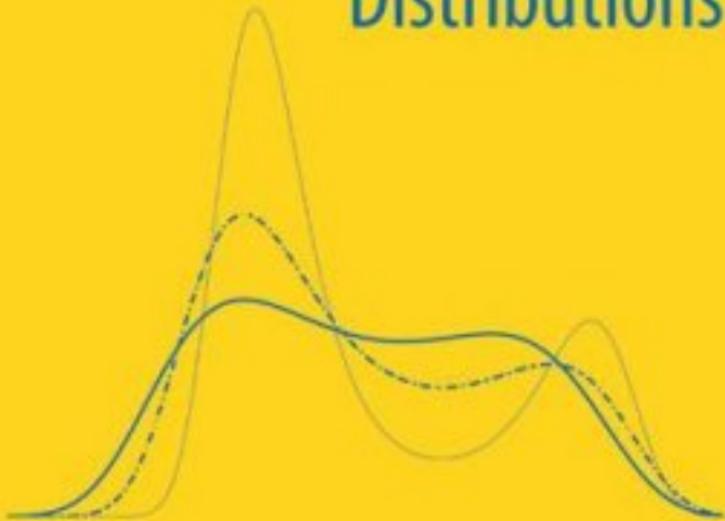


Eric Jondeau • Ser-Huang Poon
Michael Rockinger

Financial Modeling Under Non-Gaussian Distributions



Springer

Springer Finance

Editorial Board

M. Avellaneda

G. Barone-Adesi

M. Broadie

M.H.A. Davis

E. Derman

C. Klüppelberg

E. Kopp

W. Schachermayer

Springer Finance

Springer Finance is a programme of books aimed at students, academics and practitioners working on increasingly technical approaches to the analysis of financial markets. It aims to cover a variety of topics, not only mathematical finance but foreign exchanges, term structure, risk management, portfolio theory, equity derivatives, and financial economics.

- Ammann M.*, Credit Risk Valuation: Methods, Models, and Application (2001)
Back K., A Course in Derivative Securities: Introduction to Theory and Computation (2005)
Barucci E., Financial Markets Theory. Equilibrium, Efficiency and Information (2003)
Bielecki T.R. and Rutkowski M., Credit Risk: Modeling, Valuation and Hedging (2002)
Bingham N.H. and Kiesel R., Risk-Neutral Valuation: Pricing and Hedging of Financial Derivatives (1998, 2nd ed. 2004)
Brigo D. and Mercurio F., Interest Rate Models: Theory and Practice (2001, 2nd ed. 2006)
Buff R., Uncertain Volatility Models-Theory and Application (2002)
Carmona R.A. and Tehranchi M.R., Interest Rate Models: an Infinite Dimensional Stochastic Analysis Perspective (2006)
Dana R.A. and Jeanblanc M., Financial Markets in Continuous Time (2002)
Deboeck G. and Kohonen T. (Editors), Visual Explorations in Finance with Self-Organizing Maps (1998)
Delbaen F. and Schachermayer W., The Mathematics of Arbitrage (2005)
Elliott R.J. and Kopp P.E., Mathematics of Financial Markets (1999, 2nd ed. 2005)
Fengler M.R., Semiparametric Modeling of Implied Volatility (2005)
Geman H., Madan D., Pliska S.R. and Vorst T. (Editors), Mathematical Finance–Bachelier Congress 2000 (2001)
Gundlach M., Lehrbass F. (Editors), CreditRisk⁺ in the Banking Industry (2004)
Kellerhals B.P., Asset Pricing (2004)
Külpmann M., Irrational Exuberance Reconsidered (2004)
Kwok Y.-K., Mathematical Models of Financial Derivatives (1998)
Malliavin P. and Thalmaier A., Stochastic Calculus of Variations in Mathematical Finance (2005)
Meucci A., Risk and Asset Allocation (2005)
Pelsser A., Efficient Methods for Valuing Interest Rate Derivatives (2000)
Prigent J.-L., Weak Convergence of Financial Markets (2003)
Schmid B., Credit Risk Pricing Models (2004)
Shreve S.E., Stochastic Calculus for Finance I (2004)
Shreve S.E., Stochastic Calculus for Finance II (2004)
Yor M., Exponential Functionals of Brownian Motion and Related Processes (2001)
Zagst R., Interest-Rate Management (2002)
Zhu Y.-L., Wu X., Chern I.-L., Derivative Securities and Difference Methods (2004)
Ziegler A., Incomplete Information and Heterogeneous Beliefs in Continuous-time Finance (2003)
Ziegler A., A Game Theory Analysis of Options (2004)

Eric Jondeau, Ser-Huang Poon
and Michael Rockinger

Financial Modeling Under Non-Gaussian Distributions

Eric Jondeau
University of Lausanne and Swiss
Finance Institute
HEC-Department of Finance and
Insurance
CH-1015 Lausanne-Dorigny
Switzerland

Ser-Huang Poon
Manchester Business School
University of Manchester
Oxford Road
Manchester M13 9PL
UK

Michael Rockinger
University of Lausanne and Swiss
Finance Institute
HEC-Department of Finance and
Insurance
CH-1015 Lausanne-Dorigny
Switzerland

Mathematics Subject Classification (2000): 91B28, 91B84

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

Library of Congress Control Number: 2006923899

Springer Finance Series ISSN 1616-0533
ISBN-10 1-84628-419-8
ISBN-13 978-1-84628-419-9

Printed on acid-free paper

© 2007 Springer-Verlag London Limited
MATLAB® is a trademark of The MathWorks, Inc. and is used with permission. The MathWorks does not warrant the accuracy of the text or exercises in this book. This book's use or discussion of MATLAB® software or related products does not constitute endorsement or sponsorship by The MathWorks of a particular pedagogical approach or particular use of the MATLAB® software.

Apart from any fair dealing for the purposes of research or private study, or criticism or review, as permitted under the Copyright, Designs and Patents Act 1988, this publication may only be reproduced, stored or transmitted, in any form or by any means, with the prior permission in writing of the publishers, or in the case of reprographic reproduction in accordance with the terms of licences issued by the Copyright Licensing Agency. Enquiries concerning reproduction outside those terms should be sent to the publishers.

The use of registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant laws and regulations and therefore free for general use.

The publisher makes no representation, express or implied, with regard to the accuracy of the information contained in this book and cannot accept any legal responsibility or liability for any errors or omissions that may be made.

Printed in the United States of America (MVY)

9 8 7 6 5 4 3 2 1

Springer Science+Business Media, LLC
springer.com

To my family
from Eric Jondeau

In loving memory of Janet and Harry
from Ser-Huang Poon

To Laura
from Michael Rockinger

Preface

Practitioners and academic researchers who have handled financial market data know that asset return distributions do not have the bell shape that is associated with the Gaussian or normal distribution. The fact that many popular models are still based on the assumption of normality is because of the simplicity that the Gaussian model presents, and perhaps the bigger reason is the lack of understanding of the modeling, estimation, and handling of non-Gaussian distributions. The use of Gaussian models, when the asset return distributions are not normal, can be very dangerous; such a practice could lead to a wrong choice of portfolio, the underestimation of extreme losses, and hugely mispriced derivative products. Non-Gaussian distributions are the key theme of this book, which addresses the causes and the consequences of non-normality of asset returns.

Other books in this field include Campbell, Lo, and Mackinlay (1997), and Embrechts, Klüppelberg, and Mikosch (1997). In related fields, one may also cite Gouriéroux and Jasiak (2001), Tsay (2002), and Taylor (2005). On option pricing, we have the book of Schoutens (2003), and of Cont and Tankov (2004). All these books are complementary to this one and are useful for understanding some aspects of the modeling of asset returns or option prices. In this book, we cover a wide range of issues that are resulting from non-normality and time dependency of both asset returns and option prices.

The book is written for non-mathematicians who want to model financial market prices. Its emphasis is on practice. We worked hard to make the materials accessible to non-mathematicians but at the same time not to sacrifice the mathematical rigor and the complexity of the original models. This book targets practitioners in the finance industry and especially those who call themselves *Quant* and have the responsibility of managing portfolios and monitoring financial risk. We also hope the *not-so-Quant* people will find this book useful in understanding the myth of *rocket scientists'* involvement in quantitative finance. This book is suitable for use as a core text for specialist M.Sc. and Ph.D. students in Empirical Finance, Financial Econometrics and

Financial Derivatives. It is useful for mathematicians who want to know more about how their mathematical tools are applied in finance and for those who just want to know more about finance and financial markets. We provide, in Part V, some mathematical appendices that we feel are sufficient for coping with the mathematical contents of the main parts. Some basic knowledge in statistics, calculus, and probability plus plenty of determination should see the readers through the core materials. Most of the material presented in this book is extracted from our papers and material used to teach practitioners and generations of M.Sc. and Ph.D. students at University of Lausanne and Manchester Business School.

One of our main aims is to bridge the gap between the theoretical developments and the practical implementations of what many users and researchers perceived as “sophisticated” models or black boxes. We offer many empirical illustrations of the models described in this book, and especially those developed in Parts II and III.¹ Although we use only a small collection of stock market data in these examples due to the space constraint, many of the techniques and models described here could equally be applied to other financial time series, such as exchange and interest rates. But it is important to remember that models are meant to capture stylized facts. Different types of financial assets may have slightly or very different characteristics. Another disclaimer is of importance: Although we argue and provide evidence that there is a need to incorporate non-normality and time dependency in the modeling of asset returns, we do not suggest that this should always be the case. Moreover, we would not be held responsible for any problem the reader may encounter in the implementation of these techniques.

Finally, all the figures and estimations reported in this book have been produced using MATLAB.² This software provides a very efficient optimization routine, and we have made available on our web sites many MATLAB codes that are used in this book.

Eric Jondeau and Michael Rockinger both want to thank the Ecole des HEC of the University of Lausanne, the International Center for Financial Asset Management and Engineering (FAME), and the Swiss Finance Institute (SFI) for their constant support. They thank many generations of M.Sc. and Ph.D. students for the many discussions, comments and questions. Ser-Huang Poon wants to thank the Manchester Business School, and the Manchester Accounting and Finance Group in particular, for providing much support.

¹ Daily data used in this book comes from Datastream, and intradaily data comes from the Euronext Paris dataset.

² MATLAB® is a registered trademark of The MathWorks, Inc. For MATLAB product information, please contact: The MathWorks, Inc. 3 Apple Hill Drive, Natick, MA 01760-2098, USA. E-mail: info@mathworks.com. Web: www.mathworks.com.

She wants to thank also her students for many stimulating discussions and grilling reading classes. All three authors express their gratitude to Laura Rockinger for reading the entire manuscript and correcting many mistakes. We express our thanks to many other people, colleagues, friends, and family, to whom we are most grateful for their support. They are too many to name here.

Eric Jondeau, Ser-Huang Poon, and Michael Rockinger
August 2006

Contents

Part I Financial Markets and Financial Time Series

1	Introduction	3
1.1	Financial markets and financial time series	3
1.2	Econometric modeling of asset returns	4
1.3	Applications of non-Gaussian econometrics	5
1.4	Option pricing with non-Gaussian distributions	5
2	Statistical Properties of Financial Market Data	7
2.1	Definitions of returns	7
2.1.1	Simple returns	8
2.1.2	Log-returns	8
2.1.3	Stylized facts	9
2.2	Distribution of returns	10
2.2.1	Moments of a random variable	10
2.2.2	Empirical moments	14
2.2.3	Testing for normality	16
2.3	Time dependency	21
2.3.1	Serial correlation in returns	22
2.3.2	Serial correlation in volatility	23
2.3.3	Volatility asymmetry	25
2.3.4	Time-varying higher moments	26
2.4	Linear dependence across returns	26
2.4.1	Pearson's correlation coefficient	27
2.4.2	Test for equality of two correlation coefficients	28
2.4.3	Test for equality of two correlation matrices	30
2.5	Multivariate higher moments	31
2.5.1	Multivariate co-skewness and co-kurtosis	31
2.5.2	Computing moments of portfolio returns	32

3 Functioning of Financial Markets and Theoretical Models for Returns	33
3.1 Functioning of financial markets	34
3.1.1 Organization of financial markets	34
3.1.2 Examples of orders	37
3.1.3 Components of the bid-ask spread	39
3.2 Mandelbrot and the stable distribution	39
3.2.1 A puzzling result	40
3.2.2 Stable distribution	41
3.3 Clark's subordination model	44
3.3.1 The idea of the model	44
3.3.2 The density of returns under subordination	46
3.4 A bivariate mixture-of-distribution model for return and volume	48
3.4.1 A microstructure model for information arrivals	48
3.4.2 Implications of the mixture of distributions hypothesis	53
3.4.3 Testing the mixture of distribution hypothesis	57
3.4.4 Extensions	61
3.5 A model of prices and quotes in a quote-driven market	62
3.5.1 A model based on the trade flow	63
3.5.2 Estimating the parameters	66
3.5.3 The quote process	68
3.5.4 Extension to the liquidation of a large portfolio	73

Part II Econometric Modeling of Asset Returns

4 Modeling Volatility	79
4.1 Volatility at lower frequencies	79
4.2 ARCH model	81
4.2.1 Forecasting	81
4.2.2 Kurtosis of an ARCH model	82
4.2.3 Testing for ARCH effects	82
4.2.4 ARCH-in-mean model	83
4.2.5 Illustration	84
4.3 GARCH model	84
4.3.1 Forecasting	88
4.3.2 Integrated GARCH model	89
4.3.3 Estimation	89
4.3.4 Testing for GARCH effects	92
4.3.5 Software to estimate ARCH and GARCH models	92
4.3.6 Illustration	93
4.4 Asymmetric GARCH models	94
4.4.1 EGARCH model	94
4.4.2 TGARCH model	95

4.4.3	GJR model	95
4.4.4	Cox-Box transform	95
4.4.5	News impact curve	96
4.4.6	Partially non-parametric estimation	96
4.4.7	Testing for asymmetric effects	97
4.4.8	Illustration	99
4.5	GARCH model with jumps	99
4.5.1	A model with time-varying jump intensity	101
4.5.2	An empirical illustration	105
4.6	Aggregation of GARCH processes	108
4.6.1	Temporal aggregation	109
4.6.2	Cross-sectional aggregation	113
4.6.3	Estimation of the weak GARCH process	114
4.7	Stochastic volatility	115
4.7.1	From GARCH models to stochastic volatility models	115
4.7.2	Estimation of the discrete time SV model	117
4.8	Realized volatility	118
4.8.1	The difficulty to disentangle jumps	119
4.8.2	Quadratic variation	123
4.8.3	Power variation	124
4.8.4	Bipower variation	126
4.8.5	Estimation over finite time intervals	128
4.8.6	Realized covariance	135
4.8.7	Further related results	141
5	Modeling Higher Moments	143
5.1	The general problem	144
5.1.1	Higher moments of a GARCH process	145
5.1.2	Quasi Maximum Likelihood Estimation	148
5.1.3	The existence of distribution with given moments	151
5.2	Distributions with higher moments	152
5.2.1	Semi-parametric approach	153
5.2.2	Series expansion about the normal distribution	155
5.2.3	Skewed Student t distribution	159
5.2.4	Generating asymmetric distributions	166
5.2.5	Pearson IV distribution	169
5.2.6	Entropy distribution	172
5.3	Specification tests and inference	177
5.3.1	Moment specification tests	177
5.3.2	Adequacy tests based on density forecasts	179
5.3.3	Adequacy tests based on interval forecasts	180
5.4	Illustration	182
5.5	Modeling conditional higher moments	188
5.5.1	Tests for autoregressive conditional higher moments	189
5.5.2	Modeling higher moments directly	189

5.5.3	Modeling the parameters of the distribution	191
6	Modeling Correlation	195
6.1	Multivariate GARCH models	197
6.1.1	Vectorial and diagonal GARCH models	198
6.1.2	Dealing with large-dimensional systems	200
6.1.3	Modeling conditional correlation	206
6.1.4	Estimation issues	210
6.1.5	Specification tests	212
6.1.6	Test of constant conditional correlation matrix	214
6.1.7	Illustration	217
6.2	Modeling the multivariate distribution	223
6.2.1	Standard multivariate distributions	225
6.2.2	Skewed elliptical distribution	230
6.2.3	Skewed Student <i>t</i> distribution	233
6.2.4	Estimation	236
6.2.5	Adequacy tests	239
6.2.6	Illustration	240
6.3	Copula functions	240
6.3.1	Definitions and properties	241
6.3.2	Measures of concordance	242
6.3.3	Non-parametric copulas	244
6.3.4	Review of some copula families	245
6.3.5	Estimation	254
6.3.6	Adequacy tests	258
6.3.7	Modeling the conditional dependency parameter	259
6.3.8	Illustration	261
7	Extreme Value Theory	265
7.1	Univariate tail estimation	266
7.1.1	Distribution of extremes	266
7.1.2	Tail distribution	276
7.1.3	The case of weakly dependent data	291
7.1.4	Estimation of high quantiles	296
7.2	Multivariate dependence	300
7.2.1	Characterizing tail dependency	303
7.2.2	Estimation and statistical inference on $\bar{\chi}$ and χ	307
7.2.3	Modeling dependency	308
7.2.4	An illustration	309
7.2.5	Further investigations	311

Part III Applications of Non-Gaussian Econometrics

8 Risk Management and VaR	315
8.1 Definitions and measures	316
8.1.1 Definitions	316
8.1.2 Models for portfolio returns	320
8.2 Historical simulation	321
8.3 Semi-parametric approaches	322
8.3.1 Extreme Value Theory (EVT)	324
8.3.2 Quantile regression technique	328
8.4 Parametric approaches	330
8.4.1 RiskMetrics – J.P. Morgan	331
8.4.2 The portfolio-level approach	334
8.4.3 The asset-level approach	337
8.5 Non-linear models	341
8.5.1 The “delta-only” method	341
8.5.2 The “delta-gamma” method	341
8.6 Comparison of VaR models	342
8.6.1 Evaluation of VaR models	343
8.6.2 Comparison of methods	343
8.6.3 10-day VaR and scaling	344
8.6.4 Illustration	345
9 Portfolio Allocation	349
9.1 Portfolio allocation under non-normality	349
9.1.1 Direct maximization of expected utility	350
9.1.2 An approximate solution based on moments	353
9.2 Portfolio allocation under downside risk	359
9.2.1 Definition	360
9.2.2 Downside risk as an additional constraint	360
9.2.3 Downside risk as an optimization criterion	361

Part IV Option Pricing with Non-Gaussian Returns

10 Fundamentals of Option Pricing	365
10.1 Notations	366
10.2 The no-arbitrage approach to option pricing	369
10.2.1 Choice of a stock price process	369
10.2.2 The fundamental partial differential equation	371
10.2.3 Solving the fundamental PDE	373
10.2.4 The Black-Scholes-Merton formula	375
10.3 Martingale measure and BSM formula	377
10.3.1 Self-financing strategies and portfolio construction	377

10.3.2 Change of numeraire	378
10.3.3 Change of Brownian motion	378
10.3.4 Evolution of S_t under Q	379
10.3.5 The expected pay-off as a martingale	379
10.3.6 The trading strategies	380
10.3.7 Equivalent martingale measure	381
11 Non-structural Option Pricing	383
11.1 Difficulties with the standard BSM model	384
11.2 Direct estimation of the risk-neutral density	385
11.2.1 Expression for the RND	385
11.2.2 Estimating the parameters of the RND	387
11.3 Parametric methods	389
11.3.1 Mixture of log-normal distributions	389
11.3.2 Mixtures of hypergeometric functions	394
11.3.3 Generalized beta distribution	395
11.4 Semi-parametric methods	395
11.4.1 Edgeworth expansions	395
11.4.2 Hermite polynomials	399
11.5 Non-parametric methods	402
11.5.1 Spline methods	402
11.5.2 Tree-based methods	406
11.5.3 Maximum entropy principle	407
11.5.4 Kernel regression	408
11.6 Comparison of various methods	409
11.7 Relationship with real probability	414
11.7.1 The link between RNDs and objective densities	414
11.7.2 Empirical findings	416
12 Structural Option Pricing	417
12.1 Stochastic volatility model	417
12.1.1 The square root process	418
12.1.2 Solving the PDE based on characteristic function	419
12.1.3 A new partial differential equation	422
12.2 Option pricing with stochastic volatility	425
12.2.1 Hull and White (1987, 1988)	425
12.2.2 Heston (1993)	426
12.2.3 Characteristic function of the SV model	428
12.2.4 Further insights	429
12.3 Models with jumps	432
12.3.1 Stochastic process with jumps	432
12.3.2 Diffusion with double exponential jumps	434
12.3.3 Combining stochastic volatility with jumps	436
12.3.4 Jumpy affine models	440
12.4 Models with even wilder jumps: Lévy option pricing	441

12.4.1 Commonly used Lévy processes	443
12.4.2 Choice of the time-changing process	444
12.4.3 Option pricing	445
12.4.4 Pricing options with risk-neutral characteristic function	446
12.4.5 Empirical results	447

Part V Appendices on Option Pricing Mathematics

13 Brownian Motion and Stochastic Calculus	451
13.1 Law of large numbers and the central limit theorem	451
13.2 Random walks	453
13.3 Construction of the Brownian motion	453
13.4 Properties of the Brownian motion	456
13.5 Stochastic integration	457
13.6 Stochastic differential equations	459
13.7 Ito's lemma	460
13.8 Multivariate extension of Ito's lemma	462
13.9 Transition probabilities and partial differential equations	463
13.10 Kolmogorov backward and forward equations	464
13.11 PDE associated with diffusions	466
13.12 Feynman-Kac formula	468
14 Martingale and Changing Measure	471
14.1 Martingales	471
14.2 Changing probability of a normal distribution	472
14.3 Radon-Nikodym derivative	473
14.4 Girsanov's theorem	474
14.5 Martingale representation theorem	475
15 Characteristic Functions and Fourier Transforms	477
15.1 Characteristic functions	477
15.1.1 Basic properties	478
15.1.2 Moments and the characteristic function	478
15.1.3 Convolution theorem	479
15.1.4 Uniqueness	480
15.1.5 Inversion theorem	480
15.2 Fourier transform and characteristic function	483
16 Jump Processes	487
16.1 Counting and marked point process	487
16.2 The Poisson process	489
16.2.1 Construction of the Poisson distribution	489
16.2.2 Properties of the Poisson distribution	491
16.2.3 Moments of pure Poisson process	492

XVIII Contents

16.2.4	Compound Poisson process	493
16.3	The exponential distribution	494
16.3.1	Definition and properties	494
16.3.2	Moments of the exponential variable	495
16.3.3	Hazard and survivor functions	496
16.4	Duration between Poisson jumps	497
16.5	Compensated Poisson processes	498
17	Lévy Processes	501
17.1	Construction of the Lévy process	501
17.2	Properties of Lévy processes	505
References		507
Index		535

1

Introduction

This book is about the modeling of asset returns when the normality assumption does not apply. It provides an up-to-date and step-by-step description of the tools that are useful for the modeling of non-Gaussian asset return distributions and for option pricing in the non-Gaussian context.

1.1 Financial markets and financial time series

For more than four decades, distributions of financial asset returns have been known to be non-Gaussian (see Mandelbrot, 1963, and Fama, 1965). The assumption of normality is stacked against two hard facts: First, the empirical distributions of asset returns have tails thicker than those from a normal distribution and appear to be negatively skewed. This means more extreme negative values, which has a very serious implication for risk management and portfolio selection. Second, returns are time dependent. Squared returns, absolute returns, and all measures and proxies of volatility exhibit strong serial correlation. This is now known as volatility clustering or conditional heteroskedasticity (Engle, 1982).

Financial modeling is all about capturing and exploiting patterns in the data including the two phenomena mentioned above. Chapter 2 discusses the unique statistical properties of financial market data and several so-called stylized facts. These stylized facts will be the basis for Part II where each chapter will tackle some specific features of financial market returns.

Chapter 3 describes the actual functioning and the microstructure of financial markets. Here, we present some theoretical models that may help explaining why asset returns are non-normal and time dependent. The foundation is built on Clark (1973) who postulates that non-normality and volatility clustering could be due to intermittent information arrivals.

1.2 Econometric modeling of asset returns

Part II is concerned with the time series aspects of asset returns. Chapters 4, 5, and 7 cover models for the second, third, and fourth moments and the tails of return distributions. These higher moments and tail measures are the hallmarks of non-Gaussian distributions. Chapter 6 deals with the dependence structure when the higher moments display significant departure from normality and returns appear to be time dependent. The dependence among the tail observations, described in Chapter 7, is very different from the dependence, described in Chapter 6, of the central and main part of the distributions, because of the differences in the underpinning statistical theories and the important fact that financial markets do behave very differently between normal and crisis periods.

Specifically, Chapter 4 covers models for volatility that include the better known GARCH (Generalized Autoregressive and Conditional Heteroskedasticity) class of models and some new extensions such as GARCH models with jumps and realized volatility models. With high-frequency data becoming more common these days, realized volatility is expected to remain an area of active research. This chapter also describes the lesser known or lesser discussed issues on GARCH aggregation and the relationship between stochastic volatility model in continuous time and the discrete time GARCH model.

Although time-varying volatility and volatility asymmetry may produce thick-tail and asymmetric distributions in asset return, volatility alone cannot explain away all the non-normality. To fully capture return distributions, we also need models for skewness and kurtosis. Chapter 5 does exactly that by fitting time-varying higher-moment conditional models to returns. It also describes tests for the adequacy of these conditional high moment models.

Chapters 4 and 5 are concerned with univariate time series characteristics. Chapter 6 shifts the focus to the relationships between and among the asset return series. This involves two main tasks. First, we have to extend the GARCH family to a (possibly large) number of assets in order to reproduce the joint dynamic of volatility. Second, we have to capture in this multivariate framework the non-normality of returns. Here, we move from a multivariate GARCH model with normal distribution to one with skewed Student t distribution that is designed to capture both fat-tailedness and asymmetry. We also examine an alternative approach that circumvents some difficulties in designing a multivariate distribution. The so-called copula approach is a tool that is able to join any type of marginal distribution. It has many theoretical appeals. But in many finance applications in practice, integration of the joint distribution is needed. This cannot be done analytically in the copula approach and will become more and more cumbersome as the number of assets or the dimension of the problem increases.

Chapter 7 presents an approach that is very different from all the other chapters in this part. It deals with models for only the tails of the distribution. We describe in this chapter various approaches for characterizing the behavior

of extreme events. In particular, we explain how to model the distribution of maxima over subsamples and the distribution of exceedances above a high threshold. In a multivariate context, we also highlight the important difference between asymptotic dependence and asymptotic independence and describe some non-parametric statistical measures for both.

1.3 Applications of non-Gaussian econometrics

Part III presents some examples of applications of the models described in Part II. All these applications are important and of routine use in the finance industry. Readers should be convinced, after reading this part, that the non-Gaussian models are not only indispensable in financial modeling, but they can also be very rewarding. Specifically, Chapter 8 deals with risk management and the Value-at-Risk (VaR) measure introduced by the Basel Accords. The industry benchmark, the RiskMetrics model, is based on normal distributions. We describe alternative techniques that are more appropriate for non-Gaussian distributions.

Chapter 9 is concerned with portfolio construction and asset allocation. Markowitz's mean-variance analysis is appropriate for Gaussian distributions or quadratic utility function only.¹ In the context of non-normal returns, this approach may not hold anymore. The main idea is that the investor's expected utility may be approximated as a function of mean, variance, but also of higher moments of the portfolio return. A rational investor would be averse to negative skewness and high kurtosis and in favor of positive skewness. We will also show in this chapter how downside risk constraint affects asset allocation decisions.

1.4 Option pricing with non-Gaussian distributions

Part IV deals with derivative assets and considers option pricing when the underlying asset return has a non-Gaussian distribution. The seminal contributions by Black and Scholes (1973) and Merton (1973) laid the foundation of pricing by no-arbitrage and, later, pricing by equivalent martingale measure. This model, which essentially assumes normality and time independence of price changes, has been shown for a long time to be unable to reproduce some well-known stylized facts such as the volatility smile or the term structure of volatilities. As for the modeling of asset returns, option pricing models have to incorporate the volatility clustering and the non-normality of the conditional distribution. Among these models, the most well-known is the stochastic volatility model and the models with jumps. Due to the mathematical

¹ Only the mean and variance terms are relevant when asset pricing is based on a quadratic utility function.

content of these option pricing models, we have provided five support chapters in the appendices in Part V, so that Part IV is not overly cluttered by mathematical abstraction. We have made every effort in ensuring that these two parts are accessible to non-mathematician readers.

In Chapter 10, we go through the fundamental building blocks of the Black-Scholes-Merton (BSM) model and use it as an example to introduce the key mathematical concepts such as Brownian motion and stochastic calculus (Chapter 13) and martingale and changing measure (Chapter 14). The BSM model is based on the underlying asset return having a normal distribution. Almost as soon as the importance of this model was recognized, the implied volatility smile was reported, indicating that the normality assumption is inconsistent with option price data. Nevertheless and despite the BSM pricing irregularities, the popularity of the BSM model survives even today.

From the BSM model it emerges that options can be priced using risk neutral densities (RND). Breeden and Litzenberger (1978) were the first to realize that the RND can be recovered from the option prices. Not only that RND can be used for pricing other, typically the less liquid and more exotic, derivatives written on the same underlying, but also researchers have found RND to be more informative and more responsive to news than the actual densities obtained from the prices of the underlying asset. But, the most important fact is that these empirically obtained RNDs are almost exclusively non-Gaussian. Chapter 11 covers a whole range of parametric and non-parametric methods for extracting RNDs. These techniques do not assume a specific model for the underlying asset. This is the reason why we called this chapter the “non-structural” approach to option pricing.

In the last chapter, Chapter 12, we put extensions of the BSM model into what we call “structural” option pricing models: structural in the sense that we now have a specific dynamic for the underlying asset price and sometimes a specific dynamic for the volatility also. This chapter is the most mathematically demanding and would require the support of Chapters 15, 16, and 17 for the mathematically less inclined readers. But this chapter also truly reflects the non-Gaussian nature of the underlying asset distributions in that jumps of “all shapes and sizes” are permitted at both return and volatility levels. At the time of writing, this is the cutting edge of option pricing as we know it!

2

Statistical Properties of Financial Market Data

In this chapter, we describe the empirical characteristics of some statistical measures of asset returns. More specifically, in Section 2.1, we provide some definitions that are useful for the rest of the book. Although most of the statistical measures used are well-known, emphasis is placed on critical issues concerning the estimation and the applications of these statistical measures. Then, we consider the various aspects of the return distribution. In Section 2.2, we show that the standard assumption of normality is rejected for stock market returns mainly because of asymmetry and tail thickness of the return distribution. In Section 2.3, we discuss time dependence and demonstrate the strong volatility clustering in financial markets. Section 2.4 discusses correlation across asset returns, its time-varying nature, and the implication of changing dependence on asset allocation. However, correlation is not a valid measure for dependence when returns are non-normal. Other measures for dependence are provided in Chapter 6. Finally, when the distribution appears non-normal, it is very useful and relevant to consider multivariate moments such as co-skewness and co-kurtosis. This is covered in Section 2.5.

Financial modeling deals with capturing the main characteristics of return distributions. The statistical properties of financial market data identified in this chapter will serve as the backbone for the later chapters. For instance, Chapter 3 describes some theoretical models explaining some of the observed statistical properties, and Chapters 4 and 5 will develop the tools necessary for modeling time-dependency and non-normality, in practice.

2.1 Definitions of returns

Although prices are what we observe in financial markets, most empirical studies focus on returns. The reason is that, in general, prices are non-stationary whereas returns are stationary. There are several return definitions, each of which produces a different set of properties for returns. We more specifically focus on simple returns and log-returns.

2.1.1 Simple returns

The one-period simple return for holding an asset is

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}} \quad \text{or} \quad P_t = P_{t-1} (1 + R_t),$$

where P_t is the price (including dividends) of the asset at time t , and R_t is the one-period simple return from time $t-1$ to t .

If we hold the asset for k periods from $t-k$ to t , we have the k -period simple return

$$R_t [k] = \frac{P_t - P_{t-k}}{P_{t-k}},$$

or

$$P_t = P_{t-k} (1 + R_t [k]) = P_{t-k} (1 + R_{t-k+1}) \times \cdots \times (1 + R_t),$$

where $R_t [k]$ is the k -period simple return from $t-k$ to t . Therefore, the relation between the simple one-period return and k -period return is non-linear

$$1 + R_t [k] = \prod_{j=0}^{k-1} (1 + R_{t-j}).$$

In some cases, if returns are small, we may use the approximation

$$R_t [k] \approx \sum_{j=0}^{k-1} R_{t-j},$$

but it is too crude in many applications.

On the other hand, the simple return of an N -asset portfolio is simply the weighted average of the individual simple returns of the N assets, denoted $R_{i,t}$, for $i = 1, \dots, N$. Let p denote the portfolio with investment weight ω_i on asset i and $\sum_{i=1}^N \omega_i = 1$. Then, the portfolio return is

$$R_{p,t} = \sum_{i=0}^N \omega_i R_{i,t}.$$

2.1.2 Log-returns

If a bank pays an annual interest of $R_t^{(m)}$, m times a year, the interest rate for each unit of investment is by definition $R_t^{(m)} / m$, and after one year the total value of the deposit is $\left(1 + \frac{1}{m} R_t^{(m)}\right)^m$. In the limiting case where the interest is cumulated continuously ($m \rightarrow \infty$), we have

$$\lim_{m \rightarrow \infty} \left(1 + \frac{1}{m} R_t^{(m)}\right)^m = e^{r_t}.$$

We deduce that r_t , the continuously compounded return (or log-return), is

$$r_t = \log(P_t) - \log(P_{t-1}) = p_t - p_{t-1},$$

where $p_t = \log(P_t)$ is the log-price.

A key advantage of the log-return is that the multiple-period return is simply the sum of one-period returns, so that

$$r_t[k] = \log(1 + R_t[k]) = \sum_{j=0}^{k-1} \log(1 + R_{t-j}) = \sum_{j=0}^{k-1} r_{t-j}.$$

Now, the log-return of a portfolio does not have the same convenient property as the case with simple return. Indeed, the portfolio log-return $r_{p,t}$ is

$$r_{p,t} = \log\left(\sum_{i=1}^N \omega_i e^{r_{i,t}}\right) \neq \sum_{i=1}^N \omega_i r_{i,t}.$$

But this problem is usually considered as minor in empirical applications.

2.1.3 Stylized facts

At first sight, there is no reason why commodity price, stock price, or exchange rate should behave in a particular fashion. However, many empirical studies¹ have identified a set of common features among financial data that are known as *stylized facts*. Cont (2001), in particular, provides a comprehensive survey of these stylized facts, which include:

1. *Fat tails*: The unconditional distribution of returns has fatter tails than that expected from a normal distribution. This means that, if we use the normal distribution to model financial returns, we will underestimate the number and magnitude of crashes and booms.
2. *Asymmetry*: The unconditional distribution is negatively skewed, suggesting that extreme negative returns are more frequent than extreme positive returns. The asymmetry and fat-tail phenomena persist even after adjusting for conditional heteroskedasticity (or changing volatility), meaning that the conditional distribution is also non-normal.
3. *Aggregated normality*: As the frequency of the returns lengthens, the return distribution gets closer to the normal distribution.
4. *Absence of serial correlation*: Returns generally do not display significant serial correlation, except at high frequency.

¹ See, for example, Bollerslev, Chou, and Kroner (1992), Pagan (1996), Shephard (1996), Campbell, Lo, and MacKinlay (1997), Cont (2001), and Gouriéroux and Jasiak (2001a).

5. *Volatility clustering*: Volatility of returns is serially correlated, suggesting that a large (positive or negative) return tends to be followed by another large (positive or negative) return. Among the proxies for volatility, absolute returns appear to be the most strongly serially correlated.
6. *Time-varying cross-correlation*: Correlation between asset returns tends to increase during high-volatility periods, in particular during crashes.

In the following sections, we will use financial market data to illustrate these stylized facts and to see how aggregation (or lengthening of data frequency) affects these stylized facts. The data consists of log-returns on four stock market indices; namely the Standard and Poor's 500 (SP500) from the US, the DAX 30 from Germany, the FTSE All Shares from the UK, and the Nikkei 225 from Japan. The sample period covers 2 January 1980 to 31 August 2004.

2.2 Distribution of returns

Figure 2.1 displays the log-returns on four stock market indices. Extreme negative returns appear to be more pronounced than extreme positive returns. Figure 2.2 displays the histogram of the log-returns, which show the empirical distributions as asymmetric and have tails that do not vanish as fast as those of the normal distribution. This is confirmed by the minimum and maximum returns reported in Table 2.1.

Table 2.1. Minimum and maximum of daily log-returns

	SP500	DAX	FT-SE	Nikkei
Minimum	-22.833	-13.71	-11.914	-16.135
Maximum	8.709	7.553	5.698	12.430

2.2.1 Moments of a random variable

In this book, we will typically use non-Gaussian distributions, which are characterized by moments that are higher than the second order.² Consider a continuous random variable X (say the log-return series) with *cumulative distribution function (cdf)*

² In this book, we use the words *Gaussian* and *normal* interchangeably. The normal distribution was studied as early as 1733 by de Moivre. Later on, Laplace and Gauss used it to model least-square errors. Hence, the Gaussian distribution is sometimes referred to as the Gauss-Laplace distribution.

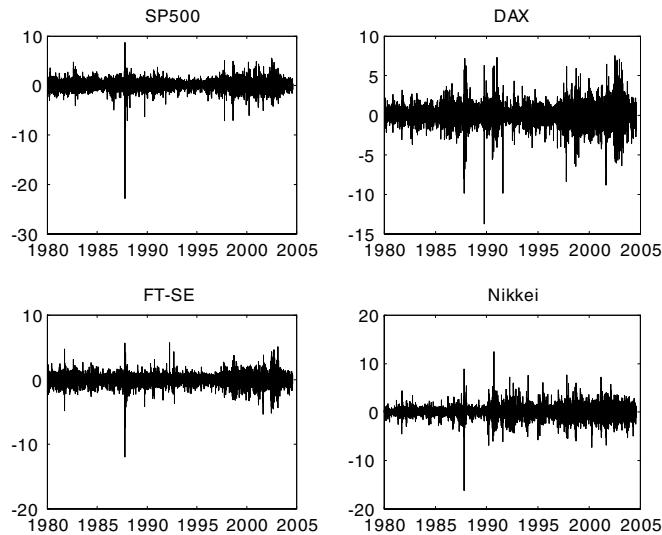


Fig. 2.1. Evolution of daily log-returns.

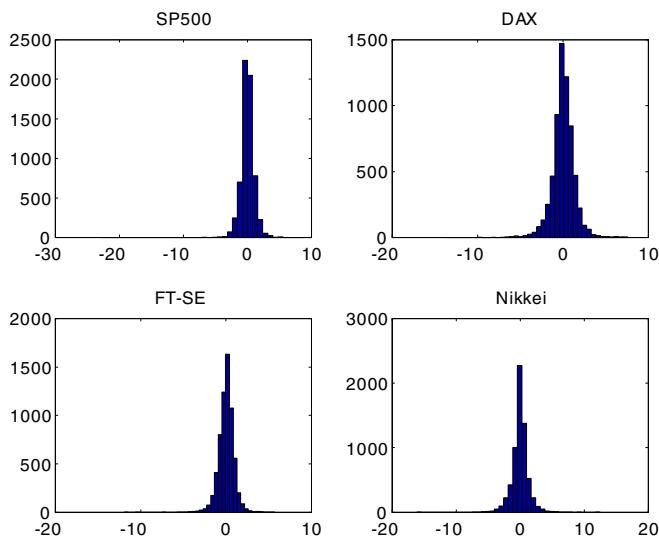


Fig. 2.2. Histogram of daily log-returns.

$$F_X(x) = \Pr[X \leq x] = \int_{-\infty}^x f_X(u) du,$$

where f_X is the *probability density function (pdf)*. The uncentered moments of X are defined as

$$m_k = E[X^k] = \int_{-\infty}^{\infty} x^k f_X(x) dx, \quad \text{for } k = 1, 2, \dots.$$

The first non-central moment $m_1 = E[X] = \mu$ is the mean of X .

Given the mean m_1 , the centered moments of X are defined as

$$\mu_k = E[(X - m_1)^k] = \int_{-\infty}^{\infty} (x - m_1)^k f_X(x) dx, \quad \text{for } k = 1, 2, \dots.$$

By construction, $\mu_1 = 0$. The second centered moment is the variance of X : $V[X] = \mu_2 = m_2 - m_1^2 = \sigma^2$. The third and fourth centered moments μ_3 and μ_4 are, respectively, the non-standardized skewness and kurtosis. The standardized skewness and kurtosis, or simply skewness and kurtosis, are defined as

$$s = Sk[X] = E\left[\left(\frac{X - \mu}{\sigma}\right)^3\right] = \frac{\mu_3}{\sigma^3},$$

$$\kappa = Ku[X] = E\left[\left(\frac{X - \mu}{\sigma}\right)^4\right] = \frac{\mu_4}{\sigma^4}.$$

Skewness, $Sk[X]$, depicts the asymmetry of the distribution. When it is negative, large negative realizations of X are more likely to appear than large positive realizations. *Kurtosis*, $Ku[X]$, captures the tail thickness of the distribution. A large kurtosis means that large realizations (either positive or negative) are more likely to be obtained than expected from a normal distribution.

Figures 2.3 and 2.4 illustrate how introducing asymmetry or fat tails would affect the shape of a distribution vis-à-vis the normal distribution. In Figure 2.3, the distribution with negative skewness leans to the right, since the long negative tail on the left has to be compensated by a greater probability mass on the shorter tail on the right. Note also that distributions with thicker tails have a thinner and higher peak in the center when compared with the normal distribution ³

³ In Figure 2.3, the asymmetric distributions are obtained from the so-called skewed Student t distribution with 10 degrees of freedom and skewness parameters -0.75 and 0.75 , respectively, for the long-dash line and the solid bold line (see Section 5.2.3). In Figure 2.4, the fat-tailed distributions are Student t distributions with 5 and 2 degrees of freedom, respectively, for the long-dash line and the solid bold line.

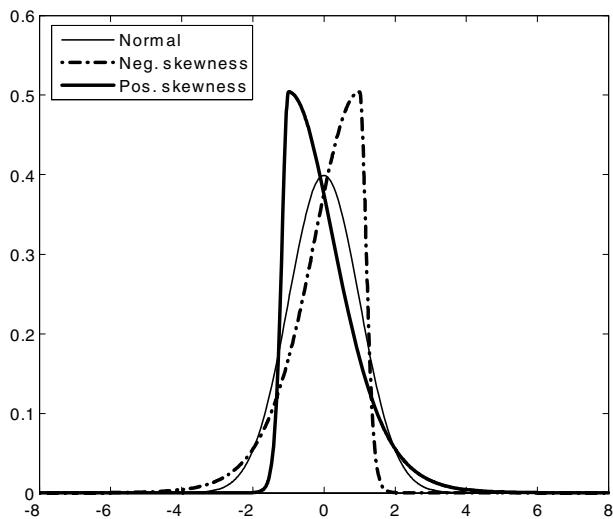


Fig. 2.3. Examples of asymmetric distributions.

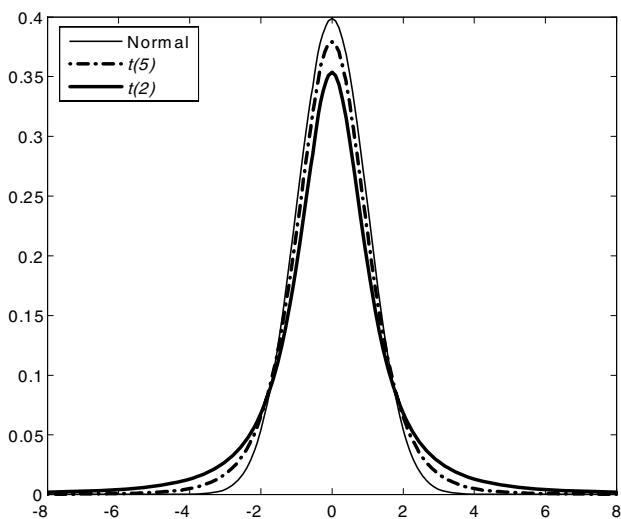


Fig. 2.4. Examples of distributions with fat tails.

The normal distribution has skewness always equal to zero and kurtosis always equal to 3. All further moments are either 0 (odd moments) or functions of μ and σ (even moments). Hence, a normal distribution is fully identified by its mean and variance. Also, as distributions are often compared with the normal distribution, the *excess kurtosis* (i.e., $\kappa - 3$) is often reported instead of the kurtosis itself. Most financial asset returns have kurtosis values greater than 3, and stock market return, in particular, is strongly featured by a negative skewness. Other financial series, such as interest rates, returns on commodities or on hedge funds, may have positive skewness. Finally, foreign exchange rates may have positive as well as negative skewness, depending on the way they are computed.

2.2.2 Empirical moments

Measures of location

Now, consider a time series of realized returns $\{r_t\}_{t=1}^T$. The most commonly used measures of location are *sample mean* \bar{r} and *median* m . The sample mean is calculated as⁴

$$\bar{r} = \hat{\mu} = \frac{1}{T} \sum_{t=1}^T r_t.$$

If $\{r_t\}_{t=1}^T$ has a symmetric distribution, then \bar{r} is the optimal measure of location. Median is defined as the 50th percentile of the sample. In other words, 50% of the sample has a value lower than $m = med[r]$, i.e.,

$$\Pr [r_t \leq m] = \Pr [r_t \geq m] = \frac{1}{2}.$$

Mean, as a measure for location, is sensitive to outliers. One erroneously recorded value could potentially move the mean away from the central part of the distribution. In contrast, the median is more robust against outliers, because it does not rely on the precise value of the realizations other than the median itself.

Measures of dispersion

Variance is a popular measure of dispersion, and it is the optimal measure for dispersion if returns have a normal distribution. The unbiased *sample variance*, $\hat{\sigma}^2$, can be calculated from a set of returns as⁵

⁴ In this book, we denote estimates with a hat $\hat{\cdot}$.

⁵ Note that $\hat{\sigma}^2$ is not the Maximum-Likelihood (ML) estimator of the variance under normality. For a normal distribution, the ML estimator for variance is $\tilde{\sigma}^2 = \frac{1}{T} \sum_{t=1}^T (r_t - \bar{r})^2$.

$$\hat{\sigma}^2 = \frac{1}{T-1} \sum_{t=1}^T (r_t - \bar{r})^2. \quad (2.1)$$

In finance, the standard deviation $\hat{\sigma}$, the square root of variance, is often used to mean volatility.⁶ Since the statistical properties of sample mean make it a very inaccurate estimate of the true mean, especially for small samples, taking deviations around zero instead of the sample mean as in (2.1) typically increases volatility forecast accuracy. There are methods for estimating volatility that are designed to exploit or reduce the influence of extremes (see Ball and Torous, 1983, Parkinson, 1980, or Garman and Klass, 1980). But these methods generally rely on the assumption of normality of returns, which is not likely in the case of financial market series.

Another interesting measure of dispersion is the *mean absolute deviation* (*MAD*). It is computed as

$$MAD = \frac{1}{T} \sum_{t=1}^T |r_t - \bar{r}|.$$

The *MAD* is more robust to outliers than the variance, because it involves lower powers of returns.

The *inter-quartile range* (*IQR*) is the difference between the 75th and 25th percentiles of the data. Since only the middle 50% of the data affects this measure, it is robust against outliers.

Higher moments

The sample (standardized) skewness and kurtosis are computed, respectively, as

$$\hat{s} = \frac{1}{T} \sum_{t=1}^T \left(\frac{r_t - \bar{r}}{\hat{\sigma}} \right)^3,$$

and

$$\hat{\kappa} = \frac{1}{T} \sum_{t=1}^T \left(\frac{r_t - \bar{r}}{\hat{\sigma}} \right)^4.$$

Under the assumption of normality, we have the following asymptotic distributions (see Kendall and Stuart, 1977)

$$\begin{aligned} \sqrt{T}(\hat{\mu} - \mu) &\Rightarrow \mathcal{N}(0, \sigma^2) & \sqrt{T}(\hat{\sigma}^2 - \sigma^2) &\Rightarrow \mathcal{N}(0, 2\sigma^4), \\ \sqrt{T}\hat{s} &\Rightarrow \mathcal{N}(0, 6) & \sqrt{T}(\hat{\kappa} - 3) &\Rightarrow \mathcal{N}(0, 24), \end{aligned}$$

⁶ While $\hat{\sigma}^2$ in equation (2.1) is an unbiased estimate of σ^2 , the square root of $\hat{\sigma}^2$ is a biased estimate of σ due to Jensen inequality. Cox and Rubinstein (1985) explain how this bias can be corrected assuming a normal distribution for r_t . However, in most cases, the impact of this adjustment is small.

where “ \Rightarrow ” denotes asymptotic convergence. Due to the large value of their asymptotic variances, sample skewness and sample kurtosis are informative only for large samples. These asymptotic distributions can be used to perform statistical tests on the distributional assumption of returns. This is the basis of the Jarque-Bera test for normality described in the next subsection.

All the statistics described above are often classified as summary statistics. Table 2.2 presents the summary statistics of the four stock market returns measured at different data frequencies. If we rank the returns based on the summary statistics, we note that the ranks produced by the mean are similar to those produced by the median. Similarly, using standard deviation or *MAD* to rank returns does not produce vastly different results. As expected, median and *MAD* are less dispersed because they are more robust against outliers. Annualized variance, derived from daily, weekly, and monthly data, is similar to the variance of annual data. Although the same is not true for skewness and kurtosis, there is no systematic pattern between, e.g., skewness of daily and annual returns. For instance, the daily skewness of SP500 is larger than the annual skewness, but the opposite is true for skewness of FT-SE.

2.2.3 Testing for normality

It has been noted for a long time that most financial asset returns are non-normal (Mandelbrot, 1963, and Fama, 1963). This non-normality is strongly featured in two statistical phenomena: (i) Extreme events occur more often than predicted by a normal distribution (Mandelbrot, 1963, Fama, 1963, Blattberg and Gonedes, 1974, Kon, 1984), and (ii) crashes occur more often than booms (Fama, 1965, Arditti, 1971, Simkowitz and Beedles, 1978, Singleton and Wingender, 1986). Phenomenon (i) corresponds to excess kurtosis or *fat tails*, whereas phenomenon (ii) is associated with negative skewness or *asymmetry*.

Following Mandelbrot (1963), two related issues have been addressed in the finance literature; does the normality assumption hold for asset returns? If not, to what extent are returns non-normal? In this chapter, we focus on unconditional normality. That is, for the moment we assume that returns distribution does not change through time. In Chapter 5, we will relax this assumption and allow distribution to vary conditionally to some distributional variable. Various tests for normality have been developed based on the moments or on the density function of the distribution, or properties of the ranked series. We will describe here tests that are based on properties of the original distribution.

Tests based on moments

The most widely used test in this category is due to Jarque and Bera (1980) and Bera and Jarque (1981) who rely on the fact that skewness and excess kurtosis are both equal to zero for the normal distribution. Jarque and Bera

Table 2.2. Summary statistics of log-returns

	SP500	DAX	FT-SE	Nikkei
Daily frequency (6,437 observations)				
Mean	0.036	0.032	0.035	0.008
Median	0.011	0.019	0.039	0.000
Std deviation	1.047	1.329	0.926	1.266
MAD	0.712	0.917	0.657	0.851
IQR	1.017	1.322	0.999	1.140
Ann. mean	9.105	7.943	8.870	2.047
Ann. std dev.	16.624	21.096	14.693	20.099
Skewness	-1.751	-0.472	-0.803	-0.158
Kurtosis	42.513	9.847	13.295	12.066
Weekly frequency (1,288 observations)				
Mean	0.180	0.159	0.176	0.041
Median	0.266	0.244	0.307	0.194
Std deviation	2.443	3.014	2.234	2.899
MAD	1.717	2.202	1.602	2.123
IQR	2.638	3.279	2.387	3.144
Ann. mean	9.377	8.253	9.176	2.150
Ann. std dev.	17.614	21.736	16.110	20.906
Skewness	-1.862	-0.588	-1.188	-0.214
Kurtosis	27.146	6.423	12.436	6.010
Monthly frequency (296 observations)				
Mean	0.769	0.683	0.737	0.167
Median	1.043	1.133	1.176	0.613
Std deviation	4.486	6.338	4.859	5.794
MAD	3.349	4.644	3.518	4.419
IQR	5.417	7.114	5.269	7.156
Ann. mean	9.231	8.202	8.845	2.006
Ann. std dev.	15.540	21.956	16.831	20.071
Skewness	-0.875	-0.849	-1.398	-0.418
Kurtosis	6.394	5.766	9.136	3.957
Annual frequency (25 observations)				
Mean	9.556	8.692	9.009	1.941
Median	10.923	11.980	13.039	7.594
Std deviation	14.598	27.419	15.819	22.914
MAD	10.931	21.854	11.555	18.914
IQR	18.821	39.918	15.979	30.030
Skewness	-0.831	-0.695	-1.412	-0.355
Kurtosis	3.299	3.047	4.420	2.374

(1980) show, for *iid* and normally distributed random variables, that the standardized skewness and excess kurtosis are asymptotically independent and have the following asymptotic distributions

$$\frac{\hat{s}}{\sqrt{6/T}} \xrightarrow[H_0]{\text{approx}} \mathcal{N}(0, 1) \quad \text{and} \quad \frac{\hat{\kappa} - 3}{\sqrt{24/T}} \xrightarrow[H_0]{\text{approx}} \mathcal{N}(0, 1).$$

Bera and Jarque's idea for testing normality is to use the *JB* test statistic defined as

$$JB = T \left[\frac{\hat{s}^2}{6} + \frac{(\hat{\kappa} - 3)^2}{24} \right],$$

which is asymptotically distributed as $\chi^2(2)$.

The most severe limitation of the *JB* test is that the asymptotic distribution holds only for very large samples. To take into account the finite sample bias, Doornik and Hansen (1994) propose an "omnibus" test for normality. First, they produce approximations for the finite sample distributions of skewness and kurtosis under the assumption of normality and with the additional assumptions that the kurtosis has a Gamma distribution and that $\hat{\kappa} > 1 + \hat{s}^2$.⁷ Using z_1 and z_2 to denote the finite-sample skewness and kurtosis, Doornik and Hansen (1994) show, with the normality assumption, that

$$\begin{aligned} \tilde{W} &= z_1^2 + z_2^2 \xrightarrow[\text{app}]{\text{approx}} \chi^2(2), \\ z_1 &= \frac{1}{\sqrt{\log(\omega)}} \log \left(g + \sqrt{1 + g^2} \right), \\ z_2 &= \left[\left(\frac{\chi}{2\alpha} \right)^{1/3} - 1 + \frac{1}{9\alpha} \right] \sqrt{9\alpha}, \end{aligned}$$

where $\xrightarrow[\text{app}]{\text{approx}}$ denotes approximate distribution and

$$\begin{aligned} g &= \hat{s}^2 \sqrt{\frac{\omega^2 - 1}{2} \frac{(T+1)(T+3)}{6(T-2)}}, \\ \omega^2 &= -1 + \sqrt{2(b_0 - 1)}, \\ \chi &= 2b_1 (\hat{\kappa} - 1 - \hat{s}^2), \\ \alpha &= b_2 + b_3 \hat{s}^2, \end{aligned}$$

and with the following correction factors for finite sample

⁷ An approximation of finite sample distribution of skewness can also be found in D'Agostino (1970).

$$\begin{aligned}
b_0 &= \frac{3(T^2 + 27T - 70)(T+1)(T+3)}{(T-2)(T+5)(T+7)(T+9)}, \\
b_1 &= \frac{(T+5)(T+7)(T^3 + 37T^2 + 11T - 313)}{12\tau}, \\
b_2 &= \frac{(T-2)(T+5)(T+7)(T^2 + 27T - 70)}{6\tau}, \\
b_3 &= \frac{(T-7)(T+5)(T+7)(T^2 + 2T - 5)}{6\tau}, \\
\tau &= (T-3)(T+1)(T^2 + 15T - 4).
\end{aligned}$$

A second drawback of the *JB* test is that the empirical skewness and kurtosis are computed for given values of mean and variance, both of which are subject to sampling errors. Richardson and Smith (1993) and Harvey (1995) independently propose joint estimation of the moments based on the GMM (Generalized Method of Moments) orthogonality conditions⁸

$$\begin{aligned}
e_{1,t} &= r_t - \bar{r}, \\
e_{2,t} &= (r_t - \bar{r})^2 - \sigma^2, \\
e_{3,t} &= (r_t - \bar{r})^3 / \sigma^3 - s, \\
e_{4,t} &= (r_t - \bar{r})^4 / \sigma^4 - \kappa.
\end{aligned}$$

Solving the orthogonality conditions by the method of moments yields a joint estimate $\hat{\theta}$ of the first four moments $\theta = (\mu, \sigma^2, s, \kappa)'$ and an associated covariance matrix Σ_T , which helps to remove the dependence on the asymptotic distribution of these moments. Under the null hypothesis of normality, we have $s = \kappa - 3 = 0$. The test is performed by a Wald test using the statistic

$$W = T \left(G \hat{\theta} \right)' \times (G \Sigma_T G)^{-1} \times \left(G \hat{\theta} \right),$$

where

$$G = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Under the null hypothesis of normality, W is distributed as a $\chi^2(2)$.

For the case of weakly dependent data, Bai and Ng (2001) provide the sampling distributions of the skewness and kurtosis estimators and propose a joint test for normality that takes this sampling distribution into account.

⁸ Richardson and Smith (1993) estimate all four moments and test whether skewness and excess kurtosis are jointly zero. On the other hand, Harvey (1995) performs the estimation under the null and uses the last two orthogonality conditions as over-identifying conditions.

Tests based on densities

The key test here is the Kolmogorov-Smirnov (KS) test that compares the empirical *cdf* $F_r(\cdot)$ of the returns with the *cdf* of the normal distribution (or any other assumed distribution) $F^*(\cdot)$ with parameter vector θ . Since the true distribution $F_r(\cdot)$ is unknown, it is approximated by the empirical *cdf* $G(\cdot)$ defined as

$$G(x) = \frac{1}{T} \sum_{t=1}^T 1_{\{r_t \leq x\}}.$$

The null hypothesis is $H_0 : G(\cdot) = F^*(x; \theta)$ for all x versus the alternative hypothesis $H_a : G(\cdot) \neq F^*(x; \theta)$ for at least one x . The KS test assumes that θ is known. If θ is unknown, the Lilliefors test, explained later on, should be used.

One of the simplest measures is the largest distance between the two functions $G(x)$ and $F^*(x; \theta)$. This is known as the *KS* test as suggested initially by Kolmogorov (1933). The test statistic is defined as

$$KS = \sup_{\{x\}} |F^*(x; \theta) - G(x)|.$$

In practice, this test is very easy to implement:

1. Sort the sample data by increasing order and denote the new sample $\{\tilde{r}_t\}_{t=1}^T$, with $\tilde{r}_1 \leq \dots \leq \tilde{r}_T$. Then, by construction, the empirical *cdf* is given by $G(\tilde{r}_t) = t/T$.
2. Evaluate the assumed *cdf* $F^*(\tilde{r}_t; \theta)$ for all values $\{\tilde{r}_t\}_{t=1}^T$. In the case of the normal distribution, we assumed that the mean μ and the variance σ^2 are known.
3. Compute the *KS* test statistic

$$KS = \sup_{\{t\}} \left| F^*(\tilde{r}_t; \theta) - \frac{t}{T} \right|.$$

The critical values of this statistic have been tabulated.

A drawback of the *KS* test is that the mean μ and the variance σ^2 are unknown and are subject to sampling errors when they are estimated empirically. In this case, we can apply the Lilliefors modified KS_L test, which involves, first of all, the estimation of $\hat{\mu}$ and $\hat{\sigma}^2$ from the data. It is then followed by steps 1 to 3 above, replacing (μ, σ^2) with $(\hat{\mu}, \hat{\sigma}^2)$ in step 2 to produce $F^*(\tilde{r}_t; \hat{\theta})$ in step 3, and a new set of critical values in the final stage. At the 95% confidence level, the new critical value for KS_L is $0.805/\sqrt{T}$.

Top panel of Figure 2.5 displays the empirical *cdf* $F^*(\tilde{r}_t; \hat{\theta})$ of the SP500 return, and the bottom panel presents the quantile plot (or QQ-plot) of the empirical and the normal distributions. The Lilliefors statistic KS_L corresponds to the maximum difference between the two curves (here, we have

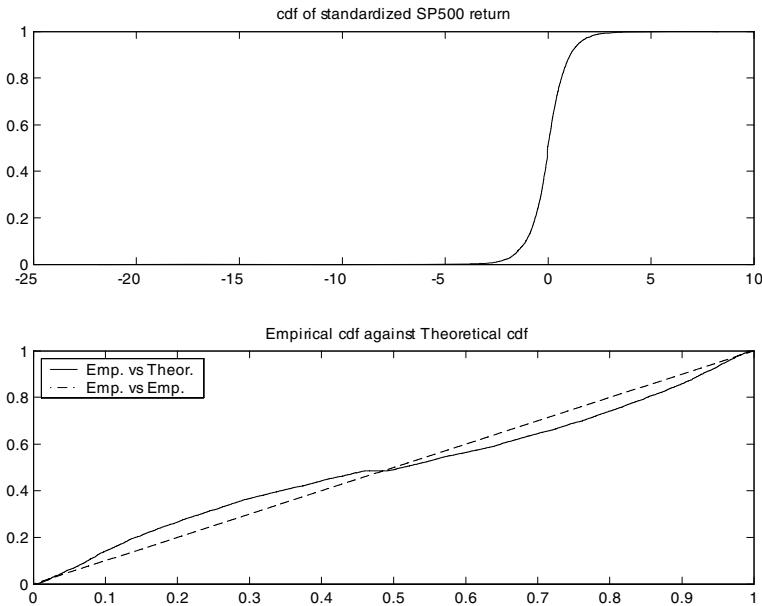


Fig. 2.5. Comparison of the empirical distribution of the SP500 with the normal distribution as the theoretical distribution.

$KS_L = 0.0692$). Given that the critical value is $0.805/\sqrt{6437} = 0.01$, the null hypothesis of normality is strongly rejected.

Table 2.3 presents various statistics to test normality as described above. The diagnosis on normality of returns is quite clear-cut. From daily to monthly frequencies, the different tests systematically reject the null hypothesis of normality. The rejection is due to both asymmetry and tail thickness of the distribution. In the case of annual returns, normality is not rejected. But in this instance, it is possible that the sample size is so small that the tests lack power and are simply unreliable.

2.3 Time dependency

Time dependency in asset returns can occur at several levels. We will discuss dependency in returns, volatility, and higher moments in turns in the following subsections. Time dependency is a crucial issue for several reasons: First, if return distribution is time dependent, then statistical tests using unconditional statistics and inferences derived thereof could be misleading. Second, if the time dependency can be fully exploited, it will help to produce better forecasts of level, volatility, and higher moments of returns. Such an improve-

Table 2.3. Normality tests for log-returns

	SP500	DAX	FT-SE	Nikkei
Daily frequency (6,437 observations)				
Skewness	-1.751	-0.472	-0.803	-0.158
$t(s)$	-57.332	-15.468	-26.302	-5.186
Kurtosis	42.513	9.847	13.295	12.066
$t(\kappa)$	647.054	112.127	168.593	148.461
JB	421965.45	12811.81	29115.35	22067.65
p-value	0.000	0.000	0.000	0.000
KS_L	0.069	0.070	0.052	0.083
Critical value	0.010	0.010	0.010	0.010
Weekly frequency (1,288 observations)				
Skewness	-1.862	-0.588	-1.188	-0.214
$t(s)$	-27.264	-8.606	-17.404	-3.138
Kurtosis	27.146	6.423	12.436	6.010
$t(\kappa)$	176.818	25.068	69.099	22.042
JB	32007.99	702.47	5077.56	495.69
p-value	0.000	0.000	0.000	0.000
KS_L	0.057	0.067	0.060	0.053
Critical value	0.022	0.022	0.022	0.022
Monthly frequency (296 observations)				
Skewness	-0.875	-0.849	-1.398	-0.418
$t(s)$	-6.135	-5.956	-9.799	-2.931
Kurtosis	6.394	5.766	9.136	3.957
$t(\kappa)$	11.900	9.696	21.512	3.355
JB	179.246	129.487	558.767	19.844
p-value	0.000	0.000	0.000	0.000
KS_L	0.051	0.062	0.079	0.060
Critical value	0.047	0.047	0.047	0.047
Annual frequency (25 observations)				
Skewness	-0.831	-0.695	-1.412	-0.355
$t(s)$	-1.662	-1.389	-2.825	-0.710
Kurtosis	3.299	3.047	4.420	2.374
$t(\kappa)$	0.299	0.047	1.420	-0.626
JB	2.850	1.932	9.996	0.895
p-value	0.085	0.155	0.001	0.327
KS_L	0.099	0.093	0.189	0.086
Critical value	0.161	0.161	0.161	0.161

ment is obviously very important for many finance applications such as risk management and asset allocation.

2.3.1 Serial correlation in returns

Here, the null hypothesis is that the first p serial correlations of returns are equal to 0, $H_0 : \rho_1 = \dots = \rho_p = 0$, where the correlation of order j is estimated by

$$\hat{\rho}_j = \frac{\sum_{t=j+1}^T (r_t - \bar{r})(r_{t-j} - \bar{r})}{\sum_{t=1}^T (r_t - \bar{r})^2} \quad \text{for } 0 \leq j < T-1.$$

A simple test of H_0 could be based on the Ljung-Box Q statistic

$$Q_p = T(T+2) \sum_{j=1}^p \frac{1}{T-j} \hat{\rho}_j^2.$$

Under the null hypothesis of no serial correlation, the Q_p statistic is asymptotically distributed as $\chi^2(p)$. A common practice is to test H_0 repeatedly using several choices of p .

2.3.2 Serial correlation in volatility

To test for time dependency in volatility, we need a time-varying measure of volatility. There are at least two possible ways to approach this; first using mean-adjusted squared returns and secondly using absolute returns. Assume that returns have the following dynamics

$$r_t = \mu + \varepsilon_t, \quad \text{with} \quad \varepsilon_t = \sigma_t z_t,$$

where μ is the constant mean, ε_t the mean adjusted returns, σ_t the time-varying volatility, and z_t is an $\mathcal{N}(0, 1)$ innovation. Then with information set \mathcal{F}_{t-1} at time $t-1$,

$$E[\varepsilon_t^2 | \mathcal{F}_{t-1}] = \sigma_t^2 E[z_t^2 | \mathcal{F}_{t-1}] = \sigma_t^2$$

because z_t^2 is distributed as $\chi^2(1)$. Therefore, ε_t^2 can be viewed as a proxy for the volatility at time t . Alternatively, omitting μ for the moment, we have $r_t \sim \mathcal{N}(0, \sigma_t^2)$ and

$$E[|r_t|] = \sigma_t \sqrt{2/\pi}.$$

Consequently, $|r_t|/\sqrt{2/\pi}$ is a proxy for σ_t . It should be noted however that these two measures are noisy estimates of conditional volatility. See Chapter 4 for more detail and more sophisticated measures of conditional volatility.

Given these crude measures of volatility, the test for serial dependence in volatility can be executed in the same manner as the test for returns in the previous subsection using Ljung-Box Q statistic. Non-zero serial correlation in squared or absolute returns is evidence of volatility time dependence. Table 2.4 reports the Ljung-Box statistic for returns, squared returns, and absolute returns for $p = 10$ (and $p = 5$ in the case of annual data). The corresponding critical values are 18.307 for $p = 10$ (and 11.071 for $p = 5$). The results indicate that returns are serially correlated when they are sampled at the daily frequency but not at other frequencies. Squared and absolute returns are more strongly correlated, at least at the daily and weekly frequencies.

Table 2.4. Ljung-Box test statistics of log-returns, squared returns, and absolute returns

Frequency	SP500	DAX	FT-SE	Nikkei
Returns				
Daily	21.205	29.882	69.534	29.347
Weekly	15.178	15.655	23.265	12.306
Monthly	9.537	5.412	6.660	7.418
Annual	11.727	4.946	2.838	2.280
Squared returns				
Daily	438.200	2327.700	2923.500	776.500
Weekly	30.947	350.231	138.645	131.836
Monthly	5.653	21.260	3.592	52.751
Annual	1.850	1.530	8.154	8.345
Absolute returns				
Daily	1784.700	4411.200	2780.600	2630.100
Weekly	194.380	537.175	190.139	238.459
Monthly	17.955	28.815	10.764	42.842
Annual	2.319	4.217	10.063	9.327

When sampled at monthly frequency, squared and absolute returns of DAX and Nikkei are still serially correlated reflecting the strong time dependence of volatility.

Indeed, the strong time dependence in volatility is an important feature in financial market modeling and gives rise to the birth of a huge class of conditional heteroskedasticity models (see Chapter 4). In fact, the autocorrelation coefficient of volatility proxies is significantly greater than zero even for very long horizons. To illustrate this point, we plot in Figure 2.6 the autocorrelogram of squared returns for up to 100 lags.⁹ The figure shows that the serial correlation of DAX squared returns is significant up to lag 100. This phenomenon is now referred to as the long memory of volatility.

There has also been some debate as to whether volatility has a unit root (Perry, 1982, and Pagan and Schwert, 1990). That is, if

$$\sigma_t = \varphi\sigma_{t-1} + \varepsilon_t,$$

the issue concerns whether or not φ is indistinguishable from 1. Much of these discussions took place before the new concept of realized volatility developed. If we apply unit root test to the volatility series computed using intraday data, we categorically reject the null hypothesis of a unit root even at the 1% level (see Section 4.8).

⁹ The 90% confidence interval reported in the figure is computed assuming that the estimator $\hat{\rho}_j$ is asymptotically normal $N(0, V)$, with V estimated by $\hat{V} = 1 + 2 \sum_{i=1}^{j-1} \hat{\rho}_i^2$.

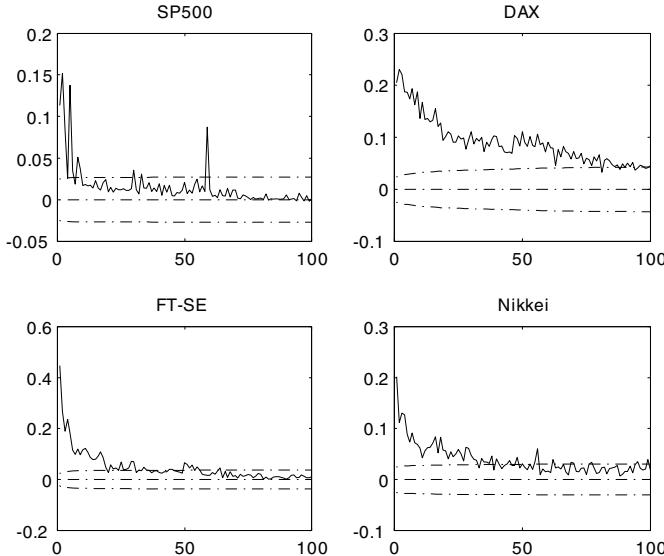


Fig. 2.6. Autocorrelogram for daily squared returns.

2.3.3 Volatility asymmetry

Another important feature of financial market volatility is that it is more affected by negative returns than by positive returns. To illustrate the point, we perform the following regressions

$$\varepsilon_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \varepsilon_{t-1}^2 \times 1_{\{\varepsilon_{t-1} < 0\}}, \quad (2.2)$$

or

$$|\varepsilon_t| = \omega + \alpha |\varepsilon_{t-1}| + \beta |\varepsilon_{t-1}| \times 1_{\{\varepsilon_{t-1} < 0\}}, \quad (2.3)$$

where α measures the direct effect of past returns, and β captures the additional impact of negative return shocks. Table 2.5 reports the regression coefficients with their standard errors reported in parentheses. First note that α and β estimates are of similar magnitude for the case of absolute returns but not for the case of squared returns. For absolute returns, α ranges between 0.12 and 0.18, whereas β has a range of similar magnitude between 0.11 and 0.14. This means that, on average, a negative return shock has twice as much impact on volatility as a positive return shock. Black (1976) and Christie (1986) call this leverage effect, relating to the fact that when equity value decreases, leverage and, hence, risk and volatility increase. Campbell and Hentschel (1992) interpret the different responses as a stronger impact of bad news than good news. This asymmetric (or leverage) effect in volatility will be investigated further in Chapter 4.

Table 2.5. Parameter estimates of volatility asymmetry regressions (2.2) and (2.3)

	SP500	DAX	FT-SE	Nikkei
Squared returns				
ω	1.434 (0.068)	0.540 (0.035)	1.310 (0.068)	2.955 (0.300)
α	0.134 (0.021)	0.237 (0.024)	0.079 (0.019)	0.061 (0.048)
β	0.105 (0.025)	0.256 (0.027)	0.205 (0.024)	-0.003 (0.050)
Absolute returns				
ω	0.694 (0.016)	0.502 (0.011)	0.669 (0.015)	0.931 (0.022)
α	0.178 (0.015)	0.180 (0.015)	0.154 (0.015)	0.127 (0.017)
β	0.135 (0.017)	0.120 (0.017)	0.123 (0.018)	0.114 (0.019)

2.3.4 Time-varying higher moments

One of the main features of asset returns is the volatility clustering, suggesting that the volatility of returns may vary over time depending on the arrival of new information. We may then ask if other moments of the distribution are also time-varying. To illustrate this issue, we plot in Figure 2.7 the evolution of the first four moments of the daily SP500 return, computed using a rolling window of 126 working days, which is about 6 months of data. The figure indicates that, beyond the October 1987 crash, there has been a recent upward trend in the evolution of the volatility (measured as the standard deviation of the series). Skewness and kurtosis are clearly affected by large jumps. The median value for the skewness is -0.11, and the median value for the kurtosis is 3.62. These values suggest that, over shorter sample periods, the normality assumption may hold. But, over a longer sample period, return is prone to large negative shocks that have a severe impact on the skewness and kurtosis estimates.

2.4 Linear dependence across returns

In this section, we turn to some stylized facts concerning the joint distribution of returns, which can be analyzed by separating the marginal distributions and the dependence structure. The marginal distributions are the univariate distributions that characterize the individual series. The dependence structure, on the other hand, describes how the individual series are related to each other, usually without the influence of the univariate distribution. Here, we focus on the dependence structure.

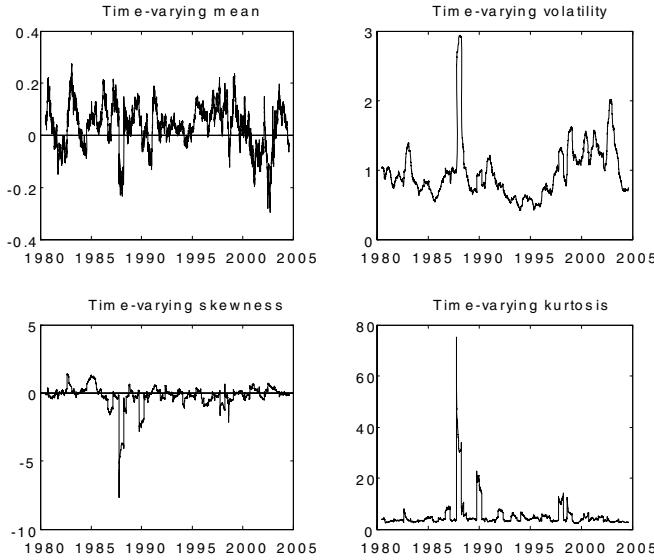


Fig. 2.7. Time-varying moments for the daily SP500 return, computed over rolling windows of 6 months.

2.4.1 Pearson's correlation coefficient

The most frequently used measure of dependence is the *linear correlation*, or Pearson's correlation

$$\rho[X, Y] = \frac{\text{Cov}[X, Y]}{\sqrt{V[X] V[Y]}},$$

where $\text{Cov}[X, Y]$ is the covariance between X and Y . The Pearson's correlation satisfies the constraint $-1 \leq \rho[X, Y] \leq 1$, and has as special cases: $\rho[X, X] = 1$ and $\rho[X, Y] = \rho[Y, X]$. Since $\rho[\alpha X + \beta, \gamma Y + \delta] = \text{sign}(\alpha\gamma) \rho[X, Y]$, correlation is invariant under strictly increasing linear transformations. When $|\rho[X, Y]| = 1$, we have perfect correlation (or perfect linear dependence), because it means that Y can be written as $Y = \alpha X + \beta$. When $\rho[X, Y] = 0$, the two series are uncorrelated. It is worth emphasizing that “no correlation” does not necessarily mean “independence”. Correlation is a good measure for dependence if returns have elliptical distributions, which include normal and Student t distributions.¹⁰ Correlation may not have any relationship with dependence for non-elliptical distributions.

A consistent estimator of the Pearson's correlation between two series $\{x_t\}_{t=1}^T$ and $\{y_t\}_{t=1}^T$ is

¹⁰ Elliptical distributions are defined and analyzed in Section 6.2.

$$\hat{\rho} = \frac{\sum_{t=1}^T (x_t - \bar{x})(y_t - \bar{y})}{\sqrt{\sum_{t=1}^T (x_t - \bar{x})^2 \sum_{t=1}^T (y_t - \bar{y})^2}},$$

where \bar{x} and \bar{y} are the sample mean of x and y , respectively. The asymptotic distribution of $\hat{\rho}$ is

$$\sqrt{T}(\hat{\rho} - \rho) \Rightarrow \mathcal{N}(0, 1).$$

Table 2.6 reports the correlation matrix for returns on the four stock market indices measured at various frequencies. All correlation coefficients reported in the table are positive. Moreover, the lower the data frequency, the higher the correlation coefficient estimate. Figure 2.8 presents the evolution of the correlation coefficients computed using a rolling period of 6 months. Except for the pair SP500 vs. Nikkei, all correlation coefficients exhibit an upward trend.

2.4.2 Test for equality of two correlation coefficients

There have been a lot of discussions in the literature on how the correlation changes when the financial markets are “agitated”. This is an important issue in portfolio diversification. If correlation increases during turbulent periods and stock market crashes, then the benefits of diversification disappear when they are most needed. Previous studies on changing correlation estimated it

Table 2.6. Correlation matrix of log-returns

	SP500	DAX	FT-SE	Nikkei
Daily frequency				
SP500	1	0.358	0.369	0.109
DAX	0.358	1	0.564	0.236
FT-SE	0.369	0.564	1	0.268
Nikkei	0.109	0.236	0.268	1
Weekly frequency				
SP500	1	0.535	0.548	0.301
DAX	0.535	1	0.592	0.339
FT-SE	0.548	0.592	1	0.357
Nikkei	0.301	0.339	0.357	1
Monthly frequency				
SP500	1	0.598	0.714	0.413
DAX	0.598	1	0.623	0.373
FT-SE	0.714	0.623	1	0.412
Nikkei	0.413	0.373	0.412	1
Annual frequency				
SP500	1	0.711	0.798	0.405
DAX	0.711	1	0.753	0.474
FT-SE	0.798	0.753	1	0.629
Nikkei	0.405	0.474	0.629	1

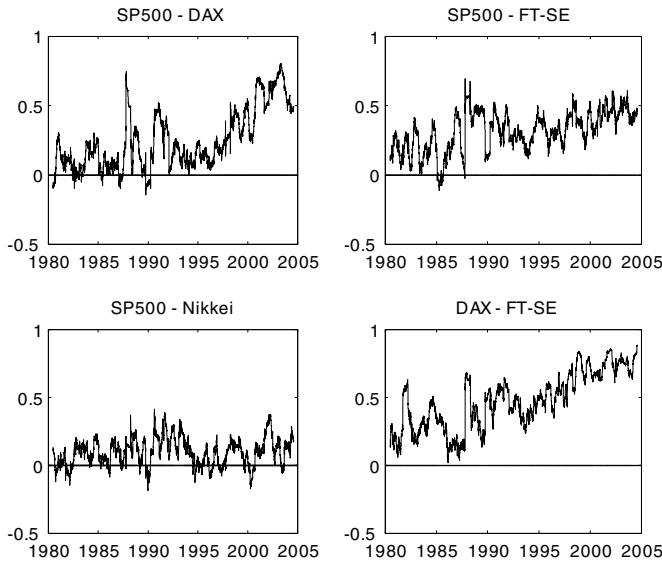


Fig. 2.8. Time-varying correlations between daily stock market return computed using a rolling window of 6 months.

either from different sample periods or conditional on one of the return series exceeding a threshold. The first type of studies includes Kaplanis (1988) and Ratner (1992) who find the correlation matrix to remain constant before and after the 1987 stock market crash. On the contrary, Koch and Koch (1991) find that correlation increases through time, but King, Sentana, and Wadhwani (1994) argue that the 1987 stock market crash is the main cause of the increase in correlation. However, Boyer, Gibson, and Loretan (1997), Loretan and English (1999), and Forbes and Rigobon (2002) show that the correlation coefficient between two series is biased, when it is computed conditionally on one of the series exceeding a threshold. Therefore, even when the breaking date is assumed to be known (corresponding to a well-established crash, for instance), unconditional correlation estimates have to be corrected before any testing procedure. As a consequence, a higher correlation coefficient could be a spurious outcome caused by higher volatility. To test for a change in the correlation relationship, it is necessary (i) to identify a data generating process that allows for the possibility of structural changes, (ii) to estimate the model parameters, and (iii) to test for changing correlations and possibly other structural breaks.

The above discussions highlight the complexities involved in estimating and testing correlation coefficients, the conditional nature of the distributional variables, and the complications caused by structural breaks. We will address

these issues in a greater detail in Chapter 6. Here, we present the simplest case when there is no structural break and changing distributional characteristics other than possibly the correlation relationship. Before executing the test on correlation, it is also a good practice to first filter out the time dependence of the individual series. In some cases, the removal of common factor dependence is helpful in identifying the true dependence structure.

