N(f; \mathsf{P})$ for the contract, can choose the time of exercising it in the most rational way.

Clearly, if it is shown at time σ when $X_\sigma^{\overline{\pi}} > f_\sigma$, then the writer obtains the net profit $X_\sigma^{\overline{\pi}} - f_\sigma$ after paying f_σ to the buyer. Hence the buyer should choose instant σ such that $X_\sigma^{\overline{\pi}} = f_\sigma$. Such an instant actually *exists* and, as follows from Theorem 4 in § 2c, it is the instant τ_0^N obtained in the course of the solution of the optimal stopping problem of finding the upper bound

$$\sup_{\tau\in\mathfrak{M}_0^N} B_0\, \mathsf{E}\, \frac{f_\tau}{B_\tau}. \tag{9}$$

5. We see from (8) that finding the price $\widetilde{\mathbb{C}}_N(f; \mathsf{P})$ reduces to the solution of the optimal stopping problem for the stochastic sequence $f_0, f_1, \dots, f_N$.

In § § 5b, c, following mainly [443], we shall consider the standard call and put options with pay-off functions $f_n = (S_n - K)^+$ and $f_n = (K - S_n)^+$ (or slightly more general functions $f_n = \beta^n(S_n - K)^+$ and $f_n = \beta^n(K - S_n)^+$), respectively. Together with the Markov property of the sequence $S = (S_n)_{n \geqslant 0}$, this special form of the functions f_n enables us to solve the optimal stopping problems in question using the 'Markovian version' of the theory of optimal stopping rules described in § 2a.5.

§ 5b. Standard Call Option Pricing

1. We consider a standard *call option* with pay-off function that has the following form at time n:

$$f_n(x) = \beta^n (x - K)^+, \qquad x \in E, \tag{1}$$

where $0 < \beta \leqslant 1$, $E = \{x = \lambda^k : k = 0, \pm 1, \dots\}$, and $\lambda > 1$.

For $0 \leqslant n \leqslant N$ we set

$$V_n^N(x) = \sup_{\tau \in \mathfrak{M}_n^N} \mathsf{E}_x (\alpha\beta)^\tau (S_\tau - K)^+, \tag{2}$$

where $S_{n+k} = S_n \lambda^{\varepsilon_{n+1} + \cdots + \varepsilon_{n+k}}$ and $S_n = x$.

It is worth noting that

$$V_n^N(x) = (\alpha\beta)^n V_0^{N-n}(x) \tag{3}$$

and, in accordance with relation (8) in § 5a (and under the assumption $S_0 = x$), the price in question is

$$\widetilde{\mathbb{C}}_N(f; \mathsf{P}) = V_0^N(x). \tag{4}$$

By Theorem 3 in § 2a and our remark upon it,

$$V_0^N(x) = Q^N g(x), \tag{5}$$

where $g(x) = (x - K)^+$ and

$$Qg(x) = \max\bigl(g(x), \alpha\beta T g(x)\bigr). \tag{6}$$

The optimal time τ_0^N exists in the class $\mathfrak{M}_0^N$ and can be found by the formula

$$\tau_0^N = \min\bigl\{0 \leqslant n \leqslant N : V_0^{N-n}(S_n) = g(S_n)\bigr\}. \tag{7}$$

Setting

$$\begin{aligned} D_n^N &= \bigl\{x \in E : V_0^{N-n}(x) = g(x)\bigr\} \\ &= \bigl\{x \in E : V_n^N(x) = (\alpha\beta)^n g(x)\bigr\}, \end{aligned} \tag{8}$$

we see that

$$\tau_0^N = \min\{0 \leqslant n \leqslant N : S_n \in D_n^N\}. \tag{9}$$

Thus, given a sequence of *stopping domains*

$$D_0^N \subseteq D_1^N \subseteq \cdots \subseteq D_N^N = E \tag{10}$$

and a sequence of *continuation domains*

$$C_0^N \supseteq C_1^N \supseteq \cdots \supseteq C_N^N = \varnothing \tag{11}$$

with $C_n^N = E \setminus D_n^N$, we can formulate the following rule for the option buyer concerning the time of exercising the contract.

If $S_0 \in D_0^N$, then $\tau_0^N = 0$; i.e., the buyer must agree at once to the pay-off $(S_0 - K)^+$.

On the other hand, if $S_0 \in C_0^N = E \setminus D_0^N$ (which is a typical situation), then the buyer must wait for the next value, S_1 and take the decision of whether $\tau_1^N = 1$ or $\tau_1^N > 1$ depending on whether $S_1 \in D_1^N$ or $S_1 \in C_1^N$, and so on.

In our case of a standard call option it is easy to describe, *in qualitative terms*, the geometry of the sets D_n^N and C_n^N, $0 \leqslant n \leqslant N$, and therefore also the buyer's strategy of choosing the exercise time.

2. We see from (4)–(7) that finding the function $V_0^N(x)$ and the instant τ_0^N reduces to finding recursively the functions $V_0^n(x) = Q^n g(x)$ for $n = 1, 2, \dots, N$.

By assumption, $0 < \beta \leqslant 1$. We claim that the case of $\beta = 1$ is elementary.

In fact, the sequence $(\alpha^n S_n)_{n\geqslant 0}$ is a martingale with respect to each measure P_x, $x \in E$; therefore $(\alpha^n(S_n - K))_{n\geqslant 0}$ is a *submartingale* and, by Jensen's inequality for the convex function $x \rightsquigarrow x^+$, the sequence $(\alpha^n(S_n - K)^+)_{n\geqslant 0}$ is also a submartingale.

Hence for each Markov time τ, $0 \leqslant \tau \leqslant N$, we have

$$\mathsf{E}_x \alpha^\tau (S_\tau - 1)^+ \leqslant \mathsf{E}_x \alpha^N (S_N - 1)^+ \tag{12}$$

by Doob's stopping theorem (Chapter V, § 3a). This immediately shows that we can take $\tau_0^N = N$ as an optimal stopping time in the problem

$$\sup_{\tau \in \mathfrak{M}_0^N} \mathsf{E}_x \alpha^\tau (S_\tau - K)^+,$$

and therefore if $S_0 = x$, then

$$\widetilde{\mathbb{C}}_N(f; \mathsf{P}) = V_0^N(x) = \mathsf{E}_x \alpha^N (S_N - K)^+. \tag{13}$$

Translating this into more practical terms we obtain the following result of R. Merton [346]:

> *if the discount factor β is equal to* 1, *then the standard American and European call options are 'the same'.*

In addition, the value of $V_0^N(x)$ can be found by formula (6) in § 4d.

3. We consider now a more interesting case of $0 < \beta < 1$.

THEOREM 1. *For each $N \geqslant 0$ there exists a sequence of numbers $x_n^N \in E \cup \{0\}$, $0 \leqslant n \leqslant N$, such that*

$$D_n^N = \{x \in E : x \in [x_n^N, \infty)\},$$
$$C_n^N = \{x \in E : x \in (0, x_n^N)\}$$

and

$$\tau_0^N = \min\{0 \leqslant n \leqslant N : S_n \in D_n^N\} = \min\{0 \leqslant n \leqslant N : S_n \in [x_n^N, \infty)\}.$$

Moreover,

$$0 = x_N^N \leqslant x_{N-1}^N \leqslant \cdots \leqslant x_0^N \tag{14}$$

and

$$V_0^N(x) = \begin{cases} g(x), & x \in D_0^N = [x_0^N, \infty), \\ Q^N g(x), & x \in C_0^N = (0, x_0^N). \end{cases} \tag{15}$$

The rational price $\widetilde{\mathbb{C}}_N(f; \mathsf{P})$ is equal to $V_0^N(S_0)$.

Proof. We set for simplicity $K = 1$, set consecutively $N = 1, 2, \dots$, and analyze $Q^n g(x)$ for $n \leqslant N$.

Let $N = 1$ and let $x = 1$ be the initial point, i.e., $x = \lambda^0$. By formula (4) in § 5a, we see for the function $g(x) = (x-1)^+$ (bearing in mind the inequality $\lambda > 1$) that

$$Tg(1) = pg(\lambda) + (1-p)g(\lambda^{-1}) = p(\lambda - 1) > 0,$$
$$Qg(1) = \max(g(1), \alpha\beta Tg(1)) = \max(0, \alpha\beta p(\lambda - 1)) = \alpha\beta p(\lambda - 1) > 0.$$

Hence the use of the operator Q 'rises' the value of $g = g(x)$ at the point $x = 1$ to $Qg(1) = \alpha\beta p(\lambda - 1)$.

In a similar way,

$$Tg(\lambda) = p(\lambda^2 - 1) \quad \text{and} \quad Qg(\lambda) = (\lambda - 1)\max\left(1, \beta\frac{\lambda - \alpha}{\lambda - 1}\right).$$

Hence if $\beta \leqslant \dfrac{\lambda - 1}{\lambda - \alpha}$, then Q does not change the value of $g(\lambda) = \lambda - 1$. However, if $\beta > \dfrac{\lambda - 1}{\lambda - \alpha}$, then Q 'rises' the value of $g(\lambda)$ to $\beta(\lambda - \alpha)$ $(> \lambda - 1)$.

Now let $x = \lambda^k$, $k > 1$. Then

$$Tg(\lambda^k) = \lambda^k\alpha^{-1} - 1 \quad \text{and} \quad Qg(\lambda^k) = \max(\lambda^k - 1, \beta(\lambda^k - \alpha)).$$

We note that

$$Qg(\lambda^k) = g(\lambda^k) \iff \lambda^k - 1 \geqslant \beta(\lambda^k - \alpha) \iff \lambda^k(1-\beta) \geqslant 1 - \alpha\beta. \qquad (16)$$

Since

$$\lambda^k(1-\beta) \geqslant 1 - \alpha\beta \implies \lambda^{k+1}(1-\beta) \geqslant 1 - \alpha\beta,$$

it follows that

$$Qg(\lambda^k) = g(\lambda^k) \implies Qg(\lambda^{k+1}) = g(\lambda^{k+1}),$$

which can be interpreted as follows: if $\lambda^k \in D_0^1$, then the points $\lambda^{k+1}, \lambda^{k+2}, \dots$ belong to D_0^1.

By (16) we obtain that for sufficiently large k the quantity λ^k belongs to D_0^1.

Let $x_0^1 = \min\{x \in E\colon Qg(x) = g(x)\}$. Then it follows from the above that $[x_0^1, \infty) \subseteq D_0^1$. Moreover, we can say that $[x_0^1, \infty) = D_0^1$.

Indeed, let $x = \lambda^k$ with $k \leqslant -1$. Then $Tg(x) = 0$, $Qg(x) = 0$, so that both instantaneous stopping at these points and continuation of observations (by one step) bring one no gains. For that reason we can ascribe the points $x = \lambda^k$ with $k \leqslant -1$ to the continuation domain C_0^1. Of course, this domain also contains $x = \lambda^0 = 1$ and the points $x = \lambda^k$ with $k > 1$ such that $\lambda^k < x_0^1$.

Thus, if $N = 1$, then

$$\tau_0^1 = \begin{cases} 0 & \text{for } S_0 \in [x_0^1, \infty), \\ 1 & \text{for } S_0 \in (0, x_0^1). \end{cases}$$

The analysis for $N = 2, 3, \dots$ proceeds in a similar way; the only difference is that while we considered the action of Q on $g(x)$ on the first step we shall now study its action on the functions $Qg(x), Q^2g(x), \dots$, each of them downwards convex (as is $g = g(x)$; see Fig. 58 below) and coinciding with $g(x)$ for large x. These properties mean that for each N there exist x_n^N such that $D_n^N = [x_n^N, \infty)$.

Not plunging any deeper into the detail of this simple analysis we note also that it is immediately clear for $N = 2$ that $D_1^2 = D_0^1$ and, therefore, $x_1^2 = x_0^1$. Considering the set $D_0^2 = \{x\colon Q^2g(x) = g(x)\} = \{x\colon g(x) \geqslant TQg(x)\}$ we see that $D_0^2 = [x_0^2, \infty)$ for some x_0^2. In addition, $0 = x_2^2 \leqslant x_1^2 \leqslant x_0^2$, so that τ_0^2 has the following structure:

$$\tau_0^2 = \begin{cases} 0 & \text{for } S_0 \in [x_0^2, \infty), \\ 1 & \text{for } S_0 \in (0, x_0^2),\ S_1 \in [x_1^2, \infty), \\ 2 & \text{for } S_0 \in (0, x_0^2),\ S_1 \in (0, x_1^2). \end{cases}$$

Fig. 57 and 58 below give one a clear notion of the structure of the stopping domains D_n^N and the continuation domains C_n^N and also of the functions $V_0^N(x) = Q^N g(x)$.

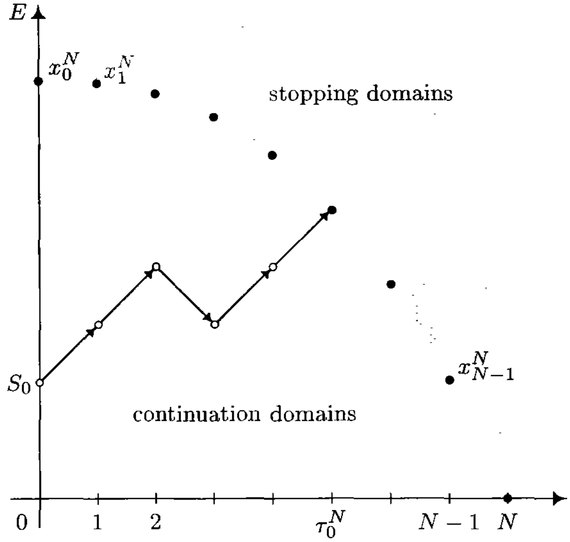


FIGURE 57. Call option. Stopping domains $D_0^N = [x_0^N, \infty), D_1^N = [x_1^N, \infty), \dots, D_N^N = [0, \infty)$, and continuation domains $C_0^N = (0, x_0^N), C_1^N = (0, x_1^N), \dots, C_N^N = \varnothing$. The trajectory $(S_0, S_1, S_2, \dots)$ leaves the continuation domains at time τ_0^N

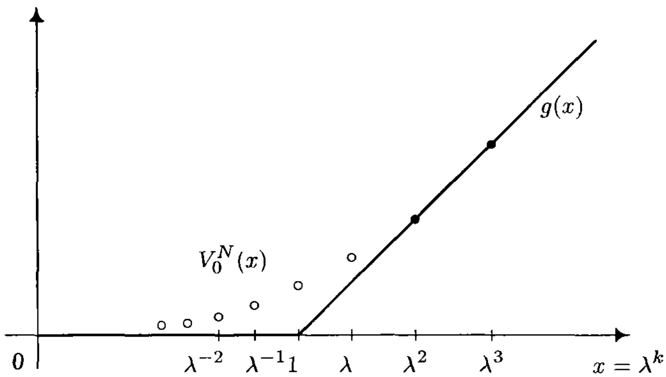


FIGURE 58. Graphs of the functions $g(x) = (x-1)^+$ and $V_0^N(x) = Q^N g(x)$ for a discounted call option with pay-offs $f_n = \beta^n g(x)$, $0 < \beta < 1$, $0 \leqslant n \leqslant N$, $\lambda > 1$

4. As follows from the above, finding the rational price $\widetilde{\mathbb{C}}_N(f;\mathsf{P})$ for $S_0 = x$ reduces to finding the functions $V_0^N(x) = Q^N g(x)$, which can be calculated recursively by the formula

$$\begin{aligned} Q^n g(x) &= \max\bigl(Q^{n-1}g(x), \alpha\beta T Q^{n-1} g(x)\bigr) \\ &= \max\bigl(g(x), \alpha\beta T Q^{n-1} g(x)\bigr). \end{aligned} \tag{17}$$

(See also [441; 2.2.1].)

Clearly, $V_0^N(x) \leqslant V_0^{N+1}(x)$, and therefore there exists

$$V^*(x) = \lim_{N\to\infty} V_0^N(x). \tag{18}$$

By Theorem 4 in Chapter V, § 6a the function $V^* = V^*(x)$ is the *smallest $\alpha\beta$-excessive majorant* of the function $g = g(x)$, i.e., $V^* = V^*(x)$ is the smallest function $U = U(x)$ such that $U(x) \geqslant g(x)$ and $U(x) \geqslant (\alpha\beta)TU(x)$. In addition, $V^* = V^*(x)$ satisfies the equation

$$V^*(x) = \max\bigl(g(x), (\alpha\beta)TV^*(x)\bigr), \tag{19}$$

following from (17) and (18).

By the same theorem $V(x)$ is just the solution of the optimal stopping problem in the class $\mathfrak{M}_0^\infty = \{\tau = \tau(\omega) \colon 0 \leqslant \tau(\omega) < \infty, \omega \in \Omega\}$, i.e.,

$$V^*(x) = \sup_{\tau\in\mathfrak{M}_0^\infty} \mathsf{E}_x(\alpha\beta)^\tau g(S_\tau). \tag{20}$$

The knowledge of $V^* = V^*(x)$ can be interesting also in the following respect: $V^*(S_0)$ is equal at the same time to the rational price

$$\widetilde{\mathbb{C}}_\infty(f;\mathsf{P}) = \inf\{y \colon \exists\, \pi \text{ with } X_0^\pi = y,\ X_\tau^\pi \geqslant \beta^\tau g(S_\tau),\ \forall\, \tau \in \mathfrak{M}_0^\infty\} \tag{21}$$

for the system $f = (f_n)$ of pay-off functions

$$f_n(x) = \beta^n (x-K)^+, \quad n \geqslant 0, \tag{22}$$

in the case when the buyer can choose *an arbitrary* stopping time τ in the set $\mathfrak{M}_0^\infty$ to exercise the contract. (The corresponding proof is similar to the proof of the theorem in § 1c.)

The analysis of options with exercise times in the set $\mathfrak{M}_0^\infty$ in place of the class $\mathfrak{M}_0^N$ with finite N can appear unappealing from the practical point of view. However, one should take into account that the discount factor β, $0 < \beta < 1$, does not allow optimal stopping times to be 'excessively large'. Incidentally, the *analytic* solution of problems of type (20) is much easier than the solution of problems of the type (2) with finite N; and if N is sufficiently large, then by $V^*(x)$ we can make a (probably, crude) guess about the values of the function $V_0^N(x)$.

5. We now proceed to the function $V^*(x)$, which, as we know, must satisfy (19).

Note that $D_0^N \supseteq D_0^{N+1}$, and therefore $x_0^N \leqslant x_0^{N+1}$. Hence three exists the limit $\lim_{N\to\infty} x_0^N = x^*$, and by (19) the function $V^*(x)$ must have the following form:

$$V^*(x) = \begin{cases} g(x), & x \geqslant x^*, \\ (\alpha\beta)TV^*(x), & x < x^*. \end{cases} \tag{23}$$

We point out that both 'boundary point' x^* and function $V^*(x)$ are *unknown* and must be found here. Problems of this sort are referred to as *free boundary* or *Stefan problems* (see, e.g., [441]).

In general, a solution $(x^*, V^*(x))$ to (23) is not necessarily unique and we may need some additional conditions to select the 'right' solution. We discuss below the arguments behind these additional conditions.

Let $C^* = (0, x^*)$ and let $D^* = [x^*, \infty)$. Then the function $V^*(x)$ in the domain C^* satisfies the equation

$$\varphi(x) = \alpha\beta T\varphi(x), \tag{24}$$

i.e. (in view of (4) in § 5a),

$$\varphi(x) = \alpha\beta\Big[p\varphi(\lambda x) + (1-p)\varphi\Big(\frac{x}{\lambda}\Big)\Big]. \tag{25}$$

In accordance with the general theory of difference equations (see, e.g., [174]) we shall seek the solutions of this equations in the form $\varphi(x) = x^\gamma$. Then γ must be a root of the equation

$$1 = \beta\big[\alpha p\lambda^\gamma + \alpha(1-p)\lambda^{-\gamma}\big]. \tag{26}$$

Recall that $p = \dfrac{r-a}{b-a}$, $b = \lambda - 1$, and $a = \lambda^{-1} - 1$. We have in fact found p from the condition

$$\mathsf{E}\frac{1+\rho_1}{1+r} = 1,$$

(see (4) in Chapter V, § 4d), i.e., the relation

$$\alpha\lambda p + \alpha(1-p)\lambda^{-1} = 1. \tag{27}$$

Comparing (26) and (27) we see that if $\beta = 1$, then (26) has the root $\gamma_1 = 1$ and another root γ_2 such that

$$\lambda^{\gamma_2} = \frac{1-p}{\lambda p}. \tag{28}$$

Since

$$\frac{1-p}{\lambda p} = \frac{\alpha\lambda - 1}{\lambda - \alpha} < 1,$$

it follows that $\gamma_2 < 0$.

Hence if $\beta = 1$, then the general solution to (26) has the following form:

$$\varphi(x) = c_1 x + c_2 x^{\gamma_2}, \tag{29}$$

where $\gamma_2 < 0$.

By the nature of the problem the required function $V^*(x)$ must be nonnegative and nondecreasing. Hence $c_2 = 0$, and therefore

$$V^*(x) = \begin{cases} x - 1, & x \geqslant x^*, \\ c_1 x, & x < x^*, \end{cases} \tag{30}$$

where x^* and c_1 are still to be defined.

By the submartingale property of the sequence $(\alpha^n (S_n - 1)^+)_{n \geqslant 0}$ with respect to each measure P_x, $x \in E$, we obtain that $x^* = \infty$, because if $\beta = 1$ for each point $x \geqslant 1$, then

$$\alpha T g(x) > g(x),$$

and therefore it would certainly be 'more advantageous' to make at least one observation than to stop immediately.

Further, $c_1 \geqslant 1$ in (30), for if $c_1 < 1$, then $x^* < \infty$.

On the other hand, c_1 cannot be larger than 1 in view of the (additional) property that $V^*(x)$ is the *smallest* α-excessive majorant of $g(x)$, and the smallest function $c_1 x$ with $c_1 \geqslant 1$ clearly has the coefficient $c_1 = 1$.

Thus, for $\beta = 1$ and $g(x) = (x-1)^+$ we have $V^*(x) = \sup\limits_{\tau \in \mathfrak{M}_0^\infty} \mathsf{E}_x \alpha^\tau g(S_\tau) = x$, and there exists no optimal stopping time (in the class $\mathfrak{M}_0^\infty$). However, for each $\varepsilon > 0$ and each $x \in E$ we can find a finite stopping time $\tau_{x,\varepsilon}$ such that

$$\mathsf{E}_x \alpha^{\tau_{x,\varepsilon}} g(S_{\tau_{x,\varepsilon}}) \geqslant V^*(x) - \varepsilon.$$

(See [441; Chapter 3] for greater detail.)

6. Assume now that $0 < \beta < 1$. Then equation (26) has two roots, $\gamma_1 > 1$ and $\gamma_2 < 0$, such that the quantities $y_1 = \lambda^{\gamma_1}$ and $y_2 = \lambda^{\gamma_2}$ that are the solutions of the quadratic equation

$$y = \beta\big[\alpha p y^2 + \alpha(1-p)\big] \tag{31}$$

can be expressed as follows:

$$y_1 = \frac{A}{2} + \sqrt{\frac{A^2}{4} - B}, \qquad y_2 = \frac{A}{2} - \sqrt{\frac{A^2}{4} - B}, \tag{32}$$

where $A = (\alpha\beta p)^{-1}$ and $B = (1-p)p^{-1}$.

Thus, if $0 < \beta < 1$, then the general solution $\varphi(x)$ of (25) can be represented as the sum:

$$\varphi(x) = c_1 x^{\gamma_1} + c_2 x^{\gamma_2}. \tag{33}$$

For the same reasons as in the case $\beta = 1$, the coefficient c_2 must here be equal to zero and it follows from (23) that the required function is

$$V^*(x) = \begin{cases} x - 1, & x \geqslant x^*, \\ c^* x^{\gamma_1}, & x < x^*, \end{cases} \tag{34}$$

where x^* and c^* are constants to be defined (see (40)–(43) below.)

To find x^* and c^* we use the observation that $V^*(x)$ must be the *smallest $\alpha\beta$-excessive majorant* of the function $g(x) = (x-1)^+$, $x \in E$.

The following reasoning shows how, in the class of functions

$$\overline{V}_{\overline{c}}(x; \overline{x}) = \begin{cases} x - 1, & x \geqslant \overline{x}, \\ \overline{c} x^{\gamma_1}, & x < \overline{x} \end{cases} \tag{35}$$

with $x, \overline{x} \in E$, $\overline{c} > 0$, and $\gamma_1 > 1$, one can find the *smallest majorant* of the function $g(x) = (x-1)^+$. (Then, of course, we must verify that the function so obtained is *$\alpha\beta$-excessive.*)

To this end we point out that for sufficiently large $\overline{c}$ the function $\varphi_{\overline{c}}(x) = \overline{c} x^{\gamma_1}$ is knowingly larger than $g(x)$ for all $x \in E$. Hence it is clear from (35) how one can find the smallest majorant $g(x)$ among the functions $\overline{V}_{\overline{c}}(x; \overline{x})$

We now choose $\overline{c}$ sufficiently large so that $\varphi_{\overline{c}}(x) > g(x)$ for all $x \in E$, and then make $\overline{c}$ smaller until, for some value of $\overline{c}_1$, the function $\varphi_{\overline{c}_1}(x)$ 'meets' $g(x)$ at some point $\overline{x}_1$.

The functions $\varphi_{\overline{c}}(x)$, $x \in E$, are convex, therefore, in principle, there can exist another point, $\overline{x}_2 \in E$, such that $\overline{x}_2 > \overline{x}_1$ and $\varphi_{\overline{c}_1}(\overline{x}_2) = g(\overline{x}_2)$.

In our case the phase space $E = \{x = \lambda^k, k = 0, \pm 1, \dots\}$ is discrete. However, if we assume that $\lambda = 1 + \Delta$, where $\Delta > 0$ is *small*, then the distance between $\overline{x}_1$ and $\overline{x}_2$ is also small and, moreover, these points 'merge' into one point, $\widetilde{x}$, as $\Delta \downarrow 0$.

Clearly, $\widetilde{x}$ is precisely the point in the interval $(0, \infty)$ where the graph of $\varphi_{\widetilde{c}}(x) = \widetilde{c} x^{\gamma_1}$ for some value of $\widetilde{c}$ *touches* the graph of $g(x) = (x-1)^+$, $x \in (0, \infty)$.

Obviously, $\widetilde{c}$ and $\widetilde{x}$ can be defined from the system of two equations

$$\varphi_{\widetilde{c}}(\widetilde{x}) = g(\widetilde{x}), \tag{36}$$

$$\left.\frac{d\varphi_{\widetilde{c}}(x)}{dx}\right|_{x=\widetilde{x}-} = \left.\frac{dg(x)}{dx}\right|_{x=\widetilde{x}+}, \tag{37}$$

which yields

$$\widetilde{x} = \frac{\gamma_1}{\gamma_1 - 1}, \qquad \widetilde{c} = \frac{(\gamma_1 - 1)^{\gamma_1 - 1}}{\gamma_1^{\gamma_1}}. \tag{38}$$

Moreover, it is clear that the function

$$\widetilde{V}(x) = \begin{cases} x - 1, & x \geqslant \widetilde{x}, \\ \widetilde{c} x^{\gamma_1}, & x < \widetilde{x}, \end{cases} \tag{39}$$

is a good approximation to the least functions of the form (35) if $\Delta > 0$ is sufficiently small. (Cf. formula (37) in Chapter VIII, § 2a).

Remark. We must point out condition (37) of '*smooth fitting*', which enters our discussion in a fairly natural way. This condition often plays the role of an *additional* requirement in optimal stopping problems and enables one to single out the 'right' solution of a problem (see [441] and Chapter VIII in the present book.)

The above-described *qualitative* method of finding the smallest majorant of $g(x)$ brings one after more minute analysis (see [443]) to the following 'optimal' values x^* and c^* of $\overline{x}$ and $\overline{c}$ ensuring that the corresponding function $V^*(x) = \overline{V}_{c^*}(x; x^*)$ is not only the smallest majorant of $g(x)$, but also an $\alpha\beta$-*excessive* majorant:

$$c^* = \min(c_1^*, c_2^*), \tag{40}$$

where

$$c_1^* = \left(\lambda^{[\log_\lambda \widetilde{x}]} - 1\right)\lambda^{-\gamma_1[\log_\lambda \widetilde{x}]}, \tag{41}$$

$$c_2^* = \left(\lambda^{[\log_\lambda \widetilde{x}]} - 1\right)\lambda^{-\gamma_1[\log_\lambda \widetilde{x}]-\gamma_1}, \tag{42}$$

and

$$x^* = \begin{cases} \lambda^{[\log_\lambda \widetilde{x}]}, & \text{if } c^* = c_1^*, \\ \lambda^{[\log_\lambda \widetilde{x}]+1}, & \text{if } c^* = c_2^* \end{cases} \tag{43}$$

(here $[y]$ is the integer part of y and $\widetilde{x}$ is defined in (38)).

It is obvious that the function $V^*(x)$ so obtained is $\alpha\beta$-excessive for $x < x^*$ because, by construction, $\alpha\beta TV^*(x) = V^*(x)$ for such x. On the other hand, if $x \geqslant x^*$, then we can verify directly that $\alpha\beta TV^*(x) \leqslant V^*(x)$ once we take into account (40)–(43) and the fact that $V^*(x) = x - 1$ for such x.

7. By Theorem 4 in § 2a our function $V^*(x)$ is precisely equal to the supremum $\sup\limits_{\tau \in \mathfrak{M}_0^\infty} \mathsf{E}_x(\alpha\beta)^\tau g(S_\tau)$, and the instant

$$\tau^* = \inf\{n \colon V^*(S_n) = g(S_n)\} = \inf\{n \colon S_n \geqslant x^*\}$$

is an optimal stopping time, provided that $\mathsf{P}_x(\tau^* < \infty) = 1$, $x \in E$.

Clearly,

$$\begin{aligned} \mathsf{P}_x(\tau^* > N) &= \mathsf{P}_x\Big(\max_{n \leqslant N} S_n < x^*\Big) \\ &= \mathsf{P}_x\Big(S_0 \max_{n \leqslant N} \lambda^{\varepsilon_1 + \cdots + \varepsilon_n} < x^*\Big) \end{aligned} \tag{44}$$

and since $\mathsf{P}(\varepsilon_i = 1) = p$ and $\mathsf{P}(\varepsilon_i = -1) = q$, the probability on the right-hand side converges to zero as $N \to \infty$ for $p \geqslant q$.

By (5) the inequality $p \geqslant q$ is equivalent to the relation

$$r \geqslant \frac{a+b}{2}. \tag{45}$$

Bearing in mind that $b = \lambda - 1$ and $a = \lambda^{-1} - 1$ we see that $\mathsf{P}_x(\tau^* < \infty) = 1$ for each $x < x^*$, provided that

$$r \geqslant \frac{\lambda + \lambda^{-1}}{2} - 1. \tag{46}$$

On the other hand, if $x \geqslant x^*$, then $\mathsf{P}_x(\tau^* = 0) = 1$ without regard to (46).

Summing up we arrive at the following result.

THEOREM 2. *Assume that* $0 < \beta < 1$ *and that* (46) *holds. Then the rational price* $\widetilde{\mathbb{C}}_\infty(f; \mathsf{P})$ *of an American call option with pay-offs* $f_n = \beta^n (S_n - 1)^+$, $n \geqslant 0$, *is described by the formula*

$$\widetilde{\mathbb{C}}_\infty(f; \mathsf{P}) = V^*(S_0),$$

where

$$V^*(S_0) = \begin{cases} S_0 - 1, & S_0 \geqslant x^*, \\ c^* S_0^{\gamma_1}, & S_0 < x^*, \end{cases}$$

and the constants c^* *and* x^* *can be found by* (40)–(43). *The optimal time for exercising the option is* $\tau^* = \inf\{n \colon S_n \geqslant x^*\}$. *In addition,*

$$V^*(S_0) = \mathsf{E}_{S_0} (\alpha\beta)^{\tau^*} (S_{\tau^*} - 1)^+.$$

§ 5c. Standard Put Option Pricing

1. The pay-off functions for a standard *put option* are as follows:

$$f_n(y) = \beta^n (K - y)^+, \qquad x \in E, \tag{1}$$

where $0 < \beta \leqslant 1$, $E = \{y = \lambda^k \colon k = 0, \pm 1, \dots\}$, $\lambda > 1$.

By analogy with the preceding section, we set

$$V_n^N(y) = \sup_{\tau \in \mathfrak{M}_n^N} \mathsf{E}_y (\alpha\beta)^\tau (K - S_\tau)^+ \tag{2}$$

and

$$V^*(y) = \sup_{\tau \in \mathfrak{M}_0^\infty} \mathsf{E}_y (\alpha\beta)^\tau (K - S_\tau)^+. \tag{3}$$

These values are of interest because

$$V_0^N(y) = \widetilde{\mathbb{C}}_N(f \colon \mathsf{P}), \qquad y = S_0, \tag{4}$$

and

$$V^*(y) = \widetilde{\mathbb{C}}_\infty(f; \mathsf{P}), \qquad y = S_0, \tag{5}$$

where the prices $\widetilde{\mathbb{C}}_N(f; \mathsf{P})$ and $\widetilde{\mathbb{C}}_\infty(f; \mathsf{P})$ for the system $f = (f_n)_{n \geqslant 0}$ of functions $f_n = f_n(y)$ defined by (1) are as in formula (7) in § 5a and formula (21) in § 5b, respectively.

THEOREM 1. *For each $N \geqslant 0$ there exists a sequence y_n^N, $0 \leqslant n \leqslant N$, with values in $E \cup \{+\infty\}$ such that*

$$D_n^N = \{y \in E\colon y \in (0, y_n^N]\}, \tag{6}$$

$$C_n^N = \{y \in E\colon y \in (y_n^N, \infty)\} \tag{7}$$

and

$$\begin{aligned}\tau_0^N &= \min\{0 \leqslant n \leqslant N\colon S_n \in D_n^N\}\\ &= \min\{0 \leqslant n \leqslant N\colon S_n \in (0, y_n^N]\}.\end{aligned}$$

Moreover,

$$y_0^N \leqslant \cdots \leqslant y_{N-1}^N \leqslant y_N^N = \infty \tag{8}$$

and

$$V_0^N(y) = \begin{cases} g(y), & y \in D_0^N = (0, y_0^N], \\ Q^N g(y), & y \in C_0^N = (y_0^N, \infty). \end{cases} \tag{9}$$

The rational price $\widetilde{\mathbb{C}}_N(f; \mathsf{P})$ is equal to $V_0^N(S_0)$.

Proof. This is similar to the case of call options discussed in § 5b; the proof is based on the analysis of the subset of points $y \in E$ at which the operators Q^n increase the value of the function $g(y)$.

It is worth noting that, of course, the operator Q raises the value of $g(y)$ at the point $y = K$ (we have assumed above for simplicity that $K = 1$) and $Qg(y) = g(y) = 0$ for $y > K$. Hence these values of $y \in E$ can be ascribed both to the stopping domains and to the continuation domains. As seen from (6) and (7), we have actually put these points in the continuation domains.

2. We consider now the question of finding the function $V^*(y)$ $\Bigl(= \lim\limits_{N\to\infty} V_0^N(y)\Bigr)$, the quantity $y^* = \lim\limits_{N\to\infty} y_0^N$, and the optimal time τ^* such that

$$V^*(y) = \mathsf{E}_y(\alpha\beta)^{\tau^*}(K - S_{\tau^*})^+ \tag{10}$$

(for simplicity we set $K = 1$ again).

Let $C^* = (y^*, \infty)$ and let $D^* = (0, y^*]$. As in § 5b, we see that the function $V^*(y)$ in the domain C^* is a solution of the equation

$$\varphi(y) = \alpha\beta\Bigl[p\varphi(\lambda y) + (1-p)\varphi\Bigl(\frac{y}{\lambda}\Bigr)\Bigr].$$

Its general solution is $c_1 y^{\gamma_1} + c_2 y^{\gamma_2}$ where $\gamma_1 > 1$ and $\gamma_2 < 0$ (see (31) and (32) in § 5b).

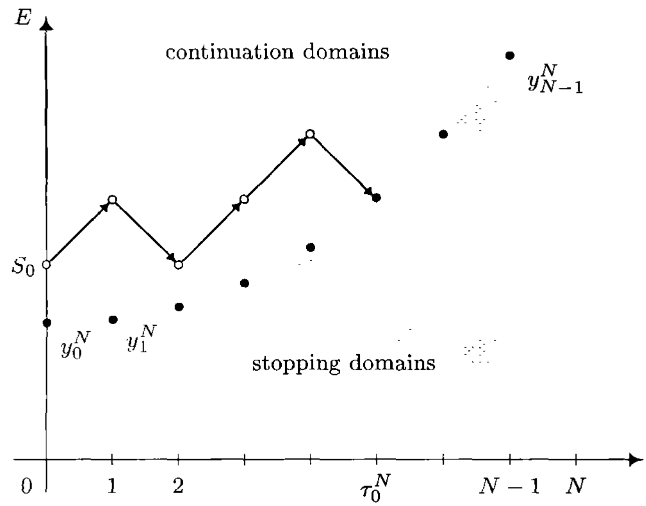


FIGURE 59. Put option. Stopping domains $D_0^N = (0, y_0^N], \dots, D_{N-1}^N = (0, y_{N-1}^N], D_N^N = (0, \infty)$, and continuation domains $C_0^N = (y_0^N, \infty), \dots, C_{N-1}^N = (y_{N-1}^N, \infty), C_N^N = \varnothing$. The trajectory $(S_0, S_1, S_2, \dots)$ leaves the continuation domains at time τ_0^N

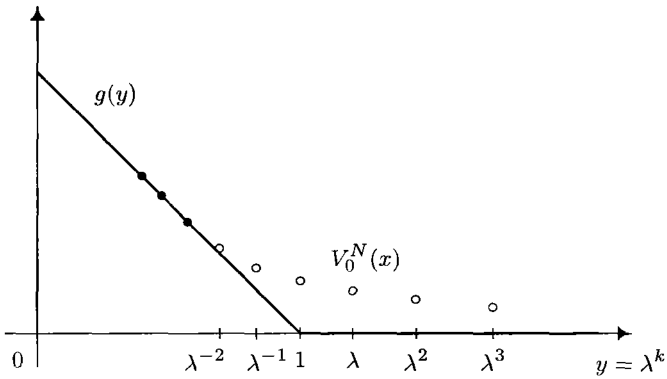


FIGURE 60. Graphs of the functions $g(y) = (1 - y)^+$ and $V_0^N(y) = Q^N g(y)$ for a put option with pay-offs $f_n = \beta^n g(y)$, where $0 < \beta \leqslant 1$, $0 \leqslant n \leqslant N$, and $\lambda > 1$

Since $V^*(y) \leqslant 1$, it follows that $c_1 = 0$, so that we must seek $V^*(y)$ in the class of functions

$$\overline{V}_{\overline{c}}(y;\overline{y}) = \begin{cases} 1-y, & y \leqslant \overline{y}, \\ \overline{c}y^{\gamma_2}, & y > \overline{y}, \end{cases} \tag{11}$$

where the 'optimal' values c^* and y^* of $\overline{c}$ and $\overline{y}$ are to be determined on the basis of the above-mentioned (§ 5b.6) additional conditions that $V^*(y) = \overline{V}_{c^*}(y;y^*)$ must be the *smallest $\alpha\beta$-excessive majorant of* $g(y) = (1-y)^+$.

Following the scheme (exposed in § 5b) of finding, for small $\Delta = 1 - \lambda > 0$, *approximations* $\widetilde{c}$ and $\widetilde{y}$ to the parameters c^* and y^*, we see that they can be determined from the system of equations

$$\varphi_{\widetilde{c}}(\widetilde{y}) = g(\widetilde{y}),$$

$$\left.\frac{d\varphi_{\widetilde{c}}(y)}{dy}\right|_{y=\widetilde{y}+} = \left.\frac{dg(y)}{dy}\right|_{y=\widetilde{y}-}. \tag{12}$$

Solving it we obtain

$$\widetilde{y} = \left|\frac{\gamma_2}{\gamma_2 - 1}\right|, \qquad \widetilde{c} = \frac{|\gamma_2|^{|\gamma_2|}}{|\gamma_2 - 1|^{|\gamma_2 - 1|}}. \tag{13}$$

Once we know the values $\widetilde{y}$ and $\widetilde{c}$ corresponding to the 'limiting' case ($\lambda \downarrow 1$) we can find (see [443]) the values of y^* and c^* in the initial, 'prelimit' scheme (with $\lambda > 1$) by the formulas

$$c^* = \min(c_1^*, c_2^*), \tag{14}$$

where

$$c_1^* = \left(1 - \lambda^{[\log_\lambda \widetilde{y}]}\right)\lambda^{-\gamma_2[\log_\lambda \widetilde{y}]}, \tag{15}$$

$$c_2^* = \left(1 - \lambda^{[\log_\lambda \widetilde{y}]+1}\right)\lambda^{-\gamma_2[\log_\lambda \widetilde{y}]-\gamma_2} \tag{16}$$

and

$$y^* = \begin{cases} \lambda^{[\log_\lambda \widetilde{y}]}, & \text{if } c^* = c_1^*, \\ \lambda^{[\log_\lambda \widetilde{y}]+1}, & \text{if } c^* = c_2^*. \end{cases} \tag{17}$$

The fact that the so-obtained smallest majorant $V^*(y)$ of $g(y) = (1-y)^+$ is $\alpha\beta$-excessive can be established by an immediate verification.

We finally observe that the condition

$$r \leqslant \frac{a+b}{2} = \frac{\lambda + \lambda^{-1}}{2} - 1 \tag{18}$$

(cf. (45) in § 5b) ensures that $\mathsf{P}_y(\tau^* < \infty) = 1$ for $y \in E$ and $\tau^* = \inf\{n\colon S_n \leqslant y^*\}$. (If $y \leqslant y^*$, then $\mathsf{P}_y(\tau^* = 0) = 1$.)

Thus, if condition (14) holds, then τ^* is optimal in the following sense: property (10) holds for all $y \in E$.

THEOREM 2. *Assume that $0 < \beta \leqslant 1$ and condition (18) holds. Then the rational price $\widetilde{\mathbb{C}}_\infty(f; \mathsf{P})$ of an American put option with pay-offs $f_n = \beta^n(1 - S_n)^+$, $n \geqslant 0$, is defined by the formula*

$$\widetilde{\mathbb{C}}_\infty(f; \mathsf{P}) = V^*(S_0), \tag{19}$$

where

$$V^*(S_0) = \begin{cases} 1 - S_0, & S_0 \leqslant y^*, \\ c^* S_0^{\gamma_2}, & S_0 > y^*, \end{cases} \tag{20}$$

and the parameters c^* and y^* can be found by (14)–(17). The optimal time of exercising the option is $\tau^* = \inf\{n : S_n \leqslant y^*\}$. Moreover,

$$V^*(S_0) = \mathsf{E}_{S_0}(\alpha\beta)^{\tau^*}(1 - S_{\tau^*})^+.$$

§ 5d. Options with Aftereffect. 'Russian Option' Pricing

1. The pay-offs f_n for the put and call options considered above have the *Markovian* structure:

$$f_n = \beta^n(S_n - K)^+ \quad \text{and} \quad f_n = \beta^n(K - S_n)^+, \tag{1}$$

respectively.

It is of interest both for theory and financial engineering to consider also options with *aftereffect*. Examples here can be options with the following pay-off functions:

$$f_n = \beta^n \Big(aS_n - \min_{0 \leqslant r \leqslant n} S_r\Big)^+, \tag{2}$$

$$f_n = \beta^n \Big(\max_{0 \leqslant r \leqslant n} S_r - aS_n\Big)^+ \tag{3}$$

or with pay-offs

$$f_n = \beta^n \left(aS_n - \sum_{k=0}^{n} S_k\right)^+, \tag{4}$$

$$f_n = \beta^n \left(\sum_{k=0}^{n} S_k - aS_n\right)^+ \tag{5}$$

where $0 \leqslant \beta \leqslant 1$, $a \geqslant 0$.

Options with pay-offs (4) and (5) are called *Asian* (call and put) options. Call and put options with pay-offs (2) and (3) have been considered for $a = 0$ in [434] and [435], where they are called '*Russian options*' (see also [118] and [283]). In what follows we stick to [283].

2. We consider the *CRR*-model in which ρ_n can take two values, $\lambda - 1$ and $\lambda^{-1} - 1$, where $\lambda > 1$. In addition, for definiteness, we shall consider an American *put option* with pay-off function (3), where β, $0 < \beta < 1$, is the discount factor.

In accordance with the general theory (see subsection 2), the rational price $\widehat{\mathbb{C}}$ of such an option can be found by the formula

$$\widehat{\mathbb{C}} = \sup_{\tau \in \mathfrak{M}_0^\infty} \mathsf{E} \alpha^\tau f_\tau, \tag{6}$$

where $\alpha = (1+r)^{-1}$ and E is averaging with respect to the martingale measure P such that p and q are described by formula (6) in § 5a.

Since

$$\widehat{\mathbb{C}} = \sup_{\tau \in \mathfrak{M}_0^\infty} \mathsf{E}(\alpha\beta)^\tau \Big(\max_{0 \leqslant r \leqslant \tau} S_r - a S_\tau \Big)^+ \tag{7}$$

and $S_n = S_0 \lambda^{\varepsilon_1 + \dots + \varepsilon_n}$, the quantity $\widehat{\mathbb{C}}$ is definitely finite ($\widehat{\mathbb{C}} \leqslant S_0$) if

$$\alpha\beta\lambda \leqslant 1. \tag{8}$$

We set $Y_n = \max\limits_{k \leqslant n} S_k$. Clearly,

$$Y_n = \max\{Y_{n-1}, S_n\}. \tag{9}$$

Moreover, $(S_n, Y_n)_{n \geqslant 0}$ is a *Markov sequence* and, in principle, one can solve the optimal stopping problem (7) on the basis of general results on optimal stopping rules for two-dimensional Markov chains. (see [441] and § 2a).

However—and this is remarkable—our *two-dimensional* Markov problem can be reduced to some *one-dimensional* Markov problem if one uses the idea of change of measures and chooses an appropriate discounting asset (numéraire). (See also Chapter VII, § 1b on this subject.)

Let $\tau \in \mathfrak{M}_0^\infty$. We recall that $B_n = B_0 \alpha^{-n}$ with $\alpha = (1+r)^{-1}$, so that

$$\begin{aligned} \mathsf{E}(\alpha\beta)^\tau \Big(\max_{0 \leqslant r \leqslant \tau} S_r - a S_\tau \Big)^+ &= \mathsf{E}(\alpha\beta)^\tau \left(\frac{\max_{0 \leqslant r \leqslant \tau} S_r}{S_\tau} - a \right)^+ S_\tau \\ &= S_0 \mathsf{E} \left[\beta^\tau \left(\frac{Y_\tau}{S_\tau} - a \right)^+ \cdot \frac{S_\tau / S_0}{B_\tau / B_0} \right]. \end{aligned} \tag{10}$$

We set $Z_n = \dfrac{S_n/S_0}{B_n/B_0}$. Then $Z_n > 0$, and the sequence $Z = (Z_n, \mathscr{F}_n, \mathsf{P})_{n \geqslant 0}$ is a P-martingale with $\mathsf{E} Z_n = 1$.

For $A \in \mathscr{F}_n$ we set

$$\widehat{\mathsf{P}}_n(A) = \mathsf{E}(Z_n I_A).$$

Clearly, the measures in the collection $(\widehat{\mathsf{P}}_n)_{n\geqslant 0}$ are *compatible* (i.e., $\widehat{\mathsf{P}}_{n+1} \,|\, \mathscr{F}_n = \widehat{\mathsf{P}}_n$, $n \geqslant 0$) and by Ionescu Tulcea's theorem on the extension of a measure (see, e.g., [439; Chapter II, § 9]) there exists a measure $\widehat{\mathsf{P}}$ (in the space $\Omega = \{-1, 1\}^\infty$) such that $\widehat{\mathsf{P}} \,|\, \mathscr{F}_n = \widehat{\mathsf{P}}_n$, $n \geqslant 0$.

Hence

$$\mathsf{E}(\alpha\beta)^\tau \Big(\max_{0\leqslant r\leqslant \tau} S_r - aS_\tau\Big)^+ = S_0 \widehat{\mathsf{E}} \beta^\tau \left(\frac{Y_\tau}{S_\tau} - a\right)^+. \tag{11}$$

We now set

$$X_n = \frac{Y_n}{S_n} \tag{12}$$

and observe that

$$X_{n+1} = \max\left(\frac{X_n}{\lambda^{\varepsilon_{n+1}}}, 1\right) \tag{13}$$

and, moreover, all X_n range in the set $\widehat{E} = \{1, \lambda, \lambda^2, \dots\}$.

With respect to the new measure $\widehat{\mathsf{P}}$ the sequence $\varepsilon = (\varepsilon_n)_{n\geqslant 1}$ is also a sequence of *independent identically distributed* (i.i.d.) random variables with

$$\widehat{p} = \widehat{\mathsf{P}}(\varepsilon_n = 1) = \mathsf{E} I_{(\varepsilon_n = 1)} \alpha \lambda^{\varepsilon_1} = \alpha\lambda p \tag{14}$$

and

$$\widehat{q} = \widehat{\mathsf{P}}(\varepsilon_n = -1) = \frac{\alpha}{\lambda}(1 - p). \tag{15}$$

We shall consider the sequence $(X_n)_{n\geqslant 0}$ defined by recursive relations (13) under the assumption that $X_0 = x \in \widehat{E}$. Let $\widehat{\mathsf{P}}_x$ be the probability distribution of this sequence. Then $X = (X_n, \widehat{\mathscr{F}}_n, \widehat{\mathsf{P}}_x)$ with $x \in \widehat{E}$ and $\widehat{\mathscr{F}}_n = \sigma(X_0, X_1, \dots, X_n)$, $n \geqslant 0$, is a Markov sequence and, therefore, to find the price $\widehat{\mathbb{C}}$ one must consider the optimal stopping problem

$$\widehat{V}(x) = \sup_{\tau \in \widehat{\mathfrak{M}}_0^\infty} \widehat{\mathsf{E}}_x \beta^\tau (X_\tau - a)^+, \qquad x \in \widehat{E}, \tag{16}$$

where $\widehat{\mathfrak{M}}_0^\infty$ is the class of finite stopping times $\tau = \tau(\omega)$ such that $\{\omega \colon \tau(\omega) \leqslant n\} \in \widehat{\mathscr{F}}_n$, $n \geqslant 0$.

The price $\widehat{\mathbb{C}}$ in question is connected with the solution $\widehat{V}(1)$ of this problem by the formula

$$\widehat{\mathbb{C}} = S_0 \widehat{V}(1). \tag{17}$$

Remark 1. Strictly speaking, the supremum in (7) is taken over the class $\mathfrak{M}_0^\infty$ and therefore formula (17) is valid if, in the definition (16) of $\widehat{V}(x)$, we consider the supremum over the wider class $\mathfrak{M}_0^\infty$ in place of $\widehat{\mathfrak{M}}_0^\infty$. However, these two suprema are the same, which follows from the general theory of optimal stopping rules for Markov sequences (see [441]) and is in effect proved below (see Remark 2).

3. Let $g(x) = (x-a)^+$, $x \in \widehat{E}$, and let

$$\widehat{V}_0^N(x) = \sup_{\tau \in \widehat{\mathfrak{M}}_0^N} \widehat{\mathsf{E}}_x \beta^\tau g(X_\tau),$$

where $\widehat{\mathfrak{M}}_0^N$ is the class of stopping times τ in $\widehat{\mathfrak{M}}_0^\infty$ such that $\tau(\omega) \leqslant N$, $\omega \in \Omega$ (see Fig. 61).

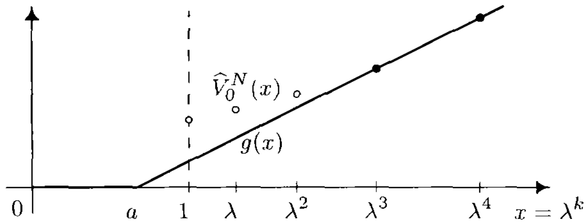


FIGURE 61. Graphs of functions $g(x) = (x-a)^+$ and $\widehat{V}_0^N(x) = \widehat{Q}^N g(x)$ for $0 < a < 1$

We also set

$$\widehat{T}f(x) = \widehat{\mathsf{E}}_x f(X_1) = \widehat{p} f\Big(\frac{x}{\lambda} \wedge 1\Big) + (1-\widehat{p}) f(\lambda x) \tag{18}$$

and

$$\widehat{Q}f(x) = \max\big(f(x), \beta \widehat{T}f(x)\big). \tag{19}$$

By Theorem 3 in § 2a and our remark upon it,

$$\widehat{V}_0^N(x) = \widehat{Q}^N g(x) \tag{20}$$

and the optimal stopping time $\widehat{\tau}_0^N \in \widehat{\mathfrak{M}}_0^N$ can be described as follows (cf. (9) in § 5b):

$$\widehat{\tau}_0^N = \min\{0 \leqslant n \leqslant N : X_n \in \widehat{D}_n^N\}, \tag{21}$$

where

$$\widehat{D}_n^N = \{x \in \widehat{E} \colon \widehat{V}_0^{N-n}(x) = g(x)\}. \tag{22}$$

Clearly, $\widehat{D}_0^N \subseteq \widehat{D}_1^N \subseteq \cdots \subseteq \widehat{D}_N^N = \widehat{E} \equiv \{1, \lambda, \lambda^2, \dots\}$.

In the same way as in § 5b, considering successively the functions $\widehat{Q}g(x), \dots, \widehat{Q}^N g(x)$ and comparing them with $g(x)$ we see that the stopping domains $\widehat{D}_n^N$ are as follows:

$$\widehat{D}_n^N = \{x \in \widehat{E} \colon x \in [\widehat{x}_n^N, \infty)\}, \tag{23}$$

where

$$1 = \widehat{x}_N^N \leqslant \widehat{x}_{N-1}^N \leqslant \cdots \leqslant \widehat{x}_0^N. \tag{24}$$

The qualitative picture of the stopping domains $\widehat{D}_n^N$ and the continuation domains $\widehat{C}_n^N = \widehat{E} \setminus \widehat{D}_n^N$ is as in Fig. 57 in § 5b (with the self-evident change of notation $S_i \to X_i$ and $E \to \widehat{E}, \dots, x_N^N = 0 \to \widehat{x}_N^N = 1$).

Remark 2. If

$$V_0^N(x) = \sup_{\tau \in \mathfrak{M}_0^N} \widehat{\mathsf{E}}_x \beta^\tau g(X_\tau),$$

then it follows by Theorem 3 in § 2a that $V_0^N(x) = \widehat{Q}^N g(x)$. Comparing this with (20) we see that $\widehat{V}_0^N(x) = V_0^N(x)$, $x \in \widehat{E}$, and that the instant $\widehat{\tau}_0^N$ defined by (21) is optimal not only in the class $\widehat{\mathfrak{M}}_0^N$, but also in the broader class $\mathfrak{M}_0^N$.

4. Since $g(x) \geqslant 0$, it follows by Theorem 4 in § 2b that $\widehat{V}(x) = \lim_{N \to \infty} \widehat{V}_0^N(x)$. We now set

$$\widehat{\tau} = \inf\{n \geqslant 0 \colon \widehat{V}(X_n) = g(X_n)\} = \inf\{n \geqslant 0 \colon X_n \in \widehat{D}\},$$

where $\widehat{D} = \{x \in E \colon x \in [x, \infty)\}$ and $\widehat{x} = \lim_{N \to \infty} \widehat{x}_0^N$.

By the same theorem, $\widehat{\tau}$ is an optimal stopping time in the problem (16), provided that $\widehat{\mathsf{P}}_x(\widehat{\tau} < \infty) = 1$, $x \in \widehat{E}$. Leaving aside for the moment this property of $\widehat{\tau}$ we proceed to $\widehat{x}$ and $\widehat{V}(x)$.

The function $\widehat{V}(x)$ satisfies the equation

$$\widehat{V}(x) = \max\big(g(x), \beta \widehat{T} \widehat{V}(x)\big), \qquad x \in \widehat{E}, \tag{25}$$

therefore, in the continuation domain $\widehat{C} = \widehat{E} \setminus \widehat{D}$ it is a solution of the equation

$$\varphi(x) = \beta \widehat{T} \varphi(x), \qquad x \in \widehat{C}, \tag{26}$$

or, in more explicitly,

$$\varphi(x) = \beta\Big[\widehat{p}\,\varphi\Big(\frac{x}{\lambda} \vee 1\Big) + (1 - \widehat{p})\varphi(\lambda x)\Big], \qquad x \in \widehat{C}. \tag{27}$$

In particular, for $x = 1$ we have

$$\varphi(1) = \beta\big[\widehat{p}\,\varphi(1) + (1 - \widehat{p})\varphi(\lambda)\big], \tag{28}$$

while for $x \geqslant \lambda$,

$$\varphi(x) = \beta\Big[\widehat{p}\,\varphi\Big(\frac{x}{\lambda}\Big) + (1 - \widehat{p})\varphi(\lambda x)\Big]. \tag{29}$$

It would be natural to seek a solution of (29) in the form of a power function x^γ (cf. § 5b.5). Then we obtain for γ the equation

$$\frac{1}{\beta} = \widehat{p}\,\lambda^{-\gamma} + (1-\widehat{p}\,)\lambda^{\gamma}, \tag{30}$$

with two solutions, $\gamma_1 < 0$ and $\gamma_2 > 1$, such that the quantities $y_1 = \lambda^{\gamma_1}$ and $y_2 = \lambda^{\gamma_2}$ are defined by the formulas

$$y_1 = \frac{A}{2} - \sqrt{\frac{A^2}{4} - B}, \qquad y_2 = \frac{A}{2} + \sqrt{\frac{A^2}{4} - B} \tag{31}$$

with

$$A = \frac{1}{(1-\widehat{p}\,)\beta} \quad \text{and} \quad B = \frac{\widehat{p}}{1-\widehat{p}}\,. \tag{32}$$

For $x \geqslant \lambda$ the general solution $\varphi(x)$ of (29) can be represented as $c\psi_b(x)$, where $\psi_b(x) = bx^{\gamma_1} + (1-b)x^{\gamma_2}$.

Since $\psi_b(1) = 1$, it follows that $c = \varphi(1)$.

Substituting $\varphi(\lambda) = \varphi(1)\psi_b(\lambda)$ in (28) and bearing in mind that $\varphi(1) \neq 0$ due to the nature of the problem, we obtain the following equation for the unknown b:

$$1 = \beta\Big\{\widehat{p} + (1-\widehat{p}\,)\Big[b\lambda^{\gamma_1} + (1-b)\lambda^{\gamma_2}\Big]\Big\}. \tag{33}$$

It has the solution

$$\widehat{b} = \frac{(1-\widehat{p}\,)\lambda^{\gamma_2} + \widehat{p} - \beta^{-1}}{(1-\widehat{p}\,)(\lambda^{\gamma_2} - \lambda^{\gamma_1})}\,. \tag{34}$$

Since γ_1 and γ_2 can be determined from (30), it easily follows that $0 < \widehat{b} < 1$.

Let $\widehat{V}_{c_0}(x) = c_0\psi_{\widehat{b}}(x)$ for $x < x_0$, where c_0 and x_0 are some constants to be determined. Clearly, the required function $\widehat{V}(x)$ belongs to the family of functions

$$\widehat{V}_{c_0}(x; x_0) = \begin{cases} (x-a)^+, & x \geqslant x_0, \\ \widehat{V}_{c_0}(x), & x < x_0. \end{cases} \tag{35}$$

Here the 'optimal' values $\widehat{c}$ and $\widehat{x}$ of c_0 and x_0 can be found by means of the following considerations: the required function $\widehat{V}(x) = \widehat{V}(x; \widehat{x})$ must be the *smallest* *β-excessive majorant* of $g(x)$, i.e., the smallest function satisfying *simultaneously* the two inequalities

$$\begin{aligned} \widehat{V}(x) &\geqslant g(x), \\ \widehat{V}(x) &\geqslant \beta\widehat{T}\widehat{V}(x) \end{aligned} \tag{36}$$

for each $x \in \widehat{E} = \{1, \lambda, \lambda^2, \dots\}$.

The solubility of this problem can be proved and the *exact* values of $\widehat{c}$ and $\widehat{x}$ can be found in precisely the same manner as for a standard call option (see § 5b.6 and [283]). Namely, if $\Delta = \lambda - 1$ is close to zero, then we can consider approximations $\widetilde{c}$ and $\widetilde{x}$ of $\widehat{c}$ and $\widehat{x}$ obtained as follows (cf. a similar procedure in § § 5b, c.)

We shall assume that the functions $\psi_{\widehat{b}}(x)$, $\widehat{V}_{c_0}(x)$, $g(x)$, and $\widehat{V}_{c_0}(x; x_0)$ on $\widehat{E} = \{1, \lambda, \lambda^2, \dots\}$ are defined by the same formulas on $[1, \infty)$.

Then the approximations $\widetilde{c}$ and $\widetilde{x}$ can be found from the additional conditions

$$\begin{aligned} \widehat{V}_{\widetilde{c}}(\widetilde{x}) &= g(\widetilde{x}), \\ \left.\frac{d\widehat{V}_{\widetilde{c}}(x)}{dx}\right|_{x=\widetilde{x}-} &= \left.\frac{dg(x)}{dx}\right|_{x=\widetilde{x}+}. \end{aligned} \tag{37}$$

Bearing in mind that

$$\widehat{V}_{\widetilde{c}}(x) = \widetilde{c}\psi_{\widehat{b}}(x) = \widetilde{c}\big[\widehat{b}x^{\gamma_1} + (1-\widehat{b})x^{\gamma_2}\big],$$

$g(x) = (x-a)^+$, and, for sure, $\widetilde{x} > a$, we see that $\widetilde{c}$ and $\widetilde{x}$ are the solutions of the system of equations

$$\begin{aligned} \widetilde{c}\big[\widehat{b}\widetilde{x}^{\gamma_1} + (1-\widehat{b})\widetilde{x}^{\gamma_2}\big] &= \widetilde{x} - a, \\ \widetilde{c}\big[\widehat{b}\gamma_1\widetilde{x}^{\gamma_1 - 1} + (1-\widehat{b})\gamma_2\widetilde{x}^{\gamma_2 - 1}\big] &= 1. \end{aligned} \tag{38}$$

In particular, for $a = 0$ we have

$$\widetilde{x} = \left(\frac{\widehat{b}}{1-\widehat{b}}\,\frac{1-\gamma_1}{\gamma_2 - 1}\right)^{\frac{1}{\gamma_2-\gamma_1}}, \tag{39}$$

$$\widetilde{c} = \frac{\widetilde{x}}{\gamma_1\widehat{b}\widetilde{x}^{\gamma_1} + \gamma_2(1-\widehat{b})\widetilde{x}^{\gamma_2}}. \tag{40}$$

The above considerations show that for positive, sufficiently small Δ the quantity $\widehat{V}_{\widetilde{c}}(1)$ is close to $\widehat{V}(1)$. Hence, in view of (17) and the equality $\widehat{V}_{\widetilde{c}}(1) = \widetilde{c}$, we obtain that $\widehat{\mathbb{C}} \approx S_0 \cdot \widetilde{c}$ for small $\Delta > 0$. (See [283] for a more detailed analysis; cf. also Chapter VIII, § 2d.)

Chapter VII. Theory of Arbitrage in Stochastic Financial Models. Continuous Time

1. Investment Portfolio in Semimartingale Models

§ 1a. Admissible Strategies. Self-Financing. Stochastic Vector Integral

In this section we shall consider models of two securities markets with continuous time:

a (B, S) market formed by a bank account B and stock $S = (S^1, \dots, S^d)$ of finitely many kinds,

and

a $(B, \mathcal{P})$ market formed also by a bank account B and, in general, a continual family of bonds $\mathcal{P} = \{\mathsf{P}(t, T) : 0 \leqslant t \leqslant T,\ T > 0\}$.

In § § 1–4 we are concerned with (B, S)-markets. A $(B, \mathcal{P})$-market has some peculiarities and we defer its discussion to § 5.

1. We consider a financial market of $d + 1$ assets $X = (X^0, X^1, \dots, X^d)$ that operates in uncertain conditions of the probabilistic character described by a filtered probability space (stochastic basis) $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geqslant 0}, \mathsf{P})$, where $(\mathscr{F}_t)_{t\geqslant 0}$ is the flow of incoming 'information'.

Our main assumption about the assets $X^i = (X^i_t)_{t\geqslant 0}$, $i = 0, 1, \dots, d$, is that they are positive *semimartingales* (see Chapter III, § 5a).

By analogy with the discrete-time case we call an arbitrary predictable (see Chapter III, § 5a) $(d + 1)$-dimensional process $\pi = (\pi^0, \pi^1, \dots, \pi^d)$, where $\pi^i = (\pi^i_t)_{t\geqslant 0}$, an *investment portfolio*, and we shall say that π describes the *strategy* (of an investor, trader, ...) in the above market.

The process $X^\pi = (X^\pi_t)_{t\geqslant 0}$, where

$$X^\pi_t = \sum_{i=0}^{d} \pi^i_t X^i_t, \tag{1}$$

or, in the vector notation,

$$X_t^\pi = (\pi_t, X_t), \tag{2}$$

is called the *value* (or the *value process*) of the portfolio π. The value $x = X_0^\pi$ is the *initial capital*, which one emphasizes sometimes by the notation $X^\pi = X^\pi(x)$.

2. In Chapter V, § 1a, in the discussion of the discrete-time case we introduced the concept of *self-financing strategy* π and explained the role of these strategies, in which all changes of the value X^π must be the results of changes in the market value (price) of the assets X^i and no in- or outflows of capital are possible.

The definition of self-financing becomes slightly more delicate in the continuous-time case. In the final analysis this is related to the problem of the description of the class of *integrable* functions with respect to the semimartingales in question.

We recall that in the discrete-time case (see Chapter V, § 1a) we say that a portfolio $\pi = (\pi^0, \pi^1, \dots, \pi^d)$ is self-financing ($\pi \in SF$) if for each $n \geqslant 1$ we have

$$X_n^\pi = X_0^\pi + \sum_{k=1}^n (\pi_k, \Delta X_k), \tag{3}$$

or, in a more expanded form,

$$X_n^\pi = X_0^\pi + \sum_{k=1}^n \sum_{i=0}^d \pi_k^i \, \Delta X_k^i. \tag{4}$$

In the same way, a reasonable definition of a *self-financing protfolio* or a *self-financing strategy* π (we shall write $\pi \in SF$) in the continuous-time case could be the equality

$$X_t^\pi = X_0^\pi + \int_0^t (\pi_s, dX_s) \tag{5}$$

for each $t > 0$. Equivalently,

$$X_t^\pi = X_0^\pi + \int_0^t \sum_{i=0}^d \pi_s^i \, dX_s^i, \tag{6}$$

which we can symbolically express as follows:

$$dX_t^\pi = (\pi_t, dX_t).$$

Of course, we must first *define* the 'stochastic vector integrals' in (5).

One way is simply to set

$$\int_0^t (\pi_s, dX_s) \equiv \sum_{i=0}^d \int_0^t \pi_s^i \, dX_s^i \tag{7}$$

by definition, i.e., to treat a 'stochastic vector integral' as the *sum of usual* 'stochastic integrals'.

This is quite sensible (and we shall use this definition) in the case of 'simple' functions; in fact, this is the most natural (if not the unique) construction suggested by the term 'integration'.

It turns out, however, that the definition (7) *does not cover* all cases when a 'stochastic vector integral' $\int_0^t (\pi_s, dX_s)$ can be defined, e.g., as a limit of some integrals of 'simple' processes $\pi(n) = (\pi_s(n))_{s\geqslant 0}$, $n \geqslant 1$, approximating $\pi = (\pi_s)_{s\geqslant 0}$ in some suitable sense.

The point here is as follows.

First, even the usual (scalar) stochastic integrals $\pi^i \cdot X_t^i \equiv \int_0^t \pi_s^i \, dX_s^i$ can be defined for a broader class of *predictable* processes π^i than the locally bounded ones discussed in our exposition in Chapter III, § 5a. (What makes *locally bounded* processes π^i attractive is the following feature: if $X^i \in \mathscr{M}_{\text{loc}}$, then the stochastic integral $\pi^i \cdot X^i$ is *also* in the class $\mathscr{M}_{\text{loc}}$; see property (c) in Chapter III, § 5a.7.)

Second, the 'component-wise' definition (7) does not take into account the possible 'interference' of the semimartingales involved; in principle, this interference can extend the class of vector-valued processes $\pi = (\pi^0, \pi^1, \dots, \pi^d)$ that can be approximated by 'simple' processes $\pi(n)$, $n \geqslant 1$.

3. We explain now the main ideas and results of 'stochastic vector integration' that takes these points into consideration; we refer to special literature for detail (see, e.g., [74], [172], [248; Chapter II], [249], [250], [303], [304], or [347]).

Let $X = (X^1, \dots, X^d)$ be a d-dimensional semimartingale admitting a decomposition

$$X = X_0 + A + M, \tag{8}$$

where $A = (A^1, \dots, A^d)$ is a process of bounded variation and $M = (M^1, \dots, M^d)$ is a local martingale ($A \in \mathscr{V}$ and $M \in \mathscr{M}_{\text{loc}}$).

Clearly, we can find a nondecreasing adapted (to the flow $\mathbb{F} = (\mathscr{F}_t)_{t\geqslant 0}$) process $C = (C_t)_{t\geqslant 0}$, $C_0 = 0$, and adapted processes $c^i = (c_t^i)$ and $c^{ij} = (c_t^{ij})$, $i, j = 1, \dots, d$, such that

$$A_t^i = \int_0^t c_s^i \, dC_s, \qquad t > 0, \tag{9}$$

while the quadratic variations satisfy the equality

$$[M^i, M^j]_t = \int_0^t c_s^{ij} \, dC_s. \tag{10}$$

(As regards the definition of the processes $[M^i, M^j]$ and their property $[M^i, M^j]^{1/2} \in \mathscr{A}_{\text{loc}}$, see § 5b, Chapter III.)

Let $\pi = (\pi^1, \dots, \pi^d)$ be a predictable process.

We shall say that

$$\pi \in L_{\mathrm{var}}(A), \tag{11}$$

if (for each $\omega \in \Omega$)

$$\int_0^t \left| \sum_{i=1}^d \pi_s^i c_s^i \right| dC_s < \infty, \qquad t > 0. \tag{12}$$

We also write

$$\pi \in L_{\mathrm{loc}}^q(M) \tag{13}$$

($q \geqslant 1$) if

$$\left[\left(\sum_{i,j=1}^d \pi^i c^{ij} \pi^j \right) \cdot C \right]^{q/2} \in \mathscr{A}_{\mathrm{loc}},$$

i.e., if there exists a sequence of Markov times τ_n approaching ∞ as $n \to \infty$ such that

$$\mathsf{E} \left[\int_0^{\tau_n} \left(\sum_{i,j=1}^d \pi_s^i c_s^{ij} \pi_s^j \right) dC_s \right]^{q/2} < \infty. \tag{14}$$

If there *exists* a representation $X = X_0 + A + M$ such that a predictable process π belongs to the class $L_{\mathrm{var}}(A) \cap L_{\mathrm{loc}}^q(M)$, then we write

$$\pi \in L^q(X) \tag{15}$$

(or also $\pi \in L^q(X; \mathsf{P}, \mathbb{F})$, if we must emphasize the role of the underlying measure P and the flow $\mathbb{F} = (\mathscr{F}_t)_{t \geqslant 0}$).

Importantly, the fact that π belongs to $L^q(X)$ is independent of the choice of the dominating process $C = (C_t)_{t \geqslant 0}$ (see, e.g., [249]).

Both in the scalar and in the vector cases one standard definition of the stochastic integral $\int_0^t (\pi_s, dX_s)$, $t > 0$, for $\pi \in L^q(X)$ consists in setting

$$\int_0^t (\pi_s, dX_s) \equiv \int_0^t (\pi_s, dA_s) + \int_0^t (\pi_s, dM_s), \tag{16}$$

where

$$\int_0^t (\pi_s, dA_s) \equiv \sum_{i=1}^d \int_0^t \pi_s^i c_s^i \, dC_s \tag{17}$$

is the sum of (trajectory-wise) Lebesgue–Stieltjes integrals and

$$\int_0^t (\pi_s, dM_s) \tag{18}$$

is the stochastic integral with respect to the local martingale $M = (M^1, \dots, M^d)$.

The definition of Lebesgue-Stieltjes integrals with respect to a process of bounded variation (for $\pi \in L_{\mathrm{var}}(A)$ and arbitrary $\omega \in \Omega$) encounters no difficulties.

The main problem here is

a) to give a definition of the (vector) integrals (18) with respect to a local martingale M (for $\pi \in L^q_{\mathrm{loc}}(M)$)

and

b) to prove the consistency of the definition (16) or, in other words, to show that the values of the resulting integrals $\int_0^t (\pi_s, dX_s)$ are independent of a particular semimartingale decomposition (8).

Remark 1. There are several pitfalls in the 'natural' definition (16).

First, one does not automatically get, say, the property of *linearity*

$$a\int_0^t (\pi'_s, dX_s) + b\int_0^t (\pi''_s, dX_s) = \int_0^t (a\pi'_s + b\pi''_s, dX_s).$$

It is not *a priori* clear whether the integrability withstands the replacement of the measure P by some equivalent measure $\widetilde{\mathsf{P}}$, i.e., whether $L^q(X; \mathsf{P}, \mathbb{F}) = L^q(X; \widetilde{\mathsf{P}}, \mathbb{F})$ and whether the values of the corresponding integrals are the same (if only P-a.s.).

Neither is it clear whether this definition is invariant under a reduction of the flow of σ-algebras $\mathbb{F} = (\mathscr{F}_t)_{t\geqslant 0}$. Namely, assume that X is a $\mathbb{F}$-semimartingale such that the X_t are $\mathscr{G}$-measurable, where $\mathbb{G} = (\mathscr{G}_t)_{t\geqslant 0}$ is a flow of σ-algebras satisfying *general conditions* (Chapter III, § 5a) and $\mathscr{G}_t \subseteq \mathscr{F}_t$, $t \geqslant 0$. It is well known (see, e.g., [249] and [250]) that X is also a $\mathbb{G}$-semimartingale in that case. Hence one would anticipate that if a process π is $\mathbb{G}$-adapted, then

$$\pi \in L^q(X; \mathsf{P}, \mathbb{F}) \implies \pi \in L^q(X; \mathsf{P}, \mathbb{G})$$

and the value of the stochastic integral is independent of the particular stochastic basis, $(\Omega, \mathscr{F}, \mathbb{F})$ or $(\Omega, \mathscr{F}, \mathbb{G})$, underlying the processes π and X.

It is shown in [74], [248], and [249] that all these properties hold nicely (for each $q \geqslant 1$).

4. We describe now the construction of the integrals (18) with respect to a local martingale M for $\pi \in L^q_{\mathrm{loc}}(M)$.

If $\pi \in L^2(M)$, i.e.,

$$\|\pi\|_{L^2(M)} \equiv \left[\mathsf{E}\int_0^\infty \left(\sum_{i,j=1}^d \pi^i_s c^{ij}_s \pi^j_s\right) dC_s\right]^{1/2} < \infty, \tag{19}$$

then the stochastic integral $(\pi \cdot M)_t \equiv \int_0^t (\pi_s, dM_s)$ can be defined as in the scalar case for square integrable martingales (see Chapter III, § 5a.4).

Namely, we find first a sequence of simple predictable vector processes $\pi(n) = (\pi^1(n), \dots, \pi^d(n))$, $n \geqslant 1$, such that

$$\|\pi - \pi(n)\|_{L^2(M)} \to 0 \quad \text{as} \quad n \to \infty. \tag{20}$$

For these processes $\pi(n)$ the integrals $(\pi(n) \cdot M)_t$ can be defined component-wise by formula (7).