The test of the equality of the correlation coefficients of two non-overlapping periods T_1 and T_2 relies on the asymptotic distribution of the correlation coefficient. Since the distribution becomes unstable as $\rho \rightarrow 1$, Fisher (1915) introduces the z -transformation and recommends that the test be conducted on $z(\hat{\rho})$ instead of $\hat{\rho}$, where $\hat{\rho}$ is the estimated correlation coefficient. Formally,

$$z(\hat{\rho}) = \frac{1}{2} \log \left(\frac{1 + \hat{\rho}}{1 - \hat{\rho}} \right).$$

Under the assumption that two samples are drawn from two independent bivariate normal distributions with the same correlation coefficient, the difference between the estimated $z(\hat{\rho})$ for the two samples converges to the normal distribution $\mathcal{N}(0, 1/(T_1 - 3) + 1/(T_2 - 3))$.

2.4.3 Test for equality of two correlation matrices

Jennrich (1970) develops a test based on the normalized difference between two correlation matrices. Let R_1 and R_2 denote the correlation matrices for two independent subsamples of equal size $T_1 = T_2 = T$. Then the normalized difference is

$$Z = \sqrt{\frac{T}{2}} R^{-1} (R_1 - R_2),$$

where $R = \frac{1}{2} (R_1 + R_2)$ is the average correlation matrix over the two subsamples and for $R = (\hat{\rho}_{ij})$, the inverse is $R^{-1} = (\hat{\rho}^{ij})$, where $\hat{\rho}^{ij}$ denotes the components of the inverse of the correlation matrix. Then the test statistic χ^2 is defined as

$$\chi^2 = \frac{1}{2} \text{tr}(Z^2) - \text{diag}(Z)' S^{-1} \text{diag}(Z), \quad (2.4)$$

where $S = (\delta_{ij} + \hat{\rho}^{ij} \hat{\rho}_{ij})$ with

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{otherwise,} \end{cases}$$

$\text{diag}(A)$ denotes the diagonal of the square matrix A in a column form and $\text{tr}(B)$ denotes the trace of B . The Jennrich test statistic χ^2 has an asymptotic chi-square distribution with $p(p - 1)/2$ degrees of freedom, where p is the number of variables or the dimension of the correlation matrix.

2.5 Multivariate higher moments

The ability to compute *multivariate higher moments* is important in some financial applications such as asset allocation. In this section, we will first define multivariate higher moments (or co-moments of higher order). Then, we will describe how the moments of a portfolio return can be expressed in a very convenient way.

2.5.1 Multivariate co-skewness and co-kurtosis

If there are n assets in the portfolio, then the (n, n^2) co-skewness matrix is defined as

$$\begin{aligned} M_3 &= E[(r - \mu)(r - \mu)' \otimes (r - \mu)'] = \{s_{ijk}\}, \\ s_{ijk} &= E[(r_i - \mu_i)(r_j - \mu_j)(r_k - \mu_k)] \quad \text{for } i, j, k = 1, \dots, n, \end{aligned}$$

where r_i denotes the individual asset return i , μ is the mean, and \otimes is the Kronecker product.¹¹ For instance, in the case of $n = 3$ assets, the resulting $(3, 9)$ co-skewness matrix is

$$\begin{aligned} M_3 &= \left[\begin{array}{ccc|ccc|ccc} s_{111} & s_{112} & s_{113} & s_{211} & s_{212} & s_{213} & s_{311} & s_{312} & s_{313} \\ s_{121} & s_{122} & s_{123} & s_{221} & s_{222} & s_{223} & s_{321} & s_{322} & s_{323} \\ s_{131} & s_{132} & s_{133} & s_{231} & s_{232} & s_{233} & s_{331} & s_{332} & s_{333} \end{array} \right] \\ &= [S_{1jk} \ S_{2jk} \ S_{3jk}], \end{aligned}$$

where S_{1jk} denotes the (n, n) matrix with elements $\{s_{1jk}\}_{j,k=1,2,3}$.

The (n, n^3) co-kurtosis matrix is defined as

$$\begin{aligned} M_4 &= E[(r - \mu)(r - \mu)' \otimes (r - \mu)' \otimes (r - \mu)'] = \{\kappa_{ijkl}\}, \\ \kappa_{ijkl} &= E[(r_i - \mu_i)(r_j - \mu_j)(r_k - \mu_k)(r_l - \mu_l)] \quad \text{for } i, j, k, l = 1, \dots, n. \end{aligned}$$

Again, with the $n = 3$ example, the $(3, 27)$ co-kurtosis matrix is

$$M_4 = [K_{11kl} \ K_{12kl} \ K_{13kl} \ | \ \dots \ | \ K_{31kl} \ K_{32kl} \ K_{33kl}],$$

where K_{11kl} denotes the (n, n) matrix with elements $\{\kappa_{11kl}\}_{k,l=1,2,3}$.

These notations are extensions of the covariance matrix M_2 . They have been used by Harvey et al. (2002), Prakash, Chang, and Pactwa (2003), and Athayde and Flôres (2004). It should be noted that, because of certain symmetries, not all the elements of these matrices have to be computed. For example, the dimension of the covariance matrix is (n, n) , but only $n(n + 1)/2$

¹¹ If $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$, then $A \otimes B = \begin{bmatrix} a & 2a & b & 2b \\ 3a & 4a & 3b & 4b \\ c & 2c & d & 2d \\ 3c & 4c & 3d & 4d \end{bmatrix}$. In general, if A has dimension $(m \times n)$ and B has dimension $(p \times q)$, $A \otimes B$ has dimension $(mp \times nq)$.

elements have to be computed. Similarly, the co-skewness and co-kurtosis matrices have dimensions (n, n^2) and (n, n^3) , but only involve, respectively, $n(n+1)(n+2)/6$ and $n(n+1)(n+2)(n+3)/24$ different elements.^{12,13}

2.5.2 Computing moments of portfolio returns

Now, using these notations from the previous subsection, moments of the portfolio return can be computed in a very tractable way. For a given portfolio weight vector α , moments of the portfolio return are

$$\begin{aligned}\mu_p &= \alpha' \mu & \sigma_p^2 &= \alpha' M_2 \alpha \\ s_p &= \alpha' M_3 (\alpha \otimes \alpha) & \kappa_p &= \alpha' M_4 (\alpha \otimes \alpha \otimes \alpha).\end{aligned}$$

The second- to fourth-order moments of the portfolio return can also be written in a slightly different form

$$\begin{aligned}\sigma_p^2 &= E \left[\sum_{i=1}^n \alpha_i (r_i - \mu_i) (r_p - \mu_p) \right] = \alpha' \Sigma_p, \\ s_p &= E \left[\sum_{i=1}^n \alpha_i (r_i - \mu_i) (r_p - \mu_p)^2 \right] = \alpha' S_p, \\ \kappa_p &= E \left[\sum_{i=1}^n \alpha_i (r_i - \mu_i) (r_p - \mu_p)^3 \right] = \alpha' K_p,\end{aligned}$$

where

$$\begin{aligned}\Sigma_p &= E [(r_i - \mu_i) (r_p - \mu_p)] = M_2 \alpha, \\ S_p &= E [(r_i - \mu_i) (r_p - \mu_p)^2] = M_3 (\alpha \otimes \alpha), \\ K_p &= E [(r_i - \mu_i) (r_p - \mu_p)^3] = M_4 (\alpha \otimes \alpha \otimes \alpha),\end{aligned}$$

are, respectively, the $(n, 1)$ vectors of covariances, co-skewness, and co-kurtosis between the asset returns and the portfolio return.¹⁴

¹² For $n = 3$, we have to compute 6 out of 9 elements for the covariance matrix, 10 out of 27 elements for the co-skewness matrix, and 15 out of 81 elements for the co-kurtosis matrix.

¹³ There are other ways of calculating co-moments. In particular, it may be possible to improve the efficiency of co-moment measures by specifying the joint distribution. For instance, Simaan (1993) assumes all returns depend on a random factor with a non-spherical distribution. When no specific structure for the conditional joint distribution of returns is assumed, the measures of co-moment may suffer from inefficiency. See also Chapter 5.

¹⁴ The notations Σ_p , S_p , and K_p are directly related to the notions of systematic risk and are widely used in the literature on higher-moment CAPM, see Kraus and Litzenberger (1976), Ingersoll (1987), Hwang and Satchell (1999), Jurczenko and Maillet (2001).

3

Functioning of Financial Markets and Theoretical Models for Returns

In various chapters of this book, we will consider modeling the temporal and cross-sectional behavior of asset returns. As already documented in the previous chapter, asset returns follow a non-normal distribution. As an intermediate step between the description of stylized features and the modeling of asset returns, it is worth considering some historical developments and theoretical considerations concerning asset price behavior. As we will discover in this chapter, besides the radical assumptions of Mandelbrot (1963), who postulated that asset returns are distributed according to the Lévy stable distribution, information arrival to financial markets will be seen as a major force that generates non-normality. Information arrival is intimately related to the way economic agents interact. This chapter aims at providing to the reader elements on the working of financial markets at an individual level. The area of finance that deals with this issue is called *market microstructure*. Given the importance of the market microstructure, it seems relevant to provide some background information. The interested reader may find more details on the organization of financial markets in Demarchi and Foucault (2000), on the market microstructure theory in O'Hara (1995), and a recent survey of the market microstructure literature in Biais, Glosten, and Spatt (2002).

In the second part of the chapter, we turn to some theoretical models that may help understanding and explaining the empirical characteristics of returns. The main stylized facts that have been identified are the non-normality of returns and the clustering of volatility. Early evidence of return non-normality is by Mandelbrot (1963) and Fama (1963, 1965). To explain why aggregated returns may have a non-normal distribution despite the central limit theorem, Mandelbrot suggests that returns may be driven by a stable distribution, because it does not converge to the normal distribution by aggregation (Section 3.2). Subsequently, several studies have proposed explanations based on the functioning of financial markets and therefore they belong to the market microstructure literature. For instance, the subordination model of Clark (1973), developed in Section 3.3, suggests that the dynamic of returns is in fact driven by a latent variable, namely the information arrival. If the

information arrival varies over time, then the distribution of returns is likely to be non-normal. This model has been extended in several directions. In particular, the latent variable may drive other key variables on financial markets, such as volume or number of trades. Therefore, some implications of this model on the joint dynamic of these variables can be tested empirically. Tauchen and Pitts (1983) focus on the dynamic of asset returns and trading volume (Section 3.4). This theoretical model is able to explain both the observed non-normality and volatility clustering. However, empirical tests of the model are not completely convincing, and several extensions of the initial model have been proposed. Easley, Kiefer, and O'Hara (1997) take another route and investigate the relation between prices and market marker's quotes in quote-driven market (Section 3.5). This model provides an interesting way to characterize and estimate the dynamic of asset returns.

Study of these models reveals contrasting approaches. In the Tauchen and Pitts (1983) model, heterogeneity of beliefs will lead to trading volume. On the other hand, the Easley, Kiefer, and O'Hara (1997) model does not deal at all with volume. It starts with an exogenous trade flow from which information is gleaned. As such, these models provide complementary insights. At the same time, it appears as somewhat frustrating that there exists no uniform setting that could give a complete description of the order book, the volume and trade intensity, as well as of the way information gets generated. In the same line of arguments, the empirical microstructure literature recognizes the importance of the trade size. The larger the trade, the more it will affect prices. Since we do not consider the impact of trades in the latter chapters of this book, it seems appropriate to describe in this chapter the recent developments on important issues, such as how to liquidate (or purchase) a large portfolio. We will discuss how to change the weights of a portfolio while minimizing the impact on the market in the last section of this chapter.

3.1 Functioning of financial markets

3.1.1 Organization of financial markets

Many different types of assets exist (stocks, currencies, derivatives, bonds, swaps, ...) and almost as many trading systems are available. There are many small details that will distinguish one market form another. Here, we do not consider the small details but provide only some global features characterizing markets.

We may distinguish, at a first level, quote-driven markets from order-driven markets. In a *quote-driven market*, an individual of some financial institution, called a dealer, will post, either by voice, upon request, or permanently, via some electronic information system, a price at which she is willing to buy (the bid price) or at which she is willing to sell (the ask price) a given quantity of the asset. If an individual, i.e., a public investor, wishes to purchase or to sell

the asset, she will have to call an agent, called a broker, who will in turn call a dealer who will eventually execute a trade.

The quotes that the dealer posts may be either firm or indicative. In the case of a firm order, another dealer may trade at the posted price up to certain quantities. Should the quantities exceed a given threshold, the dealers would communicate directly and agree on prices according to the quantity proposed. If the quotes are only indicative, a dealer will call for confirmation. Typically, various dealers are interconnected by permanently monitored communication lines. Once a trade has been made by voice, it gets confirmed by a telefax. Certain parameters of the trade (for instance, just the price) will get recorded by some information system (such as Bloomberg and Reuters). In case of disagreement, the monitoring of the phone calls allows a reconstruction of the orders.

In an *order-driven market*, each investor can trade with each other by entering their orders into an electronic trading system. The orders can be of various types. A so-called market order would lead to an immediate purchase or sale at the best available prices. A limit order, on the other hand, would indicate the willingness of some investor to trade a certain amount at some given price. It is possible, in many markets, to provide additional information such as on the length of time during which the order should exist. In such a case, the limit order will be called a stop order. The cumulation of various limit orders will yield the order book. Figure 3.1 represents a possible state of the order book. The horizontal axis represents amounts that are to be traded and the vertical axis represents prices. The upper (discontinuous) curve is called the ask side of the order book. This represents the prices and quantities at which investors would like to sell their stocks. Other traders would like to purchase stocks. This demand for stocks is represented by the lower side of the order book. In the figure, a given symbol corresponds to a given trader. The amount of symbols indicates the quantities that a trader is willing to buy or sell. Prices may only take discrete values as it is also the case in practice. For convenience, we assume that positive prices correspond to sell orders, negative prices to buy orders.¹ Inspection of the various symbols indicates that it is perfectly possible for a given trader to quote at different prices for different quantities. In the figure, we have the trader represented by + who is willing to sell 5 units for 2, and 7 units at 4. This trader also placed buy orders of 3 units at a price of -2. We notice that she is willing to sell 9 units at -2. Between these two orders, several orders of other traders are present. The reason for this is that the limit-order book enforces temporal priority. In this case the lot that arrived first is the one that will also be executed first. Here, some

¹ This implies that the actual market price is the sum of the price indicated in the figure and of some fundamental value of the stock. This type of representation is convenient to model the order book as shown by Goettler, Parlour, and Rajan (2005).

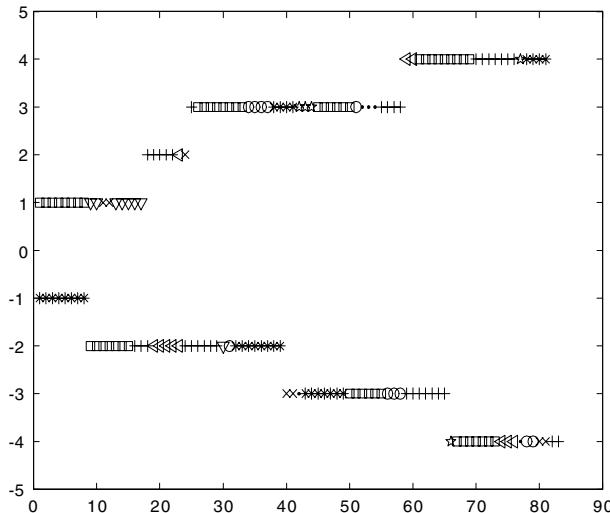


Fig. 3.1. Example of order book in an order-driven market.

traders placed their orders between the times when the trader represented by + placed her orders.

Clearly, at any point of time, one of the limit orders may be cancelled. Also, a big order may get split into small orders, thereby allowing a large trader to hide her intentions. Certain markets also allow for hidden orders. The nature of limit orders can also be rather complex in that limit orders may be conditional on some events. Some limit orders are valid for one day, and others up till the end of a month. Orders that are not traded immediately are called *booked* orders. These orders are also named *outstanding orders*.²

Roughly speaking, orders can be put into two categories. Those that should be executed at once and those that should be executed at some date in the future, when some special condition occurs. Let us turn to the first category of orders: those that should be executed immediately. The most general order in this category is the *market order*, which is an order to buy or sell immediately some stock at the best current price. If the number of shares to be bought or sold is larger than the number of shares at the best available price, then the order will *run the order book*. The so-called *all-or-none order* is instead an order that must be filled completely or the trade will not take place. In the second category, we find trades that are conditional. For instance, a *limit order at market price* is an order where only volume quoted at the best market

² A visit to various stock exchanges such as Euronext or Toronto Stock Exchange reveals that there exist many types of orders.

price will get executed. If by chance the entire order gets executed (meaning that the volume traded is smaller than the total volume at the best price), the order represents a complete fill. Otherwise it represents a partial fill. A *fill or kill (FOK)* order is an order where as big a volume as possible will be traded at a given price, but the partial fill will immediately be cancelled. Other limit orders are *day orders*, i.e., orders that are valid only for the day when they are entered. At the end of the day, such orders will be eliminated from the order book. Other limit orders such as the *good-till-cancelled orders* or the *good-till-date orders* specify a price, volume, and implicitly a time horizon. In a good-till-cancelled order, the time horizon of the order is either the time of the execution, the time of cancellation by the trader, or a pre-specified horizon set by the exchange. For instance, for the Toronto Stock Exchange, the time horizon is one year. The good-till-date order is valid until execution, or until a specified date.

3.1.2 Examples of orders

Let us imagine that a trader wishes to purchase a certain number of units of a given share. In this case, various possibilities arise:

1) If the order is an *at market price order*, then only that quantity will be traded that corresponds to the best quote. As an example, let us assume that the order book is as follows

Bid side		Ask side	
Price	Quantity	Price	Quantity
99.5	10	100.0	75
99.0	37	100.5	35
98.5	16	101.0	117

If the trader wishes to purchase 25 units at market price, then given that there are 75 units available, she will obtain these 25 units for a price of 100. After the trade, the book becomes

Bid side		Ask side	
Price	Quantity	Price	Quantity
99.5	10	100.0	50
99.0	37	100.5	35
98.5	16	101.0	117

If, however, the trader wishes to purchase 88 units at market price, she will only get 75. The remaining 13 units are then registered at the bid side. This means that the trader indicates her willingness to purchase 13 units at 100. The order book becomes

Bid side		Ask side	
Price	Quantity	Price	Quantity
100	13	100.5	35
99.5	10	101.0	117
99.0	37		
98.5	16		

2) Starting with the original book, if the order is a *market order*, then the trader wishing to purchase 100 units will receive 75 units at a price of 100 and 25 units at a price of 100.5. The total price to be paid will be

$$75 \times 100 + 25 \times 100.5 = 10012.5.$$

In this case, after the trade, the order book becomes

Bid side		Ask side	
Price	Quantity	Price	Quantity
99.5	10	100.5	10
99.0	37	101.0	117
98.5	16		

As these various examples indicate, depending on the type of order, and depending on the way an order book looks before its arrival, many issues are possible.

First, prices are not continuous real numbers but discrete ones. The minimum variation by which a price may vary is called the *tick size*. The difference between the bid and the ask prices is called the *spread*. It follows that the best spread is bigger or equal to the tick size.

If we assume that buy and sell orders arrive at random, then the simple fact that the buy price is higher than the sell order will induce a negative autocorrelation in the price process. This observation led to one of the first microstructure papers, e.g., Roll (1984).

Inspection of the various tables and of the way they evolve shows that trades will affect prices. For instance, if an institutional trader wishes to purchase a large number of units, the slope of the ask side will be of relevance. The slope of the order book will be called the *depth* of the market. The deeper the market, the lower the impact on the market of a given purchase. Also, it is likely that a large trader may wish to have information on as many quoted prices and associated quantities as possible, whereas a small trader may only be interested in the best bid and ask quotes. The amount of information that is made available to traders is not an obvious decision. Kyle (1985) provides a detailed analysis on the theoretical role of market depth.

At this stage, we may discuss a certain number of issues related to the order book. The first issue that we wish to discuss is what determines the size of the bid-ask spread.

3.1.3 Components of the bid-ask spread

Without entering into subtler aspects of microstructure, generally, we distinguish three components of the bid-ask spread.

The first one is the *inventory component*: If a trader holds a certain amount of units, if some adverse piece of news hits the market, the trader will be affected because her inventory loses value. Thus, for markets with greater price volatility, we may suspect the inventory component to be larger. Clearly, this component will be more important for a dealers' market than for a limit order market because the dealers may be expected to hold relatively large amounts of the asset in order to be able to satisfy orders as they arrive.

A second component is the *transaction-cost component*: Clearly, each trader in a given market must pay for the existence of the market. These payments may take various forms. For a floor trader, in some markets, it may be the license to trade. For a dealer, located behind a telephone, it may be the rent of her office and possibly, a certain commission paid for the supervisory institution of the market.

Last, we may mention the *asymmetric-information component*: Students of the information content of prices have been led to introduce various classes of investors. A first category of investors is the so-called liquidity trader or noise trader. These are, for instance, private investors who decide to rebalance their portfolio or who need to sell units of some financial asset to buy some other asset, say a car. The reason for their sale is purely random as opposed to being based on some news concerning the asset.

A second category of investors consists of the so-called informed traders. The term informed trader does not mean that the information has been illegally obtained, i.e., informed traders must not be confused with insider traders, whose activity is illegal. An informed trader could be an institutional investor who takes her job seriously, and who, after careful evaluation of an asset, reaches the conclusion that the asset is mispriced. If the trader judges the value to be higher than the current quoted price, it may be of interest for her to purchase the asset. This implies, of course, that the seller of the asset sold it too cheap. As important news hit a market, for instance macroeconomic news at the level of the foreign exchange or bond markets, or company specific news such as the results of the annual operations of a company, we may expect that traders do not wish to get disadvantaged because of some informational asymmetry. In this case, we may expect that the quoted bid-ask spreads will increase. Depending on the composition of the traders, informed versus non-informed, that we face, we may quote different prices. This is called the asymmetric information component.

3.2 Mandelbrot and the stable distribution

At this stage, we have provided the reader with some indications on how financial markets work. Before describing models that are based on information

arrival, we make a historical parenthesis and discuss Mandelbrot's model of asset prices.

3.2.1 A puzzling result

Assume that we sample stock prices during a given day, $t = 1, \dots, T$, where the sampling interval is given by Δ . Also, assume that the opening price, on a given day t , is written as $P_{t,0}$, and that the i -th price is $P_{t,i\Delta}$, for $i = 0, \dots, M$. We may then compute intraday returns

$$\rho_{t,i} = \log \left(\frac{P_{t,i\Delta}}{P_{t,(i-1)\Delta}} \right),$$

and we could also obtain the daily overall return as

$$r_t = \sum_{i=1}^M \rho_{t,i}.$$

Under the central limit theorem, we could then expect that the daily returns follow a normal distribution. Indeed, if the log-price increments are distributed *similarly*, with mean μ and variance σ^2 , and if the increments are not *too dependent*,³ then

$$\frac{1}{\sigma/\sqrt{M}} \left(\frac{1}{M} \sum_{i=1}^M \rho_{t,i} - \mu \right) \Rightarrow \mathcal{N}(0, 1),$$

or

$$r_t \Rightarrow \mathcal{N}(M\mu, M\sigma^2),$$

where the symbol \Rightarrow means convergence in distribution, which in turn means intuitively that if we construct a histogram of returns, for M large this histogram could not be distinguished from the normal density.

As discussed in Chapter 2, however, empirical evidence suggest that returns are highly non-normal. Various directions can be taken to solve this apparent puzzle. A first, very daring approach was the one proposed by Mandelbrot (1963). He assumes that price increments follow a stable distribution, a class of distribution described initially by Paul Lévy. One corollary of the stable distribution is that no finite second moments exist. In other words, according to Mandelbrot, we should model returns using a distribution where the variance no longer exists. Clearly, this direction is daring, because it implies that the mean-variance paradigm does not hold any longer. However,

³ Many versions of the central limit theorem exist. See for instance Billingsley (1968) or Gnedenko and Kolmogorov (1954). These versions specify how dependent realizations may be and how dissimilar, in terms of distribution, they may be so that the central limit theorem still holds.

going this way has taken research into a new direction by attracting the attention of researchers on the contributions of Lévy, an issue discussed in later chapters.

Another direction that has been suggested by Clark (1973) is that the arrival rate of information is time varying. This new assumption has led to the so-called mixture of distributions hypothesis (MDH), a model that has been relatively widely tested, at least in a univariate setting.

In this second section of the chapter, we address several issues in the theoretical modeling of asset returns. In the remainder of this section, we discuss Mandelbrot's contribution and its limitations. Then, in Section 3.3, we present the initial version of the MDH as described by Clark (1973). In Section 3.4, we turn to the bivariate extension of this model, developed by Tauchen and Pitts (1983). We discuss the tests of the MDH, proposed by Harris (1987) and Richardson and Smith (1994). Finally, Section 3.5 describes the model of Easley, Kiefer, and O'Hara (1997) that also provides a joint model for asset returns and trading volume, but in the context of a quote driven market.

3.2.2 Stable distribution

Mandelbrot (1963) kept the assumption that price increments are independent and identically distributed. By dropping the assumption that variance is finite, he proposed to use distributions belonging to the class of stable distributions that were discovered by Paul Lévy around 1920.⁴ Lévy defined and described stable distributions using the characteristic function.⁵ Its definition is that if X_1 and X_2 are two independent random variables drawn from a given stable distribution, then for c_1, c_2 , some real constants, there must exist another real constant c such that

$$c_1 X_1 + c_2 X_2 = cX, \quad (3.1)$$

where X has the same distribution as X_1 and X_2 .

As an example of stable distribution, we may mention the normal one with zero mean and variance σ^2 . Indeed, assume $X_1 \sim \mathcal{N}(0, \sigma^2)$ and $X_2 \sim \mathcal{N}(0, \sigma^2)$, both independent, then

$$c_1 X_1 + c_2 X_2 \sim \mathcal{N}(0, c\sigma^2),$$

with $c^2 = c_1^2 + c_2^2$.⁶

⁴ For an excellent text on the probabilistic developments, even though written in French, the reader may consult Lévy (1954). Curiously, Mandelbrot named this type of distribution the Pareto stable distribution.

⁵ This book contains in Chapter 15 a set of properties of characteristic functions.

⁶ There exist definitions of stability allowing for a non-zero mean. We only wish to illustrate the role that stable distributions played in finance, how they are characterized, and what their limitations are, rather than provide a full development of this theory. For a recent book on stable distributions, the reader may consult Samorodnitsky and Taqqu (1994).

The definition of a stable distribution is given in terms of its characteristic function. Let $\varphi(u)$ denote the characteristic function. From (3.1), it follows that

$$\varphi(c_1 u) \varphi(c_2 u) = \varphi(cu). \quad (3.2)$$

A modern definition of the stable characteristic function is

$$\varphi(u) = \begin{cases} \exp\left(i\delta u + \gamma^\alpha \left\{-|u|^\alpha + i\beta u |u|^{\alpha-1} \tan\left(\frac{\alpha\pi}{2}\right)\right\}\right), & \text{for } \alpha \neq 1, \\ \exp\left(i\delta u + \gamma^\alpha \left\{-|u|^\alpha - i\beta u \frac{2}{\pi} \log(|u|)\right\}\right), & \text{for } \alpha = 1, \end{cases} \quad (3.3)$$

where the term in $i\delta u$ corresponds to introducing a mean of δ . We notice that for this definition

$$c^\alpha = c_1^\alpha + c_2^\alpha.$$

For this reason, stable distributions are also sometimes called *alpha stable*. If we set $\alpha = 2$ and $\beta = 0$, in (3.3), we obtain the characteristic function of the normal distribution. As expected, the normal density is a special case of the stable family of distributions. If we set $\alpha = 1$ and $\beta = 0$, we obtain the characteristic function of the Cauchy distribution. For this situation, we may explicitly invert the characteristic function obtaining the density

$$f(x) = \frac{\gamma}{\pi(\gamma^2 + x^2)}.$$

It is well-known that the Cauchy distribution has fatter tails than the normal. This result shows that the parameter α is associated with the tail of the distribution. Given that $u/|u|$ in the expression of the characteristic function is nothing else but the function returning the sign of u , parameter β is associated with the relative importance of the two sides of the distribution. In other words, it characterizes the skewness of the distribution.

Figure 3.2 represents examples of the stable distribution. These curves are obtained by integration with the Simpson formula for the inverse transform.^{7,8} Inspection of the curves reveals that all the densities have tails that are fatter than the ones of the normal distribution obtained here with the value $\alpha = 2$.

⁷ Section 15.2 discusses how to obtain a density from a characteristic function. It turns out that the characteristic function takes a wide range as α becomes smaller and smaller. The integration uses as range for u $[-25; +25]$ for all integrations where $\alpha > 0.5$. For the case $\alpha = 0.5$, we have used as range $[-45; +45]$. Even with this range there remain some oscillations in the density.

⁸ According to the Simpson rule, to evaluate the integral $\int_L^U f(I)dI$, we construct I_j , for $j = 0, \dots, 2N$, with $I_0 = L$ and $I_{2N} = U$. The resulting spacing between adjacent I_j is denoted by δ . Also denote $f_j = f(I_j)$. The integral may then be approximated with

$$\int_L^U f(I)dI \approx \frac{\delta}{3} [f_0 + 4(f_1 + f_3 + \dots + f_{2N-1}) + 2(f_2 + f_4 + \dots + f_{2(N-1)}) + f_{2N}].$$

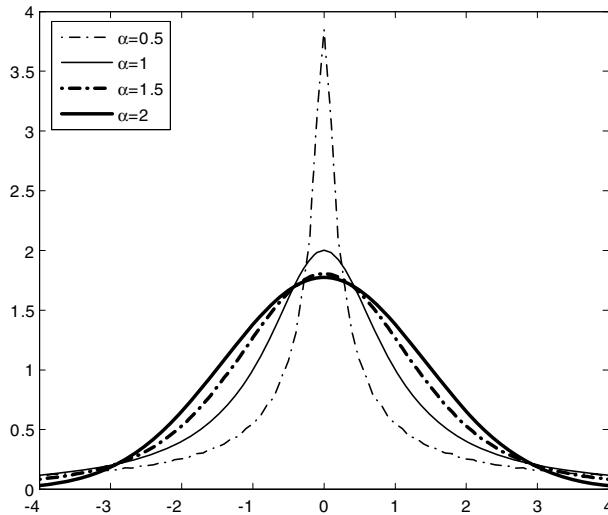


Fig. 3.2. Pdf of the stable distribution for various values of α .

The inversion formula becomes more difficult to implement for small values of α , implying that great care needs to be taken in actual implementations of the density. Indeed, large α imply a fast decrease of φ as u moves away from 0, whereas a small α implies a large domain for φ , and thus more work to select correct upper and lower bounds for the integration involved in the inverse Fourier transform needs to be provided.

In Figure 3.3, we keep the parameter α set to 1.5 and perform changes for the parameters β and γ . As this figure shows, the change in the β parameter corresponds to a change of the skewness of the distribution. A positive β is associated with a positive skewness. The parameter γ corresponds to a scaling of the curves.

Even though Mandelbrot's proposal provided an explanation of the non-normality of returns, it was met with a certain scepticism. First, as already mentioned, this distribution, not allowing for the existence of the variance, entailed that a great part of modern finance should be abandoned. For instance, mean-variance analysis builds critically on the existence of variance. Second, the characteristic function is in general not invertible so that a closed-form solution of the density is not available. This in turn implies difficulties for the numerical estimation of parameters. Some research (for instance, Fama, 1963, or Koutrouvelis, 1980) proposes multi-step algorithms to get a grip on the parameters. Clearly, using modern computing power, it would be possible to numerically invert the characteristic function and to estimate its param-

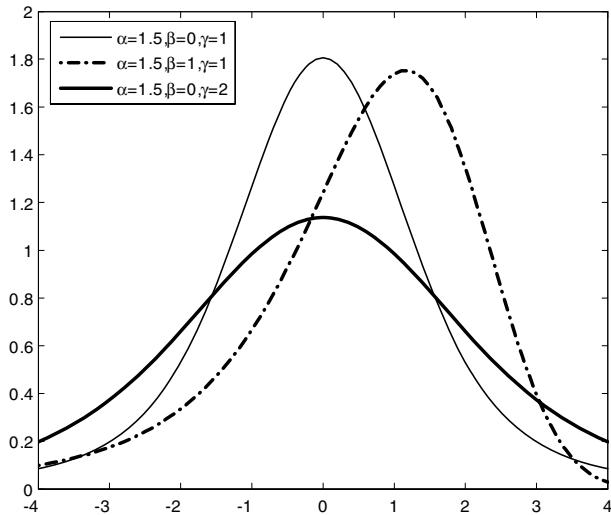


Fig. 3.3. Pdf of the stable distribution for various values of γ .

ters via maximum likelihood. Third, and this is probably the death stroke for this type of assumption, is the observation of DuMouchel (1973), by using extreme-value theory, that the tails of returns tend to have, in general, finite second and probably even third moments. Fourth, alternative methods have been advanced to describe the returns generating mechanism explaining the non-normality of asset. Among those alternative methods, the contribution of Clark (1973) certainly stands out.

3.3 Clark's subordination model

3.3.1 The idea of the model

Given the theoretical difficulties associated with the theory of stable distributions such as the non-existence of the second and higher moments and its failure to resist empirical tests, an alternative explanation to Mandelbrot's had to be developed at some point. One alternative explanation goes back to the work of Clark (1973). Concerning the arrival rate of events, Clark suggests that the clock does not always tick at a constant speed. He assumes that there are days on which more news is released than on others. Days with more news are then expected to display greater volatility.

Formally, let $p_t = \log(P_t)$ be the log-price process. Here, we assume that time is taking continuous values, $t \in \mathbb{R}^+$. The index t represents the *actual*

time, also called *calendar time*. Clark assumes that there exists a stochastic process written $T(t)$ that maps \mathbb{R}^+ into \mathbb{R}^+ . This process is assumed to be increasing in that if $t < s$ it must be that $T(t) \leq T(s)$. This map intuitively takes calendar time into what could be called *event time*.⁹ The assumption that more information becomes available as time goes by translates into the fact that $T(t)$ must be an increasing function of calendar time. Clark then supposes that the log-price process is given by

$$p_t = \mu T(t) + \sigma W_{T(t)},$$

where W_t stands for a Brownian motion.

The process p_t is said to be *subordinated* to W_t . The process $T(t)$ is called a *directing process*. We may presently ask how this new model solves the issue of non-normality of daily returns. We thus consider log-price increments over a given time span of length Δ that could correspond to a day. Then we have

$$r_t = p_t - p_{t-\Delta} = \mu(T(t) - T(t - \Delta)) + \sigma(W_{T(t)} - W_{T(t-\Delta)}).$$

Given that increments of Brownian motions are normally distributed, we obtain after introducing the notation $I_t = T(t) - T(t - \Delta)$ that

$$r_t \sim \mathcal{N}(\mu I_t, \sigma^2 I_t).$$

This expression means that the return measured over the given time span depends on the amount of activity, measured by I_t , that occurred over that interval. It also implies that conditional on the information arrival, returns are distributed as a normal distribution.

Given that each return is drawn from a different distribution, we may view the returns as being generated by a mixture of distributions. For this reason, the model of Clark is sometimes referred to as the *mixture of distributions hypothesis*.

From this stage, different roads may be taken. For instance, we could assume a given distribution for the directing process, for instance that the directing process is given by a log-normal process, a Poisson process, or any increasing process. Below we will show how the density of returns may be obtained if we use a log-normal process. A drawback of following this road is that returns will be independent. To circumvent this problem, the recent literature has proposed more complicated directing processes based on Lévy processes.¹⁰ Such directing processes assume that the information arrival rate

⁹ The expression *theta time* has also been used to indicate time measured in terms of events. Theta time is measured by cumulating squares of intraday price increments. This corresponds to the concept of realized volatility that we will also encounter in Chapter 4.

¹⁰ Another difficulty, inherent in this method, is that the density requires a numerical integration that turns out to be relatively unstable. The choice of range of the support of the characteristic function plays for instance a critical role. Should we use Fast Fourier techniques as indicated in Chapter 15, we run into a discretization bias.

changes over time. That type of model belongs to the class of stochastic volatility models, see for instance Barndorff-Nielsen and Shephard (2001). Still, a further complication comes from the empirical observation that after a large drop in returns, markets tend to be more agitated, an observation called leverage effect, a feature that needs to be modeled.

Another possibility is to assume that the information arrival rate constitutes a latent, thus unobservable variable, about which we wish to make some inference. This latent variable may also affect other observable variables such as volume or the trading activity over a given time horizon. Again this opens various directions. A first direction, followed for instance by Tauchen and Pitts (1983), is to introduce volume and to assume that both returns and volume are driven by some common source of information. A second direction, taken for instance by Jones, Kaul, and Lipson (1994), is to find variables that could proxy the information arrival rate. Candidates that could proxy this arrival rate are share volume, volume in monetary terms, number of transactions per unit of time, the average trade size, etc. A discussion of proxies for arrival rate may be found in Lo and Wang (2000).

We now discuss, in the following, these two alternative approaches. We begin in Section 3.3.2, with the consequences of assuming a log-normal distribution for the directing process. Then, in Section 3.4, we discuss the microstructure model developed by Tauchen and Pitts (1983), in which the information arrival plays the role of directing process for both asset returns and trading volume.

3.3.2 The density of returns under subordination

Using the properties of conditional probabilities and assuming that the distribution of increments of the directing process I_t follows a distribution with density $g(I_t)$, we obtain that

$$f(r_t) = \int_{I_t \in \mathbb{R}^+} f(r_t | I_t) g(I_t) dI_t.$$

Conditional on the information arrival flow I_t , the density of returns is assumed to be given by the normal distribution. If we assume that the distribution of I_t is log-normal, i.e., $\log(I_t) \sim \mathcal{N}(\mu, m_2)$, we obtain that

$$\begin{aligned} f(r_t) &= \int_{I_t \in \mathbb{R}^+} \frac{1}{\sqrt{2\pi\sigma_r^2 I_t}} \exp\left(-\frac{1}{2} \frac{(r_t - \mu_r I_t)^2}{\sigma_r^2 I_t}\right) \\ &\quad \times \frac{1}{\sqrt{2\pi m_2 I_t}} \exp\left(-\frac{1}{2} \frac{(\log(I_t) - \mu)^2}{m_2}\right) dI_t. \end{aligned} \quad (3.4)$$

To illustrate the types of shapes of the density that results from the subordination, we construct numerically the density according to (3.4). The numerical integration can be performed by approximating the integral involving I_t with

the Simpson rule. A difficulty is the choice of lower and upper bounds for the integral. Also the choice of step length between adjacent points over which to evaluate the density may be important. Given that the integrand involves a density, we may be guided by the *cdf* of I_t . We may choose as an upper bound a value such that 99.9% of the probability mass is taken into consideration. We consider 4 different models that differ by their parameters as described in Table 3.1.

In the following, we set the lower bound of the integration $L = 0.0001$, $U = 25$, and $\delta = 0.1$. We obtain Figure 3.4 representing the various densities. The benchmark model, Model 1, is traced with a continuous thin line. As we increase the variance of returns, the density becomes wider as the curve with the thick continuous line testifies. This corresponds to Model 2. The next exercise consists in increasing the average level of news μ from 1 to 2. This yields Model 3 represented with the dotted line. We notice that the mean is shifted to the right and the body of the density becomes wider. In this model news affect both the drift and the variance. We might view the feature that an increase in news shifts the density to the right as an inconvenient of this model. This inconvenient is easy to circumvent by assuming that I_t does not affect the mean of returns. The last model under consideration, represented with the dashes, Model 4, corresponds to an increase in the variance of the information arrival rate. We notice that this generates densities with particularly thick tails. This is a feature that we expect for actual return data.

Many distributional assumptions can be made concerning the information arrival. Clark (1973) assumes a log-normal arrival rate. So do Tauchen and Pitts (1983) and Foster and Viswanathan (1995). Richardson and Smith (1994) consider in addition a uniform arrival rate, a Poisson distribution, and the inverted gamma distribution. This latter model is particularly appealing because it has been shown by Blattberg and Gonedes (1974) that if returns are conditionally normal with a variance that follows the inverse gamma distribution, then the resulting distribution of returns will be the Student t distribution. As we will see in Chapter 5, the Student t distribution has many features that allow for a good modeling of asset returns. More recently, Madan and Seneta (1990) proposed the use of the gamma distribution. Use of this distribution yields the so-called variance-gamma model. The idea of changing time may also be found in the literature on Lévy processes, as we will discuss in Section 12.4.

Table 3.1. Values of the parameters for the various models considered

Model number	μ_r	σ_r	μ	$\sqrt{m_2}$
1	1	1	1	1
2	1	1.5	1	1
3	1	1	2	1
4	1	1	1	2

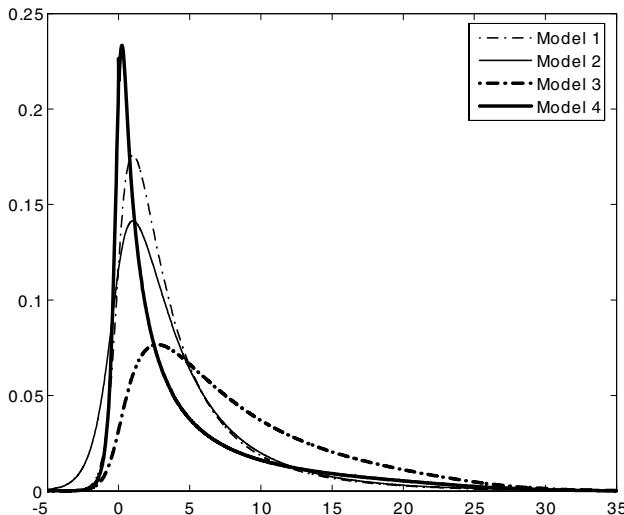


Fig. 3.4. *Pdf of various densities of returns under log-normal subordination.*

3.4 A bivariate mixture-of-distribution model for return and volume

3.4.1 A microstructure model for information arrivals

Now, we go one step further and investigate how the information arrival may affect not only the asset return but also the trading volume. The model developed by Tauchen and Pitts (1983) is rather simple, yet very rich, and it has been extended into various directions.¹¹ A first contribution of this work is that it provides a better understanding of the way volatility should behave, as the number of traders increases exogenously. Another contribution is that it helps to explain why Clark (1973) found in an empirical investigation that volume does not capture information arrival very well.

The assumptions of the Tauchen and Pitts (1983) model are the following:

- (i) There are J traders. This number is given exogenously. In other words, it is not the particular structure of a market that impacts the number of traders. The index $j = 1, \dots, J$ corresponds to the various traders.
- (ii) The trader may take long or short positions as we evolve from one equilibrium to another. When stating that the trader takes long and short positions, we understand implicitly that this model corresponds better

¹¹ See for instance Liesenfeld (2001).

to the futures market than to an equity market. One reason why this assumption needs to be made is that through time, the number of companies listed increases. To track the number of listed shares implies additional difficulties.

- (iii) It is assumed that at each day, there are I trades. The index $i = 1, \dots, I$, corresponds to a trade. When transaction i takes place, each trader has a certain reservation value p_{ij}^* . If the quoted price is p_i , then trader j has a demand given by

$$Q_{ij} = \lambda (p_{ij}^* - p_i).$$

The parameter λ is positive and constant. The specification of such a simple demand (and supply) function is to keep the model as simple as possible. We could consider that transactions costs would affect demand.

Clearly, a trader will wish to buy ($Q_{ij} > 0$) if the quoted price p_i is lower than the value the trader associates to a given asset at the instant of the i th trade, p_{ij}^* .

The way the model has been set up so far, shows that traders differ in their reservation value. It is traders' expectations that are likely to cause the difference in reservation values. Clearly, if $p_{ij}^* = p_i$, then trader j will not trade. This notion conveys the idea that if everybody has the same reservation value, or that there is consensual information, then there will be no trade. The model, in other words, does not assume that there is always a certain minimal level of trade, due to noise traders.

If some traders buy, others must sell. Hence we must have that

$$\sum_{j=1}^J Q_{ij} = 0.$$

This constraint in turn implies that

$$p_i = \frac{1}{J} \sum_{j=1}^J p_{ij}^*.$$

In equilibrium, the observed price is the average reservation value among all traders.

What is the volume traded between two time periods? To understand how to compute the equilibrium volume, consider the following numerical example:

	Position _{$i-1$}	Position _{i}	Buy(+) / sell(-)
$j = 1$	3	1	$\Delta Q_{i1} = -2$
$j = 2$	4	5	$\Delta Q_{i2} = +1$
$j = 3$	2	3	$\Delta Q_{i3} = +1$

The second and third columns represent the positions, in units of some asset, of three traders, $j = 1, 2, 3$ for two different instants, $i - 1$, and i . The last column indicates the changes required to achieve the new position.

In this example, trader one ($j = 1$) had 3 units at time $i - 1$ and she diminishes her position down to 1 unit at time i . The net volume traded is 2. It represents half of the total absolute value of the traded assets. This reasoning also holds for the other traders. As a consequence, the aggregate volume at trade i is

$$v_i = \frac{1}{2} \sum_{j=1}^J |Q_{ij} - Q_{i-1,j}| = \frac{\lambda}{2} \sum_{j=1}^J |\Delta p_{ij}^* - \Delta p_i|.$$

We notice that the volume is due to changes in reservation values with respect to their averages. As a consequence, if traders have very similar opinion (no heterogeneity of beliefs), there needs to be no volume. Yet, there may be large jumps in prices. This type of price behavior has also been discussed in an early contribution by Karpoff (1987).

To go further, especially to make the model testable, Tauchen and Pitts (1983) assume that changes in reservation values are due to global information and trader-specific information:

$$\Delta p_{ij}^* = \phi_i + \psi_{ij}.$$

The first component of the reservation price change is ϕ_i , a common information component. We may view this component as representing some macroeconomic news. The second component ψ_{ij} is some individual information. It is assumed that

$$\begin{aligned} E[\phi_i] &= E[\psi_{ij}] = 0, \\ V[\phi_i] &= \sigma_\phi^2 \quad \text{and} \quad V[\psi_{ij}] = \sigma_\psi^2, \end{aligned}$$

and that ϕ_i and ψ_{ij} are independent. We introduce $\bar{\psi}_i = \frac{1}{J} \sum_{j=1}^J \psi_{ij}$ the average individual information. Under these assumptions, the price change and volume at trade i may be written as

$$r_i = \Delta p_i = \phi_i + \frac{1}{J} \sum_{j=1}^J \psi_{ij}, \tag{3.5}$$

$$v_i = \frac{\lambda}{2} \sum_{j=1}^J |\psi_{ij} - \bar{\psi}_i|. \tag{3.6}$$

The first expression shows that common opinion and average information matters. If common information hints at good news, the asset price is to increase. If the private information of one of the traders is that the price should be higher, then the asset price will increase. This increase will, however, be averaged out by the mass of traders.

The second expression shows that heterogeneity of information is responsible for volume. To go further, the authors assume normality of all individual

components. We recall the following analytical result, useful for later on: if $X \sim \mathcal{N}(0, \sigma^2)$, then $E[X] = \sigma\sqrt{2/\pi}$. The computation of the first and second moments associated with equations (3.5) and (3.6) yields

$$\begin{aligned} m_r &= E[r_i] = 0, \\ \sigma_r^2 &= V[r_i] = \sigma_\phi^2 + \frac{\sigma_\psi^2}{J}, \\ m_v &= E[v_i] = \frac{\lambda}{2} J \times E[|\psi_{ij} - \bar{\psi}_i|] = \frac{\lambda J}{2} \sqrt{\frac{2}{\pi}} V[\psi_{ij} - \bar{\psi}_i] \\ &= \frac{\lambda J}{2} \sqrt{\frac{2}{\pi}} \left(\sigma_\psi^2 - \frac{\sigma_\psi^2}{J} \right)^{1/2} = \frac{\lambda J}{2} \sqrt{\frac{2}{\pi}} \sigma_\psi \left(\frac{J-1}{J} \right)^{1/2}, \\ \sigma_v^2 &= V[v_i] = \left(\frac{\lambda}{2} \right)^2 V[|\psi_{ij} - \bar{\psi}_i|]. \end{aligned}$$

At this stage, the model unfolds its ingenuity. For instance, we notice that as the number of traders J increases, the expected volume, μ_v , may increase, while at the same time price change variability gets smaller. These appear to be characteristics that we also observe in financial markets. The last two equations also show that larger heterogeneity of beliefs goes hand in hand with more volume and larger variability thereof. We see that this model can, therefore, be used to investigate the impact of a change in the number of traders over time. The model also has interesting properties if we focus only on a given moment of time, say a day, when the number of traders can be expected to be constant. Over a day, we observe *I price equilibria*. The expression *price equilibria* is of course used in a rather loose sense, because with equilibrium it is meant here that a trade occurred. In empirical work, we could proxy the number of price equilibria with the number of trades over a given time horizon. An alternative interpretation is that there are price equilibria and then for some of them there are trades. In this latter interpretation, we observe the trade process and have as a latent process the number of equilibria.

If we use the initial idea, then we can take the daily number of trades as a proxy for the number of daily equilibria. We then obtain that

$$r = \sum_{i=1}^I r_i \quad \text{and} \quad v = \sum_{i=1}^I v_i,$$

where I is the *mixture variable* or *directing process*. The two processes, r and v are called subordinated processes. Conditional on I being given, and assuming normality, it follows easily that

$$r \sim \mathcal{N}(0, I \sigma_r^2), \quad \text{and} \quad v \sim \mathcal{N}(I m_v, I \sigma_v^2),$$

which implies that

$$r = \sqrt{I} \sigma_r z_r, \quad z_r \sim \mathcal{N}(0, 1), \quad (3.7)$$

$$v = I m_v + \sqrt{I} \sigma_v z_v, \quad z_v \sim \mathcal{N}(0, 1). \quad (3.8)$$

The equations (3.7) and (3.8) are very interesting starting points for empirical research. See, for instance, Liesenfeld (2001). We also notice that in (3.8) volume is composed by a first component $I m_v$, which is the number of information arrivals, and $\sqrt{I} z_v$, which is the information content of the information. Given that volume involves the variable information arrival in addition to some other variable, this implies that the volume v is a rather bad substitute for information arrival. This observation justifies the finding of Clark (1973), who showed that trading volume is not a perfect proxy for the subordinating variable I .¹²

The previous model shows that if we consider just prices or asset returns, we deal with an object that is related to other market specific characteristics such as the monetary volume, the number of shares traded, or the number of transactions that take place on a given day. In other words, we consider only a partial model. Contemplation of (3.7) shows that returns depend on normal innovations and some information process. Whereas here an explicit model exists describing the origin of I , in certain models, such as ARCH or GARCH models, discussed in Chapter 4, a parametric specification is given where the directing process is a function of past returns. In stochastic volatility models, a parametric specification is also given, but in that case, a second source of randomness is introduced. It is interesting to notice that only little research effort has been made to use forecasting models of volume or trading activity to help predicting the future variance of an asset.

It is worth noticing that if the information flow I_t comes clustered, i.e., if after a first release of information further information gets released, then the return process will also exhibit clustering of volatility in the sense that calm periods will be followed by calm periods and agitated periods by agitated periods.

This observation raises the issue of the clustering and the explanatory power of the information flow. Several empirical studies investigate this issue. For instance, Ederington and Lee (1993) consider macroeconomic news announcement days, that is, they obtain the calendar of days when macroeconomic news get released and question if at moments where an information release takes place the volatility is greater. In their study, they are able to show that once such macroeconomic news are taken into account, there remains little unexplained volatility. This study got extended to interest rate markets by Fleming and Remolona (1999) who confirm the Ederington and Lee study. These studies show that pointwise announcements will create point wise peaks in volatility. It leaves open the issue if news announcements create clustered volatility. Jones, Lamont, and Lumsdaine (1998) question if macro-

¹² If we had $\sigma_v = 0$, then by taking the ratio r/\sqrt{v} we ought to obtain a normally distributed random variable. Empirically, this is not the case.

economic announcements come clustered. They show that announcement-day volatility does not persist. Given that asset return volatility clusters, they conclude that volatility that does not stem from announcement day must be serially correlated.

3.4.2 Implications of the mixture of distributions hypothesis

Following the early contribution of Clark (1973), various research aimed at testing the mixture of distribution hypothesis. The contribution of Tauchen and Pitts (1983) provided an economic model, yielding testable implications concerning returns and volume jointly.¹³ Harris (1986, 1987) extended the predictions of that model by considering a model where price increments, r_t , and volume, v_t , of a given day, are conditionally normal. Also, conditionally on the information arrival, I_t , the covariance between returns and volume is equal to zero.

Formally, we may write the model of Harris (1987) as

$$r_t \sim \mathcal{N}(m_r I_t, \sigma_r^2 I_t), \quad (3.9)$$

$$v_t \sim \mathcal{N}(m_v I_t, \sigma_v^2 I_t), \quad (3.10)$$

$$\text{Cov}[r_t, v_t | I_t] = 0. \quad (3.11)$$

The first equation corresponds to Clark's initial 1973 model. If we consider a restriction of the bivariate model, we obtain Tauchen and Pitts (1983). Harris (1987) derives a large set of conditional and unconditional moments that he uses to extend the set of stylized facts concerning asset returns and volume, already obtained by Tauchen and Pitts (1983). Richardson and Smith (1994) further built on these conditional and unconditional moments to construct a formal test of the mixture of distribution hypothesis based on the method of moments. Currently, we follow Harris (1986, 1987) and present selected moments and co-moments of this model. These theoretical moments may be then used to estimate the parameters of the model in a Generalized Method of Moments (GMM) setting.

Assume that the information arrival variable I has mean m_I and that central moments of I are denoted by $\mu_k = E[(I - m_I)^k]$. Also assume that, conditionally on I , the daily price increment satisfies

$$E[r|I] = m_r I, \quad (3.12)$$

$$V[r|I] = E[(r - m_r I)^2 | I] = \sigma_r^2 I. \quad (3.13)$$

Then, using the law of iterated expectations, we have the following unconditional moments of r (see the Appendix of Harris, 1987, for computational details)¹⁴

¹³ Liesenfeld (1998, 2001) provides direct tests of the Tauchen and Pitts (1983) model.

¹⁴ For instance, the derivation of equation (3.14) is obvious:

$$E[r] = m_r m_I, \quad (3.14)$$

$$E[(r - m_r m_I)^2] = V[r] = \sigma_r^2 m_I + m_r^2 \mu_2, \quad (3.15)$$

$$E[(r - m_r m_I)^3] = 3\mu_r \sigma_r^2 \mu_2 + m_r^3 \mu_3, \quad (3.16)$$

$$\begin{aligned} E[(r - m_r m_I)^4] &= 3\sigma_r^4 m_I^2 + 3\sigma_r^4 \mu_2 + 6m_r^2 \sigma_r^2 \mu_3 \\ &\quad + 6m_r^2 \sigma_r^2 m_I \mu_2 + m_r^4 \mu_2. \end{aligned} \quad (3.17)$$

Also, if $m_r = 0$, then

$$\begin{aligned} E[r^2] &= \sigma_r^2 m_I, \\ E[(r^2 - \sigma_r^2 m_I)^2] &= 3m_v^4 \mu_2 + 2\sigma_r^4 m_I^2. \end{aligned}$$

This last expression corresponds to the variability of squared returns.

Unconditional skewness and kurtosis are immediately obtained as

$$Sk[r] = E \left[\left(\frac{r - E[r]}{V[r]^{1/2}} \right)^3 \right] = \frac{3\sigma_r^2 m_r \mu_2 + m_r^3 \mu_3}{(\sigma_r^2 m_I + m_r^2 \mu_2)^{3/2}}, \quad (3.18)$$

$$\begin{aligned} Ku[r] &= E \left[\left(\frac{r - E[r]}{V[r]^{1/2}} \right)^4 \right] \\ &= 3 + \frac{3\sigma_r^4 \mu_2 + 6m_r^2 \sigma_r^2 \mu_3 + m_r^4 \mu_4 - 3m_r^4 \mu_2}{(\sigma_r^2 m_I + m_r^2 \mu_2)^2}, \end{aligned} \quad (3.19)$$

Therefore, if $m_r = 0$, then returns will not be skewed even if there is subordination. Equation (3.19) shows that if $m_r = 0$, kurtosis becomes

$$Ku[r] = 3 + 3 \frac{\mu_2}{m_I^2},$$

and, thus, as Harris (1987) points out, kurtosis will be larger than the one obtained for a normal distribution. For the general case, where $m_r \neq 0$, the greater the skewness and kurtosis of the news arrival rate, the greater the kurtosis of the returns. This result may also be found in Clark (1973). Kurtosis in excess over 3 is a well-documented feature of asset returns as already shown in Chapter 2.

$$E[r] = E[E[r|I]] = E[m_r I] = m_r m_I.$$

Equation (3.15) is obtained as follows

$$\begin{aligned} E[(r - m_r m_I)^2] &= E[E[(r - m_r m_I)^2|I]] = E[E[(r - E[r|I] + E[r|I] - m_r m_I)^2|I]] \\ &= E[E[(r - E[r|I])^2|I]] + 2E[E[(r - E[r|I])(E[r|I] - m_r m_I)|I]] \\ &\quad + E[(E[r|I] - m_r m_I)^2|I] \\ &= E[\sigma_r^2 I] + 0 + E[(m_r I - m_r m_I)^2] = \sigma_r^2 m_I + m_r^2 \mu_2. \end{aligned}$$

It is also interesting to investigate the joint behavior of observable series such as returns and volume. This implies that we need some results concerning covariances. Harris (1987) shows that, if conditional on I , two random variables X and Y have the distribution

$$\begin{aligned} X &\sim \mathcal{N}(m_X I, \sigma_X^2 I), & m_X \geq 0, \sigma_X > 0, \\ Y &\sim \mathcal{N}(m_Y I, \sigma_Y^2 I), & m_Y \geq 0, \sigma_Y > 0, \end{aligned}$$

with

$$\text{Cov}(X, Y|I) = 0,$$

then the following results hold¹⁵

$$\begin{aligned} \text{Cov}[X, I] &= m_X \mu_2, \\ \text{Cov}[X^2, I] &= \sigma_X^2 \mu_2 + m_X^2 (\mu_3 + 2m_I \mu_2), \\ \text{Cov}[X, Y] &= m_X m_Y \mu_2, \\ \text{Cov}[X^2, Y] &= \sigma_X^2 m_Y \mu_2 + m_Y m_X^2 (\mu_3 + 2m_I \mu_2). \end{aligned}$$

Harris (1987) also derives the expression for the correlation between a return and the information variable or between the two returns. We obtain that

$$\text{Corr}[X, I] = \frac{m_X \mu_2}{[(\sigma_X^2 m_I + m_X^2 \mu_2) \mu_2)]^{1/2}}.$$

As we could expect, if the variance of the daily information rate μ_2 tends to zero, so does the correlation between X and I . Also, if the drift of the prices m_X is zero, there will be no correlation between returns and the information flow. The sign of the correlation is determined by the sign of m_X . We also obtain

¹⁵ These results are obtained in a straightforward manner. For instance, we have

$$\begin{aligned} \text{Cov}[X, I] &= E[(X - E[X])(I - E[I])] \\ &= E[(X - E[X|I] + E[X|I] - E[X])(I - E[I])|I]] \\ &= E[(E[X|I] - E[X])(I - E[I])] \\ &= [(m_X I - m_X m_I)(I - m_I)] = m_X E[(I - m_I)^2] = m_X \mu_2. \end{aligned}$$

Similarly for $\text{Cov}[X^2, I]$. Furthermore,

$$\begin{aligned} \text{Cov}[X, Y] &= E[E[(X - E[X|I] + E[X|I] - E[X]) \\ &\quad \times (-E[Y|I] + E[Y|I] - E[Y]|I])] \\ &= E[(E[X|I] - E[X])(E[Y|I] - E[Y])] \\ &= E[(m_X I - m_X m_I)(m_Y I - m_Y m_I)] = m_X m_Y \mu_2. \end{aligned}$$

The second equality follows from the fact that in the first line, $E[X]$ and $E[X|I]$ are both known once I is known. Further, all the terms involving $X - E[X|I]$ cancel.

$$\text{Corr}[X, Y] = \frac{m_X m_Y \mu_2}{[(\sigma_X^2 m_I + m_X^2 \mu_2)(\sigma_Y^2 m_I + m_Y^2 \mu_2)]^{1/2}}.$$

If $m_X = 0$, then

$$\text{Corr}[X^2, I] = (3 + 2m_I^2 \mu_2)^{-1/2},$$

$$\text{Corr}[X^2, Y] = \mu_2 \left[(3\mu_2 + 2m_I^2) \left(\frac{\sigma_Y^2}{m_Y^2} m_I + \mu_2 \right) \right]^{-1/2}.$$

If X plays the role of a return and Y corresponds to some trading activity measure such as volume, then these computations show that return and volume should be correlated.

Finally, Harris provides autocovariances and autocorrelations. We now denote by r_t the return on a given day t , and we assume that, conditional on the information arrival rate I_t

$$r_t \sim \mathcal{N}(m_r I_t, \sigma_r^2 I_t), \quad m_r \geq 0, \text{ and } \sigma_r > 0.$$

We introduce the notation for the autocorrelation of information arrival rate

$$\gamma \equiv \text{Cov}[I_t, I_{t-1}] = E[(I_t - m_I)(I_{t-1} - m_I)].$$

Then, we deduce the following expression for the autocovariance of returns

$$\begin{aligned} \text{Cov}[r_t, r_{t-1}] &= E[(r_t - E[r_t])(r_{t-1} - E[r_{t-1}])] \\ &= E[E[(r_t - E[r_t|I_t]) + E[r_t|I_t] - E[r_t]] \\ &\quad \times (r_{t-1} - E[r_{t-1}|I_{t-1}] + E[r_{t-1}|I_{t-1}] - E[r_{t-1}])|I_t, I_{t-1}]] \\ &= E[(E[r_t|I_t] - E[r_t])(E[r_{t-1}|I_{t-1}] - E[r_{t-1}])] \\ &= E[(m_r I_t - m_r m_I)(m_r I_{t-1} - m_r m_I)] \\ &= m_r^2 \gamma. \end{aligned}$$

Concerning volatility, Harris shows that if $m_r = 0$ then

$$\begin{aligned} \text{Cov}[r_t^2, r_{t-1}^2] &= E[r_t^2 r_{t-1}^2] - (E[r_t^2])^2 = E[E[r_t^2 r_{t-1}^2 | I_t, I_{t-1}]] - \sigma_r^4 m_I^2 \\ &= E[\sigma_r^2 I_t \sigma_r^2 I_{t-1}] - \sigma_r^4 m_I^2 \\ &= \sigma_r^4 \gamma. \end{aligned}$$

The last equality stems from the fact that

$$E[I_t I_{t-1}] = E[(I_t - m_I + m_I)(I_{t-1} - m_I + m_I)] = \gamma + m_I^2.$$

From these moments, we can compute the autocorrelations

$$\text{Corr}[r_t, r_{t-1}] = \frac{m_r^2 \gamma}{\sigma_r^2 m_I + m_r^2 \mu_2} = \frac{\gamma / \mu_2}{\frac{\sigma_r^2 m_I}{m_r^2} \frac{1}{\mu_2} + 1}.$$

As long as the average information flow is strictly positive, the first term of the denominator is positive, and therefore the overall denominator will be larger than one. This result shows that the autocorrelation between returns is smaller than the autocorrelation of the information arrivals, γ/μ_2 .

Also, if $m_r = 0$ we have

$$\text{Corr} [r_t^2, r_{t-1}^2] = \frac{1}{3} \frac{\gamma}{\mu_2}.$$

This result shows that the dependency between squared returns is directly proportional to the autocorrelation of the arrivals of information.

These moments provide many empirically testable hypotheses. Harris (1987) does not provide a formal test, however, he verifies using a joint model for price and volume that many of the empirical facts are verified. In the following, using longer samples, we replicate some of his findings. Rather than working at the level of individual company data, we use aggregate data, namely the SP500 and the associated volume series.

The upper part of Figure 3.5 displays the raw volume series associated with the SP500.¹⁶ We notice that the raw series has an upward trend. This trend can be explained by the fact that the number of companies increased over the sample period and so did the capitalization and shares of the largest existing companies. Because of this trend, the raw data cannot be used. We therefore filter the data as suggested by Gallant, Rossi, and Tauchen (1992) and Bollerslev and Jubinski (1999). That is, if V_t denotes the raw volume for day t , we first perform the OLS regression

$$\log(V_t) = \alpha + \beta t + u_t,$$

and then use the filtered value

$$v_t = V_t / \exp(\hat{\alpha} + \hat{\beta}t),$$

where the parameters with a hat indicate estimated parameters.¹⁷ Graphical inspection of the lower series in Figure 3.5, representing the filtered volumes, indicates that the time trend has disappeared. There remain, however, long run oscillations in the data. Also, on some of the earlier days, the volume traded appears excessively high or low. On the other hand, not being sure that these data points are outliers, we decided to keep them in the analysis.

3.4.3 Testing the mixture of distribution hypothesis

The contribution of Richardson and Smith (1994) is to provide such a formal test. After considering the necessity, for identification purposes, to fix μ_1 to

¹⁶ The data presented here has been extracted from Datastream. It turned out that the raw series contained some days with zero volume even though the value of the index changed. The data for such days has been replaced by an interpolated value.

¹⁷ It is trivial to demonstrate that $v_t = \exp(\hat{u}_t)$.

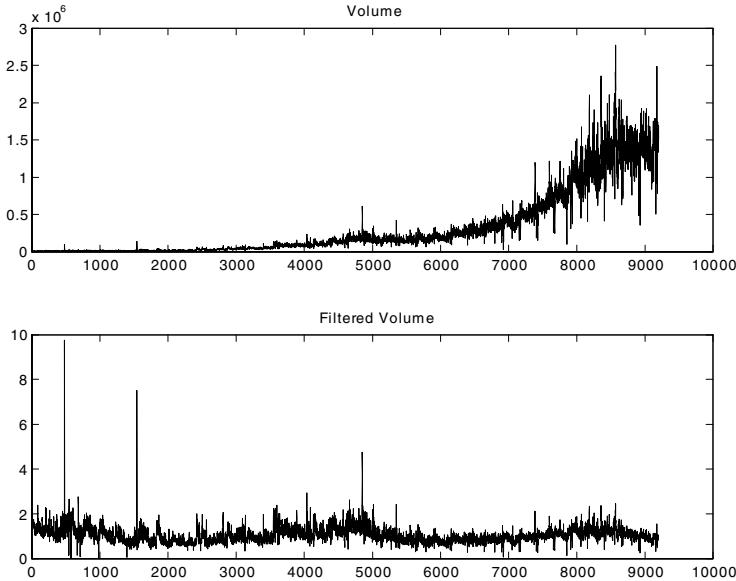


Fig. 3.5. Evolution of the daily trading volume of the SP500.

some value, say 1, they estimate various forms of the mixture of distribution hypothesis with GMM.¹⁸ The moment-generating conditions are given by

$$g_T(\theta) = \frac{1}{T} \sum_{t=1}^T M_t(\theta),$$

where

$$M_t(\theta) = \begin{bmatrix} r_t - m_r \\ v_t - m_v \\ (r_t - m_r)^2 - \sigma_r^2 - m_r^2\mu_2 \\ (v_t - m_v)^2 - \sigma_v^2 - m_v^2\mu_2 \\ (r_t - m_r)(v_t - m_v) - m_r m_v \mu_2 \\ (r_t - m_r)^4 - 3\sigma_r^4(1 + \mu_2) - \dots \\ 6m_r^2\sigma_r^2(\mu_3 + \mu_2) - m_r^4\mu_4 \\ (v_t - m_v)^4 - 3\sigma_v^4(1 + \mu_2) - \dots \\ 6m_v^2\sigma_v^2(\mu_3 + \mu_2) - m_v^4\mu_4 \\ (r_t - m_r)^2(v_t - m_v)^2 - \sigma_r^2\sigma_v^2(1 + \mu_2) - \dots \\ (\sigma_v^2 m_r^2 + \sigma_r^2 m_v^2)(\mu_3 + \mu_2) - m_r^2 m_v^2 \mu_4 \\ (r_t - m_r)^3(v_t - m_v) - 3m_r m_v \sigma_r^2(\mu_3 + \mu_2) - m_r^3 m_v \mu_4 \\ (r_t - m_r)(v_t - m_v)^3 - 3m_r m_v \sigma_v^2(\mu_3 + \mu_2) - m_r m_v^3 \mu_4 \end{bmatrix}.$$

¹⁸ The GMM estimation has been developed by Hansen (1982). See Hamilton (1994) for a well-written exposition of that estimation and testing technique.

Table 3.2. Parameter estimates of the mixture-of-distributions model

Parameter	Estimate	Std error
m_r	0.034	(1.035)
σ_r	0.104	(0.046)
m_v	0.957	(0.013)
σ_v	1.032	(0.010)
μ_2	0.003	(0.011)
μ_3	5.380	(0.121)
μ_4	0.029	(0.570)

The parameters are summarized in $\theta = (m_r, m_v, \sigma_r^2, \sigma_v^2, \mu_2, \mu_3, \mu_4)'$. Intuitively, invoking the law of large numbers, the average of sample realizations will converge to the expectations. Hence, for T large, we expect that $g_T(\theta)$ converges to $E[M_t(\theta)]$, which should be exactly equal to zero if the mixture of distributions hypothesis holds. Since the number of moment conditions exceeds the number of unknown parameters, the empirical moment conditions cannot be exactly equal to zero. However, in this case, it can be tested whether the model correctly adjusts to the data, using these so-called over-identifying restrictions. This is the basis for the Hansen's (1982) J test. It can also be tested if the parameters satisfy certain relations.

The GMM estimator of θ is obtained by minimizing the expression

$$g_T(\theta)' S^{-1} g_T(\theta),$$

where S is the weighting matrix defined as

$$S = \frac{1}{T} \sum_{t=1}^T M_t(\theta) M_t(\theta)'.$$

There are also versions of the model that correct for possible autocorrelation and heteroskedasticity in the data.¹⁹ For the model at hand, there are 10 moment conditions and 7 unknown parameters. Using the same data as previously, involving 9,200 observations, ranging from January 1, 1969 to January 26, 2005, at the aggregate level, we obtain parameter estimates summarized in Table 3.2.

Inspection of the various parameters indicates that the information flow variable has a very small variance. Indeed, the parameter μ_2 takes a value of 0.003. This remark has also been made by Richardson and Smith (1994). Similarly to their findings, the information flow variable is highly positively skewed. Such a distribution is compatible with the assumption that the arrival rate tends to be constant most of the time interrupted by periods when there is a huge increase in the information arrival flow. These increases could be explained by information releases at a firm specific level or at a macroeconomic level.

¹⁹ See Hamilton (1994).

The Hansen's J statistic for the mixture of distributions hypothesis is 23.79. This statistics is distributed as a χ^2 with 3 degrees of freedom because there are 10 moment conditions and only 7 parameters. The mixture of distribution model gets therefore rejected even at the 1% significance level. Again, this finding is in line with Richardson and Smith (1994) who use, however, data at the firm level. For about 50% of the stocks, that they consider, the MDH is not successful. This is an important result suggesting that more research is required on the theoretical front.

It is possible to test if the information flow follows certain distributions. A first test is to ask if the informational flow is uniform. To perform this test, it is necessary to exhibit relations on the parameters that may be tested. Assume that the information flow obeys the uniform distribution

$$f(I) = \frac{1}{a} \quad \text{for } 0 \leq I \leq a,$$

where, for identification purpose, we impose the condition $E[I] = 1$, so that $a = 2$. This condition implies $V[I] = \frac{1}{2} \int_0^2 u^2 du - 1 = \frac{1}{3}$. A first observation is that a uniform distribution is symmetric around its mean. Hence its third moment must be zero, providing a first condition $\mu_3 = 0$. A second condition is that the kurtosis of the information arrival rate is $Ku[I] = 9/5$. It follows that we can test the null hypothesis that the information arrival rate is uniform by testing the following moment restrictions

$$H_0 : \begin{cases} \mu_3 = 0 \\ \mu_4 - \frac{9}{5}\mu_2^2 = 0. \end{cases} \quad (3.20)$$

To do so, we consider the vector of estimated restrictions

$$Q = \begin{bmatrix} \hat{\mu}_3 \\ \hat{\mu}_4 - \frac{9}{5}\hat{\mu}_2^2 \end{bmatrix},$$

where the variables with a hat represent the variables estimated by GMM. We also consider the matrix of gradients of the two moment conditions in expression (3.20)

$$G = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & -(18/5)\hat{\mu}_2 & 0 & 1 \end{bmatrix},$$

and compute the statistics $Q'(GVG')^{-1}Q$, where V is the covariance matrix of the parameter estimates contained in the vector $\hat{\theta}$. This statistics is distributed as a $\chi^2(2)$ under the null hypothesis.

Alternatively, we may test if the information arrival process could be generated by an inverse gamma distribution. For such a distribution, Richardson and Smith (1994) derive the moment restriction

$$H_0 : (\mu_4 + 4\mu_3 + 6\mu_2 + 1)(1 - \mu_3 - \mu_2) - (1 + \mu_2)^2(1 + \mu_3 + 3\mu_2) = 0.$$

They show furthermore that if the information arrival rate follows a Poisson distribution, then the null hypothesis becomes

Table 3.3. *Test of adequation of a given distribution of the data*

Model	χ^2 statistic	p-value
Uniform	5.508	0.064
Inverse gamma	1.769	0.184
Poisson	4.784	0.091
Log-normal	4.006	0.135

$$H_0 : \begin{cases} \mu_3 - \mu_2^2 = 0, \\ \mu_4 - \mu_2^3 - 3\mu_2^2 = 0. \end{cases}$$

Last, they consider the case of a log-normal distribution given by

$$H_0 : \begin{cases} \mu_3 - \mu_2^3 - 3\mu_2^2 = 0, \\ \mu_4 + 4(1 + \mu_2)^3 + 3 - (1 + \mu_2)^6 - 6(1 + \mu_2) = 0. \end{cases}$$

Returning to the data already used in the estimation, we obtain for these various moment restrictions the test statistic W reported in Table 3.3. For this data, at the conventional level of 5%, we do not reject any of the distributions considered.

Given its intuitive appeal, the high rejection rate of the mixture of distribution model is astonishing. It suggests that an alternative specification should be worked out. Also, so far the model neglects the fact that the information arrival could be time varying. For instance, periods with greater information arrival could be followed by others. Timmermann (2000) provides moments for a Markov switching model, in the spirit of Hamilton (1989). This type of model might be able to explain the failure of the current mixture of distributions tests.

3.4.4 Extensions

Measuring the information arrival flow

So far, we have considered a model that could explain some of the stylized facts of asset returns. We may ask now if it is possible to measure directly the information flow in markets. In his seminal contribution, Clark (1973) conjectures that volume might proxy the information flow. However, he rejects that idea on empirical grounds. Another idea, proposed by Lamoureux and Lastrapes (1990, 1994), is to investigate whether volume could capture some of the information contained in the variance of asset returns. This idea is based on (3.13), which shows that the variance is conditioned on the information flow. They use a so-called GARCH model, presented in detail in Chapter 4, and show that volume is a better predictor of the daily variability than a parametric model.

In this spirit, another issue is whether there exists a variable that might be more adequate to capture the information flow than volume. This is the

question addressed in Jones, Kaul, and Lipson (1994). They relate the daily variance of an asset to the daily volume and the number of trades on a given day. They show that daily volume is not as good a measure of information arrival as the number of trades. Recently, Chan and Fong (2000) focus on non-linearity in the information-flow activity measure. They find that the activity measures, when they are located in the upper quartile, affect asset volatility by proportionally more than when they are located in a lower quartile.

Measuring durations between trades

In the sections above, we discussed the effect of trading activity measures. Indirectly related to this discussion is the issue of the duration between trades. Indeed, given a fixed time interval, the greater the information arrival, and thus the number of transactions, the shorter the time between transactions. Even though this type of question is less directly related to the economic aspects explaining the kurtosis of assets, an emerging strand of models focuses on these aspects. We may mention Engle and Russel's (1998) autoregressive conditional duration (ACD) model. Other techniques such as the compound autoregressive model (CAR), developed by Darolles, Gouriéroux, and Jasiak (2001), or the autoregressive gamma model, by Gouriéroux and Jasiak (2003), also geared at modeling this type of behavior and might provide interesting alternatives to such a model.

3.5 A model of prices and quotes in a quote-driven market

We now consider an economic model that explains the quote arrival process and the price process. Easley, Kiefer, and O'Hara (1997) (EKO thereafter) provide a description of the temporal evolution of the best bid and ask prices as well as a description of the price flow. This paper also allows for an econometric estimation of the parameters involved in the model. Even though this contribution helps to understand the temporal evolution of the best bid and ask, there remains the issue of obtaining a satisfactory model of the limit order arrival at the level of the entire order book. Such a model would, obviously, be of a complex nature.²⁰

The model considered here is one where some trader aggregates the orders and matches them. This trader is called a *market maker*. In the EKO model, the market maker is risk neutral, meaning that the prices quoted will be the expected values of the stock's fundamental value, given that the trade will be

²⁰ At the time these lines are written, one paper appears to provide a model of the dynamic for the entire order book, which is the contribution of Goettler, Parlour, and Rajan (2005). This model is, however, of significant numerical complexity. Also, it is not obvious how to estimate and to test such a model.

either a sale or a purchase. Once prices are quoted, the actual price process will be given by a pre-specified trade arrival process. This trade arrival process will take into account that there are various types of traders, i.e., informed and uninformed traders. The latter are also called noise traders. Each time a trade takes place, given that quotes have been determined, we are able to obtain a sequence of traded prices.

3.5.1 A model based on the trade flow

Figure 3.6 represents a subtree of events related to the trades of an uninformed investor. Given that the uninformed traders have strictly no information on the fundamental value of the asset, they will sell with a probability of $\varepsilon/2$, and they will buy with the same probability of $\varepsilon/2$. It is also possible that no uninformed trader trades. This event will take place with a probability of $1 - \varepsilon$.

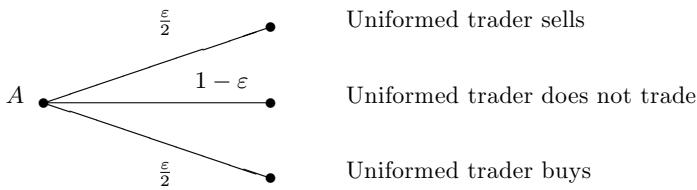


Fig. 3.6. Event subtree in the EKO model.

Figure 3.7 presents the entire event tree, referring to the subtree A at three of the terminal nodes. The first step on a given day is the release of a signal concerning the value V of the asset. If the signal is released, an event taking place with a probability of α , then the signal may be good \bar{V} or bad \underline{V} . The low signal, \underline{V} , occurs with a probability of δ and the high signal, \bar{V} , with a probability of $1 - \delta$. The signal is also fully revealing in that if the signal \underline{V} becomes known, the informed traders, if they decide to trade, will sell and not buy. As the figure indicates, the informed traders will only trade if there is a release of information. The probabilities of trading for the informed traders are symmetric in the sense that upon receiving a signal, they will trade with a probability of μ .

It is worth emphasizing, at this stage, that with the given definition of the trade flow, it is easy to simulate a sequence of trades. This is an exercise to which we presently turn. Figure 3.8 represents, along the horizontal axis, simulated daily trades. It is assumed that there are 100 trades per day, and that we consider 10 days. Each of the rows in the figure represents, therefore,

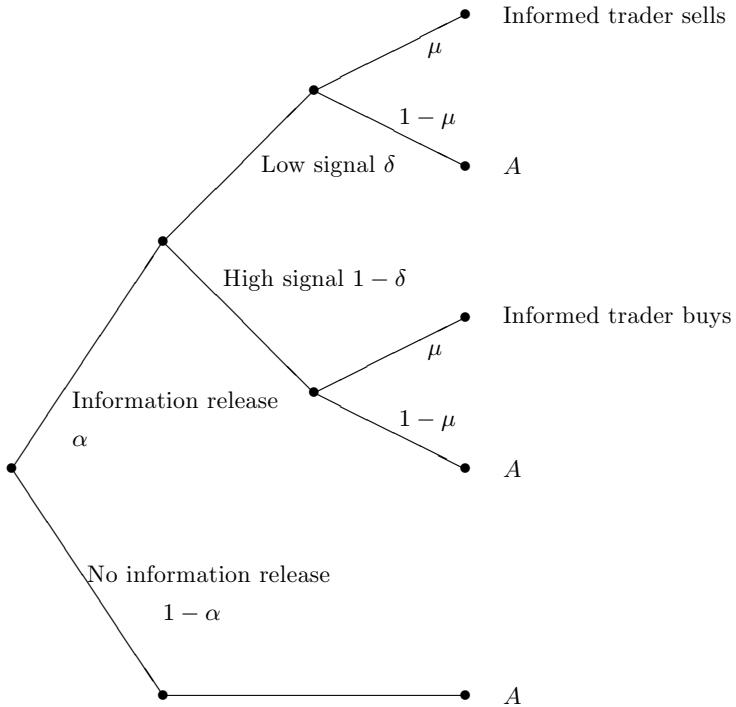


Fig. 3.7. Event tree in the EKO model.

the trade-flow as it occurs during one day. The parameters chosen are $\alpha = 0.6$, $\delta = 0.5$, $\mu = 0.2$, and $\varepsilon = 0.5$. The symbol \times represents a buy order and the symbol \square represents a sell order. A dot indicates that no trade occurred during a given time period.

By simply inspecting the sequence of trades that took place on given days, we can infer the signal that occurred on a given day. For instance, the first day, corresponding to the lowest row of symbols, is characterized by many trades, almost all of them being buy orders. This means that on this day, a good signal got revealed. Since the informed traders became informed about the signal being a good one, they bought stock all day long. Only occasionally did a sell order occur. This is when an uninformed trader got the chance of trading. If we consider the trades that took place on day 2, we notice that either there is no trade or that the pattern of buys and sells is random. This suggests that on this day no information got revealed. Only noise traders got to trade, and when they did, they traded at random.