By the Burkholder–Gundy–Davis inequality (see, e.g., [248; 2.34] or [304; Chapter 1, § 9]),

$$\mathsf{E} \sup_{u \leqslant t} \left| \int_0^u (\pi_s(n), dM_s) \right|^2 \leqslant C_2 \|\pi(n)\|_{L^2(M)}, \qquad t > 0, \tag{21}$$

with some *universal* constant C_2.

We conclude from (20) and (21) that

$$\mathsf{E} \sup_{u \leqslant t} \left| \int_0^u (\pi_s(n) - \pi_s(m), dM_s) \right|^2 \to 0 \quad \text{as} \quad m, n \to \infty.$$

Since the L^2-variables make up a complete space, there exist for each $t \geqslant 0$ a random variable, denoted by $(\pi \cdot M)_t$ or $\int_0^t (\pi_s, dM_s)$ and called the stochastic vector integral of $\pi \in L^2(M)$ with respect to the local martingale M, such that

$$\mathsf{E} \sup_{u \leqslant t} \left| \int_0^u (\pi_s(n), dM_s) - \int_0^u (\pi_s, dM_s) \right|^2 \to 0 \quad \text{as} \quad n \to \infty.$$

It is easy to see (cf. the proof of Theorem 4.40 in [250; Chapter I] that we can choose the variables $\int_0^t (\pi_s, dM_s)$, $t \geqslant 0$, such that the process $\left(\int_0^t (\pi_s, dM_s)\right)_{t \geqslant 0}$ is adapted to the flow $\mathbb{F} = (\mathscr{F}_t)_{t \geqslant 0}$ and has right-continuous trajectories with limits from the left for each $t > 0$.

Using the standard localization trick we can extend these definitions for $\pi \in L^2(M)$ to the class of predictable processes $\pi \in L^2_{\text{loc}}(M)$, i.e., to processes such that (14) holds with $q = 2$.

It is much more complicated to construct stochastic integrals with respect to local martingales M for processes π in the class $L^1_{\text{loc}}(M)$, i.e., for processes such that

$$\left[\left(\sum_{i,j=1}^d \pi_s^i c_s^{ij} \pi_s^j\right) \cdot C\right]^{1/2} \in \mathscr{A}_{\text{loc}}.$$

Even in the scalar case ($d = 1$), where this condition has the simple form

$$(\pi^2 \cdot [M, M])^{1/2} \in \mathscr{A}_{\text{loc}},$$

the construction of the stochastic integral $\pi \cdot M$ involves some refined techniques based on some properties of local martingales that are far from trivial (see [304; Chapter 2, § 2]).

Remark 2. It is worth noting in connection with the condition $(\pi^2 \cdot [M,M])^{1/2} \in \mathscr{A}_{\mathrm{loc}}$ that it definitely holds for locally bounded processes π since, as observed earlier, each local martingale M has the property $[M,M]^{1/2} \in \mathscr{A}_{\mathrm{loc}}$.

As regards different definitions of stochastic *vector* integrals $(\pi \cdot M)_t \equiv \int_0^t (\pi_s, dM_s)$ with respect to local martingales M and for $\pi \in L_{\mathrm{loc}}(M)$ $(\equiv L^1_{\mathrm{loc}}(M))$ or of stochastic integrals $(\pi \cdot X)_t \equiv \int_0^t (\pi_s, dX_s)$ with respect to semimartingales X for $\pi \in L(X)$ $(\equiv L^1(X))$, see [74], [249], and [250], where the authors establish the following, well-anticipated properties (here X and Y are semimartingales):

a) if ρ is a predictable bounded process and $\pi \in L(X)$, then $\rho\pi \in L(X)$, $\rho \in L(\pi \cdot X)$ and $(\rho\pi) \cdot X = \rho \cdot (\pi \cdot X)$;
b) if $X \in \mathscr{M}_{\mathrm{loc}}$, then $L_{\mathrm{loc}}(X) \subseteq L(X)$;
c) if $X \in \mathscr{V}$, then $L_{\mathrm{var}}(X) \subseteq L(X)$;
d) $L(X) \cap L(Y) \subseteq L(X+Y)$, and if $\pi \in L(X) \cap L(Y)$, then $\pi \cdot X + \pi \cdot Y = \pi \cdot (X+Y)$;
e) $L(X)$ is a vector space and $\pi' \cdot X + \pi'' \cdot X = (\pi' + \pi'') \cdot X$, for $\pi', \pi'' \in L(X)$.

The question whether $L(X)$ is the maximal class of processes π where a)–e) hold is discussed in [249] and [250]. Another argument in favor of its 'maximality' is the following result of J. Mémin [343]: the space $\{\pi \cdot X \colon \pi \in L(X)\}$ is closed in the space of semimartingales with respect to the *Emery topology* ([74], [138]).

5. If the process $\pi = (\pi^1, \dots, \pi^d)$ is *locally bounded* and $M \in \mathscr{M}_{\mathrm{loc}}$, then the stochastic vector integral process $\left(\int_0^t (\pi_s, dM_s)\right)_{t \geqslant 0}$ is a *local martingale* ([74], [249]). In the scalar case we mentioned this property in Chapter III, § 5a.7 (property (c)).

It is worth noting that if there is *no* local boundedness, then the stochastic integral $\int_0^t \pi_s \, dM_s$ with respect to a local martingale M *is not*, in general, a local martingale even in the scalar case, as shown by the following example.

EXAMPLE (M. Emery [137]). We consider two independent stopping times σ and τ on a probability space $(\Omega, \mathscr{F}, \mathsf{P})$ that have exponential distribution with parameter one. We set

$$M_t = \begin{cases} 0, & t < \min(\sigma, \tau), \\ 1, & t \geqslant \min(\sigma, \tau) = \sigma, \\ -1, & t \geqslant \min(\sigma, \tau) = \tau. \end{cases} \tag{22}$$

If $\mathscr{F}_t = \sigma(M_s, s \leqslant t)$, $t \geqslant 0$, then it is easy to see that $M = (M_t, \mathscr{F}_t, \mathsf{P})$ is a *martingale.*

We consider a deterministic (and therefore, predictable) process $\pi = (\pi_t)_{t \geqslant 0}$ with $\pi_0 = 0$ and $\pi_t = 1/t$ for $t > 0$.

The martingale M is a process of bounded variation and π is integrable with respect to M (in the Lebesgue–Stieltjes sense) because the random variable $\min(\sigma, \tau)$ is strictly positive with probability one.

By property b) in subsection 4, the integral $\int_0^t \pi_s \, dM_s$ with respect to M coincides with the Lebesgue–Stieltjes integral.

Let $Y_t = \int_0^t \pi_s \, dM_s$, $t \geqslant 0$. By (22) we obtain

$$Y_t = \begin{cases} 0, & t < \min(\sigma, \tau), \\ \dfrac{1}{\min(\sigma, \tau)}, & t \geqslant \min(\sigma, \tau) = \sigma, \\ -\dfrac{1}{\min(\sigma, \tau)}, & t \geqslant \min(\sigma, \tau) = \tau. \end{cases} \tag{23}$$

The process $Y = (Y_t)_{t \geqslant 0}$ is *not* a martingale because $\mathsf{E}|Y_t| = \infty$, $t > 0$. Neither is this process a local martingale because $\mathsf{E}|Y_T| = \infty$ for each stopping time $T = T(\omega)$, not identically equal to zero (see [137] for greater detail).

6. In connection with M. Emery's example there arises a natural question on conditions ensuring that a stochastic vector integral $\left(\int_0^t (\pi_s, dX_s)\right)_{t \geqslant 0}$ is a local martingale if so is the process X. We have already mentioned one such condition: the local boundedness of π.

The following result of J.-P. Ansel and C. Stricker [9; Corollaire 3.5]) gives one conditions in terms of the values of the stochastic integral itself, rather then in terms of π (this is convenient in discussions of arbitrage, as we shall see below).

THEOREM ([9]). *Let $X = (X^1, \dots, X^d)$ be a P-local martingale and let $\pi = (\pi^1, \dots, \pi^d)$ be a predictable process such that the stochastic integral $\pi \cdot X$ is well defined and bounded below by a constant ($\pi \cdot X_t \geqslant C$, $t \geqslant 0$). Then $\pi \cdot X$ is a local martingale.*

7. We now return to the issue of self-financing strategies.

DEFINITION 1. Let $X = (X^0, X^1, \dots, X^d)$ be a $(d+1)$-dimensional nonnegative semimartingale describing the prices of $d+1$ kinds of assets. We say that a strategy $\pi = (\pi^0, \pi^1, \dots, \pi^d)$ is *admissible* (relative to X) if $\pi \in L(X)$.

DEFINITION 2. An admissible strategy $\pi \in L(X)$ is said to be *self-financing* if its value $X^\pi = (X_t^\pi)_{t \geqslant 0}$ defined by (1) has a representation (5).

We denote the class of self-financing strategies by SF or $SF(X)$; cf. Chapter V, § 1a.

For discrete time, self-financing (see (3) or (4)) is equivalent to the relation

$$\sum_{i=0}^{d} (X_{k-1}^i, \Delta \pi_k^i) = 0, \qquad k \geqslant 1,$$

(formula (13) in Chapter V, § 1a).

We consider now the issue of possible analogs of this relation in the continuous-time case, under the assumption (7).

To this end we distinguish the strategies $\pi = (\pi^0, \pi^1, \dots, \pi^d)$ whose components are (predictable) processes of *bounded variation* ($\pi^i \in \mathscr{V}$, $i = 0, 1, \dots, d$). One example of such strategies are *simple* functions that we chose to be the starting point in our construction of stochastic integrals (Chapter III, § 5a).

Under the above-mentioned assumption ($\pi^i \in \mathscr{V}$) we obtain

$$d(\pi_t^i X_t^i) = \pi_t^i \, dX_t^i + X_{t-}^i \, d\pi_t^i$$

(see property b) in Chapter III, § 5b.4), so that self-financing condition (5) is equivalent to the relation

$$\sum_{i=0}^d \int_0^t X_{s-}^i \, d\pi_s^i = 0, \qquad t > 0, \tag{24}$$

or, symbolically,

$$(X_{t-}, d\pi_t) = 0. \tag{25}$$

In a more general case where $\pi = (\pi^0, \pi^1, \dots, \pi^d)$ is a *predicrtable semimartingale* we obtain (using Itô's formula in Chapter III, § 5c)

$$\begin{aligned} d(\pi_t^i X_t^i) &= \pi_{t-}^i \, dX_t^i + X_{t-}^i \, d\pi_t^i + d[X^i, \pi^i]_t \\ &= \pi_t^i \, dX_t^i - \Delta\pi_t^i \, dX_t^i + X_{t-}^i \, d\pi_t^i + d\langle X^{ic}, \pi^{ic}\rangle_t + \Delta\pi_t^i \, \Delta X_t^i \\ &= \pi_t^i \, dX_t^i + X_{t-}^i \, d\pi_t^i + d\langle X^{ic}, \pi^{ic}\rangle_t. \end{aligned} \tag{26}$$

Thus, the condition of self-financing of a semimartingale (predictable) strategy π assumes the following form:

$$\sum_{i=0}^d X_{t-}^i \, d\pi_t^i + d\langle X^{ic}, \pi^{ic}\rangle_t = 0. \tag{27}$$

In particular, if $\pi \in \mathscr{V}$, then $\langle X^{ic}, \pi^{ic}\rangle = 0$ and we derive (25) from (27).

8. The main reason why one mostly restricts oneself to the class of semimartingales in the consideration of continuous-time models of financial mathematics lies in the fact that (as we see) it is in this class that we can define *stochastic* (vector) *integrals* (which is instrumental in the description of the evolution of capital) and self-financing strategies. (This factor has been explicitly pointed out by J. Harrison, D. Kreps, and S. Pliska [214], [215], who were the first to focus on the role of semimartingales and their stochastic calculus in asset pricing.)

This does not mean at all that semimartingales are 'the utmost point'. We can define stochastic integrals for many processes that are not semimartingales, e.g., for a fractional Brownian motion and, more generally, for a broad class of Gaussian processes. Of course, the problem of functions that can be under the integral sign should be considered separately in these cases.

We recall again that in the case of (scalar) semimartingales stochastic integrals are defined for all *locally bonunded predictable* processes (Chapter III, § 5a). It is important here (in particular, for financial calculations) that the stochastic integrals of such functions are also *local martingales.*

The situation becomes much more complicated, however, if we consider locally *unbounded* functions. M. Emery's example above shows that if π is not locally bounded, then the integral process $\int_0^{\cdot}(\pi_s, dM_s)$ (even with respect to a martingale M) is not a local martingale in general.

This is in sharp contrast with the discrete-time case where each 'martingale transform' $\sum\limits_{k\leqslant\cdot}(\pi_k, \Delta M_k)$ is a local martingale. (See the theorem in Chapter II, § 1c).

The following definitions seem appropriate in this context.

DEFINITION 3. Let $X = (X_t, \mathscr{F}_t, \mathsf{P})_{t\geqslant 0}$ be a semimartingale. Then X is called a *martingale transform of order* d ($X \in \mathscr{M}T^d$) if there exists a *martingale* $M = (M^1, \dots, M^d)$ and a predictable process $\pi = (\pi^1, \dots, \pi^d) \in L_{\text{loc}}(M)$ such that

$$X_t = X_0 + \int_0^t (\pi_s, dM_s), \qquad t \geqslant 0 \tag{28}$$

(cf. Definition 7 in Chapter II, § 1c).

DEFINITION 4. Let $X = (X_t, \mathscr{F}_t, \mathsf{P})_{t\geqslant 0}$ be a semimartingale. Then X is called a *local martingale transform of order* d ($X \in \mathscr{M}T^d_{\text{loc}}$) if there exists a *local martingale* $M = (M^1, \dots, M^d)$ and a predictable process $\pi = (\pi^1, \dots, \pi^d) \in L^1_{\text{loc}}(M)$ such that the representation (28) holds.

For discrete time the classes $\mathscr{M}_{\text{loc}}$, $\mathscr{M}T^d$, and $\mathscr{M}T^d_{\text{loc}}$ are *the same* for each $d \geqslant 1$ (see the theorem in Chapter II, § 1c). This is no longer so in the general continuous-time case, as shows M. Emery's example.

9. The above-introduced concept of self-financing, which characterizes markets without in- or outflows of capital, is one possible form of financial constraints on portfolio and transactions in securities markets. In Chapter V, § 1a we considered other kinds of constraints in the discrete-time case.

Such balance conditions can be almost mechanically transferred to the continuous-time case.

For instance, if $X^0 = B$ is a bank account and $X^1 = S$ is a stock paying *dividends*, then the balance conditions can be, by analogy with Chapter V, § 1a.4, written as follows:

$$dX_t^\pi = \beta_t\, dB_t + \gamma_t(dS_t + dD_t), \tag{29}$$

where D_t is the total dividend yield (of one share) on the time interval $[0, t]$. In that case

$$d\left(\frac{X_t^\pi}{B_t}\right) = \gamma_t \left(d\left(\frac{S_t}{B_t}\right) + \frac{dD_t}{B_t}\right). \tag{30}$$

The conditions in the cases involving 'consumption' and 'operating expense' can also be appropriately reformulated (see formulas (25)–(35) and (36)–(40) in Chapter V, § 1a).

§ 1b. Discounting Processes

1. Comparing the values (prices) of different assets one usually distinguishes a 'standard', 'basis' asset and valuates other assets in its terms. For instance, the discussion of the S&P500 market of 500 different stocks (see Chapter I, § 1b or, e.g., [310] for greater detail) it is natural to take the S&P500 Index, a (weighted) average of these 500 assets, for such a basis.

In Chapter I (see § 2c) we exposed briefly a popular *CAPM* pricing model, in which one usually takes a bank account (a riskless asset) for a basis, and the 'quality' and 'riskiness' of various assets A are measured in terms of their 'betas' $\beta(A)$.

In our considerations of $d+1$ assets $X^0, X^1, \dots, X^d$ we shall agree to choose one of them, say, X^0, as the basis asset. It is usually the asset that has the 'most simple' structure. It should be pointed out, however, that, in principle, we could choose an *arbitrary* process $Y = (Y_t, \mathscr{F}_t)_{t \geqslant 0}$ to play that role, as long as it is strictly positive.

There exist also purely 'analytic' criteria for our choice of Y; namely, if such a *discounting* process (or 'numéraire' as it is often called; see, e.g., [175]) is suitably chosen, then the process $\dfrac{X^\pi}{Y}$ is sometimes more easy to manage than X^π itself; see the remark at the end of Chapter V, § 2a.

2. If $Y = (Y_t, \mathscr{F}_t)_{t \geqslant 0}$ is a positive process defined on the stochastic basis $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t \geqslant 0}, \mathsf{P})$ in addition to $X = (X^0, X^1, \dots, X^d)$, then for $i = 0, 1, \dots, d$ we set

$$\overline{X}^i = \frac{X^i}{Y}, \qquad \overline{X}^\pi = \frac{X^\pi}{Y}. \tag{1}$$

If π is a self-financing portfolio (relative to X), then one would like to know whether it is self-financing with respect to the discounted portfolio $\overline{X} = (\overline{X}^0, \overline{X}^1, \dots, \overline{X}^d)$. To this end we assume that we have property (7) in § 1a and $Y^{-1} = \dfrac{1}{Y}$ is a predictable process of bounded variation ($Y^{-1} \in \mathscr{V}$). Then

$$d\overline{X}_t^i = Y_t^{-1}\, dX_t^i + X_{t-}^i\, dY_t^{-1} \tag{2}$$

and

$$d\overline{X}_t^\pi = Y_t^{-1}\, dX_t^\pi + X_{t-}^\pi\, dY_t^{-1}. \tag{3}$$

Hence we see from condition of self-financing (6) in § 1a that

$$\begin{aligned}\sum_{i=0}^{d} \pi_t^i \, d\overline{X}_t^i &= Y_t^{-1} \sum_{i=0}^{d} \pi_t^i \, dX_t^i + \Big(\sum_{i=0}^{d} \pi_t^i X_{t-}^i \Big) \, dY_t^{-1} \\ &= Y_t^{-1} \, dX_t^{\pi} + \Big(\sum_{i=0}^{d} \pi_t^i X_{t-}^i \Big) \, dY_t^{-1}. \end{aligned} \tag{4}$$

Assume that (P-a.s.) for $t > 0$ we have

$$\sum_{s \leqslant t} \sum_{i=0}^{d} \big| \pi_s^i \, \Delta X_s^i \, \Delta Y_s^{-1} \big| < \infty. \tag{5}$$

Then

$$\Big(\sum_{i=0}^{d} \pi_t^i X_{t-}^i \Big) \, dY_t^{-1} = \Big(\sum_{i=0}^{d} \pi_t^i X_t^i \Big) \, dY_t^{-1} - \Big(\sum_{i=0}^{d} \pi_t^i \, \Delta X_t^i \Big) \, \Delta Y_t^{-1}$$

and by (4) we obtain

$$\begin{aligned}\sum_{i=0}^{d} \pi_t^i \, d\overline{X}_t^i &= Y_t^{-1} \, dX_t^{\pi} + X_t^{\pi} \, dY_t^{-1} - \Big(\sum_{i=0}^{d} \pi_t^i \, \Delta X_t^i \Big) \, \Delta Y_t^{-1} \\ &= Y_t^{-1} \, dX_t^{\pi} + X_{t-}^{\pi} \, dY_t^{-1} + \Big(\Delta X_t^{\pi} - \sum_{i=0}^{d} \pi_t^i \, \Delta X_t^i \Big) \, \Delta Y_t^{-1} \\ &= Y_t^{-1} \, dX_t^{\pi} + X_{t-}^{\pi} \, dY_t^{-1}, \end{aligned} \tag{6}$$

because $dX_t^{\pi} = \sum_{i=0}^{d} \pi_t^i \, dX_t^i$, and $\Delta X_t^{\pi} = \sum_{i=0}^{d} \pi_t^i \, \Delta X_t^i$ by the properties of stochastic integrals (see property f) in Chapter III, § 5a.7).

By (3) and (6) we obtain

$$d\overline{X}_t^{\pi} = \sum_{i=0}^{d} \pi_t^i \, d\overline{X}_t^i, \tag{7}$$

which means that the discounted portfolio $\overline{X} = (\overline{X}^0, \overline{X}^1, \dots, \overline{X}^d)$ has the self-financing property.

Remark. We assume in the above proof that Y is a *positive predictable process*, $Y^{-1} \in \mathscr{V}$, and (5) *holds*. As regards other possible conditions ensuring the preservation of self-financing after discounting, see, for instance, [175].

A classical example of a discounting process is a bank account $B = (B_t)_{t\geqslant 0}$ with

$$B_t = B_0 \exp\biggl(\int_0^t r(s)\,ds\biggr), \tag{8}$$

where $r = (r(t))_{t\geqslant 0}$ is some, *stochastic* in general, interest rate that is usually assumed to be a positive process. A bank account is a convenient 'gauge' for the assessment of the 'quality' of other assets, e.g., stock or bonds.

3. Let Y^1 and Y^2 be two discounting assets and let $t \in [0, T]$. We assume that the discounted process $\dfrac{X}{Y^1}$ is a $((d+1)$-dimensional) martingale with respect to some measure P^1 on $(\Omega, \mathscr{F}_T)$.

We now find out when there exists a measure $\mathsf{P}^2 \sim \mathsf{P}^1$ such that the discounted process $\dfrac{X}{Y^2}$ is a martingale with respect to P^2 (cf. our discussion in Chapter V, § 4).

To this end we assume that the process $\dfrac{Y^2}{Y^1}$ is a (positive) martingale with respect to P^1.

For $A \in \mathscr{F}_T$ we set

$$\mathsf{P}^2(A) = \mathsf{E}_{\mathsf{P}^1}\biggl(I_A\, \frac{Y_T^2}{Y_T^1}\Big/\frac{Y_0^2}{Y_0^1}\biggr). \tag{9}$$

Clearly, P^2 is a probability measure in $(\Omega, \mathscr{F}_T)$ and $\mathsf{P}^2 \sim \mathsf{P}^1$.

By Bayes's formula (see (4) in Chapter V, § 3a),

$$\begin{aligned}
\mathsf{E}_{\mathsf{P}^2}\biggl(\frac{X_T^i}{Y_T^2}\,\Big|\,\mathscr{F}_t\biggr) &= \mathsf{E}_{\mathsf{P}^1}\biggl(\frac{X_T^i}{Y_T^2}\cdot\frac{Y_T^2}{Y_T^1}\,\Big|\,\mathscr{F}_t\biggr)\cdot\frac{Y_t^1}{Y_t^2}\\
&= \mathsf{E}_{\mathsf{P}^1}\biggl(\frac{X_T^i}{Y_T^1}\,\Big|\,\mathscr{F}_t\biggr)\cdot\frac{Y_t^1}{Y_t^2}\\
&= \frac{X_t^i}{Y_t^1}\cdot\frac{Y_t^1}{Y_t^2} = \frac{X_t^i}{Y_t^2} \quad (\mathsf{P}^2\text{- and } \mathsf{P}^1\text{-a.s.}).
\end{aligned} \tag{10}$$

Hence the discounted process $\dfrac{X}{Y^2}$ is a martingale with respect to the measure P^2 defined by (9),

It should be taken into account that if f_T is a $\mathscr{F}_T$-measurable nonnegative random variable, then it follows from (9) and (10) (provided that $\mathscr{F}_0 = \{\varnothing, \Omega\}$) that

$$Y_0^1 \mathsf{E}_{\mathsf{P}^1}\biggl(\frac{f_T}{Y_T^1}\biggr) = Y_0^2 \mathsf{E}_{\mathsf{P}^2}\biggl(\frac{f_T}{Y_T^2}\biggr) \tag{11}$$

and (P^2-, P^1-a.s.)

$$Y_0^1 \mathsf{E}_{\mathsf{P}^1}\biggl(\frac{f_T}{Y_T^1}\,\Big|\,\mathscr{F}_t\biggr) = Y_0^2 \mathsf{E}_{\mathsf{P}^2}\biggl(\frac{f_T}{Y_T^2}\,\Big|\,\mathscr{F}_t\biggr). \tag{12}$$

§ 1c. Admissible Strategies. Some Special Classes

1. In accordance with Definition 2 in § 1a, the value $X^\pi = (X_t^\pi)_{t\leqslant T}$ of an admissible strategy $\pi \in SF(X)$ can be represented as follows:

$$X_t^\pi = X_0^\pi + \int_0^t (\pi_s, dX_s), \qquad t \leqslant T, \tag{1}$$

where $\int_0^t (\pi_s, dX_s)$ is a stochastic vector integral with respect to the (nonnegative) *semimartingale* $X = (X^0, X^1, \dots, X^d)$.

We shall assume in what follows that X^0 is positive ($X_t^0 > 0$, $t \leqslant T$) and take it as the discounting process. To avoid 'fractional' expressions $\left(\text{similar to } \dfrac{X_t^i}{X_t^0} \text{ kind}\right)$ we assume from the outset that $X_t^0 \equiv 1$, i.e., the original semimartingale $X = (1, X^1, \dots, X^d)$ is a $(d+1)$-dimensional *discounted* asset.

2. We shall now introduce several special classes of admissible strategies; their role will be completely revealed by our discussions of 'martingale criteria' of the absence of the opportunities for arbitrage (see §§ 4 and 5 below).

DEFINITION 1. For each $a \geqslant 0$ we set

$$\Pi_a(X) = \{\pi \in SF(X)\colon X_t^\pi \geqslant -a,\ t \in [0,T]\}. \tag{2}$$

The meaning of *a-admissibility* "$X_t^\pi \geqslant -a$, $t \in [0,T]$" is perfectly clear: the quantity $a \geqslant 0$ is a bound (resulting from economic considerations) on the possible losses in the process of the implementation of the strategy π.

If $a > 0$, then the value X^π can take negative values, which we interpret as borrowing (either borrowing from the bank account or short selling, say, stock)

If $a = 0$, then the full value $X_t^\pi = \sum_{i=0}^d \pi_t^i X_t^i$ must remain nonnegative for all $0 \leqslant t \leqslant T$.

The classes $\Pi_a(X)$, $a \geqslant 0$, were introduced already in the first papers ([214], [215]) devoted to *arbitrage theory*; later, they were regarded as the most natural classes of strategies where (by contrast to the well-known 'St.-Petersburg game'; see, e.g. [186; 2nd ed.]) the investor is not allowed to 'double his stake after a loss' indefinitely long (cf. Example 2 in Chapter V, § 2b).

In papers devoted to necessary and sufficient conditions of the absence of arbitrage one considers mainly the classes $\Pi_a(X)$, $a \geqslant 0$, and some their generalizations. Here we must mention first of all several papers by F. Delbaen and W. Schachermayer (see, e.g., [100], [101] and historical and bibliographical references therein).

3. The classes $\Pi_a(X)$, $a \geqslant 0$, are by far not the only 'natural' classes of admissible strategies.

The following definition is consistently used by C. A. Sin [447].

DEFINITION 2. Let $g = (g^0, g^1, \dots, g^d)$ be a $(d+1)$-dimensional vector with non-negative components and let $g(X_t) = (g, X_t)$ $\left(= \sum_{i=0}^{d} g^i X_t^i\right)$.

We set

$$\Pi_g(X) = \{\pi \in SF(X) \colon X_t^\pi \geqslant -g(X_t),\ t \in [0,T]\}. \tag{3}$$

As in the case of Definition 1, condition $X_t^\pi \geqslant -g(X_t)$, $t \in [0,T]$, is transparent: at each instant t the quantity $g(X_t)$ imposes a bound on the maximum possible losses or debt levels considered admissible in an 'economy' formed by funds g^0 in a bank account and g^i shares of each of d assets, $i = 1, \dots, d$.

Clearly, if $g^0 \geqslant a$, then $\Pi_a(X) \subseteq \Pi_g(X)$.

4. For a discussion of various issues of *arbitrage* in semimartingale models it can be useful to introduce several classes of $\mathscr{F}_T$-measurable 'test' pay-off functions $\psi = \psi(\omega)$ that can in principle be majorized by the returns $\int_0^T (\pi_s, dX_s)$ of a strategy π in one of the classes of admissible strategies introduced above.

DEFINITION 3. For $a \geqslant 0$ let

$$\Psi_a(X) = \left\{\psi \in L_\infty(\Omega, \mathscr{F}_T, \mathsf{P}) \colon \psi \leqslant \int_0^T (\pi_s, dX_s) \text{ for some strategy } \pi \in \Pi_a(X)\right\}$$

and

$$\Psi_+(X) = \left\{\psi \in L_\infty(\Omega, \mathscr{F}_T, \mathsf{P}) \colon \psi \leqslant \int_0^T (\pi_s, dX_s) \text{ for some strategy } \pi \in \Pi_+(X)\right\},$$

where $\Pi_+(X) = \bigcup_{a \geqslant 0} \Pi_a(X)$.

DEFINITION 4. For $g = (g^0, g^1, \dots, g^d)$ with $g^i > 0$, $i = 0, 1, \dots, d$, we set

$$\Psi_g(X) = \left\{\psi \in L_g(\Omega, \mathscr{F}_T, \mathsf{P}) \colon \psi \leqslant \int_0^T (\pi_s, dX_s) \text{ for some strategy } \pi \in \Pi_g(X)\right\},$$

where $L_g(\Omega, \mathscr{F}_T, \mathsf{P})$ is the set of $\mathscr{F}_T$-measurable random variables ψ such that $|\psi| \leqslant g(X_T)$.

5. As usual, we introduce the following norm in the space $(\Omega, \mathscr{F}_T, \mathsf{P})$ of random variables ψ (although it would be more accurate to speak about equivalence classes of random variables; see, e.g., [439; Chapter II, § 10]):

$$\|\psi\|_\infty \equiv \operatorname*{ess\,sup}_\omega |\psi| = \inf\{0 \leqslant c < \infty \colon \mathsf{P}(|\psi| > c) = 0\}.$$

This makes L_∞ a complete (and therefore, Banach) space.

We shall denote the closures of the sets $\Psi_a(X)$, $a \geqslant 0$, and $\Psi_+(X)$ with respect to this norm $\|\cdot\|_\infty$ by $\overline{\Psi}_a(X)$ and $\overline{\Psi}_+(X)$, respectively.

The space $\Psi_g(X)$ is endowed with the norm $\|\cdot\|_g$ defined by the formula

$$\|\psi\|_g \equiv \left\| \frac{\psi}{g(X_T)} \right\|_\infty.$$

We denote the corresponding closure of $\Psi_g(X)$ by $\overline{\Psi}_g(X)$.

2. Semimartingale Models without Opportunities for Arbitrage. Completeness

§ 2a. Concept of Absence of Arbitrage and Its Modifications

1. In the case of discrete time ($n \leqslant N < \infty$) and finitely many kinds of assets ($d < \infty$) the extended version of the *First fundamental theorem* (Chapter V, § 2e) states that for (B,S)-markets

$$ELMM \Longleftrightarrow EMM \Longleftrightarrow NA. \tag{1}$$

Here *NA* is the *No-Arbitrage* property in the sense of Definition 2 in Chapter V, § 2a. The properties *EMM* and *ELMM* mean the existence of an *Equivalent Martingale Measure* and an *Equivalent Local Martingale Measure*, respectively.

Thus, if there is no arbitrage in the market in question, then the implication $NA \Longrightarrow EMM$ means the existence of a martingale measure ($\widetilde{\mathsf{P}} \sim \mathsf{P}$), which (as shown in the previous chapter) enables one to use the well-developed machinery of martingale theory in calculations.

On the other hand, if there exists at least one martingale measure in our model of a (B,S)-market, then the implication $EMM \Longrightarrow NA$ means that we have a 'fair' market (in the sense of the absence of opportunities for arbitrage).

The first two implications $\Longrightarrow$ and $\Longleftarrow$ in (1) are also or principal importance: they show that martingale and locally martingale measures are in fact *the same* in our case.

Clearly, one would also like to have results of type (1) for the continuous-time case (if only for semimartingale models). However, the situation there is more complicated, although *in essense*, there is '*absence of arbitrage*' (in the sense of an appropriate definition) if and only if *there exists* an equivalent measure with certain 'martingale' properties (which we specify below).

It will be clear from what follows that for a satisfactory answer to the question of the validity of (1) we require different versions of the 'absence of arbitrage', depending, in the long run, on the classes of admissible strategies.

We recall in this connection that one requires essentially no constraints on the strategies $\pi = (\beta, \gamma)$ to prove (1) in the discrete-time case (apart from the standard assumptions of predictability and self-financing).

Continuous time is different: for a mere definition of self-financing we must use there stochastic vector integrals $\int_0^t (\pi_s, dX_s)$, and their existence requires the constraint of admissibility $\pi \in L(X)$ on (predictable) strategies π.

Of course, if we let into consideration only 'simple' strategies, finite linear combinations of 'elementary' ones (Chapter III, § 5a), then there are no 'technical' complications due to the definition of vector integrals (see § 1a).

Unfortunately, in the continuous-time case we *cannot* in general deduce the existence of martingale measures of measures with some or other 'martingale' properties from the absence of arbitrage in the class of 'simple' strategies. (The class of 'simple' strategies is too 'thin'!)

2. We proceed now to main definitions related to the absence of arbitrage in semimartingale models $X = (1, X^1, \dots, X^d)$, $X^i = (X_t^i)_{t \leqslant T}$, $i = 1, \dots, d$.

The following concept is, in effect, *classical* (cf. Definition 2 in Chapter V, § 2a).

DEFINITION 1. We say that the *property NA* holds at time T if for each strategy $\pi \in SF(X)$ with $X_0^\pi = 0$ we have

$$\mathsf{P}(X_T^\pi \geqslant 0) = 1 \implies \mathsf{P}(X_T^\pi = 0) = 1. \tag{2}$$

DEFINITION 2. We say that the properties NA_a and NA_+ hold if

$$\Psi_a(X) \cap L_\infty^+(\Omega, \mathscr{F}_T, \mathsf{P}) = \{0\} \tag{3}$$

or

$$\Psi_+(X) \cap L_\infty^+(\Omega, \mathscr{F}_T, \mathsf{P}) = \{0\}, \tag{4}$$

respectively, where $\Psi_a(X)$ and $\Psi_+(X)$ are as defined in § 1c and $L_\infty^+(\Omega, \mathscr{F}_T, \mathsf{P})$ is the subset of nonnegative random variables in $L_\infty(\Omega, \mathscr{F}_T, \mathsf{P})$.

It is easy to show that (4) is equivalent to the condition

$$\Psi_+^0(X) \cap L_\infty^+(\Omega, \mathscr{F}_T, \mathsf{P}) = \{0\}, \tag{5}$$

where

$$\Psi_+^0(X) = \biggl\{ \psi \in L_\infty(\Omega, \mathscr{F}_T, \mathsf{P}) \colon \psi = \int_0^T (\pi_s, dX_s) \quad \text{for some strategy } \pi \in \Pi_+(X) \biggr\}. \tag{6}$$

DEFINITION 3. We say that the property $\overline{NA}_+$ holds if

$$\overline{\Psi}_+(X) \cap L_\infty^+(\Omega, \mathscr{F}_T, \mathsf{P}) = \{0\}. \tag{7}$$

The property $\overline{NA}_+$, which is a refinement of NA_+, is consistently used by F. Delbaen and W. Schachermayer (see, e.g., [100], [101]), who call it the *NFLVR-property: No Free Lunch with Vanishing Risk.*

This name can be explained as follows.

In the discussion of the absence of arbitrage in its NA_+-version we take for 'test' functions ψ only *nonnegative* functions that are smaller or equal to the 'returns' $\int_0^T (\pi_s, dX_s)$ from the strategy $\pi \in \Pi_+(X)$.

However, in our considerations of the $\overline{NA}_+$-version of the absence of arbitrage we can take (also nonnegative) 'test' functions $\psi \in \overline{\Psi}_+(X) \cap L^+_\infty(\Omega, \mathscr{F}_T, \mathsf{P})$, e.g., ones that are the *limits* (in the norm $\|\cdot\|_\infty$) of some sequences of elements ψ^k $(k \geqslant 1)$ of $\Psi_+(X)$, which can take, generally speaking, *negative* values $\left(\text{in particular, these can be some integrals } \int_0^T (\pi_s^k, dX_s)\right)$.

Since $\|\psi^k - \psi\|_\infty \to 0$ as $k \to \infty$, we can assume that $\psi^k \geqslant -1/n$ (for all $\omega \in \Omega$), which can be interpreted as *vanishing risk.*

Remarkably, the absence of arbitrage in the $\overline{NA}_+$-version has a transparent ('martingale') criterion established in [101] (see Theorem 2 in § 2c).

3. We present now versions of the concept of absence of arbitrage based on the use of strategies in the class $\Pi_g(X)$.

DEFINITION 4. Let $g = (g^0, g^1, \dots, g^d)$, where $g^i > 0$, $i = 0, 1, \dots, d$. We say that the properties NA_g and $\overline{NA}_g$ are satisfied if

$$\Psi_g(X) \cap L^+_\infty(\Omega, \mathscr{F}_T, \mathsf{P}) = \{0\}$$

or

$$\overline{\Psi}_g(X) \cap L^+_\infty(\Omega, \mathscr{F}_T, \mathsf{P}) = \{0\},$$

respectively.

In [447], the property $\overline{NA}_g$ is called the *NFFLVR*-property: *No Feasible Free Lunch with Vanishing Risk.*

§ 2b. Martingale Criteria of the Absence of Arbitrage. Sufficient Conditions

1. We shall assume that our financial market is formed by $d+1$ assets $X = (1, X^1, \dots, X^d)$, where $X^i = (X^i_t)_{t \leqslant T}$ are nonnegative semimartingales on a filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t \leqslant T}, \mathsf{P})$ with $\mathscr{F}_T = \mathscr{F}$, $\mathscr{F}_0 = \{\varnothing, \Omega\}$.

If $\widetilde{\mathsf{P}}$ is a probability measure on $(\Omega, \mathscr{F}_T)$ such that $\widetilde{\mathsf{P}} \sim \mathsf{P}$ and X is a *martingale* with respect to this measure, i.e.,

$$X \in \mathscr{M}(\widetilde{\mathsf{P}})$$

or it is a *local martingale*, i.e.,

$$X \in \mathscr{M}_{\mathrm{loc}}(\widetilde{\mathsf{P}}),$$

then we say that we have the property

$$EMM$$

or

$$ELMM,$$

respectively.

The next theorem, which describes some *sufficient* conditions of the absence of arbitrage, is probably the most useful result of the *theory of arbitrage* in semimartingale models from the standpoint of financial mathematics and engineering.

THEOREM 1. *In the semimartingale model $X = (1, X^1, \dots, X^d)$, for each $a \geqslant 0$ and $g = (g^0, g^1, \dots, g^d)$ with $g^i \geqslant 0$, $i = 0, 1, \dots, d$, we have*

$$ELMM \implies NA_a, \tag{1}$$

$$EMM \implies NA_g. \tag{2}$$

Proof. Let X^π be the value of a strategy $\pi \in \Pi_g(X)$:

$$X_t^\pi = X_0^\pi + \int_0^t (\pi_s, dX_s), \qquad t \leqslant T. \tag{3}$$

Assume that $\widetilde{\mathsf{P}}$ is a martingale measure equivalent to P. As mentioned in Remark 1 in § 1a, the integrability of π with respect to X withstands the replacement of P by the equivalent measure $\widetilde{\mathsf{P}}$. Hence if $\pi \in \Pi_g(X)$, then the stochastic vector integral in (3) is well-defined also with respect to $\widetilde{\mathsf{P}}$.

The proof of the absence of arbitrage for the strategy $\pi \in \Pi_g(X)$ with $X_0^\pi = 0$ (in the sense of property NA_g) is based on the demonstration of the *supermartingale* property of X^π with respect to $\widetilde{\mathsf{P}}$.

Indeed, if the supermartingale property holds, then

$$\mathsf{E}_{\widetilde{\mathsf{P}}} X_T^\pi \leqslant \mathsf{E}_{\widetilde{\mathsf{P}}} X_0^\pi = 0, \tag{4}$$

so that we immediately obtain the required relation $X_T^\pi = 0$ (P- and $\widetilde{\mathsf{P}}$-a.s.) from the condition $X_T^\pi \geqslant 0$ (P and $\widetilde{\mathsf{P}}$-a.s.).

Thus, let us establish the $\widetilde{\mathsf{P}}$-supermartingale property of X^π.

If $\pi \in \Pi_g(X)$, then (P- and $\widetilde{\mathsf{P}}$-a.s.)

$$X_t^\pi = X_0^\pi + \int_0^t (\pi_s, dX_s) \geqslant -(g, X_t). \tag{5}$$

By the linearity of stochastic vector integrals from (5) we obtain

$$\int_0^t (\pi_s + g, dX_s) \geqslant -X_0^\pi - (g, X_0). \tag{6}$$

The process X is, by definition, a martingale with respect to $\widetilde{\mathsf{P}}$, and since the stochastic vector integral in (6) is uniformly bounded below, it is a local martingale by a result of J.-P. Ansel and C. Stricker (see § 1a.6), and therefore (by Fatou's lemma), a supermartingale.

Consequently,

$$X_t^\pi = X_0^\pi + \int_0^t (\pi_s + g, dX_s) - (g, X_t - X_0), \tag{7}$$

where the stochastic integral is a $\widetilde{\mathsf{P}}$-supermartingale and $(g, X_t - X_0)_{t\leqslant T}$ is a $\widetilde{\mathsf{P}}$-martingale. Hence for $\pi \in \Pi_g(X)$ the process X is a supermartingale with respect to the measure $\widetilde{\mathsf{P}}$, which together with (4) proves the required assertion (2).

To prove (1) it suffices to observe that, as follows by (6) with $g = (a, 0, \dots, 0)$, a stochastic integral (with respect to a local martingale) is itself a local martingale. Since $X^0 \equiv 1$, it follows that$(g, X_t - X_0) = 0$ for $g = (a, 0, \dots, 0)$. Hence the right-hand side of (7) is a $\widetilde{\mathsf{P}}$-local martingale, and the proof of (1) can be completed in a similar way to (2).

COROLLARY. *It follows from* (1) *that*

$$ELMM \implies NA_+. \tag{8}$$

2. Assertions (1), (2), and (8) can be improved in the following way.

THEOREM 2. *In the semimartingale model $X = (1, X^1, \dots, X^d)$ we have*

$$ELMM \implies \overline{NA}_+, \tag{9}$$

and if $g = (g^0, g^1, \dots, g^d)$ with $g^i > 0$, $i = 0, 1, \dots, d$, then

$$EMM \implies \overline{NA}_g. \tag{10}$$

Proof. Let $\psi \in \overline{\Psi}_g(X)$ with $\psi \geqslant 0$. Then there exists a sequence $(\psi^k)_{k\geqslant 1}$ of functions in $\Psi_g(X)$ such that

$$\|\psi - \psi^k\|_g = \operatorname*{ess\,sup}_\omega \left| \frac{\psi(\omega) - \psi^k(\omega)}{g(X_T(\omega))} \right| \leqslant \frac{1}{k} \to 0 \quad \text{as} \quad k \to \infty.$$

Without loss of generality we can assume that for all $\omega \in \Omega$ we have

$$-\frac{1}{k} \leqslant \frac{\psi(\omega) - \psi^k(\omega)}{g(X_T(\omega))} \leqslant \frac{1}{k}, \tag{11}$$

and therefore

$$-\frac{g(X_T(\omega))}{k} \leqslant \psi(\omega) - \frac{g(X_T(\omega))}{k} \leqslant \psi^k(\omega). \tag{12}$$

Since $\psi^k \in \Psi_g(X)$, there exists a strategy $\pi^k \in \Pi_g(X)$ such that

$$\psi^k \leqslant \int_0^T (\pi_s^k, dX_s). \tag{13}$$

Together with (12) this brings us to the inequality

$$\frac{g(X_T)}{k} \leqslant \int_0^T (\pi_s^k, dX_s), \tag{14}$$

which shows that, for the sequence of strategies $(\pi^k)_{k\geqslant 1}$, the negative part of the returns described by stochastic integrals (the 'risks') approaches zero as k increases ('vanishing risk').

Inequality (14) is clearly equivalent to the relation

$$-\frac{(g, X_0)}{k} \leqslant \int_0^T \left(\pi_s^k + \frac{g}{k}, dX_s\right). \tag{15}$$

Since $|\psi - \psi^k| \leqslant g(X_T)/k$, we obtain in view of (13), (15), and Fatou's lemma, that

$$\begin{aligned} 0 \leqslant \mathsf{E}_{\widetilde{\mathsf{P}}}\psi &= \mathsf{E}_{\widetilde{\mathsf{P}}} \lim \psi^k = \mathsf{E}_{\widetilde{\mathsf{P}}} \lim\left(\psi^k + \frac{(g, X_T - X_0)}{k}\right) \\ &= \mathsf{E}_{\widetilde{\mathsf{P}}} \varliminf\left(\psi^k + \frac{(g, X_T - X_0)}{k}\right) \leqslant \mathsf{E}_{\widetilde{\mathsf{P}}} \varliminf\left(\int_0^T \left(\pi_s^k + \frac{g}{k}, dX_s\right)\right) \\ &\leqslant \varliminf \mathsf{E}_{\widetilde{\mathsf{P}}} \int_0^T \left(\pi_s^k + \frac{g}{k}, dX_s\right) \leqslant 0, \end{aligned} \tag{16}$$

where the last inequality is a consequence of the $\widetilde{\mathsf{P}}$-supermartingale property of the stochastic integrals $\int_0^t \left(\pi_s^k + \frac{g}{k}\, , \, dX_s\right)$, $t \leqslant T$.

Hence $\mathsf{P}(\psi = 0) = \widetilde{\mathsf{P}}(\psi = 0) = 1$, which proves the implication (10).

To prove (9) we assume that $\psi \in \overline{\Psi}_+(X)$ and $\psi \geqslant 0$. Then there exists a sequence of functions $(\psi^k)_{k\geqslant 1}$ in $\Psi_+(X)$ such that

$$\|\psi - \psi^k\|_\infty \equiv \operatorname*{ess\,sup}_\omega |\psi(\omega) - \psi^k(\omega)| \leqslant \frac{1}{k} \to 0, \tag{17}$$

and moreover

$$\psi^k \leqslant \int_0^T (\pi_s^k, dX_s) \tag{18}$$

for $\pi^k \in \Pi_{a_k}(X)$ with some $a_k \geqslant 0$.

By (17) and (18) we obtain

$$-\frac{1}{k} \leqslant \int_0^T (\pi_s^k, dX_s). \tag{19}$$

Further, as in (16), we see that

$$\begin{aligned} 0 \leqslant \mathsf{E}_{\widetilde{\mathsf{P}}} \psi &= \mathsf{E}_{\widetilde{\mathsf{P}}} \lim \psi^k = \mathsf{E}_{\widetilde{\mathsf{P}}} \varliminf \psi^k \\ &\leqslant \mathsf{E}_{\widetilde{\mathsf{P}}} \varliminf \int_0^T (\pi_s^k, dX_s) \\ &\leqslant \varliminf \mathsf{E}_{\widetilde{\mathsf{P}}} \int_0^T (\pi_s^k, dX_s) \leqslant 0, \end{aligned}$$

where the last inequality is again a consequence of the $\widetilde{\mathsf{P}}$-supermartingale property of the stochastic integrals $\int_0^t (\pi_s^k, dX_s)$, $t \leqslant T$. This completes the proof.

§ 2c. Martingale Criteria of the Absence of Arbitrage. Necessary and Sufficient Conditions (a List of Results)

1. In the present section we *state* several results concerning *necessary* and *sufficient* conditions of the absence of arbitrage (in one or another sense).

We recall again the implications *EMM* $\Longleftrightarrow$ *NA* in the discrete-time case, which has become a prototype for various result about general semimartingale models.

Moreover, if the implication *EMM* $\Longrightarrow$ *NA* is easy to prove, then the proof of the reverse implication *EMM* $\Longleftarrow$ *NA*, where one must either suggest an *explicit* construction or prove the *existence* of a martingale measure, is based on designs that are far from simple (even in the seemingly simple case of discrete time!; see § 2 in Chapter V).

Little wonder, therefore, that the proof of the corresponding results in the continuous-time case (due mainly to F. Delbaen and W. Schachermayer [100], [101]) is fairly complicated and we restrict ourselves to a *list* of several interesting results, referring the reader to the indicated special literature for detail.

THEOREM 1 ([100]). a) *Let $X = (1, X^1, \dots, X^d)$ be a semimartingale with bounded components. Then*

$$\boxed{EMM \Longleftrightarrow \overline{NA}_+} \tag{1}$$

b) *Let $X = (1, X^1, \dots, X^d)$ be a semimartingale with locally bounded components. Then*

$$\boxed{ELMM \Longleftrightarrow \overline{NA}_+} \tag{2}$$

2. To state the result concerning necessary and sufficient conditions of property $\overline{NA}_+$ in *general semimaryingale* models let us introduce, following [101], the concept of σ-martingale and σ-martingale measure.

To this end we recall that for *discrete time* each local martingale X is incidentally a *martingale transform*, i.e., $X = X_0 + \gamma \cdot M$ (see the theorem in Chapter II, § 1c), where γ is a predictable sequence and M is a martingale.

Analyzing the proof of this theorem in Chapter II, § 1c one observes that each local martingale X can be represented as a martingale transform $X = X_0 + \gamma \cdot M$, where γ is a *positive* predictable sequence.

Bearing that in mind and following [101] we shall call a martingale transform with positive elements of γ a *σ-martingale*. Further, if there exists a measure $\widetilde{\mathsf{P}} \sim \mathsf{P}$ such that X is a σ-martingale with respect to it, then we shall talk about the *EσMM* property.

Using these concepts, in the discrete-time case we can now put the *First fundamental theorem* (Chapter V, §§ 2b, c; see also (1) in § 2a) in the following form:

$$\boxed{EMM \iff ELMM \iff E\sigma MM \iff NA.} \tag{3}$$

This form is useful since it suggests routes for generalizations of the *First fundamental theorem* to *continuous time*.

It was a substantial success of the authors of [101] that they have developed a clear perception of the following approach: to find necessary and sufficient conditions for the absence of arbitrage in general *semimartingale* models in its $\overline{NA}_+$-version one must turn to σ-martingales and σ-martingale measures.

We now give the corresponding definitions.

DEFINITION 1. We call a semimartingale $X = (X^1, \dots, X^d)$ a *σ-martingale* if there exist a $\mathbb{R}^d$-valued martingale $M = (M_t)_{t\leqslant T}$ and an M-integrable *predictable positive* one-dimensional process $\gamma = (\gamma_t)_{t\leqslant T}$ such that $X = X_0 + \gamma \cdot M$.

DEFINITION 2. If there exists a measure $\widetilde{\mathsf{P}} \sim \mathsf{P}$, such that a semimartingale X is a σ-martingale with respect to this measure, then we say that $\widetilde{\mathsf{P}}$ is a *semimartingale measure* and the *EσMM*-property holds.

Remark 1. As already mentioned, the term 'σ-martingale' has been introduced in [101]. Before that such processes had been called (see, e.g., [73] or [137]) *semimartingales of the class* (Σ_m). We point out that σ-martingales are special cases of martingale transforms (see Definition 3 in § 1a).

The following results can be seen as the culmination of the process of search of necessary and sufficient conditions for the absence of opportunities to arbitrage in its $\overline{NA}_+$-version.

THEOREM 2 ([101]). *In general semimartingale models*

$$\boxed{E\sigma MM \iff \overline{NA}_+} \tag{4}$$

Remark 2. M. Emery's example (§ 1a.5) shows that a σ-martingale is not necessarily a local martingale.

For a clearer insight into the connections between Theorems 1 and 2 and the corresponding results in the discrete-time case (see (3)) we reformulate them as follows.

COROLLARY. *In general semimartingale models* $X = (1, X^i)_{1\leqslant i\leqslant d}$, *where* $X^i = (X^i_t)_{t\leqslant T}$, $i = 1, \dots, d$, $T < \infty$, *we have*

$$\boxed{EMM \Longrightarrow ELMM \Longrightarrow E\sigma MM \Longrightarrow \overline{NA}_+} \tag{5}$$

For locally bounded semimartingales $X = (1, X^i)_{1\leqslant i\leqslant d}$ *with* $X^i = (X^i_t)_{t\leqslant T}$, $i = 1, \dots, d$, $T < \infty$, *we have*

$$\boxed{EMM \Longrightarrow ELMM \Longleftrightarrow E\sigma MM \Longleftrightarrow \overline{NA}_+} \tag{6}$$

For bounded semimartingales $X = (1, X^i)_{1\leqslant i\leqslant d}$ *with* $X^i = (X^i_t)_{t\leqslant T}$, $i = 1, \dots, d$, $T < \infty$, *we have*

$$\boxed{EMM \Longleftrightarrow ELMM \Longleftrightarrow E\sigma MM \Longleftrightarrow \overline{NA}_+} \tag{7}$$

3. We consider now the question of necessary and sufficient conditions for the absence of arbitrage in its $\overline{NA}_g$-version.

THEOREM 3 ([447]). *In general semimartingale models* $X = (1, X^1, \dots, X^d)$ *with* $X^i = (X^i_t)_{t\leqslant T}$, $i = 1, \dots, d < \infty$, *the condition* $\overline{NA}_g$ *for* $g = (g^0, g^1, \dots, g^d)$ *with* $g^i > 0$, $i = 0, 1, \dots, d$, *is equivalent to EMM*:

$$\boxed{EMM \Longleftrightarrow \overline{NA}_g} \tag{8}$$

We have already established the implication $\Longrightarrow$. The proof of the reverse implication is based on the following considerations.

Let $X = (1, X^1, \dots, X^d)$ be a semimartingale. Then, as shown in [447], X satisfies condition $\overline{NA}_g$ if and only if the *discounted* prices $\dfrac{X}{g(X)}$ satisfy condition $\overline{NA}_+$. Since $\dfrac{X}{g(X)}$ is a *bounded* semimartingale, it follows by (7) that there exists an equivalent measure, which proves (8).

To complete our list of results we dwell on two counter-examples.

EXAMPLE 1 ($EMM \not\Leftarrow NA$). We consider a (B, S)-market with $B_t \equiv 1$ and $S_t = W_t$, where $W = (W_t)_{t\geqslant 0}$ is a standard Wiener process (the linear Bachelier model; see Chapter VIII, § 1a). For a self-financing strategy $\pi = (\beta, \gamma)$ its value is

$$X^\pi_t = \beta_t + \gamma_t S_t = X^\pi_0 + \int_0^t \gamma_u \, dS_u.$$

We set $\tau = \inf\{t : S_t = 1\}$ and $\gamma_u = I(u < \tau)$. Then $X^\pi_\tau = X^\pi_0 + S_\tau$, so that if $X^\pi_0 = 0$, then $X^\pi_\tau = 1$ (P-a.s.).

Clearly, there exists in this case a martingale measure (namely, the Wiener measure); however, our choice of the self-financing strategy $\pi = (\beta, \gamma)$ with $\gamma_u = I(u < \tau)$ shows that we have opportunities for arbitrage.

EXAMPLE 2 (*ELMM* $\not\Rightarrow$ *NA*, *ELMM* $\Rightarrow$ $\overline{NA}_+$, but $\not\Rightarrow$ NA_g). Let $X_t^0 \equiv 1$, $t \in [0,1]$, and let

$$X_t^1 = \begin{cases} Y_{\tan(\frac{\pi t}{2})}, & 0 \leqslant t < 1, \\ 0, & t = 1, \end{cases}$$

where

$$Y_t = \exp\Bigl(W_t - \frac{1}{2}\Bigr)$$

and $W = (W_t)_{t \leqslant 1}$ is a Wiener process.

The process X^1 is a local martingale. Since $\int_0^1 \gamma_s \, dX_s^1 = 1$ for $\gamma_s \equiv 1$, it follows that we have arbitrage in its classical interpretation. Simultaneously, it follows from (2) that we have property $\overline{NA}_+$. As regards property NA_g, it breaks down here: let us set (as in [447]) $\pi_t^0 = 1$ and $\pi_t^1 = -1$. Then $X_0^\pi = 0$, $X_t^\pi = 1 - X_t^1 \geqslant -g(X_t)$ with $g(X_t) = 1 + X_t^1$ and $X_1^\pi = 1$.

4. All our previous discussions of arbitrage related to the case where the processes describing the evolution of assets were *semimartingales*. It would be natural to consider now the issues of existence or nonexistence of opportunities for arbitrage also in the *nonsemimartingale* case.

In the following two examples we consider (B, S)-markets in which the process $S = (S_t)_{t \geqslant 0}$ is constructed from a fractional Brownian motion $B^{\mathbb{H}} = (B_t^{\mathbb{H}})_{t \geqslant 0}$ with parameter $\frac{1}{2} < \mathbb{H} < 1$, which (as already mentioned in Chapter III, § 2c) is not a semimartingale, and therefore lacks a (local) martingale measure. This feature is an indirect indication (cf. Chapter I, § 2f.4) that there may be opportunities for arbitrage in the corresponding 'fractal' models; and indeed, they occur in the examples below.

EXAMPLE 3 (*NELMM* $\Rightarrow$ *A*). Consider a (B, S)-market of the following *linear* structure:

$$B_t \equiv 1, \quad S_t = 1 + B_t^{\mathbb{H}}, \qquad t \geqslant 0, \tag{9}$$

where $B^{\mathbb{H}} = (B_t^{\mathbb{H}})_{t \geqslant 0}$ is a fractional Brownian motion with $\frac{1}{2} < \mathbb{H} \leqslant 1$. (Cf. the *linear Bachelier model* in Chapter VIII, § 1a.)

We consider the (Markov) strategy $\pi = (\beta, \gamma)$ with

$$\beta_t = -(B_t^{\mathbb{H}})^2 - 2B_t^{\mathbb{H}}, \tag{10}$$

$$\gamma_t = 2B_t^{\mathbb{H}}. \tag{11}$$

Then its value

$$X_t^\pi \equiv \beta_t + \gamma_t S_t = -(B_t^{\mathbb{H}})^2 - 2B_t^{\mathbb{H}} + 2B_t^{\mathbb{H}}(1 + B_t^{\mathbb{H}}) = (B_t^{\mathbb{H}})^2.$$

Using Itô's formula (22) in Chapter VIII, § 5c we obtain

$$dX_t^\pi \equiv d(B_t^{\mathbb{H}})^2 = 2B_t^{\mathbb{H}}\, dB_t^{\mathbb{H}}.$$

From (9) and (11) we see that

$$dX_t^\pi = \gamma_t\, dS_t.$$

This means that the strategy π in question is self-financing. Since $X_0^\pi = 0$ and $X_t^\pi = (B_t^{\mathbb{H}})^2 > 0$ for $t > 0$, it follows that there occurs arbitrage in this (B, S)-market (in the class of 0-admissible strategies) at each instant $t > 0$.

From the financial point of view this is a fairly artificial example in which the prices S_t, $t > 0$, can take negative values. Our next example does not have this deficiency.

EXAMPLE 4 (*NELMM* $\Rightarrow$ *A*). Consider a (B, S)-market such that

$$dB_t = rB_t\, dt, \qquad B_0 = 1, \tag{12}$$

$$dS_t = S_t(r\, dt + \sigma\, dB_t^{\mathbb{H}}), \qquad S_0 = 1, \tag{13}$$

where $B^{\mathbb{H}} = (B_t^{\mathbb{H}})_{t \geqslant 0}$ is again a fractional Brownian motion with $\frac{1}{2} < \mathbb{H} \leqslant 1$. in view of formula (33) in Chapter III, § 5c we obtain

$$B_t = e^{rt}, \tag{14}$$

$$S_t = e^{rt + \sigma B_t^{\mathbb{H}}}. \tag{15}$$

We consider now the strategy $\pi = (\beta, \gamma)$ with

$$\beta_t = 1 - e^{2\sigma B_t^{\mathbb{H}}}, \tag{16}$$

$$\gamma_t = 2(e^{\sigma B_t^{\mathbb{H}}} - 1). \tag{17}$$

For this strategy we have

$$X_t^\pi = \beta_t B_t + \gamma_t S_t = e^{rt}(e^{\sigma B_t^{\mathbb{H}}} - 1)^2.$$

Using Itô's formula (32) in Chapter III, § 5c we obtain

$$dX_t^\pi = re^{rt}(e^{\sigma B_t^{\mathbb{H}}} - 1)^3\, dt + 2\sigma e^{rt + \sigma B_t^{\mathbb{H}}}(e^{\sigma B_t^{\mathbb{H}}} - 1)\, dB_t^{\mathbb{H}},$$

and it is easy to see that the expression on the right-hand side is just the expression for $\beta_t\, dB_t + \gamma_t\, dS_t$ that can be obtained taking account of (14)–(17).

Thus,

$$dX_t^\pi = \beta_t\, dB_t + \gamma_t\, dS_t,$$

which means that the strategy π defined by (16) and (17) is self-financing.

Since for this strategy we also have $X_0^\pi = 0$ and $X_t^\pi > 0$ for $t > 0$, this model (as also the one in Example 3) leaves space for arbitrage (in the class of 0-admissible strategies) for each $t > 0$.

§ 2d. Completeness in Semimartingale Models

1. By analogy with the terminology used in the discrete-time case (see Definition 4 in Chapter V, § 1b) we say that a semimartingale model $X = (X^0, X^1, \dots, X^d)$ is *complete* (or T-complete) if each non-negative bounded $\mathscr{F}_T$-measurable contingent claim f_T can be replicated, i.e., there exists an admissible self-financing portfolio π such that $X_T^\pi = f_T$ (P-a.s.).

Of course, this property of *replicability* depends on the class of self-financing strategies under consideration.

Recall that by the *Second fundamental theorem* (Chapter V, § § 4a, f) an arbitrage-free model with discrete time $n \leqslant N < \infty$ and finitely many assets ($d < \infty$) is complete if and only if the set of martingale measures consists of a *single* element, a measure $\widetilde{\mathsf{P}}$ equivalent to P.

2. We present below one *sufficient* condition of completeness in general semimartingale models such that the corresponding class $\mathscr{P}(\mathsf{P})$ of equivalent martingale measures is non-empty.

Under this assumption we have the following result.

THEOREM. *Assume that the set of martingale measures contains a unique measure* $\widetilde{\mathsf{P}}$. *Then there exists a strategy* π *in the class* $SF(X)$ *such that* $X_T^\pi = f_T$ (P-a.s.) *for each pay-off function* f_T *with* $\mathsf{E}_{\widetilde{\mathsf{P}}}|f_T| < \infty$.

Proof. This can be established in accordance with the following pattern (cf. the diagram in Chapter V, § 4a):

$$|\mathscr{P}(\mathsf{P})| = 1 \overset{\{1\}}{\Longrightarrow} \text{`}X\text{-representability'} \overset{\{2\}}{\Longrightarrow} \text{completeness.}$$

Here the 'X-representability' with respect to a martingale measure $\widetilde{\mathsf{P}} \in \mathscr{P}(\mathsf{P})$ means that each martingale $M = (M_t, \mathscr{F}_t, \widetilde{\mathsf{P}})_{t \leqslant T}$ defined on the same filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t \leqslant T}, \widetilde{\mathsf{P}})$ as the $\widetilde{\mathsf{P}}$-martingale X has a representation

$$M_t = M_0 + \int_0^t (\gamma_s, dX_s), \qquad t \leqslant T,$$

where $\gamma \in L(X)$ (cf. 'S-representability' in Chapter V, § 4b).

The implication $\{1\}$ follows from general results due to J. Jacod (see [248; Chapter II]); for the purposes of arbitrage theory it has been stated for the first time by J. Harrison and S. Pliska [215].

The implication $\{2\}$ can be proved as in the discrete-time case (see the proof of the lemma in Chapter V, § 4b).

As regards particular examples of complete markets, see § § 4 and 5 below.

3. *Remark* 1. The issues of the 'X-representability' of (local) on *incomplete* arbitrage-free markets have been considered, e.g., in [9].

Remark 2. We now dwell on the question of the relations between the concepts of *arbitrage*, (locally) *martingale measure*, and *completeness* the reader comes across throughout the whole book.

It should be emphasized that each of these concepts can be introduced *independently* of the others. Both arbitrage and completeness were defined in terms of the initial ('physical') probability measure P, and there was no word about a martingale measure.

In the case of discrete times and finitely many assets ($N < \infty$, $d < \infty$) it turned out (the *First fundamental theorem*) that the absence of arbitrage has a simple equivalent characterization: the *existence* of a martingale measure ($\mathscr{P}(\mathsf{P}) \neq \varnothing$), while the completeness in such arbitrage-free models turned out (the *Second fundamental theorem*) to be equivalent to the *uniqueness* of the martingale measure ($|\mathscr{P}(\mathsf{P})| = 1$).

However, if we take into consideration also more general models, then the absence of arbitrage does not require the existence of martingale measures, as shown in Example 1 in Chapter VI, § 2b.

In a similar way, the reader should not think (in connection with the *Second fundamental theorem*) that one can discuss completeness only when there are no opportunities for arbitrage or there exist martingale measures. It is formally quite possible that, e.g.,

a) there can be completeness both in arbitrage-free models and in models with arbitrage;
b) there can be completeness in the conditions where a 'classical' (i.e., nonnegative) martingale measure exists, but is not unique;
c) there can be completeness in the conditions where there exists no 'classical' martingale measure, but there exists a unique 'nonclassical' (i.e., of variable sign) martingale measure.

3. Semimartingale and Martingale Measures

§ 3a. Canonical Representation of Semimartingales. Random Measures. Triplets of Predictable Characteristics

1. In the discrete-time case we discuss the *canonical representation* in Chapter II, § 1b and Chapter V, § 3e.

We recall the essential features of this representation.

Let $H = (H_n, \mathscr{F}_n)_{n \geqslant 0}$ be a stochastic sequence, let $h_n = \Delta H_n$ $(= H_n - H_{n-1})$ for $n \geqslant 1$, and let $g = g(x)$ be a bounded truncation function, i.e., a function with compact support equal to x near the origin (one often uses the function $g(x) = xI(|x| \leqslant 1)$). Then

$$H_n = H_0 + \sum_{k=1}^{n} h_k = H_0 + \sum_{k=1}^{n} \bigl(h_k - g(h_k)\bigr) + \sum_{k=1}^{n} g(h_k), \tag{1}$$

and therefore, since the functions $g(h_k)$ are bounded, it follows from (1) by the Doob decomposition that

$$\begin{aligned} H_n = H_0 &+ \sum_{k=1}^{n} \mathsf{E}\bigl[g(h_k) \,|\, \mathscr{F}_{k-1}\bigr] \\ &+ \sum_{k=1}^{n} \bigl[g(h_k) - \mathsf{E}\bigl(g(h_k) \,|\, \mathscr{F}_{k-1}\bigr)\bigr] + \sum_{k=1}^{n} \bigl[h_k - g(h_k)\bigr]. \end{aligned} \tag{2}$$

Taking into consideration the *jump measures* $\mu_k(A) = I_A(h_k)$, $A \in \mathscr{B}(\mathbb{R} \setminus \{0\})$, $k \geqslant 1$, and their *compensators* $\nu_k(A) = \mathsf{E}\bigl(I_A(h_k) \,|\, \mathscr{F}_{k-1}\bigr) = \mathsf{P}\bigl(h_k \in A \,|\, \mathscr{F}_{k-1}\bigr)$, we see from (2) that

$$H_n = H_0 + \sum_{k=1}^{n} \int_{\mathbb{R}} g(x)\,\nu_k(dx) + \sum_{k=1}^{n} \int_{\mathbb{R}} g(x)\big(\mu_k(dx) - \nu_k(dx)\big) + \sum_{k=1}^{n} \int_{\mathbb{R}} \big(x - g(x)\big)\,\mu_k(dx). \tag{3}$$

In a similar way to Chapter V, § 3e relation (3) can be rewritten as follows:

$$H = g * \nu + g * (\mu - \nu) + (x - g) * \mu. \tag{4}$$

We call (4) the *canonical representation* of the sequence $H = (H_n, \mathscr{F}_n)_{n\geqslant 0}$.

It is worth noting that for $\mathsf{E}|h_k| < \infty$, $k \geqslant 1$, we still have the representation (4) if we take $g(x) = x$. In this case the Doob decomposition of the sequence $H = (H_n, \mathscr{F}_n)_{n\geqslant 0}$ has the following form:

$$H = x * \nu + x * (\mu - \nu). \tag{5}$$

2. We proceed now to a discussion of the canonical decomposition for *semimartingales* $H = (H_t, \mathscr{F}_t)_{t\geqslant 0}$ in the case of *continuous* time.

Let $g = g(x)$ be a truncation function. We set

$$\check{H}(g)_t = \sum_{s\leqslant t} \big[\Delta H_s - g(\Delta H_s)\big]. \tag{6}$$

Note that $\Delta H_s - g(\Delta H_s) \neq 0$, provided that $|\Delta H_s| > b$ for some $b > 0$. Since for semimartingales we have $\sum\limits_{s\leqslant t} (\Delta H_s)^2 < \infty$ (P-a.s.) for each $t > 0$ (see (24) and (25) in Chapter III, § 5b), the sums in (6) contain actually only finitely many terms and $\check{H}(g)$ is well defined as a process of bounded variation.

The process

$$H(g) = H - \check{H}(g) \tag{7}$$

has bounded jumps ($|\Delta H(g)| \leqslant b$) and therefore it is a *special* semimartingale (see Chapter III, § 5b), i.e., it has a canonical decomposition

$$H(g) = H_0 + M(g) + B(g), \tag{8}$$

where $B(g) = (B_t(g), \mathscr{F}_t)_{t\geqslant 0}$ is a predictable process of bounded variation with $B_0(g) = 0$ and $M(g) = (M_t(g), \mathscr{F}_t)_{t\geqslant 0}$ is a local martingale with $M_0(g) = 0$.

From (7) and (8) we see that

$$H = H_0 + M(g) + B(g) + \sum_{s\leqslant \cdot} \big[\Delta H_s - g(\Delta H_s)\big]. \tag{9}$$

This is a continuous analogue of (2). Now, to deduce an analog of (4) from (9) we shall need the concepts of *random measure* and its *compensator*.

3. Let $(E, \mathscr{E})$ be a measurable space.

DEFINITION 1. A *random measure* in $\mathbb{R}_+ \times E$ is a family

$$\mu = \{\mu(dt, dx; \omega);\ \omega \in \Omega\}$$

of non-negative measures $(\mathbb{R}_+ \times E, \mathscr{B}(\mathbb{R}_+) \otimes \mathscr{E})$ satisfying the condition $\mu(\{0\} \times E; \omega) = 0$ for each $\omega \in \Omega$.

EXAMPLE 1. A classical example of a random (and in addition, *integer-valued*) measure μ is the *Poisson measure* defined as follows.

For $A \in \mathscr{B}(\mathbb{R}_+) \otimes \mathscr{E}$ let

$$m(A) = \mathsf{E}\mu(A; \omega),$$

where $m(A)$ is a σ-finite (positive) measure.

We assume that $m(A) < \infty$ for each $A \in \mathscr{B}(\mathbb{R}_+) \otimes \mathscr{E}$ such that $A \subset (t, \infty) \times E$, where $t \in \mathbb{R}_+$ is arbitrary, and the random variable $\mu(A, \cdot\,)$ is independent of the particular σ-algebra $\mathscr{F}_t$.

If the measure m (of *intensivity*) has the property $m(\{t\} \times E) = 0$ for each $t \in \mathbb{R}_+$, then we call μ a *Poisson measure*. If, in addition, $m(dt, dx) = dt\, F(dx)$, where F is a positive σ-finite measure, then we call μ a *homogeneous Poisson measure*.

The attribute '*Poisson*' can be explained by the following property of this measure.

Let $(A_i)_{i \geqslant 1}$ be a sequence of pairwise disjoint measurable subsets of $\mathbb{R}_+ \times E$ such that $m(A_i) < \infty$. Then the random variables $\mu(A_i)$, $i \geqslant 1$, are independent and $\mu(A_i)$ has Poisson distribution with expectation $m(A_i)$, i.e.,

$$\mathsf{P}\big(\mu(A_i) = k\big) = \frac{e^{-m(A_i)}(m(A_i))^k}{k!}, \qquad k = 0, 1, \dots .$$

(See [250; Chapter II, § 1c].)

In Chapter III, § 5a, we introduced two σ-algebras, $\mathscr{O}$ and $\mathscr{P}$, of subsets of $\mathbb{R}_+ \times \Omega$, which we called there the σ-algebras of optional and predictable subsets.

An important role in the applications of integer-valued random measures in $\mathbb{R}_+ \times E$ is assigned to the σ-algebras $\widetilde{\mathscr{O}} = \mathscr{O} \otimes \mathscr{E}$ and $\widetilde{\mathscr{P}} = \mathscr{P} \otimes \mathscr{E}$, which are also called the *optional* and the *predictable σ-algebras of subsets* of $\mathbb{R}_+ \times \Omega \times E$.

If $W = W(t, \omega, x)$ is an optional function on $\mathbb{R}_+ \times \Omega \times E$ and μ is a random measure, then we shall define by $W * \mu = \big((W * \mu)_t(\omega), \mathscr{F}_t\big)_{t \geqslant 0}$ the random process

$$(W * \mu)_t(\omega) = \int_{(0,t] \times E} W(s, \omega, x)\, \mu(ds, dx; \omega), \tag{10}$$

where the integral is interpreted as the Lebesgue–Stieltjes integral for each $\omega \in \Omega$ and it is assumed that

$$\int_{(0,t] \times E} |W(s, \omega, x)|\, \mu(ds, dx; \omega) < \infty, \qquad t > 0. \tag{11}$$

DEFINITION 2. We say that a random measure μ is *optional* (*predictable*) if the process $W * \mu$ is optional (predictable) for each optional (predictable) function $W = W(t, \omega, x)$.

DEFINITION 3. An optional measure μ is said to be $\widetilde{\mathscr{P}}$-*σ-finite* if there exists a $\widetilde{\mathscr{P}}$-measurable partitioning $(A_n)_{n \geqslant 1}$ of the set $\mathbb{R}_+ \times \Omega \times E$ such that each variable $(I_{A_n} * \mu)_\infty$ is integrable.

The next theorem is a direct generalization of Corollary 2 (Chapter II, § 5b) to the Doob–Meyer decomposition.

THEOREM 1. *Let μ be an optional $\widetilde{\mathscr{P}}$-σ-finite random measure. Then there exists a unique (up to* P*-indistinguishability) predictable random measure ν, which is called the compensator of μ, that satisfies each of the following two equivalent conditions:*

(a) $\mathsf{E}(W * \nu)_\infty = \mathsf{E}(W * \mu)_\infty$ *for each non-negative $\widetilde{\mathscr{P}}$-measurable function W on $\mathbb{R}_+ \times \Omega \times E$;*
(b) *for each $\widetilde{\mathscr{P}}$-measurable function W on $\mathbb{R}_+ \times \Omega \times E$ such that $|W| * \mu$ is a locally integrable process, the process $|W| * \nu$ is also locally integrable and $W * \mu - W * \nu$ is a local martingale.*

The proof, together with various properties of random measures and their compensators, can be found in [250; Chapter II] or [304; Chapter 3].

Remark. One obvious property of compensator measures ν is as follows. Let $A \in \mathscr{E}$. Then the process $X = (X_t)_{t \geqslant 0}$ with $X_0 = 0$ and

$$X_t = \mu\big((0,t] \times A; \omega\big) - \nu\big((0,t] \times A; \omega\big), \qquad t > 0,$$

is a *local martingale*. The measure $\mu - \nu$ is called for that reason a (random) martingale measure.

EXAMPLE 2. For a Poisson random measure μ introduced in Example 1 its compensator ν is the intensity measure m.