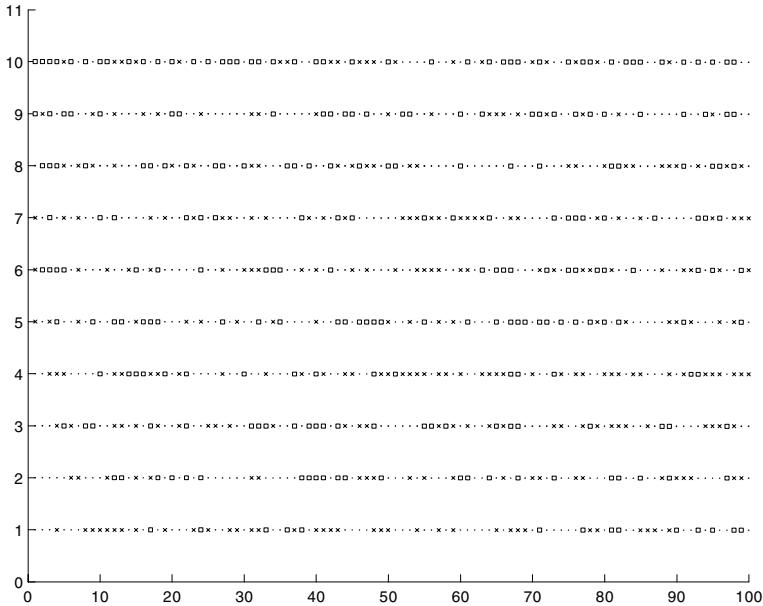


Fig. 3.8. Simulation 1 of daily trades using the EKO model.

Other sets of parameters may be used. For the set of parameters $\alpha = 0.6$, $\delta = 0.5$, $\mu = 0.4$, and $\varepsilon = 0.2$, we obtain Figure 3.9. Inspection of this figure reveals that, presently, there are days with essentially no trade, for instance days 7 and 8. These were days when no information got revealed. Presently, the parameter ε is higher than before, meaning that the uninformed trader will trade only very infrequently. Also, the informed trader will not trade when there is no signal. Thus, on days when no information got revealed, we do not expect many trades to occur. When an information is revealed, the parameter μ plays a role. This is the probability for the informed trader to trade. Since this probability has increased, we expect more informed trades to take place on days with information release. Since the signals are informative, we expect that the trades will be buy trades if news is good, and sell trades if news is bad. If we consider, for instance, day 5, we notice that nothing but sell orders took place. If we consider day 10, corresponding to the 10th row, we notice that there often are no trades and when there are any, they occur at random patterns. We conclude that on day 10, no signal got issued and that the trades are essentially noise trades. If we turn to the day before, corresponding to row 9, we notice that essentially only buy orders took place. On that day, a good signal must have been issued and informed traders decided to buy.

At this stage, the reader should understand that the trade flow is the key ingredient of this model. It is given by exogenous parameters and the trade flow has the same probability distribution from one day to the other.

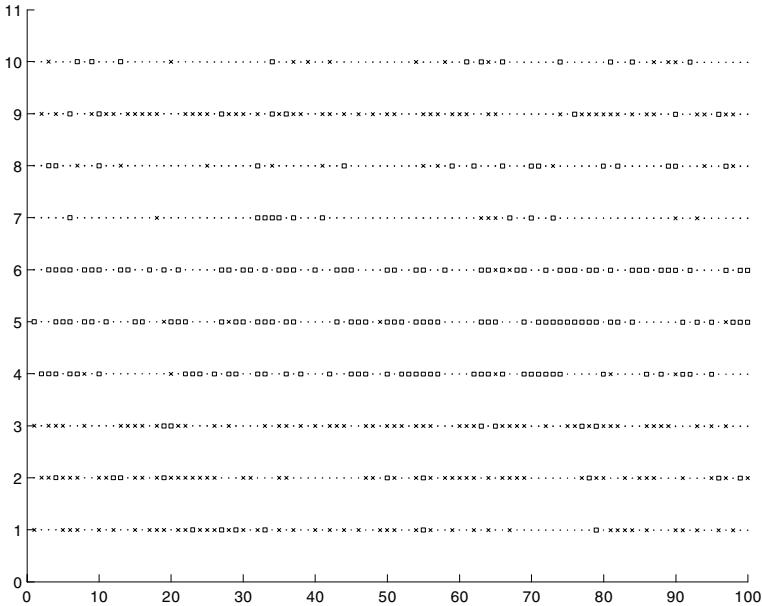


Fig. 3.9. Simulation 2 of daily trades using the EKO model.

Presently, we will again follow EKO and show how the parameters underlying the trade flow could be estimated with econometric methods. Then we will show how risk neutral market makers quote prices in this setting.

3.5.2 Estimating the parameters

The informed trader will perfectly know the signal and will then trade with a given probability. Her trades will be consistent with the signal, i.e., buy when news is good. The uninformed trader will buy or sell randomly. Each day is subdivided into intervals of equal length. During each interval, only one trader may trade. This means that trades take place sequentially.

We denote by B , S , and N the number of buy, sell, and no trade periods over a given period of time. Clearly, if the trading day has been subdivided into \bar{M} trading intervals, then $B + S + N = \bar{M}$. We regroup all the parameters into a single vector $\theta = (\alpha, \delta, \mu, \varepsilon)'$. As EKO, we denote by Ψ the signal that got revealed (no signal, \emptyset , high value of asset, H , or low value of asset, L). The unconditional probability of having B buy periods, S sell periods, and N no trade periods in the sample, i.e., the triplet $\{B, S, N\}$, is

$$\begin{aligned} \Pr[B, S, N | \theta] &= \Pr[B, S, N | \Psi = H; \theta] \times \Pr[\Psi = H; \theta] \\ &\quad + \Pr[B, S, N | \Psi = L; \theta] \times \Pr[\Psi = L; \theta] \\ &\quad + \Pr[B, S, N | \Psi = \emptyset; \theta] \times \Pr[\Psi = \emptyset; \theta]. \end{aligned} \tag{3.21}$$

Clearly, inspection of the trees reveals that

$$\Pr [\Psi = H; \theta] = \alpha(1 - \delta),$$

$$\Pr [\Psi = L; \theta] = \alpha\delta,$$

$$\Pr [\Psi = \emptyset; \theta] = 1 - \alpha.$$

Furthermore, by considering both the tree and the subtree, and counting the possibilities of the various events, we obtain

$$\begin{aligned}\Pr [B, S, N \mid \Psi = H; \theta] &= \left(\mu + (1 - \mu)\frac{\varepsilon}{2}\right)^B \left((1 - \mu)\frac{\varepsilon}{2}\right)^S ((1 - \mu)(1 - \varepsilon))^N, \\ \Pr [B, S, N \mid \Psi = L; \theta] &= \left((1 - \mu)\frac{\varepsilon}{2}\right)^B \left(\mu + (1 - \mu)\frac{\varepsilon}{2}\right)^S ((1 - \mu)(1 - \varepsilon))^N, \\ \Pr [B, S, N \mid \Psi = \emptyset; \theta] &= \left(\frac{\varepsilon}{2}\right)^B \left(\frac{\varepsilon}{2}\right)^S (1 - \varepsilon)^N = \left(\frac{\varepsilon}{2}\right)^{B+S} (1 - \varepsilon)^N.\end{aligned}$$

If we have observations of trades for various days, say $t = 1, \dots, T$, it is easy, after counting for each day (B_t, S_t, N_t) , i.e., the number of buy, sell, and no trade periods, to obtain an estimate for θ by maximizing

$$\hat{\theta} \in \arg \max_{\theta} \sum_{t=1}^T \log(\Pr[B_t, S_t, N_t \mid \theta]).$$

Using the simulation setting that also yielded the various trade flows, we estimate the parameters of the model with a simple Maximum-Likelihood procedure. Table 3.4 summarizes the results of these estimations. The first and second columns indicate the parameters chosen for the simulation. The third and fourth columns display the parameter estimates and standard errors using samples of size 20, and the last two columns show the estimates for samples of size $T = 100$.²¹ Inspection of the standard errors shows that the quality of the parameters α and δ is not as good as those concerning ε and μ .²² This is logical because α and δ are specific to given days and thus many days are required to get good estimates for the parameters. On the other hand, ε and μ are day specific, thus it is possible to obtain good estimates rather easily, because we can use all the available intraday information. As the sample size increases, the quality of the estimates increases as we would expect. Comparison of the true parameters with the estimates shows that the quality of the estimation is rather impressive.

As EKO show, it is also possible to take the model to actual data, but if this is done, several problems may arise:

²¹ In all those estimations, the parameters were constrained to belong to the set $[0.02, 0.98]$. Furthermore, the initial values are $(\alpha, \delta, \mu, \varepsilon) = (0.1, 0.5, 0.2, 0.2)$.

²² We report standard errors that are computed with the usual Hessian. Robust standard errors were very close to the ones reported.

Table 3.4. Parameter estimates of the EKO model depending on the sample size

	True value	$T = 20$		$T = 100$	
		Estimate	Std err.	Estimate	Std err.
α	0.7	0.750	(0.090)	0.693	(0.046)
δ	0.5	0.400	(0.125)	0.457	(0.061)
ε	0.2	0.205	(0.011)	0.205	(0.007)
μ	0.4	0.400	(0.013)	0.407	(0.007)

- 1) The model assumes that either a buy, or sell, or no trade may occur during a given time interval, but in practice several trades may arrive in a given interval.
- 2) The choice of time interval is not obvious, i.e., should trades get sampled each minute or each hour?
- 3) Given that the asset-specific parameters α and δ change over time, the assumption of stationarity over longer time horizons becomes questionable. This leads to a trade-off. If only few days are used to estimate the parameters, their accuracy will be low. If a large number of days are used, the parameters may have changed due to structural changes.

These caveats being made, EKO determine, using a very common stock, sampled over 5-minute intervals, that about 75% of all days information concerning the stock market becomes revealed. If an information gets revealed, about 17% of the trades will be information-based, meaning symmetrically, that 83% are noise trades. Good and bad news appear to be equally likely. The probability of an uninformed trader to participate in the market is found to be about 33%. All these parameters appear to be rather plausible.

3.5.3 The quote process

The model of EKO is most interesting in that it provides a plausible quote process, an issue to which we turn now. As stated, the market maker will quote prices conditionally on them being ask or bid prices, and conditionally on the amount of information that she has at some given point of the day. Each quote will be based on the expected value of the stock given that the next trade is either a sell or a buy. The computation of these expectations involves conditional probabilities.²³ The expressions for the best bid and ask

²³ It is useful to remember the rules of conditional probabilities. Let A and B be two events, then

$$\Pr[A|B] = \frac{\Pr[A, B]}{\Pr[B]} = \frac{\Pr[B|A] \times \Pr[A]}{\Pr[B]}.$$

If Ω is the set of all possible outcomes and \emptyset is the empty set, then

$$\Pr[A, \Omega] = \Pr[A], \quad \Pr[\Omega] = 1, \text{ and } \Pr[\emptyset] = 0.$$

prices during the day can then be derived. Let us, however, start with the first bid and ask quotes, i.e., the quotes at the opening of the market. We have

$$E[V|S] = \underline{V} \Pr[V = \underline{V}|S] + \bar{V} \times \Pr[V = \bar{V}|S]. \quad (3.22)$$

Now, if $\Psi = \emptyset$ means that there is no signal, we have

$$\begin{aligned} \Pr[V = \underline{V}|S] &= \Pr[V = \underline{V}, \Psi = \emptyset, \Psi = H, \Psi = L, S] / \Pr[S] \\ &= \Pr[V = \underline{V}|\Psi = \emptyset, S] \times \Pr[\Psi = \emptyset|S] \\ &\quad + \Pr[V = \underline{V}|\Psi = H, S] \times \Pr[\Psi = H|S] \\ &\quad + \Pr[V = \underline{V}|\Psi = L, S] \times \Pr[\Psi = L|S] \\ &= \delta \times \Pr[\Psi = \emptyset|S] + 1 \times \Pr[\Psi = L|S]. \end{aligned}$$

Here, the first line follows from the fact that Ψ can only take three values, \emptyset , H , and L . Thus, these values correspond to all possible outcomes for this random variable. The second line follows from the fact that Ψ can only take one value at a time. The last line, in turn, derives from the fact that if there is no signal, independently of the order, it adds no value if the signal is known (here to be \emptyset), then the probability of the asset having low value is δ . Also, given that the signal is assumed to be fully revealing, once $\Psi = L$ is known, it must be that $V = \underline{V}$ with a probability of one.

To compute the remaining probabilities $\Pr[\Psi = \emptyset|S]$ and $\Pr[\Psi = H|S]$, we use Bayes' rule once again. If we denote generically by X the values \emptyset , H , or L , we have

$$\Pr[\Psi = X|S] = \frac{\Pr[S|\Psi = X] \times \Pr[\Psi = X]}{\Pr[S]},$$

where

$$\begin{aligned} \Pr[S] &= \Pr[S|\Psi = \emptyset] \times \Pr[\Psi = \emptyset] + \Pr[S|\Psi = H] \times \Pr[\Psi = H] \\ &\quad + \Pr[S|\Psi = L] \times \Pr[\Psi = L]. \end{aligned}$$

From the tree, the various probabilities may now be obtained. As seen above, we have $\Pr[\Psi = H] = \alpha(1 - \delta)$, $\Pr[\Psi = L] = \alpha\delta$, and $\Pr[\Psi = \emptyset] = 1 - \alpha$. Also, $\Pr[S|\Psi = \emptyset] = \varepsilon/2$ because, if no signal occurred, only uninformed traders will trade, and in this case, they will do so with a probability of $\varepsilon/2$.

If we assume that the set of events A_j , $j = 1, \dots, n$ partitions the set Ω , meaning that if taken together, the union yields all possible events, $\bigcup_{i=1}^n A_i = \Omega$, and that no event is contained in two sets, $A_i \cap A_j = \emptyset$, then we obtain the following so-called Bayes' rules

$$\begin{aligned} \Pr[A_j|B] &= \Pr[B|A_j] \times \Pr[A_j] / \Pr[B], \\ \Pr[B] &= \Pr[B, \Omega] = \sum_{j=1}^n \Pr[B, A_j] = \sum_{j=1}^n \Pr[B|A_j] \times \Pr[A]. \end{aligned}$$

Eventually, we obtain

$$\begin{aligned}\Pr[\Psi = \emptyset | S] &= \frac{(1 - \alpha)\varepsilon/2}{\Pr[S]}, \\ \Pr[\Psi = L | S] &= \frac{(\mu + (1 - \mu)\varepsilon/2)\alpha\delta}{\Pr[S]}, \\ \Pr[\Psi = H | S] &= \frac{(1 - \alpha\mu)(1 - \delta)\varepsilon/2}{\Pr[S]},\end{aligned}$$

with

$$\Pr[S] = \mu\alpha\delta + (1 - \mu\alpha)\varepsilon/2.$$

Regrouping the various terms and using (3.22), we obtain the market maker's bid quote and ask quote

$$b_1 = E[V | S_1] = \frac{\delta\underline{V}(\alpha\mu + x) + (1 - \delta)\bar{V}x}{\delta\alpha\mu + x}, \quad (3.23)$$

$$a_1 = E[V | B_1] = \frac{\delta Vx + (1 - \delta)\bar{V}(\alpha\mu(1 - \alpha\mu) + x)}{(1 - \delta)\alpha\mu + x}, \quad (3.24)$$

where $x = (1 - \alpha\mu)\varepsilon/2$. During the day, the market maker will adjust her beliefs using all the information that occurred up to a given time of the day. During trading interval i , either N , no trade occurs, or S , a sale occurs, or B , a buy order occurs. Following EKO, it is convenient to perform a slight abuse of notation and denote by S , N , and B the number of sales, no trades, and buys that occurred up to time $B + S + N = i$. Using these notations, the trader will again quote prices given her conditional expectations similarly to the opening bid and asks. In trading interval $i + 1$, she quotes

$$\begin{aligned}b_{i+1} &= E[V | B, S + 1, N], \\ a_{i+1} &= E[V | B + 1, S, N],\end{aligned}$$

after the trading sequence B , S , and N , depending on whether the next order is a sell or a buy order. Conditional on these events, the market maker sets the bid and ask quotes as

$$\begin{aligned}b_{i+1} &= \underline{V}\Pr[V = \underline{V} | B, S + 1, N] + \bar{V}\Pr[V = \bar{V} | B, S + 1, N], \\ a_{i+1} &= \underline{V}\Pr[V = \underline{V} | B + 1, S, N] + \bar{V}\Pr[V = \bar{V} | B + 1, S, N],\end{aligned}$$

where

$$\begin{aligned}\Pr[V = \underline{V} | B, S + 1, N] &= \delta \times \Pr[\Psi = \emptyset | B, S + 1, N] \\ &\quad + 1 \times \Pr[\Psi = L | B, S + 1, N],\end{aligned}$$

because if no signal occurs, the probability of having $V = \underline{V}$ is δ . Also observing the order flow during the day does not add information to the fact

that the signal is $\Psi = \emptyset$. If a low signal occurs, given that the signal is fully revealing, it must be that with a probability of one, the value of the asset is low, \underline{V} .

To conclude, it is necessary to know the expression of $\Pr[\Psi = X|B, S, N]$ for $X = \emptyset, H, L$. Clearly, once a general formula is found for this expression, it must also hold for $S + 1, S + 2, \dots$, and similarly for values of B . We have

$$\Pr[\Psi = \emptyset|B, S, N] = \frac{\Pr[B, S, N|\Psi = 0] \times \Pr[\Psi = 0]}{\Pr[B, S, N]}.$$

The expression for $\Pr[B, S, N]$ has already been determined in (3.21) when it was discussed how to estimate the parameters relevant for the tree.

At this stage, it is possible to generate trades and thus to infer from them the type of signal received by the market. The following figures represent possible evaluations of the bid and the ask quotes throughout the day. Both figures were obtained from simulations based on the parameters $\alpha = 0.6$, $\delta = 0.5$, $\mu = 0.4$, and $\varepsilon = 0.2$. The first figure corresponds to the order flow of trading day 10, whereas the second figure corresponds to trading day 9. Both figures were obtained using as fundamental values $\underline{V} = 70$ and $\bar{V} = 130$.

Inspection of Figure 3.10 shows that day 10 was mainly characterized by no trades, and when trades took place, they were of random character. As already discussed, such a trading pattern is characteristic of days when no information is released and when trades are essentially due to noise traders. We notice on row 10 of Figure 3.9 that the first order is a buy order followed by three sell orders. We find this pattern in Figure 3.10 where the first trade takes place on the upper curve characterizing ask quotes. The following trades take place at the bid quotes. As the day evolves, the bid and ask quotes do not converge given that the trades are not informative.

Returning to Figure 3.9, day 9, we notice that on this day mainly buy orders took place. This means that informed traders were at work buying the asset based on their good-news knowledge. As Figure 3.11 displays, after a short period of hesitation, the quote process converged to the upper level of the stock value. For the given choice of parameters, this convergence is quick. For other sets of parameters, the convergence may not be so fast.

This model is therefore an important milestone on the road to obtain a better understanding of the quote process. Several aspects of the model may request further research. For instance, the theoretical model predicts for each day the same starting bid and ask prices. This problem appears to be easy to solve given that the fundamental value of the stock may be taken as changing. More severe questions may be the following: First, only one signal is released on a given day. What if there is a cascade of news? Second, the signal is fully revealing in that the firm value \underline{V} or \bar{V} becomes known by the traders. Third, are uninformed traders really as uninformed as they are modeled here? Also, may some learning occur for all agents as the day evolves. Last, the model only focuses on the best bid and best ask which is a limitation, because in practice the entire order book is observable.

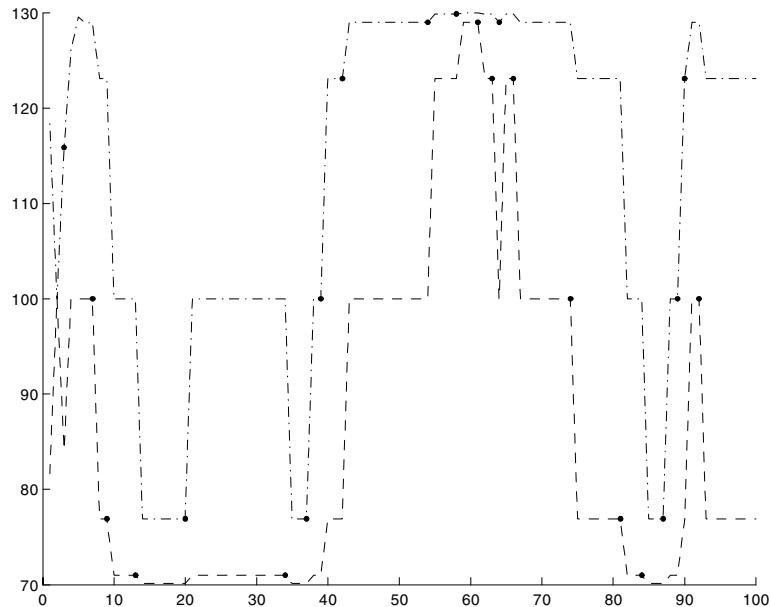


Fig. 3.10. Simulation of bid and ask quotes using the EKO model for day 10.

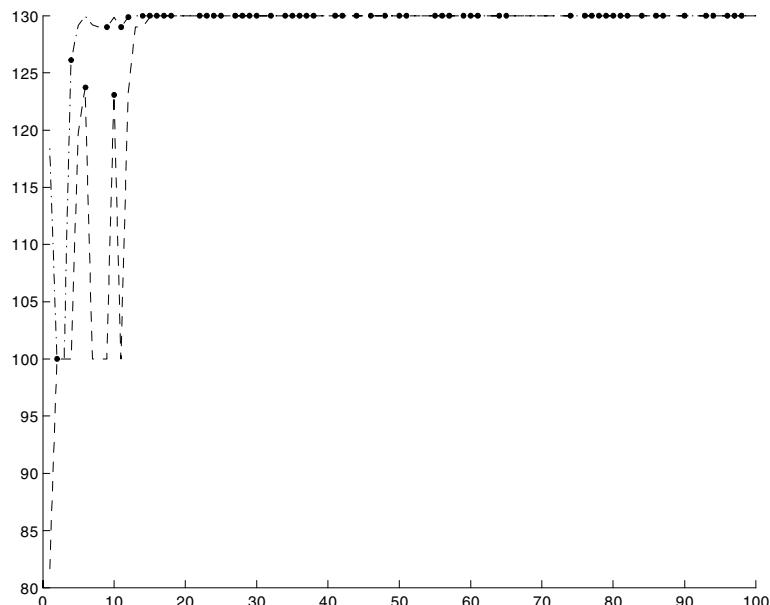


Fig. 3.11. Simulation of bid and ask quotes using the EKO model for day 9.

The EKO model that we just described provides a setting that gives a description for the best bid and ask price as well as for a price process. We will see in the next section that there exist models that consider the optimal sale, or buy, strategy, including fractioning an order over time, if traders have a market impact. To validate such models, it appears useful to have a realistic description of the entire order book. For instance, order arrivals may be modeled using a Poisson process. Such a strategy has been shown to yield an excellent econometric fit by Lo, MacKinlay, and Zhang (2002), however, the out-of-sample performance is rather weak. This reveals the importance that structural models of the order book could have. Probably more effort will be devoted to such structural models over the coming years.

3.5.4 Extension to the liquidation of a large portfolio

As discussed in the previous section, the microstructure of financial markets may have an important impact on a trader's decision such as, for instance, on a decision to liquidate a large portfolio. In this section, we wish to turn to this issue.

An early study that investigates how to optimally sell or buy a large portfolio is by Bertsimas and Lo (1998) who adopt a dynamic programming approach. Almgren and Chriss (2000) present a model within which they are able to obtain a relatively simple analytical expression for the optimal strategy. Their work has been later extended in various directions. In an empirically oriented direction, Dubil (2002) shows how the optimal liquidation strategy of Almgren and Chriss gets modified if the trader with a large position cares about Value-at-Risk rather than just about variance. In this case, the time horizon over which to liquidate becomes endogenous.

To gain some intuition on how to optimally liquidate a large portfolio, we present a simple discrete time version of the Almgren and Chriss (2000) model. The assumption of our simplified model is that there are 3 dates: 0, 1, and 2 delimitating two time periods, namely $[0, 1]$ and $[1, 2]$. The large trader holds an amount of X shares. This large trader decides to sell n_1 shares, respectively n_2 shares during each of the two periods. After time 2, the large trader no longer wishes to hold any shares at all.²⁴

The price of the stock is denoted by S_k at time $k \in \{0, 1, 2\}$. For simplicity, it is assumed that the price process evolves as a random walk: $S_k = S_{k-1} + \varepsilon_k$ where ε_k is some source of randomness with mean 0, and variance σ^2 , formally $E[\varepsilon_k] = 0$ and $V[\varepsilon_k] = \sigma^2$. In the long run, such a price description would not be realistic, however, the model may get extended to incorporate drifts, as well as positivity constraint, by assuming that prices are expressed as logs, i.e., $\log(S_k) = \log(S_{k-1}) + \varepsilon_k$.

²⁴ The distinction between temporary and permanent price impact goes back to Kraus and Stoll (1972) as well as Holthausen, Leftwich, and Mayers (1987, 1990) and Chan and Lakonishok (1993, 1995).

As the trader sells a given amount of shares in a given time period, she will permanently affect prices. This could mean that prices, in the case of a large sale, evolve as

$$S_k = S_{k-1} - an_k + \varepsilon_k, \quad a > 0.$$

For simplicity, it has been assumed here that the price impact is linear. Again, certain extensions could be made to incorporate non-linearities in the price impact. The parameter a therefore indicates by how much prices get permanently depressed in a sale of n_k shares.

As the trader sells her shares, she will also face a given demand schedule. Clearly, as she sells her shares, according to the shape of the limit order book, she will have an impact on the market. We may model this temporary impact by assuming that the average price at which the trade eventually takes place is

$$S'_k = S_{k-1} - bn_k.$$

Here, b is a scalar that indicates the temporary price impact. The larger b , the steeper the slope of the order book and the larger the impact on the price. Given that the trader only disposes of two periods during which she can sell her stocks, it must be that

$$X = n_1 + n_2.$$

Obviously, we have

$$\begin{aligned} S_1 &= S_0 - an_1 + \varepsilon_1 \\ S_2 &= S_0 - a(n_1 + n_2) + \varepsilon_1 + \varepsilon_2. \end{aligned}$$

We see that, at time 2, the prices are affected by the amount of shares traded, X , as well as by the random price changes. Given the linearity of the permanent price changes, it does not matter when the shares are actually traded.

Assuming a zero interest rate, the total revenue from the sale is

$$\begin{aligned} R &= n_1 S'_1 + n_2 S'_2 \\ &= n_1 (S_0 - bn_1) + n_2 (S_0 - an_1 + \varepsilon_1 - bn_2) \\ &= X S_0 - bn_1^2 + n_2 \varepsilon_1 - an_1 n_2 - bn_2^2. \end{aligned}$$

The expected revenue and its variance are

$$\begin{aligned} E[R] &= n_1 (S_0 - bn_1) + n_2 (S_0 - an_1 - bn_2), \\ V[R] &= n_2^2 \sigma^2. \end{aligned}$$

Almgren and Chriss (2000) assume that the trader maximizes a mean-variance utility function such as

$$U = E[R] - \frac{\lambda}{2} V[R],$$

where λ is a parameter that captures risk aversion. It is easy to express the utility as a function of n_1 only:

$$U = n_1(S_0 - bn_1) + (X - n_1)(S_0 - an_1 - b(X - n_1)) - \frac{1}{2}\lambda(X - n_1)^2\sigma^2.$$

The number of shares n_1 that maximizes utility is easily found to be

$$n_1 = \frac{2b - a + \lambda\sigma^2}{2(2b - a) + \lambda\sigma^2} X.$$

From there, it follows that

$$n_2 = X - n_1 = \frac{2b - a}{2(2b - a) + \lambda\sigma^2} X.$$

To perform a comparative static analysis, we also compute the derivatives:

$$\frac{\partial n_1}{\partial(2b - a)} = \frac{-\lambda\sigma^2 X}{D^2}, \quad (3.25)$$

$$\frac{\partial n_1}{\partial\lambda} = \frac{\sigma^2(2b - a)}{D^2}, \quad (3.26)$$

$$\frac{\partial n_1}{\partial(\sigma^2)} = \frac{\lambda(2b - a)}{D^2}, \quad (3.27)$$

where $D = 2(2b - a) + \lambda\sigma^2$. Clearly, if b increases, (3.25) shows that the number of shares sold during the first period should decrease. If the temporary price impact increases, the number of shares sold should decrease. If the permanent price impact, given by a , is higher, the future prices get depressed and hence, the more shares get sold immediately, the better.

Equation (3.26) shows that an increase of the risk aversion parameter has an impact on n_1 that depends on the sign of $2b - a$. If $b > a/2$ meaning that the temporary price impact is higher than the permanent price impact, then higher risk aversion will lead to increased sales during period 1. As (3.27) shows, a similar result is to hold if the risk level of the asset increases.

As this section shows, the liquidation of a large position or its purchase will affect the market. To model such an impact satisfactorily, the evolution of the entire order book must be understood and possibly be described by some sufficient statistics (latent factors). As demonstrated, to be able to obtain a satisfactory description of the order book, the degree of complexity increases quickly. For the moment, models that consider an optimal selling strategy take the time horizon as exogenous. The strategies considered also assume that the order book is not serially correlated and does not allow, for instance, volatility clustering. The strategies are also relatively limited in that a myopic selling strategy is obtained. There is no place for block trades in upstairs markets nor for hidden orders. It appears that in the following years, more research can be expected in this area.

In this chapter, we started by demonstrating that non-normality of asset returns is likely to be determined by information release. If we study the role of information release, we are naturally led to the microstructure literature, which investigates the interaction of investors on a market. Even though, at this stage, there does not appear to exist a general model for the order book with equilibrium considerations, at least the best bid and ask prices can be found using Bayesian learning. Once we consider the market microstructure, issues such as the optimal liquidation of a large portfolio become relevant. Too quick a sale may lead to losses due to the price impact. Too slow a sale may expose the seller to unnecessary risk. The models considered here generally address one issue at a time, and questions, such as the theoretical consequences of large sales on prices when traders learn, remain open issues.

Modeling Volatility

GARCH models have been developed to account for empirical features in the volatility of financial returns. As described in Chapter 2, return data appear stationary and barely autocorrelated. However, squared returns appear to be serially correlated. More specifically, the volatility of returns appears to cluster, so that large variations of price (positive or negative) are expected after a large variation of price (of either sign). Theoretical models justifying this volatility clustering have been discussed in Chapter 3. This serial correlation in squared returns has been initially modeled by Engle (1982) with the ARCH (AutoRegressive Conditional Heteroskedasticity) model and by Bollerslev (1986) with the Generalized ARCH (GARCH) model. Several extensions have been built on these early models to capture additional empirical features of asset returns. Section 4.2 presents the class of ARCH models, and Section 4.3 is devoted to GARCH models. In Section 4.4, we address the issues of introducing asymmetry in GARCH models. Finally, Section 4.5 describes GARCH models with jumps. Some aspects of these models are more particularly described: (1) how to forecast volatility using these models; (2) how to estimate (G)ARCH model efficiently; (3) how to test the presence of GARCH effects. Aggregation issues are discussed in Section 4.6. The multivariate extension is described in Chapter 6.

Another set of models, the so-called stochastic volatility models, has been introduced by Taylor (1982, 1986). These models, discussed in Section 4.7, introduce a second source of randomness in the price dynamics. Finally, a recent approach considers the use of intraday data to measure daily volatility. The construction and the use of this so-called realized volatility are described in Section 4.8.

4.1 Volatility at lower frequencies

In Chapter 3, we investigated the consequences of the functioning of financial markets on the intraday evolution of volatility. Often, such data is not avail-

able. Only data at some lower, say daily, frequency is available. For such a situation, so-called ARCH and GARCH models are useful in describing time-variation in conditional variance, which in turn explains, at least partially, the fat-tail phenomenon present in returns. Also, returns tend to be negatively correlated with changes in volatility, a feature that can be explained by the leverage effect (Black, 1976). It is easy to adapt GARCH models to accommodate this type of feature. This has given birth to so-called asymmetric GARCH models. As mentioned, GARCH models have been an extremely successful way to model several features of asset prices. A huge literature has emerged, and several surveys of GARCH models have been published. We may mention for instance Bollerslev, Chou, and Kroner (1992), Bera and Higgins (1993), Bollerslev, Engle, and Nelson (1994), Palm (1996), or more recently Li, Ling, and McAleer (2002). We will cover, in this chapter, the main features related to GARCH models, because these features will be building blocks for more complex models discussed in subsequent chapters.

The structure of a volatility model can be described as

$$x_t = \mu_t(\theta) + \varepsilon_t, \quad (4.1)$$

$$\varepsilon_t = \sigma_t(\theta) z_t, \quad (4.2)$$

where

$$\begin{aligned} \mu_t(\theta) &= E[x_t | \mathcal{F}_{t-1}], \\ \sigma_t^2(\theta) &= E[(x_t - \mu_t(\theta))^2 | \mathcal{F}_{t-1}]. \end{aligned}$$

In (4.1), we decompose the return x_t into a conditional mean $\mu_t(\theta)$ and a residual term ε_t . The dynamics of the conditional mean $\mu_t(\theta)$ may be an ARMA(p, q) process or could consist of seasonality features. \mathcal{F}_t is the information set available at time t . It may include current and past returns, current and past residuals, or any other variable known at time t . According to (4.2), the residual term ε_t has a volatility conditional on the information available at time $t - 1$, denoted σ_t , which may vary over time. θ is the vector of unknown parameters. The variable z_t will be assumed to follow some distribution with mean 0 and variance 1. Even though normality will be assumed in this chapter, this need not be. Other distributions such as the Student t distribution will be discussed in Chapter 5.

A volatility model is a model that describes the evolution of $\sigma_t^2(\theta)$. There are essentially two types of models for describing the dynamics of volatility:

1. In the first category, volatility is described as an exact function of a given set of variables. This category includes (G)ARCH models.
2. In the second category, volatility is described as a stochastic function. It includes Stochastic Volatility models.

The first sections of the chapter are devoted to the main aspects of the ARCH and GARCH methodology. We then turn to the stochastic volatility models, and eventually we describe the realized volatility approach.

4.2 ARCH model

The ARCH(p) model, originally introduced by Engle (1982), assumes that the conditional variance is a linear function of the past p squared innovations:

$$\sigma_t^2(\theta) = \omega + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_p \varepsilon_{t-p}^2 = \omega + \sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2. \quad (4.3)$$

According to (4.3), the conditional volatility is assumed to be a moving average of squared innovations. For this model to be well defined and the conditional variance to be positive, the parameters must satisfy the following constraints: $\omega > 0$, and $\alpha_i \geq 0$, $i = 1, \dots, p$.

The unconditional variance of innovation, denoted σ^2 , is the unconditional expectation of σ_t^2 : $\sigma^2 = E[\varepsilon_t^2] = E[E_{t-1}[\varepsilon_t^2]] = E[\sigma_t^2]$. In the case of an ARCH(p) process, it is easy to compute that $\sigma^2 = \omega / (1 - \sum_{i=1}^p \alpha_i)$. This shows that the process ε_t is covariance stationary if and only if the sum of the autoregressive parameters is less than one, i.e., $\sum_{i=1}^p \alpha_i < 1$. (See Section 4.3 below for more details on stationarity conditions.)

Although the innovation ε_t is serially uncorrelated, they are not time independent, because, defining $v_t = \varepsilon_t^2 - \sigma_t^2$, we can rewrite (4.3) as

$$\varepsilon_t^2 = \omega + \sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2 + v_t,$$

with $E_{t-1}[v_t] = 0$. Therefore, the plain ARCH(p) model can be viewed as an AR(p) model for the squared innovation ε_t^2 .

Notice that we do not describe in this section how to estimate an ARCH(p) model. Estimation is detailed in Section 4.3 devoted to GARCH models.

4.2.1 Forecasting

Forecasts of the ARCH model are obtained recursively. In the following, parameters with a hat correspond to estimates. Let t be the starting date for forecasting. Then, the 1-step ahead forecast for σ_{t+1}^2 is

$$\hat{\sigma}_t^2(1) = \hat{\omega} + \hat{\alpha}_1 \hat{\varepsilon}_t^2 + \cdots + \hat{\alpha}_p \hat{\varepsilon}_{t+1-p}^2 = \hat{\omega} + \sum_{i=1}^p \hat{\alpha}_i \hat{\varepsilon}_{t+1-i}^2,$$

where $\hat{\varepsilon}_t$ is the estimated residual. For the 2-step ahead forecast for σ_{t+2}^2 , we need a forecast of ε_{t+1}^2 . It is given by $\hat{\sigma}_t^2(1)$. We therefore obtain:

$$\hat{\sigma}_t^2(2) = \hat{\omega} + \hat{\alpha}_1 \hat{\sigma}_t^2(1) + \hat{\alpha}_2 \hat{\varepsilon}_t^2 + \cdots + \hat{\alpha}_p \hat{\varepsilon}_{t+2-p}^2.$$

The κ -step ahead forecast for $\sigma_{t+\kappa}^2$ is

$$\hat{\sigma}_t^2(\kappa) = \hat{\omega} + \hat{\alpha}_1 \hat{\sigma}_t^2(\kappa-1) + \cdots + \hat{\alpha}_p \hat{\sigma}_t^2(\kappa-p) = \hat{\omega} + \sum_{i=1}^p \hat{\alpha}_i \hat{\sigma}_t^2(\kappa-i).$$

with $\hat{\sigma}_t^2(\kappa-i) = \hat{\varepsilon}_{t+\kappa-i}^2$ if $\kappa-i \leq 0$.

4.2.2 Kurtosis of an ARCH model

ARCH models are able to generate excess kurtosis. Indeed, even if the conditional distribution for the standardized innovations $z_t = (x_t - \mu_t(\theta)) / \sigma_t(\theta)$ is assumed to be normal, the unconditional distribution for ε_t has fatter tails than the normal distribution. For instance, consider an ARCH(1) model

$$\sigma_t^2(\theta) = \omega + \alpha_1 \varepsilon_{t-1}^2,$$

where $\varepsilon_t = \sigma_t z_t$. The unconditional variance is given by $\sigma^2 = \omega / (1 - \alpha_1)$ so that we should have $\alpha_1 < 1$ for σ^2 to be finite. Now, under normality, the conditional fourth moment is given by

$$E[\varepsilon_t^4 | \mathcal{F}_{t-1}] = E[\sigma_t^4 z_t^4 | \mathcal{F}_{t-1}] = 3E[\sigma_t^4 | \mathcal{F}_{t-1}] = 3(\omega + \alpha_1 \varepsilon_{t-1}^2)^2,$$

whereas the unconditional fourth moment is

$$\begin{aligned} m_4 &= E[\varepsilon_t^4] = E[E[\sigma_t^4 z_t^4 | \mathcal{F}_{t-1}]] = 3E[(\omega + \alpha_1 \varepsilon_{t-1}^2)^2] \\ &= 3\left(\omega^2 + 2\frac{\omega^2 \alpha_1}{1 - \alpha_1} + \alpha_1^2 m_4\right) = \frac{3\omega^2(1 + \alpha_1)}{(1 - \alpha_1)(1 - 3\alpha_1^2)}. \end{aligned}$$

Since the fourth moment of ε_t is positive, we must have that $\alpha_1^2 < 1/3$. The unconditional kurtosis is

$$\kappa = \frac{m_4}{\sigma^4} = \frac{3\omega^2(1 + \alpha_1)}{(1 - \alpha_1)(1 - 3\alpha_1^2)} \times \frac{(1 - \alpha_1)^2}{\omega^2} = 3 \frac{1 - \alpha_1^2}{1 - 3\alpha_1^2},$$

which exceeds 3 for $\alpha_1 > 0$ and $3\alpha_1^2 < 1$.¹ So, the excess kurtosis is always positive and the tails of the distribution of ε_t are fatter than that of a normal distribution, even if the conditional distribution is normal.²

This result helps understanding the success of ARCH modeling, because it is able to capture both volatility clustering as well as fat-tailedness of the distribution. However, we will see in the following chapter that it is not able to capture all the fat-tailedness found in asset returns.

4.2.3 Testing for ARCH effects

A widely-used test for ARCH effects is the Lagrange-Multiplier (LM) test proposed by Engle (1982). A definite advantage of this test is that it is very easy to perform. Under the null hypothesis, the error term ε_t is assumed to be a normal white noise process $\varepsilon_t | \mathcal{F}_{t-1} \sim \mathcal{N}(0, \sigma^2)$. The alternative hypothesis is that the error term is driven by an ARCH(p) model, so that

¹ For $3\alpha_1^2 \geq 1$, the kurtosis becomes infinite.

² For a more general formulation of the ARCH model, it becomes more and more complicated to obtain the unconditional kurtosis, but the excess kurtosis is still positive. See Section 5.1.

$$\begin{aligned}\varepsilon_t &= \sigma_t z_t, \\ \sigma_t^2 &= \omega + \sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2.\end{aligned}$$

The test for ARCH(p) effects is based on the null hypothesis $H_0 : \alpha_1 = \dots = \alpha_p = 0$ against the alternative $H_a : \alpha_1 \geq 0, \dots, \alpha_p \geq 0$ with at least one strict inequality.

The LM test statistic of this hypothesis is shown to be asymptotically equivalent to the $T \times R^2$ test statistic, where T is the sample size and R^2 is computed from the regression (Engle, 1982)

$$\hat{\varepsilon}_t^2 = a_0 + a_1 \hat{\varepsilon}_{t-1}^2 + \dots + a_p \hat{\varepsilon}_{t-p}^2 + v_t.$$

Under the null of no ARCH effect, the LM and $T \times R^2$ test statistics are asymptotically distributed as a $\chi^2(p)$.

As an alternative form of the LM test, we may use the asymptotically equivalent portmanteau tests, such as the Ljung and Box (1978) statistic, for ε_t^2 .

It should be noticed that because the parameters of the ARCH model must be positive, the ARCH test should be formulated as a one-sided test. Such a test has been proposed by Demos and Sentana (1998).

4.2.4 ARCH-in-mean model

The ARCH-in-Mean model, proposed by Engle, Lilien, and Robbins (1987), was designed to capture the effect of conditional volatility on the conditional mean of the process:

$$x_t = \delta \varphi(\sigma_t) + \sigma_t \varepsilon_t, \quad (4.4)$$

where the conditional mean depends on conditional volatility σ_t . Equation (4.4) appears as a natural way to describe the trade-off between risk and expected return, which is consistent with many theories in finance. Several specifications have been proposed to incorporate volatility in the conditional equation. In particular, $\varphi(\sigma_t) = \sigma_t$ has been associated with the CAPM, whereas $\varphi(\sigma_t) = \sigma_t^2$ has been associated with the model of asset demand. Finally, $\varphi(\sigma_t) = \log(\sigma_t)$ has been estimated for instance by Engle, Lilien, and Robbins (1987).

Because of the dependence of the conditional mean on the conditional variance, several problems arise in the estimation and testing of ARCH-in-Mean models. The reason is that the second-order derivatives are not block-diagonal anymore, i.e.,

$$E \left[\frac{\partial^2 \ell_t(\theta, \delta)}{\partial \theta \partial \delta'} \right] \neq 0.$$

Consequently, the estimation of the conditional mean and the conditional variance parts of the model has to be performed jointly.

4.2.5 Illustration

As an illustration of some properties of ARCH models, we report the estimation of an ARCH(10) for the SP500, DAX, FT-SE, and Nikkei daily returns between January 1980 and August 2004. The complete model is given by

$$\begin{aligned}x_t &= \mu + \varepsilon_t, \\ \varepsilon_t &= \sigma_t z_t, \\ \sigma_t^2 &= \omega + \sum_{i=1}^{10} \alpha_i \varepsilon_{t-i}^2,\end{aligned}$$

so that the expected return is assumed to be constant over time.

Table 4.1 reports the parameter estimates and the associated t-stat for the volatility equation. We first notice that several lags of the innovation process are necessary to explain the dynamics of conditional volatility. For most markets, all lags up to the 10th are significant. This dynamics appears to be rather persistent, because the sum of parameters α_i is larger than 0.7. In the case of the Nikkei, we may even conclude that the volatility process is non-stationary, because the sum of ARCH parameters reaches its boundary of 1.

Figures 4.1 to 4.4 present the estimated dynamics of volatility for the different indices at hand. We notice that a large range of volatility patterns exists in the markets, the Nikkei being particularly agitated. Notice also that although the sum of α_i parameters is equal to 1, the Nikkei volatility does not seem to explode. These figures illustrate the volatility clustering phenomenon already mentioned. We finally observe that the October 1987 crash has a pronounced effect on the volatility of the SP500 and FT-SE daily returns. For the DAX and Nikkei daily returns, the level of volatility reached in October 1987 is similar to the level reached during other episodes.

4.3 GARCH model

Due to the large persistence in volatility, the ARCH model often requires a large p to fit the data. In such cases, it is more parsimonious to use the GARCH (Generalized ARCH) model proposed by Bollerslev (1986). The conditional variance of a GARCH(p, q) is:

$$\sigma_t^2 = \omega + \sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2. \quad (4.5)$$

For this model to be well defined and the conditional variance to be positive, the parameters must satisfy the following constraints: $\omega > 0$, $\alpha_i \geq 0$, $i = 1, \dots, p$, $\beta_j \geq 0$, for $j = 1, \dots, q$.

Table 4.1. Estimation of an ARCH(10) model for daily log-returns

	SP500	DAX	FT-SE	Nikkei
μ	0.0370 (0.0134)	0.0330 (0.0173)	0.0360 (0.0118)	0.0090 (0.0175)
ω	0.2770 (0.0192)	0.3010 (0.0243)	0.2350 (0.0161)	0.1910 (0.0166)
α_1	0.0970 (0.0144)	0.0520 (0.0136)	0.1150 (0.0161)	0.2240 (0.0194)
α_2	0.1000 (0.0155)	0.1090 (0.0166)	0.1280 (0.0178)	0.1630 (0.0184)
α_3	0.0750 (0.0148)	0.1950 (0.0197)	0.0710 (0.0160)	0.0880 (0.0158)
α_4	0.0830 (0.0154)	0.0950 (0.0172)	0.1080 (0.0178)	0.0890 (0.0152)
α_5	0.0750 (0.0151)	0.1030 (0.0166)	0.0400 (0.0143)	0.0780 (0.0146)
α_6	0.0580 (0.0137)	0.0780 (0.0158)	0.0730 (0.0161)	0.0650 (0.0138)
α_7	0.0650 (0.0147)	0.0740 (0.0152)	0.0590 (0.0156)	0.1020 (0.0162)
α_8	0.1040 (0.0165)	0.0980 (0.0175)	0.0690 (0.0155)	0.0590 (0.0140)
α_9	0.0660 (0.0151)	0.0700 (0.0160)	0.0400 (0.0131)	0.0560 (0.0135)
α_{10}	0.0570 (0.0136)	0.0160 (0.0121)	0.0370 (0.0143)	0.0760 (0.0147)
$\sum_{i=1}^{10} \alpha_i$	0.7800	0.8910	0.7380	1.0000
log-lik.	-8447.04	-9619.05	-7710.99	-9243.7

The unconditional variance is given by

$$\sigma^2 = \omega / \left(1 - \sum_{i=1}^p \alpha_i - \sum_{j=1}^q \beta_j \right).$$

Therefore, the process ε_t is covariance stationary if and only if $\sum_{i=1}^p \alpha_i + \sum_{j=1}^q \beta_j < 1$. Notice that this is a sufficient but not necessary condition for ε_t to be strictly stationary (Bollerslev, 1986, Nelson, 1990, Bougerol and Picard, 1992). To see why, assume the following GARCH(1, 1) model

$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2.$$

This model can be rewritten by successive iterations

$$\sigma_t^2 = \omega + \sigma_{t-1}^2 (\alpha_1 z_{t-1}^2 + \beta_1) = \omega \left(1 + \sum_{\tau=1}^{\infty} \prod_{k=1}^{\tau} (\alpha_1 z_{t-k}^2 + \beta_1) \right).$$

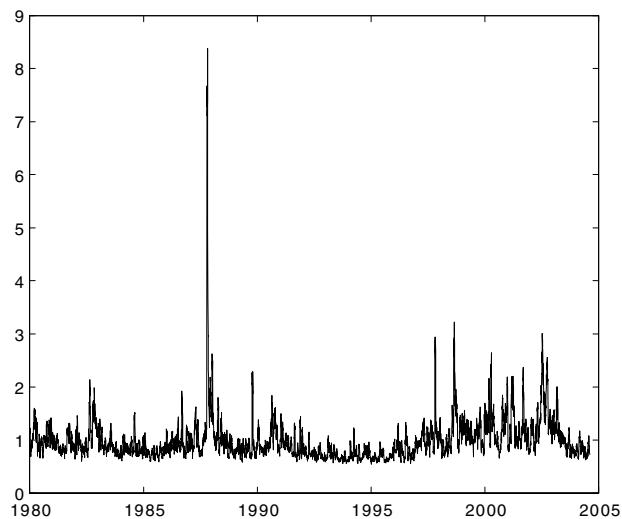


Fig. 4.1. *SP500 volatility estimated using an ARCH(10) model.*

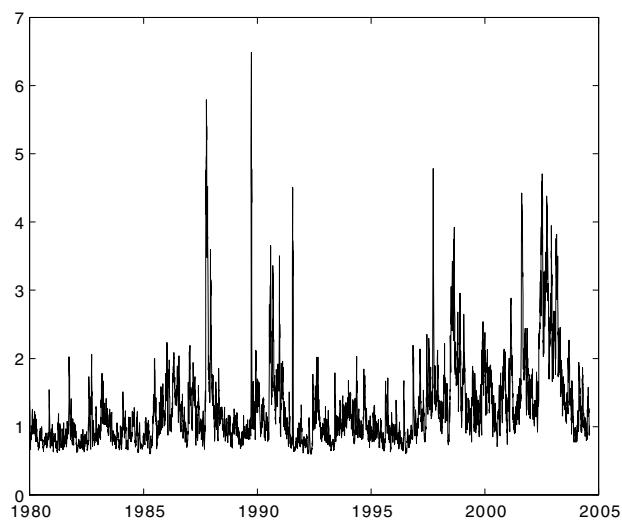


Fig. 4.2. *DAX volatility estimated using an ARCH(10) model.*

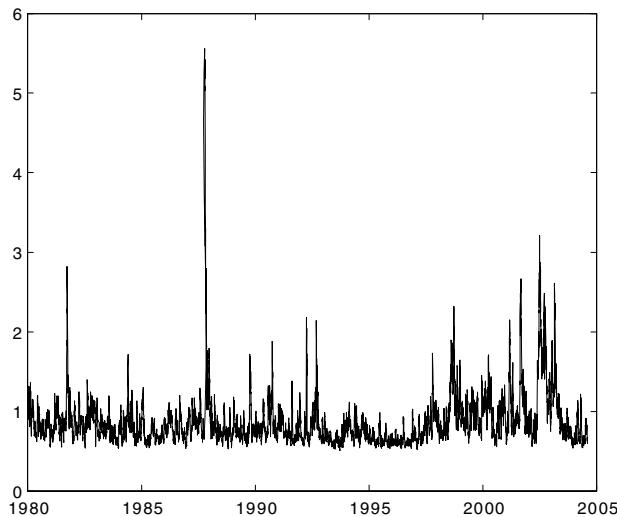


Fig. 4.3. *FT-SE volatility estimated using an $ARCH(10)$ model.*

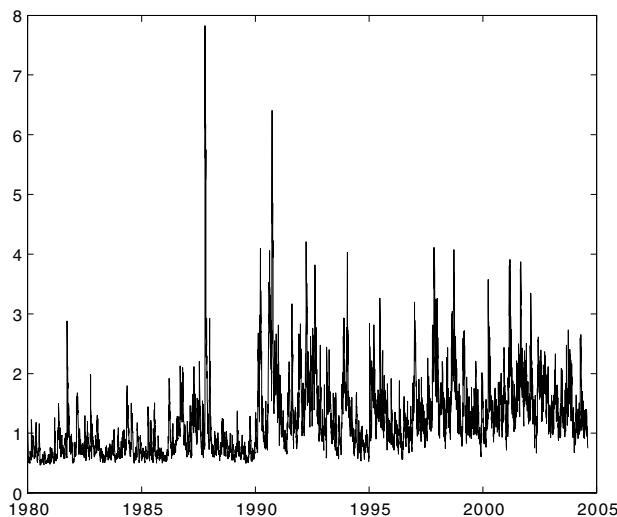


Fig. 4.4. *Nikkei volatility estimated using an $ARCH(10)$ model.*

Then, for this expression to converge toward a finite limit, so that the process ε_t is strictly stationary, we need

$$E [\log (\beta_1 + \alpha_1 z_{t-1}^2)] < 0.$$

We see that, due to the Jensen's inequality,

$$E [\log (\beta_1 + \alpha_1 z_{t-1}^2)] \leq \log (E [\beta_1 + \alpha_1 z_{t-1}^2]) = \log (\alpha_1 + \beta_1),$$

so that, even when $\alpha_1 + \beta_1 = 1$, the process is still strictly stationary.³

4.3.1 Forecasting

Forecasts of the GARCH model are obtained recursively in a similar way as for the ARCH model. Let t be the starting date for forecasting. Then, the 1-step ahead forecast for σ_{t+1}^2 is

$$\hat{\sigma}_t^2(1) = \hat{\omega} + \hat{\alpha}_1 \hat{\varepsilon}_t^2 + \hat{\beta}_1 \hat{\sigma}_t^2.$$

Since $\varepsilon_t^2 = \sigma_t^2 z_t^2$, the GARCH(1, 1) model can be rewritten as

$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 = \omega + (\alpha_1 + \beta_1) \sigma_{t-1}^2 + \alpha_1 \sigma_{t-1}^2 (z_{t-1}^2 - 1),$$

so that, at time $t + 2$, we have

$$\sigma_{t+2}^2 = \omega + (\alpha_1 + \beta_1) \sigma_{t+1}^2 + \alpha_1 \sigma_{t+1}^2 (z_{t+1}^2 - 1),$$

with $E [(z_{t+1}^2 - 1) | \mathcal{F}_t] = 0$. We deduce the following 2-step ahead forecast for σ_{t+2}^2 :

$$\hat{\sigma}_t^2(2) = \hat{\omega} + (\hat{\alpha}_1 + \hat{\beta}_1) \hat{\sigma}_t^2(1).$$

Generally speaking, the κ -step ahead forecast for $\sigma_{t+\kappa}^2$ is

$$\hat{\sigma}_t^2(\kappa) = \hat{\omega} + (\hat{\alpha}_1 + \hat{\beta}_1) \hat{\sigma}_t^2(\kappa - 1) \quad \text{for } \kappa > 1.$$

Alternatively, this last equation can be expressed as

$$\hat{\sigma}_t^2(\kappa) = \hat{\sigma}^2 + (\hat{\alpha}_1 + \hat{\beta}_1)^{\kappa-1} (\hat{\sigma}_t^2(1) - \hat{\sigma}^2) \quad \text{for } \kappa > 1,$$

where $\hat{\sigma}^2 = \hat{\omega}/(1 - \hat{\alpha}_1 - \hat{\beta}_1)$. This expression shows that $\hat{\sigma}_t^2(\kappa) \rightarrow \hat{\sigma}^2$ as $\kappa \rightarrow \infty$.

In empirical estimations, the parameters α and β tend to sum to a value that is very close to 1. It makes therefore sense to study the limit case, called an integrated GARCH model, where the sum of these parameters equals precisely one.

³ For the ARCH(1) model, the condition for strict stationary becomes

$$E [\log(\alpha_1 z_{t-1}^2)] < 0.$$

Consequently, even when $\alpha_1 = 1$, we have $E [\log(z_{t-1}^2)] \leq \log(1)$, so that the process ε_t is strictly, but not covariance, stationary.

4.3.2 Integrated GARCH model

When $\sum_{i=1}^p \alpha_i + \sum_{j=1}^q \beta_j = 1$, the unconditional variance σ^2 is not definite anymore. The process x_t is not covariance stationary anymore. However, it remains strictly stationary, because the unconditional density of ε_t does not change over time. Such a case corresponds to the Integrated GARCH (or IGARCH) described by Engle and Bollerslev (1986).

The IGARCH(1, 1) model can be written as

$$\sigma_t^2 = \omega + (1 - \beta_1) \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2,$$

with $0 < \beta_1 \leq 1$. We then see that

$$\sigma_t^2 = \omega + \sigma_{t-1}^2 + (1 - \beta_1) (\varepsilon_{t-1}^2 - \sigma_{t-1}^2).$$

So, the κ -step ahead forecast for $\sigma_{t+\kappa}^2$ is

$$\sigma_t^2(\kappa) = (\kappa - 1) \hat{\omega} + \sigma_t^2(1), \quad \text{for } \kappa \geq 1.$$

As it has been shown by Lumsdaine (1996), the statistical properties of the estimator of $(\omega, \beta_1)'$ are asymptotically normal.⁴

4.3.3 Estimation

Since this is the most widely adopted estimation, we only consider the Maximum Likelihood, ML, estimation of a GARCH(p, q) model. Assume that we have a time series of returns $\{x_1, \dots, x_T\}$, and denote $m = \max(p, q)$ the number of observations lost for initializing the process. Under the normality assumption, the likelihood function of a GARCH(p, q) model is

$$\begin{aligned} f(\varepsilon_1, \dots, \varepsilon_T | \theta) &= f(\varepsilon_T | \mathcal{F}_{T-1}) \times f(\varepsilon_{T-1} | \mathcal{F}_{T-2}) \times \dots \times f(\varepsilon_{m+1} | \mathcal{F}_m) \\ &\quad \times f(\varepsilon_1, \dots, \varepsilon_m | \theta) \\ &= \prod_{t=m+1}^T \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{\varepsilon_t^2}{2\sigma_t^2}\right) \times f(\varepsilon_1, \dots, \varepsilon_m | \theta), \end{aligned}$$

with $\theta = (\omega, \alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q)'$ the vector of unknown parameters and $f(\varepsilon_1, \dots, \varepsilon_m | \theta)$ is the joint pdf of $(\varepsilon_1, \dots, \varepsilon_m)$. In general, this term is simply dropped, to avoid specifying the joint distribution of the first observations. It is possible to justify this by observing that this term represents only one term in an usually long sequence of numbers. Omitting this term only introduces a negligible error in the likelihood. Hence, we focus on the *conditional likelihood function*

⁴ Given that regression parameters involving integrated time series have non-normal distributions, which tend to be difficult to characterize explicitly, this is rather welcome news.

$$f(\varepsilon_{m+1}, \dots, \varepsilon_T | \theta, \varepsilon_1, \dots, \varepsilon_m) = \prod_{t=m+1}^T \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{\varepsilon_t^2}{2\sigma_t^2}\right).$$

The ML estimator is obtained by maximizing this expression or, equivalently, the log-likelihood function

$$L_T(\theta | x_t, t = 1, \dots, T) = \sum_{t=m+1}^T \ell_t(\theta),$$

where

$$\ell_t(\theta) = -\frac{1}{2} \log(2\pi) - \frac{1}{2} \log(\sigma_t^2) - \frac{\varepsilon_t^2}{2\sigma_t^2}$$

is the log-likelihood of the observation t , with $\varepsilon_t = x_t - \mu_t(\theta)$.

Derivating the log-likelihood function yields the *gradient*

$$\frac{\partial \ell_t(\theta)}{\partial \theta} = -\frac{1}{2\sigma_t^2} \frac{\partial \sigma_t^2}{\partial \theta} + \frac{\varepsilon_t^2}{2\sigma_t^4} \frac{\partial \sigma_t^2}{\partial \theta} = \frac{1}{2\sigma_t^2} \frac{\partial \sigma_t^2}{\partial \theta} \left(\frac{\varepsilon_t^2}{\sigma_t^2} - 1 \right),$$

with

$$\begin{aligned} \frac{\partial \sigma_t^2}{\partial \omega} &= 1, \\ \frac{\partial \sigma_t^2}{\partial \alpha_i} &= \varepsilon_{t-i}^2, \quad \text{for } i = 1, \dots, p, \\ \frac{\partial \sigma_t^2}{\partial \beta_j} &= \sigma_{t-j}^2, \quad \text{for } j = 1, \dots, q. \end{aligned}$$

The second-order derivatives yield a matrix called *Hessian*, consisting of terms such as

$$\frac{\partial^2 \ell_t(\theta)}{\partial \theta \partial \theta'} = -\frac{1}{2\sigma_t^2} \frac{\partial \sigma_t^2}{\partial \theta} \frac{\partial \sigma_t^2}{\partial \theta'} \frac{\varepsilon_t^2}{\sigma_t^2} + \left(\frac{\varepsilon_t^2}{\sigma_t^2} - 1 \right) \frac{\partial}{\partial \theta'} \left(\frac{1}{2\sigma_t^2} \frac{\partial \sigma_t^2}{\partial \theta} \right),$$

with

$$\begin{aligned} E\left[\frac{\partial^2 \ell_t(\theta)}{\partial \theta \partial \theta'}\right] &= -E\left[\frac{1}{2\sigma_t^2} \frac{\partial \sigma_t^2}{\partial \theta} \frac{\partial \sigma_t^2}{\partial \theta'} \frac{\varepsilon_t^2}{\sigma_t^2} + \left(\frac{\varepsilon_t^2}{\sigma_t^2} - 1 \right) \frac{\partial}{\partial \theta'} \left(\frac{1}{2\sigma_t^2} \frac{\partial \sigma_t^2}{\partial \theta} \right)\right] \\ &= -E\left[\frac{1}{2\sigma_t^2} \frac{\partial \sigma_t^2}{\partial \theta} \frac{\partial \sigma_t^2}{\partial \theta'}\right]. \end{aligned}$$

If we define $\xi_t = (1, \varepsilon_{t-1}^2, \dots, \varepsilon_{t-p}^2, \sigma_{t-1}^2, \dots, \sigma_{t-q}^2)'$, we may write $\sigma_t^2 = \xi_t' \theta$, so that

$$E\left[\frac{\partial^2 \ell_t(\theta)}{\partial \theta \partial \theta'}\right] = -E\left[\frac{\xi_t \xi_t'}{2\sigma_t^2}\right].$$

The covariance matrix of the ML estimator is given by the inverse of the information matrix, which is defined as minus the expected Hessian

$$I(\theta) = -E \left[\frac{\partial^2 \ell_t(\theta)}{\partial \theta \partial \theta'} \right] = E \left[\frac{\xi_t \xi_t'}{2\sigma_t^2} \right].$$

This expression can be estimated by

$$\hat{I}(\theta) = \frac{1}{2T} \sum_{t=1}^T \frac{\hat{\xi}_t' \hat{\xi}_t}{\hat{\sigma}_t^2}.$$

As shown by Weiss (1986), provided standardized innovations have finite fourth moments, the ML estimator is consistent and has the following asymptotic distribution

$$\sqrt{T} (\hat{\theta} - \theta) \Rightarrow \mathcal{N}(0, I(\theta)^{-1}).$$

For a regression model of the type

$$\begin{aligned} x_t &= X_t \delta + \varepsilon_t, \\ \varepsilon_t &= \sigma_t z_t, \\ \sigma_t^2 &= \omega + \sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2, \end{aligned}$$

where X_t is a set of exogenous explanatory variables and δ a set of unknown parameters pertaining to the conditional mean equation, we should estimate parameters δ and θ simultaneously. Fortunately, it turns out that the second derivatives of the regression model log-likelihood are such that (Engle, 1982)

$$E \left[\frac{\partial^2 \ell_t(\theta, \delta)}{\partial \theta \partial \delta'} \right] = 0.$$

Consequently, the information matrix is block-diagonal and the two sets of parameters can be estimated separately.

In practice, the estimation of a GARCH model should be performed as follows:

1. Estimate the mean equation $x_t = \mu_t + \varepsilon_t$. Deduce $\hat{\varepsilon}_t = x_t - \hat{\mu}_t$ and $\hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^T \hat{\varepsilon}_t^2$.
2. Select initial values for θ , say $\theta^0 = (\omega^0, \alpha_1^0, \dots, \alpha_p^0, \beta_1^0, \dots, \beta_q^0)$ and set $\hat{\varepsilon}_1^2 = \dots = \hat{\varepsilon}_m^2 = \hat{\sigma}_1^2 = \dots = \hat{\sigma}_m^2 = \hat{\sigma}^2$, with $m = \max(p, q)$.
3. Compute the conditional variance $\hat{\sigma}_t^2 = \hat{\omega}^0 + \sum_{i=1}^p \alpha_i^0 \hat{\varepsilon}_{t-i}^2 + \sum_{j=1}^q \beta_j^0 \hat{\sigma}_{t-j}^2$ for $t = m+1, \dots, T$.
4. Compute the log-likelihood $L_T(\theta^0) = \sum_{t=m+1}^T \ell_t(\theta^0)$, where

$$\ell_t(\theta^0) = -\frac{1}{2} \log(2\pi) - \frac{1}{2} \log(\hat{\sigma}_t^2) - \frac{\hat{\varepsilon}_t^2}{2\hat{\sigma}_t^2}.$$

Change the value of parameters, say θ^1 , so that the log-likelihood increases $L_T(\theta^1) > L_T(\theta^0)$.

5. Iterate steps 3 and 4 until convergence of the log-likelihood to a fixed value.

4.3.4 Testing for GARCH effects

Assume we want to test the null hypothesis that the process is homoskedastic against the alternative that the variance follows a GARCH(1, 1) process. Then, the null is $H_0 : \alpha_1 = \beta_1 = 0$, against $H_a : \alpha_1 \geq 0, \beta_1 \geq 0$ with at least one strict inequality.

A difficulty in constructing a statistic for this test is that, under the null hypothesis, the model is not identified. Indeed, for a GARCH(1, 1) process $\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$, we have when $\alpha_1 = 0$

$$\sigma_t^2 = \omega + \beta_1 \sigma_{t-1}^2 = \frac{\omega}{1 - \beta_1} + \beta_1 \left(\sigma_{t-1}^2 - \frac{\omega}{1 - \beta_1} \right) = \vartheta + \beta_1 (\sigma_{t-1}^2 - \vartheta).$$

Therefore, if we set, at date $t = 0$, $\sigma_0^2 = \vartheta$, then $\sigma_t^2 = \vartheta$ for all t . Then, β_1 is unidentified under the null. Under the null, the GARCH(1, 1) effect and the ARCH(1) effect are locally equivalent (see Lee, 1991). So, the test of $H_0 : \alpha_1 = \beta_1 = 0$, against $H_a : \alpha_1 \geq 0, \beta_1 \geq 0$ with at least one strict inequality, is equivalent to the test of no ARCH(1).

4.3.5 Software to estimate ARCH and GARCH models

Many packages exist that estimate the parameters of ARCH and GARCH models. Without pretending to be complete, we wish to dress a list for the reader who is interested in the actual estimation of the models discussed in this book. First, there exist software programs that are specific to econometric time series analysis and that allow the estimation of a wide range of ARCH and GARCH models. We may mention in this list the programs Eviews and Rats. There is also a list of statistics programs that either have modules or packages that estimate such models. In this class, we have STATA, S-Plus (see also Zivot and Wang, 2003), R. This latter is an open source program based on the S language.

Then there is a list of programming languages that use matrix language, which is convenient if own development is required. We may mention in this family Ox, which is a free program. Ox being a matrix language, it is possible to extend this language by additional ad hoc, user-written, modules. A module that estimates GARCH has been written by Laurent and Peters (2005). In this family of matrix language programs we have also MATLAB and Gauss. For the former program, GARCH code may be found in the UCSD Garch toolbox, written by Kevin Sheppard.⁵ For the latter, we may consult Ronald Schoenberg's web page.⁶

⁵ See http://www.kevinsheppard.com/research/ucsd_garch/ucsd_garch.aspx. Some GARCH codes may also be found under <http://www.hec.unil.ch/mrockinger>.

⁶ See <http://faculty.washington.edu/rons/>.

For some really complex situations, we may also use Fortran code and sophisticated optimization code.⁷ Using such a relatively low-level programming language allows one to obtain a most efficient optimized compiled code. The longevity of Fortran guarantees the existence of repeatedly tested modules.

4.3.6 Illustration

Table 4.2 reports the estimation of a GARCH(1,1) model for the daily returns of our four market indices. The main interest of the GARCH model as compared with the initial ARCH model is that often the GARCH(1,1) specification is enough to capture most dynamic in volatility. Consequently, the number of unknown parameters is reduced considerably.

We notice that the parameters α_1 and β_1 are found to be rather close for all indices at hand. In most cases, estimates of the GARCH(1,1) model on returns yield $\alpha_1 + \beta_1 \approx 1$, so that the conditional variance is nearly integrated (Integrated GARCH model). We conclude that the volatility process may be unbounded although the return process is still strictly stationary.

The IGARCH process may reflect other dynamics for volatility. For instance, if the true model is a regime-switching model for volatility, estimating a GARCH model will generally result in a nearly integrated volatility process, a feature discussed in Gouriéroux and Jasiak (2001b). Finally, we observe that the log-likelihood significantly improves for all indices but the DAX, while the number of unknown parameters dramatically reduces. This suggests that the GARCH model is able to produce dynamics for the volatility that are very close to the ARCH model in a more parsimonious way.⁸

Table 4.2. Estimation of a GARCH(1,1) model for daily log-returns

	SP500	DAX	FT-SE	Nikkei
ω	0.012 (0.0014)	0.037 (0.0029)	0.023 (0.0028)	0.018 (0.0019)
α_1	0.072 (0.0016)	0.115 (0.0041)	0.102 (0.0064)	0.136 (0.0031)
β_1	0.919 (0.0031)	0.868 (0.0063)	0.872 (0.0084)	0.864 (0.0040)
$\alpha_1 + \beta_1$	0.991	0.983	0.974	1.000
log-lik.	-8398.130	-9630.780	-7686.280	-9198.150

⁷ Such as SNOPT or NPSOL, developed by Gill, Murray, and Saunders (1997, 1999). See also Jondeau and Rockinger (2003a, 2003b).

⁸ The figures for the GARCH volatility are not reproduced here, because they look extremely similar to those for the ARCH volatility.

4.4 Asymmetric GARCH models

GARCH models are able to capture volatility clustering as well as some amount of fat-tailedness. However, in these models, positive and negative past values have a symmetric effect on the conditional variance. Yet, an abundant literature has documented that negative returns “bad news” tend to be followed by larger increases in volatility than equally large positive returns “good news” (Black, 1976, Christie, 1982, or French, Schwert, and Stambaugh, 1987).

Pagan and Schwert (1990) and Engle and Ng (1993) have defined the concept of *news impact curve*, which relates past return shocks (news) to current volatility. This curve measures how new information is incorporated into volatility estimates. In the GARCH model, this curve is a quadratic function centered on $\varepsilon_{t-1} = 0$. For asymmetric models, the curve is designed to increase differently in the two directions.

Several parameterizations have been proposed to capture such asymmetry in the response of volatility to shocks. To mention some of them, we have the Exponential GARCH (EGARCH) model of Nelson (1991), the Threshold GARCH (TGARCH) of Zakoïan (1994), the GJR model of Glosten, Jagannathan, and Runkle (1993), and the Absolute GARCH (AGARCH) model of Hentschel (1995). Some general expressions have been designed to incorporate the most well-known asymmetric GARCH models (Higgins and Bera, 1992, or Hentschel, 1995).

4.4.1 EGARCH model

The Exponential GARCH (EGARCH) model has been introduced by Nelson (1991) to improve two annoying aspects of the standard GARCH model: (1) the parameters α and β have to be constrained during the course of the estimation to ensure positivity of the variance process; (2) as already mentioned, empirical evidence suggests an asymmetric response of volatility to shocks. In the EGARCH model, σ_t^2 depends on both the size and the sign of lagged shocks:

$$\log(\sigma_t^2) = \omega + \sum_{i=1}^p \alpha_i g(z_{t-i}) + \sum_{j=1}^q \beta_j \log(\sigma_{t-j}^2), \quad (4.6)$$

with $\alpha_1 = 1$ and $g(z_t) = [\gamma(|z_t| - E[|z_t|]) + \psi z_t]$. In this specification, parameters are not restricted to be non-negative, because the conditional volatility is always positive. The process is strictly and covariance stationary if and only if $\sum_{j=1}^q \beta_j < 1$. Notice that $E[|z_t|] = \sqrt{2/\pi}$ under normality.⁹

This specification introduces the desired asymmetry in the following way. Over the range $0 < z_t < \infty$, $g(z_t)$ is linear in z_t with slope $\gamma + \psi$, whereas over the range $-\infty < z_t \leq 0$, $g(z_t)$ is linear in z_t with slope $\gamma - \psi$.

⁹ For z_t distributed as a standardized Student t distribution with ν degrees of freedom, we have

$$E[|z_t|] = \frac{2\sqrt{\nu-2}\Gamma((\nu+1)/2)}{(\nu-1)\Gamma(\nu/2)\sqrt{\pi}}.$$

4.4.2 TGARCH model

The Threshold GARCH (TGARCH) model (Zakořan, 1994) is defined by:

$$\sigma_t = \omega + \sum_{i=1}^p [\alpha_i |\varepsilon_{t-i}| + \gamma_i \Pi_{t-i}^- |\varepsilon_{t-i}|] + \sum_{j=1}^q \beta_j \sigma_{t-j}, \quad (4.7)$$

with Π_t^- equal to 1 if $\varepsilon_t < 0$, and 0 otherwise.

The conditional volatility is positive when parameters satisfy $\omega > 0$, $\alpha_i \geq 0$, $\alpha_i + \gamma_i \geq 0$ and $\beta_j \geq 0$, for $i = 1, \dots, p$ and $j = 1, \dots, q$. Notice that ensuring covariance stationarity is not trivial, because it involves a rather non-linear constraint on parameters. In addition, this model does not nest the plain vanilla GARCH model.

4.4.3 GJR model

The *GJR* model (Glosten, Jagannathan, and Runkle, 1993) is closely related to the TGARCH model, because it is defined by:

$$\sigma_t^2 = \omega + \sum_{i=1}^p [\alpha_i \varepsilon_{t-i}^2 + \gamma_i \Pi_{t-i}^- \varepsilon_{t-i}^2] + \sum_{j=1}^q \beta_j \sigma_{t-j}^2. \quad (4.8)$$

The conditional volatility is positive when parameters satisfy $\omega > 0$, $\alpha_i \geq 0$, $\alpha_i + \gamma_i \geq 0$ and $\beta_j \geq 0$, for $i = 1, \dots, p$ and $j = 1, \dots, q$. The process is covariance stationary if and only if $\sum_{i=1}^p (\alpha_i + \gamma_i/2) + \sum_{j=1}^q \beta_j < 1$ (Hentschel, 1995).

4.4.4 Cox-Box transform

Hentschel (1995) has put forward that many members of the GARCH family can be embedded in a Cox-Box transform of the form (in the case $p = q = 1$)

$$\frac{\sigma_t^\gamma - 1}{\gamma} = \omega + \alpha_1 \sigma_{t-1}^\gamma f^\nu(z_{t-1}) + \beta_1 \frac{\sigma_{t-1}^\gamma - 1}{\gamma}, \quad (4.9)$$

where $f(z_t) = |z_t - b| - c(z_t - b)$ is the news impact curve introduced by Pagan and Schwert (1990) and Engle and Ng (1993). The GARCH model is obtained from $\gamma = \nu = 2$, $b = c = 0$. The EGARCH model arises as a limit case when $\gamma = 0$, $\nu = 1$ and $b = 0$. The TGARCH model corresponds to $\gamma = \nu = 1$, $b = 0$, and $|c| \leq 1$. The GJR model is obtained from $\gamma = \nu = 2$, and $b = 0$.

For a generalized error distribution (GED) with fat-tailedness parameter ν , analyzed by Nelson (1991), we have

$$E[|z_t|] = \lambda 2^{1/\nu} \Gamma(2/\nu) / \Gamma(1/\nu),$$

$$\text{where } \lambda = \left[2^{-2/\nu} \Gamma(1/\nu) / \Gamma(3/\nu) \right]^{1/2}.$$

4.4.5 News impact curve

A comparison of the GARCH and GJR models suggests an interesting metric by which to analyze the effect of news on conditional volatility. Holding constant the information dated $t - 2$ and earlier, we can examine the relation between ε_{t-1} and σ_t . This curve is called *news impact curve*. It relates past return shocks (news) to current volatility. It measures how new information is incorporated into volatility estimates.

For the GARCH(1, 1) model, the curve is a quadratic function centered on $\varepsilon_{t-1} = 0$,

$$\sigma_t^2 = A + \alpha_1 \varepsilon_{t-1}^2,$$

with $A = \omega + \beta_1 \sigma^2$ and σ^2 the unconditional variance.

For the GJR(1, 1) model, it has a minimum at $\varepsilon_{t-1} = 0$ with

$$\sigma_t^2 = \begin{cases} A + \alpha_1 \varepsilon_{t-1}^2 & \text{for } \varepsilon_{t-1} > 0, \\ A + (\alpha_1 + \gamma_1) \varepsilon_{t-1}^2 & \text{for } \varepsilon_{t-1} \leq 0, \end{cases}$$

with $A = \omega + \beta_1 \sigma^2$.

For the EGARCH(1, 1) model, it has a minimum at $\varepsilon_{t-1} = 0$ with

$$\sigma_t^2 = \begin{cases} A \exp\left(\frac{\psi+\gamma}{\sigma} \varepsilon_{t-1}\right) & \text{for } \varepsilon_{t-1} > 0, \\ A \exp\left(\frac{\psi-\gamma}{\sigma} \varepsilon_{t-1}\right) & \text{for } \varepsilon_{t-1} \leq 0, \end{cases}$$

with $A = \sigma^{2\beta_1} \exp\left(\omega - \gamma \sqrt{2/\pi}\right)$.

Figure 4.5 illustrates the difference in the news impact curve between a symmetric GARCH model, represented by a continuous line, and an asymmetric (GJR) model (with $\alpha_1 = 0.05$ and $\gamma_1 = 0.1$), represented by a dotted and dashed line. For the GARCH model, the effect on volatility of a shock is the same, be it positive or negative. For the GJR model, by contrast, we notice a much stronger effect on the volatility when the lagged shock is negative.

4.4.6 Partially non-parametric estimation

Engle and Ng (1993) have proposed an alternative approach to estimating the news impact curve based on a partially non-parametric procedure, inspired by Gouriéroux and Monfort (1992). The range of ε_t is divided into m^- intervals when ε_{t-1} is negative and m^+ intervals when ε_{t-1} is positive, with $2m = m^- + m^+$. We define the $2m+1$ break points τ_i , $i = -m^-, \dots, m^+$. Assuming $\tau_0 = 0$ allows one to test asymmetry. Finally, we define

$$\begin{aligned} P_{i,t} &= 1, \text{ if } \varepsilon_t > \tau_i \text{ and } 0 \text{ otherwise and} \\ N_{i,t} &= 1, \text{ if } \varepsilon_t \leq \tau_{-i} \text{ and } 0 \text{ otherwise.} \end{aligned}$$

Then, the piecewise linear specification of the heteroskedasticity function is

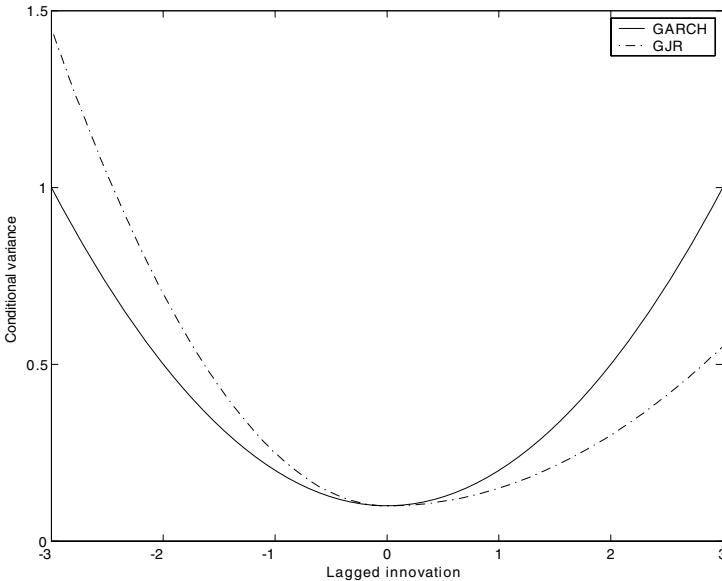


Fig. 4.5. News impact curve.

$$\sigma_t^2 = \omega + \beta\sigma_{t-1}^2 + \sum_{i=0}^{m^+} \theta_i P_{i,t-1} \times (\varepsilon_{t-1} - \tau_i) + \sum_{i=0}^{m^-} \delta_i N_{i,t-1} \times (\varepsilon_{t-1} - \tau_{-i}).$$

The functional form is a linear spline with knots at the τ_i s. Between 0 and τ_1 the slope is θ_0 , between τ_1 and τ_2 it is $\theta_0 + \theta_1$, and so forth. Between 0 and τ_{-1} , the slope is δ_0 , between τ_{-1} and τ_{-2} it is $\delta_0 + \delta_1$, and so forth. A simple approach to select the τ_i s is to use equally spaced bins with break points at $i \times \sigma$ for $i = 0, \pm 1, \pm 2, \dots$, where σ is the unconditional standard deviation of ε_t . In such case, we have

$$\sigma_t^2 = \omega + \beta\sigma_{t-1}^2 + \sum_{i=0}^m \theta_i P_{i,t-1} \times (\varepsilon_{t-1} - i\sigma) + \sum_{i=0}^m \delta_i N_{i,t-1} \times (\varepsilon_{t-1} + i\sigma).$$

4.4.7 Testing for asymmetric effects

Tests for asymmetry in the conditional variance have been proposed by Engle and Ng (1993). These tests examine whether we can predict the squared standardized residuals by some past variables that are not included in the hypothesized volatility model and may reflect an asymmetric response of volatility to shocks.

Assume that the hypothesized volatility model is a standard GARCH(1,1) model. The volatility dynamic thus depends on $\xi_{0,t} = (1, \varepsilon_{t-1}^2, \sigma_{t-1}^2)'$, so that we have $\sigma_{0,t}^2 = \xi_{0,t}' \theta_0$, with $\theta_0 = (\omega, \alpha_1, \beta_1)'$. Consider a set of

missing explanatory variables $\xi_{a,t}$ and their pertaining parameters θ_a , with $\log(\sigma_{a,t}^2) = \xi'_{a,t}\theta_a$.¹⁰ The nesting model can then be defined as

$$\log(\sigma_t^2) = \log\left(\xi'_{0,t}\theta_0\right) + \xi'_{a,t}\theta_a.$$

Now, define $z_t = \varepsilon_t/\sigma_{0,t}$ the standardized residual obtained under the hypothesized (GARCH) model, and consider the auxiliary regression

$$z_t^2 = \xi^{*'_{0,t}}\theta_0 + \xi^{*'_{a,t}}\theta_a + e_t,$$

with

$$\begin{aligned}\xi^{*}_{0,t} &= \frac{1}{\sigma_{0,t}^2} \frac{\partial \sigma_t^2}{\partial \theta_0}, \\ \xi^{*}_{a,t} &= \frac{1}{\sigma_{0,t}^2} \frac{\partial \sigma_t^2}{\partial \theta_a} = \frac{\sigma_t^2 \xi_{a,t}}{\sigma_{0,t}^2},\end{aligned}$$

and e_t the error term.

Then, we construct a Lagrange-Multiplier test statistic for the null hypothesis $H_0 : \theta_a = 0$ in the auxiliary regression above. Therefore, all variables in $\xi^{*}_{0,t}$ and $\xi^{*}_{a,t}$ are evaluated at ML estimator of θ_0 and $\theta_a = 0$, so that we finally have

$$\begin{aligned}\xi^{*}_{0,t} &= \frac{\xi_{0,t}}{\sigma_{0,t}^2}, \\ \xi^{*}_{a,t} &= \xi_{a,t},\end{aligned}$$

In addition, under the null, the left-hand side variable z_t^2 is theoretically orthogonal to the two right-hand side sets of variables. Therefore, the LM test statistic can be computed as the $T \times R^2$ associated with the regression. Under the null, the test statistic is asymptotically distributed as a χ^2 with a degree of freedom equal to the number of restrictions.

In practice, variables in $\xi_{a,t}$ have to be chosen in order to test the presence of asymmetric effects. Engle and Ng (1993) propose a series of complementary tests based on the following regressions:

$$z_t^2 = a + b\Pi_{t-1}^- + \xi^{*'_{0,t}}\theta_0 + e_t, \quad (4.10)$$

$$z_t^2 = a + b\Pi_{t-1}^-\varepsilon_{t-1} + \xi^{*'_{0,t}}\theta_0 + e_t, \quad (4.11)$$

$$z_t^2 = a + b\Pi_{t-1}^+\varepsilon_{t-1} + \xi^{*'_{0,t}}\theta_0 + e_t, \quad (4.12)$$

$$z_t^2 = a + b_1\Pi_{t-1}^- + b_2\Pi_{t-1}^-\varepsilon_{t-1} + b_3\Pi_{t-1}^+\varepsilon_{t-1} + \xi^{*'_{0,t}}\theta_0 + e_t, \quad (4.13)$$

¹⁰ For instance, the missing variables may have the form of an EGARCH(1, 1) model.

The variables involved in this model would then be

$$\xi_{a,t} = \left(\frac{\varepsilon_{t-1}}{\sigma_{t-1}}, \left(\left| \frac{\varepsilon_{t-1}}{\sigma_{t-1}} \right| - \sqrt{2/\pi} \right), \log(\sigma_{t-1}^2) \right)',$$

and the parameter set would be $\theta_a = (\tilde{\psi}, \tilde{\gamma}, \tilde{\beta}_1)'$, with $\log(\sigma_{a,t}^2) = \xi'_{a,t}\theta_a$.

where $\Pi_{t-1}^- = 1_{(\varepsilon_{t-1} \leq 0)}$ and $\Pi_{t-1}^+ = 1 - \Pi_{t-1}^-$. The first test, the *Sign bias test*, examines the impact of positive and negative shocks on the conditional variance not predicted by the model under consideration. This is done by testing whether the variable Π_{t-1}^- has a predictive power on squared standardized residuals z_t^2 (equation (4.10)). The test statistic is the t-ratio of parameter b . The *Negative size bias test* examines whether the given model can explain the different effects that large and small negative shocks have on the conditional variance. The variable used for this test is $\Pi_{t-1}^- \varepsilon_{t-1}$ (equation (4.11)). The *Positive size bias test* examines whether the linear model can explain the different effects that large and small positive shocks have on the conditional variance. The variable used for this test is $\Pi_{t-1}^+ \varepsilon_{t-1}$ (equation (4.12)). The last test considers the three previous hypotheses simultaneously. Under the null hypothesis $b_1 = b_2 = b_3 = 0$ and $\theta_0 = 0$, the test statistic is the $T \times R^2$ associated with (4.13). It is distributed as a $\chi^2(3)$. Notice that, if $\xi_{0,t}^*$ is not included in the regression (4.13), the test will be conservative.

4.4.8 Illustration

In the following of our previous estimates, we now consider the estimation of asymmetric GARCH models. Table 4.3 reports the estimation of three asymmetric GARCH(1,1) models for daily returns under study: the EGARCH, GJR, and TGARCH models. Loosely speaking, the three models provide rather similar evidence: in almost all cases, the estimates suggest, as expected, that “bad news” has a stronger effect on volatility than “good news”.¹¹

If we turn to the log-likelihood estimates, we observe that no model systematically dominates the others. For instance, the EGARCH model performs very well for the Nikkei but very badly for the FT-SE. In contrast, the GJR model captures the dynamic of the DAX volatility best.

4.5 GARCH model with jumps

As seen in Chapter 3, the release of news is likely to generate volatility. So far, in this chapter, we have addressed the issue of modeling volatility under the assumption that news arrives, in a way that may be modeled with a GARCH model. An alternative is to assume that there is a sudden information release, followed by a period where this news gets gradually incorporated into the market. This suggests that in addition to the GARCH phenomenon, it is possible that returns suddenly jump. Such a model may be reasonable because the residuals of a GARCH model are in general not distributed according to a Gaussian distribution (see Chapter 5). This observation suggests that

¹¹ Notice that a negative asymmetry translates into a negative ψ parameter for the EGARCH model but a positive γ_1 for the GJR and TGARCH models. The only exception is therefore the estimate of the EGARCH model for the FT-SE.

Table 4.3. Parameter estimates of asymmetric GARCH models

	SP500	DAX	FT-SE	Nikkei
GARCH(1, 1)				
ω	0.012 (0.0014)	0.037 (0.0029)	0.023 (0.0028)	0.018 (0.0019)
α_1	0.072 (0.0016)	0.115 (0.0041)	0.102 (0.0064)	0.136 (0.0031)
β_1	0.919 (0.0031)	0.868 (0.0063)	0.872 (0.0084)	0.864 (0.0040)
log-lik.	-8398.13	-9630.78	-7686.28	-9198.15
EGARCH(1, 1)				
ω	0.003 (0.0011)	0.013 (0.0010)	-0.036 (0.0069)	0.016 (0.0014)
γ	0.133 (0.0064)	0.174 (0.0061)	0.441 (0.0223)	0.244 (0.0061)
ψ	-0.077 (-0.0042)	-0.062 (-0.0043)	0.036 (0.0107)	-0.111 (-0.0046)
β_1	0.981 (0.0016)	0.974 (0.0023)	0.715 (0.0156)	0.974 (0.0020)
log-lik.	-8325.7	-9610.37	-7905.72	-9087.97
GJR(1, 1)				
ω	0.018 (0.0014)	0.041 (0.0031)	0.023 (0.0026)	0.02 (0.0019)
α_1	0.025 (0.0047)	0.053 (0.0074)	0.055 (0.0079)	0.053 (0.0051)
γ_1	0.095 (0.0052)	0.1 (0.0069)	0.069 (0.0067)	0.168 (0.0064)
β_1	0.911 (0.0038)	0.874 (0.0069)	0.882 (0.0084)	0.863 (0.0042)
log-lik.	-8343.49	-9595.55	-7664.5	-9103.32
TGARCH(1, 1)				
ω	0.019 (0.0029)	0.034 (0.0044)	0.027 (0.0039)	0.034 (0.0041)
α_1	0.029 (0.0057)	0.061 (0.0078)	0.07 (0.0086)	0.059 (0.0068)
γ_1	0.088 (0.0079)	0.074 (0.0082)	0.063 (0.0079)	0.129 (0.0098)
β_1	0.926 (0.0063)	0.897 (0.0085)	0.89 (0.0096)	0.876 (0.0082)
log-lik.	-8325.7	-9604.9	-7661.7	-9105.1

additional phenomena take place that are not captured by the simple GARCH model. In this section, we discuss some contributions that combine discrete GARCH models with jump models that may capture this sudden release of information.

An early contribution in this field is Press (1967) who considers the possibility that returns are generated by a superposition of a Gaussian distribution and a jump component. For estimating this model, Press (1967) proposes using the method of moments.¹² The next milestone is Jorion (1989) who assumes that returns are generated by an ARCH model that incorporates jumps. He finds that exchange rates are jumpier than stock market returns. This model has been extended by Vlaar and Palm, (1993) to GARCH models. As a more recent contribution, we may mention Chan and Maheu (2002) and Maheu and McCurdy (2004). Since their contribution nests the earlier contributions that appeared in the literature, we focus here on their model.

4.5.1 A model with time-varying jump intensity

Consider a time series of returns x_t , $t = 1, \dots, T$. The following model, named by Maheu and McCurdy (2004) GARJI, follows the spirit of the GARCH literature and superposes a compound Poisson process to the GARCH with a time-varying jump intensity. The model is

$$x_t = \mu + \varepsilon_t, \quad (4.14)$$

$$\varepsilon_t = \varepsilon_{1,t} + \varepsilon_{2,t}, \quad (4.15)$$

$$\varepsilon_{1,t} = \sigma_t z_t, \quad (4.16)$$

$$z_t \sim \mathcal{N}(0, 1), \quad (4.17)$$

$$\sigma_t^2 = \omega + g(\Lambda, \mathcal{F}_{t-1}) \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2, \quad (4.18)$$

$$\begin{aligned} g(\Lambda, \Phi_{t-1}) &= \exp(\alpha + \alpha_J E[n_{t-1} | \mathcal{F}_{t-1}]) \\ &\quad + 1_{\{\varepsilon_{t-1} < 0\}} [\alpha_N + \alpha_{N,J} E[n_{t-1} | \mathcal{F}_{t-1}]] \end{aligned} \quad (4.19)$$

Equation (4.14) indicates that returns have a mean μ and an innovation, ε_t . This innovation consists, as (4.15) indicates, of two components. The first component follows a GARCH model. This is described by the (4.16)–(4.19). The source of uncertainty of the GARCH, z_t , is assumed to be normally distributed as indicated in (4.17). The second component involves a Poisson process n_t whose specification is discussed below. The volatility dynamics in (4.19) depends on the innovation ε_t , that is both the part coming from the GARCH as well as from the jumps. This is a reasonable assumption, in that we would expect that a sudden burst of information, reflected in an increase of ε_t , affects subsequent volatility.

¹² These estimates, obtained for a simple model without the GARCH effect, may be used as starting values for more complicated models.

Equation (4.19) allows for a general news impact curve nesting the GJR and Threshold GARCH models. As in general for conditional models, the specifications depend on some information set \mathcal{F}_{t-1} . This can be past return variables or other observable variables. The only requirement, in order for some variable to belong to that information set, is that the variable is known at time t . In (4.19) figures an indicator variable $1_{\{\varepsilon < 0\}}$, taking the value of 1 if $\varepsilon < 0$, and the value 0, otherwise. The chosen specification allows for the introduction of a differential impact if past news are deemed good or bad. If past news are “business as usual”, in the sense that no jump occurred, and are positive, then the impact on current volatility will be $\exp(\alpha)\varepsilon_{t-1}^2$. If no jump took place but news is bad, the volatility impact becomes $\exp(\alpha + \alpha_N)\varepsilon_{t-1}^2$. If a jump took place, with good news, the impact is $\exp(\alpha + \alpha_J)\varepsilon_{t-1}^2$, and last, if there is a jump with bad news, then the impact becomes $\exp(\alpha + \alpha_J + \alpha_N + \alpha_{N,J})$.

The arrival rate of the jumps is assumed to follow a Poisson distribution. Leaving the derivation of the Poisson distribution to Chapter 16, we wish to mention here that the probability that $n_t = j$ jumps occurred on a given day t is given by

$$\Pr[n_t = j | \mathcal{F}_{t-1}] = e^{-\lambda_t} \frac{\lambda_t^j}{j!}. \quad (4.20)$$

The parameter λ_t measures the intensity of jumps. It measures the expected value of jump arrivals over unit time intervals. Here, it is assumed that the jump frequency is time dependent. A Poisson process with time-varying parameter is also called a non-homogenous Poisson process in the statistics literature. The jump component, $\varepsilon_{2,t}$, can be written as

$$\varepsilon_{2,t} = \sum_{k=1}^{n_t} Y_{t,k} - \lambda_t \theta, \quad (4.21)$$

$$Y_{t,k} \sim \mathcal{N}(\theta, \delta^2). \quad (4.22)$$

The first equation, (4.21), specifies that the jump component is the sum of a certain amount of shocks, assumed here to be Gaussian, with the parameters indicated in (4.22). Jumps have expected size θ and dispersion δ^2 . The term $\lambda_t \theta$ in (4.21) has been introduced so that, for identification purposes, $E[\varepsilon_{2,t}] = 0$.¹³

In order to estimate the model, it is necessary to provide an expression for the jump intensity. Maheu and McCurdy (2004) assume that it evolves through time as an autoregressive process given by

$$\lambda_t = \lambda_0 + \rho \lambda_{t-1} + \gamma \xi_{t-1}. \quad (4.23)$$

¹³ Without this condition, the expected return would be $E[x_t] = \mu + \theta \lambda_t$ and it would not be possible to distinguish the μ from the θ . By subtracting $\lambda_t \theta$, we obtain a so-called compensated Poisson process. Its expected value is zero.

The source of innovation to the jump component is assumed to be conditional on time $t - 1$ rather than being generated at time t . It is this key assumption that allows Maheu and McCurdy (2004) to estimate the entire model. Using the definition of an innovation as being the difference between the expected value and the actual value, they obtain

$$\xi_{t-1} = E[n_{t-1}|\mathcal{F}_{t-1}] - \lambda_{t-1} = \sum_{j=0}^{\infty} j \Pr[n_{t-1} = j|\mathcal{F}_{t-1}] - \lambda_{t-1}. \quad (4.24)$$

This expression could be estimated if $\Pr[n_{t-1} = j|\mathcal{F}_{t-1}]$ were known. As shown by Maheu and McCurdy (2004), this expression may be inferred using Bayes' formula. We have

$$\Pr[n_t = j|\mathcal{F}_t] = \Pr[n_t = j|x_t, \mathcal{F}_{t-1}] = \frac{\Pr[n_t = j, x_t, \mathcal{F}_{t-1}]}{\Pr[x_t, \mathcal{F}_{t-1}]} \quad (4.25)$$

$$= \frac{\Pr[x_t|n_t = j, \mathcal{F}_{t-1}] \times \Pr[n_t = j|\mathcal{F}_{t-1}] \Pr[\mathcal{F}_{t-1}]}{\Pr[x_t|\mathcal{F}_{t-1}] \times \Pr[\mathcal{F}_{t-1}]} \\ = \frac{\Pr[x_t|n_t = j, \mathcal{F}_{t-1}] \times \Pr[n_t = j|\mathcal{F}_{t-1}]}{\Pr[x_t|\mathcal{F}_{t-1}]} \quad (4.26)$$

$$= \frac{f(x_t|n_t = j, \mathcal{F}_{t-1}) \times \Pr[n_t = j|\mathcal{F}_{t-1}]}{f(x_t|\mathcal{F}_{t-1})}. \quad (4.27)$$

The first equation (4.25) indicates that the fundamental information available at time t is \mathcal{F}_t , which contains the new information on x_t and the one of the previous returns captured in \mathcal{F}_{t-1} . The next two equations use the formula of conditional probabilities. The step from (4.26) to (4.27) is just notation. From there on, it is possible to evaluate (4.24), since the various components of (4.27) are known. Indeed, conditional on knowing λ_t , σ_t , and the number of jumps that took place over a time interval, $n_t = j$, the density of x_t is

$$f(x_t|\mathcal{F}_{t-1}) = \sum_{j=0}^{\infty} f(x_t|n_t = j, \mathcal{F}_{t-1}) \times \Pr[n_t = j|\mathcal{F}_{t-1}], \quad (4.28)$$

with

$$f(x_t|n_t = j, \mathcal{F}_{t-1}) = \frac{1}{\sqrt{2\pi(\sigma_t^2 + j\delta^2)}} \exp\left(-\frac{(x_t - m_t)^2}{2(\sigma_t^2 + j\delta^2)}\right), \quad (4.29)$$

and $m_t = \mu - \theta\lambda_t + \theta j$. Equation (4.29) states that, if the number of jumps is known, the overall return will be normal. This follows from the fact that returns are the sum of the fundamental Gaussian part augmented by the Gaussian jumps. The second equation specifies the density of a return unconditionally on the number of jumps. The term $\Pr[n_t = j|\mathcal{F}_{t-1}]$ is given by (4.20). Hence, at this stage, all the components are computable.

Before proceeding, we need to provide conditions ensuring that all the dynamics are well defined. In an empirical illustration, we use a GARCH dynamics such as discussed earlier (equation (4.5)) for which positivity and stationarity conditions are easy to impose. Concerning (4.23), Chan and Maheu (2002), and Maheu and McCurdy (2004) notice that this equation may be rewritten as

$$\lambda_t = \lambda_0 + (\rho - \gamma)\lambda_{t-1} + \gamma E[n_{t-1} | \mathcal{F}_{t-1}].$$

This equation shows that a sufficient condition ensuring positivity of λ_t is that $\lambda_0 > 0$, $\rho \geq \gamma$, and $\gamma \geq 0$. The condition for stationarity is that $|\rho| < 1$. The expected level of jumps is then given by

$$E[\lambda_t] = \frac{\lambda_0}{1 - \rho}.$$

We may now discuss some of the restrictions to the model. The most important restriction is to impose constancy to the jump intensity parameter, $\lambda_t = \lambda$, for all t . In that case, it is possible to estimate the parameters with a direct Maximum Likelihood estimation. The algorithm for the estimation is similar to the one outlined for the ARCH model, see Section 4.3.3, with the difference being the density of the returns. Here the likelihood of one observation is given by (4.28). We notice that the log-likelihood will not take the usual simple expression but will consist of a sum. In theory, the sum involved in computing (4.28) is infinite. In practice, the probability of multiple jumps quickly becomes small. Hence, the infinite sum can be truncated by a small number of terms.¹⁴ This restriction corresponds to the model of Vlaar and Palm (1993). By assuming that the GARCH component is only an ARCH model, we obtain the model of Jorion (1989). Last, if we assume that the ARCH model reduces to a constant volatility, we obtain the model proposed by Press (1967).

Useful moments computed by Press (1967), see also Das and Sundaram (1997), adapted to a conditional setting, are

¹⁴ In practice, truncation to the first 10 or 20 terms of the sum appears to be largely sufficient to obtain stable estimates. Jorion (1989), using exchange rate data and asset returns, truncates after 10 terms. Maheu and McCurdy (2004) propose, for stock market data, truncation after 25 terms. In practice, it is a good idea to perform the estimation for a set of increasing terms 10, 20, 30, and to verify the stability of the estimations. It is also possible to inspect the probabilities of multiple jumps, intervening in the likelihood. This probability will be found to be very small in practical implementations.

$$\begin{aligned} E[x_t | \mathcal{F}_{t-1}] &= \mu, \\ V[x_t | \mathcal{F}_{t-1}] &= \sigma_t^2 + (\theta^2 + \delta^2)\lambda_t, \\ Sk[x_t | \mathcal{F}_{t-1}] &= \frac{\lambda_t(\theta^3 + 3\theta\delta^2)}{(\sigma_t^2 + \lambda_t\delta^2 + \lambda_t\theta^2)^{3/2}}, \\ Ku[x_t | \mathcal{F}_{t-1}] &= 3 + \frac{\lambda_t(\theta^4 + 6\theta^2\delta^2 + 3\delta^4)}{(\sigma_t^2 + \lambda_t\delta^2 + \lambda_t\theta^2)^2}, \end{aligned}$$

where $Sk[x_t | \mathcal{F}_{t-1}]$ and $Ku[x_t | \mathcal{F}_{t-1}]$ denote the conditional skewness and kurtosis, respectively. We notice that the sign of the jumps, θ , will determine the sign of the skewness.

A further useful result concerns the conditional probability of observing several jumps. This probability may be computed as

$$\Pr[n_t \geq 1 | \mathcal{F}_t] = 1 - \Pr[n_t = 0 | \mathcal{F}_t]. \quad (4.30)$$

This latter term is already computed during the estimation of the model.

4.5.2 An empirical illustration

To illustrate the working of the model with jumps, we consider the SP500 as well as the DAX daily returns. We estimate model (4.14)–(4.17) with a simple GARCH volatility given by

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2,$$

and the dynamics for the intensity parameter given by (4.23). We also estimate the parameters of the model when the volatility dynamics and the jump intensity constant over time, $\sigma_t = \sigma$, $\forall t$, and $\lambda_t = \lambda$, $\forall t$.

Table 4.4 displays the parameter estimates for the restricted and the general model. The jump size, θ , is always negative corroborating the stylized fact that returns are negatively skewed. We notice that θ is larger when the GARCH effect is introduced. For instance, for the SP500, the jump size evolves from $\theta = -0.024$, obtained for the model without GARCH nor time variation in the jump intensity, to $\theta = -0.503$ for the general model. Our interpretation is that without the GARCH dynamics, the jump component is forced to capture many of the variations due to the time-varying volatility and therefore does not necessarily correspond to actual jumps. Clearly, as the jump size becomes larger, the intensity of the jumps diminishes. We notice for instance for the SP500 that $\lambda = 0.755$ for the simple model and $E[\lambda_t] = 0.21$ for the general model. This means that, for the general model, a jump takes place about every fifth day.

When we focus on the dynamics of the jump intensity, we notice by inspecting the parameter ρ that there is important persistence in the λ parameter implying a rather smooth evolution of this parameter through time. If we consider the log-likelihoods of the restricted and general models, we notice

Table 4.4. Parameters for a GARCH model combined with a jump process with time-varying intensity and a restricted model without time-varying volatility nor time-varying jump intensity

	SP500	DAX	SP500	DAX
μ	0.035 (0.010)	0.035 (0.012)	0.036 (0.013)	0.032 (0.016)
θ	-0.503 (0.113)	-0.646 (0.093)	-0.024 (0.023)	-0.226 (0.069)
δ	1.099 (0.111)	1.191 (0.105)	1.009 (0.051)	1.981 (0.116)
λ	—	—	0.755 (0.098)	0.252 (0.035)
λ_0	0.015 (0.005)	0.016 (0.005)	—	—
ρ	0.930 (0.034)	0.944 (0.014)	—	—
γ	0.508 (0.111)	0.791 (0.160)	—	—
σ	—	—	0.480 (0.032)	0.855 (0.024)
w	0.007 (0.003)	0.009 (0.007)	—	—
α	0.028 (0.007)	0.029 (0.018)	—	—
β	0.950 (0.011)	0.950 (0.029)	—	—
log-lik.	-8363.0	-9603.9	-8761.7	-10351.1
$E[\lambda_t]$	0.2130	0.2920	—	—

a statistically significant change as we introduce jumps to a simple GARCH model and a further statistically significant change as we introduce time variation in the jump intensity.

It is also possible to estimate the temporal evolution of the probability that a jump took place over a given day, see (4.30). The evolution of the jump probability for the SP500 over the period starting with the beginning 1999 and ending with the year 2002 is provided in Figure 4.6.¹⁵ As we could expect from the relatively high parameter ρ and the ensuing smoothness of λ_t , the jump probability evolves relatively smoothly over time. Moreover, the period in the middle of 2002 related to the war in Iraq also corresponds to a period of increased jump probabilities. Last, we can consider the temporal evolution implied by this model for skewness and kurtosis. This is given in Figure 4.7.

¹⁵ We could have chosen a much larger window, however, the picture becomes very erratic if we consider too many observations.

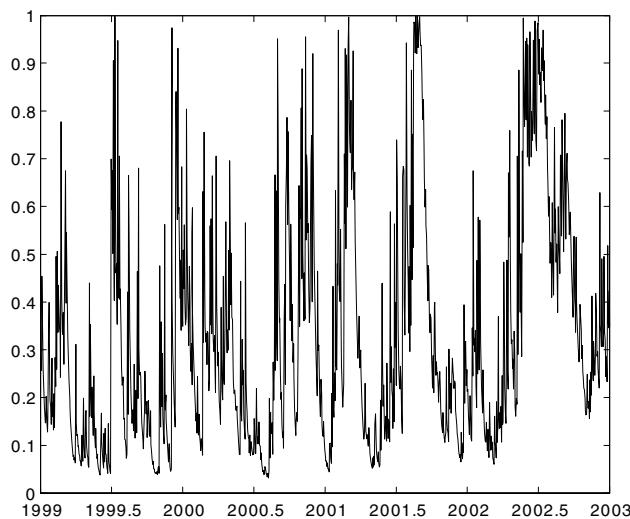


Fig. 4.6. Estimation of the jump probability $\Pr[n_t \geq 1 | I_t]$ for the SP500.

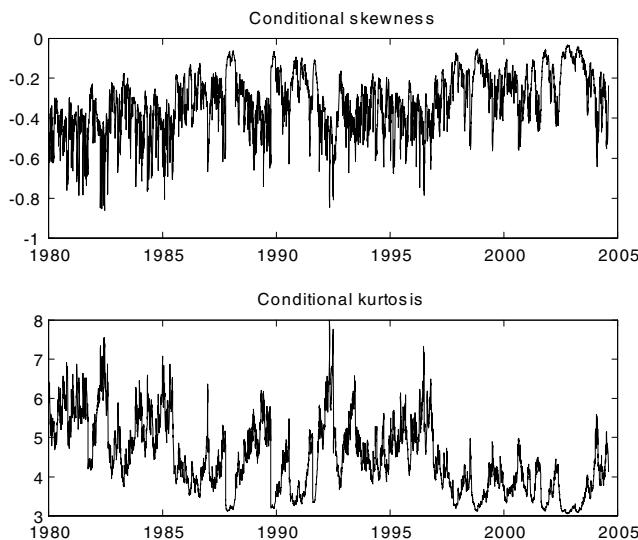


Fig. 4.7. Evolution of the conditional skewness and kurtosis for the SP500.

As these figures show, the higher moments of the data are found to evolve over time. For the October 1987 crash, we notice some limitations of this model. Indeed, the skewness parameter becomes very small, close to 0, and kurtosis approaches 3, the value a Gaussian distribution would take. This finding comes from the fact that the 1987 crash induced a huge upward jump in volatility. This increase in volatility, captured by the GARCH model, absorbs all the return variations that might otherwise be judged as a jump. This is related to an old debate of the empirical finance literature namely if we should include or not the 1987 crash in the model. Some pundits claim that any satisfactory model should be able to describe this outlier. Others are willing to discard those few observations surrounding the 1987 crash on the ground that the variation of that day cannot be encompassed in a general framework. With the model at hand, the λ_t parameter takes the value of 6, suggesting that on that day, 6 crashes took place. If this is a possibility remains an open question. We also notice that the range of variation of these two moments is rather large. The conditional skewness tends to be less negative over the recent period, while kurtosis tends to be closer to 3, suggesting that the distribution of the SP500 has been recently closer to the normal distribution. The time variation of such moments is important for portfolio allocation and will also be considered in Chapter 9.

4.6 Aggregation of GARCH processes

So far, we have seen that the description of the temporal evolution of return series, and in particular its time-varying volatility, may be relatively well captured by GARCH processes. At this stage, new questions arise. The first one concerns the evolution of the parameters of GARCH processes, as we use different time frequencies for the estimation. For instance, we may be tempted to use daily data to estimate a process that should, however, be scaled to some other frequency, say a monthly one. The answer is that the GARCH process, studied so far, does not aggregate, a result shown in an important paper by Drost and Nijman (1993). These authors introduce a new class of processes, that they name weak GARCH model, that aggregate, however.

The second question concerns the cross-section aggregation, i.e., what will be the resulting process for a portfolio of assets, if each asset has GARCH type returns? The answer is again disappointing. As shown by Nijman and Sentana (1996), a linear combination of assets, each with a GARCH dynamics, will no longer be distributed as a GARCH model in the usual sense.

In this section, we discuss the main results, relevant for the temporal as well as the cross-sectional aggregation of GARCH processes.

4.6.1 Temporal aggregation

Theoretical aspects

Drost and Nijman (1993) prove that the GARCH process, of the type considered so far, may not aggregate in a simple way.¹⁶ For this reason, they introduce different categories of GARCH processes. The GARCH models studied so far will belong to the class of strong GARCH processes. At the other extreme, Drost and Nijman (1993), aware of the temporal aggregation results of Nijman and Palm (1990), realized that projections of squared returns on ARMA processes would lead to aggregation results. For this reason, they introduced so-called weak GARCH processes that are linear projections. The following definitions give a more precise sense.

Definition 4.1. (*Strong GARCH, Drost and Nijman, 1993*) The series ε_t , with $t \in \mathbb{Z}$, and $\varepsilon_t = \sigma_t z_t$ where $\sigma_t^2 = w + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2$ is a strong GARCH process if for some parameters w , α , and β , the innovation z_t is iid with zero mean and unit variance.

Definition 4.2. (*Semi-strong GARCH*) The series ε_t , with $t \in \mathbb{Z}$, is a semi-strong GARCH process if w , α , and β may be chosen so that

$$\begin{aligned} E[\varepsilon_t | \varepsilon_{t-1}, \varepsilon_{t-2}, \dots] &= 0, \text{ and} \\ E[\varepsilon_t^2 | \varepsilon_{t-1}, \varepsilon_{t-2}, \dots] &= \sigma_t^2, \end{aligned}$$

with σ_t^2 defined as for the strong GARCH process.

Define now the linear projection operator $P[\cdot]$ as the best linear predictor of some random variable given the information structure. Formally, we define $P[x_t | \varepsilon_{t-1}, \varepsilon_{t-2}, \dots]$ the best linear predictor of x_t in terms of $(1, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-1}^2, \varepsilon_{t-2}^2, \dots)'$. Then the following definition may be introduced.

Definition 4.3. (*Weak GARCH*) The sequence ε_t , with $t \in \mathbb{Z}$, is a weak GARCH process if w , α , and β may be chosen so that

$$\begin{aligned} P[\varepsilon_t | \varepsilon_{t-1}, \varepsilon_{t-2}, \dots] &= 0, \\ P[\varepsilon_t^2 | \varepsilon_{t-1}, \varepsilon_{t-2}, \dots] &= \sigma_t^2, \end{aligned}$$

with σ_t^2 defined as for the strong GARCH process.

At this stage, Drost and Nijman (1993) assume that they observe a sample $\{\varepsilon_t\}_{t=1}^T$ of some asset returns. If P_t denotes the value of a given asset, then the continuously compounded return at time t is defined as $\varepsilon_t = \log(P_t/P_{t-1})$. This return is assumed to be sampled at, say, the daily frequency. The question is what parameters may be obtained if we had sampled at a different

¹⁶ Alternatively, we may say that the class of GARCH processes is not closed under aggregation where the notion of closedness is taken in the mathematical sense.

frequency such as weekly. For instance, we may consider returns over a lower frequency consisting of the instants $m, 2m, \dots, T$, where m is an integer and T is assumed to be a multiple of m . Drost and Nijman (1993) introduce the notation

$$\bar{\varepsilon}_{(m)t} \equiv \sum_{i=0}^{m-1} \varepsilon_{t-i} = \log \left(\frac{P_t}{P_{t-m}} \right), \quad \text{for } t = m, 2m, \dots, T.$$

Thus, $\bar{\varepsilon}_{(m)t}$ represents the return over a longer period than the initial returns series.¹⁷ The main result of their paper is given by the following proposition.

Proposition 4.4. (*Drost and Nijman, 1993*) *If the sequence ε_t , $t \in \mathbb{Z}$, is a weak*

GARCH(1, 1) process with symmetric marginal distributions, $\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2$, and unconditional kurtosis κ_ε , then $\bar{\varepsilon}_{(m)tm}$, $t \in \mathbb{Z}$, is a symmetric weak GARCH(1, 1) process with

$$\bar{\sigma}_{(m)tm}^2 = \bar{w}_{(m)} + \bar{\alpha}_{(m)} \bar{\varepsilon}_{(m)tm-m}^2 + \bar{\beta}_{(m)} \bar{\sigma}_{(m)tm-m}^2,$$

and kurtosis $\bar{\kappa}_{(m)\varepsilon}$ where

$$\begin{aligned} \bar{w}_{(m)} &= m\omega \frac{1 - (\alpha + \beta)^m}{1 - (\alpha + \beta)}, \\ \bar{\alpha}_{(m)} &= (\alpha + \beta)^m - \bar{\beta}_{(m)}, \\ \bar{\kappa}_{(m)\varepsilon} &= 3 + (\kappa_\varepsilon - 3)/m + 6(\kappa_\varepsilon - 1) \\ &\quad \times \frac{[m - 1 - m(\alpha + \beta) + (\alpha + \beta)^m] [\alpha - \alpha\beta(\alpha + \beta)]}{m^2(1 - \beta - \alpha)^2(1 - \beta^2 - 2\alpha\beta)}, \end{aligned}$$

and where $|\bar{\beta}_{(m)}| < 1$ is the solution to the quadratic equation

$$\frac{\bar{\beta}_{(m)}}{1 + \bar{\beta}_{(m)}^2} = \frac{A(\alpha + \beta)^m - B}{A\{1 + (\alpha + \beta)^{2m}\} - 2B},$$

where

$$\begin{aligned} A &= m(1 - \beta^2) + 2m(m - 1) \frac{(1 - \alpha - \beta)^2(1 - \beta^2 - 2\alpha\beta)}{(\kappa_\varepsilon - 1)[1 - (\alpha + \beta)^2]} \\ &\quad + 4 \frac{[m - 1 - m(\alpha + \beta) + (\alpha + \beta)^m] [\alpha - \alpha\beta(\alpha + \beta)]}{1 - (\alpha + \beta)^2}, \end{aligned}$$

and

$$B = \alpha - \alpha\beta(\alpha + \beta) \frac{1 - (\alpha + \beta)^{2m}}{1 - (\alpha + \beta)^2}.$$

¹⁷ Drost and Nijman (1993) introduce a distinction between flow and stock variables. For finance applications, involving asset returns, the aggregation of flows is of relevance.

Given that this proposition requires the kurtosis of the original error, κ_ε , for empirical applications the following reminder by Drost and Nijman (1993) may be useful

$$\kappa_\varepsilon = \kappa_z \frac{1 - (\alpha + \beta)^2}{1 - (\alpha + \beta)^2 - (\kappa_z - 1)\alpha^2},$$

where z_t denotes the standardized innovation. Thus, if we estimate the kurtosis of the innovations z_t , after having fitted the GARCH model, we may infer κ_ε once α and β have been estimated. Inspection of the formula for kurtosis reveals that as the aggregation step m increases, $\bar{\kappa}_{(m)\varepsilon}$ converges to 3. Thus, temporal aggregation will render the data more normal. We may intuitively view this result as resulting from the smoothing of observations as we sum the innovations. Heteroskedasticity will get smoothed. This result has been noticed already by Diebold (1988) and has been verified in Chapter 2.

An empirical illustration

To illustrate how temporal aggregation changes the parameters, we consider a situation where $(\alpha, \beta) = (0.1, 0.85)$ and $(\alpha, \beta) = (0.05, 0.94)$ as well as the cases where the kurtosis equals $\kappa_z = 3$, corresponding to the Gaussian case, and $\kappa_z = 8.75$, corresponding to the case where innovations are distributed as a Student t with 5 degrees of freedom. As in Drost and Nijman (1993), we consider various values of $m = 1, 2, 4, 5, 16, 32, 64$. The corresponding values of $(\bar{\beta}_{(m)}, \bar{\alpha}_{(m)})$ will be represented in Figure 4.8 by symbols, joined for graphical convenience by lines. The case $(\alpha, \beta) = (0.1, 0.85)$ corresponds to the upper curves represented by continuous lines. The case $(\alpha, \beta) = (0.05, 0.94)$ is represented by the dashed lower set of curves.

Among each curve, squares correspond to those pairs of the α and β parameters generated under a Gaussian scenario. Bullets are obtained under the assumption that the innovations follow a Student t with 5 degrees of freedom. We, thus, observe that the distribution and fat-tailedness of the innovations affects the properties of aggregation. Depending on the parameters α and β as well as on the kurtosis, various patterns how the parameters evolve as m increases are possible. Interestingly, therefore, the kurtosis of the innovation z_t affects the temporal aggregation properties of the parameters. In other words, in practical applications, if the kurtosis is not known, it is not possible to correctly infer how the parameters will aggregate.

We may also ask how aggregated parameters compare with those obtained for a GARCH model estimated on aggregated data. To answer this question, we estimate a symmetric GARCH(1,1) model to the SP500 returns sampled at the daily and weekly frequencies assuming normality of innovations. The sampling at weekly frequency corresponds to $m = 5$. The results are represented in Table 4.5. The table summarizes the results obtained. We notice that the kurtosis of innovations is very high in both cases. As we switch frequency, the parameters evolve in the right direction.

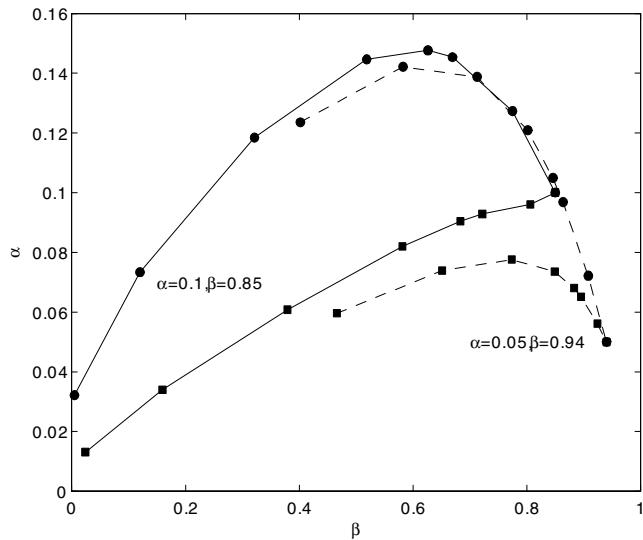


Fig. 4.8. Evolution of the parameters of a weak GARCH model under temporal aggregation. The symbols correspond (from right to left) to sampling every 1, 2, 4, 5, 8, 16, 32, 64 observations. Curves with squares correspond to Gaussian innovations and bullets to Student t distributed innovations

Table 4.5. GARCH parameters for the SP500 at both weekly and daily frequencies

Parameter	Estimate	Std error
Weekly frequency		
ω	0.2834	(0.0945)
α	0.1826	(0.0271)
β	0.7862	(0.0341)
Kurtosis, κ_z	8.21	—
Daily frequency		
w	0.0197	(0.0045)
α	0.0743	(0.0069)
β	0.9057	(0.0109)
Kurtosis, κ_z	8.71	—

To investigate how correct this switch is, we trace in Figure 4.9 the parameters $\bar{\alpha}_{(m)}$ and $\bar{\beta}_{(m)}$ obtained under temporal aggregation when we start with higher frequency data (here daily) $(\alpha, \beta) = (0.0743, 0.9057)$. The curve with the stars represents a curve corresponding to aggregated parameters for the same sequence of m as above. The parameters obtained for the GARCH model, estimated at weekly frequency, are represented by the single large diamond in the upper part of the figure. As we can see, there is a significant

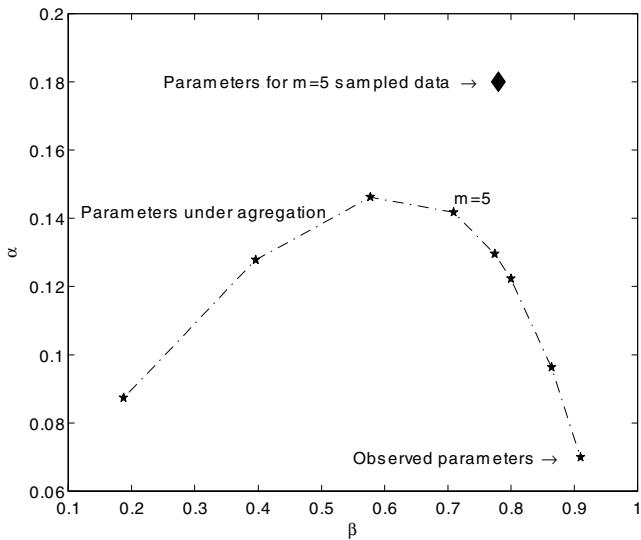


Fig. 4.9. Evolution of the parameters of a weak GARCH model under temporal aggregation.

distance between the actual parameters and those parameters that should hold under aggregation. This brings about the question of the correctness of the model.

In an interesting contribution, Drost and Werker (1996) consider a continuous time model with jumps that they discretize so as to obtain a weak GARCH representation. Estimation of this model allows these authors to provide some inference about the frequency of jumps in the data. If we contemplate the model used so far, we realize that several features that are present in the data have not been treated so far. We may mention the necessity that the fourth moment be finite (and as we will see in Chapter 7, this is not necessarily the case). The next issue is that the results are obtained for so-called weak GARCH models. These models are, however, not consistently estimated as demonstrated in a simulation exercise by Meddahi and Renault (2004). Also, these models do not allow for leverage effect. Meddahi and Renault (2004) circumvent these difficulties by building on a model in the spirit of stochastic volatility.

4.6.2 Cross-sectional aggregation

In this section, we consider the important issue of how GARCH processes aggregate across assets, an issue initially addressed by Nijman and Sentana

(1996). These authors show that the contemporaneous aggregation of independent univariate strong GARCH processes yields a weak GARCH process as defined in the previous section. This result may be seen as follows: Assume that $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$ are both generated by strong GARCH models such as

$$\varepsilon_{i,t} = \sigma_{i,t} z_{i,t}, \quad i = 1, 2,$$

where the $z_{i,t}$ are *iid* with mean 0, variance 1, and kurtosis κ_i , and that volatility is generated by the symmetric GARCH(1, 1) model

$$\sigma_{i,t}^2 = \omega_i + \alpha_i \varepsilon_{i,t-1}^2 + \beta_i \sigma_{i,t-1}^2.$$

Then Nijman and Sentana (1996) follow Bollerslev (1988) and introduce the forecast volatility error $\eta_{i,t} = \varepsilon_{i,t}^2 - \sigma_{i,t}^2 = \varepsilon_{i,t}^2 - E[\varepsilon_{i,t}^2 | \mathcal{F}_{t-1}]$, where \mathcal{F}_{t-1} represents the past information, to obtain an ARMA(1, 1) representation for $\varepsilon_{i,t}^2$, such as

$$\varepsilon_{i,t}^2 = \sigma_i^2 + [1 - (\alpha_i + \beta_i)L]^{-1}[1 - \beta_i L]\eta_{i,t},$$

where L is the lag operator, defined by $L\eta_{i,t} = \eta_{i,t-1}$, and where $\sigma_i^2 = E[\varepsilon_{i,t}^2] = \omega_i/(1 - \alpha_i - \beta_i)$. Taking the sum of the two processes yields

$$\begin{aligned} (\varepsilon_{1,t} + \varepsilon_{2,t})^2 &= \sigma_1^2 + \sigma_2^2 + [1 - (\alpha_1 + \beta_1)L]^{-1}[1 - \beta_1 L]\eta_{1,t} \\ &\quad + [1 - (\alpha_2 + \beta_2)L]^{-1}[1 - \beta_2 L]\eta_{2,t} + 2\varepsilon_{1,t}\varepsilon_{2,t}. \end{aligned}$$

Given that $\eta_{1,t}$, $\eta_{2,t}$ and the product $\varepsilon_{1,t}\varepsilon_{2,t}$ are uncorrelated, and all series are non autocorrelated, Nijman and Sentana (1996) conclude that the sum of two strong GARCH(1, 1) processes yields a weak GARCH(2, 2) process using the results from time series analysis according to which a sum of two ARMA(1, 1) processes yields an ARMA(2, 2) process. A remaining difficulty is of course to obtain a parsimonious independent representation of this process. The reason for this difficulty comes from the cross-product term $\varepsilon_{1,t}\varepsilon_{2,t}$. Furthermore, Nijman and Sentana (1996) show that, if $\alpha_1 + \beta_1 = \alpha_2 + \beta_2 = \gamma$, then the resulting process will be a weak GARCH(1, 1) process.

4.6.3 Estimation of the weak GARCH process

Clearly, the parametric specifications of a strong and a weak GARCH process are not the same. This raises as first question what would be the error if we estimated a strong GARCH process and we directly used the resulting parameters in a weak GARCH process. Drost and Nijman (1993) report that they performed simulations where the true data generating process was a strong GARCH process and where they estimate the parameters using a QML estimation. For parameters similar to the dynamic of exchange rates, they report that the results obtained in the QML estimation are close to the true parameters.

On the other hand, Nijman and Sentana (1996) construct weak GARCH processes as the sum of strong GARCH processes and by performing QML

estimation. They report a small, yet systematic bias in one of the parameters of the GARCH model.

These observations have led Francq and Zakoïan (2000) to investigate the properties of QML estimation of weak GARCH processes. For the setting of Drost and Nijman (1993), they confirm that the errors are small. For other settings and specifications, they show that the bias may be large. This in turn requires an estimation procedure for weak GARCH processes, because QML estimation appears to be troubled. The reason for this failure is that $\eta_{i,t}$ may not satisfy the properties of a martingale difference sequence, on which the QML theory in a time series context builds. Francq and Zakoïan (2000) further propose a two-stage least-squares estimation building on the observations, ε_t , and the weak GARCH representation yielding to an estimation such as

$$\varepsilon_t^2 + \sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2 = \omega + u_t + \sum_{i=1}^q \beta_i u_{t-i},$$

where u_t is now some white noise, with a constant variance.

To conclude this section, we need to mention that Meddahi and Renault (2004) propose an estimation procedure based on a state-space representation to avoid biasedness of the estimates.

4.7 Stochastic volatility

4.7.1 From GARCH models to stochastic volatility models

So far we have discussed many models in the spirit of GARCH models. Typically, for a simple model in this class, the centred returns may be written

$$\varepsilon_t = \sigma_t z_t, \tag{4.31}$$

$$\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2. \tag{4.32}$$

Inspection of these equations reveals that there is only one source of uncertainty, z_t , with $E[z_t] = 0$, and $V[z_t] = 1$. This way of modeling returns is therefore very simple, which may explain, at least partially, its tremendous success. The drawback is that this model does not allow for a specific error in the dynamics of volatility. For given parameters, if we imagine a large return, in absolute value, then the volatility equation (4.32) will imply a subsequent large volatility. In general, this is certainly a desirable feature. However, the impact of a given large return will not be the same all the time. This implies that the σ_t that this model predicts, in comparison with the true unknown volatility, will sometimes be too large and other times too small.

This observation has led Taylor (1982, 1986) to consider the first discrete time stochastic volatility (SV) model in which volatility has a specific source of randomness. This model may be expressed as follows

$$\varepsilon_t = \sigma_t z_t = z_t \exp\left(\frac{1}{2}h_t\right), \quad (4.33)$$

$$h_t = \omega + \beta h_{t-1} + v_t, \quad (4.34)$$

where v_t is an *iid* $\mathcal{N}(0, \sigma_v^2)$ process. The two processes z_t and v_t are usually assumed to be independent. Adding an error term v_t in the dynamics of volatility introduces another source of randomness in the model that may improve the description of the actual volatility. Obviously, the true volatility is never observed and, thus, h_t is a latent process that needs to get estimated. The probable improvement in terms of volatility comes at the price of higher complexity at the estimation level.

Since (4.34) represents an AR(1) process, stationarity conditions and moments are well-known. Indeed, stationarity of the h_t process imposes $|\beta| < 1$. Then, we have

$$E[h_t] = \mu_h = \frac{\omega}{1 - \beta}, \quad \text{and} \quad V[h_t] = \sigma_h^2 = \frac{\sigma_v^2}{1 - \beta^2}.$$

Therefore, since z_t is assumed to be *iid* $\mathcal{N}(0, 1)$ and uncorrelated with v_t , we can compute the moments of $\varepsilon_t = z_t \exp\left(\frac{1}{2}h_t\right)$ (see also Taylor 1986, Chapter 3),¹⁸

$$\begin{aligned} E[\varepsilon_t] &= E[z_t] E\left[\exp\left(\frac{1}{2}h_t\right)\right] = 0, \\ E[\varepsilon_t^2] &= E[z_t^2] E[\exp(h_t)] = \exp\left(\mu_h + \frac{\sigma_h^2}{2}\right), \\ E[\varepsilon_t^3] &= 0, \\ E[\varepsilon_t^4] &= E[z_t^4] E[\exp(2h_t)] = 3 \exp(2\mu_h + \sigma_h^2). \end{aligned}$$

It follows that all odd moments are zero, which results from the assumption of independence between z_t and v_t and the symmetry of z_t . Kurtosis is equal to

$$\frac{E[\varepsilon_t^4]}{(E[\varepsilon_t^2])^2} = \frac{3 \exp(2\mu_h + \sigma_h^2)}{\left(\exp\left(\mu_h + \frac{\sigma_h^2}{2}\right)\right)^2} = 3 \exp(\sigma_h^2) > 3.$$

This expression indicates that the kurtosis can become arbitrarily large.

In addition, taking the square and the logarithm of (4.33) and using (4.34) yields

$$\log(\varepsilon_t^2) = \log(\sigma_t^2) + \log(z_t^2) = h_t + \log(z_t^2).$$

Since the first component, h_t is an AR(1) process and $\log(z_t^2)$ is a white noise, the dynamics of the log of ε_t^2 will be an ARMA(1, 1) process. As Harvey, Ruiz, and Shephard (1994) indicate, extensions of the dynamics of h_t to an

¹⁸ We remind that if a random variable X is distributed as a $\mathcal{N}(\mu, \sigma^2)$, then we have the r th moment $E[(e^X)^r] = E[e^{rX}] = e^{(r\mu + r\sigma^2/2)}$.

ARIMA dynamics are straightforward. Some of the implications of changing the dynamics of h_t are discussed in that paper.

The SV model is much more flexible than the GARCH models. It has been found to fit asset returns better and have residuals closer to standard normal. In the SV (and GARCH) models, the distribution of returns has immediately fat tails. Persistence in volatility is captured by autoregressive terms β_j . Correlation between z_t and v_t produces volatility asymmetry. The SV model is therefore able to capture most statistical features of return volatility.

4.7.2 Estimation of the discrete time SV model

However, although the error term v_t makes the SV model much more flexible, it also implies that the SV model cannot be estimated directly by ML. Indeed, the process h_t is a latent variable. The likelihood function of the SV model is

$$f(\varepsilon_1, \dots, \varepsilon_T | \theta) = f(\varepsilon_T | \mathcal{F}_{T-1}) \times f(\varepsilon_{T-1} | \mathcal{F}_{T-2}) \times \dots \times f(\varepsilon_1 | \mathcal{F}_0),$$

where $\theta = (\omega, \beta, \sigma_v^2)'$ denotes the vector of unknown parameters. The difficulty is that the expression $f(\varepsilon_t | \mathcal{F}_{t-1})$ depends on an unobservable variable σ_t . To go further, we introduce the notation $\varepsilon_{\underline{T}} = \{\varepsilon_1, \dots, \varepsilon_T\}$. This allows us to express the joint likelihood of $\varepsilon_{\underline{T}}$ and $\sigma_{\underline{T}}$. Since the distribution of ε_t is known to be normal with mean 0 and variance σ_t^2 , conditionally on a given path of volatility, the joint density can be written as

$$f(\varepsilon_{\underline{T}}, \sigma_{\underline{T}}) = f(\varepsilon_{\underline{T}} | \sigma_{\underline{T}}) f(\sigma_{\underline{T}}),$$

where $f(\varepsilon_{\underline{T}} | \sigma_{\underline{T}}) = \prod_{t=1}^T f(\varepsilon_t | \sigma_t)$ and $f(\sigma_{\underline{T}})$ is some multivariate distribution. Therefore, to compute the likelihood $f(\varepsilon_{\underline{T}} | \theta)$, we may use the expression above, because the marginal is obtained from the joint density by integrating out the variable we want to get rid of, so that,

$$L_T(\theta | \varepsilon_{\underline{T}}) = f(\varepsilon_{\underline{T}} | \theta) = \int \int \dots \int f(\varepsilon_{\underline{T}} | \sigma_{\underline{T}}) f(\sigma_{\underline{T}} | \theta) d\sigma_1 d\sigma_2 \dots d\sigma_T.$$

Since this expression does not require that volatility is observable anymore, it would be possible in principle to evaluate and therefore maximize the likelihood function L_T . However, evaluating L_T using this expression would require the numerical integration of a T -dimensional integral, which is impossible in practice. Several alternative estimation procedures have therefore been proposed to simplify this problem.

A quasi-maximum likelihood estimation (QMLE) has been proposed by Harvey, Ruiz, and Shephard (1994). First, from Abramowitz and Stegun (1970, p. 943), it is known that $E[\log(z_t^2)] = -1.27$ and $V[\log(z_t^2)] = 4.93$. Introducing the notation $\xi_t \equiv \log(z_t^2) + 1.27$, which yields a mean zero innovation, and squaring the returns, we can rewrite the system (4.33)–(4.34) as

$$\begin{aligned}\log(\varepsilon_t^2) &= -1.27 + h_t + \xi_t, \\ h_t &= \omega + \beta h_{t-1} + \nu_t.\end{aligned}$$

It is possible to apply the Kalman filter to this system. Given that the Kalman filter assumes Gaussian innovations, which is clearly not the case for ξ_t , it is not possible to obtain an exact likelihood this way. Ruiz (1994) performs experiments that show that the estimates obtained via the Kalman filter have to be preferred over method-of-moments estimates.

A numerical difficulty is to deal with zero returns, i.e., when $\varepsilon_t = 0$ or cases where returns are close to zero. Breidt and Carriquiry (1996) suggest the use of the transformation

$$\log(\varepsilon_t^2) \approx \log(\varepsilon_t^2 + c \cdot \hat{\sigma}^2) - c \cdot \hat{\sigma}^2 / (\varepsilon_t^2 + c \cdot \hat{\sigma}^2),$$

where $\hat{\sigma}^2$ stands for the sample variance of the ε_t and c is a small number such as 0.02. The impact of those observations equal or close to zero will therefore be diminished. Ghysels, Harvey, and Renault (1996) suggest the use of some conditional volatility estimate instead of the constant $\hat{\sigma}^2$.

Alternative estimation procedures, based on simulation methods have been proposed. The first one is *indirect inference* (Gouriéroux, Monfort, and Renault, 1996). The second one is the *Markov Chain Monte Carlo* (Jacquier, Polson, and Rossi, 1994, see also Kim, Shephard, and Chib, 1998, or Eraker, Johannes, and Polson, 2003). Both methods are very computationally intensive even in a univariate setting. Other estimation methods are based on *simulated method of moments* (Duffie and Singleton, 1993), *efficient method of moments* (Gallant and Tauchen, 1996), or *simulated maximum likelihood* (Brandt and Santa-Clara, 2002, Durham and Gallant, 2002). Recently, there has been an interest in *empirical characteristic function methods* (Singleton, 2001; Chacko and Viceira, 2003, Jiang and Knight, 2002, Rockinger and Semenova, 2005). None of these methods is truly simple to implement. Andersen, Benzoni, and Lund (2002) compare some of these estimation techniques. Multivariate extensions appear to be very difficult to estimate. The stochastic volatility model is therefore a very important building block for empirical finance. However, because of the difficulty of estimation, it may not have attracted as much attention as it deserves.

4.8 Realized volatility

Over the recent years, it has become possible to use data at a tick-by-tick level. Such data is also called *high-frequency data*. In parallel with the availability of this data, a new research strand emerged. This strand builds on results of stochastic integration.¹⁹ The following is mainly based on Barndorff-Nielsen and Shephard (2004b, thereafter BN/S).

¹⁹ A set of worthwhile readings in the stochastic integration area is Dellacherie and Meyer (1982), Jacod and Shiryaev (2002), Protter (2004), Rogers and Williams

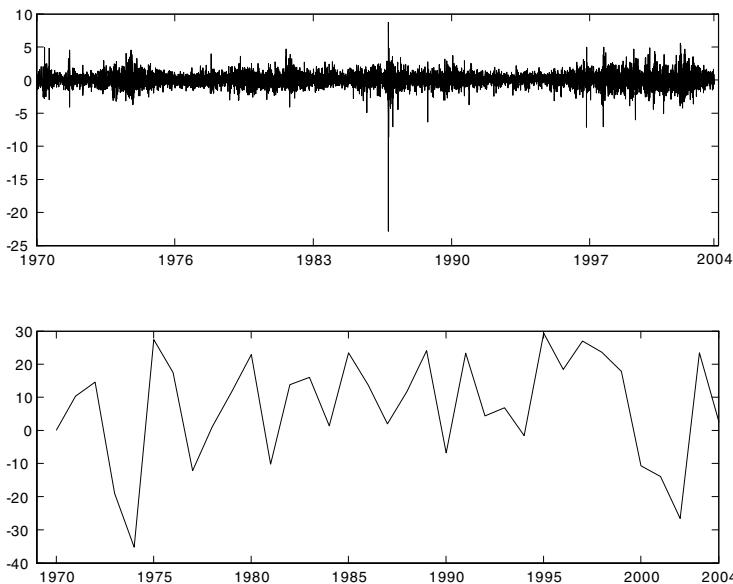


Fig. 4.10. Evolution of SP500 returns at daily and annual frequencies.

4.8.1 The difficulty to disentangle jumps

So far, probability theory tells us that in the continuous time setting, a reasonable model for asset prices is the semi-martingale. In actual empirical work, we observe, however, only discretely sampled data. This issue brings up the problem of distinguishing between the stochastic, yet continuous, part and the jumps. A recent contribution by Aït-Sahalia (2004) shows that, intuitively, data sampled at a higher and higher frequency should indeed provide information on the actual components. There are several difficulties with discrete sampling that render the detection of jumps difficult.

Figure 4.10 displays realizations of the SP500 at different frequencies.²⁰ As this figure illustrates, the shift from the daily to a yearly frequency renders the detection of the jump impossible. The crash of 1987 disappears totally at the yearly frequency. To the opposite, years without a crash, such as 2001 and 2002, saw a much larger drop in returns than 1987.

We may, therefore, reasonably ask from what timescale on does it become possible to visualize jumps in the data. Again, Aït-Sahalia (2004) provides some elements. He considers a simple model without temporal dependency of volatility in continuous time composed by a continuous part, provided by a

(2000), or the excellent recent book by Cont and Tankov (2004). Some of the following definitions may also be found in Andersen et al. (2003).

²⁰ Aït-Sahalia (2004) presents a similar plot with the Dow Jones Industrial Index for which the 1987 crash is even more pronounced.

Brownian motion, and discrete finite large jumps. Letting P_t denote the price at time t , the model that he considers assumes that the return over a given time interval $[t, t + \Delta]$ behaves as follows

$$r_{t,t+\Delta} = 100 \times \log(P_{t+\Delta}/P_t) = \mu\Delta + \sigma\sqrt{\Delta}\varepsilon_t + \sum_{s=t}^{t+\Delta} J_s N_s,$$

where ε_t is a normal $\mathcal{N}(0, 1)$ realization, N_t is a Poisson process, increasing by 1 whenever a jump takes place. When a jump takes place, it does so over a time interval of length Δ with an intensity of $\lambda\Delta$. The distribution of the jump size J is $\mathcal{N}(\beta, \eta^2)$. All sources of uncertainty, ε_t , N_t , and J_t , are assumed independent. The probability of having n jumps between t and $t + \Delta$ is

$$\Pr[N_{t+\Delta} - N_t = n] = \exp(-\lambda\Delta) \frac{(\lambda\Delta)^n}{n!}.$$

In this model, returns have, therefore, a simple continuous part, provided by the discretized Brownian motion, $\sigma\sqrt{\Delta}\varepsilon_t$, and the jump part. The question is then what the probability of observing a jump is, given that the return $r_{t,t+\Delta}$ exceeds a certain threshold, say r . Using Bayes' rule, we obtain that the probability that there is exactly one jump during the considered time interval is

$$\Pr[N_{t+\Delta} - N_t = 1 | r_{t,t+\Delta} > r] = \frac{e^{-\lambda\Delta} \lambda\Delta \left(1 - \Phi\left(\frac{r - \mu\Delta - \beta}{2(\eta^2 + \Delta\sigma^2)^{1/2}}\right)\right)}{\Pr[r_{t,t+\Delta} > r]},$$

where

$$\Pr[r_{t,t+\Delta} > r] = \sum_{n=0}^{\infty} e^{-\lambda\Delta} \frac{(\lambda\Delta)^n}{n!} \left(1 - \Phi\left(\frac{r - \mu\Delta - \beta}{2(n\eta^2 + \Delta\sigma^2)^{1/2}}\right)\right).$$

The probability of seeing more than one jump is

$$\Pr[N_{t+\Delta} - N_t \geq 1 | r_{t,t+\Delta} > r] = \frac{\sum_{n=1}^{\infty} e^{-\lambda\Delta} \frac{(\lambda\Delta)^n}{n!} \left(1 - \Phi\left(\frac{r - \mu\Delta - \beta}{2(n\eta^2 + \Delta\sigma^2)^{1/2}}\right)\right)}{\Pr[r_{t,t+\Delta} > r]}.$$

and obviously, the probability that no jump occurs is the complement of this probability. Using similar computations, we may show that the probability that two jumps occur in the time interval $[t, t + \Delta]$ is

$$\Pr[N_{t+\Delta} - N_t = 2 | r_{t,t+\Delta} > r] = \frac{e^{-\lambda\Delta} \frac{(\lambda\Delta)^2}{2!} \left(1 - \Phi\left(\frac{r - \mu\Delta - \beta}{2(2\eta^2 + \Delta\sigma^2)^{1/2}}\right)\right)}{\Pr[r_{t,t+\Delta} > r]}.$$

Aït-Sahalia calibrates the model with the values $\mu = \beta = 0$, $\sigma = 0.3$, $\lambda = 0.2$, $\eta = 0.6$, $\Delta = 1/12$. These parameters correspond to data that is calibrated annually, but for which we consider one single month as a time frame. In

order to compare returns, it is useful to standardize them. Indeed, a return that appears as a crash for a series with a small volatility may appear to be a completely normal return for a series with higher volatility. To render returns comparable, it is useful to standardize with respect to the unconditional volatility of the series. To do so, we consider u such that the threshold is given by $r = u(\Delta(\sigma^2 + (\beta^2 + \eta^2)\lambda))^{1/2}$. This parameter u will then have the units of standard deviation. Thus a return with $u = 1$ will have a size that corresponds to one unconditional standard deviation of the series.

With this choice of parameter values and scaling of the possible threshold values, Aït-Sahalia obtains the probabilities of jumps as represented in Figure 4.11, where the continuous line represents the probability of no jump, the dotted line the probability of exactly one jump, and eventually the line with dots and dashes measures the probability of having two jumps in a given month provided that we observe one return exceeding the threshold r , this corresponding to a certain standard deviation level presented along the horizontal axis.

As this figure displays, if we have a threshold of about one standard deviation, the observation of a return higher than this threshold is not very much informative about it being a jump. This probability is only about $1 - 0.98 = 2\%$. It takes relatively high values before we are able to state that a jump really occurs. Still, if we observe a return exceeding a three standard deviation threshold, the probability of this being explained by the continuous part of the process is 80%. Only at a level of about 3.5, it becomes impossible to distinguish if the return is due to a jump or to the continuous part. We may therefore conclude that the detection of jumps is a difficult issue because returns with more than 3.5 standard deviations are not very likely to take place.

The theoretical value of using data sampled at higher frequencies is demonstrated by Aït-Sahalia in Figure 4.12. This figure traces the probability that a jump occurred if we observe a return of 10% over a given time horizon. We notice that the probability of detecting a jump is very low when considering a weekly frequency. In other words, if we believe that the parameter values correspond to actual market activity, it will be nearly impossible to state if a crash happened just by eyeballing weekly returns. When considering shorter and shorter time horizons, the probability to detect jumps gets higher and higher as intuition suggests. At the daily frequency, we would think that, with 30% probability, a jump took place. It is, however, only when considering data sampled at the hourly frequency that we can actually detect a jump.

With these caveats that justify the use of high-frequency data, we now turn to tools that allow for a measurement of jumps as the data gets sampled at higher and higher frequencies. The discussion of the deep nature of jumps is not trivial. The literature on Lévy process shows that jumps can also come in the variety of *infinite activity*. This class of jumps does not appear to be detectable with the tools that follow. It appears that further research needs to be done to uncover techniques to measure such processes.

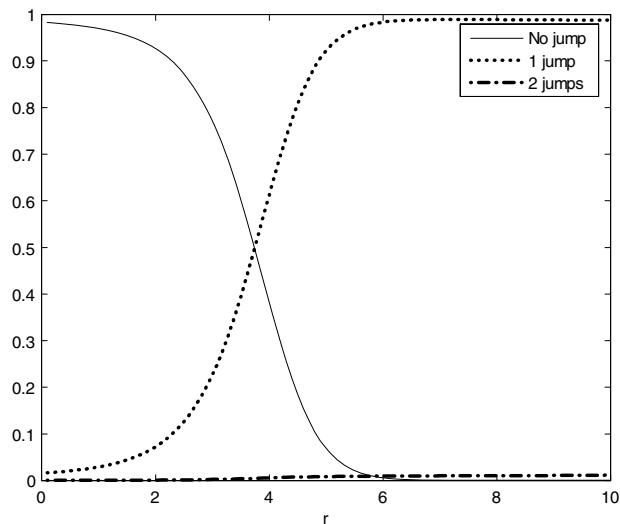


Fig. 4.11. Probability of jump, depending on the threshold r .

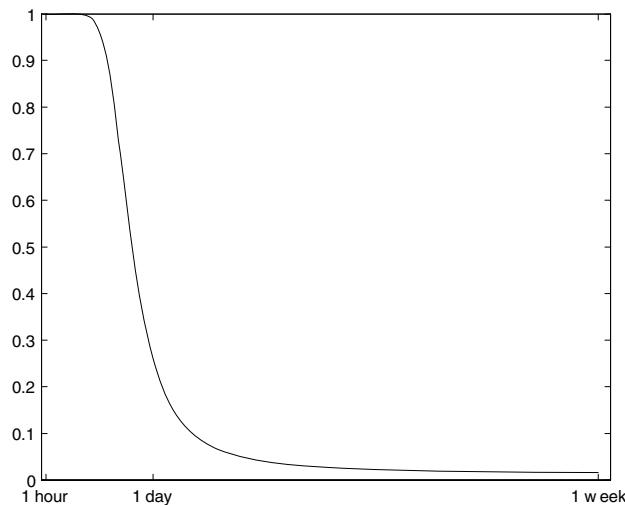


Fig. 4.12. Probability that a jump occurred if we observe a return of 10%, depending on the time horizon.

4.8.2 Quadratic variation

Consider a time horizon $[0, t]$. For this section, we will consider a partition of this interval with constant time increment, i.e., $\Delta = t/N$.²¹ Many of the results reported below have been generalized to a non-constant-step partition. In practice, with our interpretation of constant time-step, it means that data should be sampled, say, every 5, 10, or 30 minutes. Those results obtained for a non-constant time increment would hold for the raw data, that is with their original time stamp.

Let p_t denote the log-price of a given asset. The concept of quadratic variation is specified in the following definition.

Definition 4.5. *Quadratic variation of a process p_t is given as*

$$[p]_t^{[2]} \equiv \lim_{\Delta \rightarrow 0} \sum_{j=1}^{\lfloor t/\Delta \rfloor} (p_{j\Delta} - p_{(j-1)\Delta})^2,$$

where $\lfloor x \rfloor$ denotes the function returning the largest integer less than or equal to x .

The resulting process is sometimes called the *bracket process*. This expression defines the quadratic variation process as the limit case of a cumulative sum of squared returns when the sampling sequence becomes infinitely fine. The question that emerges is, what this process actually represents, to which entity does this quantity converge? A key result of probability theory states that the quadratic variation captures the quadratic variation of the local continuous martingale, as well as the measure of squared jumps that are of finite size. We may therefore write

$$[p]_t^{[2]} = [p^C]_t^{[2]} + \sum_{s \leq t} (\Delta p_s)^2,$$

where by definition $\Delta p_s = \lim_{u \rightarrow s, u < s} p_s - p_u$, and where $[p^C]_t^{[2]}$ captures the quadratic variation of the continuous part of the process. Thus, the notation above $\sum_{s \leq t} (\Delta p_s)^2$ designates the squared sum of jumps that occur in the price process. Clearly, even if the quadratic variation was known, i.e., by estimating it from the actual data, it still would not allow for disentangling the continuous and the jump parts.²²

²¹ Actually, in most of the following, we will consider limit cases where Δ will converge to zero. This means that the interval $[0, t]$ will be divided into some N intervals, with N becoming very large. We assume that the difference between $N \cdot \Delta$ and t is negligible.

²² If we use explicit models such as GARCH models combined with jump components, then it is possible to disentangle the continuous and jump parts. The issue when doing this is that strong assumptions on the stochastic volatility model and the type of jumps are made.

4.8.3 Power variation

At this stage, we have determined that the quadratic variation of the process contains the quadratic variation of the continuous part plus the measure of the jumps in the price process. A similar result can be obtained by using the concept of power variation, to which we turn now.

Definition 4.6. *The r th power variation is defined as*

$$[p]_t^{[r]} \equiv \underset{\Delta \rightarrow 0}{plim} \Delta^{1-r/2} \sum_{j=1}^{\lfloor t/\Delta \rfloor} |p_{j\Delta} - p_{(j-1)\Delta}|^r.$$

BN/S remark that the scaling factor Δ is crucial to get results that are comparable across series. The parameter r is called the power with which the variation is measured. For $r = 2$, power variation is the same as quadratic variation. Obviously, in that case, the scaling factor disappears. For $r \in (0, 2)$ the scaling factor converges to 0. For $r > 2$ it explodes as Δ converges to 0.

There remains the question what the power variation captures. To obtain further results, BN/S focus in their study of the power variation on semi-martingales with stochastic volatility. Before describing this class, let us define what semi-martingales are.

In the following, we consider, in general, time as belonging to the set $[0, \infty]$. It is assumed that there exists a filtration, $\{\mathcal{F}_t, t \geq 0\}$, describing the information available through time.

Definition 4.7. *A process M_t is a local martingale if M_0 is \mathcal{F}_0 measurable and if there exists an increasing sequence of stopping times τ_n with $\lim_{n \rightarrow \infty} \tau_n = \infty$ such that each process*

$$\{M_{\tau_i \wedge t} - M_0 = t \geq 0\}$$

is a martingale.²³

This definition states that a process is a local martingale if there exists a sequence of moments, such that for each moment, the best forecast of the future values of the process is the current value. For instance, a Brownian motion is a martingale and a local martingale. To see this, it is sufficient to take $\tau_i = i$, for $i \in \mathbb{N}$. If the time horizon is finite, $[0, T]$, then it is relatively safe to say that most local martingales encountered in finance will also be martingales.²⁴

²³ The symbols \vee and \wedge stand for the supremum and the infimum operators.

²⁴ Steve Shreve kindly communicated to us a discrete-time example of a process that is a local martingale but not a martingale. Inspection of this example shows that counter-examples are not easy to construct.

Definition 4.8. Let X_t , $t \geq 0$ be a stochastic process. X_t is a semi-martingale if it can be written as

$$X_t = X_0 + M_t + A_t,$$

where X_0 is known at time $t = 0$, where M_t is a local martingale, and where A_t is a process of finite variation, meaning that for any interval of time $[0, t]$, and for whatever partition $t_0 = 0 < t_1 < \dots < t_N = t$, whatever N ,

$$\sup \sum_{j=0}^N |A_{t_j} - A_{t_{j-1}}| < \infty.$$

By definition, we set $M_0 = A_0 = 0$.

As shown in Andersen et al. (2003), there is a no-arbitrage condition that the A_t and M_t components must satisfy. This condition states that if there exists a predictable jump in the asset price, there must be an offsetting variation in the martingale part. In our setting, the previous decomposition is the most useful to uncover, using recent convergence results, the continuous part and the jump part of a semi-martingale. A martingale is a stochastic volatility process if it can be written as

$$m_t = \int_0^t \sigma_u dW_u, \quad (4.35)$$

where σ_t is called the instantaneous volatility process and where W_t is a Brownian motion. It is assumed that σ_t is a càd-làg process, bounded away from zero. Since we can write, over a short interval of time, that $dm_t = \sigma_t dW_t$, the instantaneous variability is $(dm_t)^2 = \sigma_t^2 dt$. Over a longer time period, the variation will be defined as an integral

$$\sigma_t^{2*} = \int_0^t \sigma_u^2 du,$$

assumed to be finite, and which will be called the *integrated variance process*. Clearly, the integrated variance process measures the cumulative activity of a process where for the standard Brownian motion $\sigma_t^{2*} = t$. If the process σ_t is known, and this would be the case in a simulation exercise, then the integrated variance process may be computed with a numerical integration. A stochastic volatility process combined with a predictable component, a_t , will yield the $\mathcal{S}\mathcal{V}\mathcal{S}\mathcal{M}$ (for Stochastic Volatility Semi-Martingale) class. Processes in the class with continuous paths will be denoted by $\mathcal{S}\mathcal{V}\mathcal{S}\mathcal{M}^C$.

BN/S prove the following property: If p belongs to the class of $\mathcal{S}\mathcal{V}\mathcal{S}\mathcal{M}^C$ and (σ, a) are independent of W , then

$$[p]_t^{[r]} = \mu_r \int_0^t \sigma_s^r ds,$$

where, in this expression, $\mu_r = 2^{r/2} \Gamma(\frac{1}{2}(r+1))/\Gamma(\frac{1}{2})$. Notice that μ_r corresponds to the r th moment of a zero mean, unit variance normal process, i.e., $\mu_r = E[|u|^r]$, if $u \sim \mathcal{N}(0, 1)$. The following result demonstrates the importance of power variation. It shows that jumps will vanish if we compute power variations.

Proposition 4.9. *If $p = p^1 + p^2$ where $p^1 \in \mathcal{SVSM}^C$, i.e., $p_t^1 = a_t + m_t$, with a_t the predictable part and m_t the martingale part as in (4.35), and p^2 is a process with a finite number of jumps over finite time increments, where the processes (a_t, σ_t) are independent of W_t , as well as p^1 and p^2 are independent, then*

$$[p]_t^{[r]} = \mu_r \int_0^t \sigma_s^r ds,$$

as long as $r \in (0, 2)$.

This result means that if a process has a \mathcal{SVSM}^C component and discrete jump components, the power variation will eliminate the jump parts and only capture the continuous infinite variation variance. Intuitively, this result stems from the fact that the increment $|p_{j\Delta} - p_{(j-1)\Delta}|^r$ measures increments of the type $|p_{j\Delta}^1 - p_{(j-1)\Delta}^1|^r$ and of the type $|J|^r$ where J stands for jump. The scaling by $\Delta^{1-r/2}$ yields for the p^1 part, convergence to a well behaved integral, but for the jump part, convergence to 0.

If we consider, from here on, a small time increment $[t, t + dt]$ then, over this small time interval, for $r < 2$, it must hold that

$$[p]_{t+dt}^{[r]} - [p]_t^{[r]} = \mu_r \sigma_t^r dt \implies \sigma_t = \left[\frac{[p]_{t+dt}^{[r]} - [p]_t^{[r]}}{\mu_r dt} \right]^{1/r}.$$

These computations show that in theory at least the instantaneous volatility may be easily estimated. For practical applications, where this expression would have to be estimated from data sampled over discrete time intervals and subject to microstructure noise, the usefulness of such a computation requires further investigations.

4.8.4 Bipower variation

BN/S extend this notion of power variation by focusing on the covariance properties of adjacent increments. This yields to the notion of bipower processes. For further extensions to multipower variation, we let the reader refer to their papers. An early investigation, from a probability theoretical point of view, was led by Lepingle (1976).

Definition 4.10. *The bipower variation is defined by*

$$\{p\}_t^{[r,s]} \equiv \underset{\Delta \rightarrow 0}{plim} \Delta^{1-\frac{r+s}{2}} \sum_{j=1}^{\lfloor t/\Delta \rfloor - 1} |p_j \Delta - p_{(j-1)\Delta}|^r \cdot |p_{(j+1)\Delta} - p_{j\Delta}|^s, \quad r, s \geq 0.$$

We may trivially notice that, if $r = 0$ or $s = 0$, the bipower variation reduces to power variation. The notion of bipower variation, therefore, measures the behavior of powers of adjacent returns. If returns are in general driven by some continuous martingale component, given that this component will figure in both returns, the product will certainly capture it. In addition, following the logic of semi-martingales, there may be an occasional jump. Since the probability to have jumps in two adjacent returns tends to zero, this implies that the contribution of jumps to the bipower variation will be down-weighted. We might, therefore, intuitively expect that the bipower variation converges to an object where the jump component has been eliminated. The following key property shown by BN/S corroborates this intuition.

Proposition 4.11. *If $p = p^1 + p^2$ where $p^1 \in SVSM^C$, i.e., $p_t^1 = a_t + m_t$, with a_t the predictable part and m_t the martingale part as in (4.35), and p^2 is a process with finite jumps over finite time increments, where the processes (a_t, σ_t) are independent of W_t , as well as p^1 and p^2 are independent then, for $r + s \leq 2$, the following holds*

$$\{p\}_t^{[r,s]} = \mu_r \mu_s \int_0^t \sigma_u^{r+s} du,$$

with μ_r defined as indicated above. For $r + s > 2$, the quantity diverges.

This result is interesting in that by setting $s = r - 2$, the expression reduces to

$$\{p\}_t^{[r,2-r]} = \mu_r \mu_{2-r} \int_0^t \sigma_u^2 du.$$

This result shows that the bipower variation allows recovery of the integrated variance process, furthermore, it shows that by using the bipower variation, it is possible to get rid of the jumps. Combining these various results, we notice that the difference between power variation and bipower variation may be used to capture jumps

$$[p]_t^{[2]} - \frac{\{p\}_t^{[r,2-r]}}{\mu_r \mu_{2-r}} = \sum_{s \leq t} (\Delta p_s)^2.$$

So far, the results hold for a time period $[0, t]$. For empirical purposes, it may be interesting to consider these measures for finite time intervals. For instance, in empirical research, we may wish to compute the importance of jumps and of the continuous part, for a given day. Adapting the various concepts, introduced above, to finite time intervals is relatively easy. Rather than starting at 0, we may start at discrete points of time. Over finite time intervals, we may talk of

increments of quadratic variation, increments of bipower variation, etc. Also, rather than taking the theoretical limit cases of continuous time, by taking a discrete number of observations, it becomes possible to estimate the various ingredients by using actual data. Considering daily increments of the various variation measures, it is customary to call the estimate of the integrated daily quadratic variation *realized variance* and its square root *realized volatility*.

4.8.5 Estimation over finite time intervals

Some actual data

In this section, we indicate how to translate these various concepts into estimable entities. We also illustrate the type of results we may expect. To do so, we use a database consisting of high-frequency data of the French stock market. We focus on three companies, AXA, an insurance company, LVMH, a luxury fashion producer, as well as Société Générale, a large bank. The three companies under consideration stem from various sectors. For each of these companies, we extract data covering the second half of 2003 and the entire year 2004. The raw data is relatively large, for instance, on a typical day, there are about 12,000 transactions for AXA. From there on, we interpolated a price for predetermined 5-minute intervals.²⁵ We did this after dropping the opening price as well as those trades occurring after the official daily close, 8 hours and 30 minutes later. In our investigation, we do not work with returns from close to opening. Table 4.6 provides elementary statistics for these companies and for various frequencies of the data.

We find that the maximum return tends to be larger than the absolute value of the minimum return. This translates into the fact that during that period, the market was rather bullish. Contemplating the values of the standard deviation after converting them to a daily format, we obtain for AXA, $0.20 \times \sqrt{12} \times 8.5 = 2.01$ as a daily standard deviation using 5-minute returns, $0.66 \times \sqrt{8.5} = 1.92$ as daily standard deviation starting with hourly data, 1.73 using half-daily data, and eventually 1.67 using daily returns. These figures suggest that if we use high-frequency returns to compute daily standard deviations, we are led to an over estimation. This observation may be explained by possible autocorrelation in the data or by the presence of jumps in the data. The data appears in general positively skewed. The measures of kurtosis are larger than 3, suggesting that the data is generated by a process that is not distributed as a normal.²⁶

To examine the temporal dependency in the data, we consider first-order autocorrelation of returns, taking care not to use returns from a previous

²⁵ The companies under study contained such a huge number of trades that if we had used the raw price rather than the interpolated price, the difference would have been negligible.

²⁶ Diebold et al. (1998) notice similar problems to obtain low-frequency volatility estimates, such as monthly ones, from daily estimates of volatility.