4. We discuss now an important concept of stochastic integral $W * (\mu - \nu)$ of a $\widetilde{\mathscr{P}}$-measurable function $W = W(t, \omega, x)$ with respect to the martingale measure $\mu - \nu$.

If $|W| * \mu \in \mathscr{A}^+_{\text{loc}}$, then $|W| * \nu \in \mathscr{A}^+_{\text{loc}}$ by Theorem 1 and therefore it seems reasonable to set by definition

$$W * (\mu - \nu) = W * \mu - W * \nu. \tag{12}$$

It is easy to show that the so defined process $W * (\mu - \nu) = \big(W * (\mu - \nu)_t\big)_{t \geqslant 0}$ has the following two properties:

a) it is a purely discontinuous local martingale (see Chapter III, § 5b);

b) it makes the 'jumps'

$$\Delta\big(W * (\mu - \nu)\big)_t = \widetilde{W}_t, \tag{13}$$

where

$$\widetilde{W}_t = \int W(t,\omega,x)\,\mu(\{t\} \times dx;\omega) - \int W(t,\omega,x)\,\nu(\{t\} \times dx;\omega).$$

This properties suggests the following definition (cf. [250; Chapter II, § 1d] and [304; Chapter 3, § 5]).

DEFINITION 4. The stochastic integral $W*(\mu-\nu)$ of a $\widetilde{\mathscr{P}}$-measurable function $W = W(t,\omega,x)$ with respect to the martingale measure $\mu - \nu$ is a purely discontinuous local martingale $X = (X_t)_{t\geqslant 0}$ such that the processes $\Delta X = (\Delta X_t)_{t\geqslant 0}$ and $\widetilde{W} = (\widetilde{W}_t)_{t\geqslant 0}$ are indistinguishable.

We have already seen that if the compensator ν of a measure μ has the property $|W| * \nu \in \mathscr{A}^+_{\text{loc}}$ (or, equivalently, $|W| * \mu \in \mathscr{A}^+_{\text{loc}}$), then we can take the process $W * \mu - W * \nu$ as a purely discontinuous local martingale X.

However, the condition $|W| * \nu \in \mathscr{A}^+_{\text{loc}}$, which ensures the existence of such a process X with $\Delta X = \widetilde{W}$ can be loosened.

Here we restrict ourselves to the introduction of the necessary notation and the statements of the corresponding results. We refer to the already mentioned books [250] and [304] for detail.

Let

$$a_t(\omega) = \nu\big(\{t\} \times E;\omega\big),$$

$$q(\omega, B) = \sum_{s\in B} I\big(a_s(\omega) > 0\big)\big(1 - a_s(\omega)\big), \qquad B \in \mathscr{B}(\mathbb{R}_+),$$

$$\widehat{W}_t(\omega) = \int_E W(t,\omega,x)\,\nu\big(\{t\} \times dx;\omega\big).$$

We assume that for each finite Markov time $\tau(\omega)$ we have

$$\int_E \big|W(\tau(\omega),\omega,x)\big|\,\nu\big(\{\tau(\omega)\} \times dx;\omega\big) < \infty \quad \text{(P-a.s.)},$$

and we set

$$G(W) = \frac{(W - \widehat{W})^2}{1 + |W - \widehat{W}|} * \nu + \frac{\widehat{W}^2}{1 + |\widehat{W}|} * q. \tag{14}$$

THEOREM 2. *Let $W = W(t, \omega, x)$ be a $\widetilde{\mathscr{P}}$-measurable function such that*

$$G(W) \in \mathscr{A}^+_{\text{loc}}. \tag{15}$$

Then there exists a unique (up to stochastic indistinguishability) purely discontinuous local martingale $W * (\mu - \nu)$ such that

$$\Delta(W * (\mu - \nu)) = \widetilde{W}.$$

COROLLARY. *Let $\widehat{W} = 0$. If*

$$\frac{W^2}{1 + |W|} \in \mathscr{A}^+_{\text{loc}},$$

then the stochastic integral $W * (\mu - \nu)$ with respect to the martingale measure $\mu - \nu$ is well defined $\Big($as a purely discontinuous local martingale such that $\Delta(W * (\mu - \nu)) = \int W(t, \omega, x)\, \mu(\{t\} \times dx; \omega)\Big)$.

Remark. Let $\widehat{W} = 0$. Then

$$\big[W * (\mu - \nu), W * (\mu - \nu)\big] = W^2 * \mu, \tag{16}$$

which is a consequence of the observation that

$$\big[W * (\mu - \nu), W * (\mu - \nu)\big]_t = \sum_{0 < s \leqslant t} \big(\Delta(W * (\mu - \nu))_s\big)^2 = (W^2 * \mu)_t.$$

It clearly follows from the above equality (16) that for the *predictable* quadratic variation we have

$$\langle W * (\mu - \nu), W * (\mu - \nu)\rangle = W^2 * \nu.$$

5. Special cases of random (moreover, integer-valued) measures are presented by the *jump measures* μ^H of processes $H = (H_t, \mathscr{F}_t)_{t \geqslant 0}$ with right-continuous trajectories having limits from the left (in particular, of semimartingales):

$$\mu^H((0, t] \times A; \omega) = \sum_{0 < s \leqslant t} I_A(\Delta H_s(\omega)),$$

where $A \in \mathscr{B}(\mathbb{R} \setminus \{0\})$.

Such a random measure μ^H in $\mathbb{R}_+ \times E$ (with $E = \mathbb{R} \setminus \{0\}$) is $\widetilde{\mathscr{P}}$-σ-finite, and therefore, by the above theorem, the compensator ν^H of μ^H is well defined.

We consider the canonical decomposition (9) again.

Using the random jump measure μ^H the last term on the right-hand side of (9) can be written as follows:

$$\sum_{s\leqslant \cdot}\left[\Delta H_s - g(\Delta H_s)\right] = \big(x - g(x)\big) * \mu^H.$$

As mentioned in Chapter III, § 5b.6, each local martingale can be represented (and moreover, uniquely) as the sum of a continuous and a purely discontinuous local martingales. Hence the local martingale $M(g)$ in (9) can be represented as follows:

$$M(g)_t = M(g)_0 + M(g)_t^c + M(g)_t^d, \tag{17}$$

where $M(g)^c$ is the *continuous* and $M(g)^d$ is the *purely discontinuous* component.

The continuous local martingale $M(g)^c$ is in fact *independent* of g, and, as mentioned in Chapter III, § 5b.6, is usually denoted by H^c.

As regards the purely discontinuous component $M(g)^d$, which is a purely discontinuous local martingale, it can be represented as follows:

$$M(g)_t^d = \int_{(0,t]\times\mathbb{R}} g(x)\,d(\mu^H - \nu^H). \tag{18}$$

To prove this we must verify that, first, $G(g) \in \mathscr{A}^+_{\mathrm{loc}}$ and, second, the jumps of the local martingales on the right-hand and the left-hand sides of (18) are the same. It is proved in full in [250; Chapter II, § 2c] and [304; Chapter 3, § 5]. Here we consider one special case.

Assume that $\nu^H(\{t\}\times E;\omega) = 0$; then

$$G(g) = \frac{g^2}{1+|g|} * \nu^H$$

and the local integrability of this process is a consequence of the fact that $g = g(x)$ is a truncation function and of the inclusion $(x^2\wedge 1) * \nu^H \in \mathscr{A}^+_{\mathrm{loc}}$ holding for each semimartingale H. Hence the integral in (18) is well-defined.

Further, $\Delta M(g)^d = \Delta M(g) = g(\Delta H) - \Delta B(g)$, where

$$\Delta B(g)_t = \int_{\mathbb{R}} g(x)\,\nu^H(\{t\}\times dx;\omega) \tag{19}$$

([250; Chapter II, 2.14]). Hence, assuming that $\nu^H(\{t\}\times E;\omega) = 0$ we see that $\Delta B(g)_t = 0$, and therefore $\Delta M(g)^d = g(\Delta H)$. However, $\Delta g * (\mu^H - \nu^H)$ is also equal to $g(\Delta H)$, and the jumps of the purely discontinuous local martingales on the right-hand and the left-hand sides of (18) are the same.

Thus, from (9), in view of (17)–(18), we obtain the following representation:

$$H = H_0 + B(g) + H^c + g * (\mu^H - \nu^H) + \big(x - g(x)\big) * \mu^H, \tag{20}$$

which is called the *canonical representation* of the semimartingale H.

Comparing (20) and (4) in the discrete-time case we see that their main difference is the presence of the *continuous* component H^c in (20).

6. There exist two *predictable* terms in (20): $B(g)$ and ν^H. The third important characteristic of H is its '*angular brackets*' $\langle H^c \rangle$, which is the (predictable) compensator of the continuous locally square integrable martingale H^c.

DEFINITION 5. Let $H = (H_t, \mathscr{F}_t)_{t \geqslant 0}$ be a semimartingale and let $g = g(x)$ be a truncation function. We set $B = B(g)$, $C = \langle H^c \rangle$, and $\nu = \nu^H$.

Then we call the collection

$$\mathbb{T} = (B, C, \nu) \tag{21}$$

a *triplet of predictable characteristics* of H.

It must be noted that the components C and ν of $\mathbb{T}$ are *independent* of our choice of the truncation function $g = g(x)$. On the other hand B depends on g. Moreover, if g and g' are two distinct truncation functions, then

$$B(g) - B(g') = (g - g') * \nu.$$

7. We now discuss several properties of semimartingales that can be stated in terms of predictable characteristics B, C, and ν.

a) If H is a *semimartingale*, then

$$(x^2 \wedge 1) * \nu \in \mathscr{A}_{\mathrm{loc}},$$

i.e., the process $\left(\int_{(0,t]\times\mathbb{R}} (x^2 \wedge 1)\, d\nu\right)_{t \geqslant 0}$ is locally integrable. In other words, there exists a sequence of Markov times τ_n, $\tau_n \uparrow \infty$ (P-a.s.) such that

$$\mathsf{E} \int_{(0,t]\times\mathbb{R}} (x^2 \wedge 1)\, d\nu < \infty.$$

b) The semimartingale H is *special* (in particular, it is a local martingale) if and only if

$$\big(x^2 \wedge |x|\big) * \nu \in \mathscr{A}_{\mathrm{loc}}.$$

c) The semimartingale H is *locally square integrable* if and only if

$$x^2 * \nu \in \mathscr{A}_{\mathrm{loc}}.$$

The proof of a) is based on the inequality $\sum_{s \leqslant t} (\Delta H_s)^2 < \infty$ (P-a.s.), $t > 0$, which holds for semimartingales; see Remark 3 in Chapter III, § 5b. As regards the proofs of properties b) and c), see [250; Chapter II, § 2b].

d) If $H = H_0 + N + A$ is a canonical decomposition of a *special* semimartingale H, then

$$H = H_0 + H^c + X * (\mu - \nu) + A. \tag{22}$$

In other words, for special martingales H we can take $g(x) = x$ in the canonical decomposition (20).

e) We associate (for each $\theta \in \mathbb{R}$) the triplet $\mathbb{T} = (B, C, \nu)$ of predictable characteristics of a semimartingale H with the following predictable process of bounded variation:

$$\Psi(\theta)_t = i\theta B_t - \frac{\theta^2}{2} C_t + \int \left(e^{i\theta x} - 1 - i\theta g(x)\right) \nu\big((0,t] \times dx; \omega\big), \tag{23}$$

which is called the *cumulant* (of H); cf. Chapter III, § 1b.

Let

$$G(\theta) = \mathscr{E}\big(\Psi(\theta)\big), \tag{24}$$

where $\mathscr{E}(\Psi(\theta)) = \big(\mathscr{E}(\Psi(\theta))\big)_{t \geqslant 0}$ is the stochastic exponential constructed from $\Psi(\theta)$ (see Chapter III, § 5c, Example 1):

$$\mathscr{E}\big(\Psi(\theta)\big)_t = e^{\Psi(\theta)_t} \prod_{0 < s \leqslant t} \big(1 + \Delta\Psi(\theta)_s\big) e^{-\Delta\Psi(\theta)_s}. \tag{25}$$

THEOREM 3. *Let $\Delta\Psi(\theta)_t \neq -1$, $t > 0$. Then the following properties are equivalent:*

1) *H is a semimartingale with characteristics (B, C, ν);*
2) *for each $\theta \in \mathbb{R}$ the process*

$$\frac{e^{i\theta H_t}}{G(\theta)_t}, \qquad t \geqslant 0, \tag{26}$$

is a local martingale.

(As regards the proof based on Itô's formula for semimartingales see [250; Chapter II, § 2d].)

f) The most simple processes $H = (H_t, \mathscr{F}_t)_{t \geqslant 0}$ in the class of semimartingales are ones with *independent increments.* Their characteristic feature is that the corresponding triplet $\mathbb{T} = (B, C, \nu)$ is *non-random.* In other words, $B = (B_t)_{t \geqslant 0}$, $C = (C_t)_{t \geqslant 0}$, and the compensator measure $\nu = \nu(dt, dx)$ are *independent* of ω (see [250; Chapter II, § 4c]).

Hence the cumulant $\Psi(\theta)$ of such processes is independent of ω and if $\Delta\Psi(\theta) \neq -1$, which corresponds to the case of a process $H = (H_t)$ continuous in probability, then for $H_0 = 0$ we obtain from (22) the *Lévy–Khintchine formula*

$$\mathsf{E} e^{i\theta H_t} = e^{\Psi(\theta)_t} = \exp\left\{ i\theta B_t - \frac{\theta^2}{2} C_t + \int \left(e^{i\theta x} - 1 - i\theta g(x)\right) \nu\big((0,t] \times dx\big) \right\}. \tag{27}$$

For Lévy processes we have

$$B_t = b \cdot t, \qquad C_t = c \cdot t, \qquad \nu(dt, dx) = dt \cdot \nu(dx),$$

and

$$\mathsf{E}e^{i\theta H_t} = e^{t\psi(\theta)},$$

where

$$\psi(\theta) = i\theta b - \frac{\theta^2}{2}c + \int \left(e^{i\theta x} - 1 - i\theta g(x)\right) \nu(dx). \tag{28}$$

(The function $\psi(\theta)$ is called the cumulant along with $\Psi(\theta)_t$.)

The measure $\nu = \nu(dx)$ satisfies the conditions

$$\nu(\{0\}) = 0, \qquad (x^2 \wedge 1) * \nu < \infty \tag{29}$$

and is called the *Lévy measure*. (Cf. Chapter III, § 1b.)

g) We consider now a special case of Lévy processes: 'a Brownian motion with drift and Poisson jumps'.

More precisely, let

$$H_t = mt + \sigma W_t + \sum_{k=1}^{N_t} \xi_k, \tag{30}$$

where $W = (W_t)_{t\geqslant 0}$ is a Wiener process (Brownian motion), $\xi_1, \xi_2, \dots$ are independent identically distributed random variables with distribution function $F(x) = \mathsf{P}(\xi_1 \leqslant x)$, and $N = (N_t)_{t\geqslant 0}$ is the standard Poisson process with parameter $\lambda > 0$ ($\mathsf{E}N_t = \lambda t$). We assume that W, N, and $(\xi_1, \xi_2, \dots)$ are jointly independent.

The following chain of relation brings us easily to the canonical representation, giving us the triplet of predictable characteristics:

$$\begin{aligned}
H_t &= mt + \sigma W_t + \sum_{k=1}^{N_t} \xi_k = mt + \sigma W_t + \int_0^t \int x \, d\mu \\
&= \left(mt + \int_0^t \int g(x)\, d\nu\right) + \left(\sigma W_t + \int_0^t \int g(x)\, d(\mu - \nu)\right) + \int_0^t \int (x - g(x))\, d\mu \\
&= t\left(m + \lambda \int g(x)\, F(dx)\right) + \left(\sigma W_t + \int_0^t \int g(x)\, d(\mu - \nu)\right) + \int_0^t \int (x - g(x))\, d\mu.
\end{aligned}$$

Hence

$$B(g)_t = t\left(m + \lambda \int g(x)\, F(dx)\right),$$

$$C_t = \sigma^2 t,$$

$$d\nu = \lambda\, dt\, F(dx).$$

h) Random sequences $H = (H_n, \mathscr{F}_n)_{n\geqslant 0}$ with *discrete time* can be fitted in a natural way into continuous-time schemes (see Chapter II, § 1f). If $H_n = H_0 + \sum_{k=1}^{n} h_k$ with $h_k = \Delta H_k$, then the triplet $\mathbb{T} = (B(g), C, \nu)$ has the following structure:

$$B(g)_t = \sum_{1\leqslant k\leqslant [t]} \mathsf{E}\big[g(h_k) \mid \mathscr{F}_{k-1}\big],$$

$$C_t = 0,$$

$$d\nu\big((0,t] \times A; \omega\big) = \sum_{1\leqslant k\leqslant [t]} \mathsf{P}(h_k \in A \mid \mathscr{F}_{k-1})$$

with $A \in \mathscr{B}(\mathbb{R} \setminus \{0\})$.

§ 3b. Construction of Marginal Measures in Diffusion Models. Girsanov's Theorem

1. The results of § § 2b, c on necessary and sufficient conditions of the absence of arbitrage demonstrate that it is important for arbitrage theory to find martingale or locally martingale measures equivalent to the original probability measure

One, fairly common, method of the construction of martingale measures is based on *Girsanov's theorem* and its various generalizations. Another method, long known in actuarial studies, is based on the *Esscher transformation* (see § 3c below).

We presented Girsanov's theorem in its original formulation from [183] in Chapter III, § 3e. In the present section we shall prove it, discuss some generalizations, and suggest criteria for the local continuity and the equivalence of probability measures corresponding to diffusion and Itô processes.

2. We consider a process $X = (X_t, \mathscr{F}_t)_{t\geqslant 0}$, defined on a filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geqslant 0}, \mathsf{P})$, which is an *Itô process* (see Chapter III, § 3d) with differential

$$dX_t = a_t(\omega)\, dt + dB_t, \qquad x_0 = 0, \tag{1}$$

where $a = (a_t(\omega), \mathscr{F}_t)_{t\geqslant 0}$ is a process satisfying the condition

$$\mathsf{P}\bigg(\int_0^t |a_s(\omega)|\, ds < \infty\bigg) = 1, \qquad t \geqslant 0, \tag{2}$$

and $B = (B_t, \mathscr{F}_t)_{t\geqslant 0}$ is a standard Brownian motion.

Assume now that $\widetilde{a} = (\widetilde{a}_t(\omega), \mathscr{F}_t)_{t\geqslant 0}$ is another process and

$$\mathsf{P}\bigg(\int_0^t \big(a_s(\omega) - \widetilde{a}_s(\omega)\big)^2\, ds < \infty\bigg) = 1, \qquad t \geqslant 0. \tag{3}$$

Then there exists a well-defined process $Z = (Z_t, \mathscr{F}_t)_{t\geqslant 0}$ (cf. (21) in Chapter III, § 3d) with

$$Z_t = \exp\left\{\int_0^t (\widetilde{a}_s - a_s)\, dB_s - \frac{1}{2}\int_0^t (\widetilde{a}_s - a_s)^2\, ds\right\}, \tag{4}$$

which is a nonnegative local martingale, e.g., for the localizing times

$$\tau_k = \inf\left\{t\colon \int_0^t (\widetilde{a}_s - a_s)^2\, ds \geqslant k\right\}, \qquad k \geqslant 1.$$

By Fatou's lemma this process is a (nonnegative) supermartingale, and therefore by Doob's convergence theorem (see Chapter III, § 3b) there exists with probability one a finite limit $\lim\limits_{t\to\infty} Z_t \ (= Z_\infty)$.

Let

$$\mathsf{E} Z_\infty = 1. \tag{5}$$

(This is equivalent to the uniform integrability of the family $\{Z_t,\ t \geqslant 0\}$.) Then we can define *another* probability measure $\widetilde{\mathsf{P}}$ on $(\Omega, \mathscr{F})$ by setting

$$d\widetilde{\mathsf{P}} = Z_\infty\, d\mathsf{P}. \tag{6}$$

THEOREM 1 (I. V. Girsanov, [183]). *The process $\widetilde{B} = (\widetilde{B}_t, \mathscr{F}_t)_{t\geqslant 0}$ with*

$$\widetilde{B}_t = B_t - \int_0^t (\widetilde{a}_s - a_s)\, ds \tag{7}$$

is a standard Brownian motion with respect to the measure $\widetilde{\mathsf{P}}$ and

$$dX_t = \widetilde{a}_t(\omega)\, dt + d\widetilde{B}_t. \tag{8}$$

We present the proof in subsection 3, while here we discuss several consequences and observations.

COROLLARY 1. *Let*

$$X_t = B_t - \lambda \int_0^t a_s\, ds, \tag{9}$$

where $\lambda \in \mathbb{R}$, and let

$$Z_t^\lambda = \exp\left(\lambda \int_0^t a_s\, dB_s - \frac{\lambda^2}{2}\int_0^t a_s^2\, ds\right). \tag{10}$$

Assume that $\mathsf{E} Z_\infty^\lambda = 1$ and set $d\widetilde{\mathsf{P}}^\lambda = Z_\infty^\lambda\, d\mathsf{P}$. Then the process $X = (X_t)_{t\geqslant 0}$ is a standard Brownian motion with respect to the measure $\widetilde{\mathsf{P}}^\lambda$.

If $\mathsf{E} Z_T^\lambda = 1$ for some finite T, then $X = (X_t)_{t\geqslant 0}$ is a standard Brownian motion on the interval $[0, T]$ with respect to the measure $\widetilde{\mathsf{P}}_T^\lambda$ such that $d\widetilde{\mathsf{P}}_T^\lambda = Z_T^\lambda\, d\mathsf{P}_T$, where $\mathsf{P}_T = \mathsf{P} \,|\, \mathscr{F}_T$.

COROLARY 2. *Let $\tau = \tau(\omega)$ be a finite Markov time and let*

$$\mathsf{E}\exp\left(\lambda B_\tau - \frac{\lambda^2}{2}\tau\right) = 1. \tag{11}$$

We set $d\widetilde{\mathsf{P}}^\lambda = Z_\tau^\lambda\, d\mathsf{P}$, where $Z_\tau^\lambda = \exp\left(\lambda B_\tau - \frac{\lambda^2}{2}\tau\right)$. Then the process $X = (X_t)_{t\geqslant 0}$ with

$$X_t = B_t - \lambda\cdot(t\wedge\tau)$$

is a standard Brownian motion with respect to the measure $\widetilde{\mathsf{P}}^\lambda$.

Remark 1. We formulated Girsanov's theorem in Chapter III, § 3e under the assumptions that $0 \leqslant t \leqslant T$ and $\mathsf{E}Z_T = 1$. This is in fact a special case of the current setting, where $0 \leqslant t < \infty$, for we can set $a_t = 0$ and $\widetilde{a}_t = 0$ for $t > T$.

Remark 2. We know of some conditions ensuring property (11). These are (see, e.g., [288], [303], or [402]):

$$\mathsf{E}e^{\frac{1}{2}\tau} < \infty \quad (\text{`the } \textit{Novikov condition}\text{'}) \tag{12}$$

and

$$\sup_{t\geqslant 0}\mathsf{E}e^{\frac{1}{2}B_{\tau\wedge t}} < \infty \quad (\text{`the } \textit{Kazamaki condition}\text{'}). \tag{13}$$

Since $\sup_{t\geqslant 0}\mathsf{E}e^{\frac{1}{2}B_{\tau\wedge t}} \leqslant (\mathsf{E}e^{\frac{1}{2}\tau})^{1/2}$, condition (13) is looser than (12).

If we set, e.g., $\tau = \inf\{t\colon B_t = 1\}$, then $\mathsf{E}\sqrt{\tau} = \infty$, and therefore $\mathsf{E}e^{\frac{1}{2}\tau} = \infty$. Thus, (12) fails in this case. Nevertheless, condition (13) holds for such times, too:

$$\mathsf{E}\exp\left(B_\tau - \frac{1}{2}\tau\right) = 1.$$

If the stopping times τ are Markov with respect to the flow $(\mathscr{F}_t^B)_{t\geqslant 0}$ generated by a Brownian motion, then we can relax (12) and (13).

THEOREM 2 ([282]). *Let $\varphi = \varphi(t)$ be a nonnegative measurable function such that*

$$\varlimsup_{t\to\infty}\big(B_t - \varphi(t)\big) = +\infty \quad (\mathsf{P}\text{-a.s.}), \tag{14}$$

and let τ be a Markov time with respect to the flow $(\mathscr{F}_t^B)_{t\geqslant 0}$.

Then each condition,

$$\lim_{N\to\infty}\sup_{\sigma\in\mathfrak{M}_0^N}\mathsf{E}\exp\left\{\frac{1}{2}(\tau\wedge\sigma) - \varphi(\tau\wedge\sigma)\right\} < \infty \tag{15}$$

or

$$\lim_{N\to\infty} \sup_{\sigma\in\mathfrak{M}_0^N} \mathsf{E}\exp\Big\{\frac{1}{2}B_{\tau\wedge\sigma} - \varphi(\tau\wedge\sigma)\Big\} < \infty, \tag{16}$$

where $\mathfrak{M}_0^N$ *is the class of Markov times* σ *(with respect to* $(\mathscr{F}_t^B)_{t\geqslant 0}$*) such that* $\mathsf{P}(0 \leqslant \sigma \leqslant N) = 1$, *is sufficient for the equality*

$$\mathsf{E}\exp\Big\{B_\tau - \frac{1}{2}\tau\Big\} = 1. \tag{17}$$

Remark 3. For $Z^\lambda = (Z_t^\lambda)_{t\geqslant 0}$ defined by (10) the corresponding '*Novikov and Kazamaki conditions*' can be stated as follows:

$$\mathsf{E}\exp\Big\{\frac{\lambda^2}{2}\int_0^\infty a_s^2\,ds\Big\} < \infty, \tag{18}$$

$$\mathsf{E}\exp\Big\{\frac{\lambda}{2}\int_0^\infty a_s\,dB_s\Big\} < \infty. \tag{19}$$

As regards the proofs and extensions to processes $Z = (Z_t)_{t\geqslant 0}$ with

$$Z_t = \exp\Big\{L_t - \frac{1}{2}\langle L, L\rangle_t\Big\}, \tag{20}$$

where $L = (L_t)_{t\geqslant 0}$ is a continuous local martingale $\Big(\text{e.g., } L_t = \int_0^t a_s(\omega)\,dB_s,\ t \geqslant 0\Big)$, see [288], [303], or [402].

3. Proof of Girsanov's theorem. As in the discrete-time case (see Chapter V, § 3b), it suffices to verify that (P-a.s.)

$$\mathsf{E}_{\widetilde{\mathsf{P}}}\big[e^{i\theta(\widetilde{B}_t - \widetilde{B}_s)} \mid \mathscr{F}_s\big] = e^{-\frac{\theta^2}{2}(t-s)} \tag{21}$$

for $0 \leqslant s \leqslant t$ and $\theta \in \mathbb{R}$.

To this end we set $\alpha_s = \widetilde{a}_s - a_s$, $\widetilde{B}_t = B_t - \int_0^t \alpha_s\,ds$, and

$$Z_t = \exp\Big(\int_0^t \alpha_s\,dB_s - \frac{1}{2}\int_0^t \alpha_s^2\,ds\Big)$$

(see (7) and (10)).

By Bayes's formula (Chapter V, § 3a),

$$\mathsf{E}_{\widetilde{\mathsf{P}}}\big[e^{i\theta(\widetilde{B}_t - \widetilde{B}_s)} \mid \mathscr{F}_s\big] = \frac{1}{Z_s}\mathsf{E}_{\mathsf{P}}\big[Z_t e^{i\theta(\widetilde{B}_t - \widetilde{B}_s)} \mid \mathscr{F}_s\big], \tag{22}$$

and we claim that the right-hand side of (22) is equal to $e^{-\frac{\theta^2}{2}(t-s)}$.

For simplicity we shall consider the case of $s = 0$.

Let $U_t = e^{i\theta \widetilde{B}_t}$. Then by Itô's formula (Chapter III, § 5c) we see that

$$d(Z_t U_t) = Z_t U_t (\alpha_t + i\theta)\, dB_t - Z_t U_t \frac{\theta^2}{2}\, dt,$$

i.e.,

$$Z_t U_t = 1 + \int_0^t Z_s U_s (\alpha_s + i\theta)\, dB_s - \frac{\theta^2}{2} \int_0^t Z_s U_s\, dt.$$

Hence, using similar arguments to the proof of Lévy's theorem (Chapter III, § 5a), we obtain the following equation for $\mathsf{E}_{\mathsf{P}} Z_t U_t$:

$$\mathsf{E}_{\mathsf{P}} Z_t U_t = -\frac{\theta^2}{2} \int_0^t \mathsf{E}_{\mathsf{P}} Z_s U_s\, ds,$$

from which we conclude that

$$\mathsf{E}_{\widetilde{\mathsf{P}}} e^{i\theta \widetilde{B}_t} = \mathsf{E}_{\mathsf{P}} Z_t U_t = e^{-\frac{\theta^2}{2} t}.$$

Formula (21) can be verified in a similar way, which proves that the process $\widetilde{B} = (\widetilde{B}_t)_{t \geqslant 0}$ defined in (7) is a standard Brownian motion. Relation (8) is a consequence of (1) and (7). This complete the proof of the theorem.

4. Hence if the process X has the differential $dX_t = a_t(\omega)\, dt + dB_t$ with respect to the measure P, then its differential with respect to the measure $\widetilde{\mathsf{P}}$ defined in (6) is $dX_t = \widetilde{a}_t(\omega)\, dt + d\widetilde{B}_t$, where $\widetilde{B} = (\widetilde{B}_t)_{t \geqslant 0}$ is a Brownian motion with respect to $\widetilde{\mathsf{P}}$.

It should be noted that if all our considerations proceed for a time interval $[0, T]$, where T can be also a Markov time, then we can replace the condition $\mathsf{E} Z_\infty = 1$ by a weaker one, $\mathsf{E} Z_T = 1$.

Remark 4. Assume in Girsanov's theorem that $\widetilde{a}_t \equiv 0$, i.e.,

$$\widetilde{B}_t = B_t + \int_0^t a_s\, ds, \quad t \leqslant T$$

and

$$Z_t = \exp\left\{ -\int_0^t a_s\, dB_s - \frac{1}{2} \int_0^t a_s^2\, ds \right\}.$$

If $\mathsf{E} Z_T = 1$, then the process

$$X_t = B_t + \int_0^t a_s\, ds$$

coincides with $\widetilde{B}_t$ and is a Brownian motion with respect to the measure $\widetilde{\mathsf{P}}_T$ such that $d\widetilde{\mathsf{P}}_T = Z_T\, d\mathsf{P}$ (cf. Chapter V, § 3b), therefore it is a martingale measure.

5. Let X be an *Itô process* on the (filtered) space $(C, \mathscr{C}, (\mathscr{C}_t)_{t\geqslant 0})$ of continuous functions $\omega = (\omega_t)_{t\geqslant 0}$ with differential (1) and let $\mu^X = \operatorname{Law}(X \mid \mathsf{P})$ be the probability distribution of this process.

Girsanov's theorem is a convenient tool in answering the question when the restrictions $\mu_t^X = \mu^X \mid \mathscr{C}_t$ of the measures μ^X are absolutely continuous, equivalent, or singular with respect to the measures $\mu_t^B = \mu^B \mid \mathscr{C}_t$.

The measure μ^B is just the Wiener measure (Chapter III, § 3a), so that we are discussing the properties of the measure μ^X corresponding to the process X in its relations to the Wiener measure μ^B. If $\mu_t^X \ll \mu_t^B$ or $\mu_t^B \ll \mu_t^X$, then one would also like to have 'explicit' expressions for the Radon–Nikodym derivatives $\dfrac{d\mu_t^X}{d\mu_t^B}$ and $\dfrac{d\mu_t^B}{d\mu_t^X}$.

These issues are studied in much detail in [303; Chapter 7] for the case of Itô processes and in [250; Chapters III–V] for semimartingales (through the use of the *Hellinger distance* and *Hellinger processes*). For that reason we restrict ourselves to several results.

We consider some time interval $[0, T]$.

Assume that

$$\mathsf{P}\biggl(\int_0^T a_s^2(\omega)\, ds < \infty\biggr) = 1. \tag{23}$$

Then there exists a well-defined process $Z = (Z_t)_{t\leqslant T}$ with

$$Z_t = \exp\biggl(-\int_0^t a_s(\omega)\, dB_s - \frac{1}{2}\int_0^t a_s^2(\omega)\, ds\biggr). \tag{24}$$

THEOREM 3. *If* $\mathsf{E} Z_T = 1$, *then*

$$\mu_T^X \sim \mu_T^B \tag{25}$$

and

$$\frac{d\mu_t^B}{d\mu_t^X}(X(\omega)) = \mathsf{E}\biggl(\exp\biggl[-\int_0^T a_s(\omega)\, dX_s + \frac{1}{2}\int_0^T a_s^2(\omega)\, ds\biggr] \biggm| \mathscr{F}_T^X\biggr)(\omega), \tag{26}$$

where $\mathscr{F}_T^X = \sigma(\omega\colon X_s(\omega),\ s \leqslant T)$.

Proof. We define a measure $\widetilde{\mathsf{P}}_T$ by setting $d\widetilde{\mathsf{P}}_T = Z_T\, d\mathsf{P}_T$ (cf. (6)), where $\mathsf{P}_T = \mathsf{P} \mid \mathscr{F}_T$. The process $X = (X_t)_{t\leqslant T}$ is a Brownian motion with respect to $\widetilde{\mathsf{P}}_T$ by Girsanov's theorem, and therefore

$$\begin{aligned}
\mu_T^B(A) = \widetilde{\mathsf{P}}_T(X \in A) &= \int_{\{\omega\colon X(\omega)\in A\}} Z_T(\omega)\, \mathsf{P}(d\omega) \\
&= \int_{\{\omega\colon X(\omega)\in A\}} \mathsf{E}\bigl[Z_T \mid \mathscr{F}_T^X\bigr](\omega)\, \mathsf{P}(d\omega) \\
&= \mathsf{E}\bigl(I_A(X(\omega)) \cdot \mathsf{E}\bigl[Z_T \mid \mathscr{F}_T^X\bigr](\omega)\bigr).
\end{aligned} \tag{27}$$

Let $\Phi_T(x)$ be a $\mathscr{C}_T$-measurable functional such that $\mathsf{E}[Z_T \mid \mathscr{F}_T^X](\omega) = \Phi_T(X(\omega))$. Then from (27), using the formula of the change of variables in a Lebesgue integral (see, e.g., [439; Chapter II, § 6]) we deduce the equality

$$\mu_T^B(A) = \int_{\{\omega:\, X(\omega)\in A\}} \Phi_T(X(\omega))\,\mathsf{P}(d\omega) = \int_A \Phi_T(x)\,\mu_T^X(dx),$$

and therefore $\mu_T^B \ll \mu_T^X$. In addition,

$$\frac{d\mu_T^B}{d\mu_T^X}(x) = \Phi_T(x) \tag{28}$$

and

$$\frac{d\mu_T^B}{d\mu_T^X}\big(X(\omega)\big) = \mathsf{E}\big[Z_T \mid \mathscr{F}_T^X\big](\omega). \tag{29}$$

We claim now that $\mu_T^X \ll \mu_T^B$. To prove this we observe that $\mathsf{P}_T(Z_T(\omega) > 0) = 1$, and therefore $\mathsf{P}_T \ll \widetilde{\mathsf{P}}_T$ (cf. Chapter V, § 3a) and

$$\frac{d\mathsf{P}_T}{d\widetilde{\mathsf{P}}_T}(\omega) = Z_T^{-1}(\omega). \tag{30}$$

Hence

$$\begin{aligned}\mu_T^X(A) &= \mathsf{P}_T\big(X(\omega)\in A\big) = \mathsf{E}_{\widetilde{\mathsf{P}}_T}\big(I_A\big(X(\omega)\big)Z_T^{-1}\big)\\ &= \mathsf{E}_{\widetilde{\mathsf{P}}_T}\Big(I_A\big(X(\omega)\big)\mathsf{E}_{\widetilde{\mathsf{P}}_T}\big(Z_T^{-1} \mid \mathscr{F}_T^X\big)(\omega)\Big) = \int_A \widetilde{\Phi}_T(x)\,\mu_T^X(dx),\end{aligned} \tag{31}$$

where $\widetilde{\Phi}_T(x)$ is a $\mathscr{C}_T$-measurable functional such that

$$\mathsf{E}_{\widetilde{\mathsf{P}}_T}\big(Z_T^{-1} \mid \mathscr{F}_T^X\big)(\omega) = \widetilde{\Phi}_T\big(X(\omega)\big).$$

From (31) we see that $\mu_T^X \ll \mu_T^B$ and

$$\frac{d\mu_T^X}{d\mu_T^B}\big(X(\omega)\big) = \mathsf{E}_{\widetilde{\mathsf{P}}_T}\big[Z_T^{-1} \mid \mathscr{F}_T^X\big](\omega). \tag{32}$$

6. We consider now the special case when an Itô process X is a *diffusion-type process*, i.e., let $a_t(\omega) = A(t, X(\omega))$ in (1), where $A(t,x)$ is a nonanticipating functional ($A(t,x)$ is measurable in (t,x) and for each fixed t the functional $A(t,x)$ is $\mathscr{C}_t$-measurable in x).

In this case, for

$$dX_t = A(t,X)\,dt + dB_t, \tag{33}$$

we can deduce from Theorem 3 the following result.

THEOREM 4. *Let*

$$\mathsf{P}\left(\int_0^T A^2(t,X)\,dt < \infty\right) = 1 \tag{34}$$

and let

$$\mathsf{E}\exp\left(-\int_0^T A(t,X)\,dB_t - \frac{1}{2}\int_0^T A^2(t,X)\,dt\right) = 1 \tag{35}$$

or, equivalently,

$$\mathsf{E}\exp\left(-\int_0^T A(t,X)\,dX_t + \frac{1}{2}\int_0^T A^2(t,X)\,dt\right) = 1. \tag{36}$$

Then $\mu_T^X \sim \mu_T^B$,

$$\frac{d\mu_T^B}{d\mu_T^X}(X) = \exp\left(-\int_0^T A(t,X)\,dX_t + \frac{1}{2}\int_0^T A^2(t,X)\,dt\right), \tag{37}$$

and

$$\frac{d\mu_T^X}{d\mu_T^B}(X) = \exp\left(\int_0^T A(t,X)\,dX_t - \frac{1}{2}\int_0^T A^2(t,X)\,dt\right). \tag{38}$$

Remark 5. If we give up the condition $\mathsf{E}Z_T = 1$ in Theorem 3, then we can prove (see [303; Theorem 7.4]) that

$$\mathsf{P}\left(\int_0^T A^2(t,X)\,dt < \infty\right) = 1 \implies \mu_T^X \ll \mu_T^B \tag{39}$$

and

$$\left.\begin{aligned} \mathsf{P}\left(\int_0^T A^2(t,X)\,dt < \infty\right) = 1 \\ \mathsf{P}\left(\int_0^T A^2(t,B)\,dt < \infty\right) = 1 \end{aligned}\right\} \implies \mu_T^X \sim \mu_T^B. \tag{40}$$

7. Comparing Theorems 3 and 4 we see that while we have 'explicit' formulae (37) and (38) for diffusion-type processes, the corresponding formulas for Itô processes (see (26)) require the calculation of the conditional expectation $\mathsf{E}(\cdot\,|\,\mathscr{F}_T^X)$. The following result, which is also of independent interest, is useful in the search of 'explicit' formulas for the Radon–Nikodym derivatives also in the case of Itô processes since it allows one to 'transfer' the conditional expectation in (26) under the integral sign.

THEOREM 5. *Let X be an Itô process with differential* (1), *where*

$$\int_0^T \mathsf{E}|a_s(\omega)|\,ds < \infty. \tag{41}$$

Let $A(t,x)$ be a nonanticipating functional such that

$$A\big(t, X(\omega)\big) = \mathsf{E}(a_t \mid \mathscr{F}_t^X)(\omega). \tag{42}$$

Then the process $\overline{B} = (\overline{B}_t)_{t\leqslant T}$ with

$$\overline{B}_t = X_t - \int_0^t A\big(s, X(\omega)\big)\,ds \tag{43}$$

is a Brownian motion (with respect to the flow $(\mathscr{F}_t^X)_{t\leqslant T}$).

If, in addition to (41),

$$\mathsf{P}\bigg(\int_0^T a_s^2(\omega)\,ds < \infty\bigg) = 1, \tag{44}$$

$$\mathsf{E}\exp\bigg(-\int_0^T a_s\,dB_s - \frac{1}{2}\int_0^T a_s^2\,ds\bigg) = 1, \tag{45}$$

then $\mu_T^X \sim \mu_T^B$,

$$\mathsf{P}\bigg(\int_0^T A^2(t,X)\,dt < \infty\bigg) = 1, \tag{46}$$

$$\mathsf{P}\bigg(\int_0^T A^2(t,B)\,dt < \infty\bigg) = 1 \tag{47}$$

and

$$\frac{d\mu_T^X}{d\mu_T^B}(B) = \exp\bigg(\int_0^T A(s,B)\,dB_s - \frac{1}{2}\int_0^T A^2(s,B)\,ds\bigg), \tag{48}$$

$$\frac{d\mu_T^X}{d\mu_T^B}(X) = \exp\bigg(\int_0^T A(s,X)\,dX_s - \frac{1}{2}\int_0^T A^2(s,X)\,ds\bigg). \tag{49}$$

Proof. The fact that the process $\overline{B} = (\overline{B}_t)_{t\leqslant T}$, which in the case of $\mathscr{F}_t^{\overline{B}} = \mathscr{F}_t^X$ $(t \leqslant T)$ is called an *innovation process*, is a *Brownian motion*, can be easily proved on the basis of Itô's formula for semimartingales (Chapter III, § 3d) as applied to $e^{i\lambda(\overline{B}_t - \overline{B}_s)}$ for $t \geqslant s$ and $\lambda \in \mathbb{R}$. In fact,

$$
\begin{aligned}
e^{i\lambda(\overline{B}_t - \overline{B}_s)} = 1 + i\lambda \int_s^t e^{i\lambda(\overline{B}_u - \overline{B}_s)}\, dB_u \\
+ i\lambda \int_s^t e^{i\lambda(\overline{B}_u - \overline{B}_s)} \big[a_u(\omega) - A\big(u, X(\omega)\big)\big]\, du \\
- \frac{\lambda^2}{2} \int_s^t e^{i\lambda(\overline{B}_u - \overline{B}_s)}\, du.
\end{aligned}
\tag{50}
$$

Since

$$
\mathsf{E}\bigg(\int_s^t e^{i\lambda(\overline{B}_u - \overline{B}_s)}\, dB_u \;\bigg|\; \mathscr{F}_s^X\bigg) = 0
$$

and

$$
\begin{aligned}
&\mathsf{E}\bigg[\int_s^t e^{i\lambda(\overline{B}_u - \overline{B}_s)} \big(a_u(\omega) - A\big(u, X(\omega)\big)\big)\, du \;\bigg|\; \mathscr{F}_s^X\bigg] \\
&\qquad = \mathsf{E}\bigg[\int_s^t e^{i\lambda(\overline{B}_u - \overline{B}_s)} \mathsf{E}\Big(a_u(\omega) - A\big(u, X(\omega)\big) \;\big|\; \mathscr{F}_u^X\Big)\, du \;\bigg|\; \mathscr{F}_s^X\bigg] = 0,
\end{aligned}
$$

we see from (50) that (P-a.s.)

$$
\mathsf{E}\big(e^{i\lambda(\overline{B}_t - \overline{B}_s)} \,|\, \mathscr{F}_s^X\big) = 1 - \frac{\lambda^2}{2} \int_s^t \mathsf{E}\big(e^{i\lambda(\overline{B}_u - \overline{B}_s)} \,|\, \mathscr{F}_s^X\big)\, du,
$$

and therefore (P-a.s.)

$$
\mathsf{E}\big(e^{i\lambda(\overline{B}_t - \overline{B}_s)} \,|\, \mathscr{F}_s^X\big) = e^{-\frac{\lambda^2}{2}(t-s)}, \qquad 0 \leqslant s \leqslant t,
\tag{51}
$$

so that $\overline{B}$ is a Brownian motion. (The trajectories of $(\overline{B}_t)_{t \leqslant T}$ are continuous by (43).)

To prove the remaining assertions it suffices to observe that

$$
\frac{d\mu_T^X}{d\mu_T^B} = \frac{d\mu_T^X}{d\mu_T^{\overline{B}}} \cdot \frac{d\mu_T^{\overline{B}}}{d\mu_T^B}, \qquad \frac{d\mu_T^{\overline{B}}}{d\mu_T^B} = 1,
$$

and to use Theorems 3 and 4.

8. Theorems 1–5 can be generalized (see [303; Chapter 7] for greater detail) to multivariate processes X or to the case where (in place of diffusion coefficient 1) we allow the diffusion coefficients in (1) and (33) to depend on X and t.

We present the following result obtained in this direction.

Let $X = (X_t)_{t \leqslant T}$ be a diffusion-type process with

$$dX_t = \alpha(t, X)\,dt + \beta(t, X)\,dB_t, \tag{52}$$

where $\alpha(t, x)$ and $\beta(t, x) > 0$ are nonanticipating functionals, the stochastic integral

$$\int_0^t \beta(s, X)\,dB_s$$

is well defined for $t \leqslant T$, and $\int_0^T |\alpha(s, X)|\,ds < \infty$ (P-a.s.).

Let $\widetilde{\alpha}(t, x)$ be another nonanticipating functional with $\int_0^T |\widetilde{\alpha}(s, X)|\,ds < \infty$ (P-a.s.) such that

$$\int_0^T \left(\frac{\alpha(s, X) - \widetilde{\alpha}(s, X)}{\beta(s, X)} \right)^2 ds < \infty. \tag{53}$$

We set

$$Z_t = \exp\left\{ \int_0^t \frac{\widetilde{\alpha}(s, X) - \alpha(s, X)}{\beta(s, X)}\,dB_s - \frac{1}{2} \int_0^t \left(\frac{\widetilde{\alpha}(s, X) - \alpha(s, X)}{\beta(s, X)} \right)^2 ds \right\}. \tag{54}$$

If $\mathsf{E} Z_T = 1$ and $\widetilde{\mathsf{P}}$ is the measure such that $d\widetilde{\mathsf{P}}_T = Z_T\,d\mathsf{P}_T$, then X is a diffusion-type process with respect to $\widetilde{\mathsf{P}}_T$ and its differential is

$$dX_t = \widetilde{\alpha}(t, X)\,dt + \beta(t, X)\,d\widetilde{B}_t, \tag{55}$$

where $\widetilde{B}$ is a Brownian motion (with respect to $\widetilde{\mathsf{P}}_T$).

EXAMPLE. Let $X_t = e^{mt + \sigma B_t}$. Then

$$dX_t = X_t\left[\left(m + \frac{\sigma^2}{2} \right) dt + \sigma\,dB_t \right], \tag{56}$$

i.e., $\alpha(t, X) = \left(m + \dfrac{\sigma^2}{2} \right) X_t$ and $\beta(t, X) = \sigma X_t$. We set $\widetilde{\alpha}(t, X) \equiv 0$. By (54),

$$Z_t = \exp\left\{ \left(\frac{m}{\sigma} + \frac{\sigma}{2} \right) B_t - \frac{1}{2} \left(\frac{m}{\sigma} + \frac{\sigma}{2} \right)^2 t \right\}, \tag{57}$$

and if a measure has the differential $d\widetilde{\mathsf{P}}_T = Z_T\,d\widetilde{\mathsf{P}}$, then the process $X = (X_t)_{t \leqslant T}$ has the stochastic differential

$$dX_t = \sigma X_t\,d\widetilde{B}_t$$

with respect to $\widetilde{\mathsf{P}}_T$. In other words, the process $X = (X_t)_{t \leqslant T}$ is a *standard* geometric Brownian motion with respect to $\widetilde{\mathsf{P}}_T$ (see Chapter III, § 3a):

$$X_t = \exp\left\{ \sigma \widetilde{B}_t - \frac{\sigma^2}{2} t \right\}. \tag{58}$$

§ 3c. Construction of Martingale Measures for Lévy Processes. Esscher Transformation

1. If $X = (X_t)_{t \leqslant T}$ is a process of *diffusion* type with respect to the original measure P_T, with local characteristics $\alpha(t, X)$ and $\beta(t, X)$ (see formula (52) in § 2c), then Girsanov's theorem suggests an explicit construction of another measure $\widetilde{\mathsf{P}}_T$ such that X has local characteristics $\widetilde{\alpha}(t, X)$ and $\widetilde{\beta}(t, X)$ with respect to $\widetilde{\mathsf{P}}_T$. If now $\widetilde{\alpha}(t, X) \equiv 0$, then X is a local martingale with respect to $\widetilde{\mathsf{P}}_T$; one calls $\widetilde{\mathsf{P}}_T$ a *locally martingale* measure for that reason.

The construction of this measure proceeds by the formula

$$d\widetilde{\mathsf{P}}_T = Z_T \, d\mathsf{P}_T, \tag{1}$$

where Z_T is defined by equality (54) in the preceding section.

The Esscher transformation suggests another construction of a new measure, which is essentially based on the same idea. Namely, assume that the initial process $X = (X_t)_{t \leqslant T}$ has *independent increments* (e.g., is representable as the sum of independent random variables).

2. Recall that we encountered already the Esscher transformation and its generalization ('the conditional Esscher transformation') in Chapter V, § 2d (see, in particular, the remark in subsection 2). It should also be noted that, as a the method of the construction of 'risk-neutral' probability measures assigning 'larger weights to adverse events' and 'smaller weights to beneficial events' the Esscher transformation is known in actuarial practices since 1932, when F. Esscher's paper [144] was published.

For instance, insurance companies are based in their calculations of life-insurance premiums not on the (quite precisely known) distribution P_T of life expectancy ('mortality tables of the second kind'), but on another, different, distribution $\widetilde{\mathsf{P}}_T$ ('mortality tables of the first kind') that has the above property of shifting the balance between beneficiary and adverse events.

3. Before a discussion of the Esscher transformation in the general case we consider the following simple example (cf. Chapter V, § 2d), which is a good illustration of the true meaning of this transformation.

Let X be a real-valued random variable with Laplace transform $\Phi(\lambda) = \mathsf{E}e^{\lambda X} < \infty$, $\lambda \in \mathbb{R}$, and let $\mathsf{P} = \mathsf{P}(dx)$ be its probability distribution on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$.

We consider a *family* of probability measures $\mathsf{P}^{(a)}$, $a \in \mathbb{R}$, defined by means of the *Esscher transformation*:

$$\mathsf{P}^{(a)}(dx) = \frac{e^{ax}}{\Phi(a)} \, \mathsf{P}(dx). \tag{2}$$

Setting

$$Z^{(a)}(x) = \frac{e^{ax}}{\Phi(a)}, \tag{3}$$

we see that $Z^{(a)}(x) > 0$, $\mathsf{E}Z^{(a)}(X) = 1$, the measure $\mathsf{P}^{(a)}$ is equivalent to P, and

$$\mathsf{P}^{(a)}(dx) = Z^{(a)}(x)\,\mathsf{P}(dx). \tag{4}$$

It is also clear that

$$\Phi^{(a)}(\lambda) = \mathsf{E}_{\mathsf{P}^{(a)}} e^{\lambda X} = \frac{\mathsf{E}e^{(\lambda+a)X}}{\Phi(a)} = \frac{\Phi(a+\lambda)}{\Phi(a)}, \tag{5}$$

and therefore

$$\mathsf{E}_{\mathsf{P}^{(a)}} X = \frac{\partial \Phi^{(a)}(\lambda)}{\partial \lambda}\bigg|_{\lambda=0} = \frac{\Phi'(a)}{\Phi(a)}. \tag{6}$$

We showed in Chapter V, § 2d that if a random variable X has the properties $\mathsf{P}(X > 0) > 0$ and $\mathsf{P}(X < 0) > 0$, then the function $\Phi(a)$ attains its maximum at some point $\widetilde{a}$, where, obviously, $\Phi'(\widetilde{a}) = 0$.

Hence we obtain the expectation $\widetilde{\mathsf{E}}X \equiv \mathsf{E}_{\mathsf{P}^{(\widetilde{a})}}X = 0$ with respect to the measure $\widetilde{\mathsf{P}} = \mathsf{P}^{(\widetilde{a})}$, which one sometimes expresses by saying that $\widetilde{\mathsf{P}}$ is a 'risk-neutral' probability measure.

We can consider the property $\widetilde{\mathsf{E}}X = 0$ also as a 'single-step' version of the martingale property, which explains why one also calls $\widetilde{\mathsf{P}}$ a *martingale measure.*

4. Now let $X = (X_t)_{t \leqslant T}$ be a Lévy process on $(\Omega, \mathscr{F}_T, \mathsf{P}_T)$ with characteristic function (see (27) in § 3a here and Chapter III, § 1b)

$$\mathsf{E}e^{i\theta X_t} = e^{t\psi(\theta)}, \tag{7}$$

where the *cumulant* is

$$\psi(\theta) = i\theta b - \frac{\theta^2}{2}c + \int \bigl(e^{i\theta x} - 1 - i\theta g(x)\bigr)\,\nu(dx) \tag{8}$$

and $g(x)$ is a truncation function (e.g., $g(x) = xI(|x| \leqslant 1)$).

From (7) and (8), setting formally $\theta = -i\lambda$ we see that

$$\mathsf{E}e^{\lambda X_t} = e^{t\varphi(\lambda)}, \tag{9}$$

where

$$\varphi(\lambda) = \lambda b + \frac{\lambda^2}{2}c + \int_{\mathbb{R}} \bigl(e^{\lambda x} - 1 - \lambda g(x)\bigr)\,\nu(dx). \tag{10}$$

The easiest way to a rigorous proof of the representation (9)–(10) for the Laplace transform is based on the following observation: the process

$$Z^{(\lambda)} = \bigl(Z_t^{(\lambda)}\bigr)_{t \geqslant 0}$$

with

$$Z_t^{(\lambda)} = \exp\{\lambda X_t - t\varphi(\lambda)\} \tag{11}$$

is a martingale. (This is an immediate consequence of Itô's formula for semimartingales. Of course, one must also assume that the integral in (10) is finite.)

By analogy with (2) we introduce now for each $a \in \mathbb{R}$ the probability measure $\mathsf{P}_T^{(a)}$ defined by means of the Esscher transformation:

$$d\mathsf{P}_T^{(a)} = Z_T^{(a)}\, d\mathsf{P}_T. \tag{12}$$

THEOREM 1. *The process $X = (X_t)_{t\leqslant T}$ is also a Lévy process with respect to $\mathsf{P}_T^{(a)}$, $a \in \mathbb{R}$, which has the Laplace transform*

$$\mathsf{E}_{\mathsf{P}_T^{(a)}} e^{\lambda X_t} = e^{t\varphi^{(a)}(\lambda)}, \tag{13}$$

where

$$\varphi^{(a)}(\lambda) = \varphi(a+\lambda) - \varphi(a). \tag{14}$$

Proof. This is a consequence of Bayes's formula (Chapter V, § 3d), which states that ($\mathsf{P}_T^{(a)}$-a.s.)

$$\begin{aligned}
\mathsf{E}_{\mathsf{P}_T^{(a)}}\big(e^{\lambda(X_t-X_s)} \mid \mathscr{F}_s\big) &= \mathsf{E}_{\mathsf{P}_T^{(a)}}\left(e^{\lambda(X_t-X_s)}\frac{Z_t^{(a)}}{Z_s^{(a)}} \;\Big|\; \mathscr{F}_s\right) \\
&= \mathsf{E}\big(e^{(a+\lambda)(X_t-X_s)-\varphi(a)(t-s)} \mid \mathscr{F}_s\big) \\
&= \mathsf{E}e^{(a+\lambda)(X_t-X_s)-\varphi(a)(t-s)} \\
&= e^{(\varphi(a+\lambda)-\varphi(a))(t-s)}.
\end{aligned}$$

Hence if the measure $\mathsf{P}_T^{(a)}$ is defined by the Esscher transformation (12), then $X = (X_t)_{t\leqslant T}$ is a Lévy process also with respect to this measure and its Laplace transform is defined by (13) and (14). (Cf. Girsanov's theorem in § 3b.)

THEOREM 2. *The local characteristics $(b^{(a)}, c^{(a)}, \nu^{(a)})$ of the process $X = (X_t)_{t\leqslant T}$ with respect to the measure $\mathsf{P}_T^{(a)}$, $a \in \mathbb{R}$, can be determined from the local characteristics (b, c, ν) by the following formulas (where $g(x)$ is a truncation function):*

$$b^{(a)} = b + ac + \int_{\mathbb{R}} g(x)(e^{ax}-1)\,\nu(dx), \tag{15}$$

$$c^{(a)} = c, \tag{16}$$

$$\nu^{(a)}(dx) = e^{ax}\,\nu(dx). \tag{17}$$

Proof. By Theorem 1 the process X is a Lévy process with respect to $\mathsf{P}_T^{(a)}$ and

$$\varphi^{(a)}(\lambda) = \lambda b^{(a)} + \frac{\lambda^2}{2} c^{(a)} + \int_{\mathbb{R}} \left(e^{\lambda x} - 1 - \lambda g(x)\right) \nu^{(a)}(dx). \tag{18}$$

Bearing in mind that $\varphi^{(a)}(\lambda) = \varphi(a+\lambda) - \varphi(\lambda)$, $\lambda \in \mathbb{R}$, we immediately obtain 'transition formulas' (15)–(17) by (18) and (10).

5. Let $X = (X_t)_{t \leqslant T}$ be a Lévy process (with respect to the measure P_T). Since it is a semimartingale, this process has a (non-unique, in general) canonical decomposition $X = M + A$, where M is a local martingale and A is a process of bounded variation. We consider now the following question: what conditions on the local characteristics (b, c, ν) make X a *local martingale* (with respect to P_T).

We can express this otherwise: we are interested in the conditions making P_T a *martingale* measure for the process X.

First we must point out that if X is a *local martingale*, then it is a *special* semimartingale, and therefore, with necessity,

$$(x^2 \wedge |x|) * \nu \in \mathscr{A}_{\text{loc}} \tag{19}$$

(see (20) in § 3a).

Let X be a special semimartingale, let

$$X = N + A \tag{20}$$

be its *canonical decomposition* (where N is a local martingale and A is a predictable process of bounded variation), and let

$$X = B + X^c + g * (\mu - \nu) + (x - g) * \mu \tag{21}$$

be its *canonical representation*, which (in view of (19)) can be rewritten as follows:

$$X = B + X^c + g * (\mu - \nu) + (x - g) * (\mu - \nu) + (x - g) * \nu. \tag{22}$$

Comparing (20) and (22) we see that

$$A = B + (x - g) * \nu, \tag{23}$$

and therefore we can say that a *special semimartingale X with triplet of predictable characteristics* (B, C, ν) is a *local martingale* if and only if

$$B + (x - g) * \nu = 0. \tag{24}$$

All this shows that a Lévy process X with triplet of local characteristics (b, c, ν) is a *special semimartingale* if and only if

$$\int (x^2 \wedge |x|)\, \nu(dx) < \infty \tag{25}$$

(see (20) in § 3a), and it is, in addition, a *local martingale* if and only if

$$b + \int_{\mathbb{R}} (x - g(x))\, \nu(dx) = 0. \tag{26}$$

If $g(x) = xI(|x| \leqslant 1)$, then condition (26) assumes the following form:

$$b + \int_{|x|>1} x\, \nu(dx) = 0. \tag{27}$$

Of course, it may happen that (26) *breaks down* with respect to the initial measure P. Then one could consider the measures $\mathsf{P}_T^{(a)}$ constructed by means of the Esscher transformation. Since the 'new' local characteristics $(b^{(a)}, c^{(a)}, \nu^{(a)})$ can be defined by (15)–(17), it follows by (25) and (27) that a Lévy process is a local martingale with respect to the measure $\mathsf{P}_T^{(a)}$ if and only if

$$\int (x^2 \wedge |x|) e^{ax}\, \nu(dx) < \infty \tag{28}$$

and

$$b + ac + \int_{|x|>1} x\, \nu(dx) + \int_{\mathbb{R}} x(e^{ax} - 1)\, \nu(dx) = 0. \tag{29}$$

EXAMPLE 1. Let $X_t = mt + \sigma B_t + kN_t$, where B is a standard Brownian motion and N is a standard Poisson process with parameter $\nu > 0$ ($\mathsf{E}N_t = \nu t$). We now find a value of a such that $X = (X_t)_{t \leqslant T}$ is a local martingale with respect to the measure $\mathsf{P}_T^{(a)}$.

We represent X_t as follows:

$$X_t = (m + k\nu)t + \sigma B_t + k(N_t - \nu t), \tag{30}$$

and see that it is certainly a martingale with respect to the original measure, provided that

$$m + k\nu = 0. \tag{31}$$

Assume now that $m + k\nu \neq 0$. Then the process $(kN_t)_{t \geqslant 0}$ makes jumps of amplitude k, the Lévy measure is $\nu(dx) = \nu \cdot I_{\{k\}}(dx)$,

$$\mathsf{E}e^{\lambda(kN_t)} = e^{\nu t(e^{k\lambda} - 1)}, \tag{32}$$

and the local characteristics b and c of X (for the truncation function $g(x) = xI(|x| \leqslant k)$) have the following form:

$$b = m + k\nu,$$
$$c = \sigma^2.$$

By (29) we obtain the following equations for a (in view of our choice of the truncation function $g(x) = xI(|x| \leqslant k)$, integration over the set $\{x\colon |x| > 1\}$ is replaced by integration over $\{x\colon x > k\}$):

$$(m + k\nu) + a\sigma^2 + \nu k(e^{ak} - 1) = 0. \tag{33}$$

If $\sigma^2 \neq 0$, then this equation has a root $\widetilde{a}$, and therefore X becomes a martingale with respect to the measure $\mathsf{P}_T^{(\widetilde{a})}$. On the other hand, if $\sigma^2 = 0$, then the root $\widetilde{a}$ can be found from the equation

$$e^{ak} = -\frac{m}{k\nu}, \tag{34}$$

which is soluble if m is not zero and of distinct sign from k.

6. We assume now that the price process $S = (S_t)_{t \leqslant T}$ is generated by some Lévy process $X = (X_t)_{t \leqslant T}$:

$$S_t = e^{X_t}. \tag{35}$$

Below we consider that following question, which is important for the problem of the absence of arbitrage: is the process S a martingale with respect to the initial measure P_T or some measure $\mathsf{P}_T^{(a)}$ constructed by means of the Esscher transformation?

As in subsection 3, we start from a simple example.

Let X be a real-valued random variable and let $\Phi(a) = \mathsf{E}e^{aX}$. (We assume here that $\Phi(a) < \infty$, $a \in \mathbb{R}$.) Clearly, if $\Phi(1) = 1$, then the random variable $S = e^X$ has the 'martingale' property $\mathsf{E}S = 1$ (with respect to the original measure).

If $\Phi(1) \neq 1$, then we can look for $\widetilde{a}$ such that this 'martingale' property $\mathsf{E}_{\mathsf{P}_T^{(\widetilde{a})}} S = 1$ holds with respect to the measure $\mathsf{P}_T^{(\widetilde{a})}$ defined by the formula (2).

Since

$$\mathsf{E}_{\mathsf{P}_T^{(a)}} S = \mathsf{E}\frac{e^{(a+1)X}}{\Phi(a)} = \frac{\Phi(a+1)}{\Phi(a)}, \tag{36}$$

the value of $\widetilde{a}$ must be a root of the equation

$$\Phi(a+1) - \Phi(a) = 0. \tag{37}$$

For instance, if X is a normally distributed random variable with parameters m and σ^2, then

$$\Phi(a) = \mathsf{E}e^{aX} = e^{am+\frac{(a\sigma)^2}{2}}, \tag{38}$$

and we find from (37) that

$$\widetilde{a} = -\frac{1}{2} - \frac{m}{\sigma^2}. \tag{39}$$

We consider now the process $S = (S_t)_{t\leqslant T}$ defined in (35). The first question to ask here is about the conditions ensuring that this process is a martingale with respect to the original measure P_T.

THEOREM 3. *In order that the process $S = e^X$ be a martingale with respect to P_T it is sufficient (and also necessary) that*

$$\int_{|x|>1} e^x \, \nu(dx) < \infty \tag{40}$$

and

$$b + \frac{1}{2}c + \big(e^x - 1 - g(x)\big) * \nu = 0. \tag{41}$$

Proof. Condition (40) together with the inequality $(x^2 \wedge 1) * \nu < \infty$ ensures that the integral $(e^x - 1 - g(x)) * \nu$ is finite.

Clearly, the expectation

$$\mathsf{E}(e^{X_t - X_s} \mid \mathscr{F}_s) = \mathsf{E}e^{X_t - X_s} = e^{(t-s)\varphi(1)},$$

where $\varphi(1)$ is as in (10), is precisely the expression of the left-hand side of (41). Hence

$$\mathsf{E}(e^{X_t} \mid \mathscr{F}_s) = e^{X_s},$$

which proves that $S = e^X$ is a martingale. (The necessity of (40) and (41) is shown in [250; Chapter X, § 2a].)

7. Assume that (41) fails. Then by Theorem 1,

$$\mathsf{E}_{\mathsf{P}_T^{(a)}}(e^{X_t - X_s} \mid \mathscr{F}_s) = e^{(t-s)(\varphi(a+1)-\varphi(a))}. \tag{42}$$

Hence it is clear that if $\widetilde{a}$ is a root of the equation

$$\varphi(a+1) - \varphi(a) = 0, \tag{43}$$

then the process $S = (S_t)_{t\leqslant T}$ is a martingale with respect to the measure $\mathsf{P}_T^{(\widetilde{a})}$.

By (43), in view of (18), we obtain the following result.

THEOREM 4. *Assume that* $\widetilde{a}$ *satisfies the inequality*

$$\left|e^{\widetilde{a}x}(e^x - 1) - g(x)\right| * \nu < \infty$$

and

$$b + \left(\widetilde{a} + \frac{1}{2}\right)c + \left(e^{\widetilde{a}x}(e^x - 1) - g(x)\right) * \nu = 0. \tag{44}$$

Then the process $S = (S_t)_{t \leqslant T}$ *is a martingale with respect to the measure* $\mathsf{P}_T^{(\widetilde{a})}$.

EXAMPLE 2. Let $S_t = e^{X_t}$, where $X = (X_t)_{t \geqslant 0}$ is the process introduced in Example 1. Then equation (44) takes the following form:

$$\left(m + \frac{\sigma^2}{2}\right) + a\sigma^2 + \nu\left[e^{ak}(e^k - 1)\right] = 0. \tag{45}$$

If $\sigma = 0$ and $k \neq 0$, then (45) becomes the equation

$$e^{ak} = -\frac{m}{n(e^k - 1)}, \tag{46}$$

which certainly has a solution if $m \neq 0$ and k and m have distinct signs.

If $k = 0$, then

$$S_t = e^{mt + \sigma B_t} \tag{47}$$

is a standard geometric Brownian motion (Chapter III, § 4b). By Itô's formula

$$dS_t = S_t\left(\left(m + \frac{\sigma^2}{2}\right) dt + \sigma\, dB_t\right). \tag{48}$$

From (45) or (39) we find that $\widetilde{a} = -\dfrac{1}{2} - \dfrac{m}{\sigma^2}$ and

$$Z_T^{(\widetilde{a})} = \exp\left\{-\left(\frac{m}{\sigma^2} + \frac{\sigma}{2}\right)^2 T\right\}.$$

Since

$$\mathsf{E}_{\mathsf{P}_T^{(\widetilde{a})}} e^{\lambda X t} = \exp\left\{\frac{\sigma^2 t}{2}[-\lambda + \lambda^2]\right\},$$

it follows that

$$\mathrm{Law}(X_t \,|\, \mathsf{P}_T^{(\widetilde{a})}) = \mathscr{N}\left(-\frac{\sigma^2 t}{2}, \frac{\sigma^2 t}{2}\right).$$

This means that the process $(X_t)_{t \leqslant T}$ has the same distribution with respect to the measure $\mathsf{P}_T^{(\widetilde{a})}$ as $\left(\sigma W_t - \dfrac{\sigma^2}{2} t\right)_{t \leqslant T}$, where $W = (W_t)_{t \leqslant T}$ is a standard Wiener process (cf. the example at the end of § 3b), so that the process $S = (S_t)_{t \leqslant T}$ becomes a standard geometric Brownian motion.

Hence both constructions of a martingale measure based on the *Girsanov transformation* or the *Esscher transformation* bring us to the same results. (This is little surprise though, for the martingale measure is unique in this case and $X_t = mt + \sigma B_t$ is simultaneously a diffusion process and a process with independent increments.)

Remark. As regards the Esscher transformation and its applications to option pricing, see H. U. Gerber and E. S. W. Shiu [178], [179].

§ 3d. Predictable Criteria of the Martingale Property of Prices. I

1. We devote this and the next sections to *predictable* criteria (i.e., criteria in terms of the triplets of predictable characteristics of a process in question) ensuring that the prices, the semimartingales $S = (S_t)_{t\geqslant 0}$, are martingales or local martingales (with respect to the initial measure P or some measure $\widetilde{\mathsf{P}} \ll \mathsf{P}$; cf. Chapter V, § 3f for discrete time).

We start with the observation that different representations of prices can be convenient in different problems.

If $S = (S_t, \mathscr{F}_t)_{t\geqslant 0}$ is a semimartingale, then, by definition,

$$S_t = S_0 + a_t + m_t, \tag{1}$$

where $a = (a_t, \mathscr{F}_t)_{t\geqslant 0}$ is a process of bounded variation and $m = (m_t, \mathscr{F}_t)_{t\geqslant 0}$ is a local martingale. This decomposition is not uniquely defined. For instance, if $S = (S_t)_{t\geqslant 0}$ is a standard Poisson process ($S_0 = 0$ and $\mathsf{E}S_t = \lambda t$), then we can set either $a_t = S_t$, $m_t = 0$ or $a_t = \lambda t$, $m_t = S_t - \lambda t$ in (1).

Expansion (1) is 'additive'. However, if $S = (S_t)_{t\geqslant 0}$ is a special positive semimartingale and (1) is its decomposition with positive process $a = (a_t)_{t\geqslant 0}$, then, provided that $S_{t-} + \Delta a_t \neq 0$, S has the *multiplicative decomposition*

$$S_t = S_0 \mathscr{E}(\widehat{a})_t \mathscr{E}(\widehat{m})_t, \tag{2}$$

where

$$\mathscr{E}(g)_t = e^{g_t - \frac{1}{2}\langle g^c\rangle_t} \prod_{0<s\leqslant t} (1 + \Delta g_s) e^{-\Delta g_s} \tag{3}$$

is the stochastic exponential and the processes $\widehat{a} = (\widehat{a}_t)$ and $\widehat{m} = (\widehat{m}_t)$ can be defined by the formulas

$$\widehat{a}_t = \int_0^t \frac{da_u}{S_{u-}} \quad \text{and} \quad \widehat{m}_t = \int_0^t \frac{dm_u}{S_{u-} + \Delta a_u}\,. \tag{4}$$

Formula (2) can be easily established on the basis of Itô's formula; see the details in [304; Chapter 2, § 5].

The product of two stochastic exponentials in (2) can be, by *Yor's formula* (see (18) in Chapter III, § 3f), rewritten as a single exponential:

$$\mathscr{E}(\widehat{a})_t \mathscr{E}(\widehat{m})_t = \mathscr{E}(\widehat{H})_t, \tag{5}$$

where

$$\widehat{H}_t = \widehat{a}_t + \widehat{m}_t + [\widehat{a}, \widehat{m}]_t \tag{6}$$

and

$$[\widehat{a}, \widehat{m}]_t = \sum_{0<s\leqslant t} \Delta\widehat{a}_s \, \Delta\widehat{m}_s. \tag{7}$$

Hence we obtain the following representation for $(S_t)_{t\geqslant 0}$:

$$S_t = S_0\mathscr{E}(\widehat{H})_t, \tag{8}$$

which is very useful in the analysis of this process 'for the martingale property', because the stochastic exponential $\mathscr{E}(\widehat{H})$ is a local martingale if and only if $\widehat{H}$ is a local martingale.

We have already mentioned (Chapter II, § 1a) that, from the standpoint of *statistical analysis*, in place of (8) one could more conveniently use a representation of 'compound interest' type

$$S_t = S_0 e^{H_t} \tag{9}$$

with some semimartingale $H = (H_t)_{t\geqslant 0}$. The representation (9) is usually chosen in financial mathematics as the starting point, while the transition from (9) to (8) proceeds by the formula

$$\widehat{H}_t = H_t + \frac{1}{2}\langle H^c\rangle_t + \sum_{0<s\leqslant t}(e^{\Delta H_s} - 1 - \Delta H_s), \tag{10}$$

which can be written in the 'difference-differential' form as follows:

$$d\widehat{H}_t = dH_t + \frac{1}{2}d\langle H^c\rangle_t + \bigl(e^{\Delta H_t} - 1 - \Delta H_t\bigr). \tag{11}$$

To prove (10) we observe that, by Itô's formula for $f(H) = e^H$ we obtain

$$dS_t = S_{t-}\Bigl[dH_t + \frac{1}{2}d\langle H^c\rangle_t + (e^{\Delta H_t} - 1 - \Delta H_t)\Bigr]. \tag{12}$$

On the other hand, by (8) and the properties of stochastic exponential,

$$dS_t = S_{t-}\, d\widehat{H}_t. \tag{13}$$

Comparing (12) and (13) we arrive to formula (10).

Remark 1. The infinite sum in (10) is absolutely convergent (P-a.s.) since for each semimartingale H there exists (P-a.s.) only finitely many instants $s \leqslant t$ such that

$$|\Delta H_s| > \frac{1}{2} \quad \text{and} \quad \sum_{0<s\leqslant t}(\Delta H_s)^2 < \infty \quad \text{(P-a.s.)};$$

see Chapter II, § 5b. For the same reason the infinite product in the definition of the stochastic exponential (3) is also absolutely convergent.

2. Let H be a semimartingale and let

$$H = H_0 + B + H^c + g * (\mu - \nu) + \bigl(x - g(x)\bigr) * \mu \tag{14}$$

be its canonical representation (with respect to some truncation function $g = g(x)$; here $\mu = \mu^H$ is the jump measure of H and $\nu = \nu^H$ is its compensator; see § 3a).

By (10) and (14) we obtain the following representation for $\widehat{H}$:

$$\begin{aligned}\widehat{H} &= H + \frac{1}{2}\langle H^c\rangle + (e^x - 1 - x) * \mu \\ &= H_0 + B + H^c + \frac{1}{2}\langle H^c\rangle + g * (\mu - \nu) + \bigl(x - g(x)\bigr) * \mu + (e^x - 1 - x) * \mu.\end{aligned} \tag{15}$$

To transform the right-hand side of (15) we use the fact that $|W| * \mu \in \mathscr{A}^+_{\mathrm{loc}}$ if $|W| * \nu \in \mathscr{A}^+_{\mathrm{loc}}$ and, moreover,

$$W * (\mu - \nu) = W * \mu - W * \nu \tag{16}$$

(see § 3a).

Hence we see from (15) that if

$$\bigl(|x| I(|x| \leqslant 1) + e^x I(|x| > 1)\bigr) * \nu \in \mathscr{A}^+_{\mathrm{loc}}, \tag{17}$$

then

$$\widehat{H} = K + H_0 + H^c + (e^x - 1) * (\mu - \nu), \tag{18}$$

where $H^c + (e^x - 1) * (\mu - \nu) \in \mathscr{M}_{\mathrm{loc}}(\mathsf{P})$ and

$$K = B + \frac{1}{2}\langle H^c\rangle + \bigl(e^x - 1 - g(x)\bigr) * \nu.$$

Thus, the following result is a consequence of (18).