Table 4.6. Various statistics for high-frequency data

	AXA	LVMH	Soc. Gén.
Minimum			
5 minutes	-2.63	-1.41	-1.70
1 hour	-4.39	-2.87	-3.28
4 hours	-7.50	-5.65	-4.37
1 day	-5.21	-4.60	-5.47
Maximum			
5 minutes	3.23	1.58	1.78
1 hour	11.60	6.35	5.33
4 hours	12.55	6.48	6.45
1 day	13.96	7.52	8.82
Standard deviation			
5 minutes	0.20	0.16	0.17
1 hour	0.66	0.50	0.53
4 hours	1.23	0.97	0.97
1 day	1.67	1.31	1.35
Skewness			
5 minutes	0.12	-0.04	0.03
1 hour	1.33	0.78	0.37
4 hours	0.80	0.43	0.32
1 day	1.20	0.80	0.58
Kurtosis			
5 minutes	13.59	8.16	9.08
1 hour	33.10	13.35	9.82
4 hours	16.74	8.14	6.26
1 day	13.16	6.82	7.38
Autocorrelation			
5 minutes	-0.095	-0.129	-0.109
1 hour	0.013	0.062	0.057
4 hours	-0.018	0.080	0.059
1 day	-0.097	0.058	-0.085

day. From the table, we notice that, at the 5-minute frequency, the first-order autocorrelation is negative. This negative sign of the autocorrelation may be explained by the so-called bid-ask bounce. As trades occur, sometimes they take place at the bid and sometimes at the ask. Even if the value of a company does not change, given that the prices at which the trades occur oscillate, this translates into a negative first-order autocorrelation. As the time frequency changes, we find changing signs of the autocorrelations. Already at the hourly level, all the correlations are positive. This observation raises the issue of the optimal sampling frequency, which we will discuss in the last section of this chapter.

Realized measures

In order to implement estimation of the concepts seen above, such as the realized variance, we need to provide some notations. Assume that each day has been cut into M intervals that we suppose to be of constant step size, so that the instants where prices are measured are given by $j\delta$ where δ is the step size and $j = 0, \dots, M$. For instance, δ could correspond to 5 minutes and $M = 102$ for a day with 8 hours and 30 minutes. We index by i the i th day.

The following expression defines the day i realized variance

$$\widehat{[p]}_{i|M}^{[2]} = \sum_{j=1}^M (p_{(i-1)M\delta+\delta j} - p_{(i-1)M\delta+\delta(j-1)})^2. \quad (4.36)$$

Notice that the day figures as an index to the brackets and so does the sampling frequency. The realized volatility of day i is obtained as the square root of the realized variance:

$$\sqrt{\widehat{[p]}_{i|M}^{[2]}}. \quad (4.37)$$

The discrete measures for the realized (r th) power variation and the realized bipower variation are respectively

$$\widehat{[p]}_{i|M}^{[r]} = \delta^{1-\frac{r}{2}} \sum_{j=1}^M |p_{(i-1)M\delta+\delta j} - p_{(i-1)M\delta+\delta(j-1)}|^r, \quad (4.38)$$

$$\begin{aligned} \widehat{\{p\}}_{i|M}^{[r,s]} &= \delta^{1-\frac{r+s}{2}} \sum_{j=1}^{M-1} |p_{(i-1)M\delta+\delta j} - p_{(i-1)M\delta+\delta(j-1)}|^r \\ &\quad \times |p_{(i-1)M\delta+(\delta+1)j} - p_{(i-1)M\delta+\delta j}|^s. \end{aligned} \quad (4.39)$$

Clearly, in terms of convergence, associating a global time index t to the various days, we obtain

$$\widehat{[p]}_{i|M}^{[2]} \xrightarrow[M \rightarrow \infty]{} [p]_t^{[2]} - [p]_{t-1}^{[2]} = \int_{t-1}^t \sigma_u^2 du + \sum_{t-1 \leq s < t} (\Delta p_s)^2,$$

where $t \in (iM\delta, iM\delta + \delta)$ and $t-1 \in ((i-1)M\delta, (i-1)M\delta + \delta)$, and where $\sum_{t-1 \leq s < t} (\Delta p_s)^2$ stands for the jumps that occurred between $t-1$ and t . Also, for $r+s < 2$, we have

$$\begin{aligned} \widehat{[p]}_{i|M}^{[r]} &\xrightarrow[M \rightarrow \infty]{} [p]_t^{[r]} - [p]_{t-1}^{[r]} = \mu_r \int_{t-1}^t \sigma_u^r du, \\ \widehat{\{p\}}_{i|M}^{[r,s]} &\xrightarrow[M \rightarrow \infty]{} \{p\}_t^{[r,s]} - \{p\}_{t-1}^{[r,s]} = \mu_r \mu_s \int_{t-1}^t \sigma_u^{r+s} du. \end{aligned}$$

As formulae (4.36)–(4.39) show, it is relatively easy to compute the various expressions of interest. These expressions indicate that it is also possible to focus on a time increment such as a day rather than on an ongoing process.

To illustrate these various concepts, we consider LVMH over three days around April 1, 2004. This is a day with a relatively large negative return. Figure 4.13 traces the daily returns. As we notice, the returns are relatively large on the negative side for the three consecutive days. To provide a further insight on how the returns of these days got generated, we consider higher frequencies, namely returns at an hourly and 5-minute frequencies.

Figure 4.14 focuses on the hourly frequency. We have added bars to this figure. The longer vertical bar represents the last return of a given day. The shorter lines delimit the various hours. We included the remaining 30 minutes into the computation of the last hour. As this picture reveals, on the first day, there are many relatively small negative returns that cumulate into something larger at the daily level. On the second day, the second hour finishes with a large negative return of about -3% .

To investigate whether this is a jump or just some continuous variation, we further focus on data where the sampling takes place at a 5-minute frequency. This data is displayed in Figure 4.15. These returns reveal that the first few hours of the second day are quite agitated. To investigate if the relatively large drop of -3% stems from a single event or a succession of events, we consider also the raw price process displayed in the next figure.

Figure 4.16 represents price increments at a 5-minute interval. We notice that the prices dropped in a rather steady manner during the second hour suggesting that the -3% return is likely due to some continuous event rather than to a jump. To investigate this conjecture more formally, we turn to

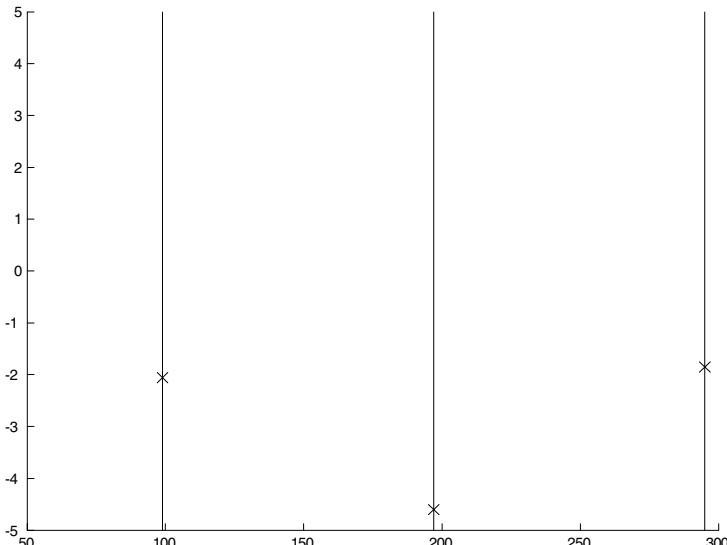


Fig. 4.13. Returns of LVMH sampled once per day.

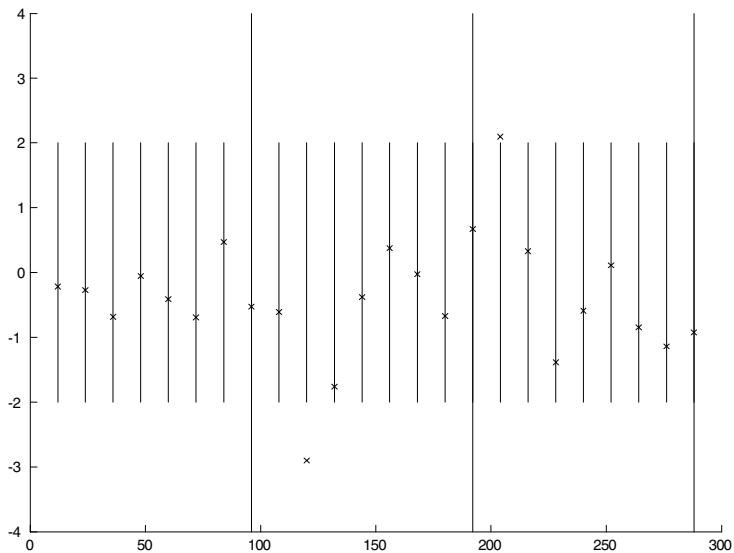


Fig. 4.14. Returns of LVMH sampled at hourly frequency.

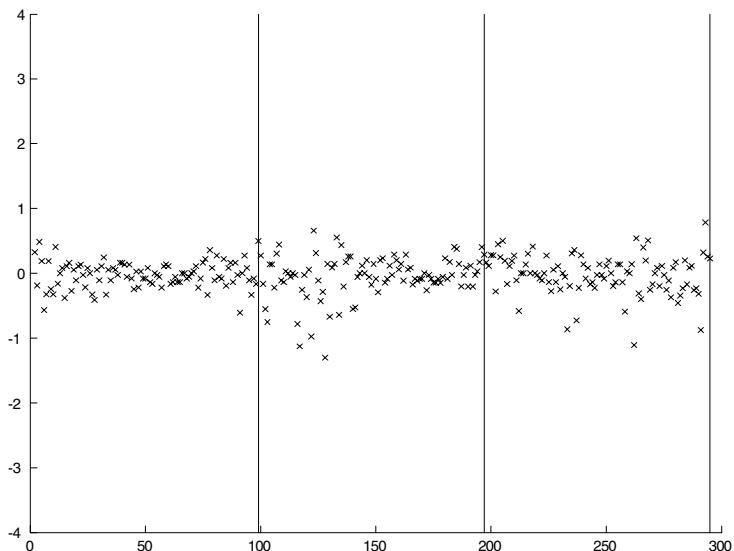


Fig. 4.15. Returns of LVMH sampled every 5 minutes.

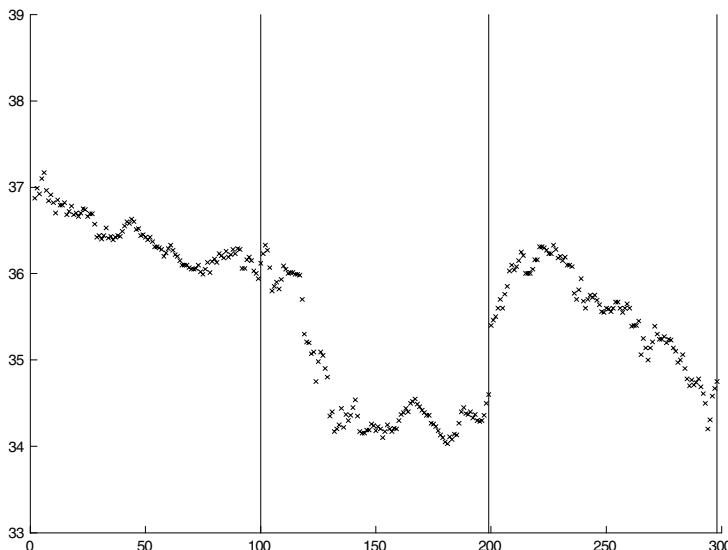


Fig. 4.16. Raw prices of LVMH sampled every 5 minutes.

the concepts of realized variance and bipower variation that should allow the detection of jumps.

Figure 4.17 represents the results of the computation of various statistics. For comparison purposes, we slightly increased the number of days. The largest price variation (on April 1, 2004) is now on day 4. The upper panel of the figure represents the daily estimate of quadratic variation, which is realized variance. This figure also contains the power variation obtained with $r = 1$ over the various days. As this figure shows, the two measures are relatively close. Obviously, given that quadratic variation captures the integrated variance and power variance, for the given power, captures the r th power of the absolute value of standard deviation, there is no reason for these two statistics to take the same value. We just notice that the two statistics evolve in the same direction.

In the lower panel of Figure 4.17, we display the difference between realized variance and realized bipower variation. As indicated by Barndorff-Nielsen and Shephard (2004a), in small samples, the estimate of the jump may yield to a negative estimate. Since the difference of these two statistics should theoretically capture the sum of squared jumps, the negativity of the measure is counterfactual. We may improve this computation by setting a lower bound at zero.

In terms of interpretation, as the lower panel of the figure indicates, there are several other days with jump measures as large as for the given day. Hence, the process for that day should be considered to be generated by a process

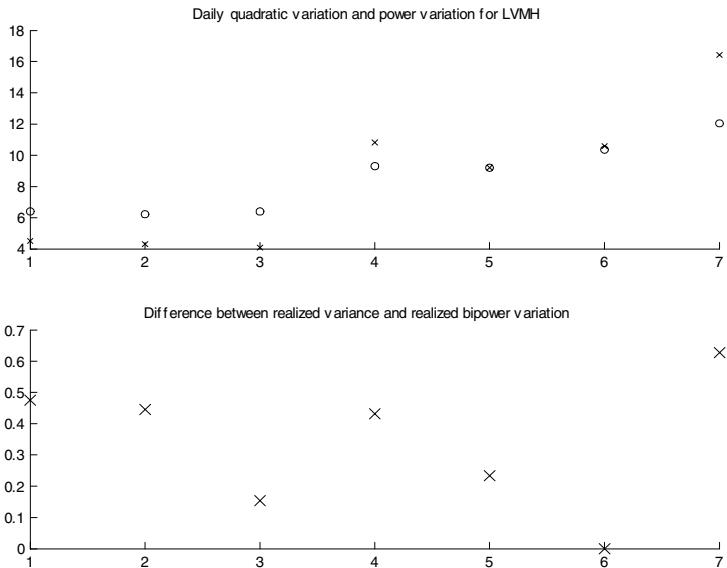


Fig. 4.17. Daily quadratic variation and bipower variation for LVMH.

without discrete jumps. Alternatively, one could imagine that the process jumps very frequently. This, however, appears in contradiction with the fact that jumps are supposed to take place only infrequently.

We may also ask how good the fit is. To do so, we compute for data, covering one year and a half, the various daily returns and standardize them by the square root of daily realized variance. Formally, for day i , we take $r_i / \sqrt{\{p\}_{i|M}^{[2]}}$, with M equal to 102 intervals. We could expect that the resulting measure is distributed as a normal if quadratic variation captures all activity. To perform this test, we trace a so-called quantile plot in Figure 4.18. As the left tail indicates, there remains some information in the returns that does not appear to be captured by realized variance, computed over a daily frequency.²⁷ An additional question that we can ask is the relation between volume and realized variance. Indeed, according to the mixture of distributions hypothesis, volume could represent the daily information content. It was already shown by Clark (1973) that even though these two measures appear correlated, this relation is not a good one. We may verify this result here by contemplating in Figure 4.19 the scatterplot between daily volumes on the horizontal axis (measured by the number of shares traded), and the realized variance along the vertical axis.

²⁷ For stock returns and exchange rates, Andersen et al. (2001) and Andersen et al. (2003) provide empirical evidence that returns scaled by realized standard deviations are approximately Gaussian.

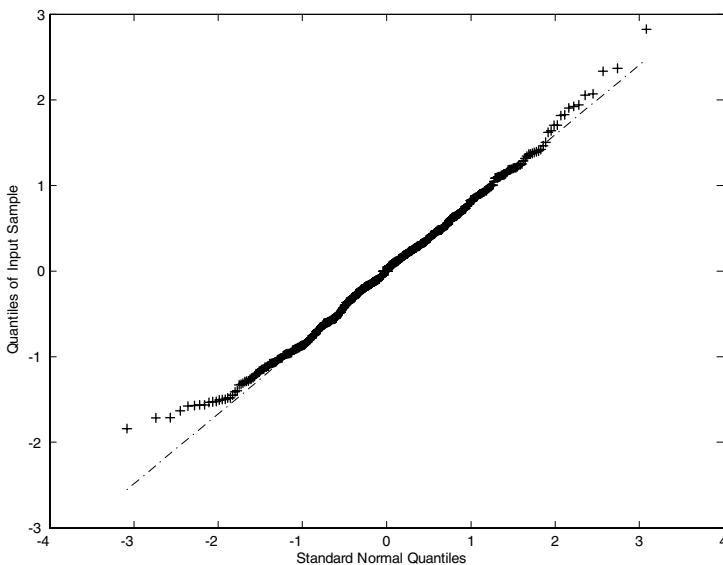


Fig. 4.18. Daily devolatilized returns.

We notice that the largest daily realized variances are associated with high volume but not the largest ones. Thus, even though there certainly exists a relation between volume and daily realized variances, this relation is far from being simple.

So far, our focus has been on univariate data. We turn now to the behavior of asset prices within a multivariate setting. In the following section, we start by providing theoretical concepts before turning to some empirical illustrations.

4.8.6 Realized covariance

Theory

It is possible to obtain a construction for high-frequency measures of covariance and correlation. Whereas Andersen et al. (2003) indicate how to compute the covariance from high-frequency data and its properties, BN/S provide a distribution theory.

Definition 4.12. Consider two processes, written p^1 and p^2 defined over some time interval $[0, t]$. The realized covariance is given by the expression

$$[p^1, p^2]_t \equiv \underset{\Delta \rightarrow 0}{\text{plim}} \sum_{j=1}^{\lfloor t/\Delta \rfloor} (p_{t_j}^1 - p_{t_{j-1}}^1)(p_{t_j}^2 - p_{t_{j-1}}^2).$$

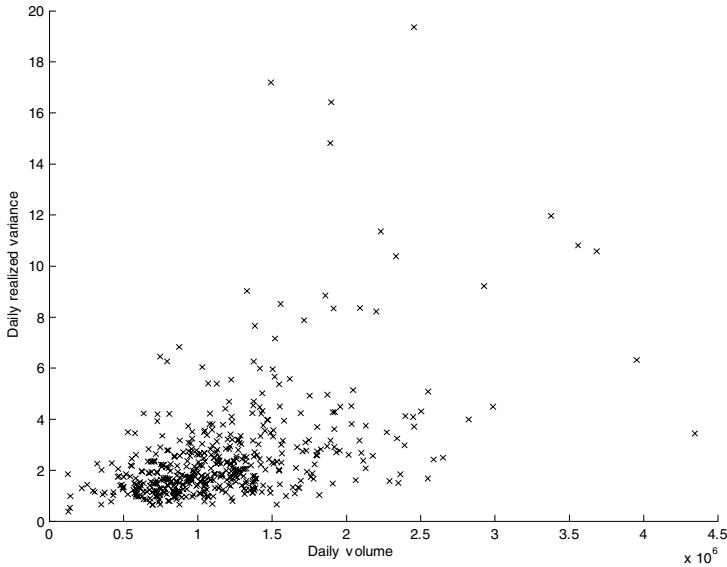


Fig. 4.19. Daily realized variance and volume for LVMH.

If both p^1 and p^2 belong to the \mathcal{SVSM}^C class, then we can always, by definition, write that

$$\begin{pmatrix} p_t^1 \\ p_t^2 \end{pmatrix} = a_t + m_t, \quad \text{and} \quad m_t = \int_0^t \theta_u dW_u,$$

where a_t is a $(2, 1)$ vector of drift terms and m_t is a two-dimensional martingale. The θ_u is a $(2, 2)$ matrix, and W_u represents a two-dimensional Brownian motion. From there on, we may write

$$\Sigma \equiv \theta_u \theta'_u = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}.$$

Clearly, σ_{12} plays the role of a covariance. It may be shown that the realized covariance converges to an integrated covariance

$$[p^1, p^2]_t = \int_0^t \sigma_{12,u} du.$$

Interestingly, if we have such estimates of a covariance matrix, it becomes possible to also discuss portfolio allocations from a high-frequency point of view. As previously, it is possible to specialize realized covariance to an interval of time. We consider a given day i that is decomposed into M time intervals, each of size δ . Then we have

$$\begin{aligned}\widehat{[p^1, p^2]}_{i|M} &= \sum_{j=1}^M \left(p_{(i-1)M\delta+\delta j}^1 - p_{(i-1)M\delta+\delta(j-1)}^1 \right) \\ &\quad \times \left(p_{(i-1)M\delta+\delta j}^2 - p_{(i-1)M\delta+\delta(j-1)}^2 \right).\end{aligned}$$

This expression is, thus, an estimate of the integrated covariance and will be called realized covariance. In the limit as M converges to infinity, realized covariance will converge to the increment of integrated covariance.

Using the realized covariance as well as realized variances, it is possible to compute a measure of realized correlation defined as

$$\frac{\widehat{[p^1, p^2]}_{i|M}}{\sqrt{\widehat{[p^1]}_{i|M} \widehat{[p^2]}_{i|M}}}.$$

We notice that this expression can be easily computed.

An empirical measure of realized correlation

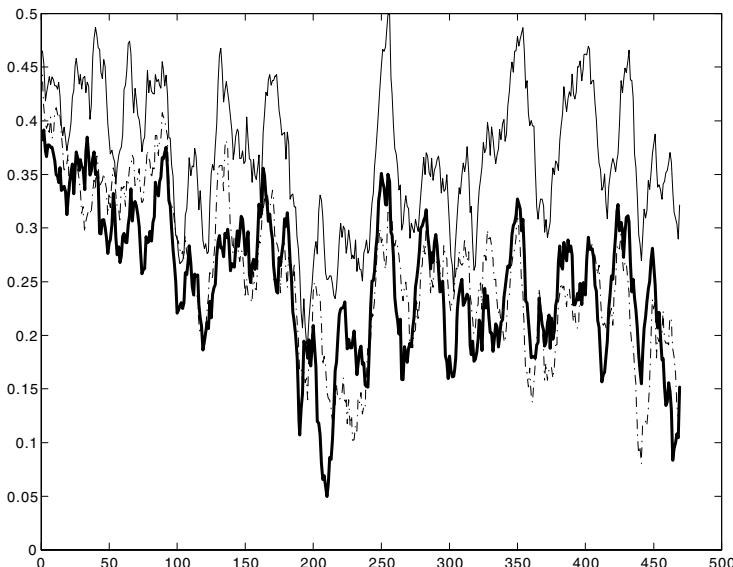
Given the concepts of realized covariance and realized variance, it is also possible to construct a realized correlation. We just measure a covariance for data sampled at some frequency over some time intervals, for instance daily, and standardize it by the corresponding realized volatility. Turning to the three companies already considered, it is possible to investigate the correlation between the assets. We take all the data over the one year and a half, extract returns at various frequencies and then compute realized covariances, realized volatilities, and the resulting realized correlation. Table 4.7 presents the resulting measures. We notice that the correlation between AXA and LVMH is about the same as between LVMH and Société Générale, if we focus on 5-minute returns and that it takes a value of about 0.3. The corresponding correlation between AXA and Société Générale is 0.41. Investigation of the table shows that as the sampling frequency increases, the correlations become larger even though the overall ranking of the correlations remains the same.

It is also possible to investigate the intertemporal evolution of correlation. In Figure 4.20, we represent daily correlations computed for the three companies under investigation. The correlations are obtained by taking 5-minute returns for each day. To ease the detection of patterns, the correlations have been smoothed with a 10-day gliding average. The thick line represents the correlations between AXA and LVMH. The thin continuous line and the dotted line correspond to the correlations between AXA and Société Générale, respectively LVMH and Société Générale.

As this figure suggests, the correlations between the various companies are systematically anchored at different levels. We also notice that the correlations vary through time. To investigate the stability of the correlations through time, we take again AXA–LVMH and consider daily correlations computed

Table 4.7. Realized correlations for high-frequency data

	AXA	LVMH	Soc. Gén.
5-minute returns			
AXA	1.00	0.30	0.41
LVMH	0.30	1.00	0.31
Soc. Gén.	0.41	0.31	1.00
Hourly returns			
AXA	1.00	0.53	0.63
LVMH	0.53	1.00	0.51
Soc. Gén.	0.63	0.51	1.00
Returns over 4 hours			
AXA	1.00	0.60	0.65
LVMH	0.60	1.00	0.53
Soc. Gén.	0.65	0.53	1.00
Daily returns			
AXA	1.00	0.63	0.69
LVMH	0.63	1.00	0.55
Soc. Gén.	0.69	0.55	1.00

**Fig. 4.20.** Daily realized correlations computed with 5-minute returns.

with data sampled at the 5-minute frequency and the hourly frequency. Again, we smooth the data with a 10-day gliding average. As Figure 4.21 represents, there is quite a significantly different pattern in the correlations. This raises the question of the frequency at which the data should be sampled for this type of measure.

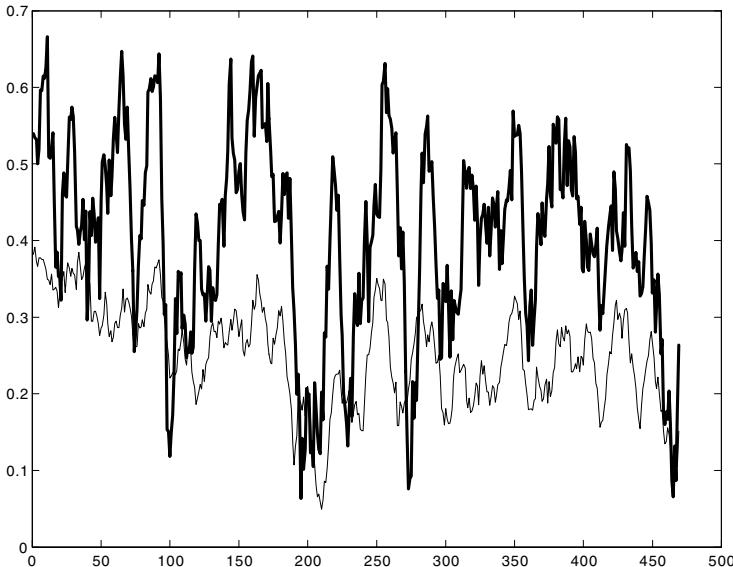


Fig. 4.21. Daily realized correlations computed at 5-minute and 1-hour frequency.

The time variability of correlation is an issue that is also addressed at the beginning of Chapter 6 with multivariate GARCH models. The link between measures of correlation, based on high-frequency data, and GARCH correlations appears to be an open issue.

Testing formally for jumps

So far, we have indicated how to decompose a price process into a continuous part and a discrete part. The following elements, drawn from Barndorff-Nielsen and Shephard (2004a), may be useful to formally test for the existence of jumps. See also BN/S.

As before, denote by $\mu_r = E[|u|^r]$, the r th moment of u . When a formula contains u , u' , and u'' , we assume that the associated distributions are all normal $\mathcal{N}(0, 1)$ and independent. If p belongs to the class of \mathcal{SVSM}^C , then as Δ converges toward 0, for any $r > 0$,

$$\frac{\Delta^{-\frac{1}{2}} \left(\{p\}_t^{[r,r]} - \mu_r^2 \int_0^t \sigma_s^{2r} ds \right)}{\sqrt{\mu_{4r}^{-1} \{V [|u|^r |u'|^r] + 2Cov [|u|^r |u'|^r, |u'|^r |u''|^r]\} [p]_t^{[4r]}}} \Rightarrow \mathcal{N}(0, 1),$$

where \Rightarrow corresponds to convergence in distribution. Even though this result has been derived for the general case with r taking a general value, this result may prove most useful for the case where $r = 1$. Indeed, this is the case

corresponding to quadratic variation. In this case, the formula allows for a construction of a confidence interval to which the true quadratic variation belongs. If the expression of $\int_0^t \sigma_s^{2r} ds$ was known, we could even consider testing for jumps. Indeed, if the statistics fall outside the confidence interval, this may be interpreted as a failure of the null hypothesis. This could then indicate the presence of a jump in the data. Since, in practice, the integral is not known, we need further results.

Fortunately, BN/S also provide results for this case, where we work with quadratic variation and bipower variation. To test if there are jumps, we need indeed a statistics concerning the difference between a quadratic variation and a bipower variation. This is where the following result is useful. This result holds for the case when the price process is again in the \mathcal{SVSM}^C class. We have

$$\frac{\Delta^{-\frac{1}{2}}}{\sqrt{\int_0^t \sigma_s^4 ds}} \left(\begin{array}{l} \sum_{j=1}^{\lfloor t/\Delta \rfloor} (p_{j\Delta} - p_{(j-1)\Delta})^2 - \int_0^t \sigma_s^2 ds \\ \mu_1^{-2} \sum_{j=1}^{\lfloor t/\Delta \rfloor - 1} |p_{j\Delta} - p_{(j-1)\Delta}| |p_{(j+1)\Delta} - p_{j\Delta}| - \int_0^t \sigma_s^2 ds \end{array} \right) \Rightarrow \mathcal{N}(0, \Sigma)$$

where

$$\Sigma = \begin{pmatrix} V[u^2] & 2\mu_1^{-2} \text{Cov}[u^2, |u| |u'|] \\ 2\mu_1^{-2} \text{Cov}[u^2, |u| |u'|] & \mu_1^{-4} \{V[|u| |u'|] + 2\text{Cov}[|u| |u', |u'| |u''|]\} \end{pmatrix}.$$

This covariance matrix may be computed once and for all numerically because none of the terms depends on the actual data. BN/S obtain

$$\Sigma = \begin{pmatrix} 2 & 2 \\ 2 & 2.60907 \end{pmatrix}.$$

These computations show that the efficiency of quadratic variation is somewhat better than of realized bipower variation. Indeed, once the variance is 2, the other time it is about 2.6. The correlation between the two statistics is equal to 0.87. This implies that the information content of both statistics will be very similar. The previous results are clearly obtained under the null hypothesis of no jumps. If we wish to test whether there are jumps, we may use the following expression:

$$\frac{\Delta^{-\frac{1}{2}}}{\sqrt{\int_0^t \sigma_s^4 ds}} \left[\sum_{j=1}^{\lfloor t/\Delta \rfloor} (p_{j\Delta} - p_{(j-1)\Delta})^2 - \mu_1^{-2} \sum_{j=1}^{\lfloor t/\Delta \rfloor - 1} |p_{j\Delta} - p_{(j-1)\Delta}| |p_{(j+1)\Delta} - p_{j\Delta}| \right] \Rightarrow \mathcal{N}(0, 0.6091).$$

Hence, as mentioned, this statistic can be used to test for the presence of jumps. To render this formula operational, it is necessary to evaluate the

integrated fourth power of volatility, also called *integrated quarticity*, that appears under the square-root. This can be done by using the fact that,

$$\lim_{\Delta \rightarrow 0} \Delta^{-1} \sum_{j=1}^{\lfloor t/\Delta \rfloor - 1} (p_{j\Delta} - p_{(j-1)\Delta})^2 (p_{(j+1)\Delta} - p_{j\Delta})^2 = \int_0^t \sigma_s^4 ds,$$

provides a feasible estimate of *realized quarticity*, an empirical measure of integrated quarticity. It is, thus, sufficient for practical purposes to approximate the integral by a finite sum obtained for a selected Δ .

4.8.7 Further related results

So far, we have discussed some recent theoretical developments. We now present some of the empirical results that have been found in the literature. First, we may mention a literature that deals with comparing volatility forecasts to actual volatility realizations. One such contribution is the one by Andersen et al. (2003). It focuses, like much of this literature, on exchange rate data. These contributions show that realized volatility is best forecast by using a vector autoregression involving multivariate realized volatility and where some long memory is allowed for by using a fractional difference, rather than by a model belonging to the GARCH family. As we have seen in the above statistical analysis of French stock market data, the sampling frequency plays an important role. Sampling data at a 5-minute frequency will not yield to the same pattern of autocorrelation than if hourly data is used. The reason for this is that at an increasing frequency, so-called *microstructure noise* comes into play. This noise may be explained by the bid-ask spread, by the fact that a given asset price may be coming from different markets, or due to the fact that prices are traded at discrete values. The contribution by Andersen, Bollerslev, and Meddahi (2004) shows that, in the presence of microstructure noise, the realized volatility forecast is even better than originally thought.

As just mentioned, the role of microstructure noise may hamper the estimation of realized volatility or of realized covariances. The question then is of course what time frequency we should use. As seen at the beginning of this section, the data should be sampled at a high enough frequency to allow for a disentangling of jumps. On the other hand, a finer and finer sampling, say at the level of minutes or seconds, will introduce an increased level of microstructure noise. Several contributions, such as Aït-Sahalia, Mykland, and Zhang (2005a and 2005b) investigate on how to improve the measure of realized volatility in the presence of microstructure noise. Their starting point is a simplified model where observed log-prices, \tilde{p}_t , are the sum of fundamental, unobservable log-prices, p_t , and some noise, ε_t . Beyond the already mentioned sources of noise, such as price discreteness, the noise may represent elements such as price impact due to trade size, inventory components of the bid-ask spread, etc. These authors adopt the following model

$$\begin{aligned}\tilde{p}_t &= p_t + \varepsilon_t, \\ dp_t &= \mu_t dt + \sigma_t dW_t,\end{aligned}$$

where ε_t can be *iid* or dependent. They show that a good estimate, named Two Scales Realized Volatility (TSRV), is based on two components. First, one computes at a given slow timescale, such as at the 5-minute frequency, different daily realized volatilities where each measure starts at a different point of time. For instance, the first such realized volatility would start at 8:00:00 and then pursue at the 5-minute frequency, the second one would start at 8:00:30 and then pursue at 5-minute frequency, and so on. These realized volatilities would then be averaged. In other words, the first measure would be obtained by subsampling. As shown by Barndorff-Nielsen and Shephard (2004a), and as has been used already by Andersen, Bollerslev, and Meddahi (2004), this estimate will not be a good estimate of integrated variance because of some discretization bias. By combining this first averaged realized volatility with a realized volatility computed at a different timescale, for instance a fast timescale such as every 10 seconds, a more precise estimate may be obtained. The proposed estimator of realized volatility is

$$\widehat{[p]}_i^{[TSRV]} = \frac{1}{K} \sum_{k=1}^K \widehat{[p]}_{k;i|M}^{[2]} - \frac{M}{M'} \widehat{[p]}_{i|M'}^{[2]}$$

where

$$\widehat{[p]}_{k;i|M}^{[2]} = \sum_{j=1}^M (p_{(i-1)M\delta+t_k+\delta j} - p_{(i-1)M\delta+t_k+\delta(j-1)})^2,$$

and where t_k , for $k = 1, \dots, K$ are starting moments during the first daily interval $[0, \delta]$, M and M' are the number of intervals taken for the slow and the fast timescale. The suggestion to use estimates based on different timescales may also be found in Bandi and Russel (2004).

A better measurement of volatility is likely to be useful for option pricing. A decomposition of prices into continuous and discrete components may improve hedging. An analysis of volatility may also shed light on the behavior of investors. One paper in this direction is the one by Andersen, Bollerslev, and Diebold (2003). The authors consider realized volatilities measured over timescales and show, in a vector autoregression framework, that realized volatility contains complementary information at each of these timescales. This suggests that there are heterogeneous investors whose time horizons are not the same. Such an observation suggests that a better understanding of investor behavior from an economic point of view may be useful.

Modeling Higher Moments

In this chapter, we investigate two important issues for the modeling of asset returns. The first issue is the modeling of the entire density of returns, so that it incorporates some of the features described in Chapter 2, in particular the asymmetry and the fat-tailedness of the distribution. In several problems in finance, it is crucial to recognize that the conditional distribution of returns is non-normal and to correctly model it. A precise knowledge of the conditional distribution is required, for instance, for asset allocation, VaR (Value-at-Risk) computation, or the valuation of contingent claims.

The second issue is the modeling of the time-variability of higher moments, which can be viewed as summaries of the density. This is a key issue for problems that involve data sampled at a rather high frequency (say daily or weekly data) or very non-normal asset returns. For instance, the modeling of conditional higher moments will be crucial in the computation of conditional VaR, which is a short-term measure typically calculated for a 1- or 10-day horizon, in which case the modeling of higher moments may substantially improve the accuracy of VaR. Let us consider now the case of asset allocation. Exactly as the model of Markovitz builds on the first two moments, mean and variance, more advanced models build on moments beyond the first two. At first glance, the existence of time-varying higher moments may be expected to strongly affect the allocation of wealth. Yet, asset allocation is often considered at a rather low frequency (with a monthly or even a quarterly horizon). At such intervals, higher moments are less likely to be time-varying. For asset allocation purpose, we may conclude that the modeling of the higher moments is crucial for asset classes with very non-normal returns, such as hedge funds or emerging market indices. A more complete description of these issues is provided in Chapters 8 and 9.

We therefore consider now the explicit modeling of the higher moments of asset returns and their conditional distributions. Since the seminal work of Engle (1982), time-varying volatility has been shown to produce fat tails in the unconditional distribution. But time-varying volatility alone is not enough to explain all the tail fatness; volatility filtered residuals still have tails fatter than

the normal distribution. Instead of considering all the possible distributions that may fit the returns data empirically, our strategy here is to focus on the two stylized facts above and set out to find distributions that can capture these two characteristics. Although there are already many fat-tailed extensions to the normal distribution (such as the Pareto and the Student t), not all of these distributions can capture asymmetry. Hence, finding a distribution with a suitable asymmetry property will be a main objective here. An important criteria in this search for a better alternative distribution is to have an as large as possible range of admissible skewness and kurtosis. Ideally, the only constraints on this domain of definition would be those ensuring that the distribution is definite. However, most distributions discussed in the literature typically impose additional restrictions on this domain of definition.

The remaining of this chapter is organized as follows. Section 5.1 presents the general problems involved in the modeling of higher moments. Section 5.2 describes a number of distributions that can be used to capture higher moments. In Section 5.3, we address the issues of specification tests and inference. We provide an illustration of some of the material presented in this Chapter in Section 5.4. Finally, Section 5.5 describes various ways to model higher moments conditional on past observations.

5.1 The general problem

Let x_t , for $t = 1, \dots, T$, be a time series of asset returns. It is convenient to break down the complete characterization of x_t into three components: (i) the conditional mean, which contains all the information about the location of the distribution, (ii) the conditional variance, which contains the scale parameter that measures the dispersion of the distribution, and (iii) the shape parameters (e.g., skewness, kurtosis) that determine the form of a conditional distribution within the general family of distributions. Thus, we may write

$$x_t = \mu_t(\theta) + \varepsilon_t, \quad (5.1)$$

$$\mu_t(\theta) = E[x_t | \mathcal{F}_{t-1}] = \mu(\theta, \mathcal{F}_{t-1}), \quad (5.2)$$

$$\varepsilon_t = \sigma_t(\theta) z_t, \quad (5.3)$$

$$\sigma_t^2(\theta) = E[(x_t - \mu_t)^2 | \mathcal{F}_{t-1}] = \sigma^2(\theta, \mathcal{F}_{t-1}), \quad (5.4)$$

$$z_t \sim g(z_t | \eta). \quad (5.5)$$

Equation (5.1) decomposes the return at time t into a conditional mean, μ_t , and an error term, ε_t . The dynamics of the conditional mean is given by (5.2). The standardized innovation, $z_t = (x_t - \mu_t(\theta)) / \sigma_t(\theta)$ has zero mean and unit variance. Equation (5.4) determines the dynamics of volatility. This may be any specification including the GARCH models. Vector θ contains all the parameters associated with the conditional mean and the conditional variance equations. Finally, equation (5.5) specifies that the standardized innovation

follows a conditional distribution g with vector of shape parameters η . In the previous chapter, the conditional distribution was assumed to be normal $\mathcal{N}(0, 1)$ with no shape parameter. In the more general case we consider now, shape parameters η will generally involve parameters capturing asymmetry and fat-tailedness of the distribution.

In this section, we address several issues related to the modeling of non-normal returns:

1. If the unconditional distribution has been found to be non-normal, do we necessarily have to assume that the conditional distribution $g(\cdot)$ is non-normal?
2. If the conditional distribution is assumed to be non-normal, do we necessarily have to model it explicitly?
3. If we have to model the conditional distribution explicitly, how far can we go in terms of asymmetry and fat-tailedness of the distribution?

5.1.1 Higher moments of a GARCH process

The idea behind the first issue is that the modeling of the conditional volatility renders the unconditional distribution of the error term ε_t non-normal, even if the conditional distribution of the innovation z_t is still assumed to be normal. This result, highlighted by Engle and González-Rivera (1991), can be of great importance for series that are characterized by a slight departure from normality. In such a case, estimating a GARCH model with a normal conditional distribution may be enough to capture the fat-tailedness of the unconditional distribution. To explain why this is the case, we need to distinguish between the conditional and the unconditional higher moments.

Conditional and unconditional skewness and kurtosis

Let σ^2 denote the unconditional variance of ε_t . The conditional skewness and kurtosis are defined conditionally to the information set \mathcal{F}_{t-1}

$$s_c = \frac{E[\varepsilon_t^3 | \mathcal{F}_{t-1}]}{\sigma_t^3},$$

$$\kappa_c = \frac{E[\varepsilon_t^4 | \mathcal{F}_{t-1}]}{\sigma_t^4},$$

whereas their unconditional counterparts are defined as

$$s_u = \frac{E[\varepsilon_t^3]}{(E[\sigma_t^2])^{3/2}} = \frac{E[\varepsilon_t^3]}{\sigma^3},$$

$$\kappa_u = \frac{E[\varepsilon_t^4]}{(E[\sigma_t^2])^2} = \frac{E[\varepsilon_t^4]}{\sigma^4}.$$

Although not shown explicitly above, the shape parameters of the unconditional distribution (s_u and κ_u) will depend on the characteristics of the conditional distribution. Indeed consider the model (5.1)–(5.5) above, so that $z_t = \varepsilon_t/\sigma_t$ is an *iid* process. Then, for any distribution such that $E[\varepsilon_t^3] < \infty$, the unconditional skewness is given by

$$s_u = \frac{E[\varepsilon_t^3]}{(E[\sigma_t^2])^{3/2}} = \frac{E[E[\varepsilon_t^3 | \mathcal{F}_{t-1}]]}{(E[\sigma_t^2])^{3/2}} = \frac{E[s_c \sigma_t^3]}{(E[\sigma_t^2])^{3/2}},$$

so that

$$s_u = s_c \frac{E[\sigma_t^3]}{(E[\sigma_t^2])^{3/2}}.$$

Because of Jensen's inequality, one has $E[\sigma_t^3] \geq (E[\sigma_t^2])^{3/2}$ as $\sigma_t > 0$, and thus $|s_u| \geq |s_c|$. Therefore, the unconditional skewness has the same sign as the conditional one, but it is possible to capture a high unconditional skewness using a conditional distribution with a smaller skewness. Notice that in cases where the conditional distribution is symmetric, introducing GARCH effects would not result in an asymmetric distribution.

Assuming that $E[\varepsilon_t^4] < \infty$, the unconditional kurtosis is given by

$$\kappa_u = \frac{E[\varepsilon_t^4]}{(E[\sigma_t^2])^2} = \frac{E[E[\varepsilon_t^4 | \mathcal{F}_{t-1}]]}{(E[\sigma_t^2])^2} = \frac{E[\kappa_c \sigma_t^4]}{(E[\sigma_t^2])^2},$$

so that

$$\kappa_u = \kappa_c \frac{E[\sigma_t^4]}{(E[\sigma_t^2])^2}.$$

Once again, Jensen's inequality implies $E[\sigma_t^4] \geq (E[\sigma_t^2])^2$ for $\sigma_t > 0$, and hence $\kappa_u \geq \kappa_c$. Therefore, even in the case where innovations z_t are assumed to be normal, a GARCH model yields an unconditional distribution with fatter tails than the normal distribution.

Conditional normal distribution

To be more explicit, consider an ARCH(1) model, with $\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2$, and normal innovations. Then, as shown in Chapter 4, we have

$$\kappa_u = 3 \frac{1 - \alpha_1^2}{1 - 3\alpha_1^2},$$

which exceeds 3 for $\alpha_1 > 0$ and $3\alpha_1^2 < 1$.

For a GARCH(1,1) model, with $\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$, Bollerslev (1986) obtains

$$\kappa_u = 3 \frac{1 - \beta_1^2 - 2\alpha_1\beta_1 - \alpha_1^2}{1 - \beta_1^2 - 2\alpha_1\beta_1 - 3\alpha_1^2},$$

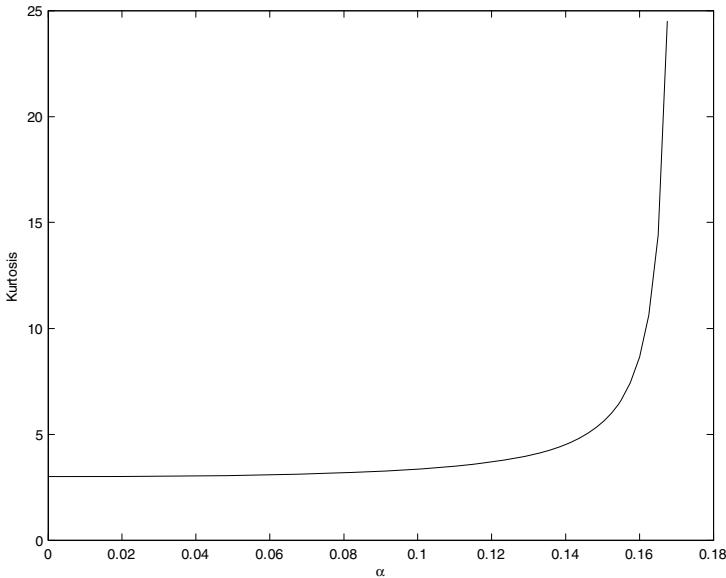


Fig. 5.1. Unconditional kurtosis of a GARCH(1,1) process when α varies (with $\beta = 0.8$).

which exceeds 3 for $\alpha_1 > 0$, $\beta_1 > 0$ and $\beta_1^2 + 2\alpha_1\beta_1 + 3\alpha_1^2 < 1$.

Figure 5.1 illustrates how the unconditional kurtosis κ_u of a GARCH(1,1) model varies when we increase the persistence of volatility. More precisely, we assume $\beta_1 = 0.8$ and vary α_1 between 0.05 and 0.16. Then, the unconditional kurtosis increases from 3 to virtually infinity. Controlling the persistence of the conditional volatility thus allows to control the fat-tailedness of the unconditional distribution.

Conditional non-normal distributions

Often, a GARCH model with a normal conditional distribution is not sufficient to capture the fat tails found in the unconditional distribution. This implies that the standardized innovation z_t is not normal and that we need a more general specification for the conditional distribution.

The first attempt to capture the excess kurtosis of GARCH standardized residuals is by Bollerslev (1987), who uses as conditional distribution $g(z_t|\eta)$ a Student t distribution. The density of the symmetric Student t distribution with zero mean and unit variance is given by

$$t(z_t|\nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi(\nu-2)\Gamma(\frac{\nu}{2})}} \left(1 + \frac{z_t^2}{\nu-2}\right)^{-\frac{\nu+1}{2}}, \quad (5.6)$$

where $\Gamma(\cdot)$ is the gamma function and ν the degree-of-freedom parameter. The kurtosis of the conditional distribution of z_t is

$$\kappa_c = 3 \frac{\nu - 2}{\nu - 4},$$

provided that $\nu > 4$. The degree-of-freedom adjustment contributes to excess kurtosis, because it is by construction larger than one, except when $\nu \rightarrow \infty$, in which case the Student t distribution collapses to a normal distribution. κ_c is always larger than 3 and decreases toward 3 when the degree-of-freedom parameter ν decreases toward 4. The unconditional kurtosis of the error term ε_t is then given by

$$\kappa_u = 3 \frac{\nu - 2}{\nu - 4} \frac{E[\sigma_t^4]}{(E[\sigma_t^2])^2}.$$

The term involving the degree-of-freedom parameter contributes to capture further excess kurtosis, because it is by construction larger than one.

Nelson (1991) also proposed the GED (*Generalized Error Distribution*) defined as

$$g(z_t|\nu) = \frac{\nu}{2^{1+1/\nu} \lambda \Gamma[1/\nu]} \exp\left(-0.5 \left| \frac{z_t}{\lambda} \right|^{\nu}\right) \quad 0 < \nu < \infty, \quad (5.7)$$

with $\lambda = (2^{-2/\nu} \Gamma[1/\nu] / \Gamma[3/\nu])^{1/2}$. When $\nu = 2$, the distribution in (5.7) reduces to the normal distribution. When $\nu < 2$ ($\nu > 2$), the conditional distribution has fatter (thinner) tails than the normal distribution. The kurtosis of the conditional distribution of z_t is

$$\kappa_c = \frac{\Gamma[1/\nu] \Gamma[5/\nu]}{(\Gamma[3/\nu])^2},$$

so that it is larger than 3 when $\nu < 2$. The unconditional kurtosis of ε_t is then given by

$$\kappa_u = \frac{\Gamma[1/\nu] \Gamma[5/\nu]}{(\Gamma[3/\nu])^2} \frac{E[\sigma_t^4]}{(E[\sigma_t^2])^2}.$$

Bai, Russell, and Tiao (2003) provide the exact representation of unconditional kurtosis for GARCH and stochastic volatility models, when innovations are conditionally non-normal. They show that the unconditional kurtosis for both models is determined jointly by the conditional distribution of z_t and the volatility clustering of ε_t .

5.1.2 Quasi Maximum Likelihood Estimation

When the conditional distribution g is not normal, the ML approach described in Section 4.3.3 cannot be directly used, because this estimation procedure assumes that the true conditional distribution of innovations is normal. However, Gouriéroux, Monfort, and Trognon (1984), Weiss (1986), and Bollerslev

and Wooldridge (1992) have shown that, provided the first and second moments are correctly specified, the parameters θ pertaining to the conditional mean and the conditional variance equations can be consistently estimated by maximizing the normal likelihood function, even if the true distribution is not normal. This procedure is called the Quasi Maximum Likelihood Estimation (QMLE). The ML and QML estimators $\hat{\theta}$ of θ are the same, because they are the solutions of the same maximization problem. However, the covariance matrices of the estimators differ, because the QML covariance matrix is computed without assuming conditional normality.

The QML estimator of θ , denoted $\hat{\theta}_{QML}$, is obtained by maximizing the conditional normal log-likelihood function, defined as

$$L_T(\theta | \underline{x}_T) = \sum_{t=1}^T \ell_t(\theta), \quad (5.8)$$

where

$$\ell_t(\theta) = -\frac{1}{2} \log(\sigma_t^2(\theta)) + \log(g(z_t)) = -\frac{1}{2} \log(2\pi) - \frac{1}{2} \log(\sigma_t^2(\theta)) - \frac{1}{2} z_t^2,$$

and $z_t = (x_t - \mu_t(\theta)) / \sigma_t(\theta)$. So the maximization problem only depends on the first and second moments. Consequently, if they are correctly specified, the QML estimation will lead to a consistent estimator $\hat{\theta}_{QML}$.

The asymptotic distribution of $\hat{\theta}_{QML}$ is given by

$$\sqrt{T}(\hat{\theta}_{QML} - \theta_0) \Rightarrow \mathcal{N}(0, \Omega), \quad (5.9)$$

where θ_0 is the true value of the parameter vector. Under normality, the asymptotic covariance matrix Ω is simply the inverse of the information matrix, as described in Section 4.3.3. When normality is not assumed, the standard errors of the QML estimator have to be “robustified”, following White (1982) and Gouriéroux, Monfort, and Trognon (1984).¹ They give asymptotically valid confidence intervals for the “pseudo-true” parameter values that minimize the information distance between the true probability measure and the normal likelihood. The robust standard errors are the square roots of the diagonal elements of the matrix

$$\Omega = A_0^{-1} B_0 A_0^{-1},$$

where A_0 is the information matrix evaluated at the true parameter vector θ_0

$$A_0 = -E \left[\frac{\partial^2 \ell_t(\theta_0)}{\partial \theta \partial \theta'} \right],$$

and B_0 is the outer product of the gradients evaluated at θ_0

¹ In the statistical literature, such standard errors are also called “sandwiched standard errors”, given that a matrix is “sandwiched” between two others.

$$B_0 = \frac{1}{T} \sum_{t=1}^T E \left[\frac{\partial \ell_t(\theta_0)}{\partial \theta'} \frac{\partial \ell_t(\theta_0)}{\partial \theta} \right],$$

with $\partial \ell_t(\theta_0) / \partial \theta'$ the score vector of $\ell_t(\theta)$.

In finite sample, the asymptotic covariance matrix is estimated by

$$\hat{\Omega}_T = \hat{A}_T^{-1} \hat{B}_T \hat{A}_T^{-1},$$

where

$$\begin{aligned}\hat{A}_T &= -\frac{1}{T} \sum_{t=1}^T \frac{\partial^2 \ell_t(\hat{\theta}_{QML})}{\partial \theta \partial \theta'}, \\ \hat{B}_T &= \frac{1}{T} \sum_{t=1}^T \frac{\partial \ell_t(\hat{\theta}_{QML})}{\partial \theta'} \frac{\partial \ell_t(\hat{\theta}_{QML})}{\partial \theta}',\end{aligned}$$

are evaluated at the QML estimates $\hat{\theta}_{QML}$.²

The QML procedure has the advantage of robustness with respect to the distributional assumption of the model. Obviously, if the distribution is not normal, the QML estimator becomes inefficient. It has been shown by Engle and González-Rivera (1991) that Gaussian QMLE has a degree of inefficiency that increases with the degree of departure from normality. Using simulation evidence, Engle and González-Rivera (1991) show that the loss of efficiency in using the Gaussian QMLE instead of the MLE with the true distribution may be as high as 84% (i.e., the variance of the QMLE can be 6.25 times larger than the variance of the MLE). This finding strongly supports the call for more efficient estimators.

A current practice, therefore, consists in using QMLE for a non-Gaussian distribution. For instance, Bollerslev (1987) estimates a GARCH model assuming a Student t distribution. In this case, the log-likelihood is maximized under the assumed distribution, using the corresponding likelihood

$$L_T(\psi | x_t, t = 1, \dots, T) = \sum_{t=1}^T \ell_t(\psi),$$

where

$$\ell_t(\psi) = -\frac{1}{2} \log(\sigma_t^2(\theta)) + \log(g(z_t(\theta) | \eta)),$$

and $\psi = (\theta', \eta')'$ is a vector of parameters. The covariance matrix of the estimator $\hat{\psi}_{QML}$ is robustified to account for departure from the assumed distribution.

² Notice that, if the innovation process is actually normally distributed, we have $A_0 = B_0$, so that $\Omega = A_0^{-1}$. Under normality, the asymptotic covariance matrix of the ML estimator is simply given by the inverse of the Hessian matrix.

The dilemma in abandoning the Gaussian QMLE, as Newey and Steigerwald (1997) point out, is that when an incorrect non-Gaussian specification is used, the estimators may no longer be consistent. They show that consistency of a non-Gaussian QMLE is achieved if either (i) the conditional mean is identically zero; or (ii) the assumed (theoretical) and true (empirical) error *pdfs* are symmetric about zero.

If the symmetry condition is not satisfied, the correct specification of the conditional mean and variance is no longer sufficient to ensure consistency of the QML estimators, as the mean and the variance need not correspond to the natural location and scale parameters.³ An additional location parameter is needed in this case to position the true distribution in order to satisfy the identification condition for consistency. The location parameter accounts for the asymmetry of z_t (i.e., the discrepancy between the conditional mean and the natural location parameter) and can be introduced in either the conditional mean equation or the distribution function. Newey and Steigerwald (1997) suggest estimating the following model

$$x_t = \mu_t(\theta) + \sigma_t(\theta)(\alpha + z_t),$$

so that

$$\ell_t(\theta, \eta, \alpha) = -\frac{1}{2} \log(\sigma_t^2(\theta)) + \log \left(g \left(\frac{x_t - \mu_t(\theta) - \alpha \sigma_t(\theta)}{\sigma_t(\theta)} \mid \eta \right) \right).$$

Hence, a crucial issue when a non-normal likelihood is used for the QMLE is whether adequation tests confirm that the assumed distribution correctly fits the data.

5.1.3 The existence of distribution with given moments

The first issue in considering alternative distributions to the normal distribution is that not all values of skewness and kurtosis are attainable. This is closely related to conditions for the existence of moments, which was first investigated by Stiltjes (1894) for the case of a density with bounded support and Hamburger (1920) for the case of unbounded support. Widder (1946) reminds that, to ensure the existence of moments, the following sequence of inequalities, involving moments μ_j and their determinants must be satisfied

$$\mu_0 \geq 0, \quad \begin{vmatrix} \mu_0 & \mu_1 \\ \mu_1 & \mu_2 \end{vmatrix} \geq 0, \quad \begin{vmatrix} \mu_0 & \mu_1 & \mu_2 \\ \mu_1 & \mu_2 & \mu_3 \\ \mu_2 & \mu_3 & \mu_4 \end{vmatrix} \geq 0, \quad \dots$$

³ For a density $g(x)$, the natural location parameter μ and scale parameter σ are those that minimize

$$E \left[-\frac{1}{2} \log(\sigma^2) + \log \left(g \left(\frac{x_t - \mu}{\sigma} \right) \right) \right].$$

where $\mu_i = \int z^i f(z) dz$. By construction, we have $\mu_0 = 1$, because the *pdf* integrates to 1. For the case where z_t are standardized innovations, we have $\mu_1 = 0$ and $\mu_2 = 1$. This implies the following relation between skewness μ_3 and kurtosis μ_4

$$\mu_3^2 < \mu_4 - 1 \quad \text{with} \quad \mu_4 > 0. \quad (5.10)$$

This relation shows that, for a given level of kurtosis, only a finite range of skewness may be attained. The curve in Figure 5.2 delimits the skewness-kurtosis boundary corresponding to the domain (5.10). For all pairs of skewness and kurtosis between the upper and lower curves, a density exists.

Any proposed extension to the normal distribution with finite third and fourth moments will have a domain of definition inside the curves, and the exact coverage will depend on the given distribution and, more precisely, on the way asymmetry and fat tails are introduced in the distribution. We will illustrate this further in the next section using several alternative distributions proposed in the literature.

5.2 Distributions with higher moments

There are several ways of dealing with asymmetry or fat tails in a distribution. First, some distributions allow asymmetry or fat tails. For instance, the skewed Student *t* distribution or the Pearson IV distribution are directly

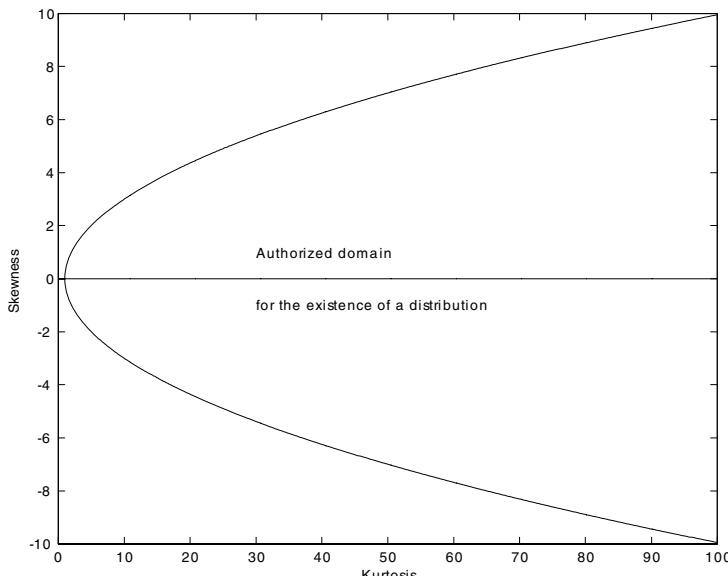


Fig. 5.2. General skewness-kurtosis boundary.

designed to incorporate such features. Second, asymmetry can also be introduced, using an expansion about a symmetric distribution. The main interest of the latter approach is that once the general framework has been developed, such a generalization can be applied to any symmetric distribution.

Engle and González-Rivera (1991) and Gallant, Hsieh, and Tauchen (1991) are among the first to address the issue of asymmetry and tail thickness in the conditional distribution; both attempt to extend from the GARCH type models. Engle and González-Rivera (1991) adopt a semi-parametric approach whereas Gallant, Hsieh, and Tauchen (1991) use a truncated Hermite polynomial expansion. Further work in this strand of the research includes Hansen (1994), Drost and Klaassen (1997), Harvey and Siddique (1999), Jondeau and Rockinger (2001, 2003a), Lambert and Laurent (2002), Premaratne and Bera (2000), and Rockinger and Jondeau (2002). Some of these approaches are presented in the following section.

Several alternative distributions have been proposed to capture asymmetry and fat tails. We do not present all possible extensions but rather a set of distributions, which have been found to have interesting properties. Other approaches may be found in Drost and Klaassen (1997) (who proposed an adaptive estimator), Brännäs and Nordman (2003) (log-generalized gamma distribution), or in Lambert and Laurent (2002) (skewed location-stable distribution).

5.2.1 Semi-parametric approach

In their so-called semi-parametric ARCH model, Engle and González-Rivera (1991) assume that first and second moments are given by a parametric ARMA process and a parametric ARCH model, respectively, while the conditional density is approximated by a non-parametric density estimator. The overall model they consider is given by (5.1)–(5.4). However the conditional distribution $g(z_t)$ is not assumed to be known but is estimated non-parametrically in a two-step procedure. The assumption is that z_t is *iid* with zero mean and unit variance. The log-likelihood function is defined by

$$\ell(\psi|x_t, t = 1, \dots, T) = \sum_{t=1}^T \ell_t(\psi), \quad (5.11)$$

where

$$\ell_t(\psi) = -\frac{1}{2} \log(\sigma_t^2(\theta)) + \log(g(z_t(\theta)|\eta)),$$

with $\psi = (\theta', \eta')$ the vector of unknown parameters.

To maximize the log-likelihood function (5.11), the following procedure is used:

- (i) An initial estimate of the set of parameters θ is given by $\tilde{\theta}$. It may be obtained by QML estimation of (5.2) and (5.4).

- (ii) The fitted residuals $\hat{\varepsilon}_t$ and the fitted variances $\hat{\sigma}_t^2(\tilde{\theta})$ are used to compute the standardized residuals $\hat{z}_t(\tilde{\theta}) = \hat{\varepsilon}_t / \hat{\sigma}_t(\tilde{\theta})$, which should have zero mean and unit variance.
- (iii) The density $g(\hat{z}_t(\theta))$ is estimated using a non-parametric method. The estimated density is denoted \hat{g} .
- (iv) The log-likelihood is computed using equation

$$L_T(\theta | x_t, t = 1, \dots, T) = \sum_{t=1}^T \ell_t(\theta),$$

where

$$\ell_t(\theta) = -\frac{1}{2} \log(\hat{\sigma}_t^2(\theta)) + \log(\hat{g}).$$

The log-likelihood function is maximized, with \hat{g} held fixed, iterating steps (ii)–(iv) until convergence.

- (v) Although the estimated density is assumed to be known when the log-likelihood is estimated, the score vector $\partial \ell_t(\theta_0) / \partial \theta'$ (that can be used for the estimation) is computed using the derivatives of $g(\hat{z}_t(\theta))$ with respect to θ . The score is therefore given by

$$\frac{\partial \ell_t(\theta_0)}{\partial \theta'} = -\frac{1}{2} \frac{1}{\sigma_t^2} \frac{\partial \sigma_t^2}{\partial \theta'} + \frac{1}{2} \left(\frac{\partial \varepsilon_t}{\partial \theta'} - \frac{1}{2} \frac{\partial \sigma_t^2}{\partial \theta'} \frac{\varepsilon_t}{\sigma_t^2} \right) \frac{\partial g(\hat{z}_t(\theta))}{\partial \theta'} \frac{1}{g(\hat{z}_t(\theta))}.$$

Engle and González-Rivera (1991) use the discrete maximum penalized likelihood estimation technique developed by Tapia and Thompson (1978) to estimate the non-parametric density. This technique works as follows. The density is approximated using the histogram of z_t with knots (n_1, \dots, n_{m-1}) and heights (p_1, \dots, p_{m-1}) over an interval (a, b) divided in m sub-intervals of length q . This approximation is illustrated in Figure 5.3.

For a sample (z_1, \dots, z_T) , the following optimization problem is solved

$$\begin{aligned} & \max_{(p_1, \dots, p_{m-1})} \sum_{t=1}^T \log(g(z_t)) - \frac{\lambda}{q} \sum_{k=1}^{m-1} (p_{k+1} - 2p_k + p_{k-1})^2, \\ & \text{subject to } q \sum_{k=1}^{m-1} p_k = 1 \quad p_k \geq 0 \quad \text{for } k = 1, \dots, m-1, \end{aligned}$$

where $p_0 = p_m = 0$ and

$$g(z) = \begin{cases} p_k + \frac{p_{k+1} - p_k}{q} (z - n_k) & \text{if } z \in [n_k; n_{k+1}), \\ 0 & \text{if } z \notin (n_0; n_m), \end{cases}$$

and λ is the penalty term to ensure smoothness of the estimate of the heights (p_1, \dots, p_{m-1}) .

The semi-parametric method avoids some problems of distribution misspecification, because using a non-normal distribution may lead to inconsistent parameter estimates if the distribution is incorrect. At the same time, the semi-parametric estimator may be more efficient (i.e., with smaller standard error) than a fully non-parametric method.

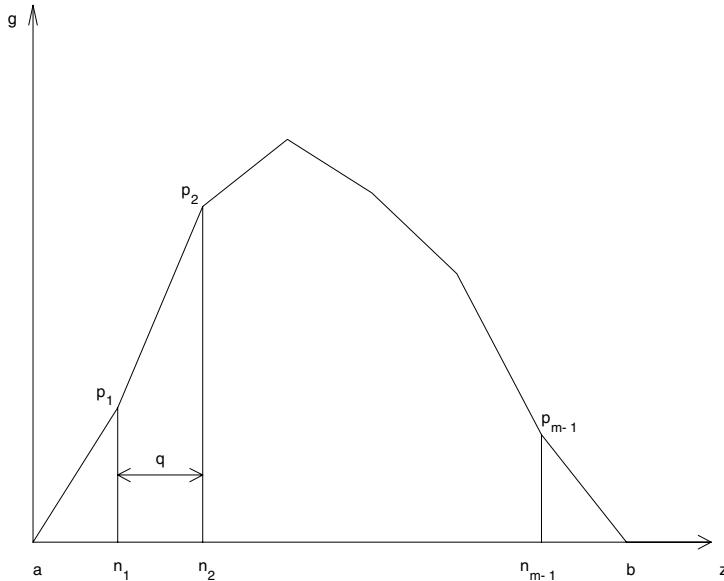


Fig. 5.3. Example of semi-parametric distribution.

5.2.2 Series expansion about the normal distribution

Early attempts to explicitly model departure from normality using series expansion around the normal distribution are by Gallant and Tauchen (1989), Gallant, Hsieh, and Tauchen (1991), and Lee and Tse (1991). Gallant, Hsieh, and Tauchen (1991) adopt a very general specification of the series expansion. Provided that large data sets are available, this semi-nonparametric approach may reveal a lot of information about the true distribution. However, such a parameterization is not parsimonious, and consequently estimation may be very computationally expensive. In addition, due to the number of parameters to be estimated, the characteristics of the distribution are virtually impossible to study.

A special case of a series expansion is the so-called Gram-Charlier expansion. Gallant and Tauchen (1989) and Lee and Tse (1991) use Gram-Charlier expansions to describe deviations from normality of innovations in a GARCH framework. Gram-Charlier expansions allow for additional flexibility over a normal distribution because they naturally introduce the skewness and kurtosis of the distribution as unknown parameters. However, being polynomial approximations, they have the drawback of yielding negative values for certain parameters. Moreover, there does not seem to be an easy and analytic characterization of those parameters for which the density will take positive values. Barton and Dennis (1952) establish conditions on the parameters guarantee-

ing positive definiteness of the underlying densities. Jondeau and Rockinger (2001) show that these conditions can be implemented numerically.

Distribution

When the true *pdf* of a random variable Z is unknown, yet believed to be rather close to a normal one, it is quite natural to approximate it with a *pdf* of the form

$$g(z|\eta) = \varphi(z)p_n(z|\eta), \quad (5.12)$$

where $\varphi(z)$ is the standard normal density with zero mean and unit variance and where $p_n(z|\eta)$ is chosen so that $g(z|\eta)$ has the same first moments as the *pdf* of Z . A widely used approximation of the true density $g(z|\eta)$ is based on the $(n+1)$ first Hermite polynomials, i.e.,

$$p_n(z|\eta) = \sum_{i=0}^n c_i He_i(z), \quad (5.13)$$

where $He_i(z)$ denotes the Hermite polynomial of order i . It is defined by $He_i(z) = (-1)^i \frac{\partial^i \varphi}{\partial z^i} \frac{1}{\varphi(z)}$.⁴

Two representations have been typically adopted in the literature, the Gram-Charlier type A expansion

$$p_4(z|\eta) = 1 + \frac{\gamma_1}{6} He_3(z) + \frac{\gamma_2}{24} He_4(z), \quad (5.14)$$

and the Edgeworth expansion

$$p_6(z|\eta) = 1 + \frac{\gamma_1}{6} He_3(z) + \frac{\gamma_2}{24} He_4(z) + \frac{\gamma_1^2}{72} He_6(z), \quad (5.15)$$

with $\eta = (\gamma_1, \gamma_2)'$.

The Edgeworth expansion (5.15) involves an additional Hermite polynomial while keeping the number of parameters constant. Therefore, the range for γ_1 and γ_2 over which positivity of the approximation is guaranteed is smaller than for the Gram-Charlier one. For this reason, we focus on the first approximation.

It can be shown that the two parameters γ_1 and γ_2 coincide with the skewness and excess kurtosis of $g(z|\eta)$, respectively. Indeed, since the approximating density is assumed to have zero mean and unit variance, we have (Johnson, Kotz, and Balakrishnan, 1994)

⁴ Straightforward computations yield the following expressions for the first six Hermite polynomials: $He_0(z) = 1$, $He_1(z) = z$, $He_2(z) = z^2 - 1$, $He_3(z) = z^3 - 3z$, $He_4(z) = z^4 - 6z^2 + 3$, $He_5(z) = z^5 - 10z^3 + 15z$, and $He_6(z) = z^6 - 15z^4 + 45z^2 - 15$.

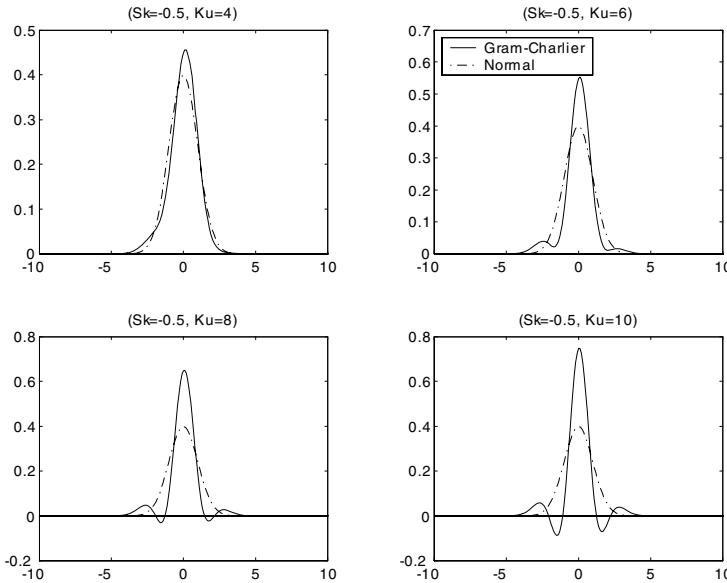


Fig. 5.4. Pdf of the Gram-Charlier distribution for various values of skewness and kurtosis (without positivity constraint).

$$\begin{aligned} \int_{-\infty}^{+\infty} z g(z|\eta) dz &= 0, & \int_{-\infty}^{+\infty} z^2 g(z|\eta) dz &= 1, \\ \int_{-\infty}^{+\infty} z^3 g(z|\eta) dz &= \gamma_1, & \text{and} & \\ \int_{-\infty}^{+\infty} z^4 g(z|\eta) dz &= 3 + \gamma_2. \end{aligned}$$

This property partly explains the success of Gram-Charlier expansions in the empirical literature, because the two additional parameters γ_1 and γ_2 are directly related to the third and fourth moments. However, Gram-Charlier expansions also have some drawbacks. A first shortcoming is that, for some (γ_1, γ_2) distant from the normal values $(0, 0)$, $g(z|\eta)$ can be negative for some z . For other pairs, the pdf $g(z|\eta)$ may be multimodal. These problems are illustrated in Figure 5.4. In addition, as will be shown below, the domain of definition for which the distribution is well defined turns out to be rather small.

Domain of definition

We focus now on implementing numerical conditions so that Gram-Charlier approximations are positive. To ensure positivity, Gallant and Tauchen (1989) suggest to square the polynomial part, $p_n(z|\eta)$, of (5.12). But by doing so, we lose the interpretation of the various parameters as moments of the density.

Jondeau and Rockinger (2001) provide some analytical results for computing the skewness-kurtosis boundary ensuring that the Gram-Charlier approximation is a density.

Some properties are useful to identify the region \mathcal{D} in the (γ_1, γ_2) -plane for which $g(z|\eta)$ is positive. For $g(z|\eta)$ to be positive, the polynomial $p_4(z|\eta)$ is required to be positive for every z , that is

$$1 + \frac{\gamma_1}{6} He_3(z) + \frac{\gamma_2}{24} He_4(z) \geq 0, \quad \forall z.$$

For a given value of z , the equation

$$p_4(z|\eta) = 1 + \frac{\gamma_1}{6} He_3(z) + \frac{\gamma_2}{24} He_4(z) = 0 \quad (5.16)$$

defines a straight line in the (γ_1, γ_2) -plane. A small deviation for z , while holding (γ_1, γ_2) fixed, will then yield a $p_4(z)$ of either positive or negative sign. Thus, we determine the set of (γ_1, γ_2) , as a function of z , such that $p_4(z|\eta)$ remains zero for small variations of z , because this set will define the requested boundary. This set is determined by the derivative of (5.16) with respect to z

$$\frac{\gamma_1}{2} He_2(z) + \frac{\gamma_2}{6} He_3(z) = 0. \quad (5.17)$$

The set of (γ_1, γ_2) that solves (5.16) and (5.17) simultaneously, also called the *envelope* of $p_4(z|\eta)$, yields a parametric representation of the boundary where $p_4(z|\eta)$ is zero for a given z . Once this boundary is determined, it remains to find that sub-region delimited by $p_4(z|\eta) = 0$ for all z .⁵

Solving the system given by (5.16) and (5.17) yields the expression for the skewness and the excess kurtosis as functions of z

$$\begin{aligned} \gamma_1(z) &= -24 \frac{He_3(z)}{d(z)}, \\ \gamma_2(z) &= 72 \frac{He_2(z)}{d(z)}, \end{aligned}$$

with $d(z) = 4(He_3(z))^2 - 3He_2(z)He_4(z)$. Straightforward computations allow us to rewrite the denominator of both expressions as $d(z) = z^6 - 3z^4 + 9z^2 + 9$. Since its minimum is attained for $z = 0$ where $d(0) = 9$, we conclude that $d(z)$ is always positive.

The sign of $\gamma_2(z)$ changes with $He_2(z) = z^2 - 1$. It is positive for $z \in (-\infty; -1] \cup [1; +\infty)$. It is negative for $z \in [-1; 1]$. Similarly, the sign of $s(z)$ changes with $He_3(z) = z^3 - 3z$. It is positive for $z \in (-\infty; -\sqrt{3}] \cup [0; \sqrt{3}]$ and negative elsewhere.

These properties are summarized in Figure 5.5. The authorized domain, ensuring the existence of the distribution, is clearly symmetric with respect

⁵ This approach has been highlighted by Barton and Dennis (1952) in a slightly different context.

to the horizontal axis. We observe that the excess kurtosis γ_2 is inside the interval $[0, 4]$. Indeed, we find that $\gamma_2(\pm\infty) = 0$ and $\gamma_2(\pm\sqrt{3}) = 4$. Since γ_2 is bounded below by 0, the kurtosis of $g(\cdot)$ will always be larger than for a normal distribution. Finally, we notice that the authorized domain for skewness and kurtosis is rather small. In many empirical applications, in particular at the daily frequency, the Gram-Charlier density would not be able to capture the skewness and kurtosis found in financial returns.

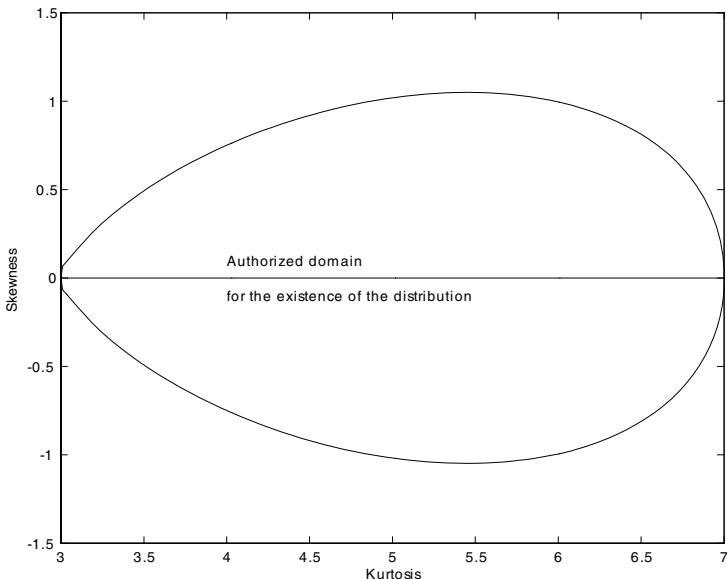


Fig. 5.5. Skewness-kurtosis boundary for positive Gram-Charlier distribution.

5.2.3 Skewed Student t distribution

The use of the Student t distribution to capture the fat tails of financial returns goes back to Bollerslev (1987) and Bollerslev and Wooldridge (1992). It is, however, a symmetric distribution, so that it cannot capture asymmetry. A generalization that fills this gap has been proposed by Hansen (1994), with the so-called skewed Student t distribution. He achieves this by introducing a generalization of the Student t distribution where asymmetries may occur, while maintaining the assumption of a zero mean and unit variance. He also illustrates how parameters, and subsequently higher moments, can be rendered time varying. Further extensions of this distribution are by Theodossiou (1998) and Jondeau and Rockinger (2003a, 2003b). A definite advantage of

the skewed Student t distribution is that multivariate extensions are available (See Section 6.2).

Distribution

In the following presentation, we directly define the distribution for the innovation Z , because this is the variable we are ultimately interested in. The skewed Student t distribution proposed by Hansen (1994) is defined by

$$g(z|\nu, \lambda) = b \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi(\nu-2)} \Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{\zeta^2}{\nu-2}\right)^{-\frac{\nu+1}{2}}, \quad (5.18)$$

where

$$\zeta = \begin{cases} (bz + a) / (1 - \lambda) & \text{if } z < -a/b, \\ (bz + a) / (1 + \lambda) & \text{if } z \geq -a/b. \end{cases}$$

The constant terms a and b are defined as

$$\begin{aligned} a &= 4\lambda c \frac{\nu-2}{\nu-1}, \\ b^2 &= 1 + 3\lambda^2 - a^2, \end{aligned}$$

and are introduced to obtain a variable Z with zero mean and unit variance,⁶ where we denote

$$c = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi(\nu-2)} \Gamma\left(\frac{\nu}{2}\right)}.$$

Here, ν denotes the degree-of-freedom parameter and λ the asymmetry parameter, so that the vector of shape parameters is $\eta = (\nu, \lambda)'$. The density is defined for $2 < \nu < \infty$ and $-1 < \lambda < 1$. Furthermore, it encompasses a large set of conventional densities. For instance, if $\lambda = 0$, Hansen's distribution is reduced to the traditional Student t distribution, without asymmetry. If, in addition, $\nu = \infty$, the Student t distribution collapses to the normal density.⁷

Figure 5.6 illustrates the *pdfs* that can be obtained with the skewed Student t distribution. Upper (lower) figures represent skewed t distributions with large negative (positive) asymmetry. Left-side (right-side) figures correspond to distributions with Gaussian (fat) tails.

⁶ As already mentioned, the mean of an asymmetric distribution is not the actual location parameter, a problem discussed by Newey and Steigerwald (1997).

⁷ As described in Section 5.2.4, Fernández and Steel (1998) also proposed a method to make any symmetric density asymmetric by some change of variable. The link between their parameter ξ (using the notation of Section 5.2.4) and the parameter λ of Hansen (1994) is simply given by

$$\lambda = \frac{\xi^2 - 1}{\xi^2 + 1}.$$

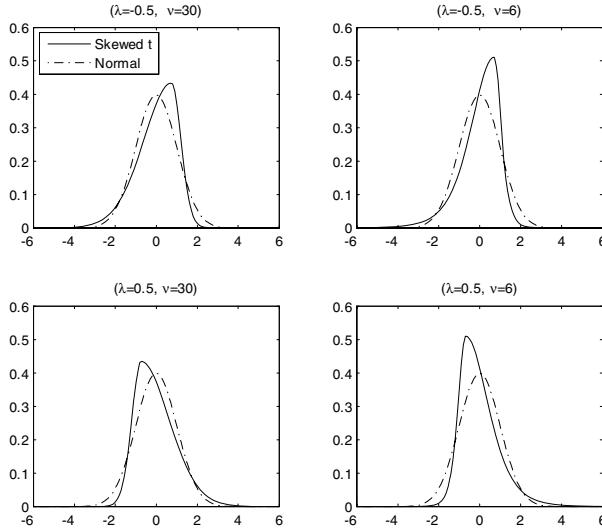


Fig. 5.6. Pdf of the skewed Student t distribution for various values of ν and λ .

It is well-known that a traditional Student t distribution with ν degrees of freedom will allow for the existence of up to ν th moments (see, for instance, Mood, Graybill, and Boes, 1974). Therefore, given the restriction $\nu > 2$, the skewed Student t distribution is well defined and its second moment exists. The higher moments can be computed as follows. Let \tilde{m}_r be the r th moment of the standard (symmetric) Student t distribution⁸

$$\tilde{m}_r = 2 \int_0^\infty x^r t(x|\nu) dx = \frac{\Gamma\left(\frac{\nu-r}{2}\right) \Gamma\left(\frac{r+1}{2}\right) (\nu-2)^{\frac{r+1}{2}}}{\sqrt{\pi(\nu-2)} \Gamma\left(\frac{\nu}{2}\right)}.$$

Then, the r th moment of the innovation $Z^* = bZ + a$ of the skewed Student t distribution can be computed as

$$M_r = E[(Z^*)^r] = \tilde{m}_r \left[(-1)^r (1-\lambda)^{r+1} + (1+\lambda)^{r+1} \right],$$

so that we have, from Jondeau and Rockinger (2003a)

⁸ This expression is obtained using the following result of Gradshteyn and Ryzhik (1994, p. 341, 3.241.4):

$$\int_0^\infty x^{\mu-1} (p + qx^\nu)^{-(n+1)} dx = \frac{1}{\nu p^{n+1}} \left(\frac{p}{q}\right)^{\frac{\mu}{\nu}} \frac{\Gamma\left(\frac{\mu}{\nu}\right) \Gamma\left(1+n-\frac{\mu}{\nu}\right)}{\Gamma(1+n)},$$

In particular, we have $\tilde{m}_1 = 2c\frac{\nu-2}{\nu-1}$, $\tilde{m}_2 = 1$, $\tilde{m}_3 = 4c\frac{(\nu-2)^2}{(\nu-1)(\nu-3)}$, and $\tilde{m}_4 = 3\frac{\nu-2}{\nu-4}$.

$$\begin{aligned} M_1 &= 4c\lambda \frac{\nu - 2}{\nu - 1} = a, \\ M_2 &= 1 + 3\lambda^2 = b^2 + a^2, \\ M_3 &= 16c\lambda(1 + \lambda^2) \frac{(\nu - 2)^2}{(\nu - 1)(\nu - 3)} \quad \text{if } \nu > 3, \\ M_4 &= 3 \frac{\nu - 2}{\nu - 4} (1 + 10\lambda^2 + 5\lambda^4) \quad \text{if } \nu > 4. \end{aligned}$$

We observe that, when $\lambda \neq 0$, Z^* does not have zero mean and unit variance anymore. This is because the transformation done to introduce asymmetry reallocates probability mass from one side of the distribution to the other. Moments of the standardized innovations, $Z = (Z^* - a)/b$, are then defined as $\mu^{(r)} = E[(Z)^r]$ with

$$E[Z] = \mu^{(1)} = 0, \tag{5.19}$$

$$V[Z] = \mu^{(2)} = 1, \tag{5.20}$$

and given that Z has zero mean and unit variance,

$$S[Z] = \mu^{(3)} = \frac{M_3 - 3aM_2 + 2a^3}{b^3}, \tag{5.21}$$

$$K[Z] = \mu^{(4)} = \frac{M_4 - 4aM_3 + 6a^2M_2 - 3a^4}{b^4}. \tag{5.22}$$

For some applications, it is very useful to have the *cdf* as well as its inverse. For instance, as will be shown in Chapter 8, VaR computation requires the expression of the inverse *cdf* of the skewed *t* distribution, because the VaR is based on the quantile of the distribution. It is shown in Jondeau and Rockinger (2003a) that the *cdf* is defined by

$$G(z) = \begin{cases} (1 - \lambda)T\left(\frac{bz+a}{1-\lambda}\sqrt{\frac{\nu}{\nu-2}}|\nu\right) & \text{if } z < -a/b, \\ (1 + \lambda)T\left(\frac{bz+a}{1+\lambda}\sqrt{\frac{\nu}{\nu-2}}|\nu\right) - \lambda & \text{if } z \geq -a/b, \end{cases}$$

where $T(x|\nu)$ is the *cdf* of the standard *t* distribution with ν degrees of freedom

$$T(x|\nu) = \int_{-\infty}^x \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\pi\nu}} \left(1 + \frac{w^2}{\nu}\right)^{-\frac{\nu+1}{2}} dw.$$

It is straightforward to show that the inverse of the *cdf* of G is written as

$$G^{-1}(y) = \begin{cases} \frac{1}{b} \left((1 - \lambda) \sqrt{\frac{\nu-2}{\nu}} T^{-1}\left(\frac{y}{1-\lambda}|\nu\right) - a \right) & \text{if } y < \frac{1-\lambda}{2}, \\ \frac{1}{b} \left((1 + \lambda) \sqrt{\frac{\nu-2}{\nu}} T^{-1}\left(\frac{y+\lambda}{1+\lambda}|\nu\right) - a \right) & \text{if } y \geq \frac{1-\lambda}{2}. \end{cases} \tag{5.23}$$

As is well-known, the inverse of the *cdf* allows the simulation of pseudo-random variates. It suffices to generate $y \sim U(0, 1)$ and to compute $z = G^{-1}(y)$. Then, z will be distributed as G .

Domain of definition

The density and the various moments do not exist for all parameters. As already mentioned, the density is defined only for $\nu > 2$ and $-1 < \lambda < 1$. Furthermore, careful scrutiny of the algebra yielding (5.21) shows that skewness exists if $\nu > 3$. Last, kurtosis in (5.22) is well defined if $\nu > 4$.

We define as \mathcal{D} the domain $(\nu, \lambda) \in [2, +\infty[\times] -1, 1[$. Given these restrictions on the underlying parameters, it is clear that the range of skewness and kurtosis will also be restricted to a certain domain. In Figure 5.7, we trace various curves relating skewness to λ varying between -1 and 1 for selected values of ν . Similarly, in Figure 5.8, we trace various curves relating kurtosis to ν varying between 4.1 and 8 for selected values of λ .

Focusing on Figure 5.7, we notice that, as ν decreases, skewness can attain very large values. On the other hand, as shown by Figure 5.8, when we increase ν , say around 8 , the tails of the density become thinner, even for extreme values of the λ parameter. This picture illustrates the fact that, for a given level of kurtosis, only a finite range of skewness exists. This feature raises the question of existence of a density for given moments.

Figure 5.9 displays the skewness-kurtosis boundary ensuring the existence of a density. The curve ABC corresponds to the theoretical domain of maximal size given by inequality (5.10). The curve DEF corresponds to the domain of skewness and kurtosis, which is attainable with a skewed Student t distribution, assuming $\nu > 2$. We notice that the kurtosis is bounded from below by

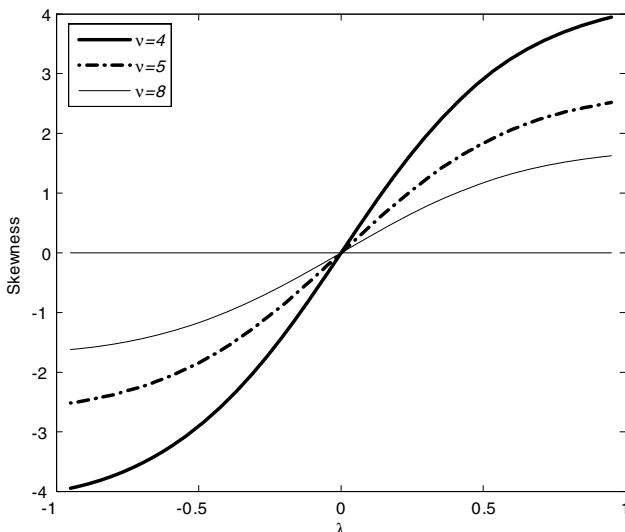


Fig. 5.7. Skewness for various values of ν .

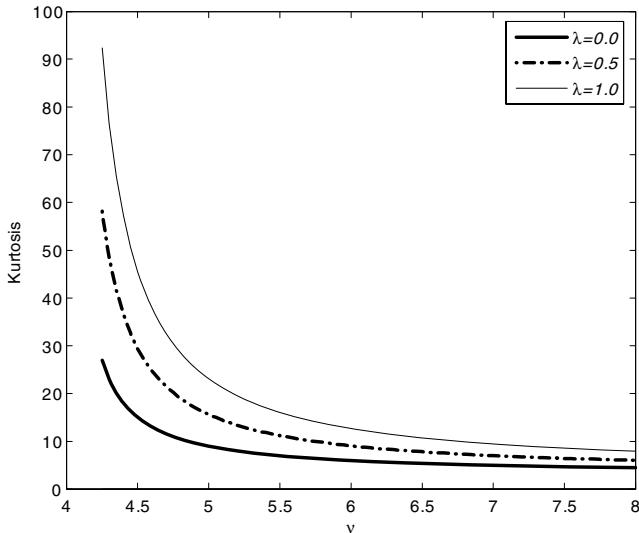


Fig. 5.8. Kurtosis for various values of λ .

3, indicating that the skewed Student t distribution does not allow tails to be thinner than those of the normal distribution.

The domain denoted \mathcal{E} corresponding to DEF in Figure 5.9 is spanned by skewness and kurtosis if both moments exist. We notice that the relation between \mathcal{D} and \mathcal{E} is not one-to-one. In particular, those points that are located in \mathcal{D} but where $\nu < 4$ have no counterpart in \mathcal{E} . The logic is that there are points in \mathcal{D} where skewness or kurtosis cease to exist, whereas in \mathcal{E} skewness and kurtosis are finite by construction. It is only when we reduce the domain \mathcal{D} to $]4, +\infty[\times]-1, 1[$ that the relation is bijective.

Alternative specifications

After the initial description of the skewed Student t distribution by Hansen (1994), several alternative specifications for an asymmetric Student t distributions have been proposed. Some of these specifications can be viewed as particular cases of the general approaches presented in the next section to generate skewness from a symmetric distribution.

Fernández and Steel (1998) have specialized their general approach to generate asymmetric distributions (based on hidden truncation) to the case of the Student t distribution

$$g(z|\nu, \xi) = \frac{2}{\xi + \frac{1}{\xi}} \left[t(z\xi|\nu) 1_{(-\infty, 0)}(z) + t\left(\frac{z}{\xi}|\nu\right) 1_{[0, \infty)}(z) \right].$$

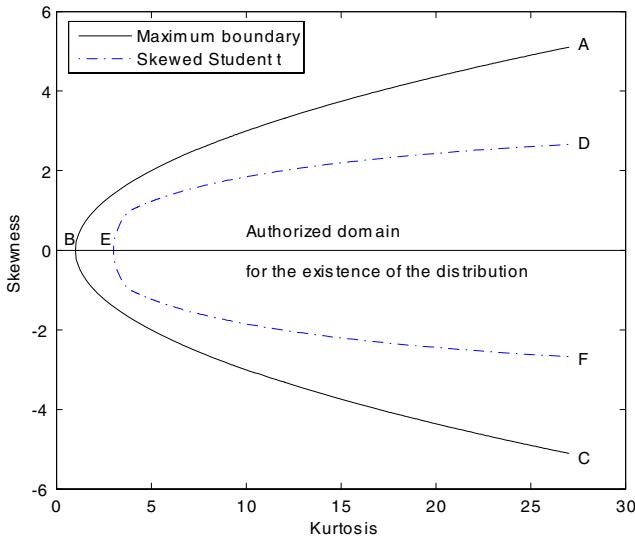


Fig. 5.9. Skewness-kurtosis boundary for skewed t distribution.

It can be easily seen that this distribution is directly related to the one proposed by Hansen (1994) through a mere change of notation of the asymmetry parameter. Lambert and Laurent (2002) have also extended this approach to the case of the skewed stable distribution and, more generally, to the case of the skewed location-scale distributions. Notice that the general approach of generating asymmetric distribution will be used in Section 6.2 to define multivariate skewed distributions.

Azzalini and Capitanio (2003) have proposed another approach to generate some asymmetry in (multivariate) distributions. In the case of the univariate Student t distribution, they propose the use of

$$g(z|\nu, \lambda) = 2t(z|\nu) T\left(\lambda z \sqrt{\frac{\nu+1}{z^2+\nu}} |\nu+1\right), \quad (5.24)$$

where $T(z|\nu)$ is the *cdf* of the Student t distribution with ν degrees of freedom. When $\lambda = 0$, $g(z)$ reduces to the standard Student t distribution. When $\lambda \rightarrow -\infty (\infty)$, we obtain a Student t distribution truncated from above (below) at zero. Parameter λ plays therefore the role of shape parameter.

Jones and Faddy (2003) have constructed the following distribution

$$g(z|\alpha, \beta) = \frac{1}{2^{\alpha+\beta-1} B(\alpha, \beta) \sqrt{\alpha+\beta}} \left(1 + \frac{z}{\sqrt{\alpha+\beta+z^2}}\right)^{\alpha+\frac{1}{2}} \\ \times \left(1 - \frac{z}{\sqrt{\alpha+\beta+z^2}}\right)^{\beta+\frac{1}{2}},$$

where $B(\alpha, \beta)$ denotes the beta function. For $\alpha = \beta$, the distribution reduces to the standard Student t distribution. When $\alpha < \beta$ (resp. $\alpha > \beta$), g is negatively (resp. positively) skewed.

Another alternative specification has been proposed by Harvey and Sidique (1999). Their so-called non-central Student t distribution allows one to model skewness but does not allow for an independent variation of skewness and kurtosis. The non-central Student t distribution is defined by two parameters: ν the degree-of-freedom parameter and δ the non-centrality parameter. This distribution is scaled to have a unit variance, while the non-centrality parameter controls the shape of the distribution. The non-central Student t distribution is defined as

$$g(z|\nu, \delta) = \frac{\nu^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)} \times \frac{\exp\left(-\frac{\delta^2}{2}\right)}{\sqrt{\pi} (\nu + z^2)^{\frac{\nu+1}{2}}} \times \sum_{i=0}^{\infty} \Gamma\left(\frac{\nu+i+1}{2}\right) \left(\frac{\delta^i}{i!}\right) \left(\frac{2z^2}{\nu+z^2}\right)^{\frac{i}{2}}.$$

Since one of the parameters is needed to control the centrality of the distribution, it is clear that this density is more restrictive as far as skewness and kurtosis are concerned.

5.2.4 Generating asymmetric distributions

Several ways of generating skewness in a symmetric distribution have been proposed. Without analyzing these various approaches in detail, we provide in this section a brief description of the two most well-known methods. See Ferreira and Steel (2004) for a recent review of some approaches developed to generate asymmetry in univariate distributions.

Hidden truncation

The first approach, based on conditioning (or hidden truncation), was introduced by Azzalini and Dalla Valle (1996) and Azzalini and Capitanio (1999) in a multivariate context (it will be described more deeply in Section 6.2).⁹ In the case of a univariate distribution, we have

$$g(z|\xi) = 2f(z)F(\xi z),$$

where f denotes a *pdf* and F its *cdf*. When $\xi = 0$, $g(z)$ reduces to the symmetric distribution f . When $\xi \rightarrow -\infty (\infty)$, we obtain a Student t distribution

⁹ See also Arnold and Beaver (2000).

truncated from above (below) at zero. The parameter ξ plays therefore the role of shape (skewness) parameter.

A difficulty with this approach is that moments are hard to compute for general distributions f . In the case of the Student t distribution, we obtain (5.24). Then, as shown by Azzalini and Capitanio (2003), the first moments are

$$\begin{aligned} E[Z] &= \mu = \delta \left(\frac{\nu}{\pi} \right)^{1/2} \frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})} \quad (\nu > 1), \\ V[Z] &= \sigma^2 = \frac{\nu}{\nu-2} - \mu^2 \quad (\nu > 2), \\ \tilde{S}[Z] &= \mu \left(\frac{\nu(3-\delta^2)}{\nu-3} - \frac{3\nu}{\nu-2} + 2\mu^2 \right) \quad (\nu > 3), \\ \tilde{K}[Z] &= \frac{3\nu^2}{(\nu-2)(\nu-4)} - \frac{4\mu^2\nu(3-\delta^2)}{\nu-3} + \frac{6\mu^2\nu}{\nu-2} - 3\mu^4 - \frac{3}{\sigma^4} \quad (\nu > 4), \end{aligned}$$

where $\delta = \xi/\sqrt{1+\xi^2}$.

Standardized skewness and kurtosis are defined in the usual way by the expressions $S[Z] = \tilde{S}[Z]/\sigma^3$ and $K[Z] = \tilde{K}[Z]/\sigma^4$. Note that we have $E[Z] = S[Z] \gtrless 0$ when $\xi \gtrless 0$. Moreover $E[Z|\xi] = -E[Z|-\xi]$ and $S[Z|\xi] = -S[Z|-\xi]$. This shows that the parameter ξ directly controls the asymmetry of the distribution.

The expressions above clearly indicate that Z does not have zero mean and unit variance anymore. The reason is that the transformation done to introduce asymmetry reallocates some probability mass from one side of the distribution to the other. More precisely, non-standardized innovations Z have mean a and variance b^2 . If we now define the standardized innovation $Z^* = (Z-a)/b$, we obtain $E[Z^*] = 0$, $V[Z^*] = 1$, $S[Z^*] = S[Z]$ and $K[Z^*] = K[Z]$. The new innovation Z^* has the same skewness and kurtosis as Z .

Inverse scale factor

For Bayesian estimation purpose, Fernández and Steel (1998) proposed another method for introducing skewness into any continuous unimodal distribution symmetric around the origin. The method involves changing the scale of the distribution at each side of the mode. The new asymmetric random variable Z is then distributed as¹⁰

¹⁰ The distribution $g(z|\xi)$ may be viewed as a distribution of the mixture (see Lambert and Laurent, 2002) below

$$z = u\xi|x| - (1-u)\frac{1}{\xi}|x|,$$

where u is the realization of a Bernoulli process with probability of success $\xi^2/(1+\xi^2)$, and x is the realization of a process with distribution f .

$$\begin{aligned} g(z|\xi) &= \frac{2}{\xi + \frac{1}{\xi}} f\left(z\xi^{-sign(z)}\right) \\ &= \frac{2}{\xi + \frac{1}{\xi}} \left[f(z\xi) 1_{(-\infty, 0)}(z) + f\left(\frac{z}{\xi}\right) 1_{[0, \infty)}(z) \right], \end{aligned}$$

where f denotes any symmetric *pdf*. The asymmetry parameter $\xi > 0$ is such that the ratio of probability masses above and below the mode is

$$\frac{\Pr[Z \geq 0]}{\Pr[Z < 0]} = \xi^2. \quad (5.25)$$

Hence, parameter ξ provides an interesting indicator of the asymmetry in the distribution. The distribution $g(z|1/\xi)$ is the symmetric of $g(z|\xi)$ with respect to the mode, with $g(z|\xi) = g(-z|1/\xi)$. Therefore, inverting ξ produces the mirror image around zero.

Moments of the asymmetric distribution $g(z|\xi)$ are easily deduced from those of the symmetric one $f(z)$. If the r th moment of a r.v. with distribution $f(\cdot)$ exists, then the associated random variable Z with distribution $g(\cdot)$ also has a finite r th moment, defined as

$$M_r = \tilde{m}_r \frac{\xi^{r+1} + \frac{(-1)^r}{\xi^{r+1}}}{\xi + \frac{1}{\xi}},$$

where

$$\tilde{m}_r = 2E[Z^r | Z > 0] = 2 \int_0^\infty u^r f(u) du$$

is the r th moment of $f(\cdot)$ truncated to the positive real values. Provided that they exist, the first four moments are then obtained as

$$\begin{aligned} E[Z] &= M_1 = \tilde{m}_1 \left(\xi - \frac{1}{\xi} \right) \equiv a, \\ V[Z] &= M_2 - M_1^2 = (\tilde{m}_2 - \tilde{m}_1^2) \left(\xi^2 + \frac{1}{\xi^2} \right) + 2\tilde{m}_1^2 - \tilde{m}_2 \equiv b^2, \\ \tilde{S}[Z] &= M_3 - 3M_1M_2 + 2M_1^3 \\ &= \left(\xi - \frac{1}{\xi} \right) \left[(\tilde{m}_3 + 2\tilde{m}_1^3 - 3\tilde{m}_1\tilde{m}_2) \left(\xi^2 + \frac{1}{\xi^2} \right) + 3\tilde{m}_1\tilde{m}_2 - 4\tilde{m}_1^3 \right], \\ \tilde{K}[Z] &= M_4 - 4M_1M_3 + 6M_2M_1^2 - 3M_1^4. \end{aligned}$$

Note that we have $S[Z|\xi] = -S[Z|1/\xi]$ and $S[Z] = 0$ when $\xi = 1$.

As before, Z does not have zero mean and unit variance anymore, but can be standardized to yield innovations $Z^* = (Z - a)/b$, with $E[Z^*] = 0$, $V[Z^*] = 1$, $S[Z^*] = S[Z]$, and $K[Z^*] = K[Z]$.

5.2.5 Pearson IV distribution

Premaratne and Bera (2000) and Brännäs and Nordman (2001) have proposed the Pearson type IV distribution to model the dynamic of asset returns. The main interest of this distribution is that the three parameters can be interpreted as the variance, skewness, and kurtosis of the process, so that it is possible to directly model these moments if required.

Distribution

This distribution is a member of the Pearson family, which is characterized as the distributions $g(z^* | \eta)$ satisfying

$$\frac{d \log(g(z^* | \eta))}{dz} = \frac{z^* - \beta}{b_0 + b_1 z^* + b_2 z^{*2}},$$

where β , b_0 , b_1 , and b_2 are parameters to be estimated. Inside this group, assuming some specific values of the parameters, we obtain the normal, gamma, Student t , or Pareto distributions. In the general case where $b_1 \neq 0$ and $b_2 \neq 0$, and where the roots of $b_0 + b_1 z^* + b_2 z^{*2}$ are imaginary, say $b + ia$ and $b - ia$, the resulting distribution is given as

$$g(z^* | \eta) = c^{-1} \left(1 + \left(\frac{z^*}{a} \right)^2 \right)^{-m} \exp \left(\delta \arctan \left(\frac{z^*}{a} \right) \right), \quad (5.26)$$

where

$$\begin{aligned} m &= -1/(2b_2), \\ \delta &= (b - \beta) / (ab_2), \end{aligned}$$

and

$$c = a \int_{-\pi/2}^{\pi/2} \cos^r(\omega) \exp(-\delta\omega) d\omega$$

is the integration constant that ensures that the density integrates to one. We also define $r = 2(m - 1)$. The vector of shape parameters is thus $\eta = (a, r, \delta)'$. This distribution is known as the Pearson IV distribution. It is of particular interest because it allows for both asymmetry and fat tails. It has been little used in empirical studies, however, mainly because its estimation raises some technical problems.

Domain of definition

We note that Z^* does not have zero mean and unit variance, because we have the following moments (see Kendall and Stuart, 1977, Premaratne and Bera, 2000)

$$\begin{aligned}
E[Z^*] &= \frac{\delta a}{r} \quad (r > 0), \\
V[Z^*] &= \frac{a^2}{r^2(r-1)} (r^2 + \delta^2) \quad (r > 1), \\
\tilde{S}[Z^*] &= \frac{4a^3\delta(r^2 + \delta^2)}{r^3(r-1)(r-2)} \quad (r > 2), \\
\tilde{K}[Z^*] &= 3 \frac{a^4(r^2 + \delta^2)[(r+6)(r^2 + \delta^2) - 8r^2]}{r^4(r-1)(r-2)(r-3)} \quad (r > 3),
\end{aligned}$$

with standardized skewness and kurtosis

$$\begin{aligned}
S[Z^*] &= s = \frac{4\delta}{r-2} \sqrt{\frac{r-1}{r^2 + \delta^2}}, \\
K[Z^*] &= \kappa = 3 \frac{(r-1)[(r+6)(r^2 + \delta^2) - 8r^2]}{(r-2)(r-3)(r^2 + \delta^2)}.
\end{aligned}$$

Parameter δ can be viewed as an asymmetry parameter, because it controls the sign of skewness. When $\delta = 0$, with $r > 2$, we obtain $S[Z^*] = s = 0$, implying symmetry of the distribution. In addition, $\delta > 0$ (< 0) entails a positive (negative) skewness. Parameter r can be viewed as a degree-of-freedom parameter, which controls for fat tails. To see this, note that we have under symmetry ($\delta = 0$)

$$\kappa = 3 \frac{r-1}{r-3},$$

so that lower values of r imply thicker tails. We obtain normality when $r \rightarrow \infty$. However, δ and r cannot be directly interpreted as a skewness and kurtosis parameter, because skewness as well as kurtosis depend on both δ and r .

Figure 5.10 displays the domain of definition of the Pearson type IV distribution. We also report the boundary for skewness and kurtosis ensuring the existence of the distribution $\kappa - s^2 - 1 = 0$ already described (see Section 5.1.3).

Estimation issue

Let us consider now the issue of estimating this model in a time-series context. It should be noted that, since the innovation Z_t^* does not have zero mean, the conditional mean of the return process $X_t = \mu_t(\theta) + \sigma_t(\theta) Z_t^*$ is given by

$$E[X_t | \mathcal{F}_{t-1}] = \mu_t(\theta) + \sigma_t(\theta) \frac{\delta a}{r}.$$

For symmetric unimodal distributions, mean, median, and mode coincide, and therefore correct specification of the conditional mean is sufficient to ensure consistent estimates of the conditional mean parameters. For asymmetric distributions, the mean and variance may not be its natural location and scale

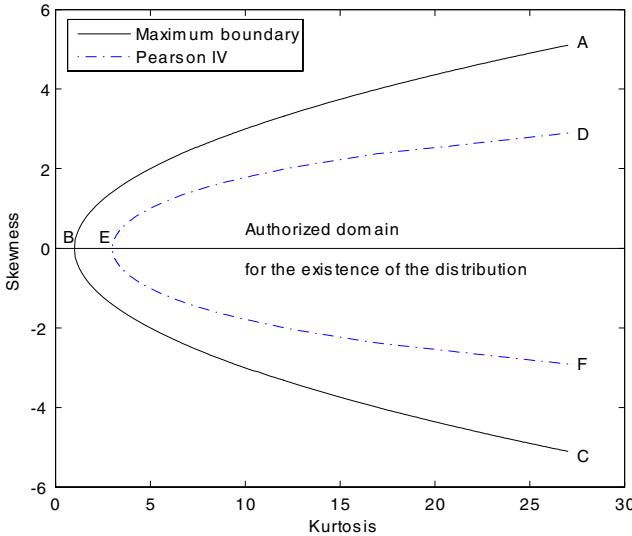


Fig. 5.10. Skewness-kurtosis boundary for the Pearson IV distribution.

parameters (Newey and Steigerwald, 1997). In this case, we have to introduce an additional parameter $\tilde{\mu}$ in the density of the innovation process to capture the fact that the mean may not be the natural location parameter. The density function (5.26) therefore becomes

$$g(z_t^* | \eta, \tilde{\mu}) = c^{-1} \left(1 + \left(\frac{z_t^* - \tilde{\mu}}{a} \right)^2 \right)^{-m} \exp \left(\delta \arctan \left(\frac{z_t^* - \tilde{\mu}}{a} \right) \right).$$

In this case, we have

$$E[Z_t^*] = \tilde{\mu} + \frac{\delta a}{r}.$$

Consequently, the conditional mean of the return process X_t becomes

$$E[X_t | \mathcal{F}_{t-1}] = \mu_t(\theta) + \sigma_t(\theta) \left(\tilde{\mu} + \frac{\delta a}{r} \right).$$

With this specification, as shown by Newey and Steigerwald (1997), the consistency of the QML estimator is ensured. Finally, the log-likelihood at time t to be maximized is given by

$$\ell_t(\psi) = -\frac{1}{2} \log (\sigma_t^2(\theta)) + \log \left(g \left(\frac{x_t - \mu_t(\theta) - \sigma_t(\theta) \tilde{\mu}}{\sigma_t(\theta)} | \eta \right) \right),$$

with $\psi = (\theta', \eta', \tilde{\mu})'$ the vector of unknown parameters.

5.2.6 Entropy distribution

Methods based on the entropy principle of Shannon (1948), and popularized by Jaynes (1957, 1982), have made their way into econometrics, e.g., Golan, Judge, and Miller (1996). At a practical level, entropy-based applications still appear to be scarce but for a few exceptions such as Zellner and Highfield (1988), Stutzer (1996), Buchen and Kelly (1996), or Ormoneit and White (1999). The difficulties with the numerical implementation of this technique may have hindered its widespread use. Yet, Rockinger and Jondeau (2002), based on Agmon, Alhassid, and Levine (1979a, 1979b), have shown that it is possible to develop a very rapid method to obtain entropy densities and that entropy densities may also be used in rather complex empirical likelihood estimations. It is worth emphasizing that this approach allows one to estimate distributions with the largest domain of definition possible compatible with the bound found in Figure 5.2.

Distribution

We assume that the econometrician is seeking a probability $p(z)$ defined over some real convex domain, \mathcal{D} , while disposing only of information on the m first moments of the probability, written as μ_i where $i = 1, \dots, m$. The construction of a probability density defined on infinitely many points with the knowledge of only a few moments is hopeless without an additional criterion. A first possibility to obtain a density, matching the given moments, is to use *ad hoc* step functions. Such an approach is implemented by Wheeler and Gordon (1969). Another criterion is given by the maximization of an entropy under the moment and density restrictions. Under this criterion, one solves

$$p \in \arg \max_{\{p\}} - \int_{z \in \mathcal{D}} p(z) \log(p(z)) dz, \quad (5.27)$$

$$\text{s.t.} \quad \int_{z \in \mathcal{D}} p(z) dz = 1, \quad (5.28)$$

$$\int_{z \in \mathcal{D}} z^i p(z) dz = \mu_i, \quad i = 1, \dots, m. \quad (5.29)$$

We refer to a density satisfying these conditions as an Entropy Density.¹¹ Jaynes (1957) notices that the entropy is a criterion where the statistician imposes a minimum amount of information. The conventional way of solving this program is to define the Hamiltonian

$$\begin{aligned} H = & - \int_{\mathcal{D}} p(z) \log(p(z)) dz - \lambda_0^* \left(\int_{\mathcal{D}} p(z) dz - 1 \right) \\ & - \sum_{i=1}^m \lambda_i \left(\int_{\mathcal{D}} z^i p(z) dz - \mu_i \right), \end{aligned}$$

¹¹ Given that a log-function is involved in (5.27), $p(z) \geq 0, \forall z \in \mathcal{D}$.

where the λ_0^* is a Lagrange parameter as are the λ_i , $i = 1, \dots, m$.

To obtain a solution of this problem, we seek a zero for the Fréchet derivative. Defining $\lambda'_0 = \lambda_0^* + 1$, we get

$$\delta H = 0 \quad \Rightarrow \quad p(z) = \exp \left(-\lambda'_0 - \sum_{i=1}^m \lambda_i z^i \right). \quad (5.30)$$

Derivation with respect to the $m + 1$ Lagrange multipliers yields the $m + 1$ conditions (5.28) to (5.29).

For small values of m , it is possible to obtain explicit solutions. If $m = 0$, meaning that no information is given, beyond the fact that one seeks a density, then we obtain the uniform distribution over \mathcal{D} . As we add the first and second moments, Golan, Judge, and Miller (1996) recall that we obtain the exponential and the normal density. The knowledge of the third or higher moment does not yield a density in closed form. Only numerical solutions may provide densities. We focus now on how densities may be obtained in a numerically efficient manner if third and higher moments are given. Some results can be found in Zellner and Highfield (1988) and Ormoneit and White (1999), while Rockinger and Jondeau (2002) describe a more efficient estimation technique.

Estimation

Substitution of (5.30) into (5.28) defines a function that, as shown later, turns out to be a *potential function*. The expression of this function is

$$P(\lambda_1, \dots, \lambda_m) \equiv \exp(-\lambda'_0) = \int_{\mathcal{D}} \exp \left(\sum_{i=1}^m \lambda_i z^i \right) dz, \quad (5.31)$$

so that

$$p(z) = \exp \left(\sum_{i=1}^m \lambda_i z^i \right) / P(\lambda_1, \dots, \lambda_m). \quad (5.32)$$

For a given set of $\lambda = (\lambda_1, \dots, \lambda_m)'$, one could evaluate (5.32) and, thus, the moment restrictions (5.29). This result suggests as a first estimation technique non-linear least squares applied to (5.29). Such an estimation yields multiple solutions and is rather slow.¹² As found by Agmon, Alhassid, and Levine (1979a, 1979b), a faster and numerically stable procedure is available. This procedure uses the physical properties of the entropy definition. In order to use this procedure, it is convenient to introduce some further results.

Since $\int_{\mathcal{D}} p(z) dz = 1$, multiplication of the right-hand side of (5.29) by this integral and the grouping under one single integral yields

¹² The technique developed by Ormoneit and White (1999) follows this approach. They show how such an algorithm may be implemented more efficiently as in Zellner and Highfield (1988), yet, they report estimations lasting several seconds whereas the ones of Rockinger and Jondeau (2002) take a fraction of a second.

$$\int_{\mathcal{D}} (z^i - \mu_i) p(z) dz = 0, \quad i = 1, \dots, m.$$

Furthermore, writing $p(z) = \exp(\lambda_0 + \sum_{i=1}^m \lambda_i(z^i - \mu_i))$, where $\lambda_0 = \lambda'_0 + \sum_{i=1}^m \lambda_i \mu_i$ indicates that the number of computations required to evaluate (5.32) subject to (5.29) may be reduced. Also, the passage from λ'_0 to λ_0 is a trivial linear transformation. Again, $p(z)$ must satisfy (5.29), and this yields a definition for λ_0

$$Q(\lambda_1, \dots, \lambda_m) \equiv \exp(-\lambda_0) = \int_{\mathcal{D}} \exp\left(\sum_{i=1}^m \lambda_i(z^i - \mu_i)\right) dz, \quad (5.33)$$

so that the probability can be rewritten as

$$p(z) = \exp\left(\sum_{i=1}^m \lambda_i(z^i - \mu_i)\right) / Q(\lambda_1, \dots, \lambda_m). \quad (5.34)$$

At this point, we have obtained two equivalent definitions for the density, namely (5.32) and (5.34). Depending on the situation, one definition or the other is useful. With the definition of (5.33), we obtain that

$$g_i \equiv \frac{\partial Q}{\partial \lambda_i} = 0 \Rightarrow \int_{\mathcal{D}} (z^i - \mu_i) p(z) dz = 0,$$

and, therefore, the zeros of the gradient of Q yield the first-order conditions. This computation validates the claim that Q defines a potential.¹³ Next, we obtain that

$$G_{ij} \equiv \frac{\partial^2 Q}{\partial \lambda_i \partial \lambda_j} = \int_{\mathcal{D}} (z^i - \mu_i)(z^j - \mu_j) p(z) dz,$$

showing that the Hessian matrix is a variance-covariance matrix. As a consequence, the Hessian matrix is symmetric and positive definite. An inverse of the Hessian will exist if the matrix is of full rank. This last condition implies that, as long as $p(z)$ is a density, the minimization of Q has a unique solution. We write the gradient of Q as g and its Hessian matrix as G .

At this stage, we have obtained the first key result, namely that the minimization of the potential function Q will yield a solution. This solution, in turn, will define an entropy distribution. We insist on the fact that the key step to obtain a solution resides in a minimization rather than in a search for a zero of a map. It turns out that, numerically, the minimization is well defined, whereas the search for a zero may even yield multiple solutions. The problem is numerically stable if Q is of full rank and if the solution is finite.

¹³ If U is an open subset of \mathbb{R}^n , a map f from U into \mathbb{R}^n is called a *vector field*. For instance, if F is a scalar function from U into \mathbb{R} , then $f = \text{grad } F$ defines a vector field. If for a given vector field f there exists a scalar function F such that $f = \text{grad } F$, then F is a *potential function* and the vector field f is said to derive from a potential.

As Agmon, Alhassid, and Levine (1979a) point out, it is not guaranteed that the minimization of the potential function will occur at finite distance. It is possible to guarantee finiteness of the solution, but to do so it is first necessary to define how to compute the integrals involved. We turn to this issue now, and we show how the existence of a finite solution can be guaranteed. The discretization of the potential (5.33) and the constraints (5.28) and (5.29) can be obtained by a Gauss-Legendre approximation (see, for instance, Davis and Polonsky, 1970) that yields

$$Q(\lambda) = \sum_{j=1}^n \exp \left(\sum_{i=1}^m \lambda_i (x_j^i - \mu_i) \right) w_j, \quad (5.35)$$

$$\sum_{j=1}^n w_j p_j = 1, \quad (5.36)$$

$$\sum_{j=1}^n w_j z_j^i p_j = \mu_i, \quad i = 1, \dots, m, \quad (5.37)$$

$$p_j \geq 0, \quad j = 1, \dots, n, \quad (5.38)$$

where $w = (w_1, \dots, w_n)'$ is a vector of weights where the integral is evaluated. Those values are tabulated, for instance, in Abramowitz and Stegun (1970). To guarantee that the Hessian is of full rank, given the way the (x_j, w_j) are obtained, it is necessary to have $2n > m$. Under this condition, even if the problem is symmetrical (for instance, because the mean and skewness is 0), the Hessian will be well defined.

This set of equations (5.36)–(5.38) can be viewed as a linear programming problem where we seek a solution to $m + 1$ equations under positivity constraints. If a solution exists, the algorithm will find it within $m+1$ and $2(m+1)$ steps. Now, if a solution exists, then it is known that Q will be minimized for some finite solution. The problem thus consists in numerically minimizing $Q(\lambda)$. As pointed out by Fletcher (1994), many algorithms are available. If the problem is known to have a single minimum, Newton's method works well. We may start with an initial value $\lambda^{(0)} = (0, \dots, 0)$ the vector with m zeros. At step k , we update the vector as

$$\lambda^{(k)} = \lambda^{(k-1)} + \delta^{(k)},$$

where $\delta^{(k)}$ is solution to $G^{(k)}\delta^{(k)} = -g^{(k)}$. This latter condition guarantees that the k th approximation in a second-order Taylor expansion of Q , that is $Q(\lambda^{(k-1)} + \delta^{(k)}) = Q(\lambda^{(k-1)}) + \delta^{(k)} g^{(k)} + \frac{1}{2} \delta^{(k)'} G^{(k)} \delta^{(k)}$, leads to a flat spot of Q , which is an extremum.

Domain of definition

We focus now, without loss of generality, on the study of the densities of a standardized variable Z that satisfy $\mu_1 = 0$, $\mu_2 = 1$, $\mu_3 = s$, and $\mu_4 = \kappa$. In

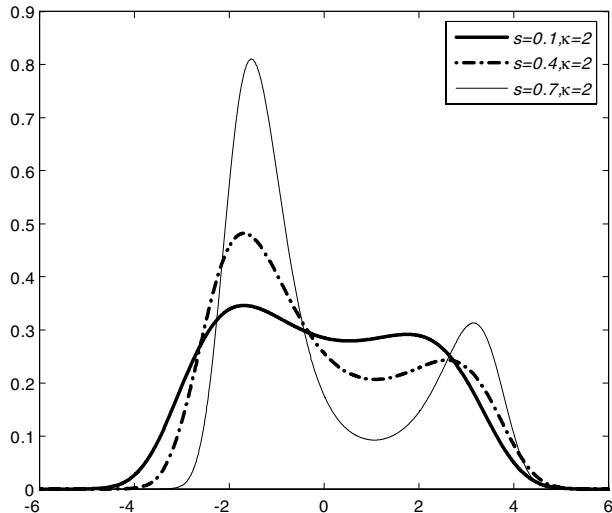


Fig. 5.11. Entropy distributions for various levels of skewness.

In this case, the parameters s and κ represent skewness and kurtosis, respectively. A solution to the minimization of the potential (5.33) subject to constraints (5.36)–(5.38) is guaranteed only if parameters s and κ are in the domain of definition \mathcal{D} . We may determine numerically the domain of definition for the entropy distribution. This may be done, for instance, using a grid-search over a large skewness-kurtosis domain where a solution to the simplex algorithm might exist. Doing so, we verify numerically that the boundary for the entropy distribution in fact corresponds to the maximum boundary $\mu_3^2 < \mu_4 - 1$.

Next, we consider how the entropy density behaves as skewness, s , and kurtosis, κ , vary. An inspection of Figures 5.11 and 5.12 reveals a rich pattern of possible densities. For densities with small kurtosis, the probability mass is squeezed towards the center. Introduction of skewness then leads to multi-modal densities. For densities with large kurtosis and skewness, given the assumed finiteness of the boundary, a small hump in the tail of the distribution will accommodate the skewness. We obtain that entropy densities may be of use when the tails of the distributions are much thinner than the tails of the normal density. Inversely, κ may become very large allowing for fat tails.

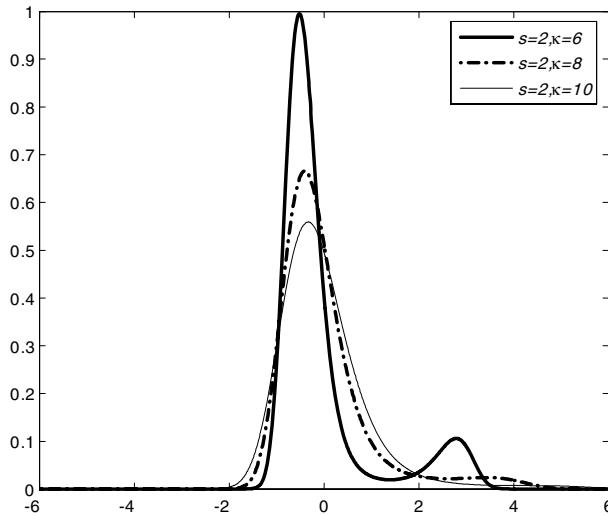


Fig. 5.12. Entropy distributions for various levels of kurtosis.

5.3 Specification tests and inference

5.3.1 Moment specification tests

The use of distributions allowing higher moments is primarily intended to capture some relevant statistical features of the data. Therefore, some conditions between theoretical and sample moments should hold. These tests imply non-linear restrictions on the parameter vector and can be viewed as mis-specification tests.¹⁴ They are related to the test already presented in Chapter 3 to test the mixture of distributions hypothesis.

To briefly describe the moment specification tests, let θ denote the $(n, 1)$ vector of parameters and $r(\theta)$ denote the $(J, 1)$ restriction functions (or moment conditions) (examples of such restrictions are given below). The null hypothesis is $H_0 : r(\theta) = 0$. The Wald test statistic is defined by $W = r(\hat{\theta})' \hat{\Omega}^{-1} r(\hat{\theta})$, where $\hat{\Omega}$ is an estimate of the (J, J) covariance matrix Ω of $r(\hat{\theta})$. Under the null hypothesis, W is asymptotically distributed as a $\chi^2(J)$.

Let $m_i(z_t, \theta)$ be the i th moment condition for the t th observation. Then, the i th restriction is defined as $r_i(\theta) = \frac{1}{T} \sum_{t=1}^T m_i(z_t, \theta)$. The empirical coun-

¹⁴ See Newey (1985), Nelson (1991), Harvey and Siddique (1999), and Brooks, Burke, and Persand (2005).

terpart is simply $r_i(\hat{\theta}) = \frac{1}{T} \sum_{t=1}^T \hat{m}_i(z_t, \hat{\theta})$, where $\hat{m}_i(z_t, \hat{\theta})$ is the estimated moment. We note $\hat{m}_j = [\hat{m}_j(z_1, \hat{\theta}) \cdots \hat{m}_j(z_T, \hat{\theta})]'$.

The covariance matrix Ω may be estimated using the derivatives of the log-likelihood function

$$\hat{\Omega} = \frac{1}{T^2} (M'M - M'D(D'D)^{-1}D'M),$$

with $M = (\hat{m}_{t,j})_{t,j}$ the (T, J) matrix containing the empirical moment conditions

$$M = [\hat{m}_1 \cdots \hat{m}_J] = \begin{bmatrix} \hat{m}_1(z_1, \hat{\theta}) & \cdots & \hat{m}_J(z_1, \hat{\theta}) \\ \vdots & \ddots & \vdots \\ \hat{m}_1(z_T, \hat{\theta}) & \cdots & \hat{m}_J(z_T, \hat{\theta}) \end{bmatrix}.$$

Each element $\hat{m}_{t,j}$ is not required to be identically zero, but the j th moment restriction states that the sum of the $\hat{m}_{t,j}$ across observations should be zero. Finally, the (T, n) matrix of derivatives D is defined as

$$D = [\hat{d}_1 \cdots \hat{d}_n] = \begin{bmatrix} \hat{d}_{1,1} & \cdots & \hat{d}_{1,n} \\ \vdots & \ddots & \vdots \\ \hat{d}_{T,1} & \cdots & \hat{d}_{T,n} \end{bmatrix},$$

where the (t, j) th element $\hat{d}_{t,j}$ is the derivative of the log-likelihood at date t with respect to the j th parameter θ_j

$$\hat{d}_{t,j} = \hat{d}_{t,j}(\hat{\theta}) = \left. \frac{\partial \ell_t(\theta)}{\partial \theta_j} \right|_{\theta=\hat{\theta}}.$$

The only remaining issue now is to determine the appropriate form of the standardized residual that should be selected as the basis for the test. Under the correct model specification, standardized residuals should be *iid* with the postulated distribution as conditional distribution. Since we consider standardized residuals, they should have zero mean and unit variance, resulting in the following orthogonality conditions for the first and the second moments

$$\begin{aligned} E[z_t] &= 0, \\ E[z_t^2 - 1] &= 0. \end{aligned}$$

Additional conditions depend on the postulated distribution, and, in particular, on whether asymmetry and fat tails are allowed. For example, a conditional normal distribution would imply that skewness and excess kurtosis are zero

$$\begin{aligned} E[z_t^3] &= 0, \\ E[z_t^4 - 3] &= 0, \end{aligned}$$

while assuming a conditional Student t distribution would imply

$$\begin{aligned} E[z_t^3] &= 0, \\ E\left[z_t^4 - 3\frac{\nu - 2}{\nu - 4}\right] &= 0. \end{aligned}$$

Additional orthogonality conditions can be imposed on the serial correlation of a given moments or on even higher moments. For instance, in the normal case, we have

$$\begin{aligned} E[z_t z_{t-j}] &= 0 \quad j = 1, 2, \dots, \\ E[(z_t^2 - 1)(z_{t-j}^2 - 1)] &= 0, \\ E[z_t^3 z_{t-j}^3] &= 0, \\ E[(z_t^4 - 3)(z_{t-j}^4 - 3)] &= 0. \end{aligned}$$

These moment conditions translate into the restrictions that the first four powers of the series should not be serially correlated.

5.3.2 Adequacy tests based on density forecasts

As pointed out by Engle and González-Rivera (1991), the QMLE has a degree of inefficiency that increases with the degree of departure from normality. Maximizing the log-likelihood using the correct distribution of Z_t is therefore likely to improve significantly the efficiency of the estimator. Unfortunately, if we use an incorrect non-normal specification, this estimator is not consistent (Newey and Steigerwald, 1997). Therefore, it appears crucial to check the validity of the distributional assumption by carrying out adequacy tests. Some of these tests are presented below.

A first natural way to test the adequacy of the estimated distribution to the data is to measure the distance between the empirical (unknown) distribution and the assumed (estimated) distribution. This type of test is similar to the Kolmogorov-Smirnov test widely used to test normality (see Section 2.2.3).

Let $f_t(z_t)$ denote the empirical (unknown) *pdf*, which we try to describe as accurately as possible using the parametric distribution $g_t(z_t)$. For instance, f_t may be the actual distribution of the SP500, and g_t may be one of the distributions described in Section 5.2. Now, define the probability integral transform as $u_t = \int_{-\infty}^{z_t} g_t(y_t) dy_t$ (Rosenblatt, 1952). We may view u_t as the value taken by the *cdf* at z_t . Then, the crucial result of this approach is the following.

Proposition 5.1. (Diebold, Gunther, and Tay, 1998). *Suppose $\{z_t\}_{t=1}^T$ is generated from the distribution $\{f_t(z_t|I_t)\}_{t=1}^T$ where $I_t = \{z_{t-1}, z_{t-2}, \dots\}$. If a sequence of density forecasts $\{g_t(z_t)\}_{t=1}^T$ coincides with $\{f_t(z_t|I_t)\}_{t=1}^T$,*

then under the usual condition of a non-zero Jacobian with continuous partial derivatives, the sequence of probability integral transforms of $\{z_t\}_{t=1}^T$ with respect to $\{g_t(z_t)\}_{t=1}^T$ is iid $U(0, 1)$, that is

$$\{u_t\}_{t=1}^T \sim \text{iid } U(0, 1).$$

In other words, if the postulated distribution $g_t(z_t)$ is correct, then its probability integral transform u_t is iid $U(0, 1)$. Diebold, Gunther, and Tay (1998) propose a two-step test. First, we test the null hypothesis that u_t is serially uncorrelated. It can be done using a standard LM test. For this purpose, Diebold, Gunther, and Tay suggest examining the autocorrelations of $(u_t - \bar{u})^i$, for $i = 1, \dots, 4$, by regressing $(u_t - \bar{u})^i$ on K lags of the variable.¹⁵ The LM test statistic is defined as $(T - K) R^2$, where R^2 is the coefficient of determination, and is distributed, under the null of no serial correlation, as a χ^2 with K degrees of freedom.

Second, we test the null hypothesis that u_t is distributed as a uniform $U(0, 1)$. For this purpose, Diebold, Gunther, and Tay divide the empirical and theoretical distribution into N cells and test whether the two distributions significantly differ on each cell. Such an approach allows a graphical representation that can be used to identify areas where the theoretical distribution fails to fit the data. The test statistic

$$DGT(N) = \sum_{n=1}^N \frac{(F_n - T/N)^2}{T/N},$$

where F_n is the number of observations in the cell n and T/N is the expected number of observations under the null. Under the null of correct specification, the test statistic $DGT(N)$ is asymptotically distributed as a $\chi^2(N - 1)$. Vlaar and Palm (1993) note that when the test statistic depends on p estimated parameters, the asymptotic distribution of $DGT(N)$ is in fact bounded between a $\chi^2(N - 1)$ and a $\chi^2(N - p - 1)$.

5.3.3 Adequacy tests based on interval forecasts

A very promising area of research for adequacy tests is the “hit” test proposed by Christoffersen (1998). He generalizes the approach based on density forecasts by focusing on conditional interval forecast. The idea is to test if the probability of being in a given interval, based on the given distribution, is compatible with the actual data.

Define the indicator variable Hit_t for a given interval forecast $(L_{t/t-1}(p), U_{t/t-1}(p))$ at time t as

$$Hit_t = \begin{cases} 1 & \text{if } z_t \in [L_{t/t-1}(p), U_{t/t-1}(p)], \\ 0 & \text{if } z_t \notin [L_{t/t-1}(p), U_{t/t-1}(p)], \end{cases}$$

¹⁵ Zero correlation is equivalent to independence only under normality. The correlogram is, therefore, only suggestive of possible independence.

where $L_{t/t-1}(p)$ and $U_{t/t-1}(p)$ denote, respectively, for the lower and upper limits of the interval.

In an unconditional setup, a natural goodness-of-fit test for a given interval forecast consists in comparing the nominal coverage $\sum_{t=1}^T \text{Hit}_t/T$ with the true coverage p . This is the test proposed by Diebold, Gunther, and Tay (1998). In a conditional setup, Christoffersen (1998) shows that testing

$$E[\text{Hit}_t | \mathcal{F}_{t-1}] = E[\text{Hit}_t | \text{Hit}_{t-1}, \text{Hit}_{t-2}, \dots] = p,$$

for all t is equivalent to testing if the sequence $\{\text{Hit}_t\}_{t=1}^T$ is *iid* Bernoulli(p). The test for conditional coverage is, thus, performed in two steps. First, we test the *unconditional coverage*, i.e., the null hypothesis that $E[\text{Hit}_t] = p$. It can be done using the likelihood-ratio test statistic

$$LR_{unc} = -2 \log \left(\frac{L(p|\text{Hit}_t, t = 1, \dots, T)}{L(\hat{\pi}|\text{Hit}_t, t = 1, \dots, T)} \right),$$

where

$$\begin{aligned} L(p|\text{Hit}_t, t = 1, \dots, T) &= (1-p)^{n_0} p^{n_1}, \\ L(\hat{\pi}|\text{Hit}_t, t = 1, \dots, T) &= (1-\hat{\pi})^{n_0} \hat{\pi}^{n_1}, \end{aligned}$$

with $\hat{\pi} = n_1 / (n_0 + n_1)$ is the MLE of π the true coverage probability. Clearly, n_0 is the number of outcomes in the interval and $n_1 = T - n_0$. Under the null hypothesis, the test statistic LR_{unc} is asymptotically distributed as a $\chi^2(1)$.

Next, we test the *independence of the Hits*. Under the alternative, the indicator variable Hit_t is assumed to be a binary first-order Markov chain, with transition probability matrix

$$\Pi_1 = \begin{bmatrix} 1 - \pi_{01} & \pi_{01} \\ 1 - \pi_{11} & \pi_{11} \end{bmatrix},$$

where $\pi_{ij} = \Pr[\text{Hit}_t = j | \text{Hit}_{t-1} = i]$. Therefore, the likelihood function for this process is

$$L(\Pi_1 | \text{Hit}_t, t = 1, \dots, T) = (1 - \pi_{01})^{n_{00}} (\pi_{01})^{n_{01}} (1 - \pi_{11})^{n_{10}} (\pi_{11})^{n_{11}},$$

where n_{ij} is the number of observations with value i followed by value j . The probability π_{ij} is estimated by $\hat{\pi}_{ij} = n_{ij} / (n_{i0} + n_{i1})$. Under the null of independence, the transition probability matrix reduces to

$$\Pi_2 = \begin{bmatrix} 1 - \pi_2 & \pi_2 \\ 1 - \pi_2 & \pi_2 \end{bmatrix},$$

with likelihood function

$$L(\Pi_2 | \text{Hit}_t, t = 1, \dots, T) = (1 - \pi_2)^{n_{00} + n_{10}} (\pi_2)^{n_{01} + n_{11}}.$$

The probability π_2 is estimated by $\hat{\pi}_2 = (n_{01} + n_{11})/T$. Then, the test statistic is

$$LR_{ind} = -2 \log \left(\frac{L(\hat{\Pi}_2 | Hit_t, t = 1, \dots, T)}{L(\hat{\Pi}_1 | Hit_t, t = 1, \dots, T)} \right).$$

Under the null hypothesis, the test statistic LR_{ind} is asymptotically distributed as a $\chi^2(1)$.

Finally, the joint test of coverage and independence, which corresponds to the *test of conditional coverage*, is given by the test statistic

$$LR_{cc} = -2 \log \left(\frac{L(p | Hit_t, t = 1, \dots, T)}{L(\hat{\Pi}_1 | Hit_t, t = 1, \dots, T)} \right).$$

which is asymptotically distributed as a $\chi^2(2)$ under the null hypothesis.

This adequacy test has been extended in a number of directions. The most interesting extension is probably the modeling of the probability of being in a given interval. Such an expansion of the information set has been proposed by Christoffersen (1998). The idea is to incorporate some additional variables in the information set such that $\mathcal{F}_{t-1} = \{X_{t-1}, \dots, X_1\}$, where X_t is a vector of observed variables. Then, if we run the regression

$$Hit_t = \alpha + \beta' f(X_{t-1}) + \varepsilon_t,$$

with f a function of past variables X_{t-1} , the test of interval forecast efficiency with respect to the information set \mathcal{F}_{t-1} is a joint test of independence ($\beta = 0$) and correct unconditional coverage ($\alpha = p$). Since the error term ε_t can be shown to be *iid*, standard OLS technique can be applied to this regression. Such a regression-based approach was also followed by Clements (2002) and Wallis (2003). Engle and Manganelli (2004) adopt a similar approach for evaluating the VaR of a portfolio in the context of quantile regressions (with their CAViaR model).

5.4 Illustration

To illustrate some aspects of the modeling of non-normal distributions, we consider the dynamic of the same four stock market indices as in previous sections, namely the SP500, the DAX, the FT-SE, and the Nikkei indices. We use daily returns over the period from January 1980 to August 2004. We remove the day corresponding to the October 1987 crash. Although it would not affect markedly the estimation of the GARCH models, it would have important consequences for the test of adequacy of the assumed distribution to the empirical distribution.

In order to provide a complete diagnosis on the series at hand, we estimate a model that is designed to capture the serial correlation and heteroskedasticity found in the daily returns. In addition, we look for a conditional distribution that is able to adjust to the empirical distribution of returns. We begin with the following AR(p)–GARCH(1, 1) model with conditionally normal innovations. The number of lags in the AR(p) process is selected in such a way that residuals ε_t can be viewed as serially uncorrelated. Consequently, the optimal number of lags is likely to depend on the return series. Table 5.1 reports the parameter estimates of the AR(p) process. It also presents the LM test for ARCH effects. We notice that the autoregressive parameters for daily returns are not very large but that some of them are strongly significant. Therefore, there is some amount of information in past returns to predict future returns. This autocorrelation may be induced by assets of different liquidity incorporating news at different speed. It should be noticed, however, that the main source of dependency in returns does not come from the correlation between returns but from the correlation between squared returns. As highlighted in the LM test, squared residuals are extremely correlated, indicating that the variability of asset returns changes over time.

Table 5.2 reports the parameter estimates of the GARCH(1, 1) process for residuals. We notice a large persistence in volatility as shown by the sum

Table 5.1. Parameter estimates of the AR(p) process for daily returns

	SP500	DAX	FT-SE	Nikkei
μ	0.043 (0.0130)	0.037 (0.0171)	0.035 (0.0118)	0.012 (0.0165)
ρ_1	0.017 (0.0127)	0.008 (0.0127)	0.061 (0.0127)	0.010 (0.0128)
ρ_2	-0.016 (0.0127)	-0.024 (0.0127)	0.010 (0.0127)	-0.066 (0.0128)
ρ_3	-0.051 (0.0127)	-0.014 (0.0127)	-0.037 (0.0127)	- (0.0127)
ρ_4	- (0.0127)	0.020 (0.0127)	0.042 (0.0127)	- (0.0127)
ρ_5	- (0.0127)	-0.008 (0.0127)	-0.019 (0.0127)	- (0.0127)
ρ_6	- (0.0127)	-0.051 (0.0127)	-0.023 (0.0127)	- (0.0127)
ρ_7	- (0.0127)	0.018 (0.0127)	0.000 (0.0127)	- (0.0127)
ρ_8	- (0.0127)	- (0.0127)	0.048 (0.0127)	- (0.0127)
LM test	1710.20	1123.75	1107.65	1125.22
$(T \times R^2)$	(0.000)	(0.000)	(0.000)	(0.000)

$\alpha_1 + \beta_1$.¹⁶ The table also provides results of the adequacy test proposed by Diebold, Gunther, and Tay (1998). In this case, the proposed model is rejected for two reasons. First, the margin u_t is not *iid*. This suggests that the AR(p)–GARCH(1, 1) is not able to capture all the time dependency of returns. Second, the margin is not uniformly distributed. This indicates that the normal distribution is not able to capture the features of the empirical distribution, in particular its asymmetry and fat-tailedness. To get a better identification of the source of rejection of the normal distribution, we compare, for various parts of the distribution of standardized residuals, the expected number of observations under the normal hypothesis and the observed number of observations under the empirical distribution. In Figure 5.13, we plot, for 50 cells of equal probability, the expected number of observations under normality (the medium horizontal line with 125 observations) as well as its confidence interval. We also plot, with diamonds, the observed number of observations, that corresponds to the empirical distribution. The figure suggests that the rejection of the normal distribution for SP500 daily returns comes from the extremes. Crashes are much more frequent and booms are less frequent in the empirical distribution than in the assumed normal distribution.

Now, we assume that innovations are not normally but t distributed, $z_t \sim t_\nu$. Thus we also estimate the degree-of-freedom parameter ν . Table 5.3 reports the corresponding parameter estimates. The degree of freedom lies within the range [6, 12], suggesting that the tails of the empirical distribution are fatter than those implied by the normal distribution. Yet, the fat-tailedness of the empirical distribution is not as extreme as the one we would have had with a stable distribution. Indeed, because our estimates of ν are clearly larger than 4, we can admit that, at least, the first four moments of the distribution do exist.¹⁷ The adequacy test reveals that the t distribution captures most features of the empirical distribution for the SP500 and the DAX indices, as illustrated in Figure 5.14. However, this distribution still fails to adequately represent the distribution of the FT-SE and Nikkei indices, because it implies too many positive extremes to be consistent with the empirical distribution.

Finally, we assume that innovations are distributed as a skewed t distribution: $z_t \sim g(\nu, \lambda)$. We, therefore, estimate the degree-of-freedom parameter ν and the asymmetry parameter λ . Table 5.4 reports the estimation of the GARCH(1, 1) model with skewed t innovation. All asymmetry parameters are negative, although it is barely significant for the SP500. For other indices, we obtain a pronounced leftward asymmetry. The adequacy test reveals that the skewed t distribution captures most features of the empirical distribution of the FT-SE and the Nikkei. Figure 5.15 shows that the fit of the empirical distribution of the SP500 by the skewed t distribution is rather good.

¹⁶ Parameter estimates slightly differ from those reported in Chapter 4 for the same series, because we now take the serial correlation in return into account.

¹⁷ An interesting property of the t distribution with ν degrees of freedom is that all moments of order above ν do not exist.

Table 5.2. Parameter estimates of the GARCH(1, 1) process under normality

	SP500	DAX	FT-SE	Nikkei
ω	0.007 (0.0014)	0.030 (0.0041)	0.022 (0.0032)	0.013 (0.0021)
α	0.048 (0.0043)	0.099 (0.0081)	0.099 (0.0085)	0.114 (0.0079)
β	0.945 (0.0047)	0.885 (0.0088)	0.876 (0.0100)	0.886 (0.0070)
$\log(L_{norm})$	-8318.10	-9543.70	-7646.58	-9126.98
Time independency				
$(u_t - \bar{u})$	7.959 (0.6329)	11.213 (0.3412)	25.567 (0.0044)	37.148 (0.0001)
$(u_t - \bar{u})^2$	38.857 (0.0000)	39.416 (0.0000)	20.656 (0.0236)	13.661 (0.1890)
$(u_t - \bar{u})^3$	24.419 (0.0066)	27.216 (0.0024)	34.567 (0.0001)	47.055 (0.0000)
$(u_t - \bar{u})^4$	22.805 (0.0115)	27.371 (0.0023)	12.443 (0.2565)	13.231 (0.2111)
Uniform(0,1)	167.758 (0.000)	86.166 (0.001)	88.887 (0.000)	161.102 (0.000)

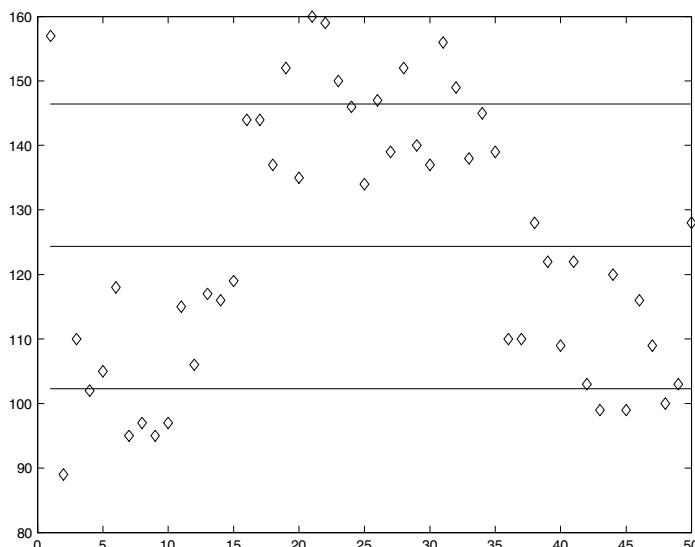
**Fig. 5.13.** Goodness-of-fit test to the $\mathcal{N}(0, 1)$ for SP500 daily returns.

Table 5.3. Parameter estimates of the GARCH(1,1) process under t innovations

	SP500	DAX	FT-SE	Nikkei
ω	0.005 (0.0015)	0.015 (0.0034)	0.015 (0.0030)	0.008 (0.0021)
α	0.044 (0.0055)	0.083 (0.0085)	0.082 (0.0085)	0.092 (0.0093)
β	0.952 (0.0059)	0.909 (0.0090)	0.899 (0.0104)	0.908 (0.0087)
ν	6.707 (0.5508)	9.808 (0.9687)	12.079 (1.4013)	6.677 (0.5256)
$\log(L_{stud})$	-8161.40	-9391.84	-7539.09	-8956.95
Time independency				
$(u_t - \bar{u})$	7.034 (0.7223)	10.624 (0.3875)	25.270 (0.0049)	34.984 (0.0001)
$(u_t - \bar{u})^2$	40.848 (0.0000)	38.271 (0.0000)	18.747 (0.0436)	16.639 (0.0827)
$(u_t - \bar{u})^3$	22.235 (0.0140)	26.082 (0.0036)	34.346 (0.0002)	46.900 (0.0000)
$(u_t - \bar{u})^4$	28.221 (0.0017)	28.634 (0.0014)	13.482 (0.1979)	22.549 (0.0125)
Uniform(0,1)	60.217 (0.131)	57.974 (0.178)	89.595 (0.000)	68.668 (0.033)

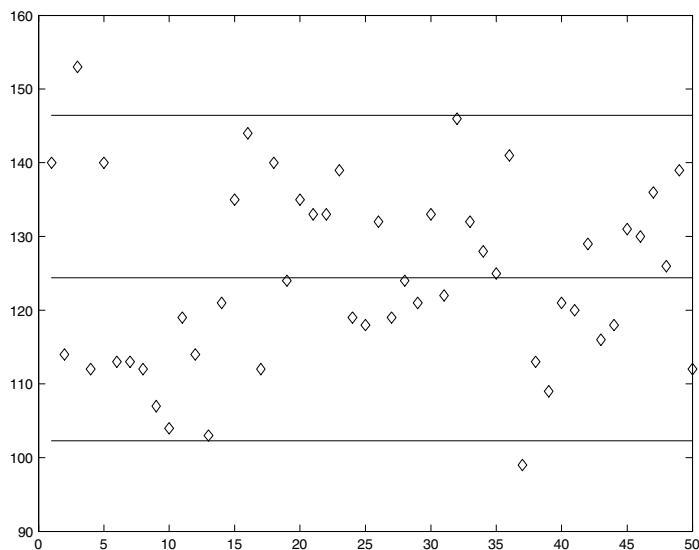
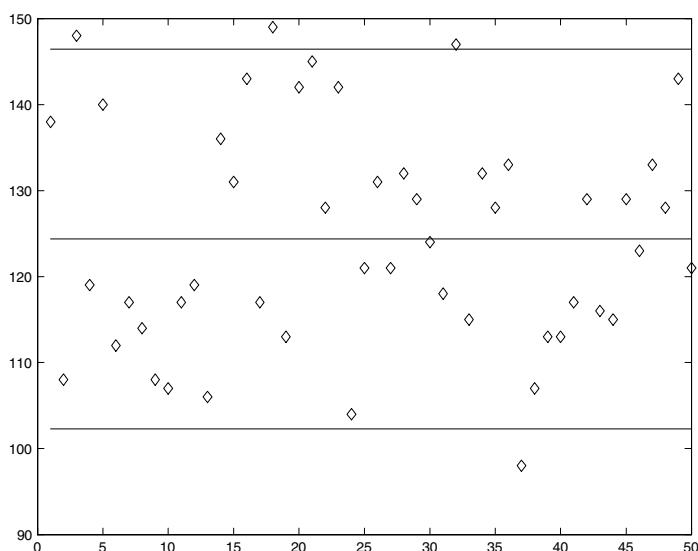
**Fig. 5.14.** Goodness-of-fit test to the t_ν for SP500 daily returns.

Table 5.4. Parameter estimates of the GARCH(1, 1) process with a skewed t

	SP500	DAX	FT-SE	Nikkei
ω	0.005 (0.0015)	0.015 (0.0033)	0.015 (0.0030)	0.009 (0.0022)
α	0.045 (0.0056)	0.083 (0.0085)	0.082 (0.0084)	0.094 (0.0093)
β	0.952 (0.0060)	0.910 (0.0090)	0.899 (0.0102)	0.906 (0.0087)
ν	6.764 (0.5636)	9.994 (1.0046)	12.583 (1.5042)	6.629 (0.5183)
λ	-0.027 (0.0149)	-0.084 (0.0189)	-0.115 (0.0189)	-0.084 (0.0174)
$\log(L_{skstud})$	-8160.07	-9381.96	-7520.76	-8945.10
Time independency				
$(u_t - \bar{u})$	7.073 (0.7186)	10.148 (0.4276)	24.773 (0.0058)	32.945 (0.0003)
$(u_t - \bar{u})^2$	41.303 (0.0000)	38.789 (0.0000)	18.952 (0.0409)	13.910 (0.1771)
$(u_t - \bar{u})^3$	21.679 (0.0168)	23.177 (0.0101)	30.186 (0.0008)	42.200 (0.0000)
$(u_t - \bar{u})^4$	28.038 (0.0018)	28.565 (0.0015)	13.580 (0.1931)	15.837 (0.1044)
Uniform(0,1)	66.520 (0.048)	49.938 (0.436)	47.594 (0.530)	34.671 (0.939)

**Fig. 5.15.** Goodness-of-fit test to the skewed t for SP500 daily returns.

To conclude, we observe that the GARCH model is unable to capture all the serial correlation in second and higher moments. Two alternative approaches may improve this disappointing result: (i) estimate an asymmetric GARCH model (such as the GJR or TARCH models), which would probably help capture at least the serial correlation in the second moment; (ii) model the dynamic of the third and fourth moments, as initially proposed in Hansen (1994) and extended in Jondeau and Rockinger (2003a) (See Section 5.5).

5.5 Modeling conditional higher moments

It is possible and sometimes necessary to go one step further and generalize the model with non-normal distribution. For a number of applications, allowing higher moments to vary over time can be very useful. The idea behind this approach is simply that the asymmetry and the fat-tailedness of returns may change over time. For asset allocation, the changing characteristics of the return distribution have to be taken into account by the investors in their decision process. In a non-normal environment, we generally recognize that investors like positive skewness and dislike kurtosis (see Chapter 9). In this context, when higher moments are time-varying, it is important to forecast their future path in order to improve the allocation of wealth. Another instance is the VaR computation of a portfolio with non-normal assets. If the distribution of asset returns varies over time, it is obviously crucial to update the measure of VaR when the asymmetry and the fat-tailedness of the distribution change.

Now, we assume that the characteristics of the shape of the distribution vary over time. Therefore, the parameter vector η is itself allowed to vary over time. Hence, we generalize model (5.1)–(5.5) as follows

$$x_t = \mu_t(\theta) + \sigma_t(\theta) z_t, \quad (5.39)$$

$$\mu_t(\theta) = E[x_t | \mathcal{F}_{t-1}] = \mu(\theta, \mathcal{F}_{t-1}), \quad (5.40)$$

$$\sigma_t^2(\theta) = E[(x_t - \mu_t)^2 | \mathcal{F}_{t-1}] = \sigma^2(\theta, \mathcal{F}_{t-1}), \quad (5.41)$$

$$z_t \sim g(z_t | \eta_t), \quad (5.42)$$

$$\eta_t = \eta(\theta, \mathcal{F}_{t-1}), \quad (5.43)$$

where the innovation, z_t , follows a conditional distribution g with time-varying parameters. The conditional dependency of shape parameters is given by (5.43). For instance, in the case of the skewed Student t distribution, η_t includes the degree-of-freedom and the asymmetric parameters.

Since the GARCH approach has been empirically successful for modeling the conditional variance, it may seem natural to extend it to include the dynamics of the shape parameters η_t as follows

$$\eta_t = \eta(z_{t-1}, z_{t-2}, \dots).$$

An important issue is to constrain the dynamics in order to ensure that the function g is always a definite distribution.

Hansen (1994), Harvey and Siddique (1999), Lambert and Laurent (2002), and Jondeau and Rockinger (2003a) have all modeled skewness and distribution parameters as extended GARCH models. This sort of extension is not as straightforward as it may seem. First it requires, at each date and each step of the optimization, to solve the non-linear problem that links parameters and moments. The second problem, as mentioned before, is that a lot of constraints have to be imposed on the time-varying higher moments to make sure that the conditional distribution is well defined. The complex boundary conditions and complex non-linear relationships between parameters make the estimation very tricky, computationally intensive, and time consuming.

5.5.1 Tests for autoregressive conditional higher moments

It is possible to test for autoregressive conditional higher moments in a similar manner to Engle's $T \times R^2$ (Lagrange Multiplier, LM) test for conditional heteroskedasticity. We recall that the latter is performed by regressing the square of non-standardized residuals $\hat{\varepsilon}_t^2$ on p lagged values. Under the null of no conditional heteroskedasticity, the LM statistic $T \times R^2$ is distributed as a $\chi^2(p)$.

Similar tests for autoregressive conditional skewness and kurtosis can be performed. They are based on the fact that, because innovations z_t are standardized, we have $E[z_t^3] = s$ and $E[z_t^4] = \kappa$. The idea is then to regress estimates of z_t^3 and z_t^4 on lagged values respectively. Therefore, if we denote \hat{z}_t the estimated standardized residual, we have the two regressions

$$\begin{aligned} \text{for skewness: } & \hat{z}_t^3 = a_0 + a_1 \hat{z}_{t-1}^3 + \cdots + a_p \hat{z}_{t-p}^3 + u_t, \\ \text{for kurtosis: } & \hat{z}_t^4 = b_0 + b_1 \hat{z}_{t-1}^4 + \cdots + b_p \hat{z}_{t-p}^4 + u_t. \end{aligned}$$

Under the null of no autoregressive conditional skewness or kurtosis, the LM statistic $T \times R^2$ is distributed as a $\chi^2(p)$.¹⁸

Note that the test of constancy of higher moments should be performed on standardized residuals z_t instead of the non-standardized error term ε_t . The reason is that the skewness and kurtosis of the non-standardized error term (i.e., $s\sigma_t^3$ and $\kappa\sigma_t^4$) are directly related to volatility, so that they will be time-varying in case of heteroskedasticity.

5.5.2 Modeling higher moments directly

Several studies have directly considered the modeling of higher moments. Harvey and Siddique (1999) focus on the modeling of the dynamics of skewness in a non-central Student t distribution. Rockinger and Jondeau (2002) model

¹⁸ This result holds if moments up to the eighth one exist. For financial data, this assumption may be wrong.

skewness and kurtosis in the context of an entropy distribution. Brooks, Burke, and Persand (2005) consider an autoregressive conditional kurtosis in the context of the standard Student t distribution.

In Harvey and Siddique (1999), the dynamics of skewness is given by

$$s_t = \alpha_0 + \alpha_1 s_{t-1} + \alpha_2 z_{t-1}^3, \quad (5.44)$$

where z_t is the standardized innovation, with zero mean and unit variance. They placed the following constraints on the parameters: $-1 < \alpha_1 < 1$, $-1 < \alpha_2 < 1$, and $-1 < \alpha_1 + \alpha_2 < 1$. Since (5.44) is modeled within a GARCH context, it is called a GARCHS(1, 1, 1) (i.e., GARCH with skewness).

Using a very similar approach, Brooks, Burke, and Persand (2002) adopt the following dynamics for kurtosis

$$\kappa_t = \beta_0 + \beta_1 \kappa_{t-1} + \beta_2 z_{t-1}^4, \quad (5.45)$$

with the following constraints on the parameters: $-1 < \beta_1 < 1$, $-1 < \beta_2 < 1$, and $-1 < \beta_1 + \beta_2 < 1$. This model is named by the authors a GARCHK(1, 1, 1) (GARCH with kurtosis).

These approaches present the advantage that higher moments are directly estimated in the course of the optimization. Therefore, higher moments are modeled in a way that is similar to the variance in the GARCH framework. However, this does not come without a price. In order to estimate the model parameters, we have to derive the relationship between the higher moments s_t and κ_t and the (time-varying) shape parameters η_t . This relationship is very likely to be non-linear, except in very special cases such as Gram-Charlier expansions or entropy densities. In the context of the Student t distribution, which is adopted by Harvey and Siddique (1999) as well as by Brooks, Burke, and Persand (2005), the relationship between the higher moments and the estimated parameters is highly non-linear. Therefore, modeling higher moments directly requires, at each date and each step of the optimization, to solve the non-linear problem that links parameters and moments. This renders the estimation computationally very intensive.

Another drawback of this approach is that the constraints, which should be imposed on the higher moments, to ensure positivity of the distribution, are not clearly established. As seen before, skewness and kurtosis cannot be freely estimated. In the specifications above (equations (5.44) and (5.45)), the only constraints on the dynamics of skewness and kurtosis are to ensure stationarity of the process, not positivity of the distribution. It turns out that the constraints, presented in Section 5.2, to ensure that the density is defined, cannot be easily imposed in this context. As shown in the context of the skewed Student t distribution (Section 5.2.3), the constraints on parameters λ and ν to ensure that the distribution is definite are easy to establish. In contrast, the constraints on s and κ are not analytically known. A similar result holds for the Gram-Charlier series expansion.

In a slightly different context, Jondeau and Rockinger (2003a) have shown that an autoregressive structure such as presented above suffers from a severe

drawback. Consider the postulated dynamics (5.44) for conditional skewness. For data with sufficient variability, as the sample increases, the model is likely to degenerate to a solution where $\alpha_2 = 0$. To understand why this is so, we may write, if t is sufficiently large and $|\alpha_1| < 1$, that

$$s_t = \alpha_0/(1 - \alpha_1) + \alpha_2 \sum_{s=0}^{\infty} \alpha_1^s z_{t-1-s}^3.$$

From this expression, we see that the mean of s_t is $\alpha_0/(1 - \alpha_1)$ and its variance is $\alpha_2^2 V[\sum_{s=0}^{\infty} \alpha_1^s z_{t-1-s}^3]$. This shows that the restriction $s < s_t < \bar{s}$ will be satisfied for $\alpha_2 = 0$ only. Otherwise, there is a non-zero probability that the constraint may be violated for some observations.¹⁹ Consequently, an autoregressive process for skewness or kurtosis is very unlikely to be a fruitful model.

In Rockinger and Jondeau (2002), the dynamics of skewness and kurtosis are also directly modeled in the context of the entropy distribution. To avoid the problem described above, a very simple specification is adopted, that does not resort to an auto-regressive process

$$\begin{aligned} s_t &= a_0 + a_1 z_{t-1}, \\ \kappa_t &= b_0 + b_1 |z_{t-1}|. \end{aligned}$$

These dynamics are estimated while numerically imposing that skewness and kurtosis be inside the domain of definition described in section 5.2.6.

5.5.3 Modeling the parameters of the distribution

In his 1994 paper, Hansen extends the model with skewed Student t distribution (cf. section 5.2.3) to the case where parameters associated with the distribution (the degree of freedom ν and the asymmetry parameter λ) are also time varying. This yields the concept of AutoRegressive Conditional Density (ARCD).

A similar approach has been adopted by Jondeau and Rockinger (2003a) and Lambert and Laurent (2002). Jondeau and Rockinger (2003a) have discussed several possible specifications for the dynamics of the degree-of-freedom parameter ν_t and the asymmetry parameter λ_t . In Table 5.5, we display these specifications. Many other specifications could be designed, involving further lags or less linear relations. We emphasize these specifications, because they highlight some difficulties that may be encountered. We assume that the coefficients are such that stationarity is guaranteed.

Model M1 specifies directly ν_t and λ_t as functions of past realizations. The advantage of this specification is that no further non-linear map is required to

¹⁹ Such a problem does not occur in the GARCH context, because the volatility at time t , σ_t , is only required to be positive. This constraint is always fulfilled because the ε_t terms are squared.

Table 5.5. Possible specifications of the model

- Model M1: $\begin{cases} \nu_t = a_1 + b_1^+ z_{t-1}^+ + b_1^- z_{t-1}^-, \\ \lambda_t = a_2 + b_2^+ z_{t-1}^+ + b_2^- z_{t-1}^-. \end{cases}$
- Model M2: $\begin{cases} \tilde{\nu}_t = a_1 + b_1^+ z_{t-1}^+ + b_1^- z_{t-1}^-, \\ \tilde{\lambda}_t = a_2 + b_2^+ z_{t-1}^+ + b_2^- z_{t-1}^-, \\ \nu_t = \mathcal{L}_{[2,30]}(\tilde{\nu}_t), \quad \lambda_t = \mathcal{L}_{]-1,1[}(\tilde{\lambda}_t). \end{cases}$
- Model M3: $\begin{cases} \nu_t = a_1 + b_1^+ z_{t-1}^+ + b_1^- z_{t-1}^- + c_1 \nu_{t-1}, \\ \lambda_t = a_2 + b_2^+ z_{t-1}^+ + b_2^- z_{t-1}^- + c_2 \lambda_{t-1}. \end{cases}$
- Model M4: $\begin{cases} \tilde{\nu}_t = a_1 + b_1^+ z_{t-1}^+ + b_1^- z_{t-1}^- + c_1 \nu_{t-1}, \\ \tilde{\lambda}_t = a_2 + b_2^+ z_{t-1}^+ + b_2^- z_{t-1}^- + c_2 \lambda_{t-1}, \\ \nu_t = \mathcal{L}_{[2,30]}(\tilde{\nu}_t), \quad \lambda_t = \mathcal{L}_{]-1,1[}(\tilde{\lambda}_t). \end{cases}$
- Model M5: $\begin{cases} \tilde{\mu}_{3t} = a_1 + b_1 z_{t-1}^3, \\ \tilde{\mu}_{4t} = a_2 + b_2 z_{t-1}^4, \\ (\mu_{3t}, \mu_{4t}) = G(\tilde{\mu}_{3t}, \tilde{\mu}_{4t}), \\ (\nu_t, \lambda_t) = F^{-1}(\mu_{3t}, \mu_{4t}). \end{cases}$
- Model M6: $\begin{cases} \tilde{\mu}_{3t} = a_1 + b_1 z_{t-1}^3 + c_1 \tilde{\mu}_{3t-1}, \\ \tilde{\mu}_{4t} = a_2 + b_2 z_{t-1}^4 + c_2 \tilde{\mu}_{4t-1}, \\ (\mu_{3t}, \mu_{4t}) = G(\tilde{\mu}_{3t}, \tilde{\mu}_{4t}), \\ (\nu_t, \lambda_t) = F^{-1}(\mu_{3t}, \mu_{4t}). \end{cases}$

obtain a description of the parameters. A drawback of this specification is that its estimation is cumbersome because the constraints $2 < \nu_t$ and $-1 < \lambda_t < 1$ must be numerically imposed. Furthermore, nothing guarantees that ν_t and λ_t will be well defined out of sample. Ad hoc techniques, such as truncation at the boundaries, could be devised for forecasting purposes.

Model M2 specifies a dynamic for unconstrained $\tilde{\nu}_t$ and $\tilde{\lambda}_t$. These unrestricted parameters get mapped into the authorized domain \mathcal{D} via the logistic map $\mathcal{L}_{[a,b]}(x) = a + \exp(x) / (1 + \exp(x)) (b - a)$. Many of the drawbacks of model M1 disappear. However, one consequence is that the impact of extreme realizations gets damped because the logistic map tends to flatten the response of variables located in its tails.