THEOREM. *Assume that condition (17) is satisfied. Then $\widehat{H} \in \mathscr{M}_{\mathrm{loc}}(\mathsf{P})$ and $S \in \mathscr{M}_{\mathrm{loc}}(\mathsf{P})$ if and only if*

$$K_t = 0 \quad (\mathsf{P}\text{-a.s.}), \qquad t > 0. \tag{19}$$

In that case the local martingale $\widehat{H}$ has the representation

$$\widehat{H} = H_0 + H^c + (e^x - 1) * (\mu - \nu). \tag{20}$$

EXAMPLE. Let H be a Lévy process with triplet (B, C, ν) of the following form:

$$B_t = bt, \qquad C_t = \sigma^2 t, \qquad \nu(dt, dx) = dt\, F(dx), \tag{21}$$

where $F = F(dx)$ is a measure such that $F(\{0\}) = 0$ and

$$\int (x^2 \wedge 1)\, F(dx) < \infty. \tag{22}$$

We can assume also a stronger version of (22):

$$\int \big(|x| I(|x| \leqslant 1) + e^x I(|x| > 1)\big)\, F(dx) < \infty. \tag{23}$$

Under this assumption the price process $S_t = S_0 e^{H_t}$ is a martingale (with respect to the initial measure P) if (b, σ^2, F) satisfies the following relation:

$$b + \frac{\sigma^2}{2} + \int \big(e^x - 1 - g(x)\big)\, F(dx) = 0. \tag{24}$$

If $B_t = B_0 e^{rt}$ is a bank account, then the discounted price process $\dfrac{S}{B} = \left(\dfrac{S_t}{B_t}\right)_{t \geqslant 0}$ is a martingale with respect to P if

$$b + \frac{\sigma^2}{2} + \int \big(e^x - 1 - g(x)\big)\, F(dx) = r. \tag{25}$$

Remark 2. In accordance with notation (10) in the preceding section the left-hand side of (24) and (25) is the value of the 'cumulant' function $\varphi(\lambda)$ for $\lambda = 1$. Hence formula (25) can be put in the following form:

$$\varphi(1) = r. \tag{26}$$

Remark 3. As Theorem 3 in § 3c shows, the condition $\int |x| I(|x| \leqslant 1)\, F(dx) < \infty$ is redundant for Lévy processes. (This condition is a result of the 'regrouping' in (15) based on formula (16).)

§ 3e. Predictable Criteria of the Martingale Property of Prices. II

1. Without assumption (19) (see the theorem in § 3d) the price process $S = S_0 e^H$ is not a local martingale with respect to the initial measure P.

However, it is sufficient in many cases (e.g., in the problem of the absence of arbitrage; see § 2b) that there exists *some* measure such that $\widetilde{\mathsf{P}} \overset{\text{loc}}{\ll} \mathsf{P}$ or $\widetilde{\mathsf{P}} \overset{\text{loc}}{\sim} \mathsf{P}$ and S is a local martingale with respect to $\widetilde{\mathsf{P}}$.

We have thoroughly discussed the question of the existence of such measures in models with discrete time (Chapter V, § § 3a–3f).

Below, we consider this question in the continuous-time case for semimartingale models.

2. Let $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geqslant 0}, \mathsf{P})$ be a stochastic basis and let $\mathsf{P}_t = \mathsf{P} \,|\, \mathscr{F}_t$ be the restriction of P to $\mathscr{F}_t$. Assume also that $\widetilde{\mathsf{P}}$ is a probability measure on $\mathscr{F}$ such that $\widetilde{\mathsf{P}} \overset{\text{loc}}{\ll} \mathsf{P}$, i.e., $\widetilde{\mathsf{P}}_t \ll \mathsf{P}_t$ for all $t \geqslant 0$. We also assume that $\mathscr{F}_0 = \{\varnothing, \Omega\}$ and $\widetilde{\mathsf{P}}_0 = \mathsf{P}_0$.

Our analysis of the existence of measures $\widetilde{\mathsf{P}}$ making one or another process a local martingale starts with *Girsanov's theorem for local martingales*, which demonstrates what happens to local martingales undergoing an absolutely continuous change of measure.

THEOREM 1. *Assume that $\widetilde{\mathsf{P}} \overset{\text{loc}}{\ll} \mathsf{P}$, let $M \in \mathscr{M}_{\text{loc}}(\mathsf{P})$ with $M_0 = 0$, and let $Z = (Z_t)_{t\geqslant 0}$, where $Z_t = \dfrac{d\widetilde{\mathsf{P}}_t}{d\mathsf{P}_t}$. Assume that the quadratic covariance $[M, Z]$ has a P-locally integrable variation and let $\langle M, Z\rangle$ be the predictable quadratic covariance (the compensator of $[M, Z]$).*

Then the process

$$\widetilde{M} = M - \frac{1}{Z_-} \cdot \langle M, Z\rangle, \tag{1}$$

is a local $\widetilde{\mathsf{P}}$-martingale and the $\widetilde{\mathsf{P}}$-characteristic $\langle \widetilde{M}^c, \widetilde{M}^c\rangle$ is the same ($\widetilde{\mathsf{P}}$-a.s.) as the P-characteristic $\langle M^c, M^c\rangle$.

Proof. In accordance with the lemma in Chapter V, § 3d,

$$XZ \in \mathscr{M} \iff X \in \mathscr{M}(\widetilde{\mathsf{P}}). \tag{2}$$

(We have stated and proved this lemma in the discrete-time case; this can be trivially extended to continuous time.)

From (2) we can easily derive the following local versions of this equivalence (see [250; Chapter III, § 3b] for the details):

$$XZ \in \mathscr{M}_{\text{loc}}(\mathsf{P}) \implies X \in \mathscr{M}_{\text{loc}}(\widetilde{\mathsf{P}}); \tag{3}$$

$$(XZ)^{T_n} \in \mathscr{M}_{\text{loc}}(\mathsf{P}) \implies X \in \mathscr{M}_{\text{loc}}(\widetilde{\mathsf{P}}), \tag{4}$$

where $(XZ)^{T_n} = (X_{t\wedge T_n} Z_{t\wedge T_n})_{t\geqslant 0}$ and $T_n = \inf(t\colon Z_t < 1/n)$.

Thus, to prove that $\widetilde{M} \in \mathscr{M}_{\text{loc}}(\widetilde{\mathsf{P}})$, it is sufficient to verify that $(\widetilde{M}Z)^{T_n} \in \mathscr{M}_{\text{loc}}(\mathsf{P})$, $n \geqslant 1$.

Let

$$A = \frac{1}{Z_-} \cdot \langle M, Z\rangle. \tag{5}$$

Then, by Itô's formula,

$$\begin{aligned}
(M - A)Z &= MZ - AZ \\
&= (M_- \cdot Z + Z_- \cdot M + [M, Z]) - (A \cdot Z + Z_- \cdot A) \\
&= \big(M_- \cdot Z + Z_- \cdot M + ([M, Z] - \langle M, Z\rangle)\big) + \langle M, Z\rangle - A \cdot Z - \langle M, Z\rangle \\
&= M_- \cdot Z + Z_- \cdot M + ([M, Z] - \langle M, Z\rangle) - A \cdot Z.
\end{aligned} \tag{6}$$

The first three terms on the right-hand side of (6) are local P-martingales. The same can be said for each $n \geqslant 1$ about the process $(A \cdot Z)^{T_n}$. Hence, by assertion (4) the process $\widetilde{M}$ belongs to $\mathscr{M}_{\mathrm{loc}}(\widetilde{\mathsf{P}})$.

Thus, M is a semimartingale with respect to the measure $\widetilde{\mathsf{P}}$, and it has the canonical decomposition

$$M = \widetilde{M} + A. \tag{7}$$

Hence by the definition of quadratic variations $[M, M]$ and $[\widetilde{M}, \widetilde{M}]$ and considering the limits as $n \to \infty$ of the Riemann sequences $S^{(n)}(M, M)$ and $S^{(n)}(\widetilde{M}, \widetilde{M})$ (see formula (10) in Chapter III, § 5b) we obtain that $[M, M] = [\widetilde{M}, \widetilde{M}]$ up to $\widetilde{\mathsf{P}}$-indistinguishability. Finally, in view of formula (22) in the same section, § 5b, of Chapter III, we conclude that the predictable quadratic variations $\langle M^c, M^c\rangle$ and $\langle \widetilde{M}^c, \widetilde{M}^c\rangle$ coincide (a.s. with respect to the measure $\widetilde{\mathsf{P}}$).

3. Let $S_t = S_0 e^{H_t}$, where $H = (H_t)_{t \geqslant 0}$ is a semimartingale, assume that $\widetilde{\mathsf{P}} \overset{\mathrm{loc}}{\ll} \mathsf{P}$, and let $Z_t = \dfrac{d\widetilde{\mathsf{P}}_t}{d\mathsf{P}_t}$.

Assume that the process $Z = (Z_t)_{t \geqslant 0}$ is generated by some P-local martingale $N = (N_t)_{t \geqslant 0}$:

$$dZ_t = Z_{t-}\, dN_t. \tag{8}$$

That is, let $Z = \mathscr{E}(N)$.

We represent S as the product $S = S_0 \mathscr{E}(\widehat{H})$, where $\widehat{H}$ can be found from H by the formula (10) in the preceding section.

Let $\widehat{H}$ be a special semimartingale with canonical decomposition

$$\widehat{H} = H_0 + \widehat{A} + \widehat{M}, \tag{9}$$

where $\widehat{M} \in \mathscr{M}_{\mathrm{loc}}(\mathsf{P})$ and $\widehat{A}$ is a predictable process of locally bounded variation.

We represent $\widehat{H}$ as follows:

$$\widehat{H} = H_0 + \widehat{A} + \widehat{M} = H_0 + \widehat{A} + \langle \widehat{M}, N\rangle + \bigl(\widehat{M} - \langle \widehat{M}, N\rangle\bigr), \tag{10}$$

and observe that

$$\frac{1}{Z_-} \cdot \langle \widehat{M}, Z\rangle = \langle \widehat{M}, N\rangle. \tag{11}$$

Then, by Theorem 1, the process $\widehat{M} - \langle \widehat{M}, N\rangle$ is a local martingale with respect to the measure $\widetilde{\mathsf{P}}$ with $d\widetilde{\mathsf{P}}_t = Z_t\, d\mathsf{P}_t$, $t \geqslant 0$.

Hence we obtain the following result.

THEOREM 2. *If $\widehat{H}$ is a special semimartingale with canonical decomposition* (9), $\widetilde{\mathsf{P}} \overset{\mathrm{loc}}{\ll} \mathsf{P}$, *and the process $Z = (Z_t)_{t \geqslant 0}$ has the representation* (8), *then*

$$\widehat{A} + \langle \widehat{M}, N\rangle = 0 \implies \widehat{H} \in \mathscr{M}_{\mathrm{loc}}(\widetilde{\mathsf{P}}). \tag{12}$$

4. Assume that condition (17) in the preceding section is satisfied and the process N has the following representation:

$$N = \beta \cdot H^c + (Y-1) * (\mu - \nu) \tag{13}$$

where $\beta = (\beta_t(\omega))_{t\geqslant 0}$ is a predictable process and $Y = Y(t,\omega,x)$, is a $\widetilde{\mathscr{P}}$-measurable function of $t \geqslant 0$, $\omega \in \Omega$, and $x \in \mathbb{R}$. (Here $\mu = \mu^H$, $\nu = \nu^H$, and we assume that the corresponding integrals with respect to H^c and $\mu-\nu$ are well defined.)

We use the representation (18) in § 3d for $\widehat{H}$:

$$\widehat{H} = H_0 + K + H^c + (e^x - 1) * (\mu - \nu). \tag{14}$$

If $\nu(\{t\} \times dx; \omega) = 0$, then we see that

$$\langle \widehat{M}, N \rangle = \beta \cdot \langle H^c \rangle + (Y-1)(e^x - 1) * \nu \tag{15}$$

(see the observation at the end of § 3a.4). By (12), (14), (15), and also relation (19) in § 3d we obtain the following result.

THEOREM 3. *Assume that the conditions* (17) *in* § 3d, $\nu(\{t\} \times dx; \omega) = 0$, *and*

$$|Y-1|\,|e^x - 1| * \nu_t < \infty \tag{16}$$

are satisfied. If, in addition,

$$B + \Big(\frac{1}{2} + \beta\Big)\langle H^c \rangle + (e^x - 1 - g(x)) * \nu + (e^x - 1)(Y-1) * \nu = 0, \tag{17}$$

then the processes $\widehat{H}$ *and* $S = S_0 \mathscr{E}(\widehat{H})$ *are local martingales with respect to the measure* $\widetilde{\mathsf{P}}$ *such that* $d\widetilde{\mathsf{P}}_t = Z_t\, d\mathsf{P}_t$, $t \geqslant 0$.

EXAMPLE. Let H be the Lévy process considered in the example of § 3d. Let $\beta_s(\omega) \equiv \beta$ and let $Y = Y(x)$. Assume also that

$$b + \Big(\frac{1}{2} + \beta\Big)\sigma^2 + \int \big(e^x - 1 - g(x)\big)\, F(dx) + \int (e^x - 1)(Y-1)\, F(dx) = 0. \tag{18}$$

Then the processes $\widehat{H}$ and $S = S_0 \mathscr{E}(\widehat{H})$ are $\widetilde{\mathsf{P}}$-local martingales.

Note that condition (18) can be also written as follows:

$$b + \Big(\frac{1}{2} + \beta\Big)\sigma^2 + \int \big((e^x - 1)Y - g(x)\big)\, F(dx) = 0. \tag{19}$$

For $\beta = 0$ and $Y = 1$ this is the same as condition (24) in § 3d.

We recall the representation for the cumulant function $\varphi(\lambda)$ in § 3c:

$$\varphi(\lambda) = \lambda b + \frac{\lambda^2}{2}\sigma^2 + \int \left(e^{\lambda x} - 1 - \lambda g(x)\right) F(dx). \tag{20}$$

Setting $\beta = \widetilde{\lambda}$ and $Y(x) = e^{\widetilde{\lambda} x}$, we find from (20) that (19) holds if $\widetilde{\lambda}$ is a root of the equation

$$\varphi(\lambda + 1) - \varphi(\lambda) = 0. \tag{21}$$

If $B_t = B_0 e^{rt}$, then the discounted prices $\dfrac{S}{B}$ form a local martingale with respect to the measure $\widetilde{\mathsf{P}}$ if $d\widetilde{\mathsf{P}}_t = Z_t \, d\mathsf{P}_t$ and $dZ_t = Z_{t-} \, dN_t$, where

$$N = \widetilde{\lambda} \cdot H^c + (e^{\widetilde{\lambda} x} - 1) * (\mu - \nu) \tag{22}$$

and $\widetilde{\lambda}$ is the root of the equation

$$\varphi(\lambda + 1) - \varphi(\lambda) = r.$$

§ 3f. Representability of Local Martingales ('$(H^c, \mu-\nu)$-Representability')

1. We assumed in the previous section that the density process $Z = (Z_t)_{t \geqslant 0}$ with $Z_t = \dfrac{d\widetilde{\mathsf{P}}_t}{d\mathsf{P}_t}$ (which is a P-local martingale) has a representation $Z = \mathscr{E}(N)$, where the P-local martingale N is a sum of two integrals, with respect to H^c and $\mu - \nu$ (see (13)).

Comparing this with the '$(\mu-\nu)$-representability' in the discrete-time case (Chapter V, S 4c) we see that the term '$(H^c, \mu-\nu)$-representability' fits very well in this context, so that we use it the title of this section.

The issue of the representability of local martingales is considered in full generality in [250; Chapter III, § 4c]. Hence we discuss here only several general results related directly to arbitrage, completeness, and the construction of probability measures that are locally absolutely continuous with respect to the original measure.

2. We note first of all that to answer in a satisfactory way the question on the representation of local martingales in terms of the local martingale H^c and the martingale measure $\mu - \nu$ we must impose certain additional restrictions on the structure of the space Ω of elementary outcomes ω. Namely, we shall assume in what follows that Ω is the *canonical space* of all right-continuous functions $\omega = (\omega_t)_{t \geqslant 0}$ that have limits from the left. (See also [250; Chapter III, 2.13] on this subject.)

We shall assume all the processes $X = (X_t(\omega))_{t \geqslant 0}$ below and, in particular, semimartingales to be canonical (i.e., $X_t(\omega) \equiv \omega_t$).

We shall take for a filtration $(\mathscr{F}_t)_{t\geqslant 0}$ the family of σ-algebras

$$\mathscr{F}_t = \bigcap_{s>t} \mathscr{F}_s^0,$$

where $\mathscr{F}_s^0 = \sigma(\omega\colon \omega_u,\ u \leqslant s)$. We also set $\mathscr{F} = \bigvee \mathscr{F}_t$.

Let P be a probability measure on $(\Omega, \mathscr{F})$, $\mathsf{P}_t = \mathsf{P}\,|\,\mathscr{F}_t$, $t \geqslant 0$, and let $H = (H_t, \mathscr{F}_t)_{t\geqslant 0}$ be a semimartingale with triplet of predictable characteristics (B, C, ν). For simplicity we assume that $H_0 = \mathrm{Const}$ (P-a.s.).

The question whether the triplet (B, C, ν) defines the measure P *unambigously* is of interest in many respects. This is not the case in general, as can be seen in the most simple, 'deterministic' examples.

For instance, let $H = (H_t)_{t\geqslant 0}$ be a solution of an (ordinary) differential equation

$$\dot{H}_t = 2|H_t|^{1/2}, \qquad H_0 = 0$$

(with non-Lipschitz right-hand side). Obviously, this equation has two solutions, $H_t^{(1)} \equiv 0$ and $H_t^{(2)} = t^2$. They are both semimartingales with respect to the measures $\mathsf{P}^{(1)}$ and $\mathsf{P}^{(2)}$, where the first is concentrated at the trajectory $\omega_t \equiv 0$ and the second at $\omega_t = t^2$. At the same time, their corresponding triplets (B, C, ν) are *the same*: $C = 0$, $\nu = 0$, and $B_t(\omega) = \int_0^t 2|\omega_s|^{1/2}\, ds$.

3. The role played by triplets and the uniqueness of probability measure in the problem of '$(H^c, \mu-\nu)$-representability' is revealed by the following result.

THEOREM 1. *Let $H = (H_t, \mathscr{F}_t)_{t\geqslant 0}$, $H_0 = \mathrm{Const}$, be a semimartingale with triplet (B, C, ν) on a filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geqslant 0}, \mathsf{P})$ and assume that the measure P is unique in the following sense: if P' is another measure such that H has the same triplet with respect to it, $\mathsf{P}' \overset{\mathrm{loc}}{\ll} \mathsf{P}$, and $\mathsf{P}'_0 = \mathsf{P}_0$, then $\mathsf{P}' = \mathsf{P}$.*

Then each local martingale $N = (N_t, \mathscr{F}_t)$ has a representation

$$N = N_0 + f \cdot H^c + W * (\mu - \nu), \tag{1}$$

where f is a predictable process with $f^2 \cdot \langle H^c \rangle \in \mathscr{A}_{\mathrm{loc}}^+$ and W is a $\widetilde{\mathsf{P}}$-predictable process with $G(W) \in \mathscr{A}_{\mathrm{loc}}^+$ (§ 3a).

The proof of this result and its generalization ('the *Fundamental representation theorem*') can be found in [250; Chapter III, § 4d].

One can deduce from this theorem the following results concerning '$(H^c, \mu-\nu)$-representability'. (They are useful, in particular, for complete arbitrage-free models)

THEOREM 2. *Let $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geqslant 0}, \mathsf{P})$ be the canonical filtered probability space.*

a) *If $H = (H_t, \mathscr{F}_t)_{t\geqslant 0}$ is a Brownian motion, then each local martingale $N = (N_t, \mathscr{F}_t)_{t\geqslant 0}$ has the following form:*

$$N = H_0 + f \cdot H, \tag{2}$$

where $f^2 \cdot \langle H\rangle \in \mathscr{A}^+_{\text{loc}}$.

b) *If a semimartingale $H = (H_t, \mathscr{F}_t)$ has independent increments, then each local martingale $N = (N_t, \mathscr{F}_t)_{t\geqslant 0}$ has a representation* (1).

The proof is an immediate consequence of Theorem 1 in view of the uniqueness of the Wiener measure and the following fact: processes with independent increments have deterministic triplets, which uniquely define the probability distributions (by the Lévy–Khintchine formula).

Note that we have already encountered assertion a) (in Chapter III, § 3c).

4. Apart from 'classical' cases a) and b) of Theorem 2, we shall now discuss briefly another case of the '$(H^c, \mu{-}\nu)$-representability' of local martingales.

We consider the stochastic differential equation

$$\begin{aligned} dH_t &= b(t, H_t)\,dt + \sigma(t, H_t)\,dB_t \\ &\quad + g\big(\delta(t, H_t, x)\big)\big(\mu(dt, dx; \omega) - \nu(dt, dx; \omega)\big) + g'\big(\delta(t, h_t, x)\big)\mu(dt, dx; \omega) \end{aligned} \tag{3}$$

(cf. Chapter III, § 3e), where b, σ, and δ are Borel functions, $g = g(x)$ is a truncation function, $g'(x) = x - g(x)$, B is a Brownian motion, and μ is homogeneous Poisson measure with compensator $\nu(dt, dx) = dt\,F(dx)$ (§ 3a). It is well known (see, e.g., [250; Chapter III, § 2c]), that for (locally) Lipschitz coefficients satisfying the condition of linear growth stochastic differential equation (3) (with initial condition $H_0 = \text{Const}$) has a unique strong solution (Chapter III, § 3c). Moreover, whatever the initial probability space on which both Brownian motion and Poisson measure are defined, the probability distribution of the solution process H on the canonical space $(\Omega, \mathscr{F})$ is uniquely defined.

The process H is a semimartingale with triplet (B, C, ν), where

$$\begin{aligned} B_t(\omega) &= \int_0^t b(s, \omega_s)\,ds, \\ C_t(\omega) &= \int_0^t \sigma^2(s, \omega_s)\,ds, \\ \nu(dt, dx; \omega) &= dt K_t(\omega_t, dx) \end{aligned}$$

and $K_t(\omega_t, A) = \int I_{A\setminus\{0\}}\big(\delta(t, \omega_t, x)\big)\,F(dx)$.

Hence, if the coefficients satisfy the above-mentioned conditions (the local Lipschitz condition and the condition of linear growth), then each local martingale $N = (N_t, \mathscr{F}_t)_{t\geqslant 0}$ admits an '$(H^c, \mu{-}\nu)$-representation'. (See [250; Chapter III, § 2a] for greater detail.)

§ 3g. Girsanov's Theorem for Semimartingales. Structure of the Densities of Probabilistic Measures

1. If $M = (M_t, \mathscr{F}_t)_{t\geqslant 0}$ is a local martingale on $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geqslant 0}, \mathsf{P})$ and $\widetilde{\mathsf{P}} \overset{\text{loc}}{\ll} \mathsf{P}$, then M is a semimartingale with respect to this measure (see (7) in § 3e).

Remarkably, each semimartingale is transformed into a semimartingale again by such a change. That is, the class of semimartingales is *stable* under locally continuous changes of measure. (This is an easy consequence of Itô's formula for semimartingales; see Chapter III, § 5c.)

From the standpoint of arbitrage the following question is of particular interest for financial mathematics: what can be said of the measures $\widetilde{\mathsf{P}}$ such that $\widetilde{\mathsf{P}} \overset{\text{loc}}{\sim} \mathsf{P}$ or $\widetilde{\mathsf{P}} \overset{\text{loc}}{\ll} \mathsf{P}$ and the semimartingale X in question (describing, e.g., the dynamics of prices) is a local martingale or a martingale with respect to $\widetilde{\mathsf{P}}$?

2. One possible approach to this problem is to describe how the canonical representation (with respect to P)

$$X = X_0 + B + X^c + g * (\mu - \nu) + \big(x - g(x)\big) * \mu \tag{1}$$

of a semimartingale X with triplet (B, C, ν) transforms under a locally continuous change of measure $\widetilde{\mathsf{P}} \overset{\text{loc}}{\ll} \mathsf{P}$ into the canonical representation

$$X = X_0 + \widetilde{B} + X^c + g * (\mu - \widetilde{\nu}) + \big(x - g(x)\big) * \mu \tag{2}$$

(with respect to $\widetilde{\mathsf{P}}$) of the same semimartingale with new triplet $(\widetilde{B}, \widetilde{C}, \widetilde{\nu})$.

Let $Z_t = \dfrac{d\widetilde{\mathsf{P}}_t}{d\mathsf{P}_t}$, $t \geqslant 0$. We set

$$\beta = \frac{d\langle Z^c, X^c\rangle}{d\langle X^c, X^c\rangle} \cdot \frac{I(Z_- > 0)}{Z_-}, \tag{3}$$

$$Y = \mathsf{E}^{\mathsf{P}}_{\mu}\left(\frac{Z}{Z_-} I(Z_- > 0) \,\Big|\, \widetilde{\mathscr{P}}\right), \tag{4}$$

where $\mathsf{E}^{\mathsf{P}}_{\mu}$ is averaging with respect to the measure M^{P}_{μ} on $(\Omega \times \mathbb{R}_+ \times E, \mathscr{F} \otimes \mathscr{B}(\mathbb{R}_+) \otimes \mathscr{E})$ defined by the formula $W * M^{\mathsf{P}}_{\mu} = \mathsf{E}(W * \mu)$ for all nonnegative measurable functions $W = W(\omega, t, x)$. (Cf. the definition of $Y_n(x, \omega)$ and $M_n(dx, d\omega)$ in Chapter V, § 3e).

The processes β and Y are crucial in the issue of the transformations that triplets undergo under changes of measure. The following result is often called *Girsanov's theorem for semimartingales.*

THEOREM 1. *Assume that* $\widetilde{\mathsf{P}} \overset{\text{loc}}{\ll} \mathsf{P}$, $Z_t = \dfrac{d\widetilde{\mathsf{P}}_t}{d\mathsf{P}_t}$, $t \geqslant 0$, *and let the processes* β *and* Y *be defined in terms of* $Z = (Z_t)_{t\geqslant 0}$ *by formulas* (3) *and* (4).

Then $\widetilde{B}$, $\widetilde{C}$, *and* $\widetilde{\nu}$, *defined by the equalities*

$$\widetilde{B} = B + \beta \cdot C + g(x)(Y-1) * \nu, \tag{5}$$

$$\widetilde{C} = C, \tag{6}$$

$$\widetilde{\nu} = Y \cdot \nu, \tag{7}$$

make up a triplet of the semimartingale X *with respect to the measure* $\widetilde{\mathsf{P}}$.

The proof of this result (not exclusively in the above case of one-dimensional semimartingales, but also for several dimensions) is presented in [250; Chapter III, § 3d] and is fairly complicated technically. Referring to this monograph for detail we comment now on the meaning of this theorem.

Note first of all that the corresponding result in the discrete-time case was proved in Chapter V, § 3e, where we explained the meaning of the discrete (relative to time) analogs of the measure M_μ^{P} and the variable Y.

Assertion (5) displays the transformation of the 'drift' component B in the triplet (B, C, ν).

Assertion (6) says that the quadratic characteristics of the continuous martingale component X^c do not change in fact under an absolutely continuous change of measure (up to $\widetilde{\mathsf{P}}$-stochastic equivalence).

Assertion (7) means that Y is just the Radon–Nikodym derivative of $\widetilde{\nu}$ with respect to ν.

3. If X is a special semimartingale, then we can set $g(x) = x$ in the canonical representation (2), so that

$$X = X_0 + B + X^c + x * (\mu - \nu). \tag{8}$$

Hence we see that X is a local martingale if $B \equiv 0$. Taken together with Theorem 1 this observation brings us to the following result.

THEOREM 2. *Let* $\widetilde{\mathsf{P}} \overset{\text{loc}}{\ll} \mathsf{P}$ *and, moreover,* $(x^2 \wedge |x|) * \widetilde{\nu} \in \mathscr{A}_{\text{loc}}^+$. *Then a special semimartingale* X *is a local martingale with respect to the measure* $\widetilde{\mathsf{P}}$ *if*

$$B + \beta \cdot C + x(Y-1) * \nu \equiv 0. \tag{9}$$

4. Formulas (3) and (4) show the way for finding β and Y once the process $Z = (Z_t)_{t\geqslant 0}$ is known. The converse question comes naturally: how, knowing β and Y, can one find the corresponding process Z?

A solution of this problem opens a way to the construction of the measure $\widetilde{\mathsf{P}}$ such that X is a local martingale with respect to it. For if β and Y satisfy (9) and we reconstruct the corresponding process Z, then, of course, taking the measure $\widetilde{\mathsf{P}}$ with $d\widetilde{\mathsf{P}}_T = Z_T\, d\mathsf{P}_T$ we see that $X = (X_t)_{t\leqslant T}$ is a local martingale on $[0,T]$.

Let X be a semimartingale defined on the canonical space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geqslant 0}, \mathsf{P})$. Assume that each P-martingale M admits an '$(X^c, \mu-\nu)$-representation':

$$M = M_0 + f \cdot X^c + W * (\mu - \nu). \tag{10}$$

(See formula (1) in § 3f.)

THEOREM 3. *Let* $\widetilde{\mathsf{P}} \overset{\text{loc}}{\ll} \mathsf{P}$, *let* $Z = (Z_t)_{t\geqslant 0}$ *be the density process, let* $\nu(\{t\} \times E;\omega) = 0$ *for* $t > 0$, *and let* β *and* Y *be defined by* (3) *and* (4). *Then (given the property of '$(X^c, \mu-\nu)$-representability') the process* Z *satisfies the relation*

$$Z = Z_0 + (Z_-\beta) \cdot X^c + Z_-(Y-1) * (\mu - \nu). \tag{11}$$

If

$$\beta^2 \cdot \langle X^c\rangle_t + (1 - \sqrt{Y}\,)^2 * \nu_t < \infty \tag{12}$$

for all $t > 0$, *then the process* $N = (N_t)_{t\geqslant 0}$ *with*

$$N_t = \beta \cdot X_t^c + (Y-1) * (\mu - \nu)_t \tag{13}$$

is a P-*local martingale. The process* $Z = (Z_t)_{t\geqslant 0}$ *is a solution of Doléans's equation*

$$dZ = Z_-\, dN \tag{14}$$

and can be represented in the following form:

$$Z_t = Z_0 \mathscr{E}(N)_t, \tag{15}$$

where

$$\mathscr{E}(N)_t = e^{N_t - \frac{1}{2}(\beta^2 C)_t} \prod_{0<s\leqslant t} (1 + \Delta N_s) e^{-\Delta N_s}. \tag{16}$$

In this statement we assume that $\nu(\{t\} \times E;\omega) = 0$. This means that the process X is *quasi-left-continuous*, i.e., for each *predictable* stopping time τ we have $\Delta X_\tau = 0$ on the set $\{\tau < \infty\}$. In the general case this theorem is stated and proved in [250; Chapter III, § 5a].

Remark. As regards direct applications of Theorems 2 and 3 to diffusion models, see the next section, § 4a.

4. Arbitrage, Completeness, and Hedge Pricing in Diffusion Models of Stock

§ 4a. Arbitrage and Conditions of Its Absence. Completeness

1. We have already discussed at length the issue of the construction of probability measures making price processes martingales or local martingales (both in the discrete and the continuous-time case). Our interest in this issue is mainly a consequence of the fact that the *existence* of equivalent martingale measures allows one to say about the absence of opportunities for arbitrage in a fairly general context (see § § 2b, c). In addition, the knowledge of *all* such measures enables one, for instance, to find the fair (rational) prices or hedging strategies, and so on, using the machinery of martingales.

In the present section we consider the issue of the absence of arbitrage in the case when prices are *Itô processes* (Chapter III, § 3d).

2. Let $(\Omega, \mathscr{F}, \mathsf{P})$ be a probability space with a Brownian motion $B = (B_t)_{t \geqslant 0}$. We shall denote by $(\mathscr{F}_t)_{t\geqslant 0}$ the *Brownian* (*Wiener*) *filtration*, i.e., the flow of σ-algebras $\mathscr{F}_t = \sigma(\mathscr{F}_t^0 \cup \mathscr{N})$, where $\mathscr{F}_t^0 = \sigma(B_s, s \leqslant t)$ and $\mathscr{N} = \{A \in \mathscr{F} \colon \mathsf{P}(A) = 0\}$. (See Chapter III, § 3a for detail). In addition, we assume that $\mathscr{F} = \bigvee \mathscr{F}_t$ $\big(\equiv \sigma\big(\bigcup \mathscr{F}_t\big)\big)$.

The filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geqslant 0}, \mathsf{P})$ satisfies the *usual conditions* (Chapter III, § 3a) and we shall regard it as the *stochastic basis* describing the probabilistic uncertainty and the structure of the flow of the incoming information.

Let $S_t = S_0 e^{H_t}$ $(S_0 > 0)$ be the price process for an asset, (some stock, say) with

$$H_t = \int_0^t \left(\mu_s - \frac{\sigma_s^2}{2}\right) dx + \int_0^t \sigma_s \, dB_s, \tag{1}$$

where $\mu = (\mu_t, \mathscr{F}_t)$ and $\sigma = (\sigma_t, \mathscr{F}_t)$ are two stochastic processes satisfying (P-a.s.) the conditions

$$\int_0^t |\mu_s| \, ds < \infty, \qquad \int_0^t \sigma_s^2 \, ds < \infty, \qquad t > 0. \tag{2}$$

By Itô's formula we obtain

$$dS_t = S_t \, d\widehat{H}_t, \tag{3}$$

where

$$\widehat{H}_t = \int_0^t \mu_s \, ds + \int_0^t \sigma_s \, dB_s, \tag{4}$$

i.e., S has the stochastic differential

$$dS_t = S_t(\mu_t \, dt + \sigma_t \, dB_t). \tag{5}$$

If $\mu_t \equiv \mu$ and $\sigma_t \equiv \sigma \neq 0$, then we obtain the *standard diffusion model* of Samuelson [420] describing the dynamics of stock prices by means of a *geometric Brownian motion* (Chapter III, § 4b):

$$dS_t = S_t(\mu \, dt + \sigma \, dB_t). \tag{6}$$

We set

$$Z_t = \exp\Bigl(-\frac{\mu}{\sigma} B_t - \frac{1}{2}\Bigl(\frac{\mu}{\sigma}\Bigr)^2 t\Bigr). \tag{7}$$

Then $\mathsf{E} Z_t = 1$ and by *Girsanov's theorem* (see Chapter III, § 3b or § 3e), for each $T > 0$ the process $S = (S_t, \mathscr{F}_t)_{t \leqslant T}$ is a martingale with respect to the measure $\widetilde{\mathsf{P}}_T$ with

$$d\widetilde{\mathsf{P}}_T = Z_T \, d\mathsf{P}_T, \tag{8}$$

where $\mathsf{P}_T = \mathsf{P} \,|\, \mathscr{F}_T$; its differential is

$$dS_t = \sigma S_t \, d\widetilde{B}_t, \tag{9}$$

where $\widetilde{B} = (\widetilde{B}_t, \mathscr{F}_t)_{t \leqslant T}$ is a standard Brownian motion with respect to $\widetilde{\mathsf{P}}_T$.

Thus, if $\mathsf{E} Z_T = 1$ then the measure $\widetilde{\mathsf{P}}_T$ on $(\Omega, \mathscr{F}_T)$ is equivalent to P_T and the process $S = (S_t, \mathscr{F}_t)_{t \leqslant T}$ becomes a martingale with respect to $\widetilde{\mathsf{P}}_T$.

It is worth noting that this measure $\widetilde{\mathsf{P}}_T$ is *unique* in the following sense: if Q_T is another measure such that $\mathsf{Q}_T \sim \mathsf{P}_T$ and $S = (S_t, \mathscr{F}_t)_{t \leqslant T}$ is a *local martingale* with respect to Q_T, then $\mathsf{Q}_T = \widetilde{\mathsf{P}}_T$. This is intimately connected with the result on the representation of local martingales in terms of the Brownian filtration (see Chapter III, § 3c); we shall prove it below, in subsection 5.

3. We consider now the case when the price process S has the differential (5). Assume that the following conditions are satisfied: P-a.s. we have

$$\sigma_t > 0 \tag{10}$$

for $t > 0$ and

$$\int_0^t \left(\frac{\mu_s}{\sigma_s}\right)^2 ds < \infty. \tag{11}$$

Then there exists a well-defined process $Z = (Z_t)_{t \geqslant 0}$ with

$$Z_t = \exp\left(-\int_0^t \frac{\mu_s}{\sigma_s}\, dB_s - \frac{1}{2}\int_0^t \left(\frac{\mu_s}{\sigma_s}\right)^2 ds\right), \tag{12}$$

that is a *positive local martingale*, and the corresponding localizing sequence $(\tau_n)_{n \geqslant 1}$ can be taken in the form

$$\tau_n = \inf\left\{t\colon \int_0^t \left(\frac{\mu_s}{\sigma_s}\right)^2 ds \geqslant n\right\}. \tag{13}$$

If $\mathsf{E} Z_T = 1$, then $Z = (Z_t)_{t \leqslant T}$ is a martingale, the measure $\widetilde{\mathsf{P}}_T$ such that $d\widetilde{\mathsf{P}}_T = Z_T\, d\mathsf{P}_T$ is a *probability* measure, $\widetilde{\mathsf{P}}_T \sim \mathsf{P}_T$, and the process $S = (S_t, \mathscr{F}_t)_{t \leqslant T}$ is a *local martingale* with respect to this measure.

To prove the last assertion we use Theorem 2 in § 3g.

Since

$$dZ_t = -Z_t \frac{\mu_t}{\sigma_t}\, dB_t, \tag{14}$$

if follows that (see formula (3) in § 3g)

$$\beta_t = \frac{d\langle Z^c, S^c\rangle}{d\langle S^c, S^c\rangle} \cdot Z_t^{-1} = -\frac{\frac{\mu_t}{\sigma_t} \cdot S_t\sigma_t}{(S_t\sigma_t)^2} = -\frac{\mu_t}{\sigma_t} \cdot \frac{1}{S_t\sigma_t}. \tag{15}$$

In the triplet (B, C, ν) of the semimartingale S with respect to the measure P we have $\nu \equiv 0$,

$$B_t = \int_0^t S_u\mu_u\, du, \quad \text{and} \quad C_t = \int_0^t S_u^2\sigma_u^2\, du. \tag{16}$$

However,

$$B_t + \int_0^t \beta_u\, dC_u = \int_0^t \left[S_u\mu_u - \frac{\mu_u S_u^2\sigma_u^2}{\sigma_u \cdot S_u\sigma_u}\right] du = 0.$$

Hence, by Theorem 2 in § 3g the process $S = (S_t, \mathscr{F}_t)_{t \leqslant T}$ is a local martingale with respect to $\widetilde{\mathsf{P}}_T$. Moreover (assuming that $\mathsf{E} Z_T = 1$), the measure $\widetilde{\mathsf{P}}_T$ is unique in the same sense as in the case of $\mu_t \equiv \mu$ and $\sigma_t \equiv \sigma$ (see subsection 5 below).

4. Now, for a market model consisting of a bank account $B(0) = (B_t(0))_{t\geqslant 0}$ with $B_t(0) \equiv 1$ and stock $S = (S_t)_{t\geqslant 0}$ with dynamics described by (5), we state conditions ensuring the absence of arbitrage (in its $\overline{NA}_+$-version; see § 2a).

Let $\pi = (\beta, \gamma)$ be a strategy and let $X^\pi = (X_t^\pi)_{t\geqslant 0}$ be its value,

$$X_t^\pi = \beta_t + \gamma_t S_t.$$

If π is a self-financing strategy, then

$$X_t^\pi = X_0^\pi + \int_0^t \gamma_u \, dS_u, \tag{17}$$

which, of course, implies that the stochastic integral in (17) must be well defined.

As is clear from § 1a, the integral in (17) is well defined for $\gamma \in L(S)$. In the current model (5) it would be reasonable to state conditions of the integrability of γ with respect to S directly, in terms of the processes $(\mu_t)_{t\leqslant T}$ and $(\sigma_t)_{t\leqslant T}$.

We assume that conditions (2), (10), and (11) are satisfied and

$$\int_0^t \gamma_u^2 \sigma_u^2 \, du < \infty \quad (\mathsf{P}\text{-a.s.}), \qquad t > 0. \tag{18}$$

By the last condition and since $S = (S_t)_{t\geqslant 0}$ is continuous, it follows that $\int_0^t S_u^2 \gamma_u^2 \sigma_u^2 \, du < \infty$ (P-a.s.), and therefore, the stochastic integral $\int_0^t S_u \gamma_u \sigma_u \, dB_u$ with respect to a Brownian motion is well defined (see Chapter III, § 3c and § 1a in the present chapter).

Further,

$$\left(\int_0^t |\gamma_u \mu_u| \, du\right)^2 \leqslant \int_0^t (\gamma_u \sigma_u)^2 \, du \cdot \int_0^t \left(\frac{\mu_u}{\sigma_u}\right)^2 du.$$

Hence it follows from (11) and (18) that the integrals $\int_0^t \gamma_u \mu_u \, du$ and $\int_0^t S_u \gamma_u \mu_u \, du$, $t > 0$, are well-defined and finite (P-a.s.).

Thus, conditions (2), (10), (11), and (18) ensure the existence of the stochastic integral in (17).

The next result, which is an immediate consequence of the implication (9) in Theorem 2 in § 2b is the best known assertion concerning the absence of arbitrage in diffusion models.

THEOREM. *Assume that stock prices $S = (S_t)_{t\geqslant 0}$ have the differential* (5) *and conditions* (2), (10), (11), *and* (18) *hold for $t \leqslant T$.*

Let $\mathsf{E} Z_T = 1$. Then the property $\overline{NA}_+$ holds and, in particular, there exist no opportunities for arbitrage in the class of a-admissible self-financing strategies for any $a \geqslant 0$.

5. We turn now to a result already mentioned in subsection 2: the measure $\widetilde{\mathsf{P}}_T$ such that $d\widetilde{\mathsf{P}}_T = Z_T\, d\mathsf{P}_T$, where Z_T is defined in (7), is *unique* in the following sense: this is a unique measure equivalent to P_T such that the process $S = (S_t, \mathscr{F}_t)_{t\leqslant T}$ is a local martingale with respect to $\widetilde{\mathsf{P}}_T$.

We shall consider a more general case, where we assume that S is as in (5) and the process Z is defined by (12).

Let Q_T be a measure *equivalent* to P_T such that $S = (S_t, \mathscr{F}_t)_{t\leqslant T}$ is a local martingale with respect to Q_T.

Then we construct the martingale

$$N_t = \mathsf{E}\biggl(\frac{d\mathsf{Q}_T}{d\mathsf{P}_T}\;\Big|\;\mathscr{F}_t\biggr), \qquad t \leqslant T. \tag{19}$$

It is *positive*, therefore there exists a process $\varphi = (\varphi_t, \mathscr{F}_t)_{t\leqslant T}$ such that

$$N_t = \exp\biggl(\int_0^t \varphi_s\, dB_s - \frac{1}{2}\int_0^t \varphi_s^2\, ds\biggr), \qquad t \leqslant T, \tag{20}$$

with $\int_0^T \varphi_s^2\, ds < \infty$ (P-a.s.), and $\mathsf{E}N_T = 1$ (see (20) in Chapter III, § 3c).

We now use Theorem 1 in § 3g, which describes the transformation of the triplet of predictable characteristics of semimartingales under absolutely continuous changes of measure.

Let $(B^{\mathsf{P}}, C^{\mathsf{P}}, \nu^{\mathsf{P}})$ be the triplet of S with respect to P. It follows from (5) that

$$B_t^{\mathsf{P}} = \int_0^t S_u \mu_u\, du, \qquad C_t^{\mathsf{P}} = \int_0^t S_u^2 \sigma_u^2\, du, \quad \text{and} \quad \nu^{\mathsf{P}} \equiv 0. \tag{21}$$

By Theorem 1 in § 3g the triplet $(B^{\mathsf{Q}}, C^{\mathsf{Q}}, \nu^{\mathsf{Q}})$ (with respect to Q) is as follows:

$$\begin{aligned} B_t^{\mathsf{Q}} &= B_t^{\mathsf{P}} + \int_0^t \beta_u\, dC^{\mathsf{P}}, \\ C_t^{\mathsf{Q}} &= C_t^{\mathsf{P}}, \\ \nu^{\mathsf{Q}} &\equiv 0, \end{aligned} \tag{22}$$

where

$$\beta_t = \frac{d\langle N^c, S^c\rangle_t}{d\langle S^c, S^c\rangle_t}\cdot\frac{1}{N_t} = \frac{\varphi_t}{S_t\sigma_t}. \tag{23}$$

Hence, from (21) and (22) we obtain

$$B_t^{\mathsf{Q}} = \int_0^t S_u[\mu_u + \varphi_u\sigma_u]\, du. \tag{24}$$

Since $S = (S_t, \mathscr{F}_t)_{t \leqslant T}$ is a local martingale with respect to Q_T, it follows that $B_t^{\mathsf{Q}} = 0$ (P-a.s.) for $t \leqslant T$. Hence we see from (24) that

$$\varphi_u(\omega) = -\frac{\mu_u(\omega)}{\sigma_u(\omega)}$$

$(\lambda \times \mathsf{P}_T)$-a.s. on $[0,T] \times \Omega$, where λ is Lebesgue measure.

Hence the processes $Z = (Z_t)_{t \leqslant T}$ and $N = (N_t)_{t \leqslant T}$ are stochastically indistinguishable, which shows that there exists a unique measure $\widetilde{\mathsf{P}}_T \sim \mathsf{P}_T$ such that $S = (S_t, \mathscr{F}_t)_{t \leqslant T}$ is a $\widetilde{\mathsf{P}}_T$-local martingale.

6. We discuss now the issue of T-completeness (see the definition in § 2d). Assume that the assumptions of the above theorem are satisfied. Assume also that for our $(B(0), S)$-market we have $B_t(0) \equiv 1$ and the process $S = (S_t)_{t \leqslant T}$ is a martingale with respect to the measure $d\widetilde{\mathsf{P}}_T = Z_T \, d\mathsf{P}_T$. Then by the uniqueness property of the (locally martingale) measure $\widetilde{\mathsf{P}}_T$ established above and in accordance with the theorem in § 2d, our diffusion model is T-complete.

A classical example of a T-complete (and arbitrage-free, in the $\overline{NA}_+$ and $\overline{NA}_g$-versions) model is, of course, the model of a geometric Brownian motion (6), which is a major factor in its popularity in financial mathematics and financial engineering.

§ 4b. Price of Hedging in Complete Markets

1. We explained the notion of hedging and methods of 'hedge pricing' in complete and incomplete markets for discrete-time case in Chapter VI.

In general semimartingale models the discussion can proceed along a parallel route, except, maybe, that we must describe concisely the classes of *admissible* strategies.

We shall stick to the diffusion model of a $(B(0), S)$-market described in § 4a.4 and use the notation from there.

Modifying slightly the definition of T-completeness given above (§ 2d) we shall say that a nonnegative $\mathscr{F}_T$-measurable pay-off function f_T with $\mathsf{E} Z_T f_T < \infty$ can be *replicated* if there exists a strategy $\pi \in \Pi_+(S)$ such that $X_T^\pi = f_T$ (P-a.s.). Clearly, if f_T is bounded, then the condition $\mathsf{E} Z_T f_T < \infty$ is satisfied.

DEFINITION. If a pay-off function f_T can be replicated, then we mean by the *price* of (perfect) *European hedging* (cf. the nomenclature in Chapter IV, § 1b), or simply the *hedging price*, the quantity

$$\mathbb{C}(f_T; \mathsf{P}) = \inf\{x \geqslant 0 \colon \exists\, \pi \in \Pi_+(S) \text{ with } X_0^\pi = x,\ X_T^\pi = f_T\}. \tag{1}$$

2. THEOREM. *Let $\widetilde{\mathsf{P}}_T$ be a unique martingale measure. Then*

$$\mathbb{C}(f_T; \mathsf{P}) = \mathsf{E}_{\widetilde{\mathsf{P}}_T} f_T \quad \big(= \mathsf{E} Z_T f_T \big). \tag{2}$$

Proof. If $\pi = (\beta, \gamma)$ is an a-admissible self-financing strategy, then

$$X_t^\pi = X_0^\pi + \int_0^t \gamma_u \, dS_u, \qquad t \leqslant T, \tag{3}$$

and (by the Ansel–Stricker theorem; see § 1a.6) $X^\pi = (X_t^\pi)_{t \leqslant T}$ is a $\widetilde{\mathsf{P}}_T$-supermartingale, so that

$$\mathsf{E}_{\widetilde{\mathsf{P}}_T} X_T^\pi \leqslant X_0^\pi. \tag{4}$$

Hence if $X_T^\pi = f_T$, then $\mathsf{E}_{\widetilde{\mathsf{P}}_T} f_T \leqslant X_0^\pi$ and

$$\mathsf{E} Z_T f_T = \mathsf{E}_{\widetilde{\mathsf{P}}_T} f_T \leqslant \mathbb{C}(f_T; \mathsf{P}). \tag{5}$$

We claim now that there exists a 0-admissible self-financing strategy $\widetilde{\pi}$ of initial value $X_0^{\widetilde{\pi}} = \mathsf{E} Z_T f_T$ that replicates f_T, i.e., $X_T^{\widetilde{\pi}} = f_T$ (P-a.s.).

We consider the process

$$X_t = \mathsf{E}(Z_T f_T \,|\, \mathscr{F}_t), \qquad t \leqslant T. \tag{6}$$

Clearly, $X = (X_t, \mathscr{F}_t)_{t \leqslant T}$ is a martingale with respect to the 'Brownian filtration' and by the representation theorem (see Chapter III, § 3c) there exists a process $\psi = (\psi_t, \mathscr{F}_t)_{t \leqslant T}$ with $\int_0^T \psi_s^2 \, ds < \infty$ such that

$$X_t = X_0 + \int_0^t \psi_s \, dB_s. \tag{7}$$

Note that the process $(Z_t^{-1} X_t)_{t \leqslant T}$ *replicates* f_T:

$$Z_T^{-1} X_T = f_T. \tag{8}$$

We now claim that there exists a 0-admissible self-financing portfolio $\widetilde{\pi} = (\widetilde{\beta}, \widetilde{\gamma})$ such that

$$X_t^{\widetilde{\pi}} = Z_t^{-1} X_t. \tag{9}$$

Since

$$dZ_t = -Z_t \frac{\mu_t}{\sigma_t} \, dB_t, \tag{10}$$

it follows by Itô's formula (Chapter III, § 3d) that

$$d(Z_t^{-1}) = Z_t^{-1}\Big(\Big(\frac{\mu_t}{\sigma_t}\Big)^2 dt + \frac{\mu_t}{\sigma_t}\,dB_t\Big) \tag{11}$$

and

$$\begin{aligned} d(Z_t^{-1}X_t) &= Z_t^{-1}dX_t + X_t d(Z_t^{-1}) + d(Z_t^{-1})\,dX_t \\ &= Z_t^{-1}(\psi_t\,dB_t) + X_t Z_t^{-1}\Big(\Big(\frac{\mu_t}{\sigma_t}\Big)^2 dt + \frac{\mu_t}{\sigma_t}\,dB_t\Big) + Z_t^{-1}\Big(\frac{\mu_t}{\sigma_t}\psi_t\Big)dt \qquad (12) \\ &= Z_t^{-1}\psi_t\Big(dB_t + \frac{\mu_t}{\sigma_t}\,dt\Big) + X_t Z_t^{-1}\frac{\mu_t}{\sigma_t}\Big(dB_t + \frac{\mu_t}{\sigma_t}\,dt\Big) \\ &= S_t(\sigma_t\,dB_t + \mu_t\,dt)\Big[S_t^{-1}Z_t^{-1}\Big(\frac{\psi_t}{\sigma_t} + X_t\frac{\mu_t}{\sigma_t^2}\Big)\Big]. \qquad (13) \end{aligned}$$

We set

$$\widetilde{\gamma}_t = S_t^{-1}Z_t^{-1}\Big(\frac{\psi_t}{\sigma_t} + X_t\frac{\mu_t}{\sigma_t^2}\Big). \tag{14}$$

Then we see that

$$d(Z_t^{-1}X_t) = \widetilde{\gamma}_t\,dS_t, \tag{15}$$

and moreover,

$$\int_0^t \widetilde{\gamma}_u^2\sigma_u^2\,du < \infty \quad (\mathsf{P}\text{-a.s.}), \qquad t \leqslant T. \tag{16}$$

(Cf. condition (18) in § 4a.)

Thus, for $t \leqslant T$,

$$Z_t^{-1}X_t = \mathsf{E}(Z_T f_T) + \int_0^t \widetilde{\gamma}_u\,dS_u. \tag{17}$$

Setting

$$\widetilde{\beta}_t = Z_t^{-1}X_t - \widetilde{\gamma}_t S_t, \tag{18}$$

we see from (15) that the strategy $\widetilde{\pi} = (\widetilde{\beta}, \widetilde{\gamma})$ is self-financing and of value $X^{\widetilde{\pi}} = (X_t^{\widetilde{\pi}})_{t\leqslant T}$ such that

$$X_0^{\widetilde{\pi}} = \mathsf{E}\big(Z_T f_T\big) \tag{19}$$

and

$$X_t^{\widetilde{\pi}} = Z_t^{-1}X_t, \qquad X_T^{\widetilde{\pi}} = f_T. \tag{20}$$

From (5), comparing (19) and (20), we derive required assertion (2). Note that the strategy $\widetilde{\pi}$ so constructed is 0-admissible because $X_t^{\widetilde{\pi}} \geqslant 0$, $t \leqslant T$.

§ 4c. Fundamental Partial Differential Equation of Hedge Pricing

1. We consider the model of a market formed by a riskless asset—a bank account with zero interest rate ($B_t(0) \equiv 1$)—and a risk asset $S = (S_t, \mathscr{F}_t)_{t\geqslant 0}$ whose dynamics is described by relation (5) in § 4a.

As follows by the theorem in § 4b, the hedging price $\mathbb{C}(f_T; \mathsf{P})$ and the corresponding hedging portfolio $\widetilde{\pi} = (\widetilde{\beta}, \widetilde{\gamma})$ can be found on the basis of the properties of the process $Y = (Y_t, \mathscr{F}_t)_{t\leqslant T}$, where

$$Y_t = Z_t^{-1}\mathsf{E}(Z_T f_T \mid \mathscr{F}_t). \tag{1}$$

In addition, the quantity

$$Y_0 = \mathsf{E}(Z_T f_T) \tag{2}$$

is precisely the price $\mathbb{C}(f_T; \mathsf{P})$,

$$Y_T = f_T, \tag{3}$$

and $Y_t = X_t^{\widetilde{\pi}}$, i.e., Y_t is the value of the hedging portfolio at time $t \leqslant T$. (These properties justify the name '*hedging-price process*' for Y. As already mentioned on several occasions, our method of finding $\mathbb{C}(f_T; \mathsf{P})$ is usually called the *martingale method.*)

In many cases we can explicitly find $\mathsf{E}(Z_T f_T)$ and, therefore, the price $\mathbb{C}(f_T; \mathsf{P})$. In particular, this can be done in the *Black–Merton–Scholes model*, where μ_t and σ_t are constants. (See Chapter VIII below).

2. F. Black and M. Scholes [44], and R. Merton [346] (1973) proposed another method for finding the price $\mathbb{C}(f_T; \mathsf{P})$ and the hedging strategy. It is based on the so-called *fundamental equation* that they have obtained.

This method, which is now widely used in financial mathematics (see, in particular, § 5c below), is essentially as follows.

We consider the process Y_t defined in (1). Since

$$Z_t^{-1} Z_T = \exp\left(- \int_t^T \frac{\mu(u, S_u)}{\sigma(u, S_u)}\, dB_u - \frac{1}{2}\int_t^T \left(\frac{\mu(u, S_u)}{\sigma(u, S_u)}\right)^2 du \right) \tag{4}$$

and $S = (S_t, \mathscr{F}_t)_{t\leqslant T}$ is a Markov process, we see that if $f_T = f(T, S_T)$, then the process $Y = (Y_t, \mathscr{F}_t)_{t\leqslant T}$ is also Markov and Y_t can be represented as $Y(t, S_t)$, where $Y(t,x)$ is a measurable function.

In [44] and [346] the authors simply started from the *assumption* that the hedging portfolio $\widetilde{\pi} = (\widetilde{\beta}, \widetilde{\gamma})$ exists and its value Y_t ($= X_t^{\widetilde{\pi}}$) at time t depends only on the last value S_t of the prices (rather than of the entire history $(S_u,\ u \leqslant t)$).

Another *a priory assumption* necessary for this method is that the function $Y(t,x)$ is in the class $C^{1,2}$. This enables one to use Itô's formula, which brings one to the following *stochastic partial differential equation* (where for simplicity we drop the arguments of functions):

$$dY = \left(\frac{\partial Y}{\partial t} + \mu S \frac{\partial Y}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 Y}{\partial S^2}\right) dt + \sigma S \frac{\partial Y}{\partial S}\, dB. \tag{5}$$

We consider now another representation for Y:

$$Y(t, S_t) = X_t^{\widetilde{\pi}} = \widetilde{\beta}_t + \widetilde{\gamma}_t S_t. \tag{6}$$

In view of self-financing,

$$dY = \widetilde{\gamma}_t\, dS_t = \widetilde{\gamma}_t S_t(\mu_t\, dt + \sigma_t\, dB_t). \tag{7}$$

Using the representations (5) and (7) of the special semimartingale $Y = Y(t, S_t)$ and the uniqueness of the representation of special semimartingales (see Chapter III, § 5b) we see that the coefficients of dB (and of dt) in (5) and (7) must be the same.

Since $S_t > 0$ for $t > 0$, it follows therefore that (P-a.s.)

$$\widetilde{\gamma}_t = \frac{\partial Y}{\partial S}(t, S_t); \tag{8}$$

moreover, the processes $(\widetilde{\gamma}_t)_{t\leqslant T}$ and $\left(\dfrac{\partial Y}{\partial S}(t, S_t)\right)_{t\leqslant T}$ are stochastically indistinguishable.

Comparing the terms with dt in (5) and (7) and taking account of (8) and the equality $\widetilde{\beta}_t = Y(t, S_t) - S_t \dfrac{\partial Y}{\partial S}(t, S_t)$, we obtain the following relation (that holds $(\lambda \times \mathsf{P})$-a.s., where λ is Lebesgue measure in $[0, T]$):

$$\frac{\partial Y}{\partial t}(t, S_t) + \frac{1}{2}\sigma^2(t, S_t) S_t^2 \frac{\partial^2 Y}{\partial S^2}(t, S_t) = 0, \qquad 0 \leqslant t < T.$$

This, in turn, must hold if $Y = Y(t, S)$ satisfies the following *Fundamental partial differential equation* for $0 \leqslant t < T$ and $0 < S < \infty$:

$$\frac{\partial Y}{\partial t}(t, S) + \frac{1}{2}\sigma^2(t, S_t) S^2 \frac{\partial^2 Y}{\partial S^2}(t, S) = 0 \tag{9}$$

with boundary-value condition

$$Y(T, S) = f(T, S), \qquad S > 0. \tag{10}$$

(Cf. the *backward Kolmogorov equation* (6) in Chapter III, § 3f).

We shall discuss the solution of this equation (which reduces—in any case, for $\sigma(t,s) \equiv \sigma = \text{Const}$—to the solution of the standard Feynman–Kac equation; see (19) in Chapter III, § 3f) in connection with the *Black–Scholes formula* in the case of $f(T,S) = (S-K)^+$, in Chapter VIII.

Here we point out the following features of this equation.

Assume that there *exists* a *unique* solution to (9)–(10). We find $\widetilde{\gamma}_t$ by (8) and define $\widetilde{\beta}_t$ by setting

$$\widetilde{\beta}_t = Y(t, S_t) - \widetilde{\gamma}_t S_t. \tag{11}$$

Clearly, the value $X_t^{\widetilde{\pi}}$ of the portfolio $\widetilde{\pi} = (\widetilde{\beta}, \widetilde{\gamma})$ is precisely $Y(t, S_t)$.

It is not *a priori* clear if this portfolio $\widetilde{\pi}$ constructed from $Y(t,S)$ is self-financing, i.e., whether

$$dY(t, S_t) = \widetilde{\gamma}_t \, dS_t. \tag{12}$$

However, this is an immediate consequence of (5) and (9):

$$\begin{aligned} dY(t,S_t) &= S_t \left(\mu_t \frac{\partial Y}{\partial S} \, dt + \sigma_t \frac{\partial Y}{\partial S} \, dB_t \right) \\ &= S_t \frac{\partial Y}{\partial S} (\mu_t \, dt + \sigma_t \, dB_t) = \widetilde{\gamma}_t \, dS_t. \end{aligned} \tag{13}$$

Thus, assume that the problem (9)–(10) has a unique solution $Y(t,S)$. Since $X_t^{\widetilde{\pi}} = Y(t,S_t)$, it follows that $X_T^{\widetilde{\pi}} = f(T, S_T)$, while $X_0^{\widetilde{\pi}} = Y(0, S_0)$ is the initial price of the portfolio $\widetilde{\pi} = (\widetilde{\beta}, \widetilde{\gamma})$.

The following heuristic arguments show that the price $X_0^{\widetilde{\pi}} = Y(0,S_0)$ (obtained in the solution of (9)–(10)) has the properties of 'rationality', 'fairness', while the portfolio $\widetilde{\pi} = (\widetilde{\beta}, \widetilde{\gamma})$ is an 'optimal' hedge.

In fact, let us interpret our problem as a search of a hedging strategy for a seller of a European call option such that the value of this strategy replicates faithfully the pay-off function $f(T, S_T)$. Our solution of (9)–(10) shows that selling this option at the price $C = Y(0, S_0)$ the seller can find a strategy $\widetilde{\pi}$, such that $X_T^{\widetilde{\pi}}$ becomes precisely equal to $f(T, S_T)$.

Assume now that the price C asked for this option contract is higher than $Y(0,S_0)$ and the buyer has accepted this price. Then, clearly, arbitrage is possible: the seller can obtain the net profit of $C - Y(0,S_0)$ meeting simultaneously the terms of the contract because there exists a hedging portfolio of initial price $Y(0,S_0)$ that replicates faithfully the pay-off function.

On the other hand, if $C < Y(0,S_0)$, then due to the uniqueness of the solution of (9)–(10) the terms of the contract will not necessarily be fulfilled (at any rate, if one must choose strategies in the Markov class).

There are several weak points in this method based on the solution of the *fundamental equation*, namely, the a priori assumptions of the 'Markovian structure' of the value of $\widetilde{\pi}$ (i.e., the representation $X_t^{\widetilde{\pi}} = Y(t, S_t)$) and of the $C^{1,2}$-regularity of $Y(t,S)$ (enabling the use of Itô's formula).

Fortunately, there exist other methods for the construction of hedging strategies and finding the 'rational' price $\mathbb{C}(f_T; \mathsf{P})$ (e.g., the 'martingale' method exposed in § 4b), which show, in particular, that a hedging portfolio does exist and its value has the form $Y(t, S_t)$ with a sufficiently smooth function Y, so that equation (9) actually holds. We present a more thorough analysis of the case of a standard European call option with $f(T, S_T) = (S_T - K)^+$ in Chapter VIII, § 1b, where we discuss and use both 'martingale' approach and approach based on the above *fundamental equation*.

3. In the above discussion we assumed that the riskless asset (a bank account) $B(0) = (B_t(0))_{t\geqslant 0}$ has the form $B_t(0) \equiv 1$. In effect, this means that we deal with discounted prices. However, in some cases one must consider the 'absolute' values of the prices rather than the 'relative', discounted ones. Here we present the corresponding modifications in the case when the bank account $B(r) = (B_t(r))_{t\geqslant 0}$ (in some 'absolute' units) has the following form:

$$B_t(r) = B_0(r) \exp\biggl(\int_0^t r_s\, ds\biggr) \tag{14}$$

(here $(r_t)_{t\geqslant 0}$ is a deterministic nonnegative function—the interest rate), and the risk asset (stock) is $S = (S_t(\mu, \sigma))_{t\geqslant 0}$, $S_0(\mu, \sigma) = S_0 > 0$, where

$$dS_t(\mu, \sigma) = S_t(\mu, \sigma)(\mu_t\, dt + \sigma_t\, dB_t). \tag{15}$$

(Our assumptions about μ_t, σ_t, and the Brownian motion $B = (B_t, \mathscr{F}_t)_{t\geqslant 0}$ are the same as in § 4a.)

Let $\widetilde{\pi} = (\widetilde{\beta}, \widetilde{\gamma})$ be a self-financing portfolio and let

$$X_t^{\widetilde{\pi}} = \widetilde{\beta}_t B_t(r) + \widetilde{\gamma}_t S_t(\mu, \sigma). \tag{16}$$

We assume that $X_t^{\widetilde{\pi}}$ is as follows:

$$X_t^{\widetilde{\pi}} = Y(t, S_t),$$

where $S_t = S_t(\mu, \sigma)$ and $Y(t, S) \in C^{1,2}$. Then for $Y = Y(t, S)$ we obtain the same equation (5).

On the other hand, since

$$dB_t(r) = r_t B_t(r)\, dt, \tag{17}$$

it follows by the property of self-financing that

$$dY(t, S_t) = dX_t^{\widetilde{\pi}} = \bigl(\widetilde{\gamma}_t \mu_t S_t + \widetilde{\beta}_t r_t B_t(r)\bigr)\, dt + \widetilde{\gamma}_t \sigma_t S_t\, dB_t. \tag{18}$$

Comparing the terms containing dB in (5) and (18) we obtain again that $\widetilde{\gamma}_t = \dfrac{\partial Y}{\partial S}(t, S_t)$.

By (16),

$$\widetilde{\beta}_t = \frac{1}{B_t(r)}\left(Y(t, S_t) - S_t \frac{\partial Y}{\partial S}(t, S_t)\right),$$

and, as in our derivation of (9), we see that the coefficients of dt in (5) and (18) are the same if $Y(t, S)$ satisfies the following *Fundamental equation* for $S \in \mathbb{R}_+$ and $0 \leqslant t < T$:

$$\frac{\partial Y}{\partial t} + rS\frac{\partial Y}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 Y}{\partial S^2} = rY, \tag{19}$$

with boundary condition $Y(T, S) = f(T, S)$, $S \in \mathbb{R}_+$.

It should be noted that $\mu = \mu(t, S)$ *does not enter* the equation or boundary condition, and therefore $Y(0, S_0)$ is independent of μ. This might appear puzzling at first glance. (On the other hand we can consider this property to be desirable because investors can have different ideas about the values of μ and σ, and therefore, different notions of the actual dynamics of the price process $S = (S_t)_{t\geqslant 0}$.) Probably, the best explanation can be given from the standpoint of the 'martingale' approach, in which the price satisfies the relations

$$Y(0, S_0) = \mathbb{C}(f_T; \mathsf{P}) = \mathsf{E}_{\widetilde{\mathsf{P}}_T} f(T, S_T)$$

(see formula (2) in § 4b) and the process $S = (S_t)_{t\leqslant T}$ is a local martingale with respect to $\widetilde{\mathsf{P}}_T$ (by Girsanov's theorem for semimartingales), with $dS_t = S_t\sigma_t\, d\widetilde{B}_t$, where $\widetilde{B}$ is a Brownian motion.

Hence we see that the price $\mathbb{C}(f_T; \mathsf{P})$ is independent of μ. However, the dependence on the volatility σ does not 'wither away' since the quadratic characteristics of continuous martingale components *do not change* under absolutely continuous changes of measure (see formula (6) in § 3g).

5. Arbitrage, Completeness, and Hedge Pricing in Diffusion Models of Bonds

§ 5a. Models without Opportunities for Arbitrage

1. In Chapter III, § 4c we considered several models of the *term structure* of prices of bond families. In particular, we observed that there existed two approaches to the description of the dynamics of bond prices $\mathsf{P}(t,T)$: the *indirect* approach (when we take some 'interest rate' process $r = (r(t))_{t\geqslant 0}$ as a 'basis' and assume that $\mathsf{P}(t,T) = F(t,r(t),T)$), and the *direct* one (when $\mathsf{P}(t,T)$ is defined in a straightforward way, as a solution to some stochastic differential equations).

These approaches bring forward distinct models. In the framework of our notion of a 'fair' market as a market without opportunities for arbitrage it would be natural to find out first of all what conditions ensure the absence of arbitrage in these models and how one can find 'explicit' expressions for the prices $\mathsf{P}(t,T)$ in these models.

2. Taking the indirect approach we assume that the *interest rate process* $r = (r(t))_{t\geqslant 0}$ is the solution of the stochastic differential equation (cf. (5) in Chapter III, § 4c)

$$dr(t) = a\big(t,r(t)\big)\,dt + b\big(t,r(t)\big)\,dW_t, \tag{1}$$

generated by a Wiener process $W = (W_t,\mathscr{F}_t)_{t\geqslant 0}$. (As regards the Brownian (Wiener) filtration $(\mathscr{F}_t)_{t\geqslant 0}$, see § 4a.) We also assume that the coefficients $a = a(t,r)$ and $b = b(t,r)$ are chosen so that equation (1) has a unique strong solution (Chapter III, § 3e).

In a natural way, one can associate with the interest rate $r = (r(t))_{t\geqslant 0}$ a *bank account*

$$B(r) = (B_t(r))_{t\geqslant 0}$$

with

$$B_t(r) = \exp\left(\int_0^t r(s)\,ds\right), \tag{2}$$

which, like in the case of stock or other assets, plays the role of a 'gauge' in the valuation of various bonds. $\Bigl($We assume throughout that $\int_0^t |r(s)|\,ds < \infty$ (P-a.s.), $t > 0.\Bigr)$

Let $\mathsf{P}(t,T)$ be the price of some T-bond (see Chapter III, § 4c), which is assumed to be $\mathscr{F}_t$-measurable for each $0 \leqslant t < T$, and $\mathsf{P}(T,T) = 1$. Below we assume also that the processes $(\mathsf{P}(t,T))_{t\geqslant 0}$ are optional for each $T > 0$. Then, in particular, $\mathsf{P}(t,T)$ is $\mathscr{F}_t$-measurable for each $T > 0$. By the meaning of $\mathsf{P}(t,T)$ as the price of bonds with $\mathsf{P}(T,T) = 1$ we must also assume that $0 \leqslant \mathsf{P}(t,T) \leqslant 1$.

Now let us introduce the discounted price

$$\overline{\mathsf{P}}(t,T) = \frac{\mathsf{P}(t,T)}{B_t(r)}, \qquad 0 \leqslant t \leqslant T. \tag{3}$$

Bearing in mind the result of the *First fundamental theorem* about the absence of arbitrage (Chapter V, § 2b) and based on the conviction that the 'existence of a martingale measure ensures (or almost ensures) the absence of arbitrage' we assume that there *exists* a martingale (or risk-neutral) measure $\widetilde{\mathsf{P}}_T$ on $\mathscr{F}_T$ such that $\widetilde{\mathsf{P}}_T \sim \mathsf{P}_T$ $(= \mathsf{P} \,|\, \mathscr{F}_T)$ and $(\overline{\mathsf{P}}(t,T), \mathscr{F}_t)_{t\leqslant T}$ is a $\widetilde{\mathsf{P}}_T$-martingale. Then we conclude directly from (3) that

$$\mathsf{E}_{\widetilde{\mathsf{P}}_T}(\overline{\mathsf{P}}(T,T) \,|\, \mathscr{F}_t) = \overline{\mathsf{P}}(t,T), \qquad t \leqslant T, \tag{4}$$

so that we have the following result.

THEOREM 1. *If there exists a martingale measure $\widetilde{\mathsf{P}}_T \sim \mathsf{P}_T$ such that the discounted process $(\overline{\mathsf{P}}(t,T), \mathscr{F}_t)_{t\leqslant T}$ is a $\widetilde{\mathsf{P}}_T$-martingale, then*

$$\mathsf{P}(t,T) = \mathsf{E}_{\widetilde{\mathsf{P}}_T}\biggl(\exp\biggl(-\int_t^T r(s)\,ds\biggr) \biggm| \mathscr{F}_t\biggr). \tag{5}$$

This is an immediate consequence of (4) and the condition $\mathsf{P}(T,T) = 1$, for we have

$$\mathsf{E}_{\widetilde{\mathsf{P}}_T}\biggl(\frac{1}{B_T(r)} \biggm| \mathscr{F}_t\biggr) = \frac{\mathsf{P}(t,T)}{B_t(r)},$$

which delivers the representation (5).

We see from (5) that if $r = (r(t))_{t\geqslant 0}$ is a Markov process with respect to $\widetilde{\mathsf{P}}$, then the price $\mathsf{P}(t,T)$ can be written as follows:

$$\mathsf{P}(t,T) = F(t, r(t), T).$$

The 'absence of arbitrage' imposes automatically certain restrictions on the function $F(t,r,T)$ (see § 5c below).

Remark 1. We point out that the price $\mathsf{P}(t,T)$ of bonds *cannot be unambigoulsy evaluated* on the basis of the state of the bank account $B(r)$ and the condition of the absence of arbitrage (more precisely, of the existence of a martingale measure). In fact there is no reason for $\widetilde{\mathsf{P}}_T$ to be unique, which means that $\mathsf{P}(t,T)$ can be realized (in the form (5)) in *various ways* depending on a particular measure $\widetilde{\mathsf{P}}_T$.

We note also that if, instead of $\mathsf{P}(T,T)=1$, we require that $\mathsf{P}(T,T)$ be equal to f_T, where f_T is $\mathscr{F}_T$-measurable and $\mathsf{E}_{\widetilde{\mathsf{P}}_T}\left|\dfrac{f_T}{B_T(r)}\right|<\infty$, then by (4) we obtain

$$\mathsf{E}_{\widetilde{\mathsf{P}}_T}\left(\frac{f_T}{B_T(r)}\;\Big|\;\mathscr{F}_t\right)=\frac{\mathsf{P}(t,T)}{B_t(r)},$$

and therefore

$$\mathsf{P}(t,T)=\mathsf{E}_{\widetilde{\mathsf{P}}_T}\left\{f_T\exp\left(-\int_t^T r(s)\,ds\right)\;\Big|\;\mathscr{F}_t\right\}. \tag{6}$$

3. We proceed now from one fixed T-bond to a *family* of T-bonds:

$$\mathcal{P}=\{\mathsf{P}(t,T);\ 0\leqslant t\leqslant T,\ T>0\}.$$

DEFINITION 1. Let P be a probability measure on $(\Omega,\mathscr{F},(\mathscr{F}_t)_{t\geqslant 0})$ with $\mathscr{F}=\bigvee\mathscr{F}_t$. We say that a measure $\widetilde{\mathsf{P}}\overset{\text{loc}}{\sim}\mathsf{P}$ (i.e., $\widetilde{\mathsf{P}}_t\sim\mathsf{P}_t$ for $t\geqslant 0$) is a *local martingale measure for the family* $\mathcal{P}$, if for each $T>0$ the discounted prices $\overline{\mathsf{P}}(t,T)=\dfrac{\mathsf{P}(t,T)}{B_t(r)}$, $t\leqslant T$, are local $\widetilde{\mathsf{P}}_T$-martingales.

To define an *arbitrage-free* $(B,\mathcal{P})$-market formed by a *bank account* B and a *family of bonds* $\mathcal{P}$ we must, first of all, discuss the notion of portfolio (strategy) in this case.

DEFINITION 2 ([38]). A strategy $\pi=(\beta,\gamma)$ in a $(B,\mathcal{P})$-market is a pair of a predictable process $\beta=(\beta_t)_{t\geqslant 0}$ and a family of finite (real-valued) Borel measures $\gamma=(\gamma_t(\,\cdot\,))_{t\geqslant 0}$ such that for all t and ω the set function $\gamma_t=\gamma_t(dT)$ is a measure in $(\mathbb{R}_+,\mathscr{B}(\mathbb{R}_+))$ with support concentrated on $[t,\infty)$ and for each $A\in\mathscr{B}(\mathbb{R}_+)$ the process $(\gamma_t(A))_{t\geqslant 0}$ is predictable.