Model M3 specifies parameters ν_t and λ_t as an autoregressive structure. As discussed above, such a specification suffers from a severe drawback, because, for data with sufficient variability, as the sample increases, the model is likely to degenerate to a solution where $b_2 = 0$.

Model M4 is similar to M3, yet it uses a non-linear map to constrain the parameters to \mathcal{D} . The model could still be estimated, but some care is needed in the interpretation of the estimates. For instance, if one estimates M4 and finds that b_1 is not statistically different from 0, then the model reduces to

$$\tilde{\nu}_t = a_1 + c_1 \tilde{\nu}_{t-1}.$$

At this stage, c_1 may be statistically significant. This may lead to the conclusion that there is persistence in the $\tilde{\nu}_t$. Such a conclusion would be erroneous, however. Indeed, if actual observations y_{t-1} do not matter, then, starting from some initial value $\tilde{\nu}_0$, the series of $\tilde{\nu}_t$ will quickly converge to its stationary level given by

$$\tilde{\nu}^* = a_1 / (1 - c_1).$$

In other words, the model where we would have estimated $\tilde{\nu}_t = \tilde{\nu}^*$ (with $b_1 = c_1 = 0$) could not be distinguished from the one obtained earlier. This implies that there exists an entire class of parameters (a_1, c_1) , all satisfying $(1 - c_1)\tilde{\nu}^* = a_1$, for which the model's characteristics are indistinguishable. The algorithm converges to one solution at random. To avoid this type of spurious finding, it is recommended to estimate M2 before M4 and to verify that past observations affect $\tilde{\nu}_t$ or $\tilde{\lambda}_t$. In no way should we trust in an estimation where c_1 or c_2 is statistically significant, yet, the parameters on the lagged innovations are not statistically significant. A further diagnostic to detect this behavior consists in changing the value of $\tilde{\nu}_1$ or of the initial value of the parameters in the numerical estimation. If the algorithm converges to significantly different values, then we should be careful about the estimated parameters.

In specification M5, the third and fourth non-central moments, s_t and κ_t , get specified using actual observations. For the model to be well defined, it must be that s_t and κ_t belong to the domain \mathcal{E} . This implies a potentially highly non-linear map G that maps some unrestricted \tilde{s}_t and $\tilde{\kappa}_t$ into \mathcal{E} . Furthermore, in order to obtain λ_t and ν_t from s_t and κ_t , it is necessary to invert a highly non-linear map that we call F in the table. Even though such an inversion could be done in theory, it will lead to a slow algorithm and also to a rather unstable estimation because the analytic computation of gradients may not be feasible. With such a specification, it is implicitly assumed that skewness and kurtosis are finite at each point of time. This observation is at odds with results from extreme value theory.

Model M6 presents the same difficulties as M5 with the added complication already discussed for model M3, in that one may find spurious dependence of skewness due to a lack of significant b_1 or b_2 estimates.

Modeling Correlation

In this chapter, we turn to the modeling of the multivariate distribution of asset returns. The most important characteristic of the multivariate distribution relies on the dependency parameter, i.e., the parameter that measures the strength of the link between two series. For a number of standard distributions (namely, those belonging to the elliptical family, which includes the normal and the Student t distributions), the dependency is simply measured by the Pearson's correlation coefficient. We often refer to correlation when meaning dependency. In practice, however, asset returns do not belong to this elliptical family and the dependency must be modeled with specific techniques. The modeling of the dependency parameter between asset returns has very important consequences in a number of financial applications.

An abundant literature has addressed the issue of how correlation between stock market returns varies when markets become agitated. The time-variability of correlation between returns is crucial from an asset management point of view. Asset allocation is often based on the use of a correlation matrix computed over a given sample period. If correlation increases during turbulent periods, the benefits of diversification would disappear when they are the most needed, i.e., during crashes. Implications for risk management are also obvious because the time-variability of the dependency parameter complicates Value at Risk (VaR) computation significantly.

As described in Chapter 2, the first tests of constancy of the dependency parameter were based on testing the equality of linear correlation coefficients computed before and after a crash. This approach has been found to be misleading, however, because conditioning the estimation of the correlation coefficient on the sample period induces an estimator bias. Subsequently, most tests of the constant correlation hypothesis have been based on the following approach: First, estimate the joint dynamics of stock-market returns and then describe how conditional correlations vary over time. There are several ways to model the joint dynamics of a number of series. The most widely used

approach is the multivariate GARCH model.¹ The central idea of the first generation of such models was that covariances had to be modeled using the same type of specification of variances in the univariate GARCH model (Kraft and Engle, 1982, and Bollerslev, Chou, and Kroner, 1992). These models, however, raise a problem of dimensionality because the number of parameters to be estimated increases dramatically with the number of series. Afterwards, most extensions have tried to reduce the computational burden.

Another difficulty of multivariate GARCH models relies on the choice of the conditional distribution of returns. When the marginal distributions are Gaussian, the extension to the multivariate case is trivial, because the joint distribution is simply a multivariate Gaussian distribution. However, for more general distributions, the multivariate extension is far from trivial. Often, it simply does not exist. In fact, an explicit multivariate extension exists only in very few cases. This is the case in particular for the Student t distribution. In other cases, a solution consists in constructing an implicit multivariate distribution by using copula functions. Due to the non-normality found in most financial return series, copula functions have had a great success, because they relate in an easy way very complicated marginal distributions. However, there is no free lunch. When copula functions are used, computing moments or, more generally, dealing with the integration of the joint distribution, becomes analytically intractable. Therefore, only numerical algorithms can be implemented. Unfortunately, in many financial applications, even with a few number of assets, such a numerical integration is too demanding to be performed in a reasonable length of time.

In this chapter, we describe how the dependency parameter may be modeled in the context of a multivariate GARCH model (Section 6.1). We present tests of constancy of the conditional dependency parameter. Notice that we will not provide a full description of the multivariate GARCH models. Rather, this section should be viewed as a brief introduction to this methodology, before applying multivariate models with non-normal distributions.² Then, we consider two aspects of the multivariate extension of non-Gaussian distributions. In Section 6.2, we consider the use of explicit multivariate distributions,

¹ An alternative approach is the multivariate Markov-switching model. For instance, Ramchand and Susmel (1998) and Ang and Bekaert (2002) test within this framework the hypothesis of a constant international conditional correlation between stock markets. Some papers also consider how correlation varies when stock-market indices are simultaneously affected by very large (positive or negative) fluctuations. Engle and Manganelli (2004) focus on the modeling of large realizations using quantile regressions. Longin and Solnik (2001), using extreme value theory, find that dependency increases more during downside movements than during upside movements. Poon, Rockinger, and Tawn (2004) provide an alternative statistical framework to test conditional dependency between extreme returns.

² In addition, it is worth emphasizing that the modeling of dependency in the context of extreme events will be detailed in Chapter 7.

which are generally difficult to estimate but allow a more efficient computation of moments. Then, in Section 6.3, we present the approach based on copula, whose estimation is much easier, although applications are more restricted.

6.1 Multivariate GARCH models

We now consider a random vector $x_t = (x_{1,t}, \dots, x_{n,t})'$ whose joint dynamics is given by

$$x_t = \mu_t(\theta) + \varepsilon_t, \quad (6.1)$$

$$\varepsilon_t = \Sigma_t^{1/2}(\theta) z_t, \quad (6.2)$$

where $\mu_t(\theta)$ denotes the $(n, 1)$ vector of conditional means, $\Sigma_t(\theta)$ denotes the (n, n) conditional covariance matrix of the error term ε_t , and θ is the vector of unknown parameters. The standardized innovation vector z_t is *iid* with mean $E[z_t] = 0$ and covariance matrix $V[z_t] = I_n$. $\Sigma_t^{1/2}(\theta)$ denotes the Cholesky decomposition of $\Sigma_t(\theta)$. In this chapter, we assume that z_t is drawn from the multivariate normal $\mathcal{N}(0, I_n)$ distribution.

Several parameterizations have been proposed for Σ_t . The main issue to be addressed is the dimensionality of the parameter vector when the number of variables n increases. Obviously, it is desirable that most statistical features highlighted in the univariate context be incorporated in the multivariate framework, in particular in terms of asymmetry and tail behavior. Some additional specific issues related to the multivariate framework have to be addressed. First, we have to deal with the conditions guaranteeing that the covariance matrix is positive definite at each date t . A second issue is whether conditional correlations have to be modeled instead of conditional covariance, and whether they have to be time-varying.

It should be noticed that we do not discuss the positivity and stationarity conditions, which have been widely studied in the literature. Very complete and comprehensive surveys of multivariate GARCH models may be found in Bollerslev, Engle, and Nelson (1994) and Bauwens, Laurent, and Rombouts (2005).

For further use, we define D_t the (n, n) diagonal matrix with the conditional variances σ_i^2 along the diagonal, so that $\{D_t\}_{ii} = \{\Sigma_t\}_{ii}$ and $\{D_t\}_{ij} = 0$, $\forall i \neq j$, for $i, j = 1, \dots, n$. We also define R_t , the (n, n) matrix of conditional correlations of ε_t , as $R_t = D_t^{-1/2} \Sigma_t D_t^{-1/2} = \{\rho_t\}_{ij}$. We deduce the $(n, 1)$ vector of normalized innovations $u_t = D_t^{-1/2} \varepsilon_t$. Notice that u_t differs from standardized innovations $z_t = \Sigma_t^{-1/2} \varepsilon_t$, because they are not orthogonalized.

6.1.1 Vectorial and diagonal GARCH models

Vech model

The first multivariate GARCH model, proposed by Kraft and Engle (1982) and Bollerslev, Chou, and Kroner (1992), assumes that each element of the covariance matrix is a linear function of the most recent past cross-products of errors and conditional variances and covariances. The Vech GARCH(p, q) model is defined as

$$\text{vech}(\Sigma_t) = \text{vech}(\Omega) + \sum_{i=1}^p A_i \text{vech}(\varepsilon_{t-i} \varepsilon'_{t-i}) + \sum_{j=1}^q B_j \text{vech}(\Sigma_{t-j}), \quad (6.3)$$

where Ω is an (n, n) positive definite and symmetric matrix, A_i and B_j are $(n(n+1)/2, n(n+1)/2)$ matrices; $\text{vech}(\cdot)$ is the operator that stacks the lower triangular elements of an (n, n) matrix as an $(n(n+1)/2, 1)$ vector. The number of parameters is $[n(n+1)/2][1 + (p+q)n(n+1)/2]$. Although this specification is very flexible, the large number of parameters (proportional to n^4) renders this model very difficult to handle. In addition, conditions that ensure that the conditional covariance matrices are positive definite are difficult to verify and impose.

Example: In the case $p = q = 1$ and $n = 2$, this model reduces to

$$\begin{pmatrix} \sigma_{1,t}^2 \\ \sigma_{12,t} \\ \sigma_{2,t}^2 \end{pmatrix} = \begin{pmatrix} \omega_{11} \\ \omega_{12} \\ \omega_{22} \end{pmatrix} + \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} \varepsilon_{1,t-1}^2 \\ \varepsilon_{1,t-1}\varepsilon_{2,t-1} \\ \varepsilon_{2,t-1}^2 \end{pmatrix} \\ + \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix} \begin{pmatrix} \sigma_{1,t-1}^2 \\ \sigma_{12,t-1} \\ \sigma_{2,t-1}^2 \end{pmatrix},$$

with 21 unknown parameters.

Diagonal vec model

In order to reduce the number of unknown parameters, Bollerslev, Engle, and Wooldridge (1988) have proposed the diagonal vec model, in which the matrices A_i^* and B_j^* are all taken to be diagonal (n, n) matrices: each element of the covariance matrix ($\sigma_{ij,t}$) only depends on the corresponding past elements $\sigma_{ij,t-1}$ and $\varepsilon_{i,t-1}\varepsilon_{j,t-1}$

$$\Sigma_t = \Omega^* + \sum_{i=1}^p A_i^* \odot (\varepsilon_{t-i} \varepsilon'_{t-i}) + \sum_{j=1}^q B_j^* \odot \Sigma_{t-j},$$

where Ω^* is an (n, n) positive definite and symmetric matrix, A_i^* and B_j^* are (n, n) symmetric matrices and \odot denotes the Hadamard product.³ This model

³ The Hadamard product defines the element-wise product of two matrices. So we have $\{A \odot B\}_{ij} = A_{ij}B_{ij}$.

has a natural interpretation, because covariances as well as variances have a GARCH-type specification. In addition, it reduces the number of unknown parameters considerably to $[n(n+1)/2](p+q+1)$, so that it is proportional to n^2 only.

Example: In the case $p = q = 1$ and $n = 2$, this model reduces to

$$\begin{aligned}\Sigma_t &= \begin{pmatrix} \omega_{11}^* & \omega_{12}^* \\ \omega_{12}^* & \omega_{22}^* \end{pmatrix} \\ &+ \begin{pmatrix} a_{11}^* & a_{12}^* \\ a_{12}^* & a_{22}^* \end{pmatrix} \odot \begin{pmatrix} \varepsilon_{1,t-1}^2 & \varepsilon_{1,t-1}\varepsilon_{2,t-1} \\ \varepsilon_{1,t-1}\varepsilon_{2,t-1} & \varepsilon_{2,t-1}^2 \end{pmatrix} \\ &+ \begin{pmatrix} b_{11}^* & b_{12}^* \\ b_{12}^* & b_{22}^* \end{pmatrix} \odot \begin{pmatrix} \sigma_{1,t-1}^2 & \sigma_{12,t-1} \\ \sigma_{12,t-1} & \sigma_{2,t-1}^2 \end{pmatrix}, \\ \Sigma_t &= \begin{pmatrix} \omega_{11}^* & \omega_{12}^* \\ \omega_{12}^* & \omega_{22}^* \end{pmatrix} + \begin{pmatrix} a_{11}^*\varepsilon_{1,t-1}^2 & a_{12}^*\varepsilon_{1,t-1}\varepsilon_{2,t-1} \\ a_{12}^*\varepsilon_{1,t-1}\varepsilon_{2,t-1} & a_{22}^*\varepsilon_{2,t-1}^2 \end{pmatrix} \\ &+ \begin{pmatrix} b_{11}^*\sigma_{1,t-1}^2 & b_{12}^*\sigma_{12,t-1} \\ b_{12}^*\sigma_{12,t-1} & b_{22}^*\sigma_{2,t-1}^2 \end{pmatrix},\end{aligned}$$

where $\Sigma_t = \begin{pmatrix} \sigma_{1,t}^2 & \sigma_{12,t} \\ \sigma_{12,t} & \sigma_{2,t}^2 \end{pmatrix}$. In such a case, there are 9 unknown parameters.

BEKK model

An alternative representation is the BEKK representation described by Engle and Kroner (1995)⁴

$$\Sigma_t = \tilde{\Omega} + \sum_{i=1}^p \tilde{A}'_i \varepsilon_{t-i} \varepsilon'_{t-i} \tilde{A}_i + \sum_{j=1}^q \tilde{B}'_j \Sigma_{t-j} \tilde{B}_j,$$

where $\tilde{\Omega}$ is an (n, n) positive definite and symmetric matrix, and \tilde{A}_i and \tilde{B}_j are (n, n) matrices. This specification involves $[n(n+1)/2] + (p+q)n^2$ unknown parameters. The main advantage of this specification is that the conditional covariance matrix is positive definite as long as $\tilde{\Omega}$ also is.

Example: In the case $p = q = 1$ and $n = 2$, this model reduces to

$$\begin{aligned}\Sigma_t &= \begin{pmatrix} \tilde{\omega}_{11} & \tilde{\omega}_{12} \\ \tilde{\omega}_{12} & \tilde{\omega}_{22} \end{pmatrix} \\ &+ \begin{pmatrix} \tilde{a}_{11} & \tilde{a}_{12} \\ \tilde{a}_{21} & \tilde{a}_{22} \end{pmatrix}' \begin{pmatrix} \varepsilon_{1,t-1}^2 & \varepsilon_{1,t-1}\varepsilon_{2,t-1} \\ \varepsilon_{1,t-1}\varepsilon_{2,t-1} & \varepsilon_{2,t-1}^2 \end{pmatrix} \begin{pmatrix} \tilde{a}_{11} & \tilde{a}_{21} \\ \tilde{a}_{12} & \tilde{a}_{22} \end{pmatrix} \\ &+ \begin{pmatrix} \tilde{b}_{11} & \tilde{b}_{12} \\ \tilde{b}_{21} & \tilde{b}_{22} \end{pmatrix}' \begin{pmatrix} \sigma_{1,t-1}^2 & \sigma_{12,t-1} \\ \sigma_{12,t-1} & \sigma_{2,t-1}^2 \end{pmatrix} \begin{pmatrix} \tilde{b}_{11} & \tilde{b}_{21} \\ \tilde{b}_{12} & \tilde{b}_{22} \end{pmatrix},\end{aligned}$$

⁴ The acronym BEKK stands for Baba, Engle, Kraft, and Kroner.

with 11 unknown parameters. To reduce the computational burden, this model may be further constrained. The diagonal BEKK model is written as

$$\begin{aligned}\Sigma_t = & \begin{pmatrix} \tilde{\omega}_{11} & \tilde{\omega}_{12} \\ \tilde{\omega}_{12} & \tilde{\omega}_{22} \end{pmatrix} \\ & + \begin{pmatrix} \tilde{a}_{11} & 0 \\ 0 & \tilde{a}_{22} \end{pmatrix} \begin{pmatrix} \varepsilon_{1,t-1}^2 & \varepsilon_{1,t-1}\varepsilon_{2,t-1} \\ \varepsilon_{1,t-1}\varepsilon_{2,t-1} & \varepsilon_{2,t-1}^2 \end{pmatrix} \begin{pmatrix} \tilde{a}_{11} & 0 \\ 0 & \tilde{a}_{22} \end{pmatrix} \\ & + \begin{pmatrix} \tilde{b}_{11} & 0 \\ 0 & \tilde{b}_{22} \end{pmatrix} \begin{pmatrix} \sigma_{1,t-1}^2 & \sigma_{12,t-1} \\ \sigma_{12,t-1} & \sigma_{2,t-1}^2 \end{pmatrix} \begin{pmatrix} \tilde{b}_{11} & 0 \\ 0 & \tilde{b}_{22} \end{pmatrix},\end{aligned}$$

while the scalar BEKK model is

$$\begin{aligned}\Sigma_t = & \begin{pmatrix} \tilde{\omega}_{11} & \tilde{\omega}_{12} \\ \tilde{\omega}_{12} & \tilde{\omega}_{22} \end{pmatrix} + \tilde{a}^2 \begin{pmatrix} \varepsilon_{1,t-1}^2 & \varepsilon_{1,t-1}\varepsilon_{2,t-1} \\ \varepsilon_{1,t-1}\varepsilon_{2,t-1} & \varepsilon_{2,t-1}^2 \end{pmatrix} \\ & + \tilde{b}^2 \begin{pmatrix} \sigma_{1,t-1}^2 & \sigma_{12,t-1} \\ \sigma_{12,t-1} & \sigma_{2,t-1}^2 \end{pmatrix}.\end{aligned}$$

Dealing with the constant term

As discussed in Engle and Mezrich (1996), these models can be estimated with the additional constraint that the long-run covariance matrix is equal to the sample covariance matrix. This approach is often called *variance targeting*. It reduces the number of parameters dramatically and often gives improved performance in finite sample. For instance, in the case of the Vech model, we have the following parameterization

$$\text{vech}(\Omega) = \left(I_n - \sum_{i=1}^p A_i - \sum_{j=1}^q B_j \right) \text{vech}(S),$$

where $S = \frac{1}{T} \sum_{t=1}^T \hat{\varepsilon}_t \hat{\varepsilon}_t'$ is the sample covariance matrix of residuals.

6.1.2 Dealing with large-dimensional systems

One general problem with multivariate GARCH models is the problem of dimensionality. The specifications described above would be very difficult to implement for large-dimensional portfolios because of the large number of parameters to be estimated. The Factor GARCH approach (and its extensions) reduces the number of variables (the factors) that have to be modeled with a multivariate GARCH dynamics. The Flexible GARCH model is based on a decentralized estimation of the covariance matrix.

Factor GARCH model

The Factor GARCH (or F-GARCH) has been proposed by Engle, Ng, and Rothschild (1990) to further reduce the number of parameters to be estimated. Engle, Ng, and Rothschild (1992) have proposed an extension of the F-GARCH model that allows distinguishing between the dynamic and static structure of asset returns. Other work based on F-GARCH models is by King, Sentana, and Wadhwani (1994) or Sentana and Fiorentini (2001).

The idea of this parameterization is that the joint dynamics of the $(n, 1)$ vector of returns x_t can be correctly described using a small number of factors K ($K < n$). This model has been used by Bollerslev and Engle (1993) to model common persistence in stock market volatilities. The description of x_t is given by the $(K, 1)$ vector of factors f_t and the (n, K) matrix B of time-invariant factor loadings

$$x_t = Bf_t + \varepsilon_t.$$

Assume that the error term ε_t has constant (n, n) conditional covariance matrix Ω , that the K factors f_t have conditional covariance matrix A_t and that ε_t and f_t are uncorrelated. Then, the conditional covariance matrix of x_t is equal to⁵

$$V_{t-1}[x_t] = \Sigma_t = \Omega + BA_tB'. \quad (6.4)$$

If we assume now that the conditional covariance matrix of factors A_t is diagonal with elements $\lambda_{k,t}$ or if off-diagonal elements are constant and combined into Ω , then the model can be simplified as

$$\Sigma_t = \Omega + \sum_{k=1}^K \beta_k \beta'_k \lambda_{k,t},$$

where β_k denotes the k th column in B .

F-GARCH models have several interesting implications (see Engle, Ng, and Rothschild, 1990). First, the conditional covariance matrix Σ_t is guaranteed to be positive semi-definite. Second, we can always construct portfolios of assets that have the same conditional variance $\lambda_{k,t}$ as factors (up to a constant term). To see this, consider the portfolio $r_{k,t} = \phi'_k x_t$ where $\phi'_k \beta_j = 1$ if $j = k$ and 0 otherwise. Then, the conditional variance of $r_{k,t}$ is given by

$$V_{t-1}[r_{k,t}] = \phi'_k \Sigma_t \phi_k = \psi_k + \lambda_{k,t},$$

with $\psi_k = \phi'_k \Omega \phi_k$. Therefore, the portfolio $r_{k,t}$ has exactly the same time variation as the k th factor, so that it can be called factor-representing portfolio. This property indicates that the information in the factor-representing portfolios is sufficient for predicting the variances and covariances of individual asset returns. If we assume that the dynamics of each component $\lambda_{k,t}$ is

⁵ In the case where ε_t and f_t are correlated with constant correlation matrix, we also obtain (6.4), with the constant matrix Ω regrouping terms of the covariance matrix of ε_t and the conditional covariance matrix of ε_t and f_t .

given by univariate GARCH(1, 1) models, we obtain the following conditional variances

$$V_{t-1}[r_{k,t}] = \omega_k + \alpha_k (\phi'_k \varepsilon_{t-1})^2 + \gamma_k V_{t-2}[r_{k,t-1}].$$

Finally, if $K < n$, we can always construct $n - K$ portfolios of assets, i.e., linear combinations of x_t , which have constant variance.

Orthogonal models

The Orthogonal GARCH model (or O-GARCH) has been proposed by Alexander and Chibumba (1997) and Alexander (2001). It assumes that the observed data can be obtained by a linear transformation of a set of uncorrelated components and the matrix of the linear transformation is an orthogonal matrix. The (n, n) covariance matrix Σ_t is generated by m univariate GARCH models, where $m \leq n$ is determined using principal component analysis. The main interest of this approach is that it avoids estimating off-diagonal components of the multivariate GARCH parameter matrices, because the model is estimated not with original data but with its principal components that are by construction unconditionally uncorrelated.

The first step is to compute the sample (n, n) correlation matrix \bar{R} of the n normalized innovations $u_t = D^{-1/2} \varepsilon_t$, where $D = \text{diag}(\sigma_1^2, \dots, \sigma_n^2)$ with σ_i^2 the sample variance of ε_t . Then, we assume that the data are generated by an orthogonal transformation of a small number K ($K \leq n$) of factors f_t . The matrix of this transformation is given by the eigenvectors of \bar{R} . The O-GARCH model is therefore defined as

$$\begin{aligned} u_t &= V_K f_t, \\ V_K &= P_K \Lambda_K^{1/2} = P_K \times \text{diag}(\lambda_1^{1/2}, \dots, \lambda_K^{1/2}), \end{aligned}$$

where Λ_K is the (K, K) diagonal matrix of eigenvalues of \bar{R} (ranked from the largest to the smallest, $\lambda_1 \geq \dots \geq \lambda_K$), P_K the associated (n, K) orthogonal matrix of eigenvectors, and $V_K = P_K \Lambda_K^{1/2}$. The unconditional covariance of f_t is the identity matrix of order K , by construction. The conditional covariance matrix of f_t is assumed to be a diagonal matrix (denoted Q_t) where each diagonal element is specified as a univariate GARCH model. The vector of factors $f_t = (f_{1,t}, \dots, f_{K,t})'$ is thus characterized by

$$\begin{aligned} E_{t-1}[f_t] &= 0, \\ V_{t-1}[f_t] &= Q_t = \text{diag}(\sigma_{f_1,t}^2, \dots, \sigma_{f_K,t}^2), \\ \sigma_{f_i,t}^2 &= \left(1 - \sum_{h=1}^p \alpha_{ih} - \sum_{h=1}^q \beta_{ih}\right) + \sum_{h=1}^p \alpha_{ih} f_{i,t-h}^2 + \sum_{h=1}^q \beta_{ih} \sigma_{f_i,t-h}^2. \end{aligned}$$

Finally, the conditional covariance matrix of ε_t is simply given by

$$\Sigma_t = V_{t-1} [\varepsilon_t] = D^{1/2} R_t D^{1/2},$$

where

$$R_t = V_{t-1} [u_t] = V_m Q_t V'_m.$$

The parameters to be estimated are the following: The individual variances σ_i^2 of ε_t (included in D), the sample correlation matrix of u_t (the decomposition of which provides us with V_m), and the parameters (α_{ih} and β_{ih} , for $i = 1, \dots, K$) of the univariate GARCH models. We notice that the individual variances σ_i^2 can be estimated by the sample analogue of D . In addition, once the number of factors K is chosen, V_K can be computed directly from the sample counterpart of \bar{R} . Consequently, the estimation burden reduces itself to the estimation of the individual GARCH models for the K factors.

The O-GARCH model implicitly assumes that the observed data can be linearly transformed into a set of uncorrelated components by means of an orthogonal matrix. These unobserved components can be interpreted as a set of uncorrelated factors that drive the market, similar to that in the F-GARCH model. The orthogonality assumption appears to be rather restrictive, however. Van der Weide (2002) has proposed the Generalized Orthogonal GARCH model (or GO-GARCH), in which innovations are linked by any possible invertible matrix. For this purpose, he argues that, when V is the map that links the uncorrelated components f_t with the observed process u_t , then there exists an orthogonal matrix U such that $V = P A^{1/2} U$, with $|U| = 1$. The O-GARCH model implicitly assumes $U = I_K$.

The matrix U can be represented as a product of $K(K - 1)/2$ rotation matrices $U = \prod_{i < j} G_{ij}(\theta_{ij})$, with $-\pi \leq \theta_{ij} \leq \pi$, where $G_{ij}(\theta_{ij})$ performs a rotation in the plane spanned by e_i and e_j over an angle θ_{ij} for $i, j = 1, \dots, n$, and e_i is the i th column of the (n, n) identity matrix. The conditional covariance matrix of ε_t has the same expression as before

$$\Sigma_t = V_{t-1} [\varepsilon_t] = D^{1/2} R_t D^{1/2},$$

where

$$R_t = V_{t-1} [u_t] = V Q_t V',$$

with $V = P A^{1/2} U$. The GO-GARCH model is stationary, provided the independent GARCH processes are stationary. In the case where $K = n$, the O-GARCH model can be viewed as a GO-GARCH for the particular choice $U = I_n$. Note also that the GO-GARCH model is a special case of the BEKK model, so that its properties can be derived from those of the BEKK model.

As for the O-GARCH model, an estimate of the parameters in D and V can be obtained from the sample analogues of D and \bar{R} . Therefore, the remaining parameters to be estimated are the parameters (α_{ih} and β_{ih} , for $i = 1, \dots, K$) of the univariate GARCH models and the $K(K - 1)/2$ rotation angles θ_{ij} . Related approaches have been followed by Vrontos, Dellaportas, and Politis (2003) and by Lanne and Saikkonen (2005).

Flexible GARCH model

Another approach can be taken in order to reduce the computational burden of large-scale multivariate GARCH models. Most models described in Section 6.1.1 are designed to reduce the number of parameters to be estimated. First-generation models impose some additional structure on the general specification. F-GARCH models summarize the information in asset returns through factors. By contrast, the flexible GARCH model proposed by Ledoit, Santa-Clara, and Wolf (2003) does not try to reduce the number of parameters but instead decentralize the estimation problem.

Assume that we have to estimate a diagonal vec model of the form

$$\sigma_{ij,t} = \omega_{ij} + \alpha_{ij}\varepsilon_{i,t-1}\varepsilon_{j,t-1} + \beta_{ij}\sigma_{ij,t-1},$$

where the conditional covariance $\sigma_{ij,t}$ between assets i and j depends on its own lag and on the cross-product between lagged innovations $\varepsilon_{i,t-1}$ and $\varepsilon_{j,t-1}$. In this case, for each (co)variance, we have 3 parameters to estimate, so that for n assets, we would have $3n(n+1)/2$ parameters. The idea of the flexible GARCH model consists in estimating the parameters $\{\omega_{ij}, \alpha_{ij}, \beta_{ij}\}$ for each (i, j) separately. Thus, the problem reduces to estimating one-dimensional (when $i = j$) or two-dimensional (when $i \neq j$) models. The difficulty then is to combine the various estimates into matrices $\hat{\Omega} = \{\hat{\omega}_{ij}\}$, $\hat{A} = \{\hat{\alpha}_{ij}\}$ and $\hat{B} = \{\hat{\beta}_{ij}\}$. Since the estimated parameters come from independent estimations, the covariance matrix Σ_t is not guaranteed to be positive semi-definite. Thus, once elements of the Σ_t matrix have been estimated, it will be necessary to use some trick to ensure positive semi-definiteness of this matrix.

The first step of the estimation of the flexible GARCH model corresponds to the (Q)ML estimation of the diagonal and off-diagonal elements of the covariance matrix. Diagonal elements are estimated using the standard univariate GARCH specification

$$\sigma_{ii,t} = \omega_{ii} + \alpha_{ii}\varepsilon_{i,t-1}^2 + \beta_{ii}\sigma_{ii,t-1},$$

with $\omega_{ii} > 0$, $\alpha_{ii} \geq 0$, $\beta_{ii} \geq 0$, and $\alpha_{ii} + \beta_{ii} < 1$. Off-diagonal elements ($\sigma_{ij,t}$) are estimated using the bivariate GARCH model for assets i and j

$$\varepsilon_{(ij),t} = \Sigma_{(ij),t}^{1/2} z_{(ij),t}$$

where

$$\varepsilon_{(ij),t} = \begin{pmatrix} \varepsilon_{i,t} \\ \varepsilon_{j,t} \end{pmatrix}, \quad z_{(ij),t} = \begin{pmatrix} z_{i,t} \\ z_{j,t} \end{pmatrix}, \quad \Sigma_{(ij),t} = \begin{pmatrix} \hat{\sigma}_{ii,t} & \sigma_{ij,t} \\ \sigma_{ij,t} & \hat{\sigma}_{jj,t} \end{pmatrix},$$

and

$$\sigma_{ij,t} = \omega_{ij} + \alpha_{ij}\varepsilon_{i,t-1}\varepsilon_{j,t-1} + \beta_{ij}\sigma_{ij,t-1}.$$

The vector $z_{(ij),t}$ is assumed to be normal $\mathcal{N}(0, I_2)$. Therefore, in the bivariate model, only the parameters pertaining to the covariance $\sigma_{ij,t}$ are estimated,

and the conditional variances $\sigma_{ii,t}$ and $\sigma_{jj,t}$ are fixed to their first-stage values $\hat{\sigma}_{ii,t}$ and $\hat{\sigma}_{jj,t}$. In order to ensure that $\Sigma_{(ij),t}$ is positive definite, we impose the following bounds during the estimation: $|\omega_{ij}| \leq \sqrt{\hat{\omega}_{ii}\hat{\omega}_{jj}}$, $0 \leq \alpha_{ij} \leq \sqrt{\hat{\alpha}_{ii}\hat{\alpha}_{jj}}$, and $0 \leq \beta_{ij} \leq \sqrt{\hat{\beta}_{ii}\hat{\beta}_{jj}}$.

Then, Ledoit, Santa-Clara, and Wolf (2003) show how to render the bivariate estimates compatible in the sense that matrix Σ_t is positive definite. Using the matrix notation for the diagonal vec model

$$\Sigma_t = \Omega + A \odot (\varepsilon_{t-1} \varepsilon'_{t-1}) + B \odot \Sigma_{t-1},$$

and denoting \div the element-wise division, they show that the conditional covariance matrix Σ_t is positive semi-definite if the three matrices $D \equiv \Omega \div (I_n - B)$, A , and B are positive semi-definite and $\alpha_{ii} + \beta_{ii} < 1$, $\forall i = 1, \dots, n$. Now, if we define $\hat{D} = \hat{\Omega} \div (I_n - \hat{B})$, we need to transform the estimated parameter matrices \hat{A} , \hat{B} , and \hat{D} in order to ensure positive semi-definiteness of the conditional covariance matrix. The new matrices \tilde{A} , \tilde{B} , and \tilde{D} are chosen to be the closest to \hat{A} , \hat{B} , and \hat{D} but such that the diagonal parameters obtained from the estimation of the univariate GARCH models remain unchanged. Formally, we have to solve the following problems

$$\min_{\tilde{D}} \left\| \tilde{D} - \hat{D} \right\|$$

s.t. \tilde{D} is positive semi-definite and $\tilde{d}_{ii} = \hat{d}_{ii}$, $\forall i = 1, \dots, n$,

$$\min_{\tilde{A}} \left\| \tilde{A} - \hat{A} \right\|$$

s.t. \tilde{A} is positive semi-definite and $\tilde{\alpha}_{ii} = \hat{\alpha}_{ii}$, $\forall i = 1, \dots, n$,

$$\min_{\tilde{B}} \left\| \tilde{B} - \hat{B} \right\|$$

s.t. \tilde{B} is positive semi-definite and $\tilde{\beta}_{ii} = \hat{\beta}_{ii}$, $\forall i = 1, \dots, n$.

Once these matrices have been obtained, we deduce $\tilde{\Omega} = \tilde{D} \odot (I_n - \tilde{B})$. One interesting property of this approach is that we have, by construction, $|\tilde{\alpha}_{ij} + \tilde{\beta}_{ij}| < 1$, $\forall i, j = 1, \dots, n$. Therefore, it is sufficient to impose that $\hat{\alpha}_{ii} + \hat{\beta}_{ii} < 1$, $\forall i = 1, \dots, n$ to ensure positive semi-definiteness of the conditional covariance matrix.

A drawback of this approach is that there is no straightforward way to compute the standard errors of the parameter estimates. The reason is that the new matrices \tilde{A} , \tilde{B} , and \tilde{D} are very nonlinear transformations of the initial matrices \hat{A} , \hat{B} , and \hat{D} for which standard errors are available. Ledoit, Santa-Clara, and Wolf (2003) suggest the use of the bootstrap procedure to obtain standard errors. It should be noticed that for large-scale multivariate GARCH models, the need for such standard errors is not clear.

6.1.3 Modeling conditional correlation

The models described in the first section can be viewed as natural extensions of the baseline univariate GARCH model. In particular, they all propose a specification for the conditional covariances that is similar to the one adopted for modeling variances in the univariate GARCH model. Unfortunately, they have some drawbacks. First, the number of unknown parameters is a power function of the number of variables, so that the estimation of these models becomes extremely difficult as n grows. Second, the derivation of the restrictions ensuring that the covariance matrix is positive definite is often difficult (except for the BEKK model).

Second-generation models focus on the dynamics of correlations rather than on the dynamics of covariances. At first sight, this task is more demanding, because it cannot be constructed as a natural generalization of the univariate GARCH model. However, due to the critical role of correlations in finance, this shift from covariances to correlations was needed.

The Constant Conditional Correlation (CCC) model

Bollerslev (1990) has suggested that the time-varying conditional covariances be parameterized in order to be proportional to the product of the corresponding conditional standard deviations. The intuition for this model is the following. Assume that $\sigma_{ij,t}$ is the covariance between two assets i and j to be modeled. Also, let $\sigma_{i,t}^2$ be the conditional variance modeled by some univariate GARCH model. Under the assumption of keeping correlation constant, denoting ρ_{ij} the constant correlation between the assets i and j , it follows that

$$\rho_{ij} = \frac{\sigma_{ij,t}}{\sigma_{i,t}\sigma_{j,t}} \quad \Rightarrow \quad \sigma_{ij,t} = \rho_{ij}\sigma_{i,t}\sigma_{j,t}.$$

Thus, knowledge of ρ_{ij} that can be computed using standardized innovations, and knowledge of the marginal GARCH models, yields a description of the time-varying covariance.

The extension to a general model with n assets is straightforward. Bollerslev (1990) introduces a time-invariant (n, n) correlation matrix with unit diagonal elements

$$R = \begin{pmatrix} 1 & \rho_{12} & \cdots & \rho_{1n} \\ \rho_{12} & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho_{n-1,n} \\ \rho_{1n} & \cdots & \rho_{n-1,n} & 1 \end{pmatrix}. \quad (6.5)$$

Therefore, the temporal variation in Σ_t is determined solely by the conditional variances

$$\Sigma_t = D_t^{1/2} R D_t^{1/2}, \quad (6.6)$$

where D_t is, as before, the diagonal matrix of conditional variances. We only need to model the dynamics of the n conditional variances and to estimate

the constant correlation matrix, so that the number of parameters to estimate reduces itself to $n(1 + p + q) + n(n + 1)/2$. An advantage of this approach is that if the conditional variances in the D_t matrices are all positive and the conditional correlation matrix R is positive definite, the sequence of conditional covariance matrices Σ_t is guaranteed to be positive definite for all t . Note, in addition, that as argued before, the correlation matrix R may be estimated in a preliminary step using the sample correlation matrix of normalized residuals.

The CCC model has been widely used in the empirical literature because of its computational simplicity. However, at least for financial returns, the assumption of constant correlation is not supported by the data. There is a huge empirical evidence that correlations vary over time (see Chapter 2). Several extensions of the CCC model have been recently proposed to allow time-varying conditional correlations.

The first models with time-varying conditional correlations have been proposed by Engle (2002), Engle and Sheppard (2001), and Tse and Tsui (2002). The basic idea is that the conditional correlation matrix R_t is in fact time varying, so that the conditional covariance matrix previously defined by (6.6) is now defined as

$$\Sigma_t = D_t^{1/2} R_t D_t^{1/2}. \quad (6.7)$$

Since the two models of Engle (2002) and Tse and Tsui (2002) are conceptually different, we present both of them in turn.

The dynamic conditional correlation (DCC) model

Engle (2002) and Engle and Sheppard (2001) have developed a model in which the conditional correlation matrix in (6.7) is defined by

$$R_t = \text{diag}(Q_t)^{-1/2} \times Q_t \times \text{diag}(Q_t)^{-1/2}, \\ Q_t = (1 - \delta_1 - \delta_2) \bar{Q} + \delta_1 (u_{t-1} u'_{t-1}) + \delta_2 Q_{t-1},$$

where \bar{Q} is the unconditional covariance matrix of $u_t = \{\varepsilon_{i,t}/\sigma_{i,t}\}_{i=1,\dots,n}$ and $\text{diag}(Q_t)$ is the (n, n) matrix with the diagonal of Q_t on the diagonal and zeros off-diagonal. The matrix \bar{Q} may be estimated by the sample analogue $\frac{1}{T} \sum_{t=1}^T \hat{u}_t \hat{u}'_t$. Parameters δ_1 and δ_2 are assumed to satisfy $0 \leq \delta_1, \delta_2 \leq 1$ and $\delta_1 + \delta_2 \leq 1$. Once these restrictions are imposed, the conditional correlation matrix R_t is guaranteed to be positive definite during the estimation.

A drawback of this approach is that only two additional parameters δ_1 and δ_2 drive the dynamics of all the correlations. Cappiello, Engle, and Sheppard (2003) have recently suggested an extension of this model in which each element of the correlation matrix has an autonomous and asymmetric dynamics

$$R_t = \text{diag}(Q_t)^{-1/2} \times Q_t \times \text{diag}(Q_t)^{-1/2}, \\ Q_t = (\bar{Q} - A' \bar{Q} A - B' \bar{Q} B - G' \bar{N} G) + A' (u_{t-1} u'_{t-1}) A \\ + B' Q_{t-1} B + G' (n_{t-1} n'_{t-1}) G,$$

where A , B and G are (n, n) diagonal parameter matrices and $n_t = 1_{\{u_t < 0\}} \odot u_t$ is the $(n, 1)$ vector that contains the normalized residual if it is negative and 0 otherwise. $\bar{Q} = E[u_t u_t']$ and $\bar{N} = E[n_t n_t']$ are estimated by their sample analogues $\frac{1}{T} \sum_{t=1}^T \hat{u}_t \hat{u}_t'$ and $\frac{1}{T} \sum_{t=1}^T \hat{n}_t \hat{n}_t'$.

The conditional covariance matrix $\Sigma_t = D_t^{1/2} R_t D_t^{1/2}$ is guaranteed to be positive definite if $(\bar{Q} - A' \bar{Q} A - B' \bar{Q} B - G' \bar{N} G)$ is positive definite. The DCC model of Engle (2002) is obtained as a special case if $G = 0$ and if A and B have $\sqrt{\delta_1}$ and $\sqrt{\delta_2}$ on the diagonal, respectively.

The time-varying correlation (TVC) model

The idea of Tse and Tsui (2002) is to specify the dynamic of the conditional correlation matrix R_t as an ARMA process

$$R_t = (1 - \theta_1 - \theta_2) R + \theta_1 R_{t-1} + \theta_2 \Psi_{t-1}, \quad (6.8)$$

where $R = \{\rho_{ij}\}$ is a time-invariant (n, n) matrix of correlations as in (6.5). The key idea of Tse and Tsui (2002) is to model the (n, n) matrix $\Psi_t = \{\psi_{ij,t}\}$ using a set of past normalized observations

$$\psi_{ij,t} = \frac{\sum_{h=0}^{m-1} u_{i,t-h} u_{j,t-h}}{\sqrt{\left(\sum_{h=0}^{m-1} u_{i,t-h}^2\right) \left(\sum_{h=0}^{m-1} u_{j,t-h}^2\right)}} \quad \text{for } 1 \leq i < j \leq n,$$

where $u_t = D_t^{-1} \varepsilon_t = (\varepsilon_{i,t} / \sigma_{i,t})_{i=1, \dots, n}$. Stated differently, Ψ_t is the sample correlation matrix of normalized residuals $E_t = (u_t, \dots, u_{t-m+1})'$. Therefore, if we define B_t the (n, n) diagonal matrix with $\left(\sum_{h=0}^{m-1} u_{i,t-h}^2\right)^{1/2}$ as i th diagonal element, we can rewrite Ψ_t as

$$\Psi_t = B_t^{-1} E_t E_t' B_t^{-1} = \{\psi_{ij,t}\}.$$

As long as $m \geq n$, the matrix Ψ_t will be in general positive definite, of course if the $u_{i,t}$ are not linearly dependent. To ensure positive definiteness of the matrix R_t , the parameters θ_1 and θ_2 have to satisfy $0 \leq \theta_1, \theta_2 \leq 1$ and $\theta_1 + \theta_2 \leq 1$. Note that the constant conditional correlation model is nested in this model, because it corresponds to the case where $\theta_1 = \theta_2 = 0$. Time-variability in the conditional correlation matrix is therefore obtained at the cost of only two additional parameters.

While the model proposed by Tse and Tsui (2002) builds on an ARMA process for the dynamics of the correlation matrix, the model of Engle (2002) is based on a GARCH-type specification for the dynamics of the covariance matrix. An advantage of the latter approach is that the dynamic of Q_t is based on a single lag of the $u_{i,t}$ terms (as in the standard GARCH(1, 1) model), whereas (6.8) requires an arbitrary number m of observations to compute sample correlations.

General dynamic covariance model

A generalization of most of the previous models has been developed by Kroner and Ng (1998). The so-called general dynamic covariance (GDC) model nests many of the existing models while including asymmetric effects. The model is defined as follows

$$\begin{aligned}\Sigma_t &= D_t R D_t + \Phi \odot \Theta_t, \\ D_t &= \{d_{ij,t}\} \text{ with } d_{ii,t} = \sqrt{\theta_{ii,t}}, \forall i \text{ and } d_{ij,t} = 0, \forall i \neq j, \\ \Theta_t &= \{\theta_{ij,t}\}, \\ \theta_{ij,t} &= \omega_{ij} + a_i' \varepsilon_{t-1} \varepsilon_{t-1}' a_i + b_i \Sigma_{t-1} b_i' \quad \forall i, j, \\ R &= \{\rho_{ij}\} \text{ with } \rho_{ii} = 1 \quad \forall i, \\ \Phi &= \{\varphi_{ij}\} \text{ with } \varphi_{ii} = 0, \forall i \text{ and } \varphi_{ij} = \varphi_{ji}, \forall i, j,\end{aligned}$$

where $a_i, b_i, i = 1, \dots, n$, are $(n, 1)$ parameter vectors, ρ_{ij} , φ_{ij} , and ω_{ij} are scalars with $\Omega = \{\omega_{ij}\}$ and R positive definite and symmetric matrices.

The GDC model has two components: the first term $D_t R D_t$ is like the constant correlation model but with the variance functions given by that of the BEKK model. The second term $\Phi \odot \Theta_t$ has zero diagonal elements but has off-diagonal elements given by the BEKK-type covariance functions. Note that the elements of Σ_t can be written as

$$\begin{aligned}\sigma_{ii,t} &= \theta_{ii,t} \quad \forall i, \\ \sigma_{ij,t} &= \rho_{ij} \sqrt{\theta_{ii,t}} \sqrt{\theta_{jj,t}} + \varphi_{ij} \theta_{ij,t} \quad \forall i \neq j.\end{aligned}$$

Thus, the GDC model is a hybrid of the constant conditional correlation model and the BEKK model.

Proposition 6.1. (Kroner and Ng, 1998) Consider the following set of conditions:

1. $\rho_{ij} = 0 \quad \forall i \neq j$.
2. $a_i = \alpha_i e_i$ and $b_i = \beta_i e_i \quad \forall i$, where e_i is the i th column of an (n, n) identity matrix, and α_i and β_i , $i = 1, \dots, n$, are scalars.
3. $\varphi_{ij} = 0 \quad \forall i \neq j$.
4. $\varphi_{ij} = 1 \quad \forall i \neq j$.
5. $A = \alpha(\omega\lambda')$ and $B = \beta(\omega\lambda')$ where $A = \{a_i\}_{i=1}^n$, $B = \{b_i\}_{i=1}^n$, ω and λ are $(n, 1)$ vectors and α and β are scalars.

The GDC model reduces to several multivariate GARCH models under different combinations of these conditions. Specifically, the GDC model becomes a restricted vech model (with the restrictions $\beta_{ij} = \beta_{ii}\beta_{jj}$ and $\alpha_{ij} = \alpha_{ii}\alpha_{jj}$) under conditions (i) and (ii), the constant conditional correlation model under conditions (ii) ad (iii), the BEKK model under conditions (i), and (iv) and the F-GARCH model under conditions (i), (iv), and (v).

Kroner and Ng (1998) also develop an extension to the GDC that allows for asymmetry. The asymmetric dynamic covariance (ADC) model has the same structure as the GDC model, except that the equation for $\theta_{ij,t}$ incorporates the leverage effect in the BEKK model, in a way close to the GJR model

$$\theta_{ij,t} = \omega_{ij} + a_i' \varepsilon_{t-1} \varepsilon_{t-1}' a_i + b_i \Sigma_{t-1} b_i' + g_i' \eta_{t-1} \eta_{t-1}' g_i \quad \forall i, j,$$

where $\eta_t = (\eta_{1,t}, \dots, \eta_{n,t})'$ with $\eta_{i,t} = \max(-\varepsilon_{i,t}, 0)$ and g_i , $i = 1, \dots, n$, are scalars.

6.1.4 Estimation issues

Maximum likelihood

We suppose now that we have a sample of size T of the $(n, 1)$ vector of observations written as $\underline{x}_T = \{x_t\}_{t=1}^T$ with conditional mean and conditional variance given by (6.1) and (6.2). Unknown parameters are regrouped in θ . Under the assumption of conditional multivariate normal distribution, the log-likelihood function for \underline{x}_T is

$$L_T(\theta | \underline{x}_T) = \sum_{t=1}^T \ell_t(\theta),$$

with

$$\ell_t(\theta) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} (\log |\Sigma_t(\theta)| + (x_t - \mu_t(\theta))' \Sigma_t^{-1}(\theta) (x_t - \mu_t(\theta))), \quad (6.9)$$

where the term $-\frac{1}{2} \log |\Sigma_t(\theta)|$ comes from the Jacobian of the transformation from the innovation process to the observed variables.

Then, the ML estimator $\hat{\theta}_{ML}$, that maximizes (6.9), is asymptotically normal with asymptotic distribution

$$\sqrt{T} (\hat{\theta}_{ML} - \theta_0) \Rightarrow \mathcal{N}(0, A_0^{-1}),$$

where A_0 is the information matrix evaluated at the true parameter vector θ_0 . See Section 4.3.3 for further details on the construction and estimation of A_0 .

Estimation of the DCC model

For the DCC model, Engle (2002) and Engle and Sheppard (2001) have also proposed a two-step estimation, based on the idea that parameters of the conditional variances (denoted θ_V) and of the conditional correlations (denoted θ_C) can be estimated separately with $\theta = (\theta_V', \theta_C')'$.⁶ A justification is

⁶ Bollerslev (1990) proposed a similar two-step approach for the estimation of the CCC model. We present here the case of the DCC model, because it nests the CCC model.

that the log-likelihood can be written as the sum of a volatility part and a correlation part. Since $\Sigma_t = D_t^{1/2} R_t D_t^{1/2}$, we have

$$\log |\Sigma_t| = \log |D_t| + \log |R_t|,$$

because D_t is diagonal, and

$$\begin{aligned} (x_t - \mu_t)' \Sigma_t^{-1} (x_t - \mu_t) &= (x_t - \mu_t)' \left(D_t^{1/2} R_t D_t^{1/2} \right)^{-1} (x_t - \mu_t) \\ &= u_t' R_t^{-1} u_t + (x_t - \mu_t)' D_t^{-1} (x_t - \mu_t) - u_t' u_t, \end{aligned}$$

where $u_t = D_t^{-1/2} \varepsilon_t$ is a vector of normalized innovations. The last two terms of the second equality are clearly equal, but this expression allows one to break down the log-likelihood in the two following terms

$$\ell(\theta_V, \theta_C | \underline{x}_T) = \ell_V(\theta_V | \underline{x}_T) + \ell_C(\theta_V, \theta_C | \underline{x}_T),$$

with

$$\begin{aligned} \ell_V(\theta_V | \underline{x}_T) &= -\frac{1}{2} \sum_{t=1}^T \left[\frac{n}{2} \log(2\pi) + \log |D_t| + (x_t - \mu_t)' D_t^{-1} (x_t - \mu_t) \right] \\ &= -\sum_{i=1}^n \left[\frac{T}{2} \log(2\pi) + \frac{1}{2} \sum_{t=1}^T \left(\log(\sigma_{i,t}^2) + \left(\frac{x_{it} - \mu_{i,t}}{\sigma_{i,t}} \right)^2 \right) \right], \\ \ell_C(\theta_V, \theta_C | \underline{x}_T) &= -\frac{1}{2} \sum_{t=1}^T (\log |R_t| + u_t' R_t^{-1} u_t - u_t' u_t). \end{aligned}$$

Notice that $\ell_V(\theta_V | \underline{x}_T)$ is simply the sum of log-likelihoods of the individual GARCH equations for each series. The second step consists in estimating the parameters pertaining to the correlation matrix, conditionally on the parameters estimated in the first stage.

Therefore, since squared residuals are not dependent on correlation parameters, these parameters can be ignored for the estimation of the conditional volatility dynamics. The two-step estimation then relies on maximizing the log-likelihood as follows. First, we estimate the volatility parameters through

$$\hat{\theta}_V \in \arg \max_{\{\theta_V\}} \ell_V(\theta_V | x_t, t = 1, \dots, T),$$

and then

$$\hat{\theta}_C \in \arg \max_{\{\theta_C\}} \ell_C(\hat{\theta}_V, \theta_C | \underline{x}_T) = -\frac{1}{2} \sum_{t=1}^T (\log |R_t| + \hat{u}_t' R_t^{-1} \hat{u}_t - \hat{u}_t' \hat{u}_t),$$

where $\hat{u}_{i,t} = (r_{i,t} - \hat{\mu}_{i,t}) / \hat{\sigma}_{i,t}$.

Engle and Sheppard (2001) show that the two-step estimator $\hat{\theta}_{TS} = (\hat{\theta}'_V, \hat{\theta}'_C)'$ is consistent and asymptotically normal, with distribution

$$\sqrt{T} (\hat{\theta}_{TS} - \theta_0) \Rightarrow \mathcal{N} (0, A_0^{-1} B_0 A_0'^{-1}),$$

where

$$A_0 = \begin{bmatrix} \frac{\partial^2 \ell_V(\theta_{V0})}{\partial \theta_V \partial \theta'_V} & 0 \\ \frac{\partial^2 \ell_C(\theta_0)}{\partial \theta_V \partial \theta'_C} & \frac{\partial^2 \ell_C(\theta_0)}{\partial \theta_C \partial \theta'_C} \end{bmatrix},$$

$$B_0 = V \left[\frac{1}{\sqrt{T}} \sum_{t=1}^T \frac{\partial \ell'_V(\theta_{V0})}{\partial \theta_V}, \frac{1}{\sqrt{T}} \sum_{t=1}^T \frac{\partial \ell'_C(\theta_0)}{\partial \theta_C} \right].$$

Due to the structure of A_0 , the asymptotic variances of the GARCH parameters $\hat{\theta}_V$ for each series are the standard robust covariance matrix estimators. For the second-stage parameters, however, the asymptotic variance involves all parameters.

Since the two sets of parameters are estimated using limited information, the estimator $\hat{\theta}_{TS}$ is not fully efficient. However, as argued by Pagan (1986), if we perform an additional iteration of the Newton-Raphson algorithm to the log-likelihood, starting at $\hat{\theta}_{TS}$, then we obtain an asymptotically efficient estimator.

6.1.5 Specification tests

Due to its computational burden, it is important to test whether a multivariate GARCH model is able to fit the data correctly. A first preliminary test would be to evaluate the ability of univariate GARCH models to correctly describe the data for each series. However, such an approach cannot be recommended, because it does not take into account the possible correlation between the series, which is actually the main characteristic of multivariate models. Therefore, multivariate tests are needed.

We do not focus in this section on the detection of non-normality in a multivariate GARCH context. The issue of dealing with non-normal distributions is addressed in the next section devoted to the modeling of multivariate models with non-normal distributions. We thus consider tests of GARCH effects, or alternatively tests of serial correlation in square standardized residuals. There are basically two types of tests. The first series of tests is based on the properties of standardized residuals, and the second test is a Portmanteau test.

Residual-based statistics

A natural diagnostic test for the multivariate GARCH model, proposed by Tse (2002), is based on the regression of cross-products of the standardized residuals $\hat{z}_{i,t} \hat{z}_{j,t}$ on own lags as follows

$$\hat{z}_{i,t} \hat{z}_{j,t} - \hat{\rho}_{ij,t} = \sum_{k=1}^p \alpha_{ij}^k (\hat{z}_{i,t-k} \hat{z}_{j,t-k} - \hat{\rho}_{ij,t-k}) + \nu_{ij,t},$$

for $i, j = 1, \dots, n$, $j > i$. The choice of explanatory variables in these regressions may theoretically differ from one exercise to the other, depending on the type of dependency we want to test. If we denote $d_{ij,t} = (\hat{z}_{i,t-1} \hat{z}_{j,t-1}, \dots, \hat{z}_{i,t-p} \hat{z}_{j,t-p})'$, the test statistic is

$$RB(p) = T \hat{\alpha}'_{ij} \hat{L}_{ij} \hat{\Omega}_{ij}^{-1} \hat{L}_{ij} \hat{\alpha}_{ij},$$

with

$$\begin{aligned} L_{ij} &= \text{plim}_{T \rightarrow \infty} \left(\frac{1}{T} \sum_{t=1}^T d_{ij,t} d'_{ij,t} \right), \\ \Omega_{ij} &= E \left[(z_{i,t} z_{j,t} - \rho_{ij,t})^2 \right] L_{ij} - Q_{ij} G Q'_{ij}, \\ Q_{ij} &= \text{plim}_{T \rightarrow \infty} \left(\frac{1}{T} \sum_{t=1}^T d_{ij,t} \frac{\partial (z_{i,t} z_{j,t} - \rho_{ij,t})}{\partial \theta'} \right), \end{aligned}$$

where θ is the vector of parameters in the GARCH model and G is the asymptotic covariance matrix of θ , such that $\sqrt{T}(\hat{\theta} - \theta) \Rightarrow \mathcal{N}(0, G)$. Tse (2002) shows that, under the null hypothesis that the specification is correct, the statistic $RB(p)$ is asymptotically distributed as a χ^2 with p degrees of freedom.

Evidently, a similar test can be performed for squared standardized residuals $\hat{z}_{i,t}^2$, using the following regression

$$\hat{z}_{i,t}^2 - 1 = \sum_{k=1}^p \alpha_{ij}^k (\hat{z}_{i,t-k}^2 - 1) + \nu_{i,t} \quad \text{for } i = 1, \dots, n.$$

Portmanteau statistics

The Ljung-Box portmanteau test for serial correlation has been extended to a multivariate context by Baillie and Bollerslev (1990) (see also Hosking, 1980). It is written as

$$H(p) = T^2 \sum_{i=1}^p \left(\frac{1}{T-i} \right) \text{tr} \left(\hat{C}_i \hat{C}_0^{-1} \hat{C}_i \hat{C}_0^{-1} \right),$$

where $\hat{C}_i = \frac{1}{T} \sum_{t=i+1}^T (x_t - \bar{x})(x_{t-i} - \bar{x})'$ is the sample autocovariance matrix of order i of x_t . Under the null hypothesis of constant correlation, the statistic $H(p)$ is distributed as a $\chi^2(n^2 p)$.

This test statistic may be used to test the presence of ARCH effects in the squared returns. In this case, we define the $n(n+1)/2$ vector of cross-products as $y_t = \text{vech}((x_t - \bar{x})(x_t - \bar{x})')$. Then, the test statistic $H(p)$ is

estimated with $\hat{C}_i = \frac{1}{T} \sum_{t=i+1}^T (y_t - \bar{y})(y_{t-i} - \bar{y})'$ as a sample autocovariance matrix for the covariances. Under the null hypothesis of no ARCH effect, the test statistic $H(p)$ is asymptotically distributed as a $\chi^2(n^2 p)$.

6.1.6 Test of constant conditional correlation matrix

Due to its computational simplicity, the CCC–GARCH model is very popular among empirical researchers. However, there are several problems that seem to be overlooked in empirical applications. First, the assumption of constant correlation is often taken for granted and seldom analyzed or tested. Second, the issue of how the assumption of a constant conditional correlation affects the dynamics of the conditional variance is rarely considered.

Alternative tests can be based on a specific parametric specification of the conditional correlation. In this case, however, implementing the test procedure is more demanding, because we have to estimate, or at least to specify, the dynamics of correlation. Therefore, we loose the main advantage of the constant correlation GARCH model, for which the sample correlation matrix is a consistent estimator of the conditional correlation matrix. Several tests have been proposed that rely on different dynamics of the correlation matrix under the alternative. The dynamic correlation models presented above evidently provide a suitable setup for testing the constant correlation hypothesis.⁷ In the following, we focus on tests that do not require the estimation of a complete GARCH model with timescale correlations.

Test based on the information matrix

Bera and Kim (2002) suggested an Information Matrix (IM) test for the constant-correlation hypothesis in a bivariate GARCH model. The basic idea is to derive a score test of the hypothesis that the variances of the parameters of interest are 0, so that it does not require the explicit specification of an alternative hypothesis. A definite advantage of this approach is thus that the test statistic does not depend on a particular specification of correlation variations.

Under the null hypothesis of constant correlation, the conditional covariance matrix Σ_t can be written as $\Sigma_t = D_t^{1/2} R D_t^{1/2}$ with

$$D_t = \begin{pmatrix} \sigma_{1,t}^2 & 0 \\ 0 & \sigma_{2,t}^2 \end{pmatrix}, \quad \text{and} \quad R = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

The matrix Σ_t is positive definite for all t , if each of the conditional variances $\sigma_{i,t}$ are positive, and if the conditional correlation matrix is positive definite, i.e., $|\rho| < 1$.

⁷ For instance, in the model by Tse and Tsui (2002), the null of constant correlation simply corresponds to $\theta_1 = \theta_2 = 0$.

The conditional log-likelihood is then defined as

$$\ell(\theta) = \frac{1}{T} \sum_{t=1}^T \ell_t(\theta),$$

with θ the vector of unknown parameters and

$$\begin{aligned}\ell_t(\theta) &= -\log(2\pi) - \frac{1}{2} \log |\Sigma_t| - \frac{1}{2} \varepsilon'_t \Sigma_t^{-1} \varepsilon_t \\ &= -\log(2\pi) - \frac{1}{2} \log(1 - \rho^2) - \frac{1}{2} u'_t R^{-1} u_t,\end{aligned}$$

with $u_t = D_t^{1/2} \varepsilon_t$ denoting normalized innovations.

Then, first and second derivatives with respect to ρ are

$$\begin{aligned}\frac{\partial \ell_t(\theta)}{\partial \rho} &= \frac{1}{1 - \rho^2} (v_{1,t} v_{2,t} + \rho), \\ \frac{\partial^2 \ell_t(\theta)}{\partial \rho^2} &= \frac{1}{(1 - \rho^2)^2} (-v_{1,t}^2 - v_{2,t}^2 + 2\rho v_{1,t} v_{2,t} + 1 + \rho^2),\end{aligned}$$

where $v_t = (u_{1,t} - \rho u_{2,t}, u_{2,t} - \rho u_{1,t}) / \sqrt{1 - \rho^2}$. This implies that the score is

$$\begin{aligned}s_t(\theta) &= \left(\frac{\partial \ell_t(\theta)}{\partial \rho} \right)^2 + \frac{\partial^2 \ell_t(\theta)}{\partial \rho^2} \\ &= \frac{1}{(1 - \rho^2)^2} (v_{1,t}^2 v_{2,t}^2 + 4\rho v_{1,t} v_{2,t} - v_{1,t}^2 - v_{2,t}^2 + 1 + 2\rho^2).\end{aligned}$$

Suppose now that we have the ML estimator $\hat{\theta}$ of the parameter vector, so that $\hat{v}_{i,t}$ is the estimate of $v_{i,t}$ with θ replaced by $\hat{\theta}$. In particular, we have $\hat{\rho} = \frac{1}{T} \sum_{t=1}^T \hat{u}_{1,t} \hat{u}_{2,t}$. Note that

$$\frac{1}{T} \sum_{t=1}^T \hat{v}_{i,t} \hat{v}_{j,t} = \begin{cases} 1 & \text{if } i = j, \\ -\hat{\rho} & \text{if } i \neq j. \end{cases}$$

Therefore, we obtain the indicator $d(\hat{\theta})$ in the IM test for the constancy of ρ_t , when $\hat{\theta}$ is the MLE of θ , as

$$d(\hat{\theta}) = (1 - \rho^2)^2 \sum_{t=1}^T s_t(\hat{\theta}) = \frac{1}{T} \sum_{t=1}^T (\hat{v}_{1,t}^2 \hat{v}_{2,t}^2 - 1 - 2\hat{\rho}^2).$$

Bera and Kim (2002) also show that the asymptotic variance of $d(\hat{\theta})$ is

$$V \left[\sqrt{T} d(\hat{\theta}) \right] = 4 (1 + 4\rho^2 + \rho^4),$$

which can be consistently estimated by substituting the MLE $\hat{\rho}$ in place of ρ .

Therefore the test statistic is

$$IMC = \frac{T d(\hat{\theta})^2}{\hat{V} [\sqrt{T} d(\hat{\theta})]} = \frac{\left(\sum_{t=1}^T (\hat{v}_{1,t}^2 \hat{v}_{2,t}^2 - 1 - 2\hat{\rho}^2) \right)^2}{4T (1 + 4\hat{\rho}^2 + \hat{\rho}^4)}.$$

It is asymptotically distributed as a $\chi^2(1)$ under the null of constant correlation.

Test based on the LM statistic

Tse (2000) proposed a test for the constant-correlation hypothesis based on the LM approach. The idea is to extend the constant-correlation model to one that includes time-varying correlations. When certain parameters in the extended model are imposed to be zero, the constant-correlation model is obtained. The extension proposed by Tse is simply

$$\begin{aligned}\rho_{ij,t} &= \rho_{ij} + \delta_{ij} \varepsilon_{i,t-1} \varepsilon_{j,t-1}, \\ \sigma_{ij,t} &= \rho_{ij,t} \sigma_{i,t} \sigma_{j,t},\end{aligned}$$

where δ_{ij} are additional parameters.

The constant-correlation hypothesis can be tested by examining the hypothesis $H_0 : \delta_{ij} = 0$ for $1 \leq i < j \leq 1$. Under the null hypothesis, there are $n(n-1)/2$ independent restrictions.

To ensure that the alternative model provides well-defined positive definite conditional covariance matrices, further restrictions have to be imposed on the parameters δ_{ij} . It is assumed that within a neighborhood of $\delta_{ij} = 0$, the optimal properties of the LM test hold under some regularity conditions.⁸

We denote θ the vector of unknown parameters, including δ_{ij} , $1 \leq i < j \leq n$. We define the $(n, 1)$ score vector

$$s = \frac{\partial \ell(\theta)}{\partial \theta} = \sum_{t=1}^T \frac{\partial \ell_t(\theta)}{\partial \theta},$$

and the (n, n) information matrix

$$V = E \left[-\frac{\partial^2 \ell(\theta)}{\partial \theta \partial \theta'} \right].$$

⁸ An alternative specification is to allow correlation to depend on the products of the lagged standardized residuals

$$\rho_{ij,t} = \rho_{ij} + \delta_{ij} z_{i,t-1} z_{j,t-1}.$$

In this case, however, because $z_{i,t}$ depends on other parameters of the model through $\sigma_{i,t}$, an analytic derivation of the LM statistic is intractable.

Then, the LM statistic for the null of constant correlation is $\hat{s}'\hat{V}^{-1}\hat{s}$, where the hats denote evaluation at $\hat{\theta}$ under the null hypothesis. Tse suggests to replace V by the sum of cross-products of the first derivatives of ℓ_t , so that

$$LMC = \hat{s}' (\hat{S}'\hat{S})^{-1} \hat{s},$$

where \hat{S} is the (T, n) matrix the rows of which are the partial derivatives of $\partial\ell_t(\theta)/\partial\theta'$. Under the regularity conditions, LMC is asymptotically distributed as a $\chi^2(n(n-1)/2)$.

Monte-Carlo simulations have been performed by Tse (2000) to compare the two test statistics IMC and LMC . Although both test statistics appear to be correctly sized under the null, the LM test is found to have better power against some alternatives and to be more robust to non-normality. It should be noticed, however, that the alternatives considered by Tse (2000) are close to the specification he proposes to model time-variability in correlations. It is unclear which statistic would perform best for alternatives that are different from those envisaged by Tse (2000).

6.1.7 Illustration

Several papers have investigated the variability of the dependency parameter over time. Hamao, Masulis, and Ng (1990), Susmel and Engle (1994), and Bekaert and Harvey (1995) measured the interdependence of returns and volatilities across stock markets. More specifically, Longin and Solnik (1995) tested the hypothesis of a constant conditional correlation between a large number of stock markets. They found that correlation generally increases in periods of high volatility of the U.S. market. Recent contributions by Kroner and Ng (1998), as well as Engle and Sheppard (2001) develop GARCH models capable of estimating and testing hypotheses of time-varying covariance matrices. Ang and Chen (2002) document that dependency between U.S. stocks and the aggregate U.S. market increases more during downside movements than during upside movements.

To illustrate some properties of the multivariate GARCH models described above, we estimate several specifications for two pairs of time series: the SP500 and DAX daily returns, and the SP500 and FT-SE returns, over the period from 1980 to 2004. We begin with first-generation models. We estimate several versions of the bivariate BEKK model, the full, diagonal, and scalar versions. Parameter estimates and log-likelihoods of these models are reported in Table 6.1. Parameter estimates indicate that, in the full BEKK model, cross-effects (α_{12} , α_{21} , β_{12} and β_{21}) are in general rather small and insignificant. This provides a rationale for the diagonal BEKK model. Now all parameters are significant. As already highlighted for the univariate GARCH models, the sum of parameters (here, $\alpha_{11}^2 + \beta_{11}^2$ and $\alpha_{22}^2 + \beta_{22}^2$, due to the structure of the model) is close to (yet smaller than) one. In addition, we notice that α_{11} and α_{22} on

one hand and β_{11} and β_{22} on the other hand are rather close, suggesting a further reduction of the number of parameters, assuming $\alpha_{11} = \alpha_{22} = \alpha$ and $\beta_{11} = \beta_{22} = \beta$. The so-called scalar BEKK model assumes that the variances and covariances have the same structure and the same parameters α and β

$$\sigma_{ij,t} = \omega_{ij} + \alpha^2 \varepsilon_{i,t-1} \varepsilon_{j,t-1} + \beta^2 \sigma_{ij,t-1}.$$

Now, we compare the dynamics of correlations given by first-generation models with those obtained with models specifically designed to capture time-varying correlation, namely the Engle's (2002) DCC model and the Tse and Tsui's (2002) TVC model. Table 6.2 reports parameter estimates corresponding to the CCC, DCC, and TVC models. First, they are estimated using the two-step approach proposed in Section 6.1.3. Therefore, the parameter estimates for the SP500 are obviously the same whatever the second return of the pair. Second, the parameters α and β in this table can be interpreted as the square of the parameters in the previous table. The dynamics of variances are very close to those obtained with BEKK models. The difference relies in the estimation of the covariances / correlations. In the BEKK models, covariances are rather smoothed, whereas correlations are very erratic (see Figures 6.1 and 6.3). With the CCC model, the correlation parameter is constant by definition. In contrast, the DCC and TVC models provide very similar and smoothed dynamics for the correlation parameter.

Figures 6.1 and 6.3 display the conditional correlation between the SP500 and the DAX, and between the SP500 and the FT-SE, respectively, as implied by the full BEKK model. As it appears clearly, these conditional correlations are very erratic. Figures 6.2 and 6.4 display the conditional correlation obtained with the CCC, DCC, and TVC models. First, we notice that the conditional correlations are now much smoother than those obtained with BEKK models. Second, we observe that the correlations implied by the DCC and TVC models are barely distinguishable from each other. Therefore, in many empirical applications, the direct modeling of the conditional correlation (through a DCC or a TVC model) is likely to provide a rather sensible estimate of the evolution in the correlation parameter.

The multivariate GARCH approach described so far assumes that the joint distribution of innovations is normal. Obviously, it is not likely to be the case in practice, because the univariate distribution of returns has generally been found to be non-normal. We therefore turn now to the modeling of the joint distribution. More precisely, we are interested in identifying some multivariate distributions that may help capturing the asymmetry and fat-tailedness of the return distribution.

Table 6.1. Parameter estimates of various specifications of the BEKK model

	SP500–DAX		SP500–FT-SE	
	Estimate	Std error	Estimate	Std error
Full BEKK model				
ω_{11}	0.0745	(0.0185)	0.0602	(0.0428)
ω_{12}	-0.0081	(0.0251)	-0.1099	(0.1213)
ω_{22}	0.1603	(0.0244)	0.1215	(0.0715)
a_{11}	0.1818	(0.0204)	0.1408	(0.0367)
a_{12}	-0.0319	(0.0628)	-0.2017	(0.0655)
a_{21}	0.0258	(0.0185)	0.0842	(0.0497)
a_{22}	0.2803	(0.0264)	0.3078	(0.0347)
b_{11}	0.9805	(0.0049)	0.9882	(0.0089)
b_{12}	0.0105	(0.0097)	0.0531	(0.0230)
b_{21}	-0.0059	(0.0045)	-0.0223	(0.0187)
b_{22}	0.9524	(0.0079)	0.9109	(0.0298)
log-lik.	-18113.3	—	-15983.0	—
Diagonal BEKK model				
ω_{11}	0.0863	(0.0139)	0.0796	(0.0144)
ω_{12}	0.028	(0.0111)	0.0395	(0.0200)
ω_{22}	0.1505	(0.0229)	0.1170	(0.0248)
a_{11}	0.1996	(0.0168)	0.2063	(0.0234)
a_{22}	0.2620	(0.0223)	0.2381	(0.0490)
b_{11}	0.9761	(0.0040)	0.9758	(0.0051)
b_{22}	0.9582	(0.0062)	0.9617	(0.0171)
log-lik.	-18120.3	—	-16100.2	—
Scalar BEKK model				
ω_{11}	0.1083	(0.0124)	0.0982	(0.0187)
ω_{12}	0.0215	(0.0089)	0.026	(0.0064)
ω_{22}	0.134	(0.0193)	0.0911	(0.0116)
a	0.2326	(0.0159)	0.2186	(0.0177)
b	0.9669	(0.0043)	0.9711	(0.0056)
log-lik.	-18131.1	—	-16106.3	—

Table 6.2. Parameter estimates of the CCC, DCC, and TVC models

	SP500–DAX		SP500–FT-SE	
	Estimate	Std error	Estimate	Std error
CCC model				
ω_1	0.0066	(0.0025)	0.0066	(0.0025)
a_1	0.0459	(0.0082)	0.0459	(0.0082)
b_1	0.9480	(0.0094)	0.9480	(0.0094)
ω_2	0.0296	(0.0113)	0.0201	(0.0059)
a_2	0.1019	(0.0221)	0.0894	(0.0135)
b_2	0.8834	(0.0213)	0.8875	(0.0168)
ρ	0.2576	(0.0174)	0.3058	(0.0258)
log-lik.	-18144.5	—	-16047.1	—
DCC model				
ω_1	0.0066	(0.0025)	0.0066	(0.0025)
a_1	0.0459	(0.0082)	0.0459	(0.0082)
b_1	0.9480	(0.0094)	0.9480	(0.0094)
ω_2	0.0296	(0.0113)	0.0201	(0.0058)
a_2	0.1019	(0.0221)	0.0894	(0.0135)
b_2	0.8834	(0.0213)	0.8875	(0.0167)
δ_1	0.0079	(0.0026)	0.0073	(0.0030)
δ_2	0.9909	(0.0032)	0.9867	(0.0055)
log-lik.	-18027.1	—	-16030.1	—
TVC model				
ω_1	0.0066	(0.0025)	0.0066	(0.0025)
a_1	0.0459	(0.0082)	0.0459	(0.0082)
b_1	0.9480	(0.0094)	0.9480	(0.0094)
ω_2	0.0296	(0.0113)	0.0201	(0.0058)
a_2	0.1019	(0.0221)	0.0894	(0.0135)
b_2	0.8834	(0.0213)	0.8875	(0.0168)
θ_1	0.0084	(0.0026)	0.0068	(0.0031)
θ_2	0.9909	(0.0029)	0.9896	(0.0055)
log-lik.	-18028.2	—	-16022.1	—

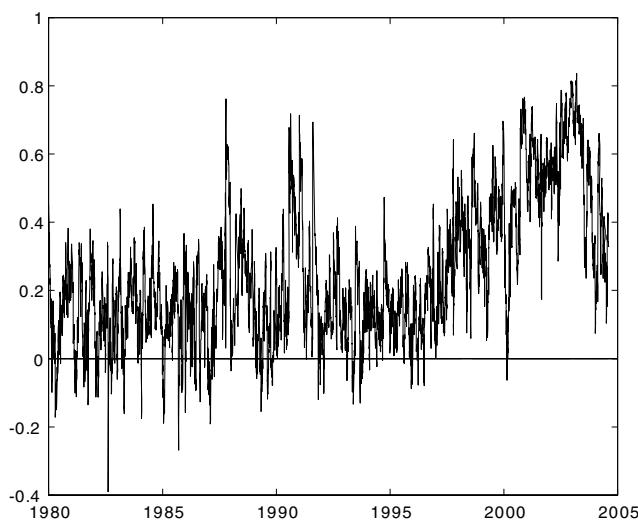


Fig. 6.1. *SP500–DAX. Conditional correlation implied by the full BEKK model.*

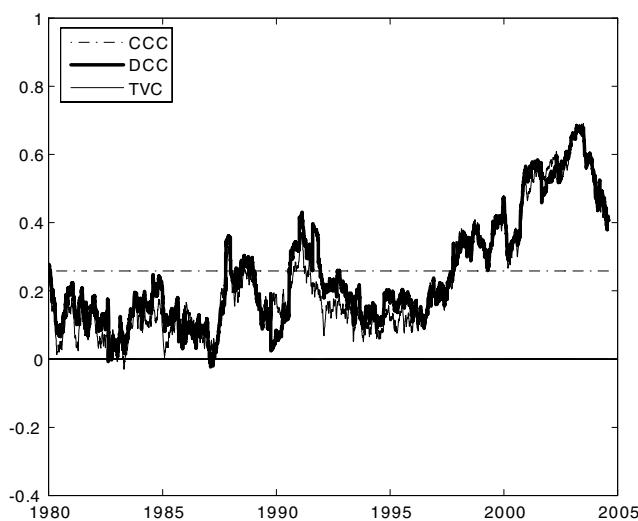


Fig. 6.2. *SP500–DAX. Conditional correlation implied by the CCC, DCC and TVC models.*

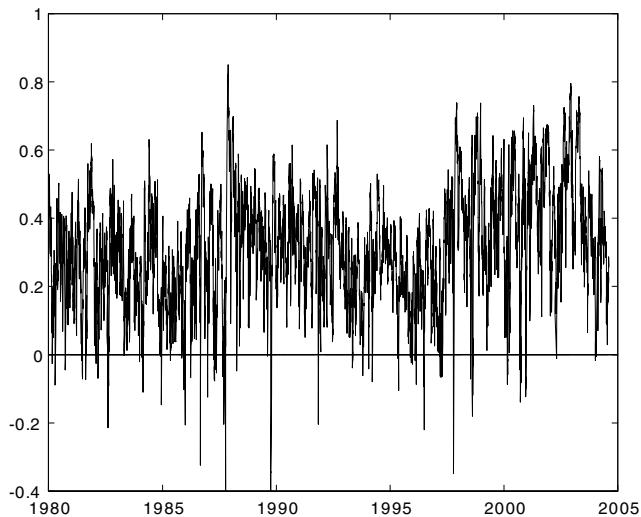


Fig. 6.3. SP500–FT-SE. Conditional correlation implied by the full BEKK model.

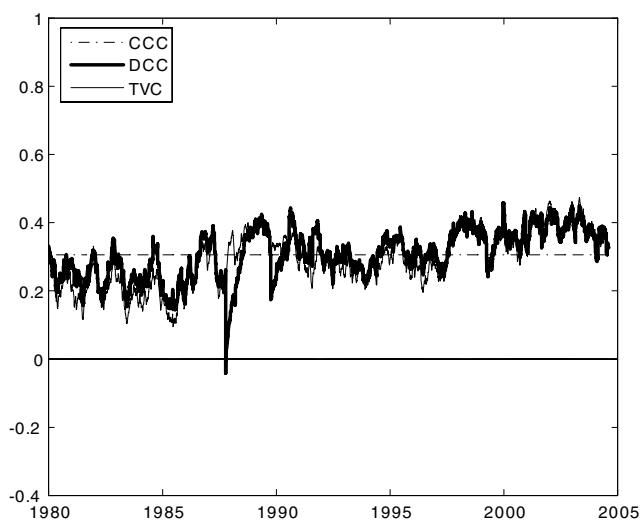


Fig. 6.4. SP500–FT-SE. Conditional correlation implied by the CCC, DCC, and TVC models.

6.2 Modeling the multivariate distribution

We now assume that the dynamic of the time-series $x_t = (x_{1,t}, \dots, x_{n,t})'$ is given by the following multivariate dynamic regression model, with time-varying means, variances, and covariances

$$x_t = \mu_t + \varepsilon_t, \quad (6.10)$$

$$\varepsilon_t = \Sigma_t^{1/2} z_t, \quad (6.11)$$

$$\mu_t = E[x_t | \mathcal{F}_{t-1}] = \mu(\theta, \mathcal{F}_{t-1}), \quad (6.12)$$

$$\Sigma_t = V[x_t | \mathcal{F}_{t-1}] = \Sigma(\theta, \mathcal{F}_{t-1}), \quad (6.13)$$

$$z_t \sim g(z_t | \eta). \quad (6.14)$$

The dynamics of the $(n, 1)$ conditional mean vector μ_t is given by (6.12), whereas the dynamics of the (n, n) conditional variance matrix Σ_t is given by (6.13). Last, standardized residuals z_t , defined as $\Sigma_t^{-1/2}(x_t - \mu_t)$, is the *iid* random vector of dimension $(n, 1)$ with a zero mean and identity variance matrix. There are several possibilities to obtain $\Sigma_t^{1/2}$. The first and probably most common one is the Cholesky decomposition, where $\Sigma_t^{1/2}$ is a lower triangular matrix. Another possibility is based on the eigenvector decomposition, $\Sigma_t = \Omega_t D_t \Omega_t$, where Ω_t is the matrix of eigenvectors, standardized to unit length, and D_t is the diagonal matrix of eigenvalues. By construction, we have $\Omega_t \Omega_t = I_t$. Then, $\Sigma_t^{1/2} = \Omega_t D_t^{1/2}$, where $D_t^{1/2}$ is the matrix whose diagonal elements are the square roots of the eigenvalues. Vector θ includes all parameters of the conditional mean and variance equations. As specified in (6.14), the conditional distribution is g , with shape parameters η .