The meaning of β and γ is transparent: β_t is the number of 'unit' bank accounts (at time t) and $\gamma_t(dT)$ is the 'number' of bonds with maturity date in the interval $[T,T+dT]$.

DEFINITION 3. The *value* of a strategy π is the (random) process $X^\pi=(X_t^\pi)_{t\geqslant 0}$ with

$$X_t^\pi=\beta_tB_t+\int_t^\infty \mathsf{P}(t,T)\,\gamma_t(dT). \tag{7}$$

(We assume here that the Lebesgue–Stieltjes integrals in (7) are defined for all t and ω.)

4. We define now a self-financing portfolio $\pi = (\beta, \gamma)$ in a $(B, \mathcal{P})$-market.

To this end, taking the *direct* approach (Chapter III, § 4c) we assume that the dynamics of the prices $\mathsf{P}(t, T)$ can be described by the HJM-model; namely, for $0 \leqslant t < T$ and $T > 0$ we have

$$d\mathsf{P}(t,T) = \mathsf{P}(t,T)\big(A(t,T)\,dt + B(t,T)\,dW_t\big), \tag{8}$$

where $W = (W_t)_{t \geqslant 0}$ is a standard Wiener process, which plays the role of the source of randomness. We must add to equations (8) the boundary conditions $\mathsf{P}(T,T) = 1$, $T > 0$. (For a discussion of the condition of the measurability of $A(t,T)$ and $B(t,T)$ and the solubility of (8), see Chapter III, § 4c).

In view of the equation

$$dB_t(r) = r(t)B_t(r)\,dt \tag{9}$$

we obtain by Itô's formula that the process

$$\overline{\mathsf{P}}(t,T) = \frac{\mathsf{P}(t,T)}{B_t(r)} \tag{10}$$

has the differential (with respect to t for each T)

$$d\overline{\mathsf{P}}(t,T) = \overline{\mathsf{P}}(t,T)\big([A(t,T) - r(t)]\,dt + B(t,T)\,dW_t\big). \tag{11}$$

We have said that a strategy $\pi = (\beta, \gamma)$ of value $X_t^\pi = \beta_t B_t + \gamma_t S_t$ in a diffusion (B, S)-market is self-financing if

$$dX_t^\pi = B_t\,dB_t + \gamma_t\,dS_t, \tag{12}$$

i.e.,

$$X_t^\pi = X_0^\pi + \int_0^t \beta_u\,dB_u + \int_0^t \gamma_u\,dS_u. \tag{13}$$

In the present case of a diffusion $(B, \mathcal{P})$-market it is reasonable to say that a strategy $\pi = (\beta, \gamma)$ of value $X_t^\pi = \beta_t B_t(r) + \int_t^\infty \mathsf{P}(t,T)\,\gamma_t(dt)$ is *self-financing* ([38]) if (in the symbolic notation)

$$dX_t^\pi = \beta_t\,dB_t(r) + \int_t^\infty d\mathsf{P}(t,T)\,\gamma_t(dT), \tag{14}$$

which (in view of (8)) should be interpreted as follows:

$$\begin{aligned} X_t^\pi = X_0^\pi + \int_0^t \beta_s\,dB_s(r) &+ \int_0^t \left[\int_s^\infty A(s,T)\mathsf{P}(s,T)\,\gamma_s(dT)\right] ds \\ &+ \int_0^t \left[\int_s^\infty B(s,T)\mathsf{P}(s,T)\,\gamma_s(dT)\right] dW_s. \end{aligned} \tag{15}$$

If a strategy $\pi = (\beta, \gamma)$ is self-financing, then its discounted value

$$\overline{X}_t^\pi = \frac{X_t^\pi}{B_t(r)} \tag{16}$$

satisfies the relation

$$d\overline{X}_t^\pi = \int_t^\infty d\overline{\mathsf{P}}(t,T)\,\gamma_t(dT). \tag{17}$$

As in (14), relation (17) is symbolic and (in view of (11)) it means that

$$\overline{X}_t^\pi = \overline{X}_0^\pi + \int_0^t \left[\int_s^\infty \big(A(s,T) - r(s)\big)\overline{\mathsf{P}}(s,T)\,\gamma_s(dT)\right] ds \\ + \int_0^t \left[\int_s^\infty B(s,T)\overline{\mathsf{P}}(s,T)\,\gamma_s(dT)\right] dW_s. \tag{18}$$

5. To state conditions for the absence of arbitrage in $(B,\mathcal{P})$-models we discuss first of all the existence of martingale measures.

To this end we define the functions $A(t,T)$ and $B(t,T)$ also for $t > T$ by setting $A(t,T) = r(t)$ and $B(t,T) = B(T,T)$.

Then we immediately see from (11) that in order that the sequence of prices $(\overline{\mathsf{P}}(t,T))_{t\leqslant T}$ be a local martingale for each $T > 0$ with respect to the original measure P it is necessary that

$$A(t,T) = r(t). \tag{19}$$

By (8) we obtain that in this case

$$d\mathsf{P}(t,T) = \mathsf{P}(t,T)\big(r(t)\,dt + B(t,T)\,dW_t\big)$$

and

$$d\overline{\mathsf{P}}(t,T) = \overline{\mathsf{P}}(t,T)B(t,T)\,dW_t. \tag{20}$$

Taking into account relations (14) and (15) in Chapter III, § 4c and the equality $\dfrac{\partial A(t,T)}{\partial T} = 0$ holding under the assumption (19) we obtain the following relation for $f(t,T)$:

$$df(t,T) = a(t,T)\,dt + b(t,T)\,dW_t,$$

where

$$a(t,T) = b(t,T)\int_0^T b(t,s)\,ds.$$

On the other hand, if (19) fails, then it would be natural (by analogy with the case of stock) to refer to the ideas underlying Girsanov's theorem.

Assume that besides the original measure P on the space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geqslant 0})$ with $\mathscr{F} = \bigvee \mathscr{F}_t$ there exists a probability measure $\widetilde{\mathsf{P}}$ such that $\widetilde{\mathsf{P}} \overset{\mathrm{loc}}{\sim} \mathsf{P}$, i.e., $\widetilde{\mathsf{P}}_t \sim \mathsf{P}_t$, $t \geqslant 0$.

We set $Z_t = \dfrac{d\widetilde{\mathsf{P}}_t}{d\mathsf{P}_t}$. Since $(\mathscr{F}_t)_{t\geqslant 0}$ is the Brownian (Wiener) filtration, it follows by the theorem on the representation of positive local martingales (see formula (22) in Chapter III, § 3c) that

$$Z_t = \exp\biggl(\int_0^t \varphi(s)\,dW_s - \frac{1}{2}\int_0^t \varphi^2(s)\,ds\biggr), \tag{21}$$

where the $\varphi(s)$ are $\mathscr{F}_s$-measurable, $\int_0^t \varphi^2(s)\,ds < \infty$ (P-a.s.), and $\mathsf{E}Z_t = 1$ for each $t > 0$.

By Girsanov's theorem (Chapter III, § 3e) the process $\widetilde{W} = (\widetilde{W}_t)_{t\geqslant 0}$ with

$$\widetilde{W}_t = W_t - \int_0^t \varphi(s)\,ds \tag{22}$$

is Wiener with respect to $\widetilde{\mathsf{P}}$. Hence

$$d\mathsf{P}(t,T) = \mathsf{P}(t,T)\bigl[\bigl(A(t,T) + \varphi(t)B(t,T)\bigr)\,dt + B(t,T)\,\widetilde{W}_t\bigr] \tag{23}$$

with respect to $\widetilde{\mathsf{P}}$ and

$$d\overline{\mathsf{P}}(t,T) = \overline{\mathsf{P}}(t,T)\bigl[\bigl(A(t,T) + \varphi(t)B(t,T) - r(t)\bigr)\,dt + B(t,T)\,\widetilde{W}_t\bigr] \tag{24}$$

(cf. (8) and (11)).

Hence (cf. (11)) the processes $(\overline{\mathsf{P}}(t,T))_{t\leqslant T}$ are local martingales with respect to $\widetilde{\mathsf{P}}$ for all $T > 0$ if and only if

$$A(t,T) + \varphi(t)B(t,T) - r(t) = 0. \tag{25}$$

By (23) we obtain in this case that

$$d\mathsf{P}(t,T) = \mathsf{P}(t,T)\bigl(r(t)\,dt + B(t,T)\,d\widetilde{W}_t\bigr), \tag{26}$$

where $\widetilde{W} = (\widetilde{W}_t)_{t\geqslant 0}$ is a Wiener process with respect to $\widetilde{\mathsf{P}}$.

6. The definition of the property of a strategy $\pi = (\beta,\gamma)$ in a $(B,\mathcal{P})$-market to be *arbitrage-free* (say, in the NA_+-version) at an instant T is as in § 1c. We say that a $(B,\mathcal{P})$-market is *arbitrage-free* if it has this quality for *all* $T > 0$.

THEOREM 2. *Let $\widetilde{\mathsf{P}} \overset{\text{loc}}{\sim} \mathsf{P}$ be a measure such that the density process $Z = (Z_t)_{t\geqslant 0}$ has the form* (21) *and condition* (25) *holds.*

Then there are no opportunities for arbitrage for any $a \geqslant 0$ in the class of a-admissible strategies π $(\overline{X}_t^\pi \geqslant -a,\ t > 0)$ such that

$$\int_0^t \left[\int_s^\infty B(s,T)\,\gamma_s(dT)\right]^2 ds < \infty, \qquad t > 0. \tag{27}$$

Proof. If (25) and (27) are satisfied, then the process $\overline{X}^\pi = (\overline{X}_t^\pi)_{t\geqslant 0}$ is a $\widetilde{\mathsf{P}}$-local martingale by relation (18).

In view of the a-admissibility $(\overline{X}_t^\pi \geqslant -a,\ t > 0)$ this process is also a supermartingale. Hence if $\overline{X}_0^\pi = 0$, then $\mathsf{E}_{\widetilde{\mathsf{P}}}\overline{X}_T^\pi \leqslant 0$ for each $T > 0$. However, $\widetilde{\mathsf{P}}(\overline{X}_T^\pi \geqslant 0) = \widetilde{\mathsf{P}}(X_T^\pi \geqslant 0) = 1$. Hence $X_T^\pi = 0$ ($\widetilde{\mathsf{P}}$- and P-a.s.) for $T > 0$, which completes the proof.

Remark 2. Assume that the function

$$\left(\frac{r(t) - A(t,T)}{B(t,T)}\right)_{t\leqslant T} \tag{28}$$

is independent of T and

$$\int_0^t \left(\frac{r(s) - A(s,T)}{B(s,T)}\right)^2 ds < \infty \qquad (\mathsf{P}\text{-a.s.}) \tag{29}$$

for each $t > 0$. In this case, looking for a measure $\widetilde{\mathsf{P}}$ with the property $\widetilde{\mathsf{P}} \overset{\text{loc}}{\sim} \mathsf{P}$ one could proceed as follows.

We denote the function in (28) by $\varphi = \varphi(t)$, $t \leqslant T$, introduce the process $Z = (Z_t)_{t\geqslant 0}$ defined by (21), and assume that $\mathsf{E}Z_t = 1$, $t > 0$. Then for each $t > 0$ the measure $\widetilde{\mathsf{P}}_t$ with $d\widetilde{\mathsf{P}}_t = Z_t\,d\mathsf{P}_t$ is a probability measure such that $\widetilde{\mathsf{P}}_t \sim \mathsf{P}_t$.

The family of measures $\{\widetilde{\mathsf{P}}_t,\ t \geqslant 0\}$ is compatible (in the sense that $\widetilde{\mathsf{P}}_s = \widetilde{\mathsf{P}}_t \mid \mathscr{F}_s$ for $s \leqslant t$) and if there *exists* a probability measure $\widetilde{\mathsf{P}}$ on $(\Omega, \mathscr{F})$ such that $\widetilde{\mathsf{P}} \overset{\text{loc}}{\sim} \mathsf{P}$, then it is the required one.

If the maturity dates of T-bonds in our $(B,\mathcal{P})$-market satisfy the inequality $T \leqslant T_0$, where $T_0 < \infty$, then we can take $\widetilde{\mathsf{P}}_{T_0}$ as the required measure $\widetilde{\mathsf{P}}$.

It is equally clear that if $Z_\infty = \lim\limits_{t\to\infty} Z_t$ and we have $\mathsf{E}Z_\infty = 1$ and $\mathsf{P}(Z_\infty > 0) = 1$, then the measure $\widetilde{\mathsf{P}}$ with $d\widetilde{\mathsf{P}} = Z_\infty\,d\mathsf{P}$ is the required martingale measure with property $\widetilde{\mathsf{P}} \overset{\text{loc}}{\sim} \mathsf{P}$.

7. We present now an example of an arbitrage-free $(B,\mathcal{P})$-model.

Following [36] and [219], we start from the forward interest rate $f(t,T)$ with stochastic differential (with respect to t for fixed T)

$$df(t,T) = a(t,T)\,dt + b(t,T)\,dW_t, \tag{30}$$

where

$$b(t,T) \equiv \sigma > 0, \tag{31}$$

$$a(t,T) = \sigma^2(T-t), \quad t < T. \tag{32}$$

Then (30) takes the following form:

$$df(t,T) = \sigma^2(T-t)\,dt + \sigma\,dW_t, \tag{33}$$

and therefore

$$f(t,T) = f(0,T) + \sigma^2 t\Big(T - \frac{t}{2}\Big) + \sigma W_t. \tag{34}$$

where $f(0,T)$ is the instantaneous forward interest rate of T-bonds on the $(B,\mathcal{P})$-market (at time $t=0$).

By (34) and the definition $r(t) = f(t,t)$ we obtain

$$r(t) = f(0,t) + \frac{\sigma^2}{2}t^2 + \sigma W_t. \tag{35}$$

Hence the interest rate $r = (r(t))_{t\geqslant 0}$ satisfies the equation

$$dr(t) = \left(\frac{\partial f(0,t)}{\partial t} + \sigma^2 t\right)dt + \sigma\,dW_t. \tag{36}$$

(Cf. the *Ho–Lee model* (12) in Chapter III, § 4c.) The coefficients $A(t,T)$ and $B(t,T)$ in (8) can be calculated on the basis of $a(t,T) = \sigma^2(T-t)$ and $b(t,T) \equiv \sigma$ in (33) as follows:

$$A(t,T) = r(t) - \int_t^T a(t,s)\,ds + \frac{1}{2}\left(\int_t^T b(t,s)\,ds\right)^2 = r(t), \tag{37}$$

$$B(t,T) = -\sigma(T-t). \tag{38}$$

Hence condition (19) holds in our $(B,\mathcal{P})$-model, and therefore the original measure P is a martingale measure and no arbitrage is possible.

The prices $\mathsf{P}(t,T)$ themselves can be found from the equation

$$d\mathsf{P}(t,T) = \mathsf{P}(t,T)\big[r(t)\,dt - \sigma(T-t)\,dW_t\big], \qquad t < T,$$

which must be solved for each $T > 0$ under the condition $\mathsf{P}(T,T) = 1$.

We can also use the equality

$$\mathsf{P}(t,T) = \exp\left(-\int_t^T f(t,s)\,ds\right), \qquad t \leqslant T \tag{39}$$

holding by (2) in Chapter III, § 4c.

By (34),

$$\begin{aligned}\int_t^T f(t,s)\,ds &= \int_t^T \left[f(0,s) + \sigma^2 t\left(s - \frac{t}{2}\right)\right] ds + \sigma(T-t)W_t \\ &= \int_t^T f(0,s)\,ds + \frac{\sigma^2}{2}tT(T-t) + \sigma(T-t)W_t.\end{aligned}$$

Consequently,

$$\begin{aligned}\mathsf{P}(t,T) &= \exp\left\{-\int_t^T f(0,s)\,ds - \frac{\sigma^2}{2}tT(T-t) + \sigma(T-t)W_t\right\} \\ &= \frac{\mathsf{P}(0,T)}{\mathsf{P}(0,t)}\exp\left\{-\frac{\sigma^2}{2}tT(T-t) + \sigma(T-t)W_t\right\}.\end{aligned} \tag{40}$$

Hence we obtain from (35) the following representation for $\mathsf{P}(t,T)$ in terms of the interest rate $r(t)$:

$$\mathsf{P}(t,T) = \frac{\mathsf{P}(0,T)}{\mathsf{P}(0,t)}\exp\left\{(T-t)f(0,T) - \frac{\sigma^2}{2}t(T-t)^2 - (T-t)r(t)\right\}. \tag{41}$$

(Cf. affine models in Chapter III, § 4c and in § 5c below.)

8. In the above diffusion models we have assumed that the interest rates $r = (r(t))$, the forward interest rates $f = (f(t,T))$, and the bond prices

$$\mathcal{P} = \{\mathsf{P}(t,T);\ 0 \leqslant t \leqslant T,\ T < \infty\}$$

themselves have a *single* source of randomness: a Wiener process $W = (W_t)_{t\geqslant 0}$.

In the vast literature on the dynamics of bond prices the authors discuss also some other models, where one Wiener process $W = (W_t)_{t\geqslant 0}$ is replaced by a *multivariate* Wiener process $W = (W^1, \dots, W^n)$. To take into account also jumps in the prices $\mathsf{P}(t,T)$ one invokes other 'sources of randomness': point processes, marked point processes, Lévy, and some other processes.

Referring the reader to the special literature (see, e.g., [36], [38], [128], and the bibliography therein), we present here only a few models equipped with such 'sources of randomness'.

In [36] and [38] the authors generalize models of type (1) by introducing models of 'diffusion with jumps' kind:

$$dr(t) = a_t\,dt + \sum_{i=1}^{d} b_t^i\,dW_t^i + \int q(t,x)\,\mu(dt,dx), \tag{42}$$

where $\mu = \mu(dt,dx)$ is an integer-valued random measure in $\mathbb{R}_+ \times \Omega \times E$ and $(W^1,\dots,W^d)$ are independent Wiener processes.

One must also make the corresponding modification in the description of the dynamics of $\mathsf{P}(t,T)$ and $f(t,T)$:

$$d\mathsf{P}(t,T) = \mathsf{P}(t,T)\Big(A(t,T)\,dt + \sum_{i=1}^{d} B^i(t,T)\,dW_t^i\Big) + \mathsf{P}(t-,T)\int_E q(t,x,T)\,\mu(dt,dx), \tag{43}$$

$$df(t,T) = a(t,T)\,dt + \sum_{i=1}^{d} b^i(t,T)\,dW_t^i + \int_E \delta(t,x,T)\,\mu(dt,dx). \tag{44}$$

9. Following [128] we consider now models with *Lévy processes* as 'sources of randomness'.

To this end we consider first equation (20), which we rewrite as follows:

$$d\mathsf{P}(t,T) = \mathsf{P}(t,T)\,d\widehat{H}(t,T), \tag{45}$$

where

$$\widehat{H}(t,T) = \int_0^t \big[r(s)\,ds + B(s,T)\,dW_s\big]. \tag{46}$$

We also set

$$H(t,T) = \int_0^t \bigg(\Big[r(s) - \frac{B^2(s,T)}{2}\Big]\,ds + B(s,T)\,dW_s\bigg). \tag{47}$$

Then (see (9)–(13) in § 3d) we have the representations

$$\mathsf{P}(t,T) = \mathsf{P}(0,T)\mathscr{E}(\widehat{H}(\,\cdot\,,T))_t \tag{48}$$

and

$$\mathsf{P}(t,T) = \mathsf{P}(0,T)e^{H(t,T)}. \tag{49}$$

In view of (47),

$$\overline{\mathsf{P}}(t,T) = \frac{\mathsf{P}(t,T)}{B_t(r)} = \overline{\mathsf{P}}(0,T)\exp\left\{\int_0^t B(s,T)\,dW_s - \frac{1}{2}\int_0^t B^2(s,T)\,ds\right\}. \tag{50}$$

If, for instance, the function $B(s,T)$, $s \leqslant T$, is bounded, then we see that the expression on the right-hand side of (50) is a martingale.

Now, we replace the Wiener process $W = (W_t)_{t\geqslant 0}$ with a Lévy process $L = (L_t)_{t\geqslant 0}$ (see Chapter III, § 1b). What can be the form of the processes $\widehat{H}(t,T)$ and $H(t,T)$ if, in place of the integrals $\int_0^t B(s,T)\,dW_s$, we consider now the integrals $\int_0^t B(s,T)\,dL_s$ interpreted as stochastic integrals over a semimartingale $L = (L_s)_{s\leqslant T}$ with bounded deterministic functions $B(s,T)$?

If the functions $B(s,T)$ are sufficiently smooth in s, then we can use N. Wiener's definition

$$\int_0^t B(s,T)\,dL_s \equiv B(t,T)L_t - \int_0^t \frac{\partial B}{\partial s}(s,T)L_s\,ds.$$

(See Chapter III, § 3c on this subject and see [128] in connection with Lévy processes.)

Let

$$\varphi(\lambda) = \lambda b + \frac{\lambda^2}{2}\sigma^2 + \int_{\mathbb{R}}\bigl(e^{\lambda x} - 1 - \lambda g(x)\bigr)\,\nu(dx) \tag{51}$$

be the cumulant function (see § 3c) of the Lévy process $L = (L_t)_{t\geqslant 0}$.

We assume that the integral in (51) is well defined and bounded for all λ such that $|\lambda| \leqslant c$, where $c = \sup\limits_{s\leqslant T}|B(s,T)|$.

By the meaning of the cumulant function

$$\mathsf{E}e^{\lambda L_t} = e^{t\varphi(\lambda)}. \tag{52}$$

Let $X_t^T = \int_0^t B(s,T)\,dL_s$, $t \leqslant T$. The process $X^T = (X_t^T)_{t\leqslant T}$ has independent increments and its triplet $(B^{X^T}, C^{X^T}, \nu^{X^T})$ of predictable characteristics can be found from the triplet (B^L, C^L, ν^L) of L (see Chapter IX, § 5a in [128] and [250]). Using Itô's formula we can see (see detail in [128]) that

$$\mathsf{E}e^{\lambda X_t^T} = \exp\left(\int_0^t \varphi(\lambda B(s,T))\,ds\right).$$

The process $\left(\exp\left(\lambda X_t^T - \int_0^T \varphi(\lambda B(s,T))ds\right)\right)_{t\leqslant T}$ is a martingale (cf. (11) in § 3c). Hence if we want that $(\overline{\mathsf{P}}(t,T))_{t\leqslant T}$ be a martingale, then it seems reasonable to generalize (50) by setting

$$\overline{\mathsf{P}}(t,T) = \overline{\mathsf{P}}(0,T)\exp\left\{\int_0^t B(s,T)\,dL_s - \int_0^t \varphi(B(s,T))\,ds\right\}. \tag{53}$$

Returning from $\overline{\mathsf{P}}(t,T)$ to the process $\mathsf{P}(t,T)$ we obtain that the $(B,\mathcal{P})$-market with

$$\mathsf{P}(t,T) = \mathsf{P}(0,T)e^{H(t,T)}, \qquad t \leqslant T, \quad T > 0, \tag{54}$$

where

$$H(t,T) = \int_0^t B(s,T)\,dL_s + \int_0^t \bigl[r(s) - \varphi(B(s,T))\bigr]\,ds, \tag{55}$$

has the following property: the discounted prices $(\overline{\mathsf{P}}(t,T))_{t\leqslant T}$ form a martingale with respect to the original measure P and there exist no opportunities for arbitrage in the class of α-admissible strategies in this market (cf. Theorem 2).

Using the connection between $\widehat{H}(t,T)$ and $H(t,T)$ (see (10) in § 3d) we obtain

$$\widehat{H}(t,T) = H(t,T) + \frac{\sigma^2}{2}\int_0^t B^2(s,T)\,ds + \sum_{0<s\leqslant t} \bigl(e^{B(s,T)\,\Delta L_s} - 1 - B(s,T)\,\Delta L_s\bigr) \tag{56}$$

and

$$d\mathsf{P}(t,T) = \mathsf{P}(t-,T)\,d\widehat{H}(t,T). \tag{57}$$

Remark 3. Starting from equations for $\mathsf{P}(t,T)$ E. Eberlein and S. Raible [128] have analyzed the structure of forward rates and interest rates $f(t,T)$ and $r(t)$ and also considered in detail *hyperbolic Lévy processes*, i.e., Lévy processes such that the random variable L_1 has hyperbolic distribution (see Chapter III, § 1d).

§ 5b. Completeness

1. Proceeding to the issue of completeness in $(B,\mathcal{P})$-models it is worth recalling that for *discrete* time $n \leqslant N < \infty$ and *finitely many* kinds of stock the completeness of an arbitrage-free market is equivalent (by the *Second fundamental theorem*) to the uniqueness of the martingale measure and to the existence of '*S-representation*' for martingales (with respect to some martingale measure).

For a diffusion $(B,\mathcal{P})$-market generated by an m-dimensional Wiener process $W = (W^1,\dots,W^m)$ we have a multidimensional analogue of Theorem 2 in Chapter III, § 3c: each local martingale $M = (M_t,\mathscr{F}_t)$ admits a *representation*

$$M_t = M_0 + \sum_{i=1}^m \int_0^t \psi_i(s)\,dW_s^i \tag{1}$$

with $\mathscr{F}_s$-measurable functions $\psi_i(s)$ such that

$$\int_0^t \psi_i^2(s)\,ds < \infty \quad (\mathsf{P}\text{-a.s.}), \qquad t > 0.$$

As for (B,S)-markets (see § 4b), this representation plays a key role in the study of completeness in $(B,\mathcal{P})$-models.

Let T_0 be a fixed instant of time and let f_{T_0} be a $\mathscr{F}_{T_0}$-measurable pay-off function. We assume that f_{T_0} is bounded ($|f_{T_0}| \leqslant C$) and we shall say that this pay-off function can be *replicated* if there exists a self-financing portfolio $\pi = (\beta, \gamma)$ such that

$$X^{\pi}_{T_0} = f_{T_0} \quad (\mathsf{P}\text{-a.s.}). \tag{2}$$

If this property holds for each T_0 and each bounded $\mathscr{F}_{T_0}$-measurable pay-off function f_{T_0}, then we say that our $(B, \mathcal{P})$-model is *complete*.

Assume that the prices $\mathsf{P}(t,T)$ of T-bonds satisfy the relations

$$d\mathsf{P}(t,T) = \mathsf{P}(t,T)\biggl(r(t)\,dt + \sum_{i=1}^{m} B_i(t,T)\,dW_t^i\biggr). \tag{3}$$

We also assume that $0 < B_i(t,T) \leqslant C = \text{Const}$.

Then the original measure P is martingale in the following sense: the family of prices $(\overline{\mathsf{P}}(t,T))_{t\leqslant T}$ is a local martingale and

$$\overline{X}_t^{\pi} = \overline{X}_0^{\pi} + \sum_{i=1}^{m} \int_0^t \biggl[\int_s^{\infty} B_i(s,T)\overline{\mathsf{P}}(s,T)\,\gamma_s(dT)\biggr]\,dW_s^i. \tag{4}$$

We set

$$M_t = \mathsf{E}\biggl(\frac{f_{T_0}}{B_{T_0}(r)} \biggm| \mathscr{F}_t\biggr), \qquad t \leqslant T_0. \tag{5}$$

Then $M = (M_t, \mathscr{F}_t)$ admits the representation (1) and comparing with (4) we see that to *replicate* the variables M_t, $t \leqslant T_0$, by the value $\overline{X}_t^{\pi}$, $t \leqslant T_0$, of some self-financing portfolio π it is necessary and sufficient that ([38])

$$\psi_i(t) = \int_t^{T_0} B_i(t,T)\overline{\mathsf{P}}(t,T)\,\gamma(dT) \tag{6}$$

($(d\mathsf{P}\times dt)$-a.s.) for $i = 1, \dots, m$ and $t \leqslant T_0$.

If there exists a solution $\{\gamma_t^*(dT),\ t \leqslant T_0,\ T \leqslant T_0\}$, then setting

$$\beta_t^* = M_t - \int_t^{T_0} \overline{\mathsf{P}}(t,T)\,\gamma_t^*(dT), \tag{7}$$

we see that $\pi^* = (\beta^*, \gamma^*)$ is a self-financing portfolio such that

$$\overline{X}_t^{\pi} = M_t, \qquad t \leqslant T_0.$$

In particular, $\overline{X}_{T_0}^{\pi^*} = \dfrac{f_{T_0}}{B_{T_0}(r)}$, and therefore $X_{T_0}^{\pi^*} = f_{T_0}$ (P-a.s.), i.e., we have T_0-*completeness*.

2. EXAMPLE ([36], [38]). Assume that there are *finitely* many, specifically, d bonds in a $(B, \mathcal{P})$-market having maturity times $T_1, \dots, T_d$. (Hence the supports of the measures $\gamma_t(dT)$ are concentrated at the points $\{T_1\}, \dots, \{T_d\}$.) There are m 'sources of randomness' and it seems plausible that we need sufficiently many kinds of bonds to replicate the pay-off function f_{T_0}: d must probably be not less than m.

Let $d = m$. Then the system (6) takes the following form:

$$\psi_i(t) = \sum_{j=1}^{d} B_i(t, T_j)\overline{\mathsf{P}}(t, T_j)\,\gamma_t(\{T_j\}), \tag{8}$$

where $i = 1, \dots, d$.

It is clear from (8) that for each $t \leqslant T_0$ this system has a solution if and only if the matrix $\|B_i(t, T_j)\|$ is invertible.

If $m = d = 1$, then the system (8) turns to the single relation

$$\psi_1(t) = B_1(t, T_1)\mathsf{P}(t, T_1)\,\gamma_t(\{T_1\}), \tag{9}$$

which means that

$$\gamma_t^*(\{T_1\}) = \frac{\psi_1(t)}{B_1(t, T_1)\mathsf{P}(t, T_1)}$$

for $t \leqslant T_1$ and $\gamma_t^*(\{T_1\}) = 0$ for $T_1 < t \leqslant T_0$.

§ 5c. Fundamental Partial Differential Equation of the Term Structure of Bonds

1. By contrast to the *direct* approach to the description of the dynamics of bond prices $\mathsf{P}(t, T)$ by stochastic differential equations (see § 5a), taking the *indirect* approach we assume that the prices $\mathsf{P}(t, T)$ have the following form:

$$\mathsf{P}(t, T) = F\big(t, r(t), T\big), \tag{1}$$

where $r(t)$ is some 'interest rate' taking, as a rule, only nonnegative values.

The indirect approach (1) was historically among the first few. Later on it has been overshadowed (first of all, in theoretical studies) by the direct approach. However, as regards obtaining simple analytic formulas, the approach (1) has retained its importance and is still popular.

2. It should be noted from the outset that this method works only under the assumption that the interest rate process $(r(t))_{t \geqslant 0}$ is a Markov process satisfying the stochastic differential equation

$$dr(t) = a\big(t, r(t)\big)\,dt + b\big(t, r(t)\big)\,dW_t \tag{2}$$

or an equation of 'diffusion-with-jumps' kind (see equation (6) in Chapter III, § 4a).

We shall assume that for each $T > 0$ the function $F^T = F(t, r, T)$ is in the class $C^{1,2}$ (in t and r). Then

$$dF^T = \left(\frac{\partial F^T}{\partial t} + a\frac{\partial F^T}{\partial r} + \frac{1}{2}b^2\frac{\partial^2 F^T}{\partial r^2}\right) dt + b\frac{\partial F^T}{\partial r}\, dW_t. \tag{3}$$

Assuming that $F^T > 0$, we now rewrite this equation as follows (cf. equation (8) in § 5a):

$$dF^T = F^T\Big(A^T(t, r(t))\, dt + B^T(t, r(t))\, dW_t\Big), \tag{4}$$

where

$$A^T(t, r) = \frac{\dfrac{\partial F^T}{\partial t} + a\,\dfrac{\partial F^T}{\partial r} + \dfrac{1}{2}b^2\,\dfrac{\partial^2 F^T}{\partial r^2}}{F^T} \tag{5}$$

and

$$B^T(t, r) = \frac{b\dfrac{\partial F^T}{\partial r}}{F^T}\,. \tag{6}$$

In finding additional conditions on the functions F^T (besides the obvious condition $F^T(T, r(T)) = F(T, r(T), T) = \mathsf{P}(T, T) = 1$) we shall be based on the condition that the $(B, \mathcal{P})$-market in question must be *arbitrage-free*. Then, comparing (4) and formula (8) in § 5a and taking account of relation (25) in § 5a we see that in order that the market be arbitrage-free there must exist a function $\varphi(t)$ such that

$$\frac{A^T(t, r) - r}{B^T(t. r)} = -\varphi(t) \tag{7}$$

for all t and T, $t \leqslant T$. (For each function $\varphi = \varphi(t)$ we can construct the corresponding 'martingale' measure $\widetilde{\mathsf{P}}$ by formula (21) in § 5a.)

In view of (5) and (6), we see from (7) that if *the functions* $F^T = F(t, r, T)$, $T > 0$, *satisfy the fundamental equation*

$$\frac{\partial F}{\partial t} + (a + \varphi b)\,\frac{\partial F}{\partial r} + \frac{1}{2}b^2\,\frac{\partial^2 F}{\partial r^2} = rF, \qquad t \leqslant T, \tag{8}$$

with boundary condition $F(T, r, T) = 1$, $T > 0$, $r \geqslant 0$, *then a* $(B, \mathcal{P})$*-market with* $\mathsf{P}(t, T) = F(t. r(t), T)$ *is arbitrage-free.*

Equation (8) is very similar to the *Fundamental equation of hedge pricing* for stock (see (19) in § 4c). However, there is a crucial difference between these cases: the function $\varphi = \varphi(t)$ in (8) cannot be *defined* in a unique way on the basis of our assumptions and must be set *a priori*. As pointed out above, the martingale measure $\widetilde{\mathsf{P}}$ is defined in terms of this function. Hence the choice of the latter is equivalent to a choice of a 'risk-neutral' measure operative, in investors' opinion, in our $(B, \mathcal{P})$-market.

3. Following notation (11) of Chapter III, § 3f, where we discussed the forward and the backward Kolmogorov equations and the probability representation of solutions to partial differential equations we set now

$$L(s,r) = \big(a(s,r) + \varphi(s)b(s,r)\big)\,\frac{\partial}{\partial r} + \frac{1}{2}b^2(s,r)\,\frac{\partial^2}{\partial r^2}\,. \tag{9}$$

The operator $L(s,r)$ is the *backward* operator of the diffusion Markov process $r = (r(t))_{t\geqslant 0}$ satisfying the stochastic differential equation

$$dr(t) = \Big(a\big(t,r(t)\big) + \varphi(t)b\big(t,r(t)\big)\Big)\,dt + b\big(t,r(t)\big)\,dW_t. \tag{10}$$

Rewriting (8) as

$$-\frac{\partial F}{\partial s} = L(s,r)F - rF, \qquad s \leqslant T, \tag{11}$$

we observe that this equation belongs (see Chapter III, § 3f) to the class of *Feynman–Kac equations* (for the diffusion process $r = (r(t))_{t\geqslant 0}$).

The probabilistic solution of this equation with boundary condition $F(T,r,T) = 1$ can be represented (cf. (19′) in Chapter III, § 3f and see [123], [170], [288] for detail) as follows:

$$F(s,r,T) = \mathsf{E}_{s,r}\left\{\exp\left(-\int_s^T r(u)\,du\right)\right\}, \tag{12}$$

where $\mathsf{E}_{s,r}$ is the expectation with respect to the probability distribution of the process $(r(u))_{s\leqslant u\leqslant T}$ such that $r(s) = r$.

Note that the formula (12), which we have derived under the assumption of the absence of arbitrage, is in perfect accord with the earlier obtained representation (5) in § 5a, because in the Markov case we have

$$\mathsf{E}\left(\exp\left(-\int_s^T r(u)\,du\right) \,\Big|\, \mathscr{F}_s\right) = \mathsf{E}\left(\exp\left(-\int_s^T r(u)\,du\right) \,\Big|\, r_s\right).$$

4. It is worth noting at this point that all the models of the dynamics of stochastic interest rates discussed in Chapter III, § 4a (see (7)–(21)) count among *diffusion Markov models* of type (10).

Their variety is primarily a result of their authors' desire to find analytically treatable models producing results compatible with actually observable data.

As noted in Chapter III, § 4c one important subclass of such analytically treatable models is formed by the (affine) models ([36], [38], [117], [119]) having the representation

$$F\big(t,r(t),T\big) = \exp\big\{\alpha(t,T) - r(t)\beta(t,T)\big\} \tag{13}$$

with deterministic functions $\alpha(t,T)$ and $\beta(t,T)$.

The model that we shall now discuss (borrowed from the above-mentioned papers) can be obtained as follows.

Assume that in (10) we have

$$a(t,r)+\varphi(t)b(t,r)=a_1(t)+ra_2(t)$$

and

$$b(t,r)=\sqrt{b_1(t)+rb_2(t)}.$$

Then (8) takes the following form:

$$\frac{\partial F}{\partial t}+(a_1+ra_2)\,\frac{\partial F}{\partial r}+\frac{1}{2}(b_1+rb_2)\,\frac{\partial^2 F}{\partial r^2}=rF,\qquad t\leqslant T. \tag{14}$$

Seeking the solution of this equation in the form (13) with $F(T,r,T)=1$ we see that $\alpha(t,T)$ and $\beta(t,T)$ are defined by $a_1(t)$, $a_2(t)$, $b_1(t)$, and $b_2(t)$ in accordance with the following relations:

$$\frac{\partial\beta}{\partial t}+a_2\beta-\frac{1}{2}b_2\beta^2=-1,\qquad \beta(T,T)=0 \tag{15}$$

and

$$\frac{\partial\alpha}{\partial t}=a_1\beta-\frac{1}{2}b_1\beta^2,\qquad \alpha(T,T)=0. \tag{16}$$

Relation (15) is the *Riccati equation*. On finding its solution $\beta(t,T)$ we can find $\alpha(t,T)$ from (16), which gives us the affine model (13) with these functions $\alpha(t,T)$ and $\beta(t,T)$.

EXAMPLE. We consider the *Vasiček model* (see (8) in Chapter III, § 4a)

$$dr(t)=\big(\overline{a}-\overline{b}r(t)\big)\,dt+\overline{c}\,dW_t,$$

where $\overline{a}$, $\overline{b}$, and $\overline{c}$ are constants.

Then we see from (15) and (16) that

$$\frac{\partial\beta}{\partial t}-\overline{b}\beta=-1,\qquad \beta(T,T)=0, \tag{17}$$

and

$$\frac{\partial\alpha}{\partial t}=\overline{a}\beta-\frac{1}{2}\overline{c}^2\beta^2,\qquad \alpha(T,T)=0. \tag{18}$$

Consequently,

$$\beta(t,T)=\frac{1}{\overline{b}}\big(1-e^{-\overline{b}(T-t)}\big) \tag{19}$$

and

$$\alpha(t,T)=\frac{\overline{c}^2}{2}\int_t^T\beta^2(s,T)\,ds-\overline{a}\int_t^T\beta(s,T)\,ds.$$

Chapter VIII. Theory of Pricing in Stochastic Financial Models. Continuous Time

1. European Options in Diffusion (B, S)-Stockmarkets

§ 1a. Bachelier's Formula

1. The content of this chapter relates to continuous time, but, conceptually, it is in direct connection with our discussion of the discrete-time case in the sixth chapter.

Here we shall be mostly interested in options. We take them as examples where one can clearly see the role of arbitrage theory and stochastic calculus and the opportunities they give one in calculations related to continuous-time financial models.

2. As mentioned before (Chapter I, § 2a) L. Bachelier was by all means the first person to describe the dynamics of stock prices using models based on 'random walks and their limit cases' (see [12]), i.e., Brownian motions in the contemporary language.

Assuming that the fluctuations of stock prices are similar to a Brownian motion, Bachelier carried out several calculations for the (rational) prices of some *options* traded in France at his time and compared their results with the actual market prices.

Formula (5) below is an updated version of several Bachelier's results on options [12], which is why we call it *Bachelier's formula.*

In the linear *Bachelier model* one considers a (B, S)-market such that the state of the bank account $B = (B_t)_{t \leqslant T}$ remains the same ($B_t \equiv 1$), while the share price $S = (S_t)_{t \leqslant T}$ can be described by a *linear Brownian motion with drift*:

$$S_t = S_0 + \mu t + \sigma W_t, \qquad t \leqslant T, \tag{1}$$

where $W = (W_t)_{t \geqslant 0}$ is a standard Wiener process (a Brownian motion) on some probability space $(\Omega, \mathscr{F}, \mathsf{P})$.

Prices in this model can also take *negative* values, therefore it cannot adequately reflect real life. Nevertheless, its discussion could be of interest from different points of view: historically, it is the first *diffusion* model, on the other hand this model is both *arbitrage-free* and *complete* (see Chapter VII).

We now set

$$Z_T = \exp\left(-\frac{\mu}{\sigma}W_T - \frac{1}{2}\left(\frac{\mu}{\sigma}\right)^2 T\right) \tag{2}$$

and let $\mathscr{F}_t$, $t \leqslant T$, be the σ-algebra generated by the values of a Wiener process $\{W_s, s \leqslant t\}$ and completed by the addition of the sets of P-probability zero.

We define the new measure $\widetilde{\mathsf{P}}_T$ on $(\Omega, \mathscr{F}_T)$ by setting

$$d\widetilde{\mathsf{P}}_T = Z_T \, d\mathsf{P}_T \tag{3}$$

(cf. (8) in Chapter VII, § 4a), where $\mathsf{P}_T = \mathsf{P} \,|\, \mathscr{F}_T$.

Note that $\widetilde{\mathsf{P}}_T$ is a *unique martingale measure* in this model (see Chapter VII, § 4a.5), i.e., the only measure such that $\widetilde{\mathsf{P}}_T \sim \mathsf{P}_T$ and the process $S = (S_t)_{t\leqslant T}$ is a $\widetilde{\mathsf{P}}$-martingale. Moreover, by Girsanov's theorem (Chapter III, § 3e or Chapter VII, § 3b),

$$\operatorname{Law}\big(S_0 + \mu t + \sigma W_t; t \leqslant T \,|\, \widetilde{\mathsf{P}}_T\big) = \operatorname{Law}\big(S_0 + \sigma W_t; t \leqslant T \,|\, \mathsf{P}_T\big). \tag{4}$$

THEOREM (Bachelier's formula). *The rational price $\mathbb{C}_T = \mathbb{C}(f_T; \mathsf{P})$ of the standard European call option with pay-off function $f_T = (S_T - K)^+$ in the model (1) is defined by the formula*

$$\boxed{\mathbb{C}_T = (S_0 - K)\Phi\left(\frac{S_0 - K}{\sigma\sqrt{T}}\right) + \sigma\sqrt{T}\,\varphi\left(\frac{S_0 - K}{\sigma\sqrt{T}}\right)} \tag{5}$$

where

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \qquad \Phi(x) = \int_{-\infty}^{x} \varphi(y)\, dy.$$

In particular, for $S_0 = K$ we have

$$\mathbb{C}_T = \sigma\sqrt{\frac{T}{2\pi}}\,. \tag{6}$$

Proof. Analyzing the proofs of the theorems in §§ 4a,b we can see that the results obtained for the model (5) (with positive prices) in § 4a remain valid for the present model (1). (The 'key' relation (14) in § 4b assumes now the form $\widetilde{\gamma}_t = Z_t^{-1}\Big(\dfrac{\psi_t}{\sigma} + X_t \dfrac{\mu}{\sigma^2}\Big)$ with Z_t as in (2).) Thus, the (B, S)-market now is arbitrage-free, T-complete, and the rational price is

$$\mathbb{C}_T = \mathsf{E}(Z_T f_T) = \mathsf{E}_{\widetilde{\mathsf{P}}_T}(f_T). \tag{7}$$

By (4) and the self-similarity of Wiener processes,

$$\begin{aligned}
\mathsf{E}_{\widetilde{\mathsf{P}}_T}(S_T - K)^+ &= \mathsf{E}_{\widetilde{\mathsf{P}}_T}\big(S_0 + \mu T + \sigma W_T - K\big)^+ \\
&= \mathsf{E}_{\mathsf{P}_T}\big(S_0 - K + \sigma W_T\big)^+ \\
&= \mathsf{E}\big(S_0 - K + \sigma\sqrt{T}\, W_1\big)^+.
\end{aligned} \tag{8}$$

Note that if ξ is a random variable with standard normal distribution $\mathcal{N}(0,1)$, then

$$\begin{aligned} \mathsf{E}(a+b\xi)^+ &= \int_{-a/b}^{\infty} (a+bx)\varphi(x)\,dx = a\Phi\Big(\frac{a}{b}\Big) + b\int_{-a/b}^{\infty} x\varphi(x)\,dx \\ &= a\Phi\Big(\frac{a}{b}\Big) - b\int_{-a/b}^{\infty} d(\varphi(x)) = a\Phi\Big(\frac{a}{b}\Big) + b\,\varphi\Big(\frac{a}{b}\Big) \end{aligned} \tag{9}$$

for $a \in \mathbb{R}$ and $b \geqslant 0$.

Setting

$$a = S_0 - K, \qquad b = \sigma\sqrt{T},$$

we obtain required formula (5) from (7), (8), and (9).

3. Now let $\widetilde{\pi} = (\widetilde{\beta}, \widetilde{\gamma})$ be a strategy (in the class of self-financing portfolios) of initial value $X_0^{\widetilde{\pi}} = \mathbb{C}_T$ that *replicates* the pay-off function f_T, i.e., let $X_T^{\widetilde{\pi}} = f_T$ (P-a.s.).

By Chapter VII, § 4b the value $X^{\widetilde{\pi}} = (X_t^{\widetilde{\pi}})_{t\leqslant T}$ of this strategy satisfies the relation

$$X_t^{\widetilde{\pi}} = \mathsf{E}_{\widetilde{\mathsf{P}}_T}(f_T \mid \mathscr{F}_t). \tag{10}$$

Since $f_T = (S_T - K)^+$ and $S = (S_t)_{t\leqslant T}$ is a Markov process, it follows that

$$\begin{aligned} X_t^{\widetilde{\pi}} &= \mathsf{E}_{\widetilde{\mathsf{P}}_T}\big((S_T - K)^+ \mid \mathscr{F}_t\big) \\ &= \mathsf{E}_{\widetilde{\mathsf{P}}_T}\big(((S_t - K) + (S_T - S_t))^+ \mid S_t\big) \\ &= \mathsf{E}(a+b\xi)^+ = a\Phi\Big(\frac{a}{b}\Big) + b\varphi\Big(\frac{a}{b}\Big), \end{aligned} \tag{11}$$

where $a = S_t - K$ and $b = \sigma\sqrt{T-t}$.

For $0 \leqslant t \leqslant T$ and $S > 0$ we set

$$C(t,S) = (S-K)\Phi\Big(\frac{S-K}{\sigma\sqrt{T-t}}\Big) + \sigma\sqrt{T-t}\,\varphi\Big(\frac{S-K}{\sigma\sqrt{T-t}}\Big). \tag{12}$$

Then we see from (11) that $X_t^{\widetilde{\pi}} = C(t, S_t)$. Simultaneously,

$$dX_t^{\widetilde{\pi}} = \widetilde{\gamma}_t\,dS_t. \tag{13}$$

By Itô's formula for $C(t, S_t)$ we obtain

$$dC(t,S_t) = \frac{\partial C}{\partial S}\,dS_t + \Big(\frac{\partial C}{\partial t} + \frac{1}{2}\sigma^2\frac{\partial^2 C}{\partial S^2}\Big)\,dt. \tag{14}$$

Comparing (14) and (13) and applying Corollary 1 to the Doob–Meyer decomposition (Chapter III, § 5b) we conclude that

$$\widetilde{\gamma}_t = \frac{\partial C}{\partial S}(t, S_t). \tag{15}$$

Differentiating the right-hand side of (12), after simple transformations we obtain

$$\widetilde{\gamma}_t = \Phi\biggl(\frac{S_t - K}{\sigma\sqrt{T-t}}\biggr). \tag{16}$$

The suitable value of $\widetilde{\beta}_t$ can be found from the observation that

$$C(t, S_t) = \widetilde{\beta}_t + \widetilde{\gamma}_t S_t. \tag{17}$$

In other words,

$$\widetilde{\beta}_t = C(t, S_t) - \widetilde{\gamma}_t S_t. \tag{18}$$

In view of (12) and (16) we see that

$$\widetilde{\beta}_t = -K\Phi\biggl(\frac{S_t - K}{\sigma\sqrt{T-t}}\biggr) + \sigma\sqrt{T-t}\,\varphi\biggl(\frac{S_t - K}{\sigma\sqrt{T-t}}\biggr). \tag{19}$$

The following peculiarities of the behavior of $\widetilde{\gamma}_t$ and $\widetilde{\beta}_t$ as $t \uparrow T$ are worth noting.

Assume that close to the terminal instant T the stock prices S_t are higher than K. Then we see from (16) and (19) that

$$\widetilde{\gamma}_t \to 1 \quad \text{and} \quad \widetilde{\beta}_t \to -K \tag{20}$$

as $t \uparrow T$. On the other hand, if $S_t < K$, then

$$\widetilde{\gamma}_t \to 0 \quad \text{and} \quad \widetilde{\beta}_t \to 0 \tag{21}$$

as $t \uparrow T$. Both relations appear to be quite matter-of-course. For if $S_t < K$ close to time T, then the pay-off function f_T vanishes and, clearly, the capital $X_T^{\widetilde{\pi}}$ equal to zero at time T is sufficient for the option writer, which is just the case if (21) holds.

On the other hand, if $S_t > K$ close to time T, then $f_T = S_T - K$, and the seller needs the capital $X_T^{\widetilde{\pi}} = S_T - K$. Since $X_t^{\widetilde{\pi}} = \widetilde{\beta}_t + \widetilde{\gamma}_t S_t$, the required amount will be available if (20) holds, since $X_t^{\widetilde{\pi}} = \widetilde{\beta}_t + \widetilde{\gamma}_t S_t \to S_T - K$.

§ 1b. Black–Scholes Formula. Martingale Inference

1. As already mentioned, the main deficiency of the linear *Bachelier model*

$$S_t = S_0 + \mu t + \sigma W_t \tag{1}$$

is that the prices S_t can assume *negative* values.

A more realistic model is that of a *geometric* (*economic* as some say, see [420]) *Brownian motion*, in which the prices can be expressed by the formula

$$S_t = S_0 e^{H_t}, \tag{2}$$

where

$$H_t = \Big(\mu - \frac{\sigma^2}{2}\Big)t + \sigma W_t. \tag{3}$$

In other words,

$$S_t = S_0 e^{\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W_t}. \tag{4}$$

Using Itô's formula (Chapter III, § 3d) we see that

$$dS_t = S_t(\mu\, dt + \sigma\, dW_t).$$

This is often expressed symbolically as

$$\frac{dS_t}{S_t} = \mu\, dt + \sigma\, dW_t,$$

which emphasizes analogy with the formula

$$\frac{\Delta S_n}{S_{n-1}} = \mu + \sigma \varepsilon_n,$$

which we used above, (e.g., in the Cox–Ross–Rubinstein model in the discrete-time case; see Chapter II, § 1e).

The model of a geometric Brownian motion (2) was suggested by P. Samuelson [420] in 1965; it underlies the *Black–Merton–Scholes model* and the famous *Black–Scholes formula* for the rational price of a standard European call option with pay-off function $f_T = (S_T - K)^+$ discovered by F. Black and M. Scholes [44] and R. Merton [346] in 1973.

2. Thus, we shall consider the Black–Merton–Scholes (B, S)-model and assume that the bank account $B = (B_t)_{t \geqslant 0}$ evolves in accordance with the formula

$$dB_t = rB_t\, dt, \tag{5}$$

whereas the stock prices $S = (S_t)_{t \geqslant 0}$ are governed by a Brownian motion:

$$dS_t = S_t(\mu\, dt + \sigma\, dW_t). \tag{6}$$

Thus, let

$$B_t = B_0 e^{rt}, \tag{7}$$

$$S_t = S_0 e^{\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W_t}. \tag{8}$$

THEOREM (*Black-Scholes formula*). *The rational price $\mathbb{C}_T = \mathbb{C}(f_T; \mathsf{P})$ of a standard European call option with pay-off function $f_T = (S_T - K)^+$ in the model (5)–(6) is described by the formula*

$$\boxed{\mathbb{C}_T = S_0\Phi\biggl(\frac{\ln\frac{S_0}{K} + T\bigl(r + \frac{\sigma^2}{2}\bigr)}{\sigma\sqrt{T}}\biggr) - Ke^{-rT}\Phi\biggl(\frac{\ln\frac{S_0}{K} + T\bigl(r - \frac{\sigma^2}{2}\bigr)}{\sigma\sqrt{T}}\biggr)} \tag{9}$$

In particular, for $S_0 = K$ and $r = 0$ we have

$$\mathbb{C}_T = S_0\biggl[\Phi\Bigl(\frac{\sigma\sqrt{T}}{2}\Bigr) - \Phi\Bigl(-\frac{\sigma\sqrt{T}}{2}\Bigr)\biggr] \tag{10}$$

and $\mathbb{C}_T \sim K\sigma\sqrt{\dfrac{T}{2\pi}}$ as $T \to 0$ (cf. formula (6) in § 1a).

We present the proof of this formula, as borrowed from [44] and [346], in the next section. Here we suggest what one would call a 'martingale' proof; it is based upon our discussion in Chapter VII.

Using notation similar to that in the preceding section we set

$$Z_T = \exp\biggl(-\frac{\mu - r}{\sigma}W_T - \frac{1}{2}\Bigl(\frac{\mu - r}{\sigma}\Bigr)^2 T\biggr), \tag{11}$$

and let $\widetilde{\mathsf{P}}_T$ be a measure on $(\Omega, \mathscr{F}_T)$ such that $d\widetilde{\mathsf{P}}_T = Z_T\, d\mathsf{P}_T$.

By Girsanov's theorem (Chapter VII, § 3b) the process $\widetilde{W} = (\widetilde{W}_t)_{t\leqslant T}$ with $\widetilde{W}_t = W_t + \dfrac{\mu - r}{\sigma}t$ is Wiener with respect to $\widetilde{\mathsf{P}}_T$, therefore

$$\begin{aligned}\operatorname{Law}(\mu t + \sigma W_t;\ t \leqslant T \mid \widetilde{\mathsf{P}}_T) &= \operatorname{Law}(rt + \sigma\widetilde{W}_t;\ t \leqslant T \mid \widetilde{\mathsf{P}}_T)\\ &= \operatorname{Law}(rt + \sigma W_t;\ t \leqslant T \mid \mathsf{P}_T).\end{aligned}$$

Hence

$$\begin{aligned}\operatorname{Law}(S_t;\ t \leqslant T \mid \widetilde{\mathsf{P}}_T) &= \operatorname{Law}\Bigl(S_0 e^{(\mu - \frac{\sigma^2}{2})t + \sigma W_t};\ t \leqslant T \mid \widetilde{\mathsf{P}}_T\Bigr)\\ &= \operatorname{Law}\Bigl(S_0 e^{(r - \frac{\sigma^2}{2})t + \sigma W_t};\ t \leqslant T \mid \mathsf{P}_T\Bigr).\end{aligned} \tag{12}$$

From the theorem in Chapter VII, § 4a we see that, taking the class of 0-admissible strategies $\pi = (\beta, \gamma)$ with $\int_0^T \gamma_u^2 S_u^2\, du < \infty$ (P-a.s.) we can describe the rational price $\mathbb{C}_T = \mathbb{C}(f_T; \mathsf{P})$ by the following formula:

$$\mathbb{C}_T = B_0\, \mathsf{E}_{\widetilde{\mathsf{P}}_T} \frac{f_T}{B_T}\,. \tag{13}$$

Since $f_T = (S_T - K)^+$ in this case, it follows in view of (12) and the **self-similarity** of the Wiener process ($\mathrm{Law}(W_T) = \mathrm{Law}(\sqrt{T}\, W_1)$) that

$$\begin{aligned}
\mathbb{C}_T &= B_0 \mathsf{E}_{\widetilde{\mathsf{P}}_T} \frac{f_T}{B_T} = e^{-rT}\, \mathsf{E}_{\widetilde{\mathsf{P}}_T} (S_T - K)^+ \\
&= e^{-rT}\, \mathsf{E}_{\widetilde{\mathsf{P}}_T} \Big(S_0 e^{\left(\mu - \frac{\sigma^2}{2}\right)T + \sigma W_T} - K \Big)^+ \\
&= e^{-rT}\, \mathsf{E}_{\mathsf{P}_T} \Big(S_0 e^{\left(r - \frac{\sigma^2}{2}\right)T + \sigma W_T} - K \Big)^+ \\
&= e^{-rT}\, \mathsf{E}_{\mathsf{P}_T} \Big(S_0 e^{\left(r - \frac{\sigma^2}{2}\right)T + \sigma \sqrt{T} W_1} - K \Big)^+ \\
&= e^{-rT}\, \mathsf{E}_{\mathsf{P}_T} \Big(S_0 e^{rT} \cdot e^{-\frac{\sigma^2}{2}T + \sigma \sqrt{T} W_1} - K \Big)^+ \\
&= e^{-rT} \mathsf{E} \Big(a e^{b\xi - \frac{b^2}{2}} - K \Big)^+ ,
\end{aligned} \tag{14}$$

where

$$a = S_0 e^{rT}, \quad b = \sigma\sqrt{T}, \qquad \xi \sim \mathcal{N}(0,1). \tag{15}$$

It is an easy calculation (similar to (9) in § 1a) that

$$\mathsf{E}\Big(a e^{b\xi - \frac{b^2}{2}} - K \Big)^+ = a\Phi\Big(\frac{\ln \frac{a}{K} + \frac{1}{2} b^2}{b} \Big) - K\Phi\Big(\frac{\ln \frac{a}{K} - \frac{1}{2} b^2}{b} \Big). \tag{16}$$

Thus, it follows from (14)–(16) that

$$\mathbb{C}_T = S_0 \Phi\Big(\frac{\ln \frac{a}{K} + \frac{1}{2} b^2}{b} \Big) - K e^{-rT} \Phi\Big(\frac{\ln \frac{a}{K} - \frac{1}{2} b^2}{b} \Big).$$

Setting here $a = S_0 e^{rT}$ and $b = \sigma\sqrt{T}$ we arrive at the *Black–Scholes formula* (9), which completes the proof.

Remark 1. Setting

$$y_{\pm} = \frac{\ln \frac{S_0}{K} + T\left(r \pm \frac{\sigma^2}{2}\right)}{\sigma\sqrt{T}},$$

we can write (9) in a more compact form:

$$\mathbb{C}_T = S_0 \Phi(y_+) - K e^{-rT} \Phi(y_-). \tag{17}$$

Let $\mathbb{P}_T$ be the rational price of the standard European *put option* with pay-off function $f_T = (K - S_T)^+$. Then, since

$$\mathbb{P}_T = \mathbb{C}_T - S_0 + K e^{-rT}$$

(cf. 'call-put parity' identity (9) in Chapter VI, § 4d), it follows that

$$\boxed{\begin{aligned}\mathbb{P}_T = -S_0\left[1-\Phi\left(\frac{\ln\frac{S_0}{K}+T\left(r+\frac{\sigma^2}{2}\right)}{\sigma\sqrt{T}}\right)\right] \\ + Ke^{-rT}\left[1-\Phi\left(\frac{\ln\frac{S_0}{K}+T\left(r-\frac{\sigma^2}{2}\right)}{\sigma\sqrt{T}}\right)\right]\end{aligned}} \tag{18}$$

or

$$\mathbb{P}_T = -S_0\Phi(-y_+) + Ke^{-rT}\Phi(-y_-). \tag{19}$$

3. The model in question is T-complete (see Definition 1 in Chapter VII, § 2d), and there exists a 0-admissible strategy $\widetilde{\pi} = (\widetilde{\beta}, \widetilde{\gamma})$ of value $X^{\widetilde{\pi}} = (X_t^{\widetilde{\pi}})_{t\leqslant T}$ such that $X_0^{\widetilde{\pi}} = \mathbb{C}_T$ and $X_T^{\widetilde{\pi}}$ replicates f_T faithfully:

$$X_T^{\widetilde{\pi}} = f_T \quad (\mathsf{P}\text{-a.s.}).$$

By the theorem in Chapter VII, § 4b,

$$\begin{aligned}
X_t^{\widetilde{\pi}} &= B_t\,\mathsf{E}_{\widetilde{\mathsf{P}}_T}\left(\frac{f_T}{B_T}\,\Big|\,\mathscr{F}_t\right) = e^{-r(T-t)}\,\mathsf{E}_{\widetilde{\mathsf{P}}_T}\big((S_T-K)^+\,|\,\mathscr{F}_t\big) \\
&= e^{-r(T-t)}\,\mathsf{E}_{\widetilde{\mathsf{P}}_T}\left(\left(S_t\cdot\frac{S_T}{S_t}-K\right)^+\,\Big|\,\mathscr{F}_t\right) \\
&= e^{-r(T-t)}\,\mathsf{E}_{\widetilde{\mathsf{P}}_T}\Big(\big(S_te^{\left(\mu-\frac{\sigma^2}{2}\right)(T-t)+\sigma(W_T-W_t)}-K\big)^+\,\Big|\,\mathscr{F}_t\Big) \\
&= e^{-r(T-t)}\,\mathsf{E}_{\widetilde{\mathsf{P}}_T}\Big(\big(S_te^{\left(\mu-\frac{\sigma^2}{2}\right)(T-t)+\sigma(W_T-W_t)}-K\big)^+\,\Big|\,S_t\Big) \\
&= e^{-r(T-t)}\,\mathsf{E}_{\mathsf{P}_T}\Big(\big(S_te^{\left(r-\frac{\sigma^2}{2}\right)(T-t)+\sigma(W_T-W_t)}-K\big)^+\,\Big|\,S_t\Big) \\
&= e^{-r(T-t)}\,\mathsf{E}\Big(\big(S_te^{r(T-t)}e^{b\xi-\frac{b^2}{2}}-K\big)^+\,\Big|\,S_t\Big) \\
&= e^{-r(T-t)}\,\mathsf{E}\Big(\big(ae^{b\xi-\frac{b^2}{2}}-K\big)^+\,\Big|\,S_t\Big),
\end{aligned} \tag{20}$$

where

$$a = S_te^{r(T-t)}, \quad b = \sigma\sqrt{T-t}, \qquad \xi \sim \mathscr{N}(0,1),$$

and the variables S_t and $\xi = W_T - W_t$ are independent with respect to the initial measure P.

Taking (16) into account, we see from (20) that the price $C(t, S_t) = X_t^{\widetilde{\pi}}$ has the following expression:

$$\begin{aligned}C(t,S_t) = S_t\Phi\left(\frac{\ln\frac{S_t}{K}+(T-t)\left(r+\frac{\sigma^2}{2}\right)}{\sigma\sqrt{T-t}}\right) \\ - Ke^{-r(T-t)}\Phi\left(\frac{\ln\frac{S_t}{K}+(T-t)\left(r-\frac{\sigma^2}{2}\right)}{\sigma\sqrt{T-t}}\right).\end{aligned} \tag{21}$$

As in § 1a (see subsection 3) we can show that, for an optimal hedging portfolio $\widetilde{\pi} = (\widetilde{\beta}_t, \widetilde{\gamma}_t)_{t \leqslant T}$, we have

$$\widetilde{\gamma}_t = \frac{\partial C}{\partial S}(t, S_t). \tag{22}$$

By (21), after simple transformations we obtain

$$\widetilde{\gamma}_t = \Phi\left(\frac{\ln \frac{S_t}{K} + (T-t)\left(r + \frac{\sigma^2}{2}\right)}{\sigma\sqrt{T-t}}\right) \tag{23}$$

(cf. formula (16) in § 1a), and since $\widetilde{\beta}_t B_t + \widetilde{\gamma}_t S_t = C(t, S_t)$, it follows that

$$\widetilde{\beta}_t = -\frac{K}{B_0} e^{-rT} \Phi\left(\frac{\ln \frac{S_t}{K} + (T-t)\left(r - \frac{\sigma^2}{2}\right)}{\sigma\sqrt{T-t}}\right). \tag{24}$$

It is easy to see that $0 < \widetilde{\gamma}_t < 1$ and $\widetilde{\beta}_t$ is always negative, which indicates borrowing from the bank account under the constraint $-\dfrac{K}{B_0} < \widetilde{\beta}_t$.

As with the *Bachelier model*, we have properties (20) and (21) in § 1a; namely,

if $t \uparrow T$ and $S_t > K$ close to time T then

$$\widetilde{\gamma}_t S_t \to S_t \quad \text{and} \quad \widetilde{\beta}_t B_t \to -K;$$

and

if $t \uparrow T$ and $S_t < K$ close to time T then

$$\widetilde{\gamma}_t S_t \to 0 \quad \text{and} \quad \widetilde{\beta}_t B_t \to 0.$$

Remark 2. The above price $C(t, S_t)$ depends, of course, also on the parameters r and σ specifying the particular model. To indicate this dependence we shall write $C = C(t, s, r, \sigma)$ (with $S_t = s$).

It is often important in practice to have a knowledge of the 'sensitivity' of $C(t, s, r, \sigma)$ to variations of the parameters t, s, r, and σ. The following functions are standard measures of this 'sensitivity' (see, e.g., [36] and [415]):

$$\theta = \frac{\partial C}{\partial t}, \qquad \Delta = \frac{\partial C}{\partial s}, \qquad \rho = \frac{\partial C}{\partial r}, \qquad V = \frac{\partial C}{\partial \sigma}.$$

(Here 'V' is pronounced 'vega'.)

For the *Black–Scholes model*, from (21) we see that

$$\begin{aligned}
\theta &= \frac{s\sigma\varphi(Y_+(T-t))}{2\sqrt{T-t}} - rKe^{-r(T-t)}\Phi\big(y_-(T-t)\big),\\
\Delta &= \Phi\big(y_+(T-t)\big),\\
\rho &= K(T-t)e^{-r(T-t)}\Phi\big(y_-(T-t)\big),\\
V &= s\varphi\big(y_+(T-t)\big)\sqrt{T-t},
\end{aligned}$$

where

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2},$$

$$y_\pm(T-t) = \frac{\ln \frac{s}{K} + (T-t)\left(r \pm \frac{\sigma^2}{2}\right)}{\sigma\sqrt{T-t}}.$$

4. Our calculations of $\mathbb{C}_T$ in (14)–(16) could be carried out in a somewhat different way, on the basis of an appropriate choice of a *discounting process* ('numéraire'), as discussed in Chapter VII, § 1b.

To this end we rewrite (13) with $f_T = (S_T - K)^+$ as follows:

$$\begin{aligned}\mathbb{C}_T &= B_0 \, \mathsf{E}_{\widetilde{\mathsf{P}}_T} \frac{(S_T - K)^+}{B_T} = B_0 \, \mathsf{E}_{\widetilde{\mathsf{P}}_T} \frac{(S_T - K)^+}{B_T} I(S_T > K) \\ &= B_0 \, \mathsf{E}_{\widetilde{\mathsf{P}}_T} \frac{S_T}{B_T} I(S_T > K) - Ke^{-rT} \mathsf{E}_{\widetilde{\mathsf{P}}_T} I(S_T > K). \end{aligned} \tag{25}$$

By (12), the calculation of $\mathsf{E}_{\widetilde{\mathsf{P}}_T} I(S_T > K)$ encounters no complications:

$$\mathsf{E}_{\widetilde{\mathsf{P}}_T} I(S_T > K) = \Phi\left(\frac{\ln \frac{S_0}{K} + T\left(r - \frac{\sigma^2}{2}\right)}{\sigma\sqrt{T}}\right). \tag{26}$$

To calculate $B_0 \, \mathsf{E}_{\widetilde{\mathsf{P}}_T} \dfrac{S_T}{B_T} I(S_T > K)$ we consider the process $\overline{Z} = (\overline{Z}_t)_{t \leqslant T}$ with

$$\overline{Z}_t = \frac{S_t/S_0}{B_t/B_0}. \tag{27}$$

It is important that $\overline{Z}$ is a *positive martingale* (with respect to the 'martingale' measure $\widetilde{\mathsf{P}}_T$) with $\mathsf{E}_{\widetilde{\mathsf{P}}_T} \overline{Z}_T = 1$. Hence we can introduce *another* measure, $\overline{\mathsf{P}}_T$, by setting

$$d\overline{\mathsf{P}}_T = \overline{Z}_T \, d\widetilde{\mathsf{P}}_T. \tag{28}$$

(The measure $\overline{\mathsf{P}}_T$ is called in [434] the *dual*—to $\widetilde{\mathsf{P}}_T$—martingale measure.)

By (7) and (8),

$$\overline{Z}_t = e^{\sigma W_t + \left(\mu - r - \frac{\sigma^2}{2}\right)t} = e^{\sigma \widetilde{W}_t - \frac{\sigma^2}{2} t},$$

where $\widetilde{W}_t = W_t + \dfrac{\mu - r}{\sigma} t$ $(t \leqslant T)$ is a Wiener process with respect to $\widetilde{\mathsf{P}}_T$.

Using Girsanov's theorem (Chapter III, § 3e or Chapter VII, § 3b) it is easy to verify that

$$\overline{W}_t = \widetilde{W}_t - \sigma t \quad \left(= W_t + \left(\frac{\mu - r}{\sigma} - \sigma\right)t\right), \quad t \leqslant T, \tag{29}$$

is a Wiener process with respect to $\overline{\mathsf{P}}_T$. In view of the above notation,

$$B_0\,\mathsf{E}_{\widetilde{\mathsf{P}}_T}\frac{S_T}{B_T}I(S_T>K)=S_0\,\mathsf{E}_{\widetilde{\mathsf{P}}_T}\overline{Z}_T I(S_T>K)=S_0\,\mathsf{E}_{\overline{\mathsf{P}}_T}I(S_T>K)$$

therefore it follows from (25) that

$$\mathbb{C}_T=S_0\,\mathsf{E}_{\overline{\mathsf{P}}_T}I(S_T>K)-Ke^{-rT}\mathsf{E}_{\widetilde{\mathsf{P}}_T}I(S_T>K). \tag{30}$$

By analogy with (12),

$$\begin{aligned}\operatorname{Law}(S_t;t\leqslant T\mid\overline{\mathsf{P}}_T)&=\operatorname{Law}\Big(S_0e^{\left(\mu-\frac{\sigma^2}{2}\right)t+\sigma W_t};\ t\leqslant T\mid\overline{\mathsf{P}}_T\Big)\\&=\operatorname{Law}\Big(S_0e^{\left(r+\frac{\sigma^2}{2}\right)t+\sigma\overline{W}_t};\ t\leqslant T\mid\overline{\mathsf{P}}_T\Big)\\&=\operatorname{Law}\Big(S_0e^{rt}e^{\sigma W_t+\frac{\sigma^2}{2}t};\ t\leqslant T\mid\mathsf{P}_T\Big).\end{aligned}\tag{31}$$

In particular, if $\xi\sim\mathcal{N}(0,1)$, then

$$\operatorname{Law}(S_T\mid\overline{\mathsf{P}}_T)=\operatorname{Law}\Big(S_0e^{\left(r+\frac{\sigma^2}{2}\right)T}\cdot e^{\sigma\sqrt{T}\xi}\mid\mathsf{P}_T\Big).$$

Hence

$$\mathsf{E}_{\overline{\mathsf{P}}_T}I(S_T>K)=\Phi\bigg(\frac{\ln\frac{S_0}{K}+T\left(r+\frac{\sigma^2}{2}\right)}{\sigma\sqrt{T}}\bigg). \tag{32}$$

A combination of (30), (26), and (32) proves the *Black-Scholes formula* (9) for $\mathbb{C}_T$ in a different way, as promised at the beginning of the subsection.

§ 1c. Black–Scholes Formula. Inference Based on the Solution of the Fundamental Equation

1. We now present the original proof of the *Black–Scholes formula* for the *rational price* of option contracts, suggested independently by F. Black and M. Scholes in [44] and R. Merton in [346] (1973).

Of course, the first question before the authors was about the definition of the *rational price*. Their (remarkable in simplicity and efficiency) idea was that this must be just the minimum level of capital allowing the option writer to build a hedging portfolio.

More formally, this can be explained as follows.

Consider a European option contract with maturity date T and pay-off function f_T. Then the *rational (fair) price* Y_t of this contract at a time t, $0\leqslant t\leqslant T$, is (by the *definition* of F. Black, M. Scholes, and R. Merton) the price of a *perfect* European hedge

$$\mathbb{C}_{[t,T]} = \inf\{x \colon \exists\, \pi \text{ with } X_t^\pi = x \text{ and } X_T^\pi = f_T \ (\mathsf{P}\text{-a.s.})\}. \tag{1}$$

(Cf. the appropriate definitions in Chapter VI, § 1b and Chapter VII, § 4b; we also denoted before $\mathbb{C}_{[0,T]}$ by $\mathbb{C}_T$.)

In general, it is not a priori clear whether there exist *perfect* hedges.

The results of Chapter VII, §§ 4a, b show that such hedges do exist in our model of a (B,S)-market and, moreover, $Y_t = \mathbb{C}_{[t,T]}$ is equal to the quantity $B_t \,\mathsf{E}_{\widetilde{\mathsf{P}}_T}\left(\dfrac{f_T}{B_T} \;\Big|\; \mathscr{F}_t\right)$, where $\widetilde{\mathsf{P}}_T$ is the *martingale* measure. This is why we called our scheme in § 1b a 'martingale' inference.

Papers [44], [346] were written before the development of the 'martingale' approach and used another method for the calculation of $Y_t = \mathbb{C}_{[t,T]}$, which we describe now.

Since both process $S = (S_t)_{t\geqslant 0}$ and pay-off function $f_T = (S_T - K)^+$ are Markov, it is natural to *assume* that the $\mathscr{F}_t$-measurable variable Y_t depends on the 'past' only through S_t:

$$Y_t = Y(t, S_t).$$

Assuming that the function $Y = Y(t,S)$ on $[0,T) \times (0,\infty)$ is, in addition, sufficiently smooth (more precisely, $Y \in C^{1,2}$), the authors of [44] and [346] obtain the following *fundamental equation*:

$$\frac{\partial Y}{\partial t} + rS\frac{\partial Y}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 Y}{\partial S^2} = rY \tag{2}$$

with boundary condition

$$Y(T,S) = (S-K)^+. \tag{3}$$

(One can find a derivation of (2) in Chapter VII, § 4c; see equation (9) there.)

The next step to the *Black–Scholes formula* (that is, to the expression for $Y(0,S_0)$) is to find a solution of (2)–(3).

Equation (2) is of *Feynman–Kac* kind (see (19) in Chapter III, § 3f) and it can be solved using the standard techniques developed for such equations.

We consider the new variables

$$\theta = \sigma^2(T-t), \tag{4}$$

$$Z = \ln S + \left(r - \frac{\sigma^2}{2}\right)(T-t) \tag{5}$$

and set

$$V(\theta, Z) = e^{r(T-t)}Y(t,S). \tag{6}$$

In these variables (2)–(3) is equivalent to the following problem:

$$\frac{\partial V}{\partial \theta} - \frac{1}{2}\frac{\partial^2 V}{\partial Z^2} = 0, \tag{7}$$

$$V(0,Z) = (e^Z - K)^+. \tag{8}$$

Relation (7) is the *heat equation*, and by formula (17′) in Chapter III, § 3f the solution to (7)–(8) can be expressed as follows:

$$V(\theta, Z) = \mathsf{E}\big(e^{W_\theta + Z} - K\big)^+, \tag{9}$$

where $W = (W_\theta)$ is a standard Wiener process.

We set

$$a = e^{Z+\frac{\theta}{2}}, \quad b = \sqrt{\theta}, \quad \text{and} \quad \xi \sim \mathcal{N}(0,1).$$

Then

$$\begin{aligned}
\mathsf{E}\big(e^{W_\theta+Z} - K\big)^+ &= \mathsf{E}\Big(e^Z \cdot e^{\sqrt{\theta}W_1} - K\Big)^+ \\
&= \mathsf{E}\Big(e^{Z+\frac{\theta}{2}} \cdot e^{\sqrt{\theta}W_1 - \frac{(\sqrt{\theta})^2}{2}} - K\Big)^+ \\
&= \mathsf{E}\Big(ae^{b\xi - \frac{b^2}{2}} - K\Big)^+.
\end{aligned} \tag{10}$$

Using formula (16) in § 1b we see that

$$\mathsf{E}\big(e^{W_\theta+Z} - K\big)^+ = e^{Z+\frac{\theta}{2}}\Phi\bigg(\frac{Z - \ln K + \theta}{\sqrt{\theta}}\bigg) - K\Phi\bigg(\frac{Z - \ln Z}{\sqrt{\theta}}\bigg). \tag{11}$$

Finally, using notation (4) and (5), by (6) and (11) we obtain the formula

$$\begin{aligned}
Y(t,S) &= e^{-r(T-t)}V(\theta, Z) \\
&= S\Phi\bigg(\frac{\ln\frac{S}{K} + (T-t)\big(r + \frac{\sigma^2}{2}\big)}{\sigma\sqrt{T-t}}\bigg) \\
&\quad - Ke^{-r(T-t)}\Phi\bigg(\frac{\ln\frac{S}{K} + (T-t)\big(r - \frac{\sigma^2}{2}\big)}{\sigma\sqrt{T-t}}\bigg).
\end{aligned} \tag{12}$$

(Cf. (21) in § 1b.)

Setting here $t = 0$ and $S = S_0$, we obtain the required *Black–Scholes formula* (formula (9) in § 1b). As shown in § 1b, the portfolio $\widetilde{\pi} = (\widetilde{\beta}_t, \widetilde{\gamma}_t)_{t\leqslant T}$ with $\widetilde{\gamma}_t = \dfrac{\partial Y}{\partial S}(t, S_t)$, and $\widetilde{\beta}_t = Y(t,S_t) - S_t\widetilde{\gamma}_t$ is a hedge of value $X_T^{\widetilde{\pi}}$ replicating faithfully the pay-off function $f_T = (S_T - K)^+$.

2. In conclusion we make several observations concerning the above two derivations of the *Black–Scholes formula.*

The 'martingale' inference in § 1b is based on the existence of a unique martingale measure in the model of a (B, S)-market in question. This means the absence of arbitrage and enables one to calculate the rational price $\mathbb{C}_T$ by the formula $\mathbb{C}_T = B_0 \, \mathsf{E}_{\widetilde{\mathsf{P}}_T} \dfrac{f_T}{B_T}$, which (for $f_T = (S_T - K)^+$) gives one just the *Black–Scholes formula.*

The approach based on the solution of the '*fundamental equation*' brings one to the same formula. It is worth noting that the lack of arbitrage and the existence of perfect hedges are reflected there by the fact that, due to the unique solubility of (2)–(3), the resulting price $Y(0, S_0)$ is automatically 'arbitrage-free', 'fair': if the price asked for the option is lower than $Y(0, S_0)$ then the seller cannot in general fulfill his obligations, while if it is higher than $Y(0, S_0)$ then the seller will for sure cash a net profit ('have a free lunch'). See Chapter V, § 1b for detail.

§ 1d. Black–Scholes Formula. Case with Dividends

1. We assume again that a (B, S)-market can be described by relations (5) and (6) in § 1b, but the stock also brings dividends (cf. Chapter V, § 1a.6).

More precisely, this means the following. If $S = (S_t)_{t \geqslant 0}$ is the stock *market price* then *with dividends taken into account* the capital $\widetilde{S} = (\widetilde{S}_t)_{t \geqslant 0}$ of the stockholder is assumed to evolve (after discounting) by the formula

$$d\left(\frac{\widetilde{S}_t}{B_t}\right) = d\left(\frac{S_t}{B_t}\right) + \frac{\delta S_t \, dt}{B_t}. \tag{1}$$

Here $\delta \geqslant 0$ is a parameter characterizing the rate of dividend payments. If $B_t \equiv 1$, then it follows from (1) that

$$d\widetilde{S}_t = dS_t + \delta S_t \, dt, \tag{2}$$

so that the increase over time dt in the capital of the stockholder is the sum of the increase dS_t in its market price and the dividends $\delta S_t \, dt$ proportional to S_t.

Since $dS_t = S_t(\mu \, dt + \sigma \, dW_t)$ and

$$d\left(\frac{S_t}{B_t}\right) = \frac{S_t}{B_t}\big((\mu - r) \, dt + \sigma \, dW_t\big), \tag{3}$$

it follows by (1) that

$$d\left(\frac{\widetilde{S}_t}{B_t}\right) = \frac{S_t}{B_t}\big[(\mu - r + \delta) \, dt + \sigma \, dW_t\big]. \tag{4}$$

We set

$$\overline{W}_t = W_t + \frac{\mu - r + \delta}{\sigma} t \tag{5}$$

and

$$\overline{Z}_T = \exp\biggl(-\frac{\mu - r + \delta}{\sigma} W_T - \frac{1}{2}\Bigl(\frac{\mu - r + \delta}{\sigma}\Bigr)^2 T\biggr). \tag{6}$$

Then, defining the measure $\overline{\mathsf{P}}_T$ by the formula

$$d\overline{\mathsf{P}}_T = \overline{Z}_T \, d\mathsf{P}_T,$$

we see by Girsanov's theorem (see Chapter III, § 3e) that $\overline{W} = (\overline{W}_t)_{t \leqslant T}$ is a Wiener process with respect to $\overline{\mathsf{P}}_T$. Hence

$$\begin{aligned}\mathrm{Law}(\mu t + \sigma W_t;\ t \leqslant T \mid \overline{\mathsf{P}}_T) &= \mathrm{Law}\bigl((r-\delta)t + \sigma \overline{W}_t;\ t \leqslant T \mid \overline{\mathsf{P}}_T\bigr) \\ &= \mathrm{Law}\bigl((r-\delta)t + \sigma W_t;\ t \leqslant T \mid \mathsf{P}_T\bigr)\end{aligned}$$

and

$$\mathrm{Law}(S_t;\ t \leqslant T \mid \overline{\mathsf{P}}_T) = \mathrm{Law}\bigl(S_0 e^{(r-\delta-\frac{\sigma^2}{2})t + \sigma W_t};\ t \leqslant T \mid \mathsf{P}_T\bigr). \tag{7}$$

Let $X_t^\pi = \beta_t B_t + \gamma_t \widetilde{S}_t$, $t \leqslant T$, be the value of a self-financing strategy $\pi = (\beta,\gamma)$. Since the discounted capital $\left(\dfrac{X_t^\pi}{B_t}\right)_{t \leqslant T}$ is a martingale with respect to $\overline{\mathsf{P}}_T$, π belongs to the class of 0-admissible strategies with $\int_0^T \gamma_u^2 \widetilde{S}_u^2 \, du < \infty$ (P-a.s.), it follows that

$$\mathsf{E}_{\overline{\mathsf{P}}_T} \frac{X_T^\pi}{B_T} = \frac{X_0^\pi}{B_0}\,.$$

Hence (cf. (13) in § 1b) we obtain that the rational price $\mathbb{C}_T(\delta; r)$ of a call option is expressed by the formula

$$\mathbb{C}_T(\delta; r) = B_0 \mathsf{E}_{\overline{\mathsf{P}}_T} \frac{f_T}{B_T}\,, \tag{8}$$

where $f_T = (S_T - K)^+$.

In view of (7) and formula (16) in § 1b, we see from (8) that

$$\begin{aligned}\mathbb{C}_T(\delta; r) &= e^{-rT} \mathsf{E}_{\overline{\mathsf{P}}_T}\Bigl(S_0 e^{(\mu - \frac{\sigma^2}{2})T + \sigma W_T} - K\Bigr)^+ \\ &= e^{-rT} \mathsf{E}\Bigl(S_0 e^{(r-\delta-\frac{\sigma^2}{2})T + \sigma W_T} - K\Bigr)^+ \\ &= e^{-rT} \mathsf{E}\Bigl(S_0 e^{(r-\delta-\frac{\sigma^2}{2})T + \sigma\sqrt{T} W_1} - K\Bigr)^+ \\ &= S_0 e^{-\delta T} \Phi\biggl(\frac{\ln\frac{S_0}{K} + T(r - \delta + \frac{\sigma^2}{2})}{\sigma\sqrt{T}}\biggr) \\ &\qquad - K e^{-rT} \Phi\biggl(\frac{\ln\frac{S_0}{K} + T(r - \delta - \frac{\sigma^2}{2})}{\sigma\sqrt{T}}\biggr).\end{aligned} \tag{9}$$

Now let $\mathbb{P}_T(\delta; r)$ be the corresponding price of a put option in the case of dividend payments. It is easy to see that $\mathbb{C}_T(\delta; r)$ and $\mathbb{P}_T(\delta; r)$ are connected by the following identity of 'call-put parity' (cf. (9) in Chapter VI, § 4d):

$$\mathbb{P}_T(\delta; r) = \mathbb{C}_T(\delta; r) - S_0 e^{-\delta T} + K e^{-rT}. \tag{10}$$

Comparing (9) and formula (9) in § 1b for $\mathbb{C}_T(0; r)$ ($\equiv \mathbb{C}_T$) and taking (10) into account we arrive at the following conclusion.

THEOREM. *The rational prices $\mathbb{C}_T(\delta; r)$ and $\mathbb{P}_T(\delta; r)$ of call and put options in the case of dividend payments are described by the formulas*

$$\boxed{\mathbb{C}_T(\delta; r) = e^{-\delta T} \mathbb{C}_T(0; r - \delta)} \tag{11}$$

and

$$\boxed{\mathbb{P}_T(\delta; r) = e^{-\delta T} \mathbb{P}_T(0; r - \delta)} \tag{12}$$

where $\mathbb{C}_T(0; r - \delta)$ and $\mathbb{P}_T(0; r - \delta)$ can be defined by the right-hand sides of (9) *and formula* (18) *in* § 1b (*'the case without dividends'*) *with r replaced by $r - \delta$.*

2. American Options in Diffusion (B, S)-Stockmarkets. Case of an Infinite Time Horizon

§ 2a. Standard Call Option

1. In the considerations of options and other derivative financial instruments one must sharply distinguish between two cases: of the time parameter t ranging over a *finite* interval $[0, T]$, and of t in the *infinite interval* $[0, \infty)$. Of course, the second case smacks of idealization, but it is much easier to study than the first case, when the decisions taken at time t depend significantly on the time $T - t$ *remaining* till the expiration of the contract.

This explains why we start with the discussion of the second case. We consider finite intervals $[0, T]$ in § 3.

2. We assume that we have a standard Wiener process $W = (W_t)_{t \geqslant 0}$ on a filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t \geqslant 0}, \mathsf{P})$ and that our diffusion (B, S)-market has the following structure:

$$dB_t = r B_t \, dt, \quad B_0 > 0, \tag{1}$$

$$dS_t = S_t(\mu \, dt + \sigma \, dW_t), \quad S_0 > 0. \tag{2}$$

For a *standard discounted call option* the pay-off function has, by definition, the following structure:

$$f_t = e^{-\lambda t} g(S_t), \tag{3}$$

where $g(x) = (x - K)^+$, $x \in E = (0, \infty)$, $\lambda \geqslant 0$.

By analogy with the discrete-time case we set

$$V^*(x) = \sup B_0 \, \widetilde{\mathsf{E}}_x \frac{f_\tau}{B_\tau} , \tag{4}$$

where the supremum is taken over the class of finite stopping times

$$\mathfrak{M}_0^\infty = \bigl\{ \tau = \tau(\omega) \colon 0 \leqslant \tau(\omega) < \infty, \ \omega \in \Omega \bigr\}, \tag{5}$$

and $\widetilde{\mathsf{E}}_x$ is the expectation with respect to the martingale measure $\widetilde{\mathsf{P}}_x$ such that the process $S = (S_t)_{t\geqslant 0}$ has a stochastic differential

$$dS_t = S_t(r\,dt + \sigma\,dW_t), \qquad S_0 = x \tag{6}$$

with respect to this measure.

To simplify the notation we assume from the very beginning that $\mu = r$. Making this assumption we can drop the sign '$\sim$' in our notation for $\widetilde{\mathsf{P}}_x$ and $\widetilde{\mathsf{E}}_x$.

Thus, let

$$V^*(x) = \sup_{\tau\in\mathfrak{M}_0^\infty} \mathsf{E}_x e^{-(\lambda+r)\tau}(S_\tau - K)^+. \tag{7}$$

It is reasonable in many problems to consider, alongside $\mathfrak{M}_0^\infty$, also the class

$$\overline{\mathfrak{M}}_0^\infty = \{\tau = \tau(\omega)\colon 0 \leqslant \tau(\omega) \leqslant \infty,\ \omega\in\Omega\}$$

of Markov times that can also assume the value $+\infty$ and to set

$$\overline{V}^*(x) = \sup_{\tau\in\overline{\mathfrak{M}}_0^\infty} \mathsf{E}_x e^{-(\lambda+r)\tau}(S_\tau - K)^+ I(\tau<\infty). \tag{8}$$

Finding $V^*(x)$ and $\overline{V}^*(x)$ is in direct relation to the calculations for standard American call options because the values of $V^*(x)$ and $\overline{V}^*(x)$ are just the rational prices, provided that the buyer can choose the exercise time in the class $\mathfrak{M}_0^\infty$ or $\overline{\mathfrak{M}}_0^\infty$ and $S_0 = x$. (The case of $\tau = \infty$ corresponds to ducking the exercise of the option.) The proof of this assertion in the discrete-time case can be carried out in the same way as the proof of Theorem 1 in Chapter VI, § 2c. The changes in the continuous-time case are not very essential: see, for instance, [33], [265], or [281] for greater detail. Moreover, if τ^* and $\overline{\tau}^*$ are optimal times delivering the solutions to (7) and (8), respectively, then they are also the optimal strike times (in the classes $\mathfrak{M}_0^\infty$ and $\overline{\mathfrak{M}}_0^\infty$).

3. Embarking on the discussion of optimal stopping problems (7) and (8) we single out the (noninteresting) case of $\lambda = 0$ first.

In that case

$$e^{-rt}(S_t - K)^+ = \Bigl(S_0 e^{\sigma W_t - \frac{\sigma^2}{2}t} - Ke^{-rt}\Bigr)^+,$$

which shows that the process $(e^{-rt}(S_t-K)^+)_{t\geqslant 0}$ is a *submartingale* and therefore if $\tau\in\mathfrak{M}_0^T$, i.e., $\tau(\omega)\leqslant T$ for $\omega\in\Omega$, then

$$\mathsf{E}_x e^{-r\tau}(S_\tau - K)^+ \leqslant \mathsf{E}_x e^{-rT}(S_T - K)^+ \leqslant x. \tag{9}$$

By the *Black–Scholes formula* (see (9) in § 1b),

$$\mathsf{E}_x e^{-rT}(S_T - K)^+ \to x \quad \text{as} \quad T\to\infty, \tag{10}$$

for each $r \geqslant 0$.

Since $V^*(x) = \lim_{T\to\infty} V_T^*(x)$ in our case, where

$$V_T^*(x) = \sup_{\tau\in\mathfrak{M}_0^T} \mathsf{E}_x e^{-(\lambda+r)\tau}(S_\tau - K)^+ \tag{11}$$

(see [441; Chapter 3] and cf. Chapter VI, § 5b), it follows by (9) and (10) that if $\lambda = 0$ and $r \geqslant 0$, then 'the observations must be continued as long as possible'. More precisely, for each $x > 0$ and each $\varepsilon > 0$ there exists a deterministic instant $T_{x,\varepsilon}$ such that

$$\mathsf{E}_x e^{-rT_{x,\varepsilon}}(S_{T_{x,\varepsilon}} - K)^+ \geqslant x - \varepsilon$$

and $T_{x,\varepsilon} \to \infty$ as $\varepsilon \to 0$.

4. We formulate now the main results obtained for the optimal stopping problems (7) and (8) in the case $\lambda > 0$.

THEOREM. *If $\lambda > 0$, then for each $x \in (0,\infty)$ we have*

$$V^*(x) = \overline{V}^*(x) = \begin{cases} c^* x^{\gamma_1}, & x < x^*, \\ x - K, & x \geqslant x^*, \end{cases} \tag{12}$$

where

$$\gamma_1 = \left(\frac{1}{2} - \frac{r}{\sigma^2}\right) + \sqrt{\left(\frac{1}{2} - \frac{r}{\sigma^2}\right)^2 + \frac{2(\lambda+r)}{\sigma^2}}\,, \tag{13}$$

$$c^* = \gamma_1^{-\gamma_1}\left(\frac{\gamma_1 - 1}{K}\right)^{\gamma_1 - 1}, \tag{14}$$

$$x^* = K\,\frac{\gamma_1}{\gamma_1 - 1}\,. \tag{15}$$

There exists an optimal stopping time in the class $\overline{\mathfrak{M}}_0^\infty$, namely, the time

$$\tau^* = \inf\{t \geqslant 0\colon S_t \geqslant x^*\}. \tag{16}$$

Moreover,

$$\mathsf{P}_x(\tau^* < \infty) = \begin{cases} 1 & \text{if } r \geqslant \dfrac{\sigma^2}{2} \text{ or } x \geqslant x^*, \\[2ex] \left(\dfrac{x}{x^*}\right)^{1-\frac{2r}{\sigma^2}} & \text{if } r < \dfrac{\sigma^2}{2} \text{ and } x < x^*. \end{cases} \tag{17}$$

We present two proofs of these results below. The first is based on the 'Markovian' approach to optimal stopping problems and is conceptually similar to the proof in the discrete-time case (see Chapter VI, § 5b). The second is based on some 'martingale' ideas used in [32] and on the transition to the 'dual' probability measure (see § 1b.4).

5. The first proof. We consider optimal stopping problems that are slightly more general than (7) and (8).

Let

$$V^*(x) = \sup_{\tau \in \mathfrak{M}_0^\infty} \mathsf{E}_x e^{-\beta\tau} g(S_\tau), \tag{18}$$

$$\overline{V}^*(x) = \sup_{\tau \in \overline{\mathfrak{M}}_0^\infty} \mathsf{E}_x e^{-\beta\tau} g(S_\tau) I(\tau < \infty) \tag{19}$$

be the prices in the optimal stopping problem for the Markov process $S=(S_t, \mathscr{F}_t, \mathsf{P}_x)$, $x \in E = (0,\infty)$, where P_x is the probability distribution for the process S with $S_0 = x$, $\beta > 0$, and $g = g(x)$ is some Borel function.

If $g = g(x)$ is nonnegative and continuous, then the general theory of optimal stopping rules for Markov processes (see [441; Chapter 3] and cf. Theorem 4 in Chapter VI, § 2a) says that:

(a) $V^*(x) = \overline{V}^*(x)$, $x \in E$; (20)

(b) $V^*(x)$ is the *smallest β-excessive majorant* of $g(x)$, i.e.. the smallest function $V(x)$ such that

$$V(x) \geqslant g(x) \quad \text{and} \quad V(x) \geqslant e^{-\beta t} T_t V(x), \tag{21}$$

where $T_t V(x) = \mathsf{E}_x V(S_t)$;

(c) $V^*(x) = \lim\limits_n \lim\limits_N Q_n^N g(x)$, (22)

where

$$Q_n g(x) = \max\bigl(g(x), e^{-\beta\cdot 2^n} T_{2^{-n}} g(x)\bigr); \tag{23}$$

(d) if $\mathsf{E}_x\Bigl[\sup\limits_t e^{-\beta t} g(S_t)\Bigr] < \infty$, then for each $\varepsilon > 0$ the instant

$$\tau_\varepsilon = \inf\bigl\{t\colon V^*(S_t) \leqslant e^{-\beta t} g(S_t) + \varepsilon\bigr\} \tag{24}$$

is an ε-optimal stopping time in the class $\mathfrak{M}_0^\infty$, i.e., $\mathsf{P}_x(\tau_x < \infty) = 1$, $x \in E$, and $V^*(x) - \varepsilon \leqslant \mathsf{E}_x e^{-\beta\tau_\varepsilon} g(S_{\tau_\varepsilon})$;

(e) if

$$\tau_0 = \inf\bigl\{t\colon V^*(S_t) \leqslant e^{-\beta t} g(S_t)\bigr\}$$

is an optimal stopping time ($\mathsf{P}_x(\tau_0 < \infty) = 1$, $x \in E$), then it is *optimal* in the class $\mathfrak{M}_0^\infty$:

$$V^*(x) = \mathsf{E}_x e^{-\beta\tau_0} g(S_{\tau_0}), \qquad x \in E;$$

moreover, if τ_1 is another optimal stopping time, then $\mathsf{P}_x(\tau_0 \leqslant \tau_1) = 1$, $x \in E$, i.e., τ_0 is the *smallest* optimal stopping time.

Let $C^* = \{x \in E\colon V^*(x) > g(x)\}$ and let $D^* = \{x \in E\colon V^*(x) = g(x)\}$.

It is easy to see from (22) and (23) (cf. Chapter VI, § 5b) that the structure of $V^* = V^*(x)$ is rather simple: this is a *downwards convex* function on $E = (0,\infty)$

majorizing $g = g(x)$. In addition, there exists x^* such that $C^* = \{x \colon x < x^*\}$ and $D^* = \{x \colon x \geqslant x^*\}$.

Hence the solution of the problem (7) and (8) is reduced to finding x^* and, of course, the function $V^*(x)$ $(= \overline{V}^*(x))$.

Analyzing the arguments used in the corresponding discrete-time problem in Chapter VI, § 5b.6, one easily understands that the required value of x^* and the function $V^*(x)$, the *smallest* $(\lambda + r)$-excessive majorant of $g(x)$, must be solutions of the following *Stephan*, or *free-boundary, problem* (see [441; 3.8]):

$$L\widetilde{V}(x) = (\lambda + r)\widetilde{V}(x), \quad x < \widetilde{x}, \tag{25}$$

$$\widetilde{V}(x) = g(x), \quad x \geqslant \widetilde{x}, \tag{26}$$

$$\left.\frac{d\widetilde{V}(x)}{dx}\right|_{x \uparrow \widetilde{x}} = \left.\frac{dg(x)}{dx}\right|_{x \downarrow \widetilde{x}}, \tag{27}$$

where

$$L = rx\,\frac{\partial}{\partial x} + \frac{\sigma^2}{2}\,x^2\,\frac{\partial^2}{\partial x^2} \tag{28}$$

is the infinitesimal operator (see [126]) of the process $S = (S_t)_{t \geqslant 0}$ with stochastic differential

$$dS_t = S_t(r\,dt + \sigma\,dW_t).$$

We shall seek a solution of (25) (in the so far unknown domain $(0, \widetilde{x})$) in the following form:

$$V(x) = cx^\gamma. \tag{29}$$

Then we obtain the following equation for γ:

$$\gamma^2 - \left(1 - \frac{2r}{\sigma^2}\right)\gamma - \frac{2(\lambda + r)}{\sigma^2} = 0. \tag{30}$$

To simplify the notation we assume that $\sigma^2 = 1$. $\Bigl($If $\sigma^2 \neq 1$, then one must make the change $r \to \dfrac{r}{\sigma^2}$, $\lambda \to \dfrac{\lambda}{\sigma^2}$ in the answers.$\Bigr)$

Equation (30) with $\sigma^2 = 1$ has two roots,

$$\gamma_1 = \left(\frac{1}{2} - r\right) + \sqrt{\left(\frac{1}{2} - r\right)^2 + 2(\lambda + r)} \tag{31}$$

and

$$\gamma_2 = \left(\frac{1}{2} - r\right) - \sqrt{\left(\frac{1}{2} - r\right)^2 + 2(\lambda + r)}. \tag{32}$$

Since $\lambda > 0$, it follows that $\gamma_1 > 1$. (If $\lambda = 0$, then $\gamma_1 = 1$.) The root γ_2 is negative. Hence the general solution of (25) is as follows:

$$\widetilde{V}(x) = c_1 x^{\gamma_1} + c_2 x^{\gamma_2}. \tag{33}$$

As in the discrete-time case (Chapter VI, § 5b), we see from (33) that $c_2 = 0$ since otherwise $V(x) \to \infty$ as $x \downarrow 0$, which is impossible in our context (we must have $V^*(x) \geqslant 0$ and $V^*(x) \leqslant x$).

Thus, $\widetilde{V}(x) = c_1 x^{\gamma_1}$ for $x < \widetilde{x}$, where c_1 and the 'free' boundary $\widetilde{x}$ are the unknowns. To find them we use condition (26) and condition (27) of 'smooth pasting'.

Condition (26) shows that

$$c_1 \widetilde{x}^{\gamma_1} = x - K. \tag{34}$$

Condition (27) takes the form

$$c_1 \gamma_1 \widetilde{x}^{\gamma_1 - 1} = 1. \tag{35}$$

We see from these two relations that

$$\widetilde{x} = K \frac{\gamma_1}{\gamma_1 - 1}, \qquad c_1 = \gamma_1^{-\gamma_1} \left(\frac{\gamma_1 - 1}{K} \right)^{\gamma_1 - 1}. \tag{36}$$

Thus, the solution $\widetilde{V}(x)$ to (25)–(27) can be represented as follows:

$$\widetilde{V}(x) = \begin{cases} c_1 x^{\gamma_1}, & x < \widetilde{x}, \\ x - K, & x \geqslant \widetilde{x}, \end{cases} \tag{37}$$

where $\widetilde{x}$ and c_1 are defined by (36).

Remark. If $K = 1$, then $\widetilde{V}(x)$ is just the function $\widetilde{V}(x)$ defined by (39) in Chapter VI, § 5b, which is not that surprising if we take into account (22) and our method of finding $\widetilde{V}(x)$ in Chapter VI.

We shall have proved the theorem once we shall have shown that the function $\widetilde{V}(x)$ so obtained is the price $V^*(x)$ (see (7)) and the time

$$\widetilde{\tau} = \inf\{t \geqslant 0 \colon S_t \geqslant \widetilde{x}\}$$

is optimal in the class $\overline{\mathfrak{M}}_0^\infty$ (and in the class $\mathfrak{M}_0^\infty$ if $\mathsf{P}_x(\widetilde{\tau} < \infty) = 1$).

To this end it is obviously suffices to use the following test: for $x \in E = (0, \infty)$ we must have

(A) $\widetilde{V}(x) = \mathsf{E}_x e^{-(\lambda + r)\widetilde{\tau}} (S_{\widetilde{\tau}} - K)^+ I(\widetilde{\tau} < \infty)$

and

(B) $\widetilde{V}(x) \geqslant \mathsf{E}_x e^{-(\lambda + r)\tau} (S_\tau - K)^+ I(\tau < \infty)$ for each $\tau \in \overline{\mathfrak{M}}_0^\infty$.

Further, since $(S_{\widetilde{\tau}} - K)^+ I(\widetilde{\tau} < \infty) = \widetilde{V}(S_{\widetilde{\tau}}) I(\widetilde{\tau} < \infty)$ and $\widetilde{V}(x) \geqslant (x - K)^+$, for (A) and (B) we must verify the following conditions:

(A′) $\widetilde{V}(x) = \mathsf{E}_x e^{-(\lambda+r)\widetilde{\tau}} \widetilde{V}(S_{\widetilde{\tau}}) I(\widetilde{\tau} < \infty)$

and

(B′) $\widetilde{V}(x) \geqslant \mathsf{E}_x e^{-(\lambda+r)\tau} \widetilde{V}(S_\tau) I(\tau < \infty)$ for each $\tau \in \overline{\mathfrak{M}}_0^\infty$ and $x \in E$.

Usually, the verification of (A′) and (B′) is based on Itô's formula for $\widetilde{V}(x)$ (more precisely, we use its generalization, the *Itô–Meyer formula*). It proceeds as follows.

Let $V = V(x)$ be a function in the class C^2, i.e., a function with continuous second derivative. Then the 'classical' Itô's formula (Chapter III, § 5c) for the function $F(t,x) = e^{-(\lambda+r)t} V(x)$ and the process $S = (S_t)_{t \geqslant 0}$ brings one to the following representation:

$$e^{-(\lambda+r)t} V(S_t) = V(S_0) + \int_0^t e^{-(\lambda+r)u} \Big[LV(S_u) - (\lambda + r) V(S_u) \Big]\, du + \int_0^t e^{-(\lambda+r)u} \sigma S_u V'(S_u)\, dW_u. \tag{38}$$

Considering now the function $\widetilde{V}(x)$ defined in (37) we can observe that it is in the class C^2 for $x \in E = (0,\infty)$ outside *one point* $x = \widetilde{x}$, so that one can anticipate (38) also for $V(x) = \widetilde{V}(x)$ if one chooses a suitable interpretation of the derivative at $x = \widetilde{x}$.

In our case $\widetilde{V}(x)$ is (downwards) convex and its first derivative $\widetilde{V}'(x)$ is well defined and continuous for all $x \in E = (0,\infty)$; its second derivative $\widetilde{V}''(x)$ is defined for $x \in E = (0,\infty)$ distinct from $\widetilde{x}$, where we have the well-defined limits

$$\widetilde{V}''_-(\widetilde{x}) = \lim_{x \uparrow \widetilde{x}} \widetilde{V}''(x) \quad \text{and} \quad \widetilde{V}''_+(\widetilde{x}) = \lim_{x \downarrow \widetilde{x}} \widetilde{V}''(x).$$

There exists a generalization of Itô's formula in stochastic calculus obtained by P.-A. Meyer for a function $V(x)$ that is a *difference of two convex functions.* (See, e.g. [248; (5.52)] or the *Itô–Meyer formula* in [395; IV].)

Our function $\widetilde{V}(x)$ is (downwards) *convex* and for $F(t,x) = e^{-(\lambda+r)t} \widetilde{V}(x)$ and $S = (S_t)_{t \geqslant 0}$ we have the *Itô–Meyer formula*, which looks the same as (38), but where $\widetilde{V}''(\widetilde{x})$ is replaced by, say, $\widetilde{V}''_-(\widetilde{x})$.

Having agreed about this we obtain

$$e^{-(\lambda+r)t} \widetilde{V}(S_t) - \widetilde{V}(S_0) = \int_0^t e^{-(\lambda+r)u} \big[L\widetilde{V}(S_u) - (\lambda + r)\widetilde{V}(S_u) \big]\, du + M_t, \tag{39}$$

where

$$M_t = \int_0^t e^{-(\lambda+r)u} \sigma S_u \widetilde{V}'(S_u)\, dW_u. \tag{40}$$

It is worth noting that

$$L\widetilde{V}(x) - (\lambda + r)\widetilde{V}(x) = 0 \tag{41}$$

for $x < \widetilde{x}$ (by (25)), and it is a straightforward calculation that this equality holds also for $x = \widetilde{x}$, while for $x > \widetilde{x}$ we have

$$L\widetilde{V}(x) - (\lambda + r)\widetilde{V}(x) \leqslant 0. \tag{42}$$

By (39), (41), and (42) we obtain (for $S_0 = x$)

$$\widetilde{V}(x) \geqslant e^{-(\lambda+r)t}\widetilde{V}(S_t) - M_t. \tag{43}$$

As is clear from (40), the process $M = (M_t)_{t\geqslant 0}$ is a local martingale. Let (τ_n) be some localizing sequence and let $\tau \in \overline{\mathfrak{M}}_0^\infty$. Then by (43) we obtain

$$\begin{aligned}
\widetilde{V}(x) &\geqslant \mathsf{E}_x e^{-(\lambda+r)(\tau_n\wedge\tau)}\widetilde{V}(S_{\tau_n\wedge\tau}) - \mathsf{E}M_{\tau_n\wedge\tau} \\
&= \mathsf{E}_x e^{-(\lambda+r)(\tau_n\wedge\tau)}\widetilde{V}(S_{\tau_n\wedge\tau}) \\
&= \mathsf{E}_x e^{-(\lambda+r)(\tau_n\wedge\tau)}\widetilde{V}(S_{\tau_n\wedge\tau})I(\tau < \infty)
\end{aligned}$$

and, by Fatou's lemma,

$$\begin{aligned}
\widetilde{V}(x) &\geqslant \varliminf_n \mathsf{E}_x e^{-(\lambda+r)(\tau_n\wedge\tau)}\widetilde{V}(S_{\tau_n\wedge\tau})I(\tau < \infty) \\
&\geqslant \mathsf{E}_x e^{-(\lambda+r)\tau}\widetilde{V}(S_\tau)I(\tau < \infty),
\end{aligned}$$

which proves (B$'$).

We now claim (A$'$).

If $x \in \widetilde{D} = \{x\colon x \geqslant \widetilde{x}\}$, then $\mathsf{P}_x(\widetilde{\tau} = 0) = 1$, and property (A$'$) is obvious.

Now let $\widetilde{x} \in \widetilde{C} = \{x\colon x < \widetilde{x}\}$. Then by (41) and (39) we obtain

$$\widetilde{V}(x) = e^{-(\lambda+r)(\tau_n\wedge\widetilde{\tau})}\widetilde{V}(S_{\tau_n\wedge\widetilde{\tau}}) - M_{\tau_n\wedge\widetilde{\tau}},$$

so that

$$\begin{aligned}
\widetilde{V}(x) &= \mathsf{E}_x e^{-(\lambda+r)(\tau_n\wedge\widetilde{\tau})}\widetilde{V}(S_{\tau_n\wedge\widetilde{\tau}}) \\
&= \mathsf{E}_x e^{-(\lambda+r)(\tau_n\wedge\widetilde{\tau})}\widetilde{V}(S_{\tau_n\wedge\widetilde{\tau}})I(\widetilde{\tau} < \infty) \\
&\qquad + \mathsf{E}_x e^{-(\lambda+r)(\tau_n\wedge\widetilde{\tau})}\widetilde{V}(S_{\tau_n})I(\widetilde{\tau} = \infty).
\end{aligned} \tag{44}$$

Since

$$
\begin{aligned}
0 &\leqslant e^{-(\lambda+r)(\tau_n\wedge\widetilde{\tau})}\widetilde{V}(S_{\tau_n\wedge\widetilde{\tau}})I(\widetilde{\tau}<\infty) \\
&\leqslant \sup_{t\leqslant\widetilde{\tau}}\big[e^{-(\lambda+r)t}\widetilde{V}(S_t)\big]I(\widetilde{\tau}<\infty) \\
&\leqslant \sup_{t\leqslant\widetilde{\tau}}\big[e^{-\lambda t}e^{\sigma W_t-\frac{\sigma^2}{2}t}\big]I(\widetilde{\tau}<\infty) \\
&\leqslant \sup_{t\geqslant 0}\big[e^{-\lambda t}e^{\sigma W_t-\frac{\sigma^2}{2}t}\big]
\end{aligned}
$$

and

$$\mathsf{E}\sup_{t\geqslant 0}\big[e^{-\lambda t}e^{\sigma W_t-\frac{\sigma^2}{2}t}\big]<\infty \tag{45}$$

(see Corollary 2 to Lemma 1 below), it follows from Lebesgue's dominated convergence theorem that

$$\lim_n \mathsf{E}_x e^{-(\lambda+r)(\tau_n\wedge\widetilde{\tau})}\widetilde{V}(S_{\tau_n\wedge\widetilde{\tau}})I(\widetilde{\tau}<\infty)=\mathsf{E}_x e^{-(\lambda+r)\widetilde{\tau}}\widetilde{V}(S_{\widetilde{\tau}})I(\widetilde{\tau}<\infty). \tag{46}$$

Further, $\widetilde{V}(S_{\tau_n})\leqslant\widetilde{V}(\widetilde{x})<\infty$ on the set $\{\omega\colon \widetilde{\tau}=\infty\}$, therefore

$$\lim_n \mathsf{E}_x e^{-(\lambda+r)\tau_n}\widetilde{V}(S_{\tau_n})I(\widetilde{\tau}=\infty)=0. \tag{47}$$

Required property (A$'$) is a consequence of (44), (46), and (47).

To complete the proof of the theorem we must verify property (46) and prove (17) (with $\tau^*=\widetilde{\tau}$). We shall establish the following result to this end.

LEMMA 1. *For $x\geqslant 0$, $\mu\in\mathbb{R}$, and $\sigma>0$ we have*

$$\mathsf{P}\Big(\max_{s\leqslant t}(\mu s+\sigma W_s)\leqslant x\Big)=\Phi\Big(\frac{x-\mu t}{\sigma\sqrt{t}}\Big)-e^{\frac{2\mu x}{\sigma^2}}\Phi\Big(\frac{-x-\mu t}{\sigma\sqrt{t}}\Big), \tag{48}$$

where $\Phi(x)=\dfrac{1}{\sqrt{2\pi}}\displaystyle\int_{-\infty}^{x}e^{-\frac{y^2}{2}}\,dy.$

Proof. For simplicity, let $\sigma^2=1$. By Girsanov's theorem (Chapter III, § 3e or Chapter VII, § 3b)

$$
\begin{aligned}
&\mathsf{P}\Big(\max_{s\leqslant t}(\mu s+W_s)>x,\ \mu t+W_t\leqslant x\Big) \\
&\quad=\mathsf{E}I\Big(\max_{s\leqslant t}(\mu s+W_s)>x,\ \mu t+W_t\leqslant x\Big) \\
&\quad=\mathsf{E}\exp\Big(\mu W_t-\frac{\mu^2}{2}t\Big)\,I\Big(\max_{s\leqslant t}W_s>x,\ W_t\leqslant x\Big).
\end{aligned} \tag{49}
$$

We set $T_x = \inf\{t \geqslant 0\colon W_t = x\}$. Then, by D. André's *reflection principle*

$$\widetilde{W}_t = W_t I(t \leqslant T_x) + (2x - W_t) I(t > T_x) \tag{50}$$

is also a Wiener process (see Chapter III, § 3b, and also [124], [266], and [439]).

By (49) and (50) we obtain

$$\begin{aligned}
&\mathsf{P}\Bigl(\max_{s\leqslant t}(\mu s + W_s) \leqslant x\Bigr)\\
&\quad= \mathsf{P}(\mu t + W_t \leqslant x) - \mathsf{P}\Bigl(\max_{s\leqslant t}(\mu s + W_s) > x,\ \mu t + W_t \leqslant x\Bigr)\\
&\quad= \Phi\Bigl(\frac{x-\mu t}{\sqrt{t}}\Bigr) - \mathsf{E}\exp\Bigl(\mu W_t - \frac{\mu^2}{2}t\Bigr) I\Bigl(\max_{s\leqslant t} W_s > x,\ W_t \leqslant x\Bigr)\\
&\quad= \Phi\Bigl(\frac{x-\mu t}{\sqrt{t}}\Bigr) - \mathsf{E}\exp\Bigl(\mu \widetilde{W}_t - \frac{\mu^2}{2}t\Bigr) I\Bigl(\max_{s\leqslant t} \widetilde{W}_s > x,\ \widetilde{W}_t \leqslant x\Bigr)\\
&\quad= \Phi\Bigl(\frac{x-\mu t}{\sqrt{t}}\Bigr) - \mathsf{E}\exp\Bigl(\mu(2x - W_t) - \frac{\mu^2}{2}t\Bigr) I(W_t \geqslant x)\\
&\quad= \Phi\Bigl(\frac{x-\mu t}{\sqrt{t}}\Bigr) - e^{2\mu x}\mathsf{E}\exp\Bigl(\mu W_t - \frac{\mu^2}{2}t\Bigr) I(W_t \geqslant x)\\
&\quad= \Phi\Bigl(\frac{x-\mu t}{\sqrt{t}}\Bigr) - e^{2\mu x}\mathsf{P}(\mu t + W_t \geqslant x)\\
&\quad= \Phi\Bigl(\frac{x-\mu t}{\sqrt{t}}\Bigr) - e^{2\mu x}\Phi\Bigl(\frac{-x-\mu t}{\sqrt{t}}\Bigr).
\end{aligned}$$

The proof is complete.

COROLLARY 1. *If $\mu < 0$, then*

$$\mathsf{P}\Bigl(\sup_{t\geqslant 0}(\mu t + \sigma W_t) \leqslant x\Bigr) = 1 - \exp\Bigl\{\frac{2\mu x}{\sigma^2}\Bigr\}. \tag{51}$$

If $\mu \geqslant 0$, then

$$\mathsf{P}\Bigl(\sup_{t\geqslant 0}(\mu t + \sigma W_t) \leqslant x\Bigr) = 0. \tag{51$'$}$$

COROLLARY 2 (to the proof of (45)). *Setting $\mu = -\Bigl(\lambda + \dfrac{\sigma^2}{2}\Bigr)$ in* (50) *we obtain*

$$\mathsf{P}\Bigl(\sup_{t\geqslant 0}\Bigl[\sigma W_t - \Bigl(\lambda + \frac{\sigma^2}{2}\Bigr)t\Bigr] \leqslant x\Bigr) = 1 - \exp\Bigl\{-\Bigl(1 + \frac{2\lambda}{\sigma^2}\Bigr)x\Bigr\}. \tag{52}$$

Hence if $\lambda > 0$, then (45) *holds.*

COROLLARY 3 (to the proof of (17)). *Let* $S_t = xe^{rt} \cdot e^{\sigma W_t - \frac{\sigma^2}{2}t}$ *and assume that* $x^* > x$. *Then one sees from* (51) *with* $\mu = r - \dfrac{\sigma^2}{2} < 0$ *that*

$$\mathsf{P}(\tau^* = \infty) = \mathsf{P}\biggl(\sup_{t\geqslant 0}\Bigl[\sigma W_t + \Bigl(r - \frac{\sigma^2}{2}\Bigr)t\Bigr] \leqslant \ln\frac{x^*}{x}\biggr) = 1 - \Bigl(\frac{x}{x^*}\Bigr)^{1-\frac{2r}{\sigma^2}}, \tag{53}$$

which proves (17) *for* $r < \dfrac{\sigma^2}{2}$ *and* $x < x^*$. *This formula is obvious for* $x \geqslant x^*$, *while for* $x < x^*$ *and* $\mu = r - \dfrac{\sigma^2}{2} \geqslant 0$ *formula* (17) *is a consequence of* (51′).

All this completes the first proof of the theorem.

6. The second proof. Let $\beta = \lambda + r$, assume that $\lambda > 0$, let γ_1 be as in (31), and let $S_0 = 1$.

Setting

$$Z_t = e^{-\beta t} S_t^{\gamma_1}, \tag{54}$$

we obtain

$$Z_t = \exp\Bigl\{\gamma_1\sigma W_t - \frac{(\gamma_1\sigma)^2}{2}t\Bigr\}. \tag{55}$$

Hence $Z = (Z_t)$ is a P-*martingale* and

$$e^{-\beta t}(S_t - K)^+ = S_t^{-\gamma_1}(S_t - K)^+ Z_t.$$

Setting, in addition,

$$G(x) = x^{-\gamma_1}(x - K)^+,$$

we see that

$$\begin{aligned}\overline{V}^*(1) &= \sup_{\tau\in\overline{\mathfrak{M}}_0^\infty} \mathsf{E}e^{-(\lambda+r)\tau}(S_\tau - K)^+ I(\tau < \infty)\\ &= \sup_{\tau\in\overline{\mathfrak{M}}_0^\infty} \mathsf{E}G(S_\tau)Z_\tau I(\tau < \infty).\end{aligned} \tag{56}$$

The process $S = (S_t)_{t\geqslant 0}$ under consideration is generated by a Wiener process $W = (W_t)_{t\geqslant 0}$; without loss of generality we can assume that $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geqslant 0}, \mathsf{P})$ is a coordinate Wiener filtered space, i.e., $\Omega = C[0,\infty)$ is the space of continuous functions $\omega = (\omega(t))_{t\geqslant 0}$, $\mathscr{F}_t = \sigma(\omega\colon \omega(s), s \leqslant t)$, $\mathscr{F} = \bigvee \mathscr{F}_t$, and P is the Wiener measure.

Let $\widetilde{\mathsf{P}}$ be a measure in $(\Omega, \mathscr{F})$ such that the process $\widetilde{W} = (\widetilde{W}_t)_{t\geqslant 0}$ with $\widetilde{W}_t = W_t - (\gamma_1\sigma)t$ is a Wiener process with respect to $\widetilde{\mathsf{P}}$.

If $\mathsf{P}_t = \mathsf{P} \mid \mathscr{F}_t$ and $\widetilde{\mathsf{P}}_t = \widetilde{\mathsf{P}} \mid \mathscr{F}_t$ are the restrictions of P and $\widetilde{\mathsf{P}}$ to $\mathscr{F}_t$, then $\widetilde{\mathsf{P}}_t \sim \mathsf{P}_t$, and the Radon–Nikodym derivative is

$$\frac{d\widetilde{\mathsf{P}}_t}{d\mathsf{P}_t} = Z_t, \tag{57}$$

where Z_t is defined by (55). (See, for instance, Theorem 2 in Chapter III, § 3e.)

Hence, if $A \in \mathscr{F}_t$, then

$$\widetilde{\mathsf{E}} I_A = \mathsf{E} I_A Z_t,$$

where $\widetilde{\mathsf{E}}$ is averaging with respect to $\widetilde{\mathsf{P}}$, and if $A \in \mathscr{F}_\tau$, then

$$\widetilde{\mathsf{E}} I_A I_{(\tau<\infty)} = \mathsf{E} Z_\tau I_A I_{(\tau<\infty)} \tag{58}$$

(cf. formula (2) in Chapter V, § 3a).

Hence we obtain that if $f = f(\omega)$ is a nonnegative $\mathscr{F}_\tau$-measurable function, then

$$\widetilde{\mathsf{E}} f I_{(\tau<\infty)} = \mathsf{E} Z_\tau f I_{(\tau<\infty)}. \tag{59}$$

Taken together with (56), this shows that

$$\overline{V}^*(1) = \sup_{\tau \in \overline{\mathfrak{M}}_0^\infty} \widetilde{\mathsf{E}} G(S_\tau) I(\tau < \infty). \tag{60}$$

In other words, the optimal stopping problem (8) is equivalent (for $x = 1$) to another problem, (60), which can be easily solved by the following arguments. We consider the function $G(x) = x^{-\gamma_1}(x - K)^+$. This function attains its maximum on $E = (0, \infty)$ at the point $x^* = K\dfrac{\gamma_1}{\gamma_1 - 1}$ (cf. (15)), and

$$\max_{x \in E} G(x) = c^* \quad \big(= G(x^*)\big), \tag{61}$$

where c^* is defined by (14). Hence by (60) we obtain

$$\overline{V}^*(1) \leqslant c^* \sup_{\tau \in \overline{\mathfrak{M}}_0^\infty} \widetilde{\mathsf{E}} I(\tau < \infty) \leqslant c^*. \tag{62}$$

Let $\tau^* = \inf\{t \geqslant 0 \colon S_t \geqslant x^*\}$ and let the initial value S_0 be equal to $1 \leqslant x^*$. Since $\lambda > 0$ by assumption, it follows that $x^* < \infty$.

LEMMA 2. 1) *For $\lambda > 0$ we have*

$$\widetilde{\mathsf{P}}(\tau^* < \infty) = 1. \tag{63}$$

2) *If $\lambda > 0$ and, in addition, $r \geqslant \dfrac{\sigma^2}{2}$, then*

$$\mathsf{P}(\tau^* < \infty) = 1. \tag{64}$$

Proof. With respect to $\widetilde{\mathsf{P}}$, the process $\widetilde{W}_t = W_t - (\gamma_1\sigma)t$ $(t \geqslant 0)$ is Wiener, and by Girsanov's theorem

$$\begin{aligned}\widetilde{\mathsf{P}}(\tau^* < \infty) &= \widetilde{\mathsf{P}}\Big(\max_{t\geqslant 0} S_t \geqslant x^*\Big) = \widetilde{\mathsf{P}}\bigg(\max_{t\geqslant 0}\Big[\sigma W_t + \Big(r - \frac{\sigma^2}{2}\Big)t\Big] \geqslant \ln x^*\bigg)\\ &= \widetilde{\mathsf{P}}\bigg(\max_{t\geqslant 0}\Big[\sigma \widetilde{W}_t + \Big(\gamma_1\sigma^2 + r - \frac{\sigma^2}{2}\Big)t\Big] \geqslant \ln x^*\bigg)\\ &= \widetilde{\mathsf{P}}\bigg(\max_{t\geqslant 0}\Big[\sigma W_t + \Big(\gamma_1\sigma^2 + r - \frac{\sigma^2}{2}\Big)t\Big] \geqslant \ln x^*\bigg) = 1,\end{aligned}$$

where the last equality follows from (51) and the relation

$$\gamma_1\sigma^2 + r - \frac{\sigma^2}{2} = \sigma^2\sqrt{\Big(\frac{1}{2} - \frac{r}{\sigma^2}\Big)^2 + \frac{2(\lambda + r)}{\sigma^2}} > 0.$$

Thus, we have proved (63). Property (64) was established by Corollary 3. This completes the proof of Lemma 2.

We return to (62). Since $\widetilde{\mathsf{P}}(\tau^* < \infty) = 1$ and $G(S_{\tau^*}) = G(x^*) = c^*$, it follows that

$$\widetilde{\mathsf{E}}G(S_{\tau^*}) = c^*$$

and by (62) we obtain

$$\begin{aligned}\overline{V}^*(1) &= \widetilde{\mathsf{E}}G(S_{\tau^*}) = \widetilde{\mathsf{E}}G(S_{\tau^*})I(\tau^* < \infty)\\ &= \mathsf{E}e^{-(\lambda+r)\tau^*}(S_{\tau^*} - K)I(\tau^* < \infty) = c^*.\end{aligned}$$

This gives us the second proof of formula (12) (for $x = 1$) and shows that τ^* is an optimal stopping time. If $r \geqslant \dfrac{\sigma^2}{2}$, then $\mathsf{P}(\tau^* < \infty) = 1$, and therefore τ^* is in this case an optimal stopping time in the class $\mathfrak{M}_0^\infty$.

§ 2b. Standard Put Option

1. We can consider *put options* with pay-off functions $f_t = e^{-\lambda t}g(S_t)$, where $g(x) = (K - x)^+$, $x \in E = (0, \infty)$, in the same way as we have considered call options. For this reason we restrict ourselves to the statements of results and the main points of their proofs.

We shall consider a diffusion (B, S)-market described by the representations (1) and (2) in § 2a. Let

$$U_*(x) = \sup_{\tau\in\mathfrak{M}_0^\infty} \mathsf{E}_x e^{-(\lambda+r)\tau}(K - S_\tau)^+, \tag{1}$$

$$\overline{U}_*(x) = \sup_{\tau\in\overline{\mathfrak{M}}_0^\infty} \mathsf{E}_x e^{-(\lambda+r)\tau}(K - S_\tau)^+ I(\tau < \infty). \tag{2}$$

THEOREM. *Assume that* $\lambda \geqslant 0$. *Then*

$$U_*(x) = \overline{U}_*(x) = \begin{cases} c_* x^{\gamma_2}, & x > x_*, \\ K - x, & x \leqslant x_*, \end{cases} \tag{3}$$

where

$$\gamma_2 = \left(\frac{1}{2} - \frac{r}{\sigma^2}\right) - \sqrt{\left(\frac{1}{2} - \frac{r}{\sigma^2}\right) + \frac{2(\lambda + r)}{\sigma^2}}, \tag{4}$$

$$c_* = |\gamma_2|^{|\gamma_2|}\left(\frac{K}{1 + |\gamma_2|}\right)^{1+|\gamma_2|}, \tag{5}$$

$$x_* = K\frac{|\gamma_2|}{1 + |\gamma_2|}. \tag{6}$$

There exists an optimal stopping time in the class $\overline{\mathfrak{M}}_0^\infty$, *namely,*

$$\tau_* = \inf\{t \geqslant 0\colon S_t \leqslant x_*\}. \tag{7}$$

Moreover,

$$\mathsf{P}_x(\tau_* < \infty) = \begin{cases} 1 & \text{if } r \leqslant \dfrac{\sigma^2}{2} \text{ or } x \leqslant x_*, \\ \left(\dfrac{x_*}{x}\right)^{\frac{2r}{\sigma^2} - 1}, & \text{if } r > \dfrac{\sigma^2}{2} \text{ and } x > x_*. \end{cases} \tag{8}$$

This result is even slightly more simple than the theorem in § 2a: the function $g(x) = (K - x)^+$ is now *bounded.*

By analogy with the corresponding discrete-time problem (Chapter VI, § 5c) it would be natural to assume that the domains of continued observations C_* and the stopping domain D_* have the following form:

$$C_* = \{x \in E\colon x > x_*\} = \{x \in E\colon U_*(x) > g(x)\}$$

and

$$D_* = \{x \in E\colon x \leqslant x_*\} = \{x \in E\colon U_*(x) = g(x)\},$$

where x_* and $U_*(x)$ are equal to the solutions ($\widetilde{x}$ and $\widetilde{U}(x)$) of the *Stephan problem*

$$L\widetilde{U}(x) = (\lambda + r)\widetilde{U}(x), \qquad x > \widetilde{x}, \tag{9}$$

$$\widetilde{U}(x) = g(x), \qquad x \leqslant \widetilde{x}, \tag{10}$$

$$\left.\frac{d\widetilde{U}(x)}{dx}\right|_{x\downarrow\widetilde{x}} = \left.\frac{dg(x)}{dx}\right|_{x\uparrow\widetilde{x}}. \tag{11}$$

In our case the *bounded* solutions of (9) have the form $\widetilde{U}(x) = \widetilde{c}x^{\gamma_2}$ for $x > \widetilde{x}$, where γ_2 is the negative root of the square equation (30) in § 2a given by (4). Using (10) and (11), we can find the values of $\widetilde{c}$ and $\widetilde{x}$, which are expressed by the right-hand sides of (5) and (6).

We can prove the equality $U_*(x) = \widetilde{U}(x)$ and the optimality of τ_* by testing the conditions (A) and (B), using arguments similar to the ones presented in § 2a. The proof of (8) is based on Lemma 1 in § 2a and the corollaries to it.

Remark. The 'martingale' (i.e., the second) proof of the above theorem is based on the observation that in our case the process $Z = (Z_t)_{t\geqslant 0}$ with

$$Z_t = e^{-\beta t} S_t^{\gamma_2}, \qquad \beta = \lambda + r,$$

is a martingale and

$$Z_t = \exp\left\{\gamma_2 \sigma W_t - \frac{(\gamma_2\sigma)^2}{2} t\right\}.$$

Hence

$$e^{-\beta t}(K - S_t)^+ = S_t^{-\gamma_2}(K - S_t)^+ Z_t \leqslant c_* Z_t,$$

and we can complete as in the case of call options (§ 2a).

§ 2c. Combinations of Put and Call Options

1. In practice, as mentioned in Chapter VI, § 4e, alongside various kinds of options, one often encounters their *combinations*. One example here can be a *strangle option*, a combination of call and put options with different strike prices.

In this section we present a calculation for an American strangle option, making again the assumption that it can be exercised at *arbitrary* time on $[0, \infty)$ and the structure of the (B, S)-market is as described by relations (1)–(2) in § 2a.

In other words, we assume that

$$B_t = B_0 e^{rt} \tag{1}$$

and

$$S_t = S_0 \exp\left\{\sigma W_t + \left(\mu - \frac{\sigma^2}{2}\right)t\right\}, \tag{2}$$

where $W = (W_t)_{t\geqslant 0}$ is a standard Wiener process and $\mu = r$. The original measure P is martingale in this case.

For a discounted strangle option the pay-off function is as follows (cf. (3) in § 2a):

$$f_t = e^{-\lambda t} g(S_t), \qquad t \geqslant 0,$$

where

$$g(s) = \begin{cases} K_1 - s, & s \leqslant K_1, \\ 0, & K_1 < s < K_2, \\ s - K_2, & s \geqslant K_2. \end{cases} \tag{3}$$

In accordance with the general theory (Chapter VII, § 4 and Chapter VI, § 2), the price is

$$V^*(x) = \sup_{\tau \in \mathfrak{M}_0^\infty} B_0 \mathsf{E}_x \frac{f_\tau}{B_\tau} = \sup_{\tau \in \mathfrak{M}_0^\infty} \mathsf{E}_x e^{-(\lambda+r)\tau} g(S_\tau), \tag{4}$$

where $\mathfrak{M}_0^\infty = \{\tau = \tau(\omega) \colon 0 \leqslant \tau(\omega) < \infty,\ \omega \in \Omega\}$ is the class of finite stopping times and E_x is averaging under the assumption that $S_0 = x \in E = (0, \infty)$.

2. To determine the price $V^*(x)$ and the corresponding optimal stopping time we use the 'martingale' trick from [32], which we have already used in the preceding sections (e.g., in the 'second proof' in § 2a.6). We shall assume here that $S_0 = 1$ and the strike prices K_1 and K_2 satisfy the inequality $K_1 < 1 < K_2$.

Let

$$\gamma_1 = \Big(\frac{1}{2} - \frac{r}{\sigma^2}\Big) + \sqrt{\Big(\frac{1}{2} - \frac{r}{\sigma^2}\Big)^2 + \frac{2(\lambda + r)}{\sigma^2}}\,, \tag{5}$$

$$\gamma_2 = \Big(\frac{1}{2} - \frac{r}{\sigma^2}\Big) - \sqrt{\Big(\frac{1}{2} - \frac{r}{\sigma^2}\Big)^2 + \frac{2(\lambda + r)}{\sigma^2}} \tag{6}$$

be the roots of the square equation (30) in § 2a.

As shown in § § 2a, b, the processes $M_t^{(1)} = e^{-\beta t} S_t^{\gamma_1}$ and $M_t^{(2)} = e^{-\beta t} S_t^{\gamma_2}$, $t \geqslant 0$, with $\beta = \lambda + r$ are P-martingales. Hence the nonnegative process $M_t(p) = pM_t^{(1)} + (1-p)M_t^{(2)}$ is a P-martingale for each $0 \leqslant p \leqslant 1$, and

$$\begin{aligned} V^*(1) &= \sup_{\tau \in \mathfrak{M}_0^\infty} \mathsf{E}_1 e^{-(\lambda+r)\tau} g(S_\tau) \\ &= \sup_{\tau \in \mathfrak{M}_0^\infty} \mathsf{E} M_\tau(p) \frac{g(S_\tau)}{pS_\tau^{\gamma_1} + (1-p)S_\tau^{\gamma_2}}\,. \end{aligned} \tag{7}$$

Proceeding as in § 2a.6 we consider measures $\widetilde{\mathsf{P}}(p)$ such that

$$\frac{d\widetilde{\mathsf{P}}_t(p)}{d\mathsf{P}_t} = M_t(p). \tag{8}$$

Then we conclude from (7) that

$$V^*(1) = \sup_{\tau \in \mathfrak{M}_0^\infty} \mathsf{E}_{\widetilde{\mathsf{P}}(p)} \frac{g(S_\tau)}{pS_\tau^{\gamma_1} + (1-p)S_\tau^{\gamma_2}}\,, \tag{9}$$

where $\mathsf{E}_{\widetilde{\mathsf{P}}(p)}$ is averaging with respect to $\widetilde{\mathsf{P}}(p)$.

The next step is to choose a suitable value of p in $[0, 1]$ (which we denote by p^* in what follows) such that we can solve the corresponding optimal stopping problem (9).

As shown in [32], the following system of equations for (p, s_1, s_2):

$$\frac{s_2 - K_2}{ps_2^{\gamma_1} + (1-p)s_2^{\gamma_2}} = \frac{K_1 - s_1}{ps_1^{\gamma_1} + (1-p)s_1^{\gamma_2}}, \tag{10}$$

$$\frac{s_2}{s_2 - K_2} = \frac{p\gamma_1 s_2^{\gamma_1} + (1-p)\gamma_2 s_2^{\gamma_2}}{ps_2^{\gamma_1} + (1-p)s_2^{\gamma_2}}, \tag{11}$$

$$\frac{s_1}{s_1 - K_1} = \frac{p\gamma_1 s_1^{\gamma_1} + (1-p)\gamma_2 s_1^{\gamma_2}}{ps_1^{\gamma_1} + (1-p)s_1^{\gamma_2}}, \tag{12}$$

where $p \in [0, 1]$, $s_2 > K_2$, and $s_1 < K_1$, has a unique solution (p^*, s_1^*, s_2^*).

Let

$$c^* = \frac{s_2^* - K_2}{p^*(s_2^*)^{\gamma_1} + (1-p^*)(s_2^*)^{\gamma_2}} \qquad \left(= \frac{K_2 - s_1^*}{p^*(s_1^*)^{\gamma_1} + (1-p^*)(s_1^*)^{\gamma_2}} \right).$$

A simple analysis shows that

$$\sup_{s \geqslant 1} \frac{g(s)}{p^* s^{\gamma_1} + (1-p^*)s^{\gamma_2}} = \sup_{s \leqslant 1} \frac{g(s)}{p^* s^{\gamma_1} + (1-p^*)s^{\gamma_2}} \quad (= c^*).$$

In addition, the function $G(s) = \dfrac{g(s)}{p^* s^{\gamma_1} + (1-p^*)s^{\gamma_2}}$ takes its maximum at some points $s_1^* < K_1$ and $s_2^* > K_2$.

Hence, by (9) we obtain

$$V^*(1) \geqslant c^*. \tag{13}$$

We set

$$\tau^* = \inf\{t \colon S_t = s_1^* \text{ or } S_t = s_2^*\}.$$

By the properties of a Brownian motion with drift, $\mathsf{P}(\tau^* < \infty) = 1$. Hence $\mathsf{E}_{\widetilde{\mathsf{P}}(p^*)} G(S_{\tau^*}) = c^*$, and therefore

$$V^*(1) = c^*$$

and τ^* is the *optimal* stopping time.

Remark. If $K_1 = K_2$, then the strangle option becomes a straddle option (see Chapter VI, § 4e.2).

§ 2d. Russian Option

1. We consider a diffusion (B, S)-market with

$$dB_t = rB_t\,dt, \qquad B_0 > 0, \tag{1}$$

and

$$dS_t = S_t(r\,dt + \sigma\,dW_t), \qquad S_0 > 0, \tag{2}$$

or, equivalently,

$$B_t = B_0 e^{rt},$$
$$S_t = S_0 e^{rt} \cdot e^{\sigma W_t - \frac{\sigma^2}{2} t}.$$

Since

$$\frac{S_t}{B_t} = \frac{S_0}{B_0} e^{\sigma W_t - \frac{\sigma^2}{2} t}, \tag{3}$$

the process $\frac{S}{B} = \left(\frac{S_t}{B_t}\right)_{t \geqslant 0}$ is a martingale with respect to the original measure P.

Let

$$f_t = e^{-\lambda t} g_t(S), \qquad t \geqslant 0, \tag{4}$$

where

$$g_t(S) = \left(\max_{u \leqslant t} S_u - aS_t\right)^+, \qquad a \geqslant 0. \tag{5}$$

American options with pay-off functions (4) (which are called *Russian options* [434], [435]) belong to the class of put options *with aftereffect* and *discounting.* (Cf. Chapter VI, § 5d.)

Using the same notation (E_x, $\mathfrak{M}_0^\infty$, $\overline{\mathfrak{M}}_0^\infty, \dots$) as in § § 2a, b we set

$$U_*(x) = \sup_{\tau \in \mathfrak{M}_0^\infty} \mathsf{E}_x e^{-r\tau} f_\tau(S) \tag{6}$$

and

$$\overline{U}_*(x) = \sup_{\tau \in \overline{\mathfrak{M}}_0^\infty} \mathsf{E}_x e^{-r\tau} f_\tau(S) I(\tau < \infty). \tag{7}$$

By contrast to 'one-dimensional' optimal stopping problems for a Markov process $S = (S_t)_{t \geqslant 0}$ considered in § § 2a, b, the problems (6) and (7) are 'two-dimensional' in the following sense: the functionals $f_t(S)$ depend on a *two-dimensional* Markov process $\left(S_t, \max_{u \leqslant t} S_u\right)$.

It is remarkable, however, that using the methods of 'change of measure' one can reduce these 'two-dimensional' problems to 'one-dimensional' ones, which makes it possible to find explicit expressions for $U_*(x)$ $(= \overline{U}_*(x))$ and optimal stopping times.

2. We have already explained rather explicitly the idea of this reduction of 'two-dimensional' problems to 'one-dimensional' in Chapter VI, § 5d, for the discrete-time case.

Here we proceed as in § 2a.5 and assume that $(\Omega, \mathscr{F}, (\mathscr{F}_t), \mathsf{P})$ is the coordinate Wiener space.

Let $\widehat{\mathsf{P}}$ be a measure in $(\Omega, \mathscr{F})$ such that its restriction $\widehat{\mathsf{P}}_t$ is equivalent to P_t for each t and

$$\frac{d\widehat{\mathsf{P}}_t}{d\mathsf{P}_t} = Z_t, \tag{8}$$

where

$$Z_t = e^{\sigma W_t - \frac{\sigma^2}{2}t} \quad \left(= \frac{S_t/S_0}{B_t/B_0}\right), \qquad t \geqslant 0. \tag{9}$$

The process $\widehat{W} = (\widehat{W}_t)_{t\geqslant 0}$ with

$$\widehat{W}_t = W_t - \sigma t \tag{10}$$

is *Wiener* with respect to $\widehat{\mathsf{P}}$ by Girsanov's theorem, and for $\tau \in \overline{\mathfrak{M}}_0^\infty$ we have

$$\begin{aligned}
\mathsf{E}_x e^{-(\lambda+r)\tau} g_\tau(S) I(\tau < \infty) &= x\mathsf{E}_x e^{-\lambda\tau} \frac{S_\tau/S_0}{B_\tau/B_0} \frac{g_\tau(S)}{S_\tau} I(\tau < \infty) \\
&= x\mathsf{E}_x e^{-\lambda\tau} Z_\tau \frac{(\max_{u\leqslant\tau} S_u - aS_\tau)^+}{S_\tau} I(\tau < \infty) \\
&= x\widehat{\mathsf{E}} e^{-\lambda\tau} \left[\frac{\max_{u\leqslant\tau} S_u}{S_\tau} - a\right]^+ I(\tau < \infty).
\end{aligned} \tag{11}$$

We consider now the process $(\psi_t)_{t\geqslant 0}$ of the variables

$$\psi_t = \frac{\max(\max_{u\leqslant t} S_u, S_0\psi_0)}{S_t}, \tag{12}$$

where $\psi_0 \geqslant 1$.

Clearly, if $\psi_0 = 1$, then

$$\psi_t = \frac{\max_{u\leqslant t} S_u}{S_t}, \tag{13}$$

and therefore

$$\widehat{\mathsf{E}} e^{-\lambda\tau} \left[\frac{\max_{u\leqslant t} S_u}{S_\tau} - a\right]^+ I(\tau < \infty) = \widehat{\mathsf{E}} e^{-\lambda\tau} [\psi_\tau - a]^+ I(\tau < \infty). \tag{14}$$

Let $\widehat{\mathsf{P}}_\psi$ be the probability distribution of the process $(\psi_t)_{t\geqslant 0}$ under the assumption $\psi_0 = \psi \geqslant 1$; we consider the following optimal stopping problems:

$$\widehat{U}(\psi) = \sup_{\tau\in\overline{\mathfrak{M}}_0^\infty} \widehat{\mathsf{E}}_\psi e^{-\lambda\tau} [\psi_\tau - a]^+ \tag{15}$$

and

$$\widehat{\overline{U}}(\psi) = \sup_{\tau \in \overline{\mathfrak{M}}_0^\infty} \widehat{\mathsf{E}}_\psi e^{-\lambda\tau}[\psi_\tau - a]^+ \, I(\tau < \infty), \tag{16}$$

which can be regarded as pricing problems for discounted American *call option* with process of stock prices $(\psi_t)_{t\geqslant 0}$ and $B_t \equiv 1$. (We point out that our initial problem relates to *put options*!)

By (6), (7), and (15), (16) we obtain (in view of (11) and (14)) the equality

$$U_*(x) = x\widehat{U}(1), \qquad \overline{U}_*(x) = x\widehat{\overline{U}}(1). \tag{17}$$

3. Before we state the main results on the solution of the optimal stopping problems (15) and (16) we discuss some properties of the process $(\psi_t)_{t\geqslant 0}$.

LEMMA. 1) *The process $(\psi_t)_{t\geqslant 0}$ is a diffusion Markov process on the phase space $E = [1, \infty)$ with respect to the measure $\widehat{\mathsf{P}}$, with instantaneous reflection at the point $\{1\}$.*

2) *The process $(\psi_t)_{t\geqslant 0}$ has the stochastic differential*

$$d\psi_t = -\psi_t\big(r\,dt + \sigma\,d\widehat{W}_t\big) + d\varphi_t, \tag{18}$$

where $(\varphi_t)_{t\geqslant 0}$ is a nondecreasing process increasing on the set $\{(\omega, t)\colon \psi_t(\omega) = 1\}$, and $\widehat{W} = (\widehat{W}_t)_{t\geqslant 0}$ is a Wiener process (with respect to $\widehat{\mathsf{P}}$).

3) *If $q = q(\psi)$ is a function on $E = [1, \infty)$ such that $q \in C^2$ on $(1, \infty)$ and there exists $q'(1+) \equiv \lim_{\psi \downarrow 1} q'(\psi)$, then*

$$Lq(\psi) = -r\psi\,\frac{dq}{d\psi} + \frac{\sigma^2}{2}\,\psi^2\,\frac{d^2 q}{d\psi^2}, \qquad \psi > 1, \tag{19}$$

and

$$q'(1+) = 0. \tag{20}$$

Proof. By (12),

$$\begin{aligned} \psi_{t+\Delta} &= \max\left\{\frac{\max_{u\leqslant t+\Delta} S_u}{S_{t+\Delta}}, \frac{S_0\psi_0}{S_{t+\Delta}}\right\} \\ &= \max\left\{\frac{\max_{u\leqslant t} S_u}{S_t \cdot S_{t+\Delta}/S_t}, \frac{S_0\psi_0}{S_t \cdot S_{t+\Delta}/S_t}, \frac{\max_{t<u\leqslant t+\Delta} S_u/S_t}{S_{t+\Delta}/S_t}\right\} \\ &= \max\left\{\psi_t \cdot \frac{1}{S_{t+\Delta}/S_t}, \frac{\max_{t<u\leqslant t+\Delta} S_u/S_t}{S_{t+\Delta}/S_t}\right\}. \end{aligned} \tag{21}$$

Note that for $t < u \leqslant t + \Delta$ we have

$$\frac{S_u}{S_t} = \exp\left\{\sigma\big(\widehat{W}_u - \widehat{W}_t\big) + \left(r + \frac{\sigma^2}{2}\right)(u - t)\right\}.$$

Hence, taking into account that $\widehat{W}$ is a Wiener process with respect to $\widehat{\mathsf{P}}$, we obtain the following 'Markovian' property:

$$\operatorname{Law}\big(\psi_{t+\Delta} \mid \mathscr{F}_t, \widehat{\mathsf{P}}\big) = \operatorname{Law}\big(\psi_{t+\Delta} \mid \psi_t, \widehat{\mathsf{P}}\big).$$

To deduce (18) we set

$$N_t = \max\left\{\max_{u \leqslant t} S_u, S_0\psi_0\right\}. \tag{22}$$

Clearly, $N = (N_t)_{t \geqslant 0}$ is a nondecreasing process of bounded variation.

By (2) and (10),

$$dS_t = S_t\big[(r + \sigma^2)\,dt + \sigma\,d\widehat{W}_t\big] \tag{23}$$

and

$$d\left(\frac{1}{S_t}\right) = -\frac{1}{S_t}\big[r\,dt + \sigma\,d\widehat{W}_t\big]. \tag{24}$$

Hence, by Itô's formula

$$d\psi_t = N_t\,d\left(\frac{1}{S_t}\right) + \frac{1}{S_t}\,dN_t = -\psi_t\big[r\,dt + \sigma\,d\widehat{W}_t\big] + \frac{dN_t}{S_t}\,, \tag{25}$$

or, in the integral form,

$$\psi_t = \psi_0 - r\int_0^t \psi_u\,du - \sigma\int_0^t \psi_u\,d\widehat{W}_u + \int_0^t \frac{dN_u}{S_u}\,. \tag{26}$$

We now set

$$\varphi_t = \int_0^t \frac{dN_u}{S_u} \tag{27}$$

and note that $dN_u(\omega) = 0$ on the set $\{(\omega, u)\colon \psi_u(\omega) > 1\}$ $\Big($in the sense of the equality $\int_0^t I(\psi_u(\omega) > 1)\,dN_u(\omega) = 0$, $t > 0\Big)$. Hence

$$\varphi_t = \int_0^t I(\psi_u = 1)\,\frac{dN_u}{S_u}\,, \tag{28}$$

which shows even more clearly that the process $(\varphi_t)_{t \geqslant 0}$ changes the value only when the process $(\psi_t)_{t \geqslant 0}$ arrives at the boundary point $\{1\}$.

We claim that

$$\int_0^t I(\psi_u = 1)\, du = 0 \quad (\widehat{\mathsf{P}}\text{-a.s.}) \tag{29}$$

for each $t > 0$.

By Fubini's theorem

$$\widehat{\mathsf{E}} \int_0^\infty I(\psi_u = 1)\, du = \int_0^\infty \widehat{\mathsf{E}} I(\psi_u = 1)\, du = \int_0^\infty \widehat{\mathsf{P}}(\psi_u = 1)\, du = 0.$$

because $\widehat{\mathsf{P}}(\psi_u = 1) = 0$, which is a consequence of the fact that the distribution of the pair $\left(\max\limits_{s \leqslant u} \widehat{W}_s, \widehat{W}_u\right)$ has a density.

Thus, the process $(\psi_t)_{t \geqslant 0}$ stays zero time ($\widehat{\mathsf{P}}$-a.s.) at $\{1\}$, so that this point is an *instaneously reflecting* boundary ([239; Chapter IV, § 7]).

4. THEOREM. *Assume that $\lambda > 0$, $a \geqslant 0$, and $\psi \geqslant 1$. Then*

$$\widehat{U}(\psi) = \widehat{\widehat{U}}(\psi) = \begin{cases} (\widehat{\psi} - a) \cdot \dfrac{\gamma_2 \psi^{\gamma_1} - \gamma_1 \psi^{\gamma_2}}{\gamma_2 \widehat{\psi}^{\gamma_1} - \gamma_1 \widehat{\psi}^{\gamma_2}}, & \psi < \widehat{\psi}, \\ \psi - a, & \psi \geqslant \widehat{\psi}, \end{cases} \tag{30}$$

where

$$\gamma_k = \frac{A}{2} + (-1)^k \sqrt{\left(\frac{A}{2}\right)^2 + B}, \qquad k = 1, 2, \tag{31}$$

are the roots of the square equation

$$\gamma^2 - A\gamma - B = 0 \tag{32}$$

with

$$A = 1 + \frac{2r}{\sigma^2}, \qquad B = \frac{2\lambda}{\sigma^2};$$

the 'threshold' $\widehat{\psi}$ is the solution of the transcendental equation

$$\psi^{\gamma_1}\left(1 - \frac{1}{\gamma_1} - \frac{a}{\psi}\right) = \psi^{\gamma_2}\left(1 - \frac{1}{\gamma_2} - \frac{a}{\psi}\right) \tag{33}$$

in the domain $\psi > a$. If $a = 0$, then

$$\widehat{\psi} = \left| \frac{\gamma_2}{\gamma_1} \cdot \frac{\gamma_1 - 1}{\gamma_2 - 1} \right|^{\frac{1}{\gamma_2 - \gamma_1}}.$$

The stopping time

$$\widehat{\tau} = \inf\{t \geqslant 0 \colon \psi_t \geqslant \widehat{\psi}\}$$

has the property $\mathsf{P}_\psi(\widehat{\tau} < \infty) = 1$, $\psi \geqslant 1$, *and it is optimal both in the class* $\mathfrak{M}_0^\infty$ *and in* $\overline{\mathfrak{M}}_0^\infty$.

As in § § 2a, b we present two proofs, one ('Markovian') based on the solution of a Stephan problem and another based on 'martingale' considerations.

The first proof. The same arguments as in § § 2a, b, based on the representation (22) in § 2a, show that the domain of continued observations $\widehat{C}$ and the stopping domain $\widehat{D}$ in (15) must have the following form:

$$\widehat{C} = \{\psi \geqslant 1 \colon \psi < \widehat{\psi}\} = \{\psi \geqslant 1 \colon \widehat{U}(\psi) > g(\psi)\}$$

and

$$\widehat{D} = \{\psi \geqslant 1 \colon \psi \geqslant \widehat{\psi}\} = \{\psi \geqslant 1 \colon \widehat{U}(\psi) = g(\psi)\},$$

where $g(\psi) = (\psi - a)^+$.

As in § § 2a, b, $\widehat{U}(\psi)$ and the unknown threshold $\widehat{\psi}$ make up a solution of the *Stephan problem*

$$L\widehat{U}(\psi) = \lambda \widehat{U}(\psi), \quad 1 < \psi < \widehat{\psi}, \tag{34}$$

$$\widehat{U}'(1+) = 0, \tag{35}$$

$$\widehat{U}(\psi) = g(\psi), \quad \psi \geqslant \widehat{\psi}, \tag{36}$$

$$\left.\frac{d\widehat{U}(\psi)}{d\psi}\right|_{\psi \uparrow \widehat{\psi}} = \left.\frac{dg(\psi)}{d\psi}\right|_{\psi \downarrow \widehat{\psi}}, \tag{37}$$

where the operator L is defined by (19).

We seek a solution to (34) in the form $\widehat{U}(\psi) = \psi^\gamma$. Then we obtain square equation (32) for γ, which has the roots $\gamma_1 < 0$ and $\gamma_2 > 1$ described by (31).

Thus, equation (34) in the domain $\{\psi \colon 1 < \psi < \widehat{\psi}\}$ has the following general solution:

$$\widehat{U}(\psi) = c_1 \psi^{\gamma_1} + c_2 \psi^{\gamma_2}, \tag{38}$$

where c_1 and c_2 are some constants.

To find $\widehat{\psi}$ and the constants c_1 and c_2 we can use three additional conditions, (35), (36), and (37), which, in view of (38), take the following form:

$$c_1 \gamma_1 + c_2 \gamma_2 = 0, \tag{35$'$}$$

$$c_1 \widehat{\psi}^{\gamma_1} + c_2 \widehat{\psi}^{\gamma_2} = \widehat{\psi} - a, \tag{36$'$}$$

$$c_1 \gamma_1 \widehat{\psi}^{\gamma_1 - 1} + c_2 \gamma_2 \widehat{\psi}^{\gamma_2 - 1} = 1. \tag{37$'$}$$

By (36′) and (37′),

$$c_1 = \frac{a\gamma_2 + (1-\gamma_2)\widehat{\psi}}{(\gamma_1 - \gamma_2)\widehat{\psi}^{\gamma_1}} \quad \text{and} \quad c_2 = \frac{a\gamma_1 + (1-\gamma_1)\widehat{\psi}}{(\gamma_2 - \gamma_1)\widehat{\psi}^{\gamma_2}}. \tag{39}$$

By (35′),

$$c_1 = -\frac{\gamma_2}{\gamma_1} c_2, \tag{40}$$

which yields equation (33) for $\widehat{\psi}$. If $a = 0$, then it follows from this equation that

$$\widehat{\psi} = \left| \frac{\gamma_2}{\gamma_1} \cdot \frac{\gamma_1 - 1}{\gamma_2 - 1} \right|^{\frac{1}{\gamma_2 - \gamma_1}}. \tag{41}$$

Finally, from (38) and (39), in view of (33), we see that

$$\widehat{U}(\psi) = (\widehat{\psi} - a) \cdot \frac{\gamma_2 \psi^{\gamma_1} - \gamma_1 \psi^{\gamma_2}}{\gamma_2 \widehat{\psi}^{\gamma_1} - \gamma_1 \widehat{\psi}^{\gamma_2}}$$

in the domain $\widehat{C} = \{\psi : \psi < \widehat{\psi}\}$.

We now claim that $\widehat{\mathsf{P}}_\psi(\widehat{\tau} < \infty) = 1$ for $\psi \geqslant 1$. To this end it suffices to show that

$$\widehat{\mathsf{P}}\left(\sup_{t \geqslant 0} \left(\frac{\sup_{u \leqslant t} S_u}{S_t} \right) \geqslant \widehat{\psi} \right) = 1 \tag{42}$$

for each $\widehat{\psi} > 1$. (Property (42) is obvious for $\widehat{\psi} = 1$.)

We have

$$\frac{\sup_{u \leqslant t} S_u}{S_t} = \exp Y_t,$$

where

$$Y_t = \sup_{u \leqslant t} \left[\sigma(\widehat{W}_u - \widehat{W}_t) + \left(r + \frac{\sigma^2}{2} \right)(u - t) \right].$$

We consider the sequence of stopping times $(\sigma_k)_{k \geqslant 0}$ such that

$$\begin{aligned}
\sigma_0 &= 0, \\
\sigma_1 &= \inf\{t \geqslant 1 : Y_t = 0\}, \\
&\dots \\
\sigma_{k+1} &= \inf\{t \geqslant \sigma_k + 1 : Y_t = 0\}, \dots .
\end{aligned}$$

Then for $\widehat{y} = \ln \widehat{\psi} \geqslant 0$ we have

$$\left\{ \omega : \sup_{t \leqslant \infty} Y_t(\omega) \geqslant \widehat{y} \right\} = \bigcup_{k \geqslant 0} \left\{ \omega : \sup_{\sigma_k \leqslant t < \sigma_{k+1}} Y_t(\omega) \geqslant \widehat{y} \right\}.$$

The events $\left\{\omega\colon \sup_{\sigma_k \leqslant t < \sigma_{k+1}} Y_t(\omega) \geqslant \widehat{y}\right\}$ are independent for distinct k and their probabilities are equal and positive. Hence $\widehat{\mathsf{P}}\left\{\omega\colon \sup_{t<\infty} Y_t(\omega) \geqslant \widehat{y}\right\} = 1$ by the Borel–Cantelli lemma, so that $\widehat{\mathsf{P}}_{\psi}(\widehat{\tau} < \infty) = 1$.

It remains to show that $\widehat{\tau} = \inf\{t \geqslant 0\colon \psi_t \geqslant \widehat{\psi}\}$ is the optimal time in the problems (15) and (16).

One way to prove this is to test properties (A) and (B) in § 2a, which can be done in the same way as for call and put options. (See [444] for the detail.) We now present another proof, based on 'martingale' considerations (cf. § 2a.5 and [32]).

The second proof. We assume for simplicity that $a = 0$. (As regards the general case of $a \geqslant 0$, see [32].)

We set

$$M_t = e^{-\lambda t}\psi_t h(\psi_t) \tag{43}$$

and find a function $h = h(\psi)$, $\psi \geqslant 1$, such that $M = (M_t)_{t\geqslant 0}$ is a local martingale with respect to the measure $\widetilde{\mathsf{P}}$.

Using Itô's formula for $e^{-\lambda t}\psi_t h(\psi_t)$ we see that

$$d\bigl(e^{-\lambda t}\psi_t h(\psi_t)\bigr) = e^{-\lambda t}\psi_t\bigl[A_t\,dt + B_t\bigl(-\sigma\,d\widehat{W}_t + d\varphi_t\bigr)\bigr], \tag{44}$$

where

$$A_t = -(\lambda + r)h(\psi_t) + (\sigma^2 - r)\psi_t h'(\psi_t) + \frac{1}{2}\sigma^2\psi_t^2 h''(\psi_t), \tag{45}$$

$$B_t = h'(\psi_t) + h(\psi_t). \tag{46}$$

By (44), to make $M = (M_t)_{t\geqslant 0}$ a local martingale we can take $h = h(\psi)$, $\psi \geqslant 1$, equal to the solution of the problem

$$\frac{1}{2}\sigma^2\psi^2 h''(\psi) + (\sigma^2 - r)\psi h'(\psi) - (\lambda + r)h(\psi) = 0, \qquad \psi > 1, \tag{47}$$

with boundary condition

$$h'(1+) + h(1+) = 0. \tag{48}$$

We rewrite (47) as

$$\psi^2 h''(\psi) + 2\left(1 - \frac{r}{\lambda^2}\right)\psi h'(\psi) - 2\left(\frac{\lambda + r}{\sigma^2}\right) = 0 \tag{49}$$

and seek its solution in the form $h(\psi) = \psi^x$. Then we obtain the following square equation for x:

$$x^2 + x(1 - 2r) - 2(\lambda + r) = 0. \tag{50}$$

Comparing it with (32):

$$\gamma^2 - \gamma(1+2r) - 2\lambda = 0, \tag{51}$$

we see that (51) turns into (50) after the change $\gamma = x+1$, so that the roots x_i and γ_i of these equations are related by the formulas $\gamma_i = x_i + 1$, $i = 1, 2$.

The general solution of (49) (for $\sigma^2 = 1$) is as follows:

$$h(\psi) = d_1 \psi^{x_1} + d_2 \psi^{x_2}, \qquad \psi > 1.$$

We consider $h = h(\psi)$ such that (48) holds. Then

$$d_1 = \frac{1+x_2}{x_2 - x_1} = \frac{\gamma_2}{\gamma_2 - \gamma_1},$$
$$d_2 = \frac{1+x_1}{x_1 - x_2} = \frac{\gamma_1}{\gamma_1 - \gamma_2}.$$

Consequently,

$$h(\psi) = \frac{1}{\gamma_2 - \gamma_1}\left[\gamma_2 \psi^{\gamma_1 - 1} - \gamma_1 \psi^{\gamma_2 - 1}\right]. \tag{52}$$

We have $\gamma_1 < 0$, $\gamma_2 > 1$, and $h'(\psi) = 0$ for $\psi = \widetilde{\psi}$, where $\widetilde{\psi}$ can be found from the equation

$$\widetilde{\psi}^{\gamma_2 - \gamma_1} \frac{\gamma_1(\gamma_2 - 1)}{\gamma_2(\gamma_1 - 1)} = 1. \tag{53}$$

Comparing (53) and (41) we observe that the quantity $\widetilde{\psi}$ is the same as $\widehat{\psi}$ in (41). Moreover, the function $h(\psi)$ takes its *minimum* at the point $\widehat{\psi}$.

We see from (43)–(48) that for the function $h = h(\psi)$ so chosen the process

$$M_t = e^{-\lambda t} \psi_t h(\psi_t), \qquad t \geqslant 0,$$

is a nonnegative local martingale, and therefore a supermartingale. Hence for each $\tau \in \mathfrak{M}_0^\infty$ and $\psi_0 = 1$ we have

$$\begin{aligned}
\widehat{\mathsf{E}}_1 e^{-\lambda\tau} \psi_\tau &= \widehat{\mathsf{E}}_1 h^{-1}(\psi_\tau) M_\tau \leqslant \widehat{\mathsf{E}}_1 h^{-1}(\widehat{\psi}) M_\tau \\
&= h^{-1}(\widehat{\psi}) \widehat{\mathsf{E}}_1 M_\tau \leqslant h^{-1}(\widehat{\psi}) \widehat{\mathsf{E}}_1 M_0 = h^{-1}(\widehat{\psi}) \\
&= \frac{\gamma_2 - \gamma_1}{\gamma_2 \widehat{\psi}^{\gamma_1 - 1} - \gamma_1 \widehat{\psi}^{\gamma_2 - 1}} = \widehat{\psi} \cdot \frac{\gamma_2 - \gamma_1}{\gamma_2 \widehat{\psi}^{\gamma_1} - \gamma_1 \widehat{\psi}^{\gamma_2}}.
\end{aligned}$$

If $\psi_0 = 1$, then $\widehat{\tau} = \inf\{t \geqslant 0 \colon \psi_t \geqslant \widehat{\psi}\}$ is finite with probability one as shown above ($\widehat{\mathsf{P}}_1(\widehat{\tau} < \infty) = 1$), and

$$\widehat{\mathsf{E}}_1 e^{-\lambda\widehat{\tau}} \psi_{\widehat{\tau}} = \widehat{\mathsf{E}}_1 h^{-1}(\psi_{\widehat{\tau}}) M_{\widehat{\tau}} = h^{-1}(\widehat{\psi}) \quad \left(= \widehat{U}(1)\right)$$

for this stopping time, which means precisely the optimality of $\widehat{\tau}$ in the class $\mathfrak{M}_0^\infty$ for $\psi_0 = 1$. (Similar arguments hold for each $\psi_0 \leqslant \widehat{\psi}$.)

We proceed now to our original problems (6) and (7). By (11), setting $a = 0$ we obtain

$$\mathsf{E}_x e^{-(\lambda+r)\tau} g_\tau(S) I(\tau < \infty) = x \widehat{\mathsf{E}} e^{-\lambda\tau} \psi_\tau I(\tau < \infty). \tag{54}$$

Here $\psi_0 = 1$ and, as shown above, $\widehat{\tau} = \inf\{t\colon \psi_t \geqslant \widehat{\psi}\}$ is the optimal stopping time in the following sense:

$$\sup_{\tau \in \overline{\mathfrak{M}}_0^\infty} \widehat{\mathsf{E}} e^{-\lambda\tau} \psi_\tau I(\tau < \infty) = \widehat{\mathsf{E}} e^{-\lambda\widehat{\tau}} \psi_{\widehat{\tau}} I(\widehat{\tau} < \infty) = \widehat{\mathsf{E}} e^{-\lambda\widehat{\tau}} \psi_{\widehat{\tau}} \tag{55}$$

and

$$\sup_{\tau \in \mathfrak{M}_0^\infty} \widehat{\mathsf{E}} e^{-\lambda\tau} \psi_\tau I(\tau < \infty) = \widehat{\mathsf{E}} e^{-\lambda\widehat{\tau}} \psi_{\widehat{\tau}}. \tag{56}$$

Hence $\widehat{\tau}$ is the optimal stopping time in the problem (7).

Arguments used above, in the proof of the property $\widehat{\mathsf{P}}_\psi(\widehat{\tau} < \infty) = 1$, are suitable also for the analysis of the process $(\psi_t)_{t \geqslant 0}$, where they show that $\widehat{\tau}$ is finite almost everywhere also with respect to P. Hence $\widehat{\tau}$ is an optimal stopping time in our original problems (6) and (7).

3. American Options in Diffusion (B, S)-Stockmarkets. Finite Time Horizons

§ 3a. Special Features of Calculations on Finite Time Intervals

1. If the time horizon is infinite, i.e., the exercise times take values in the set $[0, \infty)$, then one often manages to describe completely the price structure of American options and the corresponding domains of continued and stopped observations. For instance, in all the cases considered in § 2 we found both the price $V^*(x)$ and the boundary point x^* in the phase space $E = \{x : x > 0\}$ between the domain of continued observations and the stopping domain.

We note that this is feasible because a geometric Brownian motion $S = (S_t)_{t \geqslant 0}$ is a *homogeneous* Markov process and there are no *constraints* on exercise times, so that the resulting problem has

elliptic type.

The situation becomes much more complicated if the time parameter t ranges over a *bounded* interval $[0, T]$.

The corresponding optimal stopping problem is *inhomogeneous* in that case and we must consider problems that have

parabolic type

from the analytic standpoint. As a result, in place of a boundary point x^* we encounter in the corresponding problems an *interface* function $x^* = x^*(t)$, $0 \leqslant t \leqslant T$, separating the domain of continued observations and the stopping domain in the phase space $[0, T) \times E = \{(t, x) : 0 \leqslant t < T, \ x > 0\}$. (Cf. Figs. 57 and 59 in Chapter VI, § § 5b, c.)

We also point out that, although the theory of optimal stopping rules in the continuous-time case (see, for instance, the monograph [441]) proposes *general methods* of the search of optimal stopping times, we do not know of many *concrete* problems (e.g., concerning options) where one can find *precise* analytic formulas describing the boundary functions $x^* = x^*(t)$, $0 \leqslant t < T$, the prices, and so on.

In practice—e.g., in calculations for American options in some market—one usually resorts to quantization (with respect to the time and/or the phase variables) and finds approximate values of, say, the boundary function and prices by backward induction (see Chapter VI, § 2a).

Of course, this does not eliminate interest to exact (or 'almost' exact) solutions; in this direction we must first of all discuss several relevant questions of the *theory* of optimal stopping problems on finite time intervals and, in particular, one very common method based on the reduction of such problems to *Stephan problems* (often called also *free-boundary* problems) for partial differential equations.

2. For definiteness, we consider a (B, S)-market described by relations (1) and (2) in § 2a, where we assume that the time parameter t belongs to $[0, T]$, $\mu = r$, and the pay-off functions have the form $f_t = e^{-\lambda t} g(S_t)$, where $\lambda \geqslant 0$, the Borel function $g(x)$ is nonnegative, and $x \in E = (0, \infty)$.

Let

$$V(T, x) = B_0 \mathsf{E}_x \frac{f_T}{B_T} \tag{1}$$

and

$$V^*(T, x) = B_0 \sup_{\tau \in \mathfrak{M}_0^T} \mathsf{E}_x \frac{f_\tau}{B_\tau} \tag{2}$$

be the *rational* prices of European and American options, respectively. In relations (1) and (2) the symbol E_x means averaging with respect to the original measure (which is martingale since $\mu = r$) under the assumption $S_0 = x$.

Remark. We proved formula (1) for $V(T, x)$ in Chapter VII, § 4b. The proof of (2) is based on the optional decomposition and its idea is the same as for discrete time (see Chapter VI, §§ 2c, 5a). See, for instance, [281] for the detail specific to continuous time.

3. For $t \geqslant 0$ and $x \in E = (0, \infty)$ we set

$$V(t, x) = \mathsf{E}_x e^{-\beta t} g(S_t) \tag{3}$$

and

$$V^*(t, x) = \sup_{\tau \in \mathfrak{M}_0^t} \mathsf{E}_x e^{-\beta \tau} g(S_\tau), \tag{4}$$

where $\beta = \lambda + r$ and $x = S_0$.

In the discussion of the case where $t \in [0, T]$ it is also useful to introduce the functions

$$Y(t, x) = V(T - t, x) \tag{5}$$

and

$$Y^*(t, x) = V^*(T - t, x), \tag{6}$$

where $T - t$ is the remaining time.

Clearly,

$$Y(t,x) = \mathsf{E}_{t,x} e^{-\beta(T-t)} g(S_T) \tag{7}$$

and

$$Y^*(t,x) = \sup_{\tau \in \mathfrak{M}_t^T} \mathsf{E}_{t,x} e^{-\beta(\tau-t)} g(S_\tau), \tag{8}$$

where $\mathsf{E}_{t,x}$ is averaging with respect to the original (martingale) measure under the assumption $S_t = x$, and $\mathfrak{M}_t^T$ is the class of optimal stopping times $\tau = \tau(\omega)$ such that $t \leqslant \tau \leqslant T$.

We considered the functions $V = V(t,x)$ (in the case of a Brownian motion and for $\beta = 0$) in Chapter III, § 3f, in connection with the probabilistic representation of the solution of a Cauchy problem. The same arguments (see Chapter III, § 3f.5 for greater detail) show that the function $V = V(t,x)$ (provided that it belongs to the class $C^{1,2}$) satisfies for $t > 0$ and $x \in E$ the equation

$$\frac{\partial V}{\partial t} + \beta V = LV, \tag{9}$$

where

$$LV(t,x) = rx\frac{\partial V}{\partial x} + \frac{1}{2}\sigma^2 x^2 \frac{\partial^2 V}{\partial x^2}, \tag{10}$$

with initial condition

$$V(0,x) = g(x). \tag{11}$$

By (5), (9), and (11) we obtain that the function $Y = Y(t,x)$ satisfies for $t < T$ the equation

$$-\frac{\partial Y}{\partial t} + \beta Y = LY \tag{12}$$

and the boundary condition

$$Y(T,x) = g(x). \tag{13}$$

Recall that we met already *fundamental equation* (12) (which is in fact a Feynman–Kac equation—see Chapter III, § 3f) in § 1c, in our discussion of the methods used by F. Black, M. Scholes [44], and R. Merton [346] to calculate the rational price ($V(T,x) = Y(0,x)$) of a standard European call option with $g(x) = (x-K)^+$ and $\lambda = 0$.

4. We proceed now to the question of the rational price $V^*(T,x) = Y^*(0,x)$.

We set

$$\tau_0^T = \inf\{0 \leqslant t \leqslant T : Y^*(t,S_t) = g(S_t)\} \tag{14}$$

and

$$D_t^T = \{x \in E : Y^*(t,x) = g(x)\}, \tag{15}$$

$$C_t^T = \{x \in E : Y^*(t,x) > g(x)\}. \tag{16}$$

For $s \leqslant t$ we have $Y^*(s,x) = V^*(T-s,x) \geqslant V^*(T-t,x) = Y^*(t,x)$. Hence for $0 \leqslant s \leqslant t < T$ we obtain

$$D_0^T \subseteq D_s^T \subseteq D_t^T$$

and

$$C_0^T \supseteq C_s^T \supseteq C_t^T.$$

For $t = T$ we clearly have $D_T^T = E$ and $C_T^T = \varnothing$.

The domains

$$D^T = \{(t,x)\colon t \in [0,T),\ x \in D_t^T\}$$

and

$$C^T = \{(t,x)\colon t \in [0,T),\ x \in C_t^T\}$$

in the phase space $[0,T) \times E$ are called the *stopping* domain and the domains of *continued observations*, respectively. This is related to the fact that in 'typical' optimal stopping problems for Markov processes the stopping time τ_0^T is optimal (see, e.g., Theorem 6(3) in [441], Chapter III, § 4):

$$\mathsf{E}_x e^{-\beta\tau_0^T} g(S_{\tau_0^T}) = V^*(T,x). \tag{17}$$

Since

$$\tau_0^T = \inf\{0 \leqslant t \leqslant T\colon S_t \in D_t^T\}, \tag{18}$$

it is clear why we call $D^T = \bigcup_{t<T} (\{t\} \times D_t^T)$ the *stopping domain*: if $(t, S_t) \in D^T$, then the observations terminate. (It seems reasonable to include in this domain also the 'terminal' set $T \times D_T^T = T \times E$.)

The domains D^T and C^T can have very complicated structure depending on the properties of the functions $g = g(x)$ and, of course, of the process $S = (S_t)_{t \leqslant T}$. For instance, as subsets of $[0,T) \times E$ these domains can be *multiply connected* or consist of several stopping 'islands', and so on.

On the other hand, for standard call and put options with $g(x) = (x-K)^+$ and $g(x) = (K-x)^+$, respectively, the domains D^T and C^T are simply connected (see § 3c below).

For these options the *boundary* ∂D^T of the *stopping domain* can be represented as follows:

$$\partial D^T = \{(t,x)\colon t \in [0,T),\ x = x^*(t)\},$$

where we have

$$x^*(t) = \inf\{x \in E\colon Y^*(t,x) = (x-K)^+\},$$

for a call option and

$$x^*(t) = \sup\{x \in E\colon Y^*(t,x) = (K-x)^+\}$$

for a put option.

§ 3b. Optimal Stopping Problems and Stephan Problems

1. It follows from our discussion that for a description of the structure of the optimal stopping time τ_0^T and the domains of continued and stopped observations one must find the function $V^* = V^*(t,x)$ or, equivalently, the function $Y^*(t,x) = V^*(T-t,x)$.

There exist various characterizations of these function in the general theory of optimal stopping rules.

For instance, it is known (see [441]) that the function $Y^* = Y^*(t,x)$ is the *smallest β-excessive majorant* of the (nonnegative Borel) function $g = g(x)$. In other words, the function $Y^* = Y^*(t,x)$ is the *smallest* among all functions $F = F(t,x)$ such that

$$e^{-\beta\Delta} T_\Delta F(t,x) \leqslant F(t,x), \qquad x \in E, \tag{1}$$

for $0 \leqslant t \leqslant t+\Delta \leqslant T$, where $T_\Delta F(t,x) = \mathsf{E}_{t,x} F(t+\Delta, S_{t+\Delta})$, $x = S_t$, and

$$g(x) \leqslant F(t,x), \qquad x \in E, \quad 0 \leqslant t \leqslant T. \tag{2}$$

It follows, in particular, that

$$\max\{g(x),\, e^{-\beta\Delta} T_\Delta Y^*(t,x)\} \leqslant Y^*(t,x). \tag{3}$$

For small $\Delta > 0$ and $t = 0, \Delta, \dots, [T/\Delta]\Delta$ one would expect the function $Y^*(t,x)$ to be 'close' to

$$Y_\Delta^*(t,x) = \sup_{\tau \in \mathfrak{M}_t^T(\Delta)} \mathsf{E}_{t,x} e^{-\beta(\tau - t)} g(S_\tau), \tag{4}$$

where $\mathfrak{M}_t^T(\Delta)$ are stopping times τ such that $\tau = k\Delta$, $k = 0, 1, \dots, [T/\Delta]$, $t \leqslant \tau \leqslant T$, and $\{\omega\colon \tau \leqslant k\Delta\} \in \mathscr{F}_{k\Delta}(\Delta)$, $\mathscr{F}_{k\Delta}(\Delta) = \sigma\{\omega\colon S_\Delta, S_{2\Delta}, \dots, S_{k\Delta}\}$.

As shows the theory of optimal stopping rules, this conjecture on the 'closeness' of these functions for small $\Delta > 0$ has a rigorous formulation (see [441; Chapter III, § 2]). Further, for $Y_\Delta^*(t,x)$ with $t = 0, \Delta, \dots, [T/\Delta]\Delta$ we have the recursive relations

$$Y_\Delta^*(t,x) = \max\{g(x),\, e^{-\beta\Delta} \mathsf{E}_{t,x} Y_\Delta^*(t+\Delta, S_{t+\Delta})\} \tag{5}$$

(see the discrete-time case in Chapter VI, § 2a and [441], Chapter II, § 4), therefore, assuming that $Y^*(t,x)$ is sufficiently smooth and using Taylor's formula we see from (5) that

$$Y^*(t,x) = \max\left\{g(x),\, (1-\beta\Delta)\left[Y^*(t,x) + \left(\frac{\partial Y^*(t,x)}{\partial t} + LY^*(t,x)\right)\Delta\right] + o(\Delta)\right\}, \tag{6}$$

where

$$LY^*(t,x) = rx \frac{\partial Y^*(t,x)}{\partial x} + \frac{1}{2}\sigma^2 x^2 \frac{\partial^2 Y^*(t,x)}{\partial x^2}. \tag{7}$$

By (6) we obtain that in the domain where $Y^*(t,x) > g(x)$ (i.e., in the domain of continued observations) the function $Y^* = Y^*(t,x)$ satisfies the equation

$$-\frac{\partial Y^*}{\partial t} + \beta Y^* = LY^*. \tag{8}$$

Remark 1. Equation (8) is similar in form to (12) in § 3a; this is not that surprising for the following reasons.

Assume that there exists optimal stopping time τ_t^T in the class $\mathfrak{M}_t^T$. Then

$$Y^*(t,x) = \mathsf{E}_{t,x} e^{-\beta(\tau_t^T - t)} g(S_{\tau_t^T}).$$

Since

$$Y(t,x) = \mathsf{E}_{t,x} e^{-\beta(T-t)} g(S_T),$$

it is intuitively clear that if $(t,x) \in C_t^T$, then the (backward) equations with respect to t and x for the functions $Y^*(t,x)$ and $Y(t,x)$ must be the same since their coefficients are determined by the local characteristics of *the same* two-dimensional process $(u, S_u)_{t \leqslant u \leqslant T}$ in a neighborhood of the *initial* point $(t, x = S_t)$.

Remark 2. There exists an extensive literature devoted to the derivation of equation (8) for $Y^*(t,x)$ in the domain of continued observations; e.g., the monographs [266], [287], [441], [478] and the papers [33], [66], [134], [135], [179], [247], [265], [272], [340], [363], [467].

2. We considered already the connection between optimal stopping problems and Stephan problems in Chapter IV, § 5, in our discussion of American options in a binomial (B,S)-market. In the continuous-time case this connection has apparently been revealed for the first time in ***statistical sequential analysis***, in testing problems for statistical hypotheses about the drift of a Wiener process ([67], [300], [349], [440]; see also the historico-bibliographical passages in [116] and [441]).

One of the first papers in the finances literature considering Stephan (free-boundary) problems was that of H. McKean [340], devoted to the rational price of American warrants.

3. In mathematical physics, the *Stephan problems* arise in the study of phase transitions ([413], [463]). One simple example of such a *two-phase Stephan problem* is, for instance, as follows.

Assume that our 'time-state' space $\mathbb{R}_+ \times E = \{(t,x)\colon t \geqslant 0,\ x > 0\}$ is partitioned between two phases,

$$C^{(1)} = \{(t,x)\colon t \geqslant 0,\ 0 < x < x(t)\}$$

and

$$C^{(2)} = \{(t,x)\colon t \geqslant 0,\ x(t) \leqslant x < \infty\},$$

where $x = x(t)$, $t \geqslant 0$, is the interface (e.g., between ice and liquid in freezing water). For each phase $C^{(i)}$ the temperature $u(t,x)$ at x at time t, $i = 1,2$, is assumed to satisfy 'its own' heat equation

$$c_i \rho_i \frac{\partial u}{\partial t} = k_i \frac{\partial^2 u}{\partial x^2}, \qquad i = 1,2, \tag{9}$$

where (using physical terminology) the c_i are the specific heats, the ρ_i are the densities of the phases, and the k_i are the thermal conductivities (see, e.g., [335; vol. 5, p. 324]).

Equations (9) are considered for

the *boundary* condition $u(0,t) = \text{Const}$,

the *initial* condition $u(x,0) = \text{Const}$

and, e.g., for the following

condition at the *interface* surface:

$$u\big(t, x(t-)\big) = u\big(t, x(t+)\big), \tag{10}$$

$$\frac{\partial u}{\partial x}\big(t, x(t-)\big) = \frac{\partial u}{\partial x}\big(t, x(t+)\big), \tag{11}$$

for $t > 0$, with additional assumption $x(0) = 0$.

Under these assumptions the *Stephan problem* consists in finding the interface $x = x(t)$, $t \geqslant 0$, and the function $u = u(t,x)$ describing the temperature schedule in the phases.

4. We have presented an example of a Stephan problem borrowed from *mathematical physics* to emphasize its common and distinct traits in comparison with the Stephan problems arising in the search of optimal stopping rules and, in particular, in connection with American options.

As already mentioned, for standard put and call options we also encounter a *two-phase* situation: looking for optimal stopping rules we can restrict ourselves to the consideration of only *two* simply connected phases, the domain C^T of continued observations, where $Y^*(t,x)$ satisfies equation (8), and the domain D^T, where $Y^* = Y^*(t,x)$ coincides with the function $g = g(x)$.

In the next section we formulate precisely the corresponding Stephan problems for these two options and describe the qualitative features of the corresponding solutions $Y^* = Y^*(t,x)$ and $x^* = x^*(t)$.

§ 3c. Stephan Problem for Standard Call and Put Options

1. A call option. We assume that a (B,S)-market can be described by relations (1) and (2) in § 2a with $\mu = r$, $0 \leqslant t \leqslant T$, and the pay-off function $f_t = e^{-\lambda t} g(S_t)$, where $\lambda \geqslant 0$ and $g(x) = (x-K)^+$, $x \in E = (0,\infty)$. The main results about this option are as follows.

1) The *rational price* $V^*(T,x)$, $x = S_0$, of such a (discounted) option is defined, as mentioned in § 3a, by the formula

$$V^*(T,x) = \sup_{\tau \in \mathfrak{M}_0^T} \mathsf{E}_x e^{-\beta\tau} g(S_\tau), \tag{1}$$

where $\beta = \lambda + r$ and E_x is averaging with respect to the original (martingale) measure under the assumption $S_0 = x$.

2) For $t \in [0,T]$ and $x \in E$ let

$$Y^*(t,x) = \sup_{\tau \in \mathfrak{M}_t^T} \mathsf{E}_{t,x} e^{-\beta(\tau - t)} g(S_\tau), \tag{2}$$

where $\mathsf{E}_{t,x}$ is averaging with respect to the (martingale) measure under the assumption $x = S_t$.

The function $Y^* = Y^*(t,x)$ is the *smallest β-excessive majorant of the function* $g(x)$ (see § 3b.1).

3) The rational price is

$$V^*(T,x) = Y^*(0,x), \tag{3}$$

and the *rational time* for stopping the buyer's observations and exercising the option is

$$\tau_T^* = \inf\{0 \leqslant t \leqslant T \colon Y^*(t,S_t) = g(S_t)\} \tag{4}$$

(we denoted this time by τ_0^T in § 3a) or, equivalently,

$$\tau_T^* = \inf\{0 \leqslant t \leqslant T \colon (t,S_t) \in D^T \cup \{(T,x) \colon x \in E\}\}. \tag{5}$$

4) The *stopping domain* D^T and the *domain of continued observations* C^T are *simply connected* and have the following structure:

$$D^T = \bigcup_{0 \leqslant t < T} \{(t,x) \colon Y^*(t,x) = g(x)\}, \tag{6}$$

$$C^T = \bigcup_{0 \leqslant t < T} \{(t,x) \colon Y^*(t,x) > g(x)\}. \tag{7}$$

5) The function $Y^* = Y^*(t,x)$ on $[0,T) \times E$ belongs to the class $C^{1,2}$.

For each fixed $x \in E$ the function $Y^*(\cdot,x)$ is *nonincreasing* in t; for each fixed $t \in [0,T)$ the function $Y^*(t,\cdot)$ is *nondecreasing* and *convex* (downwards) in x.

6) The *interface function* $x^* = x^*(t)$ is nonincreasing on $[0,T)$, and the sets C_t^T and D_t^T have the following form for $t < T$:

$$C_t^T = \{x \in E \colon S_t < x^*(t)\},$$
$$D_t^T = \{x \in E \colon S_t \geqslant x^*(t)\}.$$

For $t = T$ we have $C_T^T = \varnothing$ and $D_T^T = E$.

If $\lambda = 0$, then $x^*(t) = \infty$ for $t < T$, which corresponds to the equalities

$$C_t^T = E, \qquad D_t^T = \varnothing$$

for each $t < T$. In other words, for $t < T$ observations should go on whatever the prices S_t can be, which is a consequence of the fact that the process $(e^{-rt}(S_t - K)^+)_{t\geqslant 0}$ is a submartingale, so that by Doob's optimal stopping theorem,

$$\mathsf{E}_x e^{-r\tau}(S_\tau - K)^+ \leqslant \mathsf{E}_x e^{-rT}(S_T - K)^+$$

for each $\tau \in \mathfrak{M}_0^T$.

A similar result is known in the discrete-time case (R. Merton, [346]), and we have interpreted it as follows (see Chapter VI, § 5b): the *standard American and European call options 'coincide'*.

7) The function $Y^* = Y^*(t,x)$, $t \in [0,T]$, $x \in E$, and the interface function $x^* = x^*(t)$, $0 \leqslant t < T$, are the solutions of the following 'two-phase' *Stephan* (or *free-boundary*) *problem*:

$$-\frac{\partial Y^*(t,x)}{\partial t} + \beta Y^*(t,x) = LY^*(t,x) \tag{8}$$

in the domain $C^T = \{(t,x)\colon x < x^*(t),\ t \in [0,T)\}$;

$$Y^*(t,x) = g(x) \tag{9}$$

in the domain $D^T \cup \{(T,x)\colon x \in E\}$; while at the interface surface $x^* = x^*(t)$, $0 \leqslant t < T$, we have the *Dirichlet condition* (P. G. L. Dirichlet)

$$Y^*\big(t, x^*(t)\big) = g\big(x^*(t)\big) \tag{10}$$

and the *Neumann-type condition* (K. Neumann)

$$\frac{\partial Y^*(t,x)}{\partial x}\bigg|_{x\uparrow x^*(t)} = \frac{dg(x)}{dx}\bigg|_{x\downarrow x^*(t)}, \tag{11}$$

(called above the *condition of smooth pasting*).

We shall now make several observations concerning the above results; for detailed proofs of these results see the papers mentioned at the end of § 3b.1.

We discussed already the validity of (1) at the end of § 3a. The fact that τ_T^* is optimal follows from the general theory of optimal stopping rules (see, e.g., [441; Chapter III, § 3]). As regards the smoothness of $Y^*(t,x)$ and the derivation of (8), see, e.g., [247], [363], and [467].

While condition (10) is fairly natural. *condition of smooth pasting* (11) is not that obvious. In [200] and [441; Chapter III, § 8] we have shown that (11) must hold at the boundary of the stopping domain under rather general assumptions.

We also recall that we have already encountered conditions of smooth pasting in our considerations of approximations in discrete-time problems (§ 5 in Chapter VI) and in our discussions of American options in the case of an infinite time horizon (§ 2 in this chapter).

It is worth noting that while in the standard Stephan problem of mathematical physics that we considered in the previous section *each phase* satisfies an equation 'of its own', in the optimal stopping problems we have a differential equation for $Y^*(t,x)$ only in *one phase* (in the domain of continued observations), while in the other (the stopping domain) $Y^*(t,x)$ must coincide with the fixed function $g(x)$.

We note also that $\left.\dfrac{dg(x)}{dx}\right|_{x\downarrow x^*(t)} = 1$ in our case of $g(x) = (x-K)^+$ because $x^*(t) > K$, $0 \leqslant t < T$. (It is easy to deduce the last inequality from the fact that $Y^* = Y^*(t,x)$ is a β-excessive majorant of the function $g = g(x)$.)

As concerns the solubility of the Stephan problem (8)–(11) and the properties of the interface function $x^* = x^*(t)$ see [467] and [363] (see also the comments to the second paper).

2. Put options. In that case $g(x) = (K-x)^+$. Properties 1)–4) remain valid and the function $Y^* = Y^*(t,x)$ is in the class $C^{1,2}$ again. For each fixed $x \in E$ the function $Y^*(\,\cdot\,,x)$ is nonincreasing in t; for each fixed $t \in [0,T)$ the function $Y^*(t,\,\cdot\,)$ is nonincreasing and (downwards) convex in x.

For each $\lambda \geqslant 0$ the sets C_t^T and D_t^T have the following form for $t < T$:

$$C_t^T = \{x \colon S_t > x^*(t)\},$$
$$D_t^T = \{x \colon S_t \leqslant x^*(t)\}.$$

For $t = T$ we have $C_T^T = \varnothing$ and $D_T^T = E$.

The interface function $x^* = X^*(t)$ is nondecreasing in t; if $\lambda = 0$, then $\lim\limits_{t\uparrow T} x^*(t) = K$.

The Stephan problem for $Y^*(t,x)$ and $x^*(t)$ can be formulated in a similar manner. Here conditions (8), (9), and (10) are the same, while (11) takes the following form for $0 \leqslant t < T$:

$$\left.\frac{dY^*(t,x)}{dx}\right|_{x\downarrow x^*(t)} = \left.\frac{dg(x)}{dx}\right|_{x\uparrow x^*(t)},$$

where

$$\left.\frac{dg(x)}{dx}\right|_{x\uparrow x^*(t)} = -1$$

because $g(x) = (K-x)^+$ and $x^*(t) < K$.

One can find additional information on the properties of the functions $Y^*(t,x)$ and $x^*(t)$ in the paper [363] devoted to standard American put options and containing an extensive bibliography on other types of options.

§ 3d. Relations between the Prices of European and American Options

1. We have previously mentioned that American options are in fact more commonly traded than European ones. However, while for the latter we have such remarkable results as, e.g., the *Black–Scholes formulas*, calculations for American options in problems with finite time horizon encounter great analytic difficulties. On the final count, this is due to difficulties with the solution of the corresponding Stephan problems.

It is clear from (1) and (2) in § 3a that $V^*(T,x) \geqslant V(T,x)$; this is, of course, fairly natural since, by condition, American options encourage the buyer not merely to wait for the execution but to choose its timing.

In this section we present several results on the relations between the prices of standard call and put options with pay-off functions $g(x) = (x-K)^+$ and $g(x) = (K-x)^+$, respectively.

We shall assume that $\lambda = 0$. Then formulae (1) and (2) in § 3a can be written as follows:

$$V(T,x) = \mathsf{E}_x e^{-rT} g(S_T) \tag{1}$$

and

$$V^*(T,x) = \sup_{\tau \in \mathfrak{M}_0^T} \mathsf{E}_x e^{-r\tau} g(S_\tau), \tag{2}$$

where $x = S_0$.

2. The question of the relation between the prices $V(T,x)$ and $V^*(T,x)$ for $g(x) = (x-K)^+$ is very easy to answer. In that case

$$V(T,x) = V^*(T,x), \tag{3}$$

and $\tau_T^* = T$ is an optimal stopping time in the class $\mathfrak{M}_0^T$ (see § 3c).

Remark. We emphasize that if $\lambda > 0$, then (3) does not hold any longer because the process $(e^{-(\lambda+r)t}(S_t - K)^+)_{t\geqslant 0}$ is no longer a submartingale (cf. § 3c).

We proceed now to the question of the size of the 'deficiency'

$$\Delta_0^T(x) \equiv V^*(T,x) - V(T,x) \tag{4}$$

for a standard call option ($g(x) = (K-x)^+$). We set $\lambda = 0$ and denote by $x^* = x^*(t)$, $0 \leqslant t < T$, the function defining the interface between the stopping domain and the domain of continued observations for optimal stopping time τ_T^*.

THEOREM. *For a standard call option,*

$$\Delta_0^T(x) = rK\mathsf{E}_x \int_0^T e^{-ru} I(S_u < x^*(u))\,du. \tag{5}$$

COROLLARY 1. *Let $\mathbb{P}_T$ and $\mathbb{P}_T^*$ be the rational prices of European and American call options ($\mathbb{P}_T = V(T, S_0)$, $\mathbb{P}_T^* = V^*(T, S_0)$). Then*

$$\begin{aligned} \mathbb{P}_T^* &= \mathbb{P}_T + rK\mathsf{E}_{S_0} \int_0^T e^{-ru} I(S_u < x^*(u))\,du \\ &= Ke^{-rT}\Phi(-y_-) - S_0\Phi(-y_+) + rK \int_0^T e^{-ru}\Phi\bigl(-y_-(u, x^*(u))\bigr)\,du, \end{aligned} \tag{6}$$

where (cf. the notation in § 1b)

$$y_\pm = \frac{\ln \dfrac{S_0}{K} + T\Bigl(r \pm \dfrac{\sigma^2}{2}\Bigr)}{\sigma\sqrt{T}}$$

and

$$y_-\bigl(u, x^*(u)\bigr) = \frac{\ln \dfrac{S_0}{x^*(u)} + u\Bigl(r - \dfrac{\sigma^2}{2}\Bigr)}{\sigma\sqrt{u}}. \tag{7}$$

Proof. Let

$$Y(t, x) = \mathsf{E}_{t,x} e^{-r(T-t)} g(S_T) \tag{8}$$

and let

$$Y^*(t, x) = \sup_{\tau \in \mathfrak{M}_t^T} \mathsf{E}_{t,x} e^{-r(\tau - t)} g(S_\tau), \tag{9}$$

where $g(x) = (K - x)^+$. Then for

$$\Delta_t^T(x) \equiv Y^*(t, x) - Y(t, x) \tag{10}$$

we have

$$e^{-rt}\Delta_t^T(x) = \mathsf{E}_{t,x}\bigl\{e^{-r\tau_t^T} g(S_{\tau_t^T}) - e^{-rT} g(S_T)\bigr\}, \tag{11}$$

where τ_t^T is the optimal stopping time in the problem (9).

By Itô's formula (Chapter III, § 5c)

$$d\bigl(e^{-ru}(K - S_u)^+\bigr) = e^{-ru}\,d(K - S_u)^+ - re^{-ru}(K - S_u)^+\,du, \tag{12}$$

and by the Itô–Meyer formula for convex functions (see Chapter VII, § 4a; [395; Chapter IV]; cf. also Tanaka's formula (17) in Chapter III, § 5c)

$$d(K - S_u)^+ = -I(S_u < K)\, dS_u + \frac{1}{2} L_u(K), \tag{13}$$

where

$$L_u(K) = \lim_{\varepsilon \downarrow 0} \frac{1}{2\varepsilon} \int_0^u I(|S_t - K| \leqslant \varepsilon)\, dt \tag{14}$$

is the local time of process $S = (S_t)_{t \geqslant 0}$ (on $[0, u]$) at the level K.

By (11)–(13) we obtain

$$\begin{aligned} e^{-rT} \Delta_t^T(x) &= -\mathsf{E}_{t,x} \int_{\tau_t^T}^T d\big(e^{-ru}(K - S_u)^+\big) \\ &= -\mathsf{E}_{t,x} \int_{\tau_t^T}^T e^{-ru} \big\{ -I(S_u < K)\, dS_u + \tfrac{1}{2} dL_u(K) \\ &\qquad\qquad - r(K - S_u) I(S_u < K)\, du \big\} \\ &= \mathsf{E}_{t,x} \int_{\tau_t^T}^T e^{-ru} \big\{ -dL_u(K) + I(S_u < K) \\ &\qquad\qquad \times \big[rS_u\, du + \sigma S_u\, dW_u + (rK - rS_u)\, du \big] \big\} \\ &= \mathsf{E}_{t,x} \int_{\tau_t^T}^T e^{-ru} \big\{ rK I(S_u < K)\, du - dL_u(K) \big\}. \end{aligned} \tag{15}$$

For $t \leqslant T$ we set

$$A_t = \int_{\tau_0^T}^{\tau_t^T} e^{-ru} \big\{ rK I(S_u < K)\, du - dL_u(K) \big\}. \tag{16}$$

Then, since $\tau_T^T = T$, it follows from (15) that

$$e^{-rt} \Delta_t^T(x) = \mathsf{E}_{t,x} [A_T - A_t]. \tag{17}$$

We represent now A_t as follows:

$$A_t = A_t^1 + A_t^2,$$

where

$$A_t^1 = \int_{\tau_0^T}^{\tau_t^T} e^{-ru} I\big(S_u \leqslant x^*(u)\big) \big\{ rK I(S_u < K)\, du - dL_u(K) \big\},$$

$$A_t^2 = \int_{\tau_0^T}^{\tau_t^T} e^{-ru} I\big(S_u > x^*(u)\big) \big\{ rK I(S_u < K)\, du - dL_u(K) \big\}.$$

Since $x^*(u) < K$ for all $u < T$, it follows that

$$A_t^1 = \int_{\tau_0^T}^{\tau_t^T} e^{-ru} I(S_u \leqslant x^*(u)) rK\,du$$
$$= rK \int_0^t e^{-ru} I\big(S_u < x^*(u)\big)\,du. \tag{18}$$

The process $A^1 = (A_t^1)_{t \leqslant T}$ is a predictable submartingale and, by Corollary 2 in Chapter III, § 5b, it is the compensator of itself. A slightly more refined analysis (see [134], [135], and [363]) shows that the compensator of the process $A^2 = (A_t^2)_{t \leqslant T}$ vanishes. Hence

$$e^{-rt}\Delta_t^T(x) = \mathsf{E}_{t,x}[A_T - A_t] = \mathsf{E}_{t,x}[A_T^1 - A_t^1]$$
$$= rK\,\mathsf{E}_{t,x} \int_t^T e^{-ru} I\big(S_u < x^*(u)\big)\,du \tag{19}$$

so that (5) holds for $\Delta_0^T(x)$.

Finally, formula (6) in Corollary 1 is a consequence of (5) and relation (18) in § 1b for $\mathbb{P}_T$. This completes the proof of the theorem and Corollary 1.

COROLLARY 2. *The functions $Y^*(t,x)$ and $x^*(t)$ are connected by the relation*

$$Y^*\big(t, x^*(t)\big) = K - x^*(t), \qquad t \leqslant T, \tag{20}$$

which can be regarded as an integral relation for the interface function $x^* = x^*(t)$, $t \leqslant T$, with $x^*(T) \equiv \lim_{t \to T} x^*(t)$.

It should be pointed out that the function $Y^*(t,x)$ is in fact also unknown. In place of this function one considers in practice approximations $Y_\Delta^*(t,x)$ calculated by means of backward induction (see § 3b.1). Replacing $Y^*(t,x)$ by $Y_\Delta^*(t,x)$ in (19) we obtain a function $x_\Delta^* = x_\Delta^*(t)$, $t \leqslant T$, which is taken for an approximation to $x^* = x^*(t)$, $t \leqslant T$.

4. European and American Options in a Diffusion $(B, \mathcal{P})$-Bondmarket

§ 4a. Option Pricing in a Bondmarket

1. So far we have considered only options in (B, S)-stockmarkets, whereas one encounters most various kinds of them in practice: options on eurodollars, futures, currency, etc., even options on options. Besides standard put and call options their various combinations are also traded. Many financial instruments in the option family have an elaborate structure as regards the corresponding pay-off functions and the securities underlying option contracts.

The diversity of options, some of which are said to be 'exotic', can be judged by their names: up-and-out put, up-and-in put, down-and-out call, down-and-in call, barrier option, Bermuda option, Rainbow option, Russian option, knock-out option, digital option, all-or-nothing, one-touch all-or-nothing, supershares, ... (see [232], [414], or [415]).

In our discussions of pricing for these options and other derivative instruments we should point out that the general methods remain the same as in the models considered by F. Black, M. Scholes, and R. Merton ([44], [346]) in the case of a (B, S)-stockmarket. Again, two approaches are possible: the *martingale* approach and the approach based on direct considerations of the '*fundamental equation*' (cf. § § 1b, c).

2. The discussion that follows relates to calculations for standard European and American options in the case where in place of a (B, S)-market we consider a $(B, \mathcal{P})$-market consisting of a bank account $B = (B_t)_{t \leqslant T}$ and *one* bond of maturity T with price described by a (positive) process $\mathsf{P} = (\mathsf{P}(t, T))_{t \leqslant T}$ satisfying the condition $\mathsf{P}(T, T) = 1$.

In accordance with Chapter III, § 4a and Chapter VII, § 5a, we shall take for our description of the $(B, \mathcal{P})$-market the *indirect* approach, where we assume that the state of the bank account $(B_t)_{t \leqslant T}$ can be expressed by the formula

$$B_t = B_0 \exp\biggl(\int_0^t r(s)\, ds\biggr) \tag{1}$$

with some stochastic process of *interest rate* $r = (r(t))_{t \leqslant T}$.

As regards the dynamics of the bond price process $\mathsf{P} = (\mathsf{P}(t, T))_{t \leqslant T}$ we assume that the discounted prices

$$\overline{\mathsf{P}}(t, T) = \frac{\mathsf{P}(t, T)}{B_t}, \qquad t \leqslant T, \tag{2}$$

make up a *martingale* with respect to the *initial* measure on $(\Omega, \mathscr{F}_T, (\mathscr{F}_t)_{t \leqslant T})$.

By Theorem 1 in Chapter VII, § 5a we obtain

$$\mathsf{P}(t, T) = \mathsf{E}\biggl(\exp\biggl(-\int_t^T r(s)\, ds\biggr) \biggm| \mathscr{F}_t\biggr), \tag{3}$$

and by Theorem 2 in the same section the $(B, \mathcal{P})$-market in question is *arbitrage-free* (e.g., in the NA_+-version of this concept).

3. From (1) and (3) we see that the dynamics of the processes $(B_t)_{t \leqslant T}$ and $(\mathsf{P}(t, T))_{t \leqslant T}$ in our $(B, \mathcal{P})$-market depends considerably on the structure of the process $r = (r(t))_{t \leqslant T}$.

Our main assumption about this process is that it is a *diffusion Gauss-Markov process* described by the stochastic differential equation

$$dr(t) = \bigl(\alpha(t) - \beta(t) r(t)\bigr)\, dt + \gamma(t)\, dW_t, \tag{4}$$

with Wiener process $(W_t)_{t \leqslant T}$ and the (nonrandom) initial condition $r(0) = r_0$. We assume that the functions $\alpha(t)$, $\beta(t)$, and $\gamma(t)$ are deterministic and

$$\int_0^T \bigl(|\alpha(t)| + |\beta(t)| + \gamma^2(t)\bigr)\, dt < \infty. \tag{5}$$

Under these assumptions equation (4) has a unique (strong) solution

$$r(t) = g(t)\biggl\{r_0 + \int_0^t \frac{\alpha(s)}{g(s)}\, ds + \int_0^t \frac{\gamma(s)}{g(s)}\, dW_s\biggr\}, \tag{6}$$

where

$$g(t) = \exp\biggl(-\int_0^t \beta(s)\, ds\biggr) \tag{7}$$

is the *fundamental solution* of the equation

$$g(t) = 1 - \int_0^t \beta(s) g(s)\, ds. \tag{8}$$

Remark 1. Recalling our discussion in Chapter III, § 4a we see that (4) is just the *Hull–White model*, of which the *Merton*, the *Vasiček*, and the *Ho–Lee models* are particular cases (see formulas (14), (7), (8), and (12) in Chapter III, § 4a).

4. Since $r = (r(t))_{t \leqslant T}$ is a Markov process, it follows that

$$\mathsf{P}(t,T) = \mathsf{E}\biggl[\exp\biggl(-\int_t^T r(s)\,ds\biggr)\,\bigg|\,r(t)\biggr]. \tag{9}$$

We set $I(t,T) = \int_t^T r(s)\,ds$. Then we see easily from (6) that

$$\mathsf{E}\bigl(I(t,T)\,|\,r(t)\bigr) = r(t)\int_t^T \frac{g(u)}{g(t)}\,du + \int_t^T \biggl[\int_t^u \frac{g(u)}{g(s)}\alpha(s)\,ds\biggr]\,du, \tag{10}$$

$$\mathsf{D}\bigl(I(t,T)\,|\,r(t)\bigr) = \int_t^T \biggl[\int_s^T \frac{g(u)}{g(s)}\gamma(s)\,du\biggr]^2\,ds. \tag{11}$$

Hence it is easy to obtain by (3) the following representation for

$$\mathsf{P}(t,T) = \mathsf{E}\bigl[\exp\bigl(-I(t,T)\bigr)\,|\,r(t)\bigr] = \exp\Bigl(\frac{1}{2}\mathsf{D}\bigl(I(t,T)\,|\,r(t)\bigr) - \mathsf{E}\bigl(I(t,T)\,|\,r(t)\bigr)\Bigr):$$

$$\mathsf{P}(t,T) = \exp\bigl(A(t,T) - r(t)B(t,T)\bigr), \tag{12}$$

where

$$A(t,T) = \frac{1}{2}\int_t^T \biggl[\int_s^T \frac{g(u)}{g(s)}\gamma(s)\,du\biggr]^2\,ds - \int_t^T \biggl[\int_t^u \frac{g(u)}{g(s)}\alpha(s)\,ds\biggr]\,du \tag{13}$$

and

$$B(t,T) = \int_t^T \frac{g(u)}{g(t)}\,du. \tag{14}$$

Remark 2. Using the terminology of Chapter III, § 4c we call the models with prices $\mathsf{P}(t,T)$ representable as in (12) *single-factor affine models*. Our additional assumption that $r = (r(t))_{t \leqslant T}$ is a *Gauss-Markov* process enables one to carry out fairly complete calculations for European and American options in $(B,\mathcal{P})$-markets for such models (often called also *single-factor Gaussian models*). We devote § § 4b, c below to these problems.

Remark 3. As regards the agreement between various models describing the dynamics of stock prices and empirical data, see, for instance, [257].

§ 4b. European Option Pricing in Single-Factor Gaussian Models

1. We shall consider a model of a $(B, \mathcal{P})$-market consisting of a bank account and a bond and completely determined by a *single factor*, the interest rate $r = (r(t))_{t \leqslant T}$ that is a *Gauss–Markov* process satisfying stochastic differential equation (4) in § 4a with (nonrandom) initial condition $r(0) = r_0$.

Let T^0 be some time ($T^0 < T$) treated as the *time of exercising* a European option with pay-off function $f_{T^0} = (\mathsf{P}(T^0, T) - K)^+$ for a call option and $f_{T_0} = (K - \mathsf{P}(T^0, T))^+$ for a put option.

THEOREM. *The rational price $\mathbb{C}^0(T^0, T)$ of a standard call option in the single-factor Gaussian model of a $(B, \mathcal{P})$-market in question is described by the formula*

$$\boxed{\mathbb{C}^0(T^0, T) = \mathsf{P}(0, T)\Phi(d_+) - K\mathsf{P}(0, T^0)\Phi(d_-)} \tag{1}$$

where

$$d_\pm = \frac{\ln \dfrac{\mathsf{P}(0, T)}{K\mathsf{P}(0, T^0)} \pm \dfrac{1}{2}\sigma^2(T^0, T)B^2(T^0, T)}{\sigma(T^0, T)B(T^0, T)}, \tag{2}$$

$$B(T^0, T) = \int_{T^0}^{T} \frac{g(u)}{g(T^0)}\, du, \tag{3}$$

$$\sigma(T^0, T) = \left(\int_{T^0}^{T} \left[\int_s^T \frac{g(u)}{g(s)} \gamma(s)\, du \right]^2 ds \right)^{1/2}, \tag{4}$$

$$g(u) = \exp\left(- \int_0^u \beta(s)\, ds \right). \tag{5}$$

The rational price $\mathbb{P}^0(T^0, T)$ of a standard put option is described by the formula

$$\boxed{\mathbb{P}^0(T^0, T) = K\mathsf{P}(0, T^0)\Phi(-d_-) - \mathsf{P}(0, T)\Phi(-d_+)} \tag{6}$$

Before proceeding to the proof of (1) and (6), we point out that these relations are very similar to the formula for the rational prices $\mathbb{C}(T)$ and $\mathbb{P}(T)$ in the case of stock (see (9) and (18) in § 1b).

This is not that surprising given that, similarly to the prices S_t in the Black–Merton–Scholes model, the prices $\mathsf{P}(t,T)$ in the present model have a *logarithmically normal* structure:

$$\ln \mathsf{P}(t,T) = A(t,T) - r(t)B(t,T),$$

where $(r(t))_{t\leqslant T}$ is a Gaussian process,

$$\ln \frac{S_t}{S_0} = \Bigl(r - \frac{\sigma^2}{2}\Bigr)t + \sigma W_t,$$

and $(W_t)_{t\leqslant T}$ is a Wiener (and therefore also Gaussian) process.

It is maybe more surprising that it took so long between the publication of the Black–Scholes formula in 1973 and 1989, when F. Jamshidian published the paper [256] containing formulas (1) and (6) for the Vasiček model ($\alpha(t)\equiv\alpha$, $\beta(t)\equiv\beta$, $\gamma(t)\equiv\gamma$; see (4) in § 4a and (8) in Chapter III, § 4a). We follow mainly [257] in our proof below.

2. By pricing theory for complete arbitrage-free markets (see § 5 in Chapter VII), assuming that the initial probability measure on $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\leqslant T})$, $\mathscr{F}_T = \mathscr{F}$, is a martingale measure and setting

$$R(t) = \exp\Bigl(-\int_0^t r(u)\,du\Bigr), \tag{7}$$

we obtain

$$\begin{aligned}\mathbb{C}^0(T^0,T) &= \mathsf{E}R(T^0)(\mathsf{P}(T^0,T)-K)^+\\ &= \mathsf{E}\Bigl(I\bigl(\mathsf{P}(T^0,T)>K\bigr)R(T^0)\bigl(\mathsf{P}(T^0,T)-K\bigr)\Bigr)\\ &= \mathsf{E}\Bigl(I\bigl(\mathsf{P}(T^0,T)>K\bigr)R(T^0)\mathsf{P}(T^0,T)\Bigr) - K\mathsf{E}\Bigl(I\bigl(\mathsf{P}(T^0,T)>K\bigr)R(T^0)\Bigr).\end{aligned} \tag{8}$$

Clearly, we have the coincidence of the events

$$\bigl\{\mathsf{P}(T^0,T)>K\bigr\} = \bigl\{A(T^0,T) - r(T^0)B(T^0,T) > \ln K\bigr\} = \bigl\{r(T^0)\leqslant r^*\bigr\}, \tag{9}$$

where

$$r^* = \frac{\ln K - A(T^0,T)}{-B(T^0,T)} \tag{10}$$

and $A(t,T)$ and $B(t,T)$ are defined by (13) and (14) in § 4a.

Let

$$\xi = r(T^0), \qquad \eta = \int_0^T r(u)\,du, \qquad \zeta = \int_0^{T^0} r(u)\,du.$$

Then by (8) and (9) we obtain

$$\mathbb{C}^0(T^0, T) = \mathsf{E}\big(I(\xi \leqslant r^*)e^{-\eta}\big) - K\mathsf{E}\big(I(\xi \leqslant r^*)e^{-\zeta}\big). \tag{11}$$

For a further simplification of this formula we bear on the following result, which can be verified by a direct calculation (see [257; Lemma 4.2]).

LEMMA. *Let (X, Y) be a Gaussian pair of random variables with vector of mean values (μ_X, μ_Y) and covariance matrix $\begin{pmatrix} \sigma_X^2, & \rho_{XY} \\ \rho_{XY}, & \sigma_Y^2 \end{pmatrix}$. Then*

$$\mathsf{E}I(X \leqslant x)\exp(-Y) = \exp\Big(\frac{1}{2}\sigma_Y^2 - \mu_Y\Big)\,\Phi(\widetilde{x}) \tag{12}$$

and

$$\mathsf{E}I(X \leqslant x)X\exp(-Y) = \exp\Big(\frac{1}{2}\sigma_Y^2 - \mu_Y\Big)\cdot\Big\{(\mu_X - \rho_{XY})\Phi(\widetilde{x}) - \sigma_X\varphi(\widetilde{x})\Big\}, \tag{13}$$

where

$$\widetilde{x} = \frac{x - (\mu_X - \rho_{XY})}{\sigma_X},$$

$$\varphi(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}, \qquad \Phi(x) = \int_{-\infty}^x \varphi(y)\,dy.$$

Taking formulas (6), (10), and (11) in § 4a into account it is easy to calculate that

$$\mu_\xi = \mathsf{E}r(T^0) = g(T^0)\Big(r_0 + \int_0^{T^0}\frac{1}{g(s)}\alpha(s)\,ds\Big),$$

$$\mu_\eta = \mathsf{E}\int_0^T r(u)\,du = r_0\int_0^T g(u)\,du + \int_0^T\Big[\int_0^u \frac{g(u)}{g(s)}\alpha(s)\,ds\Big]\,du,$$

$$\mu_\zeta = \mathsf{E}\int_0^{T^0} r(u)\,du = r_0\int_0^{T^0} g(u)\,du + \int_0^{T^0}\Big[\int_0^u \frac{g(u)}{g(s)}\alpha(s)\,ds\Big]\,du$$

and

$$
\begin{aligned}
\sigma_\xi^2 &= \mathsf{D}r(T^0) = \int_0^{T^0} \gamma^2(s)\left(\frac{g(T^0)}{g(s)}\right)^2 ds,\\
\sigma_\eta^2 &= \mathsf{D}\int_0^T r(u)\,du = \int_0^T \left[\int_s^T \frac{g(u)}{g(s)}\gamma(s)\,du\right]^2 ds,\\
\sigma_\zeta^2 &= \mathsf{D}\int_0^{T^0} r(u)\,du = \int_0^{T^0} \left[\int_s^{T^0} \frac{g(u)}{g(s)}\gamma(s)\,du\right]^2 ds,\\
\rho_{\xi\zeta} &= \mathsf{Cov}\left(r(T^0), \int_0^{T^0} r(u)\,du\right)\\
&= \int_0^{T^0} \left(\gamma^2(s)\frac{g(T^0)}{g(s)} \int_s^{T^0} \frac{g(u)}{g(s)}\,du\right) ds,\\
\rho_{\xi\eta} &= \mathsf{Cov}\left(r(T^0), \int_0^{T} r(u)\,du\right)\\
&= \mathsf{Cov}\left(r(T^0), \int_0^{T^0} r(u)\,du\right) + \mathsf{Cov}\left(r(T^0), \int_{T^0}^{T} r(u)\,du\right)\\
&= \rho_{\xi\zeta} + \sigma_\xi^2 \int_{T^0}^{T} \frac{g(u)}{g(T^0)}\,du.
\end{aligned}
$$

By (11) and (12) we obtain

$$
\begin{aligned}
\mathbb{C}^0(T^0,T) &= \mathsf{E}\big(I(\xi \leqslant r^*)e^{-\eta}\big) - K\mathsf{E}\big(I(\xi \leqslant r^*)e^{-\zeta}\big)\\
&= \exp\Big(\frac{1}{2}\sigma_\eta^2 - \mu_\eta\Big)\,\Phi\left(\frac{r^* - (\mu_\xi - \rho_{\xi\eta})}{\sigma_\xi}\right)\\
&\quad - K\exp\Big(\frac{1}{2}\sigma_\zeta^2 - \mu_\zeta\Big)\,\Phi\left(\frac{r^* - (\mu_\xi - \rho_{\xi\zeta})}{\sigma_\xi}\right).
\end{aligned}
\tag{14}
$$

Substituting here the above values of μ_ξ, μ_η, μ_ζ, σ_ξ, σ_η, σ_ζ, $\rho_{\xi\eta}$, and $\rho_{\xi\zeta}$, after some algebraic transformations (see [257; Appendix 4b]) we obtain required formula (1).

Relation (6) is seen to be an immediate consequence of (1) once one takes into account the equality

$$
\big(K - \mathsf{P}(T^0,T)\big)^+ = \big(\mathsf{P}(T^0,T) - K\big)^+ - \mathsf{P}(T^0,T) + K.
$$

(Cf., e.g., the derivation of (9) in Chapter VI, § 4d.) This completes the proof.

3. It follows from (6) that the rational price $\mathbb{P}^0(T^0, T)$ can be determined from the 'initial' prices $\mathsf{P}(0, T^0)$, $\mathsf{P}(0, T)$, the constant K, and the quantity $\sigma(T^0, T)B(T^0, T)$, which, in its turn is defined by the coefficients $\beta(s)$ and $\gamma(s)$ for $T^0 \leqslant s \leqslant T$.

In the case of the Vasiček model we have $\beta(s) \equiv s$, $\gamma(s) \equiv \gamma$, and it is easy to see that

$$\sigma(T^0, T)B(T^0, T) = \frac{\gamma}{\beta}\big(1 - e^{-\beta(T-T^0)}\big)\Big(\frac{1}{2\beta}\big(1 - e^{-2\beta T^0}\big)\Big)^{1/2}.$$

The initial prices $\mathsf{P}(0, T^0)$ and $\mathsf{P}(0, T)$ can be found from the following formula (see (12) in § 4a):

$$\mathsf{P}(0, t) = \exp\{A(0, t) - r_0 B(0, t)\},$$

where

$$A(0, t) = \frac{1}{\beta}\big[1 - e^{-\beta t} - \beta t\big]\left[\frac{\beta}{\alpha} - \frac{\gamma^2}{2\beta^2}\right] - \frac{\gamma^2}{4\beta^3}\big[1 - e^{-\beta t}\big]^2,$$

$$B(0, t) = \frac{1}{\beta}\big[1 - e^{-\beta t}\big].$$

§ 4c. American Option Pricing in Single-Factor Gaussian Models

1. We continue our discussion of single-factor Gaussian $(B, \mathcal{P})$-models. We found formulas for $\mathbb{C}^0(T^0, T)$ and $\mathbb{P}^0(T^0, T)$ in § 4b; now let $\mathbb{C}^*(T^0, T)$ and $\mathbb{P}^*(T^0, T)$ be the corresponding rational prices of American (call and put) options. Here we assume that the exercise times belong to the class

$$\mathfrak{M}_0^{T^0} = \{\tau = \tau(\omega) \colon 0 \leqslant \tau(\omega) \leqslant T^0,\ \omega \in \Omega\}.$$

The $(B, \mathcal{P})$-market in question is both arbitrage-free and complete, and in accordance with the general theory of pricing in such models (see Chapter VI, § 2 and Chapter VII, § 5),

$$\mathbb{C}^*(T^0, T) = \sup_{\tau \in \mathfrak{M}_0^{T^0}} \mathsf{E}\exp\Big(-\int_0^\tau r(u)\,du\Big)\big(\mathsf{P}(\tau, T) - K\big)^+ \tag{1}$$

and

$$\mathbb{P}^*(T^0, T) = \sup_{\tau \in \mathfrak{M}_0^{T^0}} \mathsf{E}\exp\Big(-\int_0^\tau r(u)\,du\Big)\big(K - \mathsf{P}(\tau, T)\big)^+ \tag{2}$$

Since

$$\mathsf{P}(\tau, T) = \exp\big(A(\tau, T) - r(\tau)B(\tau, T)\big), \tag{3}$$

equalities (1) and (2) are related to standard optimal stopping problems of finding the suprema

$$\sup_{\tau \in \mathfrak{M}_0^{T^0}} \mathsf{E}\exp\Big(-\int_0^\tau r(u)\,du\Big)G(\tau, T; r(\tau)) \tag{4}$$

for Markov processes $r = (r(t))_{t \leqslant T}$ and nonnegative functions $G(\tau, T; r(\tau))$. There exists a well-developed theory for such problems (see, for instance, [441]).

2. The question of the value of $\mathbb{C}^*(T^0, T)$ is very simple because, as it is also for standard call options in (B, S)-markets, the process

$$\left(\exp\left(-\int_0^t r(u)\,du\right)(\mathsf{P}(t,T)-K)^+\right)_{t\leqslant T}$$

is a submartingale and $\mathbb{C}^*(T^0, T) = \mathbb{C}^0(T^0, T)$ by Doob's stopping theorem (we interpreted this equality in the following way in Chapter VI, § 5b and in § 3b of the present chapter: an American call option coincides in effect with a European option).

Proceeding to the prices $\mathbb{P}^*(T^0, T)$, we consider the variables

$$Y^*(t,r) = \sup_{\tau\in\mathfrak{M}_t^{T^0}} \mathsf{E}_{t,r} \exp\left(-\int_t^\tau r(u)\,du\right) G\big(\tau, T; r(\tau)\big), \tag{5}$$

where $\mathsf{E}_{t,r}$ is averaging under the assumption $r(t) = r$, $\mathfrak{M}_t^{T^0}$ is the class of stopping times $\tau = \tau(\omega)$ such that $t \leqslant \tau(\omega) \leqslant T^0$, and

$$G\big(t, T; r(t)\big) = (K - \mathsf{P}(t,T))^+ = \Big(K - \exp\big(A(t,T) - r(t)B(t,T)\big)\Big)^+$$

with $A(t,T)$ and $B(t,T)$ as in formulas (13) and (14) in § 4a.

We set

$$C^T = \big\{(t,r)\colon Y^*(t,r) > G(t,T;r),\ 0 \leqslant t < T, r > 0\big\}$$

and

$$D^T = \big\{(t,r)\colon Y^*(t,r) = G(t,T;r),\ 0 \leqslant t < T, r > 0\big\}.$$

On the basis of the characterization of the prices $Y^*(t,r)$ as the smallest excessive majorants of the functions $G(t,T;r)$ (see [340], [363], [441; Chapter III], [467], [478]), we can show that there exists a continuous interface function $r^* = r^*(t)$, $t < T^0$, such that the domains C^T and D^T (of the continuation of observations and their stopping, respectively) have the following form:

$$C^T = \big\{(t,r)\colon r(t) < r^*(t),\ 0 \leqslant t < T,\ r > 0\big\}$$

and

$$D^T = \big\{(t,r)\colon r(t) \geqslant r^*(t),\ 0 \leqslant t < T,\ r > 0\big\}.$$

Here the function $Y^* = Y^*(t,r)$ and the interface function $r^* = r^*(t)$ are solutions of the following *Stephan problem*:

$$\frac{\partial Y^*(t,r)}{\partial t} + LY^*(t,r) - rY^*(t,r) = 0, \qquad (t,r) \in C^T,$$

where

$$LY^*(t,r) = \big(\alpha(t) - \beta(t)r\big)\frac{\partial Y^*(t,r)}{\partial r} + \frac{1}{2}\gamma^2(t)\frac{\partial^2 Y^*(t,r)}{\partial r^2}\,;$$

$$Y^*(t,r) = G(t,T;r) \tag{6}$$

in the domain D^T, and we have the *condition of smooth pasting* on ∂D^T:

$$\frac{\partial Y^*(t,r)}{\partial r}\bigg|_{r\uparrow r^*(t)} = \frac{\partial G(t,T;r)}{\partial r}\bigg|_{r\downarrow r^*(t)}. \tag{7}$$

We know of no precise analytic solution to this problem (as, incidentally, also in the case of (B,S)-models; see § 3c). At the same time, in view of the abundance of American options in real life one would like to have an idea on the difference between the price $\mathbb{P}^*(T^0,T)$ of an American option and the price $\mathbb{P}^0(T^0,T)$ of a European option, and on the behavior of the interface $r^* = r^*(t)$, $t < T^0$.

Arguments similar to the ones we used in § 3d, while looking for relations between the prices of American and European options in a (B,S)-market, can be used for $(B,\mathcal{P})$-markets in our case and bring us to the following result (cf. (19) in § 3d): for $0 \leqslant t < T^0$ we have

$$Y^*(t,r) = Y^0(t,r) + K\int_t^{T^0} \mathsf{E}_{t,r}\left\{\exp\left(-\int_t^s r(u)\,du\right)r(s)I\big(r(s) \geqslant r^*(s)\big)\right\}ds, \tag{8}$$

where

$$\begin{aligned} Y^0(t,r) &= \mathsf{E}_{t,r}\exp\left(-\int_t^{T^0} r(u)\,du\right)\big(K - \mathsf{P}(T^0,T)\big)^+ \\ &= \mathsf{E}_{t,r}\exp\left(-\int_t^{T^0} r(u)\,du\right)\Big(K - \exp\big(A(T^0,T) - r(T^0)B(T^0,T)\big)\Big). \end{aligned} \tag{9}$$

Using formulas (12) and (13) in § 4b and the expressions there for $\mu_\xi, \mu_\eta, \dots$, after simple transformations we obtain

$$\begin{aligned} Y^*(t,r) &= Y^0(t,r) \\ &\quad + K\int_t^{T^0} \mathsf{P}(t,s)\big\{\Phi\big(-\nu^*(t,s)\big)f(t,s) + \sigma(t,s)\varphi\big(\nu^*(t,s)\big)\big\}\,ds \end{aligned} \tag{10}$$

for $0 \leqslant t \leqslant T^0$, where

$$\begin{aligned} \sigma^2(t,s) &= \mathsf{D}\big(r(s)\,|\,r(t) = r\big), \\ f(t,s) &= -\frac{\partial}{\partial s}\ln \mathsf{P}(t,s), \\ \nu^*(t,s) &= \frac{r^*(s) - f(t,s)}{\sigma(t,s)}. \end{aligned}$$

(See a detailed derivation of (8) and (10) in [257].)

In particular, since

$$\mathbb{P}^*(T^0, T) = Y^*(0, r_0) \quad \text{and} \quad \mathbb{P}^0(T^0, T) = Y^0(0, r_0),$$

it follows that

$$\mathbb{P}^*(T^0, T) = \mathbb{P}^0(T^0, T) + K \int_0^{T^0} \mathsf{P}(0, s)\big\{\Phi\big(-\nu^*(0, s)\big) f(0, s) + \sigma(0, s)\varphi\big(\nu^*(0, s)\big)\big\}\, ds.$$

In conclusion we point out that (10) enables one to find (by means of backward induction), at least approximately, also the values of the price $\mathbb{P}^*(T^0, T)$ and the interface function $r^* = r^*(t)$, $t < T^0$.

As regards various methods of American option pricing (including the 'case with dividends'), see, for instance, [28], [29], [56], [57], [179], [257], [376], [478], or [479].

Bibliography

1. M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*, Dover, New York, 1968.
2. *Adaptive Computational Methods in Finance and Trading. Report of the Intertek Group. Foreword by R. B. Olsen*, "Olsen and Associates", Research Institute for Applied Economics, Zürich, December, 1994.
3. A. R. Admati and P. Pfleiderer, "A theory of intraday trading patterns", *Review of Financial Studies* **1** (1988), 3–40.
4. V. Akgiray and G. G. Booth, "The stable-law model of stock returns", *Journal of Business and Economic Statistics* **6** (1988), 51–57.
5. P. S. Alexandrov and A. Ya. Khintchine, "Andrei Nikolaevich Kolmogorov", *Uspekhi Mat. Nauk* **8** (1953), no. 3, 178–200. (Russian)
6. L. Andersen, J. Andreasen, and R. Brotherton-Ratcliffe, *The Passport Option.* Preprint, Aarhus University, Aarhus, February, 1997.
7. T. Andersen and T. Bollerslev, *Intraday Seasonality and Volatility Persistence in Foreign Exchange and Equity Markets.* Working paper no. 193, Kellogg Graduate School of Management, Northwestern University, Evanston, IL, 1994.
8. A. A. Anis and E. H. Lloyd, "The expected value of the adjusted rescaled Hurst range of independent normal summands", *Biometrika* **63** (1976), no. 1, 111–116.
9. J.-P. Ansel and C. Stricker, "Couverture des actifs contingents et prix maximum", *Annales de l'Institut Henri Poincaré* **30** (1994), no. 2, 303–315.
10. J.-P. Ansel and C. Stricker, "Quelques remarques sur un théorème de Yan", *Lecture Notes in Mathematics* **1426** (1990), 226–274.
11. V. I. Arkin and I. V. Evstigneev, *Probabilistic models of control and economic dynamics*, Nauka, Moscow, 1979. (Russian)
12. L. Bachelier, "Théorie de la spéculation", *Annales de l'Ecole Normale Supérieure* **17** (1900), 21–86; English transl. in [78].
13. R. Baillie and T. Bollerslev, "Intra-day and inter market volatility in foreign exchange rates", *Review of Economic Studies* **58** (1990), 565–585.
14. R. Baillie and T. Bollerslev, "The daily message in exchange rates: a conditional variance tale", *Journal of Business and Economic Statistics* **7** (1989), 297–305.
15. R. T. Baillie, T. Bollerslev, and H.-O. Mikkelsen, "Fractionally integrated generalized autoregressive conditional heteroskedasticity", *Journal of Econometrics* **74** (1996), no. 1, 3–30.
16. G. L. Baker and J. P. Gollub, *Chaotic Dynamics*, Cambridge Univ. Press, Cambridge, 1996.

17. V. Balakrishnan, C. Nicolis, and G. Nicolis, *Extreme Value Distributions in Chaotic Dynamics*. Preprint ULB–Cenoli, no. 95-1, Centre for Nonlinear Phenomena and Complex Systems, Université Libre de Bruxelles, Bruxelles, January, 1995.
18. C. A. Ball, "A review of stochastic volatility models with application to option pricing", *Financial Markets, Institutions and Instruments* **2** (1993), no. 5, 55–69.
19. C. A. Ball and W. N. Torous, "Bond price dynamics and options", *Journal of Financial and Quantitative Analysis* **18** (1983), 517–531.
20. M. T. Barlow, "One-dimensional stochastic differential equations with no strong solution", *Journal of the London Mathematical Society* **26** (1982), 335–347.
21. O. E. Barndorff-Nielsen, "Exponentially decreasing distributions for the logarithm of particle size", *Proceedings of the Royal Society. London. Ser. A* **353** (1977), 401–419.
22. O. E. Barndorff-Nielsen, *Gaussian\\Inverse Gaussian Processes and the Modeling of Stock Returns*. Preprint, Aarhus University, Aarhus, October, 1994; see also O. E. Barndorff-Nielsen, "Processes of normal inverse Gaussian type", *Finance and Stochastics* **2** (1998), no. 1, 41–68.
23. O. E. Barndorff-Nielsen, "Hyperbolic distributions and distributions on hyperbolae", *Scandinavian Journal of Statistics* **5** (1978), 151–157.
24. O. E. Barndorff-Nielsen, "Normal inverse Gaussian distributions and stochastic volatility modelling", *Scandinavian Journal of Statistics* **24** (1996), 1–13.
25. O. E. Barndorff-Nielsen and P. Blæsild, "Hyperbolic distributions and ramifications: Contributions to theory and application", in *Statistical Distributions in Scientific Work* (ed. C. Taillie et al.), vol. 4, Reidel, Dordrecht, 1981, pp. 19–44.
26. O. E. Barndorff-Nielsen, J. L. Jensen, and M. Sørensen, "Wind shear and hyperbolic distributions", *Boundary-Layer Meteorology* **49** (1989), 417–431.
27. A. Barnea and D. Downes, "A reexamination of the empirical distribution of stock price changes", *Journal of American Statistical Association* **68** (1973), 348–350.
28. G. Barone-Adesi and R. Elliott, "Approximations for the values of American options", *Stochastic Analysis and Applications* **9** (1991), no. 2, 115–131.
29. G. Barone-Adesi and R. E. Whaley, "Efficient analytic approximation of American option values", *Journal of Finance* **42** (1987), no. 2, 301–320.
30. M. Basseville and I. Nikiforov, *Detection of Abrupt Changes*, Prentice Hall, Englewood Cliffs, NJ, 1993.
31. M. Baxter and A. Rennie, *Financial Calculus. An Introduction to Derivative Pricing*, Cambridge Univ. Press, Cambridge, 1996.
32. M. Beibel and H. R. Lerche, "A new look at optimal stopping problems related to mathematical finance", *Statistica Sinica* **7** (1997), no. 1, 93–108.
33. A. Bensoussan, "On the theory of option pricing", *Acta Applicandae Mathematicae* **2** (1984), 139–158.
34. A. Benveniste, M. Métivier, and P. Priouret, *Adaptive Algorithms and Stochastic Approximations*, Springer-Verlag, Berlin, 1987.
35. P. L. Bernstein, *Capital Ideas*, The Free Press, New York, 1992.
36. T. Björk, *Interest Rate Theory*. Working paper no. 72, School of Economics, Stockholm, 1996.
37. T. Björk, G. Di Masi, Yu. Kabanov, and W. Runggaldier, "Towards a general theory of bond markets", *Finance and Stochastics* **2** (1997), no. 1, 141–174.
38. T. Björk, Yu. Kabanov, and W. Runggaldier, "Bond market structure in the presence of marked point processes", *Mathematical Finance* **7** (1997), no. 2, 211–239.
39. P. Billingsley, *Convergence of Probability Measures*, Wiley, New York, 1968.

40. G. W. Bishop Jr., *Charles H. Dow and the Dow Theory*, Appleton-Century-Crofts, New York, 1960.
41. F. Black, "The holes in Black–Scholes", *RISK-magazin* (March, 1988).
42. F. T. Black, E. Derman, and W. Toy, "A one-factor model of interest rate and its application to Treasury bond options", *Financial Analysts Journal* (1990), 33–39.
43. F. Black and P. Karasinski, "Bond and option pricing when short rates are lognormal", *Financial Analysts Journal* (1991), 52–59.
44. F. Black and M. Scholes, "The pricing of options and corporate liabilities", *Journal of Political Economy* **81** (1973), no. 3, 637–659.
45. Yu. N. Blagoveshchenskij and I. L. Legostaeva, "Parametric bounds for a nonparametric trend", *Soviet Mathematics, Dokl.* **25** (1982), 714-718.
46. R. C. Blattberg and N. J. Gonedes, "A comparison of the stable and the Student distributions as statistical models for stock prices", *Journal of Business* **47** (1974), 244–280.
47. S. Bochner, "Subordination of non-Gaussian stochastic processes", *Proceedings of the National Academy of Sciences of the USA* **48** (1962), 19–22.
48. T. Bollerslev, "Generalized autoregressive conditional heteroscedasticity", *Journal of Econometrics* **31** (1986), 307–327.
49. A. Boness, A. Chen, and S. Jatusipitak, "Investigations of nonstationary prices", *Journal of Business* **47** (1974), 518–537.
50. A. N. Burenin, *Futures, Forward and Option Markets*, Trivola, Moscow, 1995. (Russian)
51. A. A. Borovkov, *Probability Theory*, Nauka, Moscow, 1986. (Russian)
52. N. L. Bowers Jr., H. U. Gerber, D. C. Hickman, D. A. Jones, and C. J. Nesbitt, *Actuarial Mathematics*, The Society of Actuaries, Itasca, IL, 1986.
53. G. E. P. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco, 1970.
54. A. Brace and M. Musiela, "A multifactor Gauss–Markov implementation of Heath, Jarrow, and Morton", *Mathematical Finance* **4** (1994), no. 3, 259–283.
55. R. A. Brealey and S. C. Myers, *Principles of Corporate Finance*, McGraw-Hill, New York, 1988.
56. M. Brennan and E. S. Schwartz, "A continuous time approach to the pricing of bonds", *Journal of Banking and Finance* **3** (1979), 133–155.
57. M. Brennan and E. S. Schwartz, "The valuation of American put options", *Journal of Finance* **32** (1977), 449–462.
58. W. A. Brock, W. D. Dechert, J.-A. Scheinkman, and B. Le Baron, "A test for independence based on the correlation dimension", *Econometrics Reviews* **15** (1996), no. 3, 197–235.
59. W. A. Brock, D. A. Hsieh, and B. Le Baron, *Nonlinear Dynamics, Chaos, and Instability: Statistical Theory and Economic Evidence*, MIT Press, Cambridge, MA, 1991.
60. W. A. Brock and S. M. Potter, "Diagnostic testing for nonlinearity, chaos, and general dependence in time-series data", in *Nonlinear Modeling and Forecasting (Proceedings of a Workshop on Nonlinear Modeling and Forecasting, September 1990, Santa Fe, New Mexico)* (eds. M. Casdagli and S. Eubank), Addison-Wesley, Redwood City, CA, 1990, pp. 137–159.
61. W. A. Brock and C. L. Sayers, "Is the business cycle characterized by deterministic chaos?", *Journal of Monetary Economics* **22** (1988), 71–90.

62. P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*, Springer-Verlag, New York, 1991.
63. V. Brousseau and M. O. Czarnecki, *Modelling Exchange Rates: the Stable Model.* Preprint, Ecole Polytechnique, Paris, August 16, 1994.
64. J. B. Burnham, "Current Structure and Recent Developments in Foreign Exchange Markets", in *Recent Developments in International Banking and Finance* (ed. S. J. Khonry), Elsevier Sci. Publ. North-Holland, Amsterdam, 1991, pp. 123–163.
65. *A Concise Dictionary of Business*, Market House Books, Ltd., 1991.
66. P. Carr, R. Jarrow, and R. Myneni, "Alternative characterizations of American put options", *Mathematical Finance* **2** (1992), no. 2, 87–106.
67. H. Chernoff, "Sequential tests for the mean of a normal distribution", in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, Univ. of California Press, 1961, pp. 79–91.
68. E. M. Chetyrkin, *Methods of financial and commercial pricing*, Business Rech Delo, Moscow, 1992. (Russian)
69. K. C. Chan, G. A. Karolyi, F. Longstaff, and A. B. Sanders, "An empirical comparison of alternative models of the short-term interest rates", *Journal of Finance* **47** (1992), no. 3, 1209–1227.
70. L. Chen, *A Three-Factor Model of the Term Structure of Interest Rates.* Preprint, Federal Reserve Board, Washington, USA, July, 1995.
71. P. Chen and R. H. Day (eds.), *Nonlinear Dynamics and Evolutionary Economics*, MIT Press, Cambridge, MA.
72. J. Y. Choi, D. Salandro, and K. Shastri, "On the estimation of bid-ask spreads: Theory and evidence", *Journal of Financial and Quantitative Analysis* **23** (1988), 219–230.
73. C. S. Chou, "Caractérisation d'une classe de semimartingales", *Lecture Notes in Mathematics* **721** (1979), 250–252.
74. C. S. Chou, P.-A. Meyer, and C. Stricker, "Sur les intégrales stochastiques de processus prévisibles non bornés", *Lecture Notes in Mathematics* **784** (1980), 128–139.
75. Y. S. Chow, H. Robbins, and D. Siegmund, *Great Expectations: The Theory of Optimal Stopping*, Houghton Mifflin Comp., Boston, 1971.
76. Z. Ciesielski, "Hölder conditions for realizations of Gaussian processes", *Transactions of the American Mathematical Society* **99** (1961), 403–413.
77. J. M. C. Clark, "The representation of functionals of Brownian motion by stochastic integrals", *Annals of Mathematical Statistics* **41** (1970), 1282–1295; Correction: *ibid* **42** (1971), 1778.
78. P. H. Cootner (ed.), *The Random Character of Stock Market Prices*, MIT Press, Cambridge, MA, 1964.
79. T. Copeland and J. Weston, *Financial Theory and Corporate Policy*, Addison-Wesley, Reading, MA, 1988.
80. J. C. Cox, J. E. Ingersoll Jr., and S. A. Ross, "An analysis of variable rate loan contracts", *Journal of Finance* **35** (1980), 389–403.
81. J. C. Cox, J. E. Ingersoll Jr., and S. A. Ross, "A theory of the term structure of interest rates", *Econometrica* **53** (1985), no. 2, 385–407.
82. J. C. Cox, R. A. Ross, and M. Rubinstein, "Option pricing: a simplified approach", *Journal of Financial Economics* **7** (1979), no. 3, 229–263.
83. J. C. Cox and M. Rubinstein, *Options Markets*, Prentice Hall, Englewood Cliffs, NJ, 1985.

84. A. Cowles, "Can stock market forecasters forecast?", *Econometrica* **1** (1933), 309–324.
85. A. Cowles, "Stock Market Forecasting", *Econometrica* **12** (1944), no. 3/4, 206–214.
86. A. Cowles and H. E. Jones, "Some a posteriori probabilities in stock market action", *Econometrica* **5** (1937), 280–294.
87. S. D. Cummins and H. Geman, "An Asian option approach to the valuation of insurance futures contracts", *Review of Futures Markets* **13** (1994), 517–557.
88. M. O. Czarnecki, *Modelisation des cours boursiers: le modèle stable (utilisation pratique et pricing d'option).* Preprint, Laboratoire d'Econométrie de l'Ecole Polytechnique, Paris, 24 juin 1994.
89. M. M. Dacorogna, U. A. Müller, P. Embrechts, and G. Samorodnitsky, *Moment Condition for the HARCH(k) Models.* Preprint, "Olsen & Associates", Zürich, May 30, 1995.
90. M. M. Dacorogna, U. A. Müller, R. J. Nagler, R. B. Olsen, and O. V. Pictet, "A geographical model for the daily and weekly seasonal volatility in the foreign exchange market", *Journal of International Money and Finance* **12** (1993), no. 4, 413–438.
91. M. M. Dacorogna, U. A. Müller, O. V. Pictet, and C. G. de Vries, *The Distribution of Extremal Foreign Exchange Rate Returns in Extremely Large Data Sets.* Preprint UAM, 1992-10-22, "Olsen & Associates", Research Institute for Applied Economics, Zürich, March 17, 1995.
92. R. C. Dalang, A. Morton, and W. Willinger, "Equivalent martingale measures and no-arbitrage in stochastic securities market models", *Stochastics and Stochastics Reports* **29** (1990), no. 2, 185–201.
93. R.-A. Dana and M. Jeanblanc-Picqué, *Marchés financiers en temps continu (Valorisation et équilibre)*, Economica, Paris, 1994.
94. H. David, H. Hartley, E. Pearson, "The distribution of the ratio, in a single normal sample, of range to standard deviation", *Biometrika* **41** (1954), 482-493.
95. R. H. Day, "Complex economic dynamics: obvious in history, generic in theory, elusive in data", [383], 1–15.
96. G. J. Deboeck (ed.), *Trading on the Edge: Neural, Genetic, and Fuzzy Systems for Chaotic Financial Markets*, Wiley, New York, 1994.
97. F. Delbaen, "Representing martingale measures when asset prices are continuous and bounded", *Mathematical Finance* **2** (1992), no. 2, 107–130.
98. F. Delbaen and W. Schachermayer, "Arbitrage and free lunch with bounded risk for unbounded continuous processes", *Mathematical Finance* **4** (1994), no. 4, 343–348.
99. F. Delbaen and W. Schachermayer, *A Compactness Principle for Bounded Sequences of Martingales with Applications.* Preprint, ETH, Zürich, November, 1996.
100. F. Delbaen and W. Schachermayer, "A general version of the fundamental theorem of asset pricing", *Mathematische Annalen* **300** (1994), no. 3, 463–520.
101. F. Delbaen and W. Schachermayer, *The Fundamental Theorem of Asset Pricing for Unbounded Stochastic Processes.* Preprint, ETH-Zentrum, Zürich, 1997.
102. C. Dellacherie and P.-A. Meyer, *Probabilités et potentiel. Ch. I à IV*, Hermann, Paris, 1975.
103. C. Dellacherie and P.-A. Meyer, *Probabilités et potentiel. Ch. V à VIII*, Hermann, Paris, 1980.
104. R. L. Devaney, *An Introduction to Chaotic Dynamical Systems*, The Benjamin/Cummings Publ. Co., Menlo Park, CA, 1986.

105. C. G. De Vries, "On the relation between *GARCH* and stable processes", *Journal of Econometrics* **48** (1991), no. 3, 313–324.
106. Z. Ding, C. W. J. Granger, and R. F. Engle, "A long memory property of stock market returns and a new model", *Journal of Empirical Finance* **1** (1993), 83–106.
107. A. K. Dixit and R. S. Pindyck, *Investment under Uncertainty*, Princeton Univ. Press, Princeton, 1994.
108. C. D. Campbell, R. G. Campbell, and E. G. Dolan, *Money, Banking and Monetary Policy*, The Dryden Press, London, 1988.
109. J. L. Doob, *Stochastic Processes*, Wiley, New York, 1953.
110. J. L. Doob, *Classical Potential Theory and Its Probabilistic Counterpart*, Springer-Verlag, New York, 1984.
111. L. Dothan, "On the term structure of interest rates", *Journal of Financial Economics* **6** (1978), 59–69.
112. M. U. Dothan, *Prices in Financial Markets*, Oxford Univ. Press, Oxford, 1990.
113. R. Douady, *Cylindrical Brownian Motions and Yield Curve Smoothing*. Preprint, New York University, Courant Institute of Mathematical Sciences, New York, February, 1996.
114. F. C. Drost and T. E. Nijman, "Temporal aggregation of *GARCH* processes", *Econometrica* **61** (1993), no. 4, 909–927.
115. F. C. Drost and B. J. M. Werker, "Closing the *GARCH* gap: Continuous time *GARCH* modeling", *Journal of Econometrics* **74** (1996), no. 1, 31–57.
116. L. E. Dubins, L. A. Shepp, and A. N. Shiryaev, "Optimal stopping rules and maximal inequalities for Bessel processes", *Theory of Probability and its Applications* **38** (1993), no. 2, 226–261.
117. D. Duffie, *Dynamic Asset Pricing Theory*, Princeton Univ. Press, Princeton, 1992.
118. J. D. Duffie and J. M. Harrison, "Arbitrage pricing of a Russian option and perpetual lookback options", *The Annals of Applied Probability* **3** (1993), no. 3, 641–651.
119. D. Duffie and R. Kan, "A yield-factor model of interest rates", *Mathematical Finance* **6** (1996), no. 4, 379–406.
120. W. Du Mouchel, "Stable distributions in statistical inference. I: Symmetric stable distribution compared to other symmetric long-tailed distributions", *Journal of American Statistical Association* **68** (1973), 469–477.
121. B. Dupire, "Model Art", *RISK-magazin* **6** (1993), 118–124.
122. B. Dupire, "Pricing with a smile", *RISK-magazin* **7** (1994), no. 1, 18–20.
123. R. Durrett, *Brownian Motion and Martingales in Analysis*, Wadsworth, Belmont, CA, 1984.
124. R. Durrett, *Probability: Theory and Examples*, Duxbury Press, Belmont, CA, 1995.
125. P. Dybvig and S. Ross, "Arbitrage", in *The New Palgrave: A Dictionary of Economics* (eds. J. Eatwell, M. Milgate, and P. Newman), vol. 1, Macmillan, London, pp. 100–106.
126. E. B. Dynkin, *Markov Processes*, Fismatgiz, Moscow, 1963. (Russian)
127. E. Eberlein and U. Keller, "Hyperbolic distributions in finance", *Bernoulli* **1** (1995), no. 3, 281–299.
128. E. Eberlein and S. Raible, *Term Structure Models Driven by General Lévy Processes*. Preprint no. 33, Universität Freiburg, Institut für Mathematische Stochastik, Freiburg i. Br., 1996, pp. 1–25.
129. L. H. Ederington and J. H. Lee, "How markets process information: News releases and volatility", *Journal of Finance* **48** (1993), 1161–1191.

130. R. D. Edwards and J. Magee, *Technical Analysis of Stock Trends*, Springfield, MA, 1958.
131. A. Einstein, *Investigation on the Theory of the Brownian Movement* (ed. R. Fürth), Dover, New York, 1956.
132. A. Einstein, "On the movement of small particles suspended in a stationary liquid demanded by the molecular-kinetic theory of heat", *Annalen der Physik* **17** (1905), 549–560.
133. N. El Karoui and H. Geman, "A stochastic approach to the pricing of FRN's", *RISK-magazin* **4** (1991).
134. N. El Karoui and I. Karatzas, "A new approach to the Skorokhod problem and its applications", *Stochastics and Stochastics Reports* **34** (1991), no. 1/2, 57–82; Correction: *ibid* **36** (1991), no. 3/4, 265.
135. N. El Karoui, R. Myneni, and R. Viswanathan, *The Probabilistic Theory of the American Option*. Preprint, Université de Paris VI, Paris, 1991.
136. N. El Karoui and M. C. Quenez, "Dynamic programming and pricing of contingent claims in an incomplete market", *SIAM Journal on Control and Optimization* **33** (1995), no. 1, 29–66.
137. M. Emery, "Compensation de processus V.F. non localement intégrables", *Lecture Notes in Mathematics* **784** (1980), 152–160.
138. M. Emery, "Metrisabilité de quelques espaces de processus aléatoires", *Lecture Notes in Mathematics* **784** (1980), 140–147.
139. M. Emery, *Stochastic Calculus in Manifolds (With an Appendix: A Short Presentation of Stochastic Calculus by P.-A. Meyer)*, Springer-Verlag, Berlin, 1989.
140. R. F. Engle, "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation", *Econometrica* **50** (1982), no. 4, 987–1008.
141. R. F. Engle and T. Bollerslev, "Modelling the persistence of conditional variances", *Econometrics Reviews* **5** (1986), 1–50.
142. R. F. Engle, T. Ito, and Lin Wen-Ling, "Meteor showers or heat waves? Heteroskedastic intra-daily volatility in the foreign exchange market", *Econometrica* **58** (1990), 525–542.
143. R. F. Engle and J. R. Russell, "Forecasting transaction rates: The autoregressive conditional duration model", [393] **4**.
144. F. Esscher, "On the probability function in the collective theory of risk", *Skandinavisk Aktuarietidskrift* **15** (1932), 175–195.
145. C. J. G. Evertsz, "Self-similarity of high-frequency USD/DEM exchange rates", [393] **3**.
146. C. J. G. Evertsz and K. Berkner, "Large deviation and self-similarity analysis of curves: DAX stock prices", *Chaos, Solutions and Fractals* **6** (1995), 121–130.
147. E. F. Fama, "Efficient capital markets: A review of theory and empirical work", *Journal of Finance* **25** (1970), 383–417.
148. E. F. Fama, "Portfolio analysis in a stable Paretian market", *Management Science* **311** (1965), no. 2, 409–419.
149. E. F. Fama, "Risk, return and equilibrium", *Journal of Political Economy* **79** (1971), 30–55.
150. E. F. Fama, "The behavior of stock market prices", *Journal of Business* **34** (1965), 420–429.
151. E. F. Fama and M. H. Miller, *The Theory of Finance*, Holt, Rinehard and Winston, Inc., New York, 1972.

152. E. F. Fama and R. Roll, "Parameter estimates for symmetric stable distributions", *Journal of American Statistical Association* **66** (1971), 331–338.
153. E. F. Fama and R. Roll, "Some properties of symmetric stable distributions", *Journal of American Statistical Association* **63** (1968), 817–836.
154. F. Family and T. Vicsek (eds.), *Dynamics of Fractal Surfaces*, World Scientific Publ., Singapore, 1991.
155. J. Feder, *Fractals*, Plenum Press, New York, 1988.
156. W. Feller, *An Introduction to Probability Theory and its Applications*, vol. 2, Wiley, New York, 1966.
157. W. Feller, "The asymptotic distribution of the range of sums of independent random variables", *Annals of Mathematical Statistics* **22** (1951), no. 3, 427–432.
158. B. D. Fielitz and J. P. Rozelle, "Stable distributions and the mixtures of distributions hypotheses for common stock returns", *Journal of American Statistical Association* **78** (1983), no. 381, 28–36.
159. I. Fisher, *The Theory of Interests*, Macmillan, New York, 1930.
160. M. D. Flood, "Market structure and inefficiency in the foreign exchange market", *Journal of International Money and Finance* **13** (1994), no. 2, 131–138.
161. A. D. Fokker, "Die mittlere Energie rotierender Dipole im Strahlungsfeld", *Annalen der Physik* **43** (1914), 810–820.
162. H. Föllmer, *Probabilistic Aspects of Options*. Preprint, Helsinki University, Helsinki, January, 1990.
163. H. Föllmer and Yu. M. Kabanov, "Optional decomposition and lagrange multipliers", *Finance and Stochastics* **2** (1998), no. 1, 69–81.
164. H. Föllmer and Yu. M. Kabanov, *Optional Decomposition Theorems in Discrete Time*. Preprint, Humboldt University, Berlin, 1996.
165. H. Föllmer and D. Kramkov, "Optional decompositions under constraints", *Probability Theory and Related Fields* **109** (1997), no. 1, 1–25.
166. H. Föllmer, Ph. Protter, and A. N. Shiryaev, "Quadratic covariation and an extention of Itô's formula", *Bernoulli* **1** (1995), no. 1/2, 149–170.
167. H. Föllmer and M. Schweizer, "Hedging of contingent claims under incomplete information", in *Applied Stochastic Analysis (Stochastic Monographs, vol. 5)* (eds. M. H. A. Davis and R. J. Elliott), Gordon and Breach, London, 1991, pp. 389–414.
168. H. Föllmer and D. Sondermann, "Hedging of non-redundant contingent claims", in *Contributions to Mathematical Economics* (eds. A. Mas-Colell and W. Hildenbrand), North-Holland, Amsterdam, 1986, pp. 205–223.
169. K. R. French, G. W. Schwert, and R. Stambaugh, "Expected stock returns and volatility", *Journal of Financial Economics* **19** (1987), 3–29.
170. A. Friedman, *Stochastic Differential Equations and Applications*, vol. 1, 2, Academic Press, New York, 1975, 1976.
171. M. Frittelli and P. Lakner, "Arbitrage and free lunch in a general financial market model; the fundamental theorem of asset pricing", [336], 89–92.
172. L. I. Galtchouk, "Structure of some martingales", in *Proceedings of the School-seminairein Druskininkai*, vol. 1, 1974.
173. B. Gamrowski and S. T. Rachev, *Modèles Financiers Utilisant les Lois Stables*. Preprint.
174. A. O. Gel'fond, *Calculus of Finite Differences*, Authorized English translation of the 3rd Russian edition, Hindustan Publ. Corp., Delhi, India, 1971.

175. H. Geman, N. El Karoui, and J. C. Rochet, "Changes of numéraire, changes of probability measure and option pricing", *Journal of Applied Probability* **32** (1995), 443–458.
176. T. J. George, G. Kaul, and M. Nimalendran, "Estimation of the bid-ask spread and its components: a new approach", *Review of Financial Studies* **4** (1991), 623–656.
177. H. U. Gerber and E. S. W. Shiu, "Martingale approach to pricing American options", *ASTIN Bulletin* **24** (1994), 195–200.
178. H. U. Gerber and E. S. W. Shiu, "Option pricing by Esscher transforms", *Transactions of the Society of Actuaries* **46** (1994), 99–191.
179. R. Geske and H. E. Johnson, "The American put options valued analytically", *Journal of Finance* **39** (1984), 1511–1524.
180. S. Ghashghaie, W. Breymann, J. Peinke, P. Talkner, and Y. Dodge, "Turbulent cascades in foreign exchange markets", *Nature* **381** (1996. 27 June), 767–770.
181. E. Ghysels and J. Jasiak, "Trading patterns: Time deformation and stochastic volatility in foreign exchange markets", [393] **1**.
182. I. I. Gihman and A. V. Skorohod, *Stochastic Differential Equations*, Springer-Verlag, Berlin, 1972.
183. I. V. Girsanov, "On transforming a certain class of stochastic processes by absolutely continuous substitution of measures", *Theory of Probability and its Applications* **5** (1962), no. 3, 285–301.
184. L. Glosten and L. Harris, "Estimating the components of the bid-ask spread", *Journal of Financial Economics* **21** (1988), 123–142.
185. L. Glosten and P. Milgrom, "Bid, ask, and transaction prices in a special market with heterogeneously informed traders", *Journal of Financial Economics* **14** (1985 (March)), 21–42.
186. B. V. Gnedenko, *The Theory of Probability*, Mir Publ., Moscow, 1988.
187. B. V. Gnedenko, "Sur la distribution limite du terme maximum d'une série aléatoire", *Annals of Mathematics* **44** (1943), 423–453.
188. B. V. Gnedenko and A. N. Kolmogorov, *Limit Distributions for Sums of Independent Random Variables*, Addison-Wesley Publ. Comp., Cambridge, 1954.
189. C. A. Goodhart, "'News' and the foreign exchange market", *Proc. Manchester Statist. Soc.* (1989), 1–79.
190. C. A. Goodhart and A. Demos, "Reuters screen images of the foreign exchange market: the deutschenmark/doller spot rate", *Journal of International Securities Markets* **4** (1990), 333–348.
191. C. A. Goodhart and L. Figliuoli, "Every minute counts in financial markets", *Journal of International Money and Finance* **10** (1991), 23–52.
192. C. A. E. Goodhart and M. O'Hara, "High frequency data in financial markets: issues and applications", [393], Introductory Lecture.
193. Ch. Gouriéroux, *Modèles ARCH et applications financières*, Economica, Paris, 1992.
194. C. Gouriéroux and J.-P. Laurent, *Dynamic Hedging in Discrete Time.* Preprint, CREST, Paris, 1995.
195. C. Gouriéroux, J.-P. Laurent, and H. Pham, *Quadratic Hedging and Numéraire.* Preprint, Université de Marne-la-Vallée, Equipe d'Analyse et de Mathématiques Appliquées, October, 1995.
196. C. W. J. Grangel and P. Newbold, *Forecasting Economic Time Series*, Academic Press, New York, 1977.

197. C. W. J. Granger and O. Morgenstern, *Predictability of Stock Market Prices*, D. C. Heath & Co., Lexington, MA, 1970.
198. C. W. J. Granger and T. Teräsvirta, *Modelling Nonlinear Economic Relationships*, Oxford Univ. Press, Oxford, 1993.
199. P. Grassberger and I. Procaccia, "Measuring the strangeness of strange attractors", *Physica D* **9** (1983), 189–208.
200. B. I. Grigelionis and A. N. Shiryaev, "On Stefan's problem and optimal stopping rules for Markov processes", *Theory of Probability and its Applications* **11** (1966), no. 4, 541–558.
201. G. R. Grimmet and D. R. Stirzaker, *Probability and Random Processes*, Clarendon Press, Oxford, 1982.
202. D. Guégan, *Séries chronologiques non linéaires à temps discret*, Economica, Paris, 1994.
203. D. M. Guillaume, M. M. Dacorogna, R. R. Davé, U. A. Müller, R. B. Olsen, O. V. Hamon, and B. Jacquillat, *Le marché français des actions*, Presses Universitaires de France, Paris, 1992.
204. D. M. Guillaume, M. M. Dacorogna, R. D. Davé, U. A. Müller, R. B. Olsen, and O. V. Pictet, "From the bird's eye to the microscope: A survey of new stylized facts of the intra-daily foreign exchange markets", *Finance and Stochastics* **1** (1997), no. 2, 95–129.
205. D. M. Guillaume, O. V. Pictet, and M. M. Dacorogna, "On the intra-day performance of *GARCH* processes", [393] **3**.
206. E. J. Gumbel, *Statistics of Extremes*, Columbia Univ. Press, New York, 1960.
207. R. Hagerman, "More evidence on the distribution of security returns", *Journal of Finance* **33** (1978), 1213–1221.
208. J. Hale and H. Kogak, *Dynamics and Bifurcations*, Springer-Verlag, New York, 1991.
209. C. Halgreen, "Self-decomposability of the generalized inverse Gaussian and hyperbolic distributions", *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **47** (1979), 13–17.
210. P. Hall, "On some simple estimates of an exponent of regular variation", *Journal of the Royal Statistical Society. Ser. B* **44** (1982), no. 1, 37–42.
211. J. D. Hamilton, *Time Series Analysis*, Princeton Univ. Press, Princeton, NJ, 1994.
212. H. J. Hannan, *Multiple Time Series*, Wiley, New York, 1970.
213. A. T. Hansen, *Complete Market Pricing in the Wiener Filtration without Existence of a Martingale Measure*. Preprint, Aarhus University, Dept. of Operation Research, Aarhus, 1996.
214. J. M. Harrison and D. M. Kreps, "Martingales and arbitrage in multiperiod securities markets", *Journal of Economic Theory* **20** (1979), 381–408.
215. J. M. Harrison and S. R. Pliska, "Martingales and stochastic integrals in the theory of continuous trading", *Stochastic Processes and their Applications* **11** (1981), no. 3, 215–260.
216. C. R. Harvey and R. D. Huang, "Volatility in the foreign currency futures market", *Review of Financial Studies* **4** (1991), 543–569.
217. J. Hasbrouck, "Trades, Quotes, Inventories and Information", *Journal of Financial Economics* **22** (1988), 229–252.
218. J. A. Hausman, A. W. Lo, and A. C. McKinlay, *An Odered Probit Analysis of Transaction Stock Prices*. Working paper no. 3888, National Bureau of Economic Research, 1991.

219. D. Heath, R. Jarrow, and A. Morton, "Bond pricing and the term structure of interest rates", *Econometrica* **60** (1992), no. 1, 77–106.
220. S. I. Heston, "A closed forms solution for options with stochastic volatility with applications to bond and currency options", *Review of Financial Studies* **6** (1992), no. 2, 333–343.
221. *High Frequency Data in Finance*, Data set from "Olsen & Associates", 1993. E-mail: hfdf@olsen.ch.
222. R. C. Hilborn, *Chaos and Nonlinear Dynamics*, Oxford Univ. Press, Oxford, 1994.
223. B. M. Hill, "A simple general approach to inference about the tail of a distribution", *The Annals of Statistics* **5** (1975), no. 3, 1163–1173.
224. T. Ho and S. Lee, "Term structure movements and pricing interest rates contingent claims", *Journal of Finance* **41** (1986), 1011–1029.
225. J. Hoffman-Jørgensen, "Stable densities", *Theory of Probability and its Applications* **38** (1993), no. 2, 350–355.
226. N. Hofmann, E. Platen, and M. Schweizer, "Option pricing under incompleteness and stochastic volatility", *Mathematical Finance* **2** (1992), 153–187.
227. K. C. Hogan Jr. and M. Melvin, "Sources of meteor showers and heat waves in the foreign exchange market", *Journal of International Economics* **37** (1994), 239–247.
228. J. C. van Horne, *Financial Market Rates and Flows*, Prentice Hall, Englewood Cliffs, NJ, 1984.
229. Der-Ann Hsu, R. B. Miller, and D. W. Wichern, "On the stable Paretian behavior of stock-market prices", *Journal of American Statistical Association* **69** (1974), no. 345, 108–113.
230. R. D. Huang and R. W. Masulis, "Spreads, dealer competition, and market regimes: a market microstructure analysis or FX trading", [393] **4**.
231. G. Huberman, "A simple approach to arbitrage pricing theory". *Journal of Economic Theory* (March, 1982), 183–191.
232. M. Hudson, "The value in going out", in *From Black–Scholes to black holes*, Risk/ Finex, London/New York, 1992, pp. 183–186.
233. J. C. Hull, *Options, Futures, and Other Derivative Securities*, Prentice Hall, Englewood Cliffs, NJ, 1993.
234. J. Hull and A. White, "Pricing interest rate derivative securities", *Review of Financial Studies* **3** (1990), no. 5, 573–592.
235. J. Hull and A. White, "The pricing of options on assets with stochastic volatilities", *Journal of Finance* **42** (1987), 281–308.
236. H. Hurst, "Long-term storage capacity of reservoirs", *Transactions of American Society of Civil Engineers* **116** (1951), 770–808.
237. R. G. Ibbotson and R. A. Sinquefield, *Stocks, Bonds, Bills and Inflation: 1986 Year Book*, Ibbotson & Associates, Chicago, 1986.
238. I. A. Ibragimov and K. E. Cernin, "On the unimodality of stable laws", *Theory of Probability and its Applications* **4** (1961), no. 4, 417–419.
239. N. Ikeda and S. Watanabe, *Stochastic Differential Equations and Diffusion Processes*, North-Holland, Amsterdam, 1989.
240. J. E. Ingersoll, *Theory of Financial Decision Making*, Rowman and Littlefield, London, 1987.
241. A. D. Ioffe and V. M. Tikhomirov, *Theorie der Extremalaufgaben*, VEB Deutscher Verlag der Wissenschaften, Berlin, 1979.

242. K. Itô, "On a stochastic integral equation", *Japan Academy. Proceedings* **22** (1946), 32–35.
243. K. Itô, "On stochastic differential equations", *Memoirs of the American Mathematical Society* **4** (1951), 1–89.
244. K. Itô, "Stochastic integral", *Imperial Academy. Tokyo. Proceedings* **20** (1944), 519–524.
245. K. Itô and H. P. McKean, *Diffusion Processes and Their Sample Paths*, Springer-Verlag, New York, 1965.
246. T. Ito, R. F. Engle, and Lin Wen-Ling, "Where does the meteor shower come from? The role of stochastic policy coordination", *Journal of International Economics* **32** (1992), 221–240.
247. S. D. Jacka, "Optimal stopping and the American put", *Mathematical Finance* **1** (1991), no. 2, 1–14.
248. J. Jacod, "Calcul stochastique et problèmes de martingales", *Lecture Notes in Mathematics* **714** (1979), 1–539.
249. J. Jacod, "Intégrales stochastiques par rapport à une semimartingale vectorielle et changements de filtration", *Lecture Notes in Mathematics* **784** (1980), 161–172.
250. J. Jacod and A. N. Shiryaev, *Limit Theorems for Stochastic Processes*, Springer-Verlag, Berlin, 1987.
251. J. Jacod and A. N. Shiryaev, "Local martingales and the fundamental asset pricing theorems in the discrete-time case", *Finance and Stochastics* **2** (1998), no. 3, 259–273.
252. E. Jacquier, N. G. Polson, and P. E. Rossi, "Bayesian analysis of stochastic volatility models", *Journal of Business and Economic Statistics* **12** (1994), no. 4, 371–417.
253. A. Janicki and A. Weron, *Simulation and Chaotic Behavior of α-Stable Stochastic Processes*, M. Dekker, New York, 1994.
254. A. Jakubowski, J. Mémin, and G. Pages, "Convergence en loi des suites d'intégrales stochastiques sur l'espace $\mathbb{D}^1$ de Skorokhod", *Probability Theory and Related Fields* **81** (1989), no. 1, 111–137.
255. B. A. Jensen and J. A. Nielsen, *The Structure of Binomial Lattice Models for Bonds*. Working paper no. 91.1, Copenhagen Business School, Aarhus University, 1991.
256. F. Jamshidian, "An exact bond option formula", *Journal of Finance* **44** (1989), no. 1, 205–209.
257. P. L. Jørgensen, *American Option Pricing*. Preprint, The Faculty of Business Administration, The Aarhus School of Business, 1994.
258. P. Kabaila, "On the asymptotic efficiency of estimators of the parameters of *ARMA* processes", *Journal of Time Series Analysis* **4** (1983), 37–49.
259. Yu. M. Kabanov and D. O. Kramkov, "No-arbitrage and equivalent martingale measures: An elementary proof of the Harrison–Pliska theorem", *Theory of Probability and its Applications* **39** (1994), no. 3, 523–527.
260. Yu. M. Kabanov and D. O. Kramkov, "Asymptotic arbitrage in large financial markets", *Finance and Stochastics* **2** (1998), no. 2, 143–172.
261. Yu. M. Kabanov and D. O. Kramkov, "Large financial markets: Asymptotic arbitrage and contiguity", *Theory of Probability and its Applications* **39** (1994), no. 1, 182–187.
262. J. Kallsen and M. S. Taqqu, *Option Pricing in ARCH-type Models*. Preprint, July, 1994.
263. E. Kamke, *Differentialgleichungen. Loesungsmethoden und Lösungen*, Akad. Verlagsgesellschaft Becker & Eler Kom.-Ges., Leipzig, 1943.

264. M. Kanter, "Stable densities under change of scale and total variation inequalities", *The Annals of Probability* **3** (1975), no. 4, 697–707.
265. I. Karatzas, "On the pricing of American options", *Applied Mathematics and Optimization* **17** (1988), 37–60.
266. I. Karatzas and S. E. Shreve, *Brownian Motion and Stochastic Calculus*, Springer-Verlag, Berlin, 1988.
267. I. Karatzas and S. Shreve, *Methods of Mathematical Finance*, Columbia Univ. Press, New York, 1995.
268. T. Kariya, *Quantitative Methods for Portfolio Analysis*, Kluwer Acad. Publ., Dordrecht, 1993.
269. M. G. Kendall, "The analysis of economic time-series. Part 1. Prices", *Journal of the Royal Statistical Society* **96** (1953), 11–25.
270. M. Kendall, *Time-Series*, Charles Griffin, London, 1973.
271. M. Kendall and A. Stuart, *The Advanced Theory of Statistics*, vol. 3, Charles Griffin, London, 1966.
272. I. J. Kim, "The analytic valuation of American options", *Review of Financial Studies* **3** (1990), 547–572.
273. Klein I., Schachermayer W., "Asymptotic arbitrage in noncomplete large financial markets", *Theory of Probability and its Applications* **41** (1996), no. 4, 780–788.
274. C. C. Klimasauskas, "Neural Network Techniques", [96].
275. P. E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*, Springer-Verlag, Berlin, 1992.
276. A. N. Kolmogorov, "The local structure of turbulence in incompressible viscous fluid for very large Reynold's numbers", *Comptes Rendus (Dokl.) de l'Académie des Sciences de l'URSS* **30** (1941), 301–305.
277. A. N. Kolmogorov, *Selected Works. Vol. I: Mathematics and Mechanics* (ed. V. M. Tikhomirov), Kluwer Acad. Publ., Dordrecht, 1991.
278. A. N. Kolmogorov, *Selected Works. Vol. II: Probability Theory and Mathematical Statistics* (ed. A. N. Shiryaev), Kluwer Acad. Publ., Dordrecht, 1992.
279. A. N. Kolmogoroff, "Wienersche Spiralen und einige andere interessante Kurven im Hilbertschen Raum", *Comptes Rendus (Dokl.) de l'Académie des Sciences de l'URSS* **26** (1940), no. 2, 115–118.
280. A. N. Kolmogoroff, "Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung", *Mathematische Annalen* **104** (1931), 415–458. (See also [278].)
281. D. O. Kramkov, "Optional decomposition of supermartingales and hedging contingent claims in incomplete security markets", *Probability Theory and Related Fields* **105** (1996), no. 4, 459–479.
282. D. O. Kramkov and A. N. Shiryaev, *On Sufficient Conditions of Uniform Integrability of the Exponential Martingales.* Preprint, Steklov Mathematical Institute, Moscow, 1996; to appear in *Proceedings of the Second Congress of the European Mathematical Society (Budapest, 1996).*
283. D. O. Kramkov and A. N. Shiryaev, "On the rational pricing of the 'Russian option' for the symmetrical binomial model of a (B, S)-market", *Theory of Probability and its Applications* **39** (1994), no. 1, 153–162.
284. C. G. Krouse, *Capital Markets and Prices. Valuing Uncertain Income Streams*, North-Holland, Amsterdam, 1986.
285. D. Kreps, "Arbitrage and equilibrium in economies with infinitely many commodities", *Journal of Mathematical Economics* **8** (1981), 15–35.

286. P. Krugman, "Target zones and exchange rate dynamics", *Quarterly Journal of Economics* **106** (1991), no. 3, 669–682.
287. N. V. Krylov, *Controlled Processes of Diffusion Type*, Nauka, Moscow, 1977. (Russian)
288. N. V. Krylov, *Introduction to the Theory of Diffusion Processes*, Amer. Math. Soc., Providence, RI, 1995.
289. U. Küchler, K. Neumann, M. Sørensen, and A. Streller, *Stock Returns and Hyperbolic Distributions.* Discussion paper no. 23, Humboldt University, Berlin, 1994.
290. T. L. Lai, "Sequential change-point detection in quality control and dynamic systems", *Journal of the Royal Statistical Society. Ser. B* **57** (1995), no. 4, 613–658.
291. T. L. Lai and D. Siegmund, "Fixed accuracy estimation of an autoregressive parameter", *The Annals of Statistics* **11** (1983), no. 2, 478–485.
292. P. Lakner, "Martingale measure for a class of right-continuous processes", *Mathematical Finance* **3** (1993), no. 1, 43–53.
293. D. Lamberton and B. Lapeyre, *Introduction to Stochastic Calculus Applied to Finance*, Chapman & Hall, London, 1996.
294. O. Lanford, "A computer-assisted proof of the Feigenbaum conjectures", *American Mathematical Society. Bulletin* **6** (1982), no. 3, 427–434.
295. P. Langevin, "Sur la théorie du mouvement brownien", *Comptes Rendus de l'Académie des Sciences. Paris* **146** (1908), 530–533.
296. *Latin Dictionary (founded on Andrew's edition of Freund's latin dictionary revised, enlarged and in great part rewritten by Ch. T. Lewis, Ch. Short)*, Oxford Univ. Press, Oxford, 1966.
297. I. L. Legostaeva and A. N. Sirjaev, "Minimax weights in a trend detection problem of a random process", *Theory of Probability and its Applications* **16** (1971), no. 2, 344–349.
298. P. Lévy, *Processus stochastiques et mouvement brownien*, Gauthier-Villars, Paris, 1948.
299. S. J. Lin, "Stochastic analysis of fractional Brownian motions", *Stochastics and Stochastics Reports* **55** (1995), no. 1+2, 121–140.
300. D. V. Lindley, "Dynamic programming and decision theory", *Applied Statistics* **10** (1961), no. 1, 39–51.
301. J. Lintner, "The valuation of risky assets and the selection of risky investments on stock portfolios and capital budgets", *Review of Economics and Statistics* **47** (February, 1965), 13–34.
302. O. Linton, "Adaptive estimation in *ARCH* Models", *Econometric Theory* **9** (1993), 539–569.
303. R. S. Liptser and A. N. Shiryaev, *Statistics of random processes. I. General theory; II. Applications*, Springer-Verlag, New York, 1977, 1978.
304. R. S. Liptser and A. N. Shiryaev, *Theory of Martingales*, Kluwer Acad. Publ., Dordrecht, 1989.
305. T. Liu, C. W. J. Granger, and W. P. Heller, "Using the correlation exponent to decide whether an economic series is chaotic", [383].
306. A. Lo, "Neural networks and other nonparametric techniques in economics and finance", in *Blending Quantitative and Traditional Equity Analysis*, Association for Investment Management and Research, 1994, pp. 25–36.

307. P. R. Locke and C. L. Sayers, "Intra-day futures price volatility: information effects and variance persistence", *Journal of Applied Econometrics* **8** (1993), 15–30. (See also [383], 213–228.)
308. L. J. Lockwood and S. C. Linn, "An examination of stock market return volatility during overnight and intraday periods, 1964–1989", *Journal of Finance* **45** (1990), 591–601.
309. A. M. Loosignian, *Foreign Exchange Futures*, Dow Jones-Irwin, Homewood, IL, 1981.
310. A. M. Loosignian, *Stock Index Futures (Buying and Selling the Market Averages)*, Addison-Wesley, Reading, MA, 1985.
311. H.-W. Lorenz, *Nonlinear Dynamic Economics and Chaotic Motion*, Lecture Notes in Economics and Mathematical Systems, vol. 334, Springer-Verlag, New York, 1989.
312. B. Malkiel, *A Random Walk down Wall Street*, W. W. Norton & Co., London, 1991.
313. A. G. Malliaris and W. A. Brock, *Stochastic Methods in Economics and Finance*, North-Holland, Amsterdam, 1982.
314. B. B. Mandelbrot, *Fractals: Form, Chance, and Dimension*, W. H. Freeman & Co., San Francisco, 1977.
315. B. B. Mandelbrot, *Les Objets Fractals*, Flammarion, Paris, 1975.
316. B. B. Mandelbrot, "Limit theorems on the self-normalized range for weakly and strongly dependent process", *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **31** (1975), 271–285.
317. B. B. Mandelbrot, "Robustness of the rescaled range R/S in the measurement of noncyclic long-run statistical dependence", *Water Resources Research* **5** (1969), no. 5, 967–988.
318. B. B. Mandelbrot, "Some noises with $1/f$ spectrum: a bridge between direct current and white noise", *IEEE Transactions on Information Theory* (April, 1967).
319. B. B. Mandelbrot, "Statistical methodology for non-periodic cycles: from the covariance to R/S analysis", *Annals of Economic and Social Measurement* **1** (1972), no. 3, 259–290.
320. B. B. Mandelbrot, *The Fractal Geometry of Nature*, W. H. Freeman & Co., San Francisco, 1982.
321. B. B. Mandelbrot, "The Pareto–Lévy law and the distribution of income", *International Economic Review* (1960), no. 1.
322. B. B. Mandelbrot, "The stable Paretian income distribution when the apparent exponent is near two", *International Economic Review* (1963), no. 4.
323. B. B. Mandelbrot, "The variation of certain speculative prices", *Journal of Business* **36** (1963), 394–419.
324. B. B. Mandelbrot, "The variation of some other speculative prices", *Journal of Business* **40** (1967), 393–413.
325. B. B. Mandelbrot, "Une classe de processus stochastiques homothétiques à soi: application à la loi climatologique de H. E. Hurst", *Comptes Rendus de l'Académie des Sciences, Paris* **240** (1965), 3274–3277.
326. B. B. Mandelbrot, "When can price be arbitrated efficiently? A limit of the validity of the random walk and martingale models", *Review of Economics and Statistics*, **53** (1971), 225–236.
327. B. B. Mandelbrot and H. M. Taylor, "On the distribution of stock price difference", *Operations Research* **15** (1967), no. 6, 1057–1062.
328. B. B. Mandelbrot and J. W. van Ness, "Fractional Brownian motions, fractional noises and applications", *SIAM Review* **10** (1968), no. 4, 422–437.

329. B. B. Mandelbrot and J. R. Wallis, "Computer experiments with fractional Gaussian noises. I, II, III", *Water Resources Research* **5** (1969), 228–267.
330. R. N. Mantegna and H. E. Stanley, "Scaling behaviour in the dynamics of an economic index", *Nature* **376** (1995), 46–49.
331. H. Markowitz, *Mean-Variance Analysis in Portfolio Choice and Capital Markets*, Blackwell, Cambridge, MA, 1990.
332. H. Markowitz, "Portfolio selection", *Journal of Finance* **7** (March, 1952), 77–91.
333. H. Markowitz, *Portfolio Selection. Efficient Diversification of Investments*, Wiley, New York, 1959.
334. J. D. Martin, S. H. Cox Jr., and R. D. McMinn, *The Theory of Finance. Evidence and Applications*, The Dryden Press, London, 1988.
335. I. M. Vinogradov, V. I. Bityutskov, and Yu. V. Prokhorov (eds.), *Mathematical Encyclopedia*, vol. 1–5, Sovet. Ehntsiklopediya, Moscow, 1977–1985. (Russian)
336. M. H. A. Davis, D. Duffie, W. H. Fleming, S. E. Shreve (eds.), *Mathematical Finance*, Based on the proceedings of a workshop, held at IMA, University of Minnesota, Minneapolis, MN, USA 1992/93, Springer-Verlag, New York, 1995.
337. S. D. Howison, F. P. Kelly, P. Wilmott (eds.), *Mathematical Models in Finance*, Chapman & Hall, London, 1995.
338. Z. G. Maimin, "On detection of the deterministic component of a random field", *Theory of Probability and its Applications* **19** (1974), no. 3, 506–521.
339. J. H. McCuloch, "Simple consistent estimators of stable distribution parameters", *Communications in Statistics —Simulation and Computation* **15** (1986), no. 4, 1109–1136.
340. H.-P. McKean, "A free boundary problem for the heat equation arising from a problem of mathematical economics", *Industrial Management Review* **6** (1965), 32–39.
341. A. V. Melnikov, *Financial Markets: Stochastic Analysis and Pricing of Derivatives*, TVP Sci. Publ., Moscow, 1997. (Russian)
342. A. V. Melnikov and A. N. Shiryaev, "Criteria for the absence of arbitrage in the financial market", in *Frontiers in Pure and Applied Probability. II (Proceedings of the Fourth Russian-Finish Symposium on Probability Theory and Mathematical Statistics, Moscow, October 3–8, 1993)*, TVP Sci. Publ., Moscow, 1996, pp. 121–134.
343. J. Mémin, "Espaces de semi martingales et changements de probabilité", *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **52** (1980), no. 1, 9–39.
344. Merrill Lynch, Pierce, Fenner, Smith, Inc., *Security Risk Evaluation*, October, 1986.
345. R. Merton, *Continuous-Time Finance*, Blackwell, Cambridge, MA/Oxford, UK, 1990.
346. R. C. Merton, "Theory of rational option pricing", *Bell Journal of Economics and Management Science* (1973), no. 4 (Spring), 141–183.
347. P.-A. Meyer, "Notes sur les intégrales stochastiques. I. Intégrales hilbertiennes", *Lecture Notes in Mathematics* **581** (1977), 446–461.
348. M. G. Meyers, *A Financial History of the United States*, Columbia Univ. Press, New York, 1970.
349. V. S. Mikhalevich, "Bayesian testing of two hypothesis on mean value of normal process", *Vestnik Kiev University* **I** (1958), no. 1, 101–104.
350. M. Miller and F. Modigliani, "Dividend policy, growth, and the valuation of shares", *Journal of Business* **34** (October, 1961), 411–433.
351. T. C. Mills, *The Econometric Modelling of Financial Time Series*, Cambridge Univ. Press, Cambridge, 1995.

352. S. Mittnik and S. T. Rachev, "Stable Modelling in Finance. Special issue of the journal", *Mathematical and Computer Modelling* (1997).
353. S. Mittnik and S. T. Rachev, "Modeling asset returns with alternative stable distributions", *Econometrics Reviews* **12** (1993), no. 3, 261–330.
354. S. Mittnik and S. Rachev, "Stable distributions for asset returns", *Applied Mathematics Letters* **2** (1989), no. 3, 301–304.
355. F. Modigliani and M. Miller, "Corporation income taxes and the cost capital: a correction", *American Economic Review* **53** (June, 1963), 433–443.
356. F. Modigliani and M. Miller, "The cost of capital, corporation finance, and the theory of investment", *American Economic Review* **48** (June, 1958), 261–297.
357. K. M. Morris and A. M. Siegel, *Guide to Understanding Money and Investing*, Lightbulb Press, 1993.
358. T. Mullin (ed.), *The Nature of Chaos*, Clarendon Press, Oxford, 1994.
359. M. Musiela and D. Sondermann, *Different Dynamical Specifications of the Term Structure of Interest Rates and their Implications.* Preprint, Bonn University, Dept. of Statistics, Bonn, 1993.
360. U. A. Müller, M. M. Dacorogna, R. D. Davé, R. B. Olsen, O. V. Pictet, J. E. von Weizsäcker, *Volatilities of Different Time Resolutions—Analyzing the Dynamics of Market Components.* Preprint UAM. 1995-01-12, "Olsen & Associates", Research Institute for Applied Economics, Zürich, March 20, 1995. (See also [393] **1**; Journal of Empirical Finance, 1997.)
361. U. A. Müller, M. M. Dacorogna, R. D. Davé, O. V. Pictet, R. B. Olsen, and J. R. Ward, *Fractals and Intrinsic Time—A Challenge to Econometricians*, Technical Report UAM 1993-08-16, "Olsen & Associates", Research Institute for Applied Economics, Zürich, 1993.
362. U. A. Müller, M. M. Dacorogna, R. B. Olsen, O. V. Pictet, M. Schwarz, C. Morgenegg, "Statistical study of foreign exchange rates, empirical evidence of a price change scaling law, and intra-day analysis", *Journal of Banking and Finance* **14** (1990), 1189–1208.
363. R. Myneni, "The pricing of the American option", *Annals of Applied Probability* **2** (1992), no. 1, 1–23.
364. D. B. Nelson, "*ARCH* models as diffusion approximations", *Journal of Econometrics* **45** (1990), 7–38.
365. D. B. Nelson, "Asymptotic filtering theory for multivariate *ARCH* models", *Journal of Econometrics* **71** (1996), no. 1/2, 1–47.
366. D. B. Nelson, "Conditional heteroskedasticity in asset returns: a new approach", *Econometrics* **59** (1990), 347–370.
367. D. B. Nelson, "Filtering and forecasting with misspecified *ARCH* model", *Journal of Econometrics* **52** (1992), 61–90.
368. A. A. Novikov, "On an identity for stochastic integrals", *Theory of Probability and its Applications* **17** (1972), 717–720.
369. *Handbook of FX dealers*, Crocus International, Verba, Moscow, 1992. (Russian)
370. J. O'Brein and S. Shrivastava, *Financial Analysis and Security Trading*, Carnegie Mellon University, Pittsburg, PA, 1995.
371. M. F. M. Osborne, "Brownian motion in the stock market", *Operations Research* **7** (1959), 145–173. (See also [78], 100–128.)
372. A. R. Pagan, "Econometric issues in the analysis of regressions with generated regressors", *International Economic Review* **25** (1984), 221–247.

373. A. R. Pagan and G. W. Schwert, "Alternative models for conditional stock volatility", *Journal of Econometrics* **45** (1990), 267–290.
374. R. E. A. C. Paley and N. Wiener, "Fourier transforms in the complex domain", *American Mathematical Society Colloquium Publications* **19** (1934).
375. R. E. A. C. Paley, N. Wiener, and A. Zygmund, "Notes on random functions", *Mathematische Zeitschrift* (1933), no. 37, 647–668.
376. M. Parkinson, "Option pricing: The American put", *Journal of Business* **50** (1977), 21–36.
377. V. Pareto, *Cours d'Economie Politique*, Lausanne, Switzerland, 1897.
378. H.-O. Peitgen, H. Jürgens, D. Saupe, *Chaos and Fractals*, Springer-Verlag, New York, 1992.
379. H.-O. Peitgen and P. H. Richter, *The Beauty of Fractals. Images of Complex Dynamical Systems*, Springer-Verlag, Berlin, 1986.
380. R.-F. Peltier and J. L. Véhel, *A New Method for Estimating the Parameter of Fractional Brownian Motion.* Preprint, INRIA, Rocquencourt, 1996.
381. R.-F. Peltier and J. L. Véhel, *Multifractional Brownian Motion: Definition and Preliminary Results.* Preprint no. 2645, INRIA, Rocquencourt, August, 1995.
382. D. B. Percival, "Three curious properties of the sample variance and autocovariance for stationary processes with unknown mean", *American Statistician* **47** (1993), 274–276.
383. M. H. Pesaran and S. M. Potter (eds.), *Nonlinear Dynamics, Chaos and Econometrics*, Wiley, New York, 1993.
384. M. H. Pesaran and H. Samiei, "Estimating limited-dependent rational expections models with an application to exchange rate determination in a target zone", *Journal of Econometrics* **53** (1992), 141–163.
385. E. E. Peters, *Chaos and Order in the Capital Markets: A New View of Cycles, Prices, and Market Volatility*, Wiley, New York, 1991.
386. E. E. Peters, *Fractal Market Analysis: Applying Chaos Theory to Investment and Economics*, Wiley, New York, 1994.
387. N. Ya. Petrakov and V. I. Rotar', *Factor of Uncertainty and Control of Economical Systems*, Nauka, Moscow, 1985. (Russian)
388. P. F. Petrochenko (ed.), *English-Russian Business Glossary*, Art-Biznes-Centr, Moscow, 1992. (Russian)
389. M. Planck, "Über einen Satz der statistischen Dynamic und seine Erweiterung in der Quantentheorie", *Sitzungsberichte der Preussischen Akademie der Wissenschaften* (1917), 324–341.
390. P. Praetz, "The distribution of share price changes", *Journal of Business* **45** (1972), 49–55.
391. M. Priestley, *Non-Linear and Nonstationary Time Series*, Academic Press, New York, 1988.
392. J.-L. Prigent, *Incomplete Markets: Convergence of Options Values under the Minimal Martingale Measure.* Preprint no. 9526, THEMA, University of Cergy-Pontoise; also in *Proceedings of the 13th International Conference AFFI (Geneva, 1996).*
393. *Proceedings of the First International Conference on High Frequency Data in Finance (HFDF-I, March 29–31, 1995): Introductory Lecture and 4 volumes*, "Olsen & Associates", Research Institute for Applied Economics, Zürich.
394. Ju. V. Prohorov, "Convergence of random processes and limit theorems in probability theory", *Theory of Probability and its Applications* **1** (1956), no. 2, 157–214. (Russian)

395. Ph. Protter, *Stochastic Integration and Differential Equations: A New Approach*, Springer-Verlag, Berlin, 1990.
396. Ph. Protter and D. Talay, "The Euler scheme for Lévy driven stochastic differential equations", *Annals of Probability* **25** (1997), no. 1, 393–423.
397. T. Puu, *Nonlinear Economic Dynamics*, Springer-Verlag, New York, 1989. (Lecture Notes in Economics and Mathematical Systems, no. 336.)
398. J. B. Qiggins, "Option values under stochastic volatility. Theory and empirical estimates", *Journal of Financial Economics* **19** (1987), 351–372.
399. R. Rabemananjara and J. M. Zakoian, "Threshold *ARCH* models and asymmetries in volatility", in [383].
400. S. T. Rachev and L. Rüschendorf, "Models for option prices", *Theory of Probability and its Applications* **39** (1994), no. 1, 120–152.
401. K. Redhead and S. Huhhes, *Financial Risk Management*, 1996.
402. D. Revuz and M. Yor, *Continuous Martingales and Brownian Motion*, Springer-Verlag, Berlin, 1991.
403. M. Ridley, "The mathematics of markets", *The Economist* (October, 1993).
404. *Risk Metrics*, Morgan Guaranty Trust Company, New York, November, 1994.
405. H. V. Roberts, "Stock-market 'patterns' and financial analysis: Methodological suggestions", *Journal of Finance* **14** (1959), 1–10.
406. R. T. Rockafellar, *Convex Analysis*, Princeton Univ. Press, Princeton, NJ, 1970.
407. L. C. G. Rogers, "Equivalent martingale measures and no-arbitrage", *Stochastics and Stochastics Reports* **51** (1994), no. 1+2, 41–50.
408. L. C. G. Rogers and D. Talay (eds.), *Numerical Methods in Finance*, Cambridge Univ. Press, Cambridge, 1997.
409. B. A. Rogozin, "On some classes of processes with independent increments", *Theory of Probability and its Applications* **10** (1965), 479–483.
410. R. Roll and S. A. Ross, "An empirical investigation of the arbitrage pricing theory", *Journal of Finance* **35** (1980), 1073–1103.
411. S. A. Ross, "Information and volatility: The no-arbitrage martingale approach to timing and resolution irrelevancy", *Journal of Finance* **44** (1989), 1–18.
412. S. A. Ross, "The arbitrage theory of capital asset pricing", *Journal of Economic Theory* **13** (1976), 341–360.
413. L. N. Rubinstein, *Stefan Problem*, Zvaigzne, Riga, 1967. (Russian)
414. M. Rubinstein, *Exotic Options*. Working paper no. 220, Institute of Business and Economic Research, University of California, Berkeley, December, 1991.
415. M. Rubinstein, "Guiding force", in *From Black-Scholes to black holes*, Risk/Finex, London/New York, 1992, pp. 39–48.
416. J. Sacks and D. Ylvisaker, "Linear estimation for approximately linear models", *Annals of Statistics* **6** (1978), no. 5, 1122–1137.
417. G. Samorodnitsky, "A class of shot noise models for financial applications", [393] **3**.
418. G. Samorodnitsky and M. S. Taqqu, *Stable Non-Gaussian Random Processes*, Chapman & Hall, New York, 1994.
419. P. A. Samuelson, "Proof that properly anticipated prices fluctuate randomly", *Industrial Management Review* **6** (1965), 41–49.
420. P. A. Samuelson, "Rational theory of warrant pricing", *Industrial Management Review* **6** (1965), 13–31.

421. K. Sandmann and D. Sondermann, "A note on the stability of lognormal interest rate models and the pricing of Eurodollar futures", *Mathematical Finance* **7** (1997), no. 2, 119-125.
422. K. Sandmann and D. Sondermann, "A term structure model and the pricing of interest rate derivatives", *Review of Futures Markets* **12** (1993), no. 2, 391–423.
423. Sato Ken-iti, *Lévy Processes on the Euclidean Spaces.* Preprint, Institute of Mathematics, University of Zürich, Zürich, 1995.
424. W. Schachermayer, "A Hilbert space proof of the fundamental theorem of asset pricing in finite discrete time", *Insurance: Mathematics & Economics* **11** (1992), no. 4, 249-257.
425. W. Schachermayer, "Martingale measure for discrete-time processes with infinite horizon", *Mathematical Finance* **4** (1994), no. 1, 25–55.
426. W. M. Schmidt, "On a general class of one-factor models for the term structure of interest rates", *Finance and Stochastics* **1** (1997), no. 1, 3–24.
427. R. Schnidrig and D. Würtz, "Investigation of the volatility and autocorrelation function of the USD/DEM exchange rate on operational time scales", [393] **3**.
428. H. G. Schuster, *Deterministic Chaos. An Introduction*, Physik Verlag, Weinheim, 1984.
429. M. Schweizer, "On the minimal martingale measure and the Föllmer–Schweizer decomposition", *Stochastic Analysis and Applications* **13** (1995), no. 5, 573–599.
430. M. Schweizer, "Variance-optimal hedging in discrete time", *Mathematics of Operations Research* **20** (1995), no. 1, 1–32.
431. G. W. Schwert and P. J. Segnin, "Heteroskedasticity in stock returns", *Journal of Finance* **45** (1990), 1129–1155.
432. L. O. Scott, "Option pricing when the variance changes randomly: Theory, estimation and an application", *Journal of Financial and Quantitative Analysis* **22** (1987), 419–438.
433. W. F. Sharpe, "Capital asset prices: A theory of market equilibrium under conditions of risk", *Journal of Finance* **19** (September, 1964), 425–442.
434. L. A. Shepp and A. N. Shiryaev, "A new look at pricing of the 'Russian option'", *Theory of Probability and its Applications* **39** (1994), no. 1, 103–119.
435. L. A. Shepp and A. N. Shiryaev, "The Russian option: Reduced regret", *Annals of Applied Probability* **3** (1993), no. 3, 631–640.
436. R. J. Shiller, *Stock Market Volatility*, MIT Press, Cambridge, MA.
437. D. Shimko, *Finance in Continuous Time. A Primer*, Kolb Publ. Co., Miami, 1992.
438. A. N. Shiryaev, "Actuarial and financial business: Current state of art and perspectives of development (Notes of communication presented at Constituent Conference of Russian Society of Actuaries, September 14, 1994)", *Surveys in Applied and Industrial Mathematics* (TVP Sci. Publ., Moscow) **1** (1994), no. 5, 684–697. (Russian)
439. A. N. Shiryaev, *Probability. 2nd ed.*, Springer-Verlag, New York, 1995.
440. A. N. Shiryaev, "The problem of the most rapid detection of a disturbance in a stationary process", *Soviet Mathematics, Dokl.* **2** (1961), 795–799.
441. A. N. Shiryaev, *Optimal stopping rules*, Springer-Verlag, New York, 1978.
442. A. N. Shiryaev, "Stochastic problems of financial mathematics", *Surveys in Applied and Industrial Mathematics* (TVP Sci. Publ., Moscow) **1** (1994), no. 5, 780–820. (Russian)

443. A. N. Shiryaev, Yu. M. Kabanov, O. D. Kramkov, and A. V. Mel'nikov, "Toward the theory of pricing of options of both European and American types. I. Discrete time", *Theory of Probability and its Applications* **39** (1994), no. 1, 14–60.
444. A. N. Shiryaev, Yu. M. Kabanov, O. D. Kramkov, and A. V. Mel'nikov, "Toward the theory of pricing of options of both European and American types. II. Continuous time", *Theory of Probability and its Applications (Special Issue)* **39** (1994), no. 1, 61–102.
445. A. N. Shiryaev and V. G. Spokoiny, "On sequential estimation of an autoregressive parameter", *Stochastics and Stochastics Reports* **60** (1997), no. 3+4, 219–240.
446. A. V. Skorokhod, *Stochastic processes with independent increments*, Nauka, Moscow, 1967. (Russian)
447. C. A. Sin, *Strictly Local Martingales and Hedge Ratios on Stochastic Volatility Models*, Cornell University, Graduate School, Ithaka, NY, 1966.
448. R. L. Smith, "Estimating dimension in noisy chaotic time series", *Journal of the Royal Statistical Society. Ser. B* **54** (1992), no. 2, 329–351.
449. R. L. Smith, "Optimal estimation of fractal dimension", in *Nonlinear Modeling and Forecasting (SFI Studies in the Science of Complexity. Proceedings, vol. 12)* (eds. M. Casdagli and S. Eubank), Addison-Wesley, Reading, MA, 1992.
450. R. Sobel, *Inside Wall Street: Continuity and Change in the Financial District*, W. W. Norton & Co., New York, 1977.
451. G. Soros, *The Alchemy of Finance. Reading the Mind of the Market*, Wiley, New York, 1994.
452. V. G. Spokoiny and A. N. Shiryaev, *Statistical Experiments and Decisions: Asymptotic Theory* (in print).
453. E. M. Stein and C. J. Stein, "Stock prices distributions with stochastic volatility: an analytic approach", *Review of Financial Studies* **4** (1991), no. 4, 727–752.
454. "Stochastic Aspects of Financial Mathematics", *Theory of Probability and its Applications (Special Issue)* **39** (1994), no. 1.
455. C. Stricker, "Arbitrage et lois de martingale", *Annales de l'Institut Henri Poincaré* **26** (1990), no. 2, 451–460.
456. S. H. Strogatz, *Nonlinear Dynamics and Chaos*, Addison-Wesley, Reading, MA.
457. D. W. Stroock and S. R. S. Varadhan, "Diffusion processes with continuous coefficients. I, II", *Communications on Pure and Applied Mathematics* **22** (1969), no. 3, 345–400; no. 4, 479–530.
458. L. E. O. Svensson, "Target zones and interest rate variability", *Journal of International Economics* **31** (1991), 27–54.
459. M. S. Taqqu, "Self-similar processes", in *Encyclopaedia of Statistical Sciences* (eds. S. Kotz and N. Johnson), vol. 8, Wiley, New York, pp. 352–357.
460. S. Taylor, *Modeling Financial Time Series*, Wiley, New York, 1986.
461. H. Tong, *Nonlinear Time Series*, Oxford Univ. Press, Oxford, 1990.
462. H. Tong, *Nonlinear Time Series: A Dynamical System Approach*, Clarendon Press, Oxford, 1990.
463. A. N. Tikhonov and A. A. Samarskiji, *Equations of Mathematical Physics*, Pergamon Press, Oxford, 1963.
464. A. Timmermann, "Scales and stock markets", *Nature* **376** (1995), 18–19.
465. A. Tversky, *The Psychology of Risk. Quantifying the Market Risk Premium Phenomena for Investment Decision Making*, Institute of Chartered Financial Analysts, Charlottesville, VA, 1990.

466. G. E. Uhlenbeck and L. S. Ornstein, "On the theory of Brownian motion", *Physical Review* **36** (1930), 823–841.
467. P. L. J. van Moerbeke, "On optimal stopping and free boundary problems", *Archive for Rational Mechanics and Analysis* **60** (1976), no. 2, 101–148.
468. H. R. Varian (ed.), *Economic and Financial Modeling with Mathematica (TELOS—The Electronic Library of Science)*, Springer-Verlag, Berlin, 1993.
469. S. R. Vegle (ed.), *Stocks, Bonds, Options, Futures*, Institute of Finance, Prentice Hall, New York, 1987.
470. A. D. Venttsel, *A Course in the Theory of Stochastic Processes*, McGraw Hill, New York, 1981.
471. A. Yu. Veretennikov, "On the strong solutions of stochastic differential equations", *Theory of Probability and its Applications* **24** (1980), 354–366.
472. O. Vasicek, "An equilibrium characterization of the term structure", *Journal of Financial Economics* **5** (1977), 177–188.
473. R. F. Voss, "$1/f$ noise and fractals in economic time series", in *Fractal Geometry and Computer Graphics* (eds. J. L. Encarnação, H.-O. Peitgen, and G. Englert), Springer-Verlag, Berlin, 1992, pp. 45–52.
474. C. Walter, "Levy-stability under addition and fractal structure of markets", in *5th AFIR International Colloquium (Bruxelles, September, 1995)*.
475. H. Weyl, "Bemerkungen zum Begriff der Differential-Quotenten gebrochener Ordnung", *Viertelschr. Naturforsch. Ges. Zürich* **62** (1967), 296–302.
476. N. Wiener, "Differential space", *Journal of Mathematical Physics. Math. Inst. Tech.* **2** (1923), 131 174.
477. J. B. Wiggins, "Option values under stochastic volatility. Theory and empirical evidence", *Journal of Financial Economics* **19** (1987), 351–372.
478. P. Wilmott, J. Dewynne, and S. Howison, *Option Pricing: Mathematical Models and Computation*, Oxford Financial Press, Oxford, 1993.
479. P. Wilmott, S. Howison, and J. Dewynne, *The Mathematics of Financial Derivatives (A Student Introduction)*, Cambridge Univ. Press, Cambridge, 1996.
480. H. Working, "A random-difference series for use in the analysis of time series", *Journal of American Statistical Association* **29** (1934), 11–24.
481. B. Zhou, *Forecasting Foreign Exchange Rates Subject to De-Volatilization.* Working paper no. 3510, Sloan School of Management, MIT, Cambridge, MA, 1992. (See also [393] bf4.)
482. S. Z. Zinov'ev, "Infrastructural factor of the securities market", *Business Express* (25.06.1996), no. 23. (Russian)
483. V. M. Zolotarev, "Distribution of superposition of infinitely divisible processes", *Theory of Probability and its Applications* **3** (1958), no. 2, 185–188.
484. V. M. Zolotarev, *One-dimensional Stable Distributions*, Amer. Math. Soc., Providence, RI, 1986.
485. A. K. Zvonkin, "A transformation of the phase space of a diffusion process that removes the drift", *Mathematics of the USSR, Sbornik* **22** (1975), 129–149.

Index

Index of Symbols

ISBN 981-02-3605-0

9 789810 236052

3907 hc

http://www.worldscientific.com/