As in the univariate case, the choice of the conditional distribution $g(\cdot)$ is crucial. Engle and González-Rivera (1991) and Newey and Steigerwald (1997) have shown that the following results hold: (i) Under the assumption of a correct specification of the conditional mean and variance matrix, the ML estimation, assuming z_t to be *iid* with a Gaussian distribution, provides consistent estimators, even when the Gaussian assumption does not hold. (ii) The ML estimator relying on a Gaussian distribution is inefficient, however, with the degree of inefficiency increasing with the degree of departure from normality. (iii) The ML estimation, assuming z_t to be *iid* with a non-Gaussian distribution, provides more efficient estimators than the Gaussian ML, when the assumption made on the innovation process holds. (iv) When the assumption made on the innovation process does not hold, the ML estimation relying on a non-Gaussian distribution provides inconsistent estimators.

Another difficulty in the multivariate case comes from the way dependency between variables is introduced. In the Gaussian framework, dependency is introduced through the covariance matrix. The multivariate Gaussian distribution with zero mean and identity covariance matrix is defined as

$$g(z_t) = \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2} z_t' z_t\right).$$

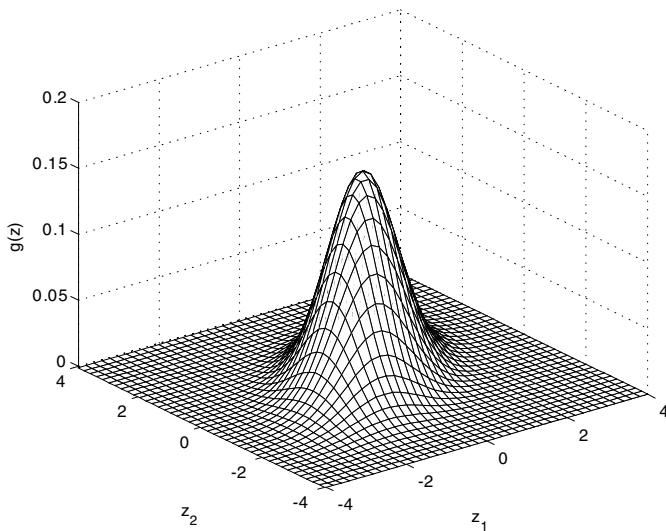


Fig. 6.5. Pdf of the bivariate Gaussian distribution, with $\rho = 0.5$.

Therefore, the Gaussian distribution of the observations x_t with conditional mean μ_t and covariance matrix Σ_t is given by

$$f(x_t) = \frac{1}{(2\pi)^{n/2}} \frac{1}{|\Sigma_t|^{1/2}} \exp\left(-\frac{1}{2} (x_t - \mu_t)' \Sigma_t^{-1} (x_t - \mu_t)\right).$$

Given that the Gaussian distribution only depends on μ_t and Σ_t , the natural measure of dependency is Σ_t . For the general case, dependency may be introduced in various ways.

Figure 6.5 shows the *pdf* of the bivariate Gaussian distribution. The two variables are supposed to have a zero mean and unit variance. The correlation is 0.5. The contour of the distribution is displayed in Figure 6.6.

One of the strong limitations of the multivariate Gaussian distribution is that it does not allow any dependence between the two variables in the tails (for more details, see Section 6.3), a feature that has been found to play a central role in the joint modeling of asset returns, especially in the context of VaR applications. In addition, as seen in the univariate context, the Gaussian distribution is symmetric, so that it is unable to capture the observed asymmetry in the return distribution.

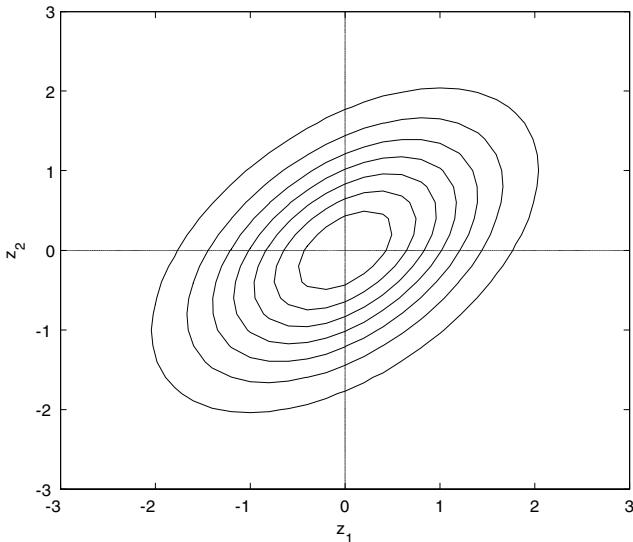


Fig. 6.6. Contour of the pdf of the bivariate Gaussian distribution, with $\rho = 0.5$.

6.2.1 Standard multivariate distributions

Multivariate Student *t* distribution

The multivariate Student *t* distribution is a distribution that naturally generates dependency in the tails. Yet, this extension to the Student *t* distribution depends on the way in which dependency between innovations is introduced. Assume that the n innovations are expressed as $Z_i = U_i / \sqrt{S_i^2 \nu_i}$, $i = 1, \dots, n$, where the U_i s and S_i s are all mutually independent, with U_i having a univariate Gaussian distribution, and S_i^2 a χ^2 distribution with ν_i degrees of freedom. In this case, because the Z_i s are independent, the joint density function is simply the product of the individual Student *t* density functions.

Now, there are basically two ways to introduce dependence into the joint distribution of returns X :

First, dependence can be introduced in the innovation process Z , by assuming that the χ^2 -variables S_i^2 that appear in the definition of the Z_i s are the same for each component, with ν degrees of freedom, so that $Z_i = U_i / \sqrt{S^2 \nu}$, $i = 1, \dots, n$. We may also assume that U_1, \dots, U_n have a joint multivariate distribution with covariance matrix R . Since the U_i s are normalized, $R = \{\rho_{ij}\}$ is also the correlation matrix of (U_1, \dots, U_n) . In this case, the multivariate Student *t* distribution, with dependent components and degree-of-freedom parameter ν , is

$$g(z|\nu) = \frac{\Gamma(\frac{\nu+n}{2})}{(\pi(\nu-2))^{n/2} \Gamma(\frac{\nu}{2}) |R|^{1/2}} \left(1 + \frac{1}{\nu-2} z' R^{-1} z\right)^{-\frac{\nu+n}{2}}.$$

When the U_i s are mutually independent ($R = I_n$), the distribution reduces to

$$g(z|\nu) = \frac{\Gamma(\frac{\nu+n}{2})}{(\pi(\nu-2))^{n/2} \Gamma(\frac{\nu}{2})} \left(1 + \frac{z' z}{\nu-2}\right)^{-\frac{\nu+n}{2}}. \quad (6.15)$$

Each marginal distribution is symmetric, with zero mean, unit variance (this is obtained by using the terms $(\nu-2)$ rather than ν in 6.15), zero skewness, and a kurtosis equal to $3(\nu-2)/(\nu-4)$. When the Student t distribution is introduced to capture fat-tailedness of the series, however, there is neither theoretical nor empirical reason for the degree-of-freedom parameters to be equal for all series. Assuming a unique degree of freedom ν for all components would therefore introduce an artificial dependence in the system.

Alternatively, the Gaussian variables U_i and the χ^2 -distributed variables S_i^2 can be assumed to be independent components, with dependence introduced for the X s via some linear structure involving a Choleski decomposition of the covariance matrix. If the degrees of freedom ν_i are different for each variable, the Student t distribution with independent components is written as

$$g(z|\nu_1, \dots, \nu_n) = \prod_{i=1}^n \frac{\Gamma(\frac{\nu_i+1}{2})}{\sqrt{\pi(\nu_i-2)} \Gamma(\frac{\nu_i}{2})} \left(1 + \frac{z_i^2}{\nu_i-2}\right)^{-\frac{\nu_i+1}{2}}, \quad (6.16)$$

while the multivariate return process X is assumed to be correlated through a non-diagonal covariance matrix Σ , such that the distribution of X , denoted $f(x|\nu_1, \dots, \nu_n)$, is deduced from the relation $X = \Sigma^{1/2} Z$.

Now, each marginal distribution is symmetric, with zero mean, unit variance, zero skewness, and a kurtosis equal to $3(\nu_i-2)/(\nu_i-4)$. Clearly, in empirical applications, the equality of the ν_i s can be explicitly tested in a second step. An undesirable property of this distribution is that it is not a member of the elliptical family described below (Section 6.2.1). As a consequence, as described in Section 6.2.4, in a multivariate GARCH model with skewed Student t innovations, the estimation of the GARCH parameters and of the shape parameters cannot be performed separately.

Figures 6.7 and 6.9 shows the *pdf* of the two types of bivariate Student t distributions. The first one corresponds to the distribution with dependent components (equation (6.15)) with degree of freedom $\nu = 6$ and correlation between the z_i s equal to $\rho_{12} = 0.5$. The second is the distribution with independent components (equation (6.16)) with degrees of freedom $\nu_1 = \nu_2 = 6$ and correlation between the x_i s equal to $\rho_{12} = 0.5$. In the two cases, the two variables are supposed to have a zero mean and unit variance. The contours of the two distributions are displayed in Figures 6.8 and 6.10 respectively.

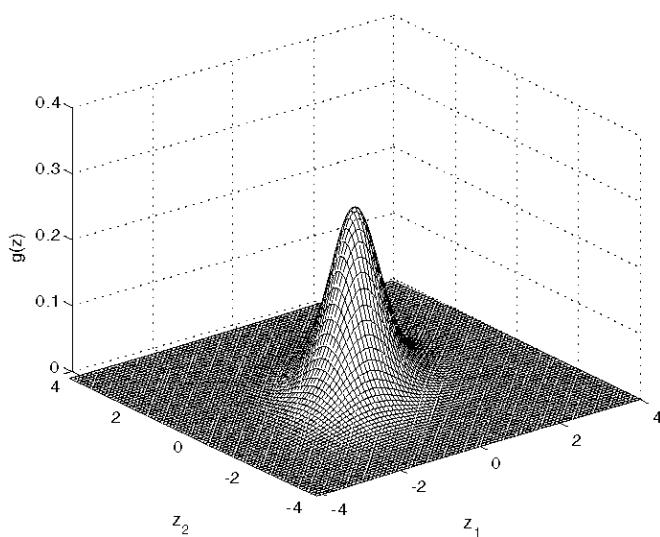


Fig. 6.7. Pdf of the Student t distribution with dependent components and $\rho_{12} = 0.5$.

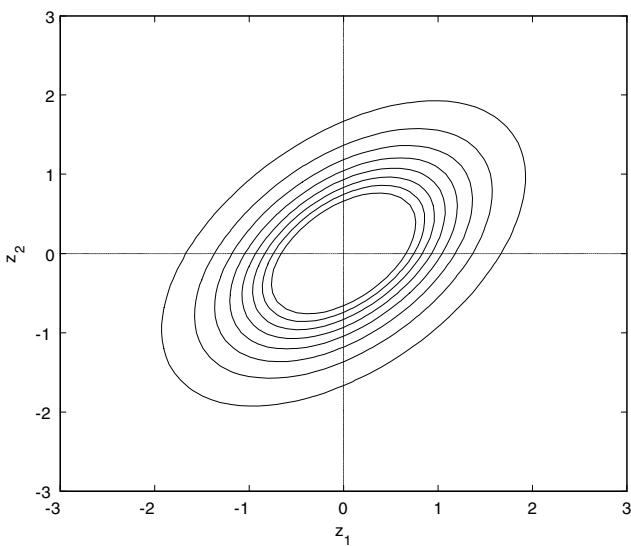


Fig. 6.8. Contour of the pdf of the Student t distribution with dependent components and $\rho_{12} = 0.5$.

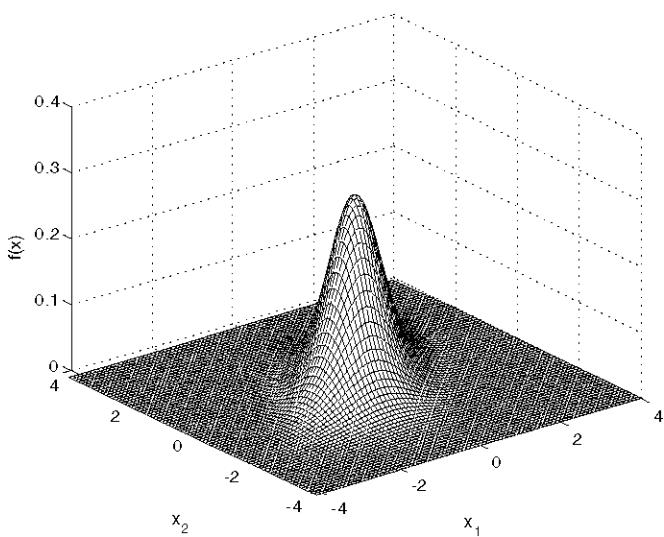


Fig. 6.9. Pdf of the Student t distribution with independent components and $\rho_{12} = 0.5$.

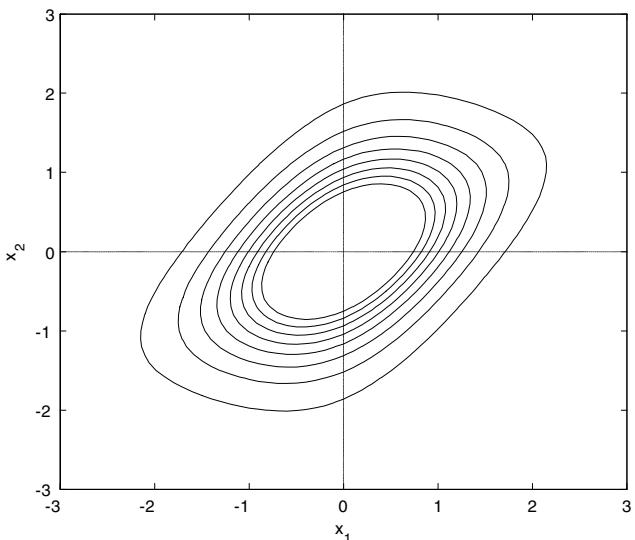


Fig. 6.10. Contour of the pdf of the Student t distribution with independent components and $\rho_{12} = 0.5$.

Multivariate Student t distributions are able to capture the fat-tailedness of the empirical distribution of asset returns. However, they are not designed to take asymmetry into account. Before turning to asymmetric distributions, we briefly describe the class of multivariate elliptical distributions, which includes the Gaussian and Student t distributions. Indeed, a series of papers have proposed a general class of multivariate distributions with asymmetry, based on elliptical distributions (see Section 6.2.2).

Elliptical distributions

A well-studied class of multivariate distributions is the class of elliptical distributions. For more detail, see Fang, Kotz, and Ng (1990). An n -dimensional vector Z is said to be elliptically distributed with location vector μ and (n, n) dispersion matrix Σ , if the density is

$$g(z | \mu, \Sigma) = |\Sigma|^{-1/2} f^{(n)}((z - \mu)' \Sigma^{-1} (z - \mu)), \quad (6.17)$$

for some density generating function $f^{(n)}(u)$, $u \geq 0$, such that

$$\int_0^\infty u^{n/2-1} f^{(n)}(u) du = \frac{\Gamma(n/2)}{\pi^{n/2}},$$

so that $f^{(n)}$ is a spherical n -dimensional density. We denote this function $Z \sim El_n(\mu; \Sigma; f^{(n)})$ with pdf $g_{f^{(n)}}(.)$ and cdf $G_{f^{(n)}}(.)$.

Elliptical distributions have several interesting properties. In particular, if $Z \sim El_n(\mu; \Sigma; f^{(n)})$, then for any (k, n) matrix A with rank $k \leq n$ and any $(k, 1)$ vector b , we have $AZ + b \sim El_k(A\mu + b; A\Sigma A'; f^{(k)})$. Another property is that if Z is partitioned as

$$Z = \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \sim El_n \left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}; \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}; f^{(n)} \right),$$

with Z_i an $(n_i, 1)$ vector (with $n = n_1 + n_2$), then Z_i is also elliptically distributed with $Z_i \sim El_{n_i}(\mu_i; \Sigma_i; f^{(n_i)})$.

Instances of elliptical distribution are the multivariate Gaussian distribution, for which

$$f^{(n)}(u) = \frac{e^{-u/2}}{(2\pi)^{n/2}},$$

so that

$$g(z | \mu, \Sigma) = \frac{1}{(2\pi)^{n/2}} |\Sigma|^{-1/2} \exp \left(-\frac{1}{2} (z - \mu)' \Sigma^{-1} (z - \mu) \right).$$

Similarly, the multivariate Student t distribution is defined by

$$f^{(n)}(u) = \left(1 + \frac{u}{\nu}\right)^{-(\nu+n)/2} \quad \nu > 0,$$

so that

$$g(z|\mu, \Sigma) = \frac{\Gamma(\frac{\nu+n}{2})}{(\pi\nu)^{n/2} \Gamma(\frac{\nu}{2})} |\Sigma|^{-1/2} \left(1 + \frac{(z - \mu)' \Sigma^{-1} (z - \mu)}{\nu}\right)^{-\frac{\nu+n}{2}}.$$

Notice that this specification slightly differs from (6.15). As explained above, (6.15) ensures that the variance of Z is by construction equal to one, while here we have $V[Z] = \nu / (\nu - 2)$.

6.2.2 Skewed elliptical distribution

The multivariate skewed Gaussian distribution has been first studied by Azzalini and Dalla Valle (1996), Azzalini and Capitanio (1999), and extended to the Student t distribution by Azzalini and Capitanio (2003). Branco and Dey (2001) have introduced a general class of multivariate skew-elliptical distributions.

This approach builds on the univariate skewed Gaussian distribution, defined by Azzalini (1985). A random variable Z has a so-called *skewed Gaussian distribution* if its *pdf* is defined by

$$g(z) = 2\varphi(z)\Phi(\lambda z),$$

where $\varphi(z)$ is the standard Gaussian *pdf*, and $\Phi(z)$ is the corresponding *cdf*. Some intuition for this specification is provided below. When $\lambda = 0$, $g(z)$ is the standard normal distribution. When $\lambda \rightarrow -\infty (+\infty)$, we obtain a Gaussian distribution truncated from above (below) at zero. The parameter λ plays the role of shape parameter.

Skewed Gaussian distribution

This distribution has been extended in several ways. In particular, Azzalini and Dalla Valle (1996) have considered the case of the multivariate skewed Gaussian distribution. Branco and Dey (2001) have extended this approach to the elliptical distribution in a multivariate framework.

The multivariate extension of the skewed Gaussian distribution has been proposed by Azzalini and Dalla Valle (1996), and Azzalini and Capitanio (1999). If Z is a $(n, 1)$ random vector with mean zero and, unit variances and correlation matrix R , its *pdf* is defined by

$$g(z) = 2\varphi_n(z|R)\Phi(\lambda'z),$$

where $\varphi_n(z|R)$ is the multivariate Gaussian *pdf* with covariance matrix R , and λ is a $(n, 1)$ vector.

When the random vector Z is not assumed to have zero means and unit variances, we can define the skewed Gaussian distribution as follows. Let

$\Sigma = \{\sigma_{ij}\}$ be the (n, n) covariance matrix of Z , $D = \text{diag}(\sigma_{11}, \dots, \sigma_{nn})$ the matrix containing variances, and $R = D^{-1/2} \Sigma D^{-1/2}$ the associated correlation matrix. Let ξ be the $(n, 1)$ location parameter vector. Then, the random variable Z has a skew normal distribution if its *pdf* may be written as

$$g(z) = 2\varphi_n(z - \xi|\Sigma) \Phi\left(\lambda' D^{-1/2}(z - \xi)\right).$$

We then write $Z \sim \mathcal{SN}_n(\xi, \Sigma, \lambda)$, referring to ξ , Σ , and λ as the location, dispersion, and shape (or skewness) parameters, respectively.

Such a distribution can be derived in several ways, described by Azzalini and Capitanio (1999). A first way of generating a skewed Gaussian distribution is *conditioning*. Suppose that U_0 is a scalar random variable and U is an n -dimensional variable, such that the joint distribution is a multivariate Gaussian distribution

$$\begin{pmatrix} U_0 \\ U \end{pmatrix} \sim \mathcal{N}_{n+1}(0, \Sigma^*) \quad \text{with} \quad \Sigma^* = \begin{pmatrix} 1 & \delta' \\ \delta & R \end{pmatrix},$$

where Σ^* is a full-rank covariance matrix. Then, the distribution of $U | U_0 > 0$ is $\mathcal{SN}_n(0, R, \lambda)$, where λ is defined as

$$\lambda' = \frac{\delta' R^{-1}}{(1 - \delta' R^{-1} \delta)^{1/2}}.$$

Alternatively, the random variable defined by

$$Z = \begin{cases} U & \text{if } U_0 > 0, \\ -U & \text{if } U_0 < 0, \end{cases}$$

is also distributed as a skewed Gaussian variate.

The skewed Gaussian distribution can also be obtained by *transformation*. Suppose that

$$\begin{pmatrix} U_0 \\ U \end{pmatrix} \sim \mathcal{N}_{n+1}(0, \Sigma^*) \quad \text{with} \quad \Sigma^* = \begin{pmatrix} 1 & 0 \\ 0 & \Psi \end{pmatrix},$$

where Ψ is a full-rank covariance matrix. Also define the $(n, 1)$ vector

$$Z_j = \delta_j |U_0| + \sqrt{1 - \delta_j^2} U'_j,$$

with weights given by $-1 < \delta_j < 1$ for $j = 1, \dots, n$. Then, (Z_1, \dots, Z_n) has an n -dimensional skewed Gaussian distribution with parameters that are functions of the δ 's and Ψ .

Skewed elliptical distribution

Branco and Dey (2001), Sahu, Dey, and Branco (2003), and Azzalini and Capitanio (2003) have proposed a general class of multivariate skewed elliptical

distributions, based on the approach developed by Azzalini and Dalla Valle (1996) and Azzalini and Capitanio (1999). The standard multivariate elliptical distribution is defined in Section 6.2.1.

Suppose that U_0 is a scalar random variable and U is a n -dimensional variable, such that the joint distribution is a multivariate elliptical distribution

$$U^* = \begin{pmatrix} U_0 \\ U \end{pmatrix} \rightarrow El_{n+1} \left(\mu^*, \Sigma^*; f^{(n+1)} \right),$$

with

$$\mu^* = \begin{pmatrix} 0 \\ \mu \end{pmatrix} \quad \text{and} \quad \Sigma^* = \begin{pmatrix} 1 & \delta' \\ \delta & \Sigma \end{pmatrix},$$

where Σ^* is a full-rank covariance matrix. Then, $Y = U | U_0 > 0$ has a skewed elliptical distribution, such that $Y \sim SE_n(\mu, \Sigma, \delta; f^{(n+1)})$, with location μ , scale Σ , characteristic function φ , and skewness parameter δ . The *pdf* of Y is denoted

$$g_Y(y) = 2g_{f^{(n)}}(y) G_{f_{q(y)}}(\lambda'(y - \mu)),$$

where $g_{f^{(n)}}(\cdot)$ is the *pdf* of $El_n(\mu, \Sigma; f^{(n)})$ with generator function $f^{(n)}(\cdot)$ and $G_{f_{q(z)}}(\cdot)$ is the *cdf* of a univariate elliptical distribution $El_1(0, 1; f_{q(z)})$ with $f_{q(z)}$ as the generator function. In addition, λ is defined as

$$\lambda' = \frac{\delta' \Sigma^{-1}}{(1 - \delta' \Sigma^{-1} \delta)^{1/2}},$$

and

$$f_{q(y)}(u) = \frac{f^{(n+1)}(u + q(y))}{f^{(n)}(q(y))},$$

with $q(y) = (y - \mu)' \Sigma^{-1} (y - \mu)$.

A special case is the skewed Student t distribution. The *pdf* of Y can be written as

$$g_Y(y) = 2t_n(y, \nu) T_1 \left(\delta' D^{-1/2} (y - \mu) \left(\frac{\nu + n}{q(y) + \nu} \right)^{1/2}; \nu + n \right),$$

where

$$t_n(y, \nu) = \frac{1}{|\Sigma|^{1/2}} g_n(q(y); \nu)$$

is the density function of a n -dimensional Student t variate with ν degrees of freedom and $T_1(x; \nu + n)$ denotes the univariate Student t distribution function with $\nu + n$ degrees of freedom.

6.2.3 Skewed Student t distribution

There are other ways to generate multivariate skewed distributions. For instance, Bauwens and Laurent (2005) have generalized the approach proposed by Fernández and Steel (1998) in the univariate context, presented in Section 5.2.4. The method consists in changing the shape of the distribution at each side of the mode. If $h(y)$ is a symmetric multivariate distribution with zero mean and identity covariance matrix, the new $(n, 1)$ random vector Z has an asymmetric distribution with the same mode as $h(y)$

$$g(z|\xi) = \left(\prod_{i=1}^n \frac{2}{\xi_i + \frac{1}{\xi_i}} \right) h(y),$$

where $y = (y_1, \dots, y_n)'$, with

$$y_i = \begin{cases} z_i^* \xi_i & \text{if } z_i < 0, \\ z_i^*/\xi_i & \text{if } z_i \geq 0. \end{cases} \quad (6.18)$$

Such a specification has several interesting properties. First, the marginal densities have the same patterns as the ones defined in the univariate case by Fernández and Steel (1998) (see Section 5.2.4). Second, ξ_i is a measure of the asymmetry of the marginal density of Z_i^* . Third, the r th moment of Z_i^* is given by

$$E[(Z_i^*)^r | \xi_i] = \frac{\xi_i^{r+1} + \frac{(-1)^r}{\xi_i^{r+1}}}{\xi_i + \frac{1}{\xi_i}} 2E[(Z_i^*)^r | Z_i > 0].$$

Finally, the components of Z^* are uncorrelated, because those of Y are uncorrelated by assumption. Consequently, if a is the vector of means and b the vector of standard deviations of Z^* , then Z^* can be standardized by the transformation $Z = (Z^* - a) \div b$, where \div denotes the element-wise division.

This general formulation has been specialized by Bauwens and Laurent (2005) and Jondeau and Rockinger (2005) to the case of the multivariate skewed Student t distribution. This distribution is therefore a multivariate extension of the distribution proposed by Hansen (1994), Fernández and Steel (1998), and Jondeau and Rockinger (2003a), and is presented in Section 5.2.4.⁹

The case with dependent components

Following the two approaches described in Section 6.2.1, we now define the multivariate distribution assuming dependent or independent components. We begin with the specification with dependent components, which is given as

⁹ As discussed in Section 5.2, the skewed t distribution proposed by Fernández and Steel (1998) is directly related to the distribution proposed by Hansen (1994) through a change of notation of the asymmetry parameter. In this section, we use the notation of Fernández and Steel (1998) for the asymmetry parameter.

$$g(z|\nu, \xi_1, \dots, \xi_n) = c \left(\prod_{i=1}^n \frac{2b_i}{\xi_i + \frac{1}{\xi_i}} \right) \left(1 + \frac{y'y}{\nu - 2} \right)^{-\frac{\nu+n}{2}}, \quad (6.19)$$

where $y = (y_1, \dots, y_n)'$ is defined in (6.18) with $z_i^* = (b_i z_i + a_i)$, and

$$\begin{aligned} c &= \frac{\Gamma(\frac{\nu+n}{2})}{(\pi(\nu-2))^{n/2} \Gamma(\frac{\nu}{2})}, \\ a_i &= \frac{\Gamma(\frac{\nu-1}{2}) \sqrt{\nu-2}}{\sqrt{\pi} \Gamma(\frac{\nu}{2})} \left(\xi_i - \frac{1}{\xi_i} \right), \\ b_i^2 &= \left(\xi_i^2 + \frac{1}{\xi_i^2} - 1 \right) - a_i^2. \end{aligned}$$

Notice that the standardized variables $Z_i = (Z_i^* - a_i)/b_i$ have zero mean and unit variance. This formulation is an extension of the multivariate Student t distribution with dependent components, defined in (6.15).

The case with independent components

We now describe the asymmetric extension of the multivariate Student t distribution with independent components, defined in (6.16). It is expressed as

$$g(z|\nu_1, \dots, \nu_n, \xi_1, \dots, \xi_n) = \prod_{i=1}^n \frac{2b_i c_i}{\xi_i + \frac{1}{\xi_i}} \left(1 + \frac{y_i^2}{\nu_i - 2} \right)^{-\frac{\nu_i+1}{2}}, \quad (6.20)$$

where y_i is given by (6.18) with $z_i^* = (b_i z_i + a_i)$, and

$$\begin{aligned} c_i &= \frac{\Gamma(\frac{\nu_i+1}{2})}{\sqrt{\pi(\nu_i-2)} \Gamma(\frac{\nu_i}{2})}, \\ a_i &= \frac{\Gamma(\frac{\nu_i-1}{2}) \sqrt{\nu_i-2}}{\sqrt{\pi} \Gamma(\frac{\nu_i}{2})} \left(\xi_i - \frac{1}{\xi_i} \right), \\ b_i^2 &= \left(\xi_i^2 + \frac{1}{\xi_i^2} - 1 \right) - a_i^2. \end{aligned}$$

It is worth emphasizing that, although the standardized variables $Z_i = (Z_i^* - a_i)/b_i$ are independent by construction, returns can be modeled as dependent through their covariance matrix. Assume that returns are defined as $X = \Sigma^{1/2} Z^*$. The moments of returns can, therefore, be computed in the following way. We denote $\Sigma^{1/2} = (\omega_{ij})_{i,j=1,\dots,n}$ the Choleski decomposition of the covariance matrix of returns, so that $X_i = \sum_{r=1}^n \omega_{ir} Z_r^*$.¹⁰ In addition,

¹⁰ For ease of exposition, we assume that $E[X] = \mu = 0$ and we intentionally omit the dependence of the covariance matrix Σ and hence of the ω_{ij} s with respect to time.

we recall that because innovations Z^* are assumed to be independent, their first moments are given by $\mu_i^{(1)} = 0$, $\mu_i^{(2)} = 1$ and (see Section 5.2.4)

$$\begin{aligned}\mu_i^{(3)} &= (M_{i,3} - 3a_i M_{i,2} + 2a_i^3) / b_i^3, \\ \mu_i^{(4)} &= (M_{i,4} - 4a_i M_{i,3} + 6a_i^2 M_{i,2} - 3a_i^4) / b_i^4.\end{aligned}$$

The (i, j) component of the covariance matrix of returns is given by

$$E [X_i X_j] = \sum_{r=1}^n \omega_{ir} \omega_{jr}.$$

The (i, j, k) component of the third central moments of returns is

$$E [X_i X_j X_k] = \sum_{r=1}^n \omega_{ir} \omega_{jr} \omega_{kr} \mu_r^{(3)}.$$

Finally, the (i, j, k, l) component of the fourth central moments of returns is

$$E [X_i X_j X_k X_l] = \sum_{r=1}^n \omega_{ir} \omega_{jr} \omega_{kr} \omega_{lr} \mu_r^{(4)} + \sum_{r=1}^n \sum_{\substack{s=1 \\ s \neq r}}^n \psi_{rs} \mu_r^{(2)} \mu_s^{(2)},$$

where $\psi_{rs} = \omega_{ir} \omega_{jr} \omega_{ks} \omega_{ls} + \omega_{ir} \omega_{js} \omega_{kr} \omega_{ls} + \omega_{ir} \omega_{js} \omega_{kr} \omega_{ls}$.

Analytical expressions of moments are quite cumbersome to derive, yet their numerical computation is very fast, because only matrix manipulations are required. For an n -variable system, the dimension of the covariance matrix is (n, n) , but only $n(n+1)/2$ elements have to be computed. Similarly, the co-skewness matrix has dimension (n, n, n) , but only $n(n+1)(n+2)/6$ elements have to be computed. Finally, the co-kurtosis matrix has dimension (n, n, n, n) , but only $n(n+1)(n+2)(n+3)/24$ elements have to be computed.¹¹

As it clearly appears from these expressions, co-skewness between excess returns depends on individual skewness of innovations and correlations between returns (through the ω_{ij} s). Co-kurtosis between excess returns depends on individual kurtosis and volatilities of innovations and correlations between returns. Such a time-variability is likely to have two sources. On one hand, the covariance matrix between excess returns is time-varying, so that the ω_{ij} elements themselves are time-varying. On the other hand, skewness and kurtosis of innovations may be time-varying, for instance, if the degree-of-freedom parameter (ν) or the asymmetry parameter (ξ) vary over time. In our illustration (Section 6.2.6), we do not consider the latter case, but persistence in conditional higher moments has been found, in a univariate context, by Jondeau and Rockinger (2003a).

¹¹ For $n = 5$, one has 15 different elements for the covariance matrix, 35 elements for the co-skewness matrix, and 70 elements for the co-kurtosis matrix (whereas these matrices have 25, 125, and 625 elements, respectively).

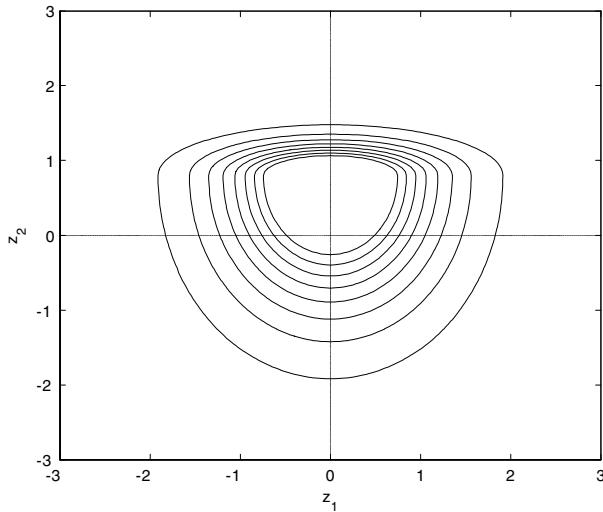


Fig. 6.11. Contour plot of the bivariate skewed Student t distribution with dependent components and $\rho = 0$.

Comparison

Figures 6.11 to 6.13 show the *pdf* and the contour plot of the two types of bivariate skewed Student t distributions (6.19) and (6.20). The two variables have a zero mean and unit variance. The degree-of-freedom parameter ν is equal to 6 in all cases. The asymmetry parameters are equal to $\xi = (1; 0.5)$ so that the first component is symmetric but the second one is markedly asymmetric. In Figure 6.11, the correlation coefficient ρ_{12} is assumed to be 0, whereas in Figures 6.12 and 6.13, it is equal to 0.5.

6.2.4 Estimation

The estimation of the multivariate model with non-Gaussian multivariate distribution, given by (6.10)–(6.14), raises some additional difficulties as compared with the Gaussian case. Unknown parameters are now θ (for the dynamics of the conditional mean and conditional covariance matrix) and η (for the conditional distribution). Define $\xi = (\theta', \eta')'$ the vector of unknown parameters. The ML estimator of the parameter vector ξ is obtained by maximizing

$$L_T(\xi | \underline{x}_T) = \sum_{t=1}^T \ell_t(\xi),$$

where

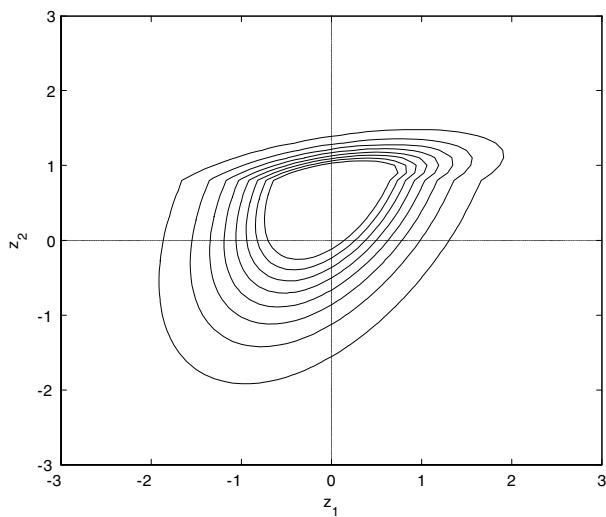


Fig. 6.12. Contour plot of the bivariate skewed Student t distribution with dependent components and $\rho = 0.5$.

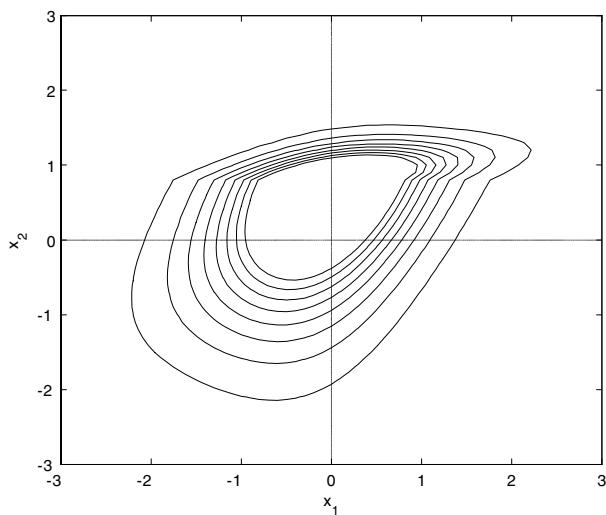


Fig. 6.13. Contour plot of the bivariate skewed Student t distribution with independent components and $\rho = 0.5$.

$$\ell_t(\xi) = -\frac{1}{2} \log |\Sigma_t(\theta)| + g\left(\Sigma_t(\theta)^{-1/2}(x_t - \mu_t(\theta))|\xi\right).$$

In general, the maximization of the log-likelihood has to be performed in one step, because the two sets of parameters θ and η interact.

A noticeable exception is the class of elliptical distributions. In this case, the distribution only involves $(x_t - \mu_t(\theta))' \Sigma_t^{-1}(\theta) (x_t - \mu_t(\theta))$, which can be rewritten as $z_t(\theta)' z_t(\theta)$ with $z_t(\theta) = \Sigma_t(\theta)^{-1/2}(x_t - \mu_t(\theta))$. Therefore, a two-step estimation can be performed, in which the two sets of parameters are estimated separately. The parameters θ are estimated using QML, assuming normality of the innovation process. Therefore, it is obtained by maximizing

$$L_T(\theta|\underline{x}_T) = \sum_{t=1}^T \ell_t(\theta),$$

where

$$\ell_t(\theta) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} (\log |\Sigma_t(\theta)| + (x_t - \mu_t(\theta))' \Sigma_t^{-1}(\theta) (x_t - \mu_t(\theta))).$$

This procedure yields a consistent estimator of θ even if the true distribution is not normal. This consistency result has been proven by Bollerslev and Wooldridge (1992) and Jeantheau (1998). Once the parameter vector $\hat{\theta}$ has been estimated, the standardized residuals are obtained as $\hat{z}_t(\hat{\theta}) = \hat{\Sigma}_t(\hat{\theta})^{-1/2}(x_t - \hat{\mu}_t(\hat{\theta}))$. Then, parameters η are estimated by maximizing the log-likelihood function, assuming now that innovations are drawn from the multivariate distribution $g(z_t|\eta)$

$$\tilde{L}_T(\hat{\theta}, \eta|\underline{x}_T) = \sum_{t=1}^T \hat{\ell}_t(\hat{\theta}, \eta),$$

where

$$\hat{\ell}_t(\hat{\theta}, \eta) = -\frac{1}{2} \log |\Sigma_t(\hat{\theta})| - \frac{1}{2} \sum_{t=1}^T \log(g(\hat{z}_t(\hat{\theta}))|\eta).$$

Such an approach can be adopted for the estimation of GARCH models with Gaussian or Student t innovations. However, it would not apply for the skewed Student t distribution, because it does not belong to the elliptical family. In such case, a joint estimation is required.¹² See Section 6.1.4 for further details on the estimation of the covariance matrix of $\hat{\theta}$.

Hafner and Rombouts (2003) have recently proposed an extension of the semi-parametric estimation technique of Engle and González-Rivera (1991) to

¹² Notice that, in the case of the Student t distribution, parameters θ and η can be estimated separately. However, if we estimate a DCC model, it is not possible to estimate the parameters pertaining to the variances and to the correlations separately anymore. The reason is that the log-likelihood cannot be broken down into variances and correlations components.

multivariate GARCH models. As in the univariate framework (Section 8.3), this approach is based on the following steps. The first step consists in estimating the model by QML, providing an estimate of the parameters θ , say $\tilde{\theta}$. Then, the fitted residuals $\hat{\varepsilon}_t$ and the fitted variances $\hat{\sigma}_t^2(\tilde{\theta})$ are used to compute the standardized residuals $\hat{z}_t(\tilde{\theta}) = \hat{\varepsilon}_t / \hat{\sigma}_t(\tilde{\theta})$, which should have zero mean and unit variance. In a third step, the density $g(\hat{z}_t(\tilde{\theta}))$ is estimated using a non-parametric method. The estimated density is denoted \hat{g} . Finally, the parameters of the GARCH model are estimated by maximizing the log-likelihood function, with \hat{g} held as fixed.

6.2.5 Adequacy tests

In Section 5.3, we described some tests aimed at assessing the ability of a given univariate distribution to capture the stylized facts of the empirical distribution. Most of these tests have been extended to the multivariate context. For instance, Diebold, Hahn, and Tay (1999) proposed an extension of the test introduced by Diebold, Gunter, and Tay (1998). The basic idea consists in writing the joint distribution $g_t(z_{1,t}, \dots, z_{n,t})$ as the product of conditional distributions, as in

$$g_t(z_{1,t}, \dots, z_{n,t}) = g_t(z_{n,t}|z_{n-1,t}, \dots, z_{1,t}) \times \dots \times g_t(z_{2,t}|z_{1,t}) \times g_t(z_{1,t}).$$

Then, at each period, it is possible to transform each component of the vector $(z_{1,t}, \dots, z_{n,t})'$ by its corresponding conditional distribution. We then obtain n series of $u_{i,t} = \int_{-\infty}^{z_{i,t}} g_t(y_{i,t}|y_{i-1,t}, \dots, y_{1,t}) dy_{i,t}$ that should be found to be *iid* $U(0, 1)$, both individually and jointly, if the model is correct.

One difficulty occurs in the multivariate case, because there are $n!$ ways to factor the joint distribution in terms of conditional distributions. For instance, in the bivariate case, we can write the joint distribution as

$$g_t(z_{1,t}, z_{2,t}) = g_t(z_{2,t}|z_{1,t}) \times g_t(z_{1,t}),$$

or

$$g_t(z_{1,t}, z_{2,t}) = g_t(z_{1,t}|z_{2,t}) \times g_t(z_{2,t}).$$

The first decomposition would provide us with probability integral transforms $u_{1,t}$ and $u_{2|1,t}$ and the second would provide $u_{2,t}$ and $u_{1|2,t}$, where we denote for instance

$$u_{2|1,t} = \int_{-\infty}^{z_{2,t}} g_t(y_{2,t}|y_{1,t}) dy_{2,t}.$$

Finally, we have to test if u_1 , $u_{2|1}$, u_2 and $u_{1|2}$ are each *iid* $U(0, 1)$ and if $(u_1, u_{2|1})$ and $(u_2, u_{1|2})$ are also *iid* $U(0, 1)$. Consequently, for large numbers of variables, implementing the adequacy test may be quite cumbersome.

6.2.6 Illustration

To illustrate the working of the Student t and skewed Student t distributions in the context of multivariate GARCH models, we use once again two pairs of time series: the SP500 and DAX daily returns and the SP500 and FT-SE returns, over the period from January 1980 to August 2004. We consider the DCC model of Engle (2002), whose estimation under normality is reported in Table 6.2, and we estimate the same specification assuming Student t and skewed Student t innovations (with dependent components) in turn. Table 6.3 reports parameter estimates corresponding to these various models.¹³ The log-likelihood indicates that the model with Student t innovations very markedly dominates the model with Gaussian innovations (compare with Table 6.2). The degree-of-freedom parameter ν is in the same range of values as found previously, in the univariate context, in Section 5.4.

Introducing an asymmetric t distribution also helps improving the fit of the empirical joint distribution, although less significantly. This finding can be explained by the weak asymmetry in the distribution of the SP500 return. Such evidence has been already observed in the univariate estimation. Finally, we do not display the dynamics of the conditional correlation, because there is no noticeable difference with the one obtained under normality (Section 6.1).

6.3 Copula functions

In many situations where marginal distributions are not Gaussian, it is simply impossible to define a joint distribution. This is the case when we want to link two variables that have different marginal distributions (for instance, a Student t variate and a Pareto variate). This is also the case for a large number of marginal distributions, for which a multivariate extension does not exist. In such contexts, a solution is to use copula functions. These functions have the property to relate two marginal distributions instead of the two series directly. Therefore, once margins have been computed, no reference is made to their true functional form. This is the reason why copula functions are able to relate any kind of margin.

The textbooks for the analysis of copula functions are Joe (1997) and Nelsen (1999). Some surveys also provide valuable information on copulas, see in particular Riboulet, Roncalli, and Bouyé (2000), and Embrechts, Lindskog, and McNeil (2003). It should be noticed at this point that copula functions have been abundantly used to investigate the behavior of the tails of a multivariate distribution. This issue is addressed in detail in Section 7.2.

¹³ Notice that the parameter estimates of the SP500 obtained for the models with Student t and skewed Student t innovations differ for the two pairs of returns. The reason is that in such cases, two-step estimation is precluded, because the log-likelihood cannot be broken down into the variance part and the correlation part. However, the difference remains barely noticeable.

Table 6.3. Parameter estimates of the DCC model under various distributions

	SP500–DAX		SP500–FT-SE	
	Estimate	Std. err.	Estimate	Std. err.
DCC with Student t innovations				
ω_1	0.0058	(0.0017)	0.0061	(0.0018)
α_1	0.0407	(0.0053)	0.0414	(0.0060)
β_1	0.9527	(0.0062)	0.9512	(0.0070)
ω_2	0.0152	(0.0033)	0.0148	(0.0031)
α_2	0.0730	(0.0082)	0.0727	(0.0084)
β_2	0.9185	(0.0088)	0.9087	(0.0107)
ν	7.5728	(0.5586)	8.1771	(0.6486)
δ_1	0.0090	(0.0026)	0.0078	(0.0078)
δ_2	0.9898	(0.0031)	0.9848	(0.0188)
log-lik.	-17599.1	—	-15639.1	—
DCC with skewed Student t innovations				
ω_1	0.0058	(0.0019)	0.0062	(0.0020)
α_1	0.0409	(0.0056)	0.0417	(0.0064)
β_1	0.9524	(0.0068)	0.9509	(0.0077)
ω_2	0.0152	(0.0040)	0.0145	(0.0034)
α_2	0.0722	(0.0095)	0.0719	(0.0090)
β_2	0.9193	(0.0108)	0.9099	(0.0118)
ν	7.6718	(0.5665)	8.2845	(0.6580)
ξ_1	0.9790	(0.0141)	0.9882	(0.0139)
ξ_2	0.9412	(0.0156)	0.9180	(0.0159)
δ_1	0.0090	(0.0026)	0.0073	(0.0087)
δ_2	0.9898	(0.0032)	0.9858	(0.0213)
log-lik.	-17593.4	—	-15626.5	—

6.3.1 Definitions and properties

The study of copulas is quite a recent phenomenon in statistics. Hence, it is not astonishing that copulas have only recently found their way into empirical finance. In order to understand their usefulness, consider two random variables X and Y with marginal distributions, or *margins*, $F(x) = \Pr[X \leq x]$ and $G(y) = \Pr[Y \leq y]$. In this paper, we assume that the cumulative distribution functions (*cdf*) are continuous. The random variables may also have joint distribution function, $H(x, y) = \Pr[X \leq x, Y \leq y]$. All the distribution functions, $F(\cdot)$, $G(\cdot)$, and $H(\cdot, \cdot)$ have as range the interval $[0, 1]$. In some cases, a multivariate distribution exists, so that the function $H(\cdot, \cdot)$ has an explicit expression. One such case is the multivariate normal distribution. In many cases, however, a description of the margins $F(\cdot)$ and $G(\cdot)$ is relatively easy to obtain, whereas an explicit expression of the joint distribution $H(\cdot, \cdot)$ may be difficult to obtain. This is where copulas are useful because they link margins into a multivariate distribution function.

Definition 6.2. A bivariate copula is a function $C : [0, 1] \times [0, 1] \rightarrow [0, 1]$ with the three following properties:

1. $C(u, v)$ is increasing in u and v .
2. $C(0, v) = C(u, 0) = 0$, $C(1, v) = v$, $C(u, 1) = u$.
3. $C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0$, $\forall u_1, u_2, v_1, v_2$ in $[0, 1]$ such that $u_1 \leq u_2$ and $v_1 \leq v_2$.

Property 1 states that, when one marginal distribution is constant, the joint probability will increase provided that the other marginal distribution increases. Property 2 states that if one margin has zero probability to occur, then it must be the same for the joint occurrence. Also, if on the contrary, one margin is certain to occur, then the probability of a joint occurrence is determined by the remaining margin probability. Property 3 indicates that, if u and v both increase, then the joint probability also increases. This property is, therefore, a multivariate extension of the condition that a *cdf* is increasing.

Furthermore, if we set $u = F(x)$ and $v = G(y)$, then $C(F(x), G(y))$ yields a description of the joint distribution of X and Y . Having obtained this intuitive definition, we now propose the following important theorem, proven in Sklar (1959) and Schweizer and Sklar (1974).

Theorem 6.3 (Sklar's theorem). *Let F and G be the marginal distributions of X and Y , respectively, and let H be the joint distribution function of (X, Y) . Then, there exists a copula C such that, for all real numbers (x, y) ,*

$$H(x, y) = C(F(x), G(y)). \quad (6.21)$$

Furthermore, if F and G are continuous, then C is unique. Conversely, if F and G are the distributions of X and Y , respectively, then the function H defined by (6.21) is a joint distribution function with marginal distributions F and G .

The density of a copula is related to its *cdf* through the following relation

$$c(u, v) \equiv \frac{\partial^2 C(u, v)}{\partial u \partial v}.$$

Similarly, the density h of the distribution H is defined by the relationship

$$h(x, y) = c(F(x), G(y)) \times f(x) \times g(y).$$

Notice that many results developed in this section extend to a higher dimensional framework. Some of the results, however, hold in the bivariate framework only. In many cases, the ease of interpretation of the dependency parameter does not hold when there are more than two margins.

6.3.2 Measures of concordance

For a number of standard distributions (namely, the elliptical family, which includes the Gaussian and the Student t distributions), dependency is naturally

measured by Pearson's correlation coefficient. However, when other distributions are considered, alternative measures are needed to characterize the link between time series. Most of the results in this section are drawn from Nelsen (1999) and Embrechts, Lindskog, and McNeil (2003).

Following Nelsen (1999, p. 136), two pairs of random variables (X, Y) and (\tilde{X}, \tilde{Y}) are concordant if $X < \tilde{X}$ implies $Y < \tilde{Y}$ or if $X > \tilde{X}$ implies $Y > \tilde{Y}$ and discordant if $X < \tilde{X}$ implies $Y > \tilde{Y}$ or if $X > \tilde{X}$ implies $Y < \tilde{Y}$.

Kendall's tau for the random variables X and Y is defined as the probability of concordance minus the probability of discordance (or non-covariation) of two independent pairs of random variables (X, Y) and (\tilde{X}, \tilde{Y})

$$\tau[X, Y] = \Pr[(X - \tilde{X})(Y - \tilde{Y}) > 0] - \Pr[(X - \tilde{X})(Y - \tilde{Y}) < 0].$$

Spearman's rho for the random variables X and Y is defined as

$$\varrho_S[X, Y] = 3 \left(\Pr[(X - \tilde{X})(Y - Y') > 0] - \Pr[(X - \tilde{X})(Y - Y') < 0] \right),$$

where (X, Y) , (\tilde{X}, \tilde{Y}) , and (X', Y') are three independent copies. Since \tilde{X} and Y' are independent, Spearman's rho can be viewed as the distance between the distribution of X and Y and independence. Spearman's rho was first proposed in 1904 by the psychologist C. Spearman. Similar to Kendall's tau, it is related to the probabilities of concordance and discordance. The distinction is that one pair (X, Y) has distribution $H(x, y)$, and the second pair, say (\tilde{X}, \tilde{Y}) , is a pair of independent random variables with same margins as X and Y , meaning that (\tilde{X}, \tilde{Y}) has distribution function $F(x)G(y)$. To render this concept operational, it is convenient to consider three independent pairs of random variables (X, Y) , (\tilde{X}, \tilde{Y}) and (X', Y') each drawn from $H(x, y)$. The assumption that all pairs are independent means that \tilde{X} can be viewed as drawn from $F(x)$ and Y' from $G(y)$. The issue then is to see if (X, Y) and (\tilde{X}, Y') are concordant.

Theorem 6.4. (Schweizer and Wolff, 1981) Let X and Y be continuous random variables whose copula is C . Then Kendall's tau and Spearman's rho for X and Y are defined as

$$\begin{aligned} \tau(X, Y) &= 4 \int \int_{[0,1]^2} C(u, v) dC(u, v) - 1 = 4E[C(U, V)] - 1, \\ \varrho_S(X, Y) &= 12 \int \int_{[0,1]^2} u v dC(u, v) - 3 = 12E[U, V] - 3. \end{aligned}$$

Note that Spearman's rho can be written as $\varrho_S[X, Y] = \rho[F(X), G(Y)]$, and may thus be viewed as Pearson's correlation of the marginal distributions $F(X)$ and $G(Y)$.

Nelsen (1999) shows that Kendall's tau and Spearman's rho satisfy conditions required to be concordance measures. If $\kappa_{X,Y}$ denote Kendall's tau or

Spearman's rho between X and Y , we have in particular that $-1 \leq \kappa_{X,Y} \leq 1$, $\kappa_{X,X} = 1$, $\kappa_{X,-X} = -1$, $\kappa_{X,Y} = \kappa_{Y,X}$ and $\kappa_{-X,Y} = \kappa_{X,-Y} = -\kappa_{X,Y}$. For continuous variables, all values of Kendall's tau and Spearman's rho in the interval $[-1, 1]$ can be obtained by a suitable choice of the underlying copula.

While Pearson's correlation is a natural scalar measure of dependence in elliptical distributions (for instance, the Gaussian or the Student t distributions), using it as a measure of dependence in more general situations might prove misleading. Indeed, Pearson's correlation has some undesirable properties. First, it is possible that the correlation $\rho[X, Y]$ between two random variables X and Y is equal to zero while the two variables are not independent. Second, the range of permissible values for the correlation is not necessarily $[-1, 1]$. Indeed, depending on the marginal distributions of the two variables, the actual range may be much smaller than $[-1, 1]$. Third, Pearson's correlation is not invariant for an increasing transform of X and Y . Finally, we may have that $\rho[X, Y] > 0$, while X and Y do not have necessarily a positive dependence.

6.3.3 Non-parametric copulas

A first approach to the modeling of non-linear dependence consists in non-parametrically estimating the unrestricted joint density (Silverman, 1986, Härdle, 1990, Scott, 1992). This method has been used by Deheuvels (1979, 1981) and Fermanian and Scaillet (2003) to deduce a non-parametric estimate of the associated unrestricted copula. The advantage of this approach is that it does not require any additional assumption on the non-linear dependence. However, it suffers from the drawbacks of any non-parametric approach. In particular, the interpretation of the patterns of non-linear dependence is often complicated, and it is likely to provide inaccurate and erratic results, even in the bivariate case. This approach has been recently improved by the work of Gagliardini and Gouriéroux (2006), who propose an intermediate approach in which the joint density is constrained and depends on a small number of one-dimensional functional parameters, yielding efficient non-parametric estimators for the one-dimensional functional parameters, which characterize non-linear dependence.

The non-parametric copula has been proposed by Deheuvels (1979, 1981). It is defined as

$$C_T \left(\frac{t_1}{T}, \frac{t_2}{T} \right) = \frac{1}{T} \sum_{t=1}^T 1_{\{x_t \leq x_{t_1,T}, y_t \leq y_{t_2,T}\}} \quad \text{with } 1 \leq t_1, t_2 \leq T,$$

where $x_{t,T}$ denotes the order statistics, i.e., $x_{1,T} \leq \dots \leq x_{T,T}$ are ordered realizations. The empirical copula frequency is then defined as

$$c_T \left(\frac{t_1}{T}, \frac{t_2}{T} \right) = \begin{cases} \frac{1}{T} & \text{if } (x_{t_1,T}, y_{t_2,T}) \text{ belongs to the sample} \\ 0 & \text{otherwise,} \end{cases}$$

so that C_T and c_T are related through the relation

$$C_T\left(\frac{t_1}{T}, \frac{t_2}{T}\right) = \sum_{p=1}^{t_1} \sum_{q=1}^{t_2} c_T\left(\frac{p}{T}, \frac{q}{T}\right).$$

Sample versions of Kendall's tau and Spearman's rho are given by

$$\begin{aligned} \tau = \frac{2T}{T-1} \sum_{t_1=1}^T \sum_{t_2=1}^T \sum_{p=1}^{t_1} \sum_{q=1}^{t_2} & \left[c_T\left(\frac{t_1}{T}, \frac{t_2}{T}\right) \cdot c_T\left(\frac{p}{T}, \frac{q}{T}\right) \right. \\ & \left. - c_T\left(\frac{t_1}{T}, \frac{q}{T}\right) \cdot c_T\left(\frac{p}{T}, \frac{t_2}{T}\right) \right], \end{aligned}$$

and

$$\varrho_S = \frac{12}{T^2 - 1} \sum_{t_1=1}^T \sum_{t_2=1}^T \left[C_T\left(\frac{t_1}{T}, \frac{t_2}{T}\right) - \frac{t_1}{T} \cdot \frac{t_2}{T} \right].$$

A non-parametric estimation of copulas has been developed by Fermanian and Scaillet (2003), using a kernel-based approach. Such an approach has the advantage to provide a smooth reconstitution of the copula function without assuming any particular parametric structure on the dependence structure between margins.

6.3.4 Review of some copula families

We provide in this section a few examples of copula functions that have been studied and estimated in the empirical finance literature. Such copulas can be found, for instance, in Riboulet, Roncalli, and Bouyé (2000), Jondeau and Rockinger (2006b), and Patton (2006). The copula functions described in this section cover some of the most used distribution in the literature. For a more general description of copula functions, see Joe (1997) and Nelsen (1999). We do not consider in this review how margins are defined. They may be parametric as well as non-parametric.

Table 6.4 reports, for the copula functions considered below, the range of possible values that can be reached by the concordance measures. κ denotes the Kendall's tau or the Spearman's rho. This table indicates that some copula functions may be able to reproduce only positive or only negative dependence. For instance, the Clayton, Gumbel or Marshall-Olkin copula only have positive dependence.

Elliptical copulas

The class of elliptical distributions includes several well-known multivariate distributions, such as the Gaussian and the Student t distributions. Let X be a n -dimensional random vector, $\mu \in \mathbb{R}^n$ and Σ a (n, n) non-negative definite,

Table 6.4. Description of some usual copula functions

Copula	Parameter space	Range for κ
Gaussian copula	$-1 < \rho < 1$	$-1 < \kappa < 1$
Student t copula	$-1 < \rho < 1$ and $n > 2$	$-1 < \kappa < 1$
Frank copula	$-\infty < \theta < \infty$	$-1 < \kappa < 1$
Clayton copula	$0 < \theta < \infty$	$0 < \kappa \leq 1$
Gumbel copula	$1 \leq \theta < \infty$	$0 \leq \kappa < 1$
Plackett copula	$0 < \theta < \infty$ and $\theta \neq 1$	$-1 < \kappa < 1$
Marshall-Olkin copula	$0 \leq \alpha \leq 1$ and $0 \leq \beta \leq 1$	$0 < \kappa \leq 1$

symmetric matrix. Then, if the characteristic function $\varphi_{X-\mu}(t)$ of $(X - \mu)$ is a function of the quadratic form $t'\Sigma t$, i.e., $\varphi_{X-\mu}(t) = \phi(t'\Sigma t)$, then X has an elliptical distribution with parameters μ , Σ , and ϕ . In addition, if X has a density, it is of the form $|\Sigma|^{-1/2} g((X - \mu)'\Sigma^{-1}(X - \mu))$, for some non-negative function g of one scalar variable. Therefore, the contours of equal density form ellipsoids in \mathbb{R}^n .

Gaussian copula

The Gaussian copula is defined by the following *cdf*

$$C_\rho(u, v) = \Phi_\rho(\Phi^{-1}(u), \Phi^{-1}(v)) \\ = \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{s^2 - 2\rho st + t^2}{2(1-\rho^2)}\right) ds dt,$$

where Φ_ρ is the bivariate standardized Gaussian *cdf* with Pearson's correlation $\rho \in [-1; 1]$ and Φ^{-1} denotes the inverse of the distribution function of the univariate standard normal distribution.

The density of the Gaussian copula is given by

$$c_\rho(u, v) = \frac{1}{|R|^{1/2}} \exp\left(-\frac{1}{2}\psi' (R^{-1} - I_2)\psi\right),$$

where $\psi = (\Phi^{-1}(u), \Phi^{-1}(v))'$, and R is the $(2, 2)$ correlation matrix between u and v with ρ as correlation parameter.

Kendall's tau and Spearman's rho are given respectively by

$$\tau(C_\rho) = \frac{2}{\pi} \arcsin(\rho), \\ \varrho_S(C_\rho) = \frac{6}{\pi} \arcsin(\rho/2).$$

Figure 6.14 displays the Gaussian copula $c_\rho(u, v)$ for $\rho = 0.5$. Figure 6.15 presents the density h of the two-dimensional distribution defined as: $h(x, y) = c_\rho(F(x), G(y)) \times f(x) \times g(y)$. In the first case, the two variables are supposed to be distributed as $\mathcal{N}(0, 1)$, so that the Gaussian copula is equivalent to the multivariate Gaussian distribution. In the second case, X is a t_3 while Y is a $\mathcal{N}(0, 1)$.

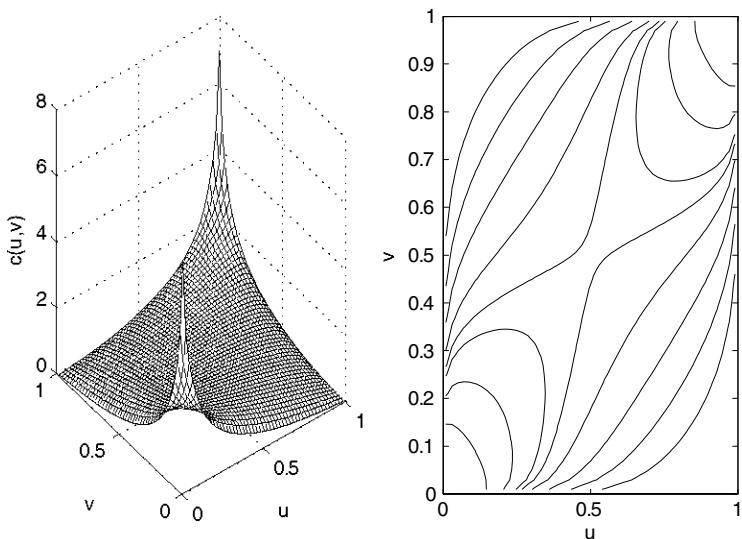


Fig. 6.14. Gaussian copula and its contour plot.

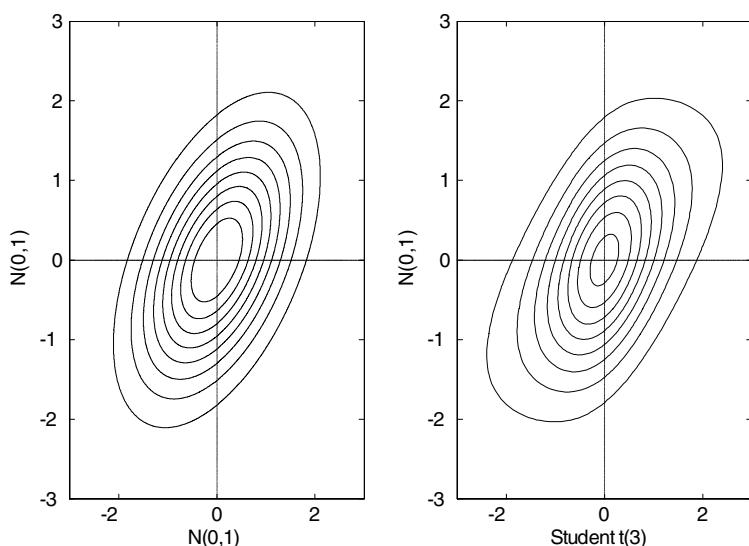


Fig. 6.15. Contour plot of the density of the Gaussian copula.

Student t copula

The Student *t* copula is defined by

$$\begin{aligned} C_{\rho,n}(u, v) &= t_{\rho,n}(t_n^{-1}(u), t_n^{-1}(v)) \\ &= \int_{-\infty}^{t_n^{-1}(u)} \int_{-\infty}^{t_n^{-1}(v)} \frac{\Gamma(\frac{n+2}{2})}{\Gamma(\frac{n}{2}) \pi n \sqrt{1-\rho^2}} \left(1 + \frac{\psi' R^{-1} \psi}{n}\right)^{-\frac{n+2}{2}} d\psi, \end{aligned}$$

where $\psi = (t_n^{-1}(u), t_n^{-1}(v))'$, $t_{\rho,n}$ is the bivariate Student *t cdf* with n degrees of freedom and correlation ρ , and t_n^{-1} denotes the inverse of the distribution function of the univariate Student *t* distribution with n degrees of freedom. The density of the Student *t* copula is given by

$$c_{\rho,n}(u, v) = \frac{1}{\sqrt{|R|}} \frac{\Gamma(\frac{n+2}{2}) \Gamma(\frac{n}{2})}{\left(\Gamma(\frac{n+1}{2})\right)^2} \frac{(1 + \frac{1}{n} \psi' R^{-1} \psi)^{-\frac{n+2}{2}}}{\prod_{i=1}^2 (1 + \frac{1}{n} \psi_i^2)^{-\frac{n+1}{2}}}.$$

As for the Gaussian copula, Kendall's tau is given by

$$\tau(C_\rho) = \frac{2}{\pi} \arcsin(\rho).$$

The analytic expression for Spearman's rho is unknown. Consequently, it must be computed numerically.

Figure 6.16 represents the density of the Student *t* copula for $\nu = 3$ and $\rho = 0.5$, corresponding to a Kendall's tau equal to 0.333. The left figure corresponds to Gaussian margins, whereas the right figure corresponds to a Gaussian and a Student *t* with $\nu = 3$ margins.

Archimedean copulas

Unlike the copulas described so far, the Archimedean ones are not derived from multivariate distribution functions. An advantage is that most Archimedean copulas have closed form expressions. A disadvantage is that multivariate extensions of Archimedean copulas are somewhat difficult to establish.

Theorem 6.5. (Nelsen, 1999, p. 91) *Let φ be a continuous, strictly decreasing function from $[0, 1]$ to $[0, \infty)$ such that $\varphi(1) = 0$ and let φ^{-1} be the inverse of φ . Then, the function from $[0, 1]^2$ to $[0, 1]$ given by*

$$C(u, v) = \varphi^{-1}(\varphi(u) + \varphi(v))$$

is a copula if and only if φ is convex.

The function φ is called a generator of the copula. Assuming that φ^{-1} is twice continuously differentiable, the copula density is

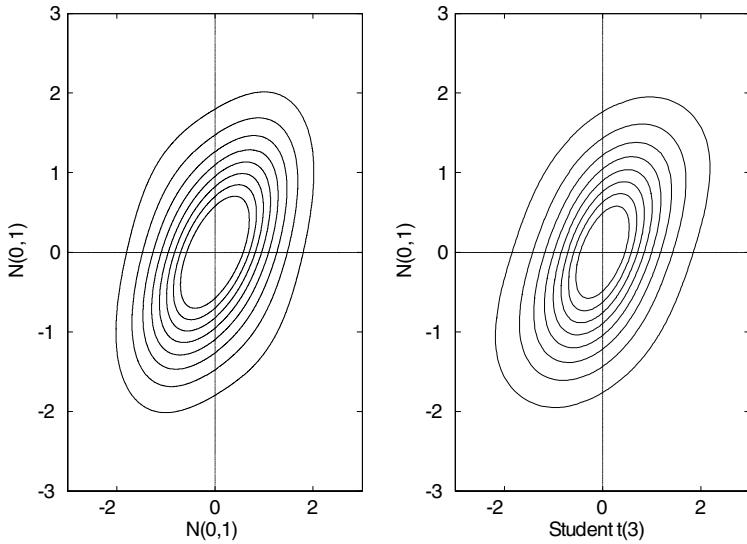


Fig. 6.16. Contour plot of the density of the Student t copula.

$$c(u, v) = \frac{\varphi^{-1''}(\varphi(u) + \varphi(v))}{\varphi^{-1'}(\varphi'(u)) \varphi^{-1'}(\varphi'(v))}.$$

In this case, Kendall's tau is given by

$$\tau(C) = 1 + 4 \int_0^1 \frac{\varphi(t)}{\varphi'(t)} dt.$$

Many additional details on Archimedean copulas are provided by Nelsen (1999) and Embrechts, Lindskog, and McNeil (2003).

Frank copula

When $\varphi(t) = \log(e^{-\theta} - 1) - \log(e^{-\theta t} - 1)$, for $\theta \neq 0$, we obtain the Frank family of copulas. The Frank copula is defined by

$$C_\theta(u, v) = -\frac{1}{\theta} \log \left(1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1} \right),$$

and

$$c_\theta(u, v) = \frac{\theta (1 - e^{-\theta}) e^{-\theta(u+v)}}{[(1 - e^{-\theta}) - (1 - e^{-\theta u})(1 - e^{-\theta v})]^2}.$$

The copula is defined for $\theta \neq 0$. Kendall's tau is given by

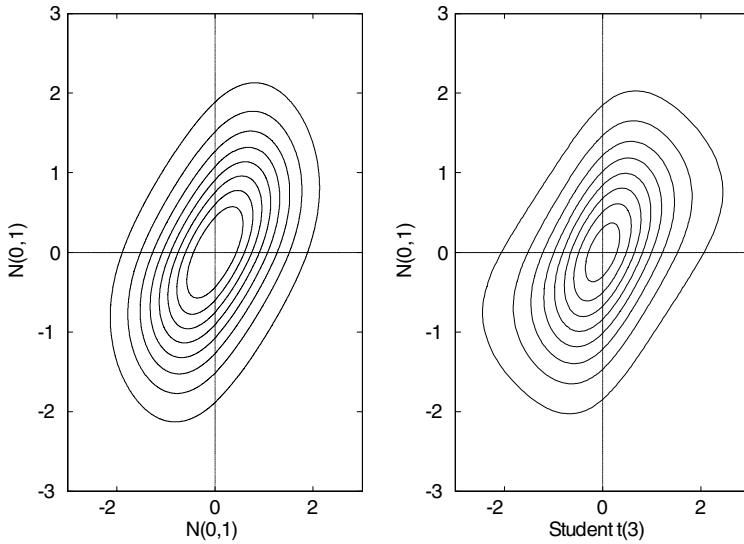


Fig. 6.17. Contour plot of the density of the Frank copula.

$$\tau(C_\theta) = 1 - 4 \frac{1 - D_1(\theta)}{\theta},$$

and Spearman's rho is

$$\varrho_S(C_\theta) = 1 - 12 \frac{D_1(\theta) - D_2(\theta)}{\theta},$$

where

$$D_k(x) = \begin{cases} \frac{k}{x^k} \int_0^x \frac{t^k}{e^t - 1} dt & \text{if } x \geq 0, \\ \frac{k|x|}{1+k} + \frac{k}{|x|^k} \int_x^0 \frac{t^k}{e^t - 1} dt & \text{if } x < 0 \end{cases}$$

is the Debye function (see Abramowitz and Stegun, 1970).

Figure 6.17 represents the density of the Clayton copula for $\theta = 3$, corresponding to a Kendall's tau equal to 0.307.

Clayton copula

When $\varphi(t) = (t^{-\theta} - 1)/\theta$, for $\theta \in [-1; \infty) \setminus \{0\}$, we obtain the Clayton family of copulas. The Clayton copula is defined by

$$C_\theta(u, v) = \max \left(\left(u^{-\theta} + v^{-\theta} - 1 \right)^{-1/\theta}, 0 \right).$$

For $\theta > 0$, the copula simplifies to

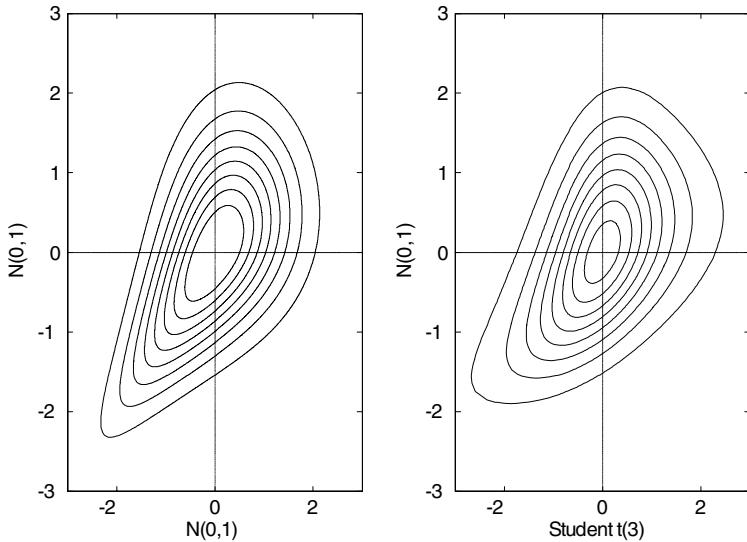


Fig. 6.18. Contour plot of the density of the Clayton copula.

$$C_\theta(u, v) = (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}.$$

The density of the copula is given by

$$c_\theta(u, v) = (1 + \theta)(uv)^{-\theta-1} (u^{-\theta} + v^{-\theta} - 1)^{-2-1/\theta}.$$

Kendall's tau is given by

$$\tau(C_\theta) = \frac{\theta}{\theta + 2}.$$

Figure 6.18 represents the density of the Clayton copula for $\theta = 1$, corresponding to a Kendall's tau equal to 0.333. The figure confirms that the Clayton copula generates dependence in the lower-tail but not in the upper tail. In order to also generate some dependence in the upper tail, we can use the so-called rotated copula, defined as follows: if (U, V) has copula $C_\theta(u, v)$, then $(1 - U, 1 - V)$ is distributed according to the rotated copula $C_\theta^R(u, v)$. For the Clayton copula, we have

$$C_\theta^R(u, v) = u + v - C_\theta(1 - u, 1 - v),$$

with density

$$c_\theta^R(u, v) = c_\theta(1 - u, 1 - v).$$

The rotated copula has dependence in the upper tail.

Gumbel copula

When $\varphi(t) = (-\log t)^\theta$, for $\theta \in [1; \infty)$, we obtain the Gumbel family of copulas. The Gumbel copula is defined by

$$C_\theta(u, v) = \exp \left\{ - \left[(-\log u)^\theta + (-\log v)^\theta \right]^{1/\theta} \right\},$$

and

$$c_\theta(u, v) = \frac{C_\theta(u, v) [\log u \times \log v]^{\theta-1}}{uv \left[(-\log u)^\theta + (-\log v)^\theta \right]^{2-1/\theta}} \times \\ \left\{ \left[(-\log u)^\theta + (-\log v)^\theta \right]^{1/\theta} + \theta - 1 \right\}.$$

Kendall's tau is given by

$$\tau(C_\theta) = 1 - \frac{1}{\theta}.$$

Figure 6.19 represents the density of the Gumbel copula for $\theta = 1.5$, corresponding to a Kendall's tau equal to 0.333. As for the Clayton copula, it is possible to define a rotated Gumbel copula with

$$C_\theta^R(u, v) = u + v - 1 + C_\theta(1 - u, 1 - v),$$

with density

$$c_\theta^R(u, v) = c_\theta(1 - u, 1 - v).$$

Plackett copula

The Plackett copula, proposed by Plackett (1965), is defined by

$$C_\theta(u, v) = \frac{1}{2(\theta - 1)} \times \\ \left[1 + (\theta - 1)(u + v) - \sqrt{[1 + (\theta - 1)(u + v)]^2 - 4uv\theta(\theta - 1)}, \right]$$

and

$$c_\theta(u, v) = \frac{\theta[1 + (u - 2uv + v)(\theta - 1)]}{\left([1 + (\theta - 1)(u + v)]^2 - 4uv\theta(\theta - 1) \right)^{\frac{3}{2}}}.$$

The copula is defined for $\theta \in [0; \infty)$. Note that the Spearman's rho of the Plackett copula is simply derived from the dependence parameter θ as

$$\rho_\theta = \frac{\theta + 1}{\theta - 1} - \frac{2\theta \log(\theta)}{(\theta - 1)^2}. \quad (6.22)$$

Figure 6.20 represents the density of the Plackett copula for $\theta = 3$, corresponding to a Spearman's rho equal to 0.352.

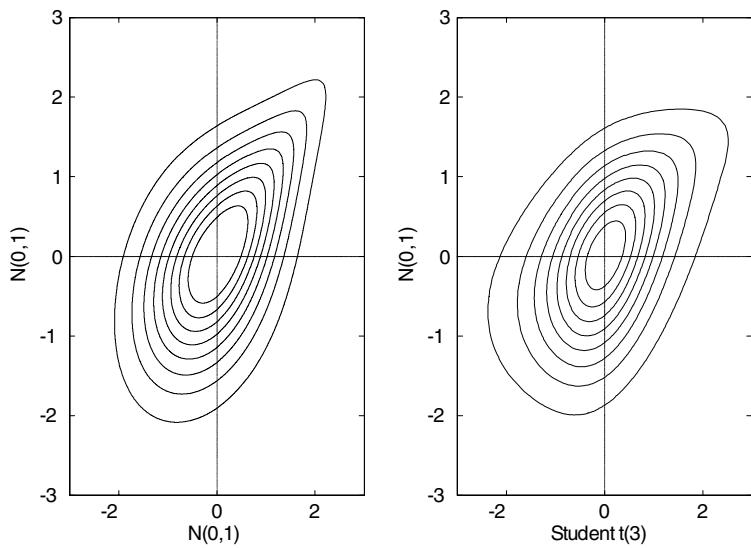


Fig. 6.19. Contour plot of the density of the Gumbel copula.

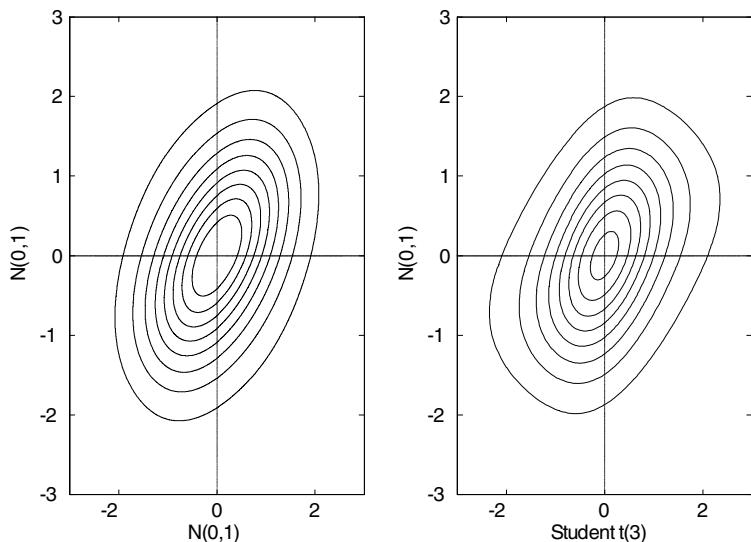


Fig. 6.20. Contour plot of the density of the Plackett copula.

Marshall-Olkin copula

The Marshall-Olkin copula is defined by the function

$$C_{\alpha,\beta}(u, v) = \min(u^{1-\alpha}v, uv^{1-\beta}) = \begin{cases} u^{1-\alpha}v & \text{if } u^\alpha \geq v^\beta, \\ uv^{1-\beta} & \text{if } u^\alpha \leq v^\beta, \end{cases}$$

with $u = \bar{F}(x)$ and $v = \bar{G}(y)$, where $\bar{F}(x) = \Pr[X > x]$ and $\bar{G}(y) = \Pr[Y > y]$ denote the marginal survival functions. The domain of definition of the parameters α and β is: $0 \leq \alpha, \beta \leq 1$.

The density of the copula is defined by

$$c_{\alpha,\beta}(u, v) \equiv \frac{\partial^2 C_{\alpha,\beta}(u, v)}{\partial u \partial v} = \begin{cases} u^{-\alpha} & \text{if } u^\alpha > v^\beta, \\ v^{-\beta} & \text{if } u^\alpha < v^\beta, \end{cases}$$

so that the Marshall-Olkin copulas have a singular component, with a mass concentrated on the curve $u^{-\alpha} = v^{-\beta}$.

Kendall's tau and Spearman's rho are obtained, applying the results of the previous section (Embrechts, Lindskog, and McNeil, 2003)

$$\begin{aligned} \tau(C_{\alpha,\beta}) &= 4 \int \int_{[0,1]^2} C(u, v) \, dC(u, v) - 1 = \frac{\alpha\beta}{\alpha + \beta - \alpha\beta}, \\ \varrho_S(C_{\alpha,\beta}) &= 12 \int \int_{[0,1]^2} uv \, dC(u, v) - 3 = \frac{3\alpha\beta}{2\alpha + 2\beta - \alpha\beta}. \end{aligned}$$

All values in the interval $[0, 1]$ can be obtained for both the Kendall's tau and the Spearman's rho.

Figure 6.21 represents the density of the Marshall-Olkin copula for $\lambda_1 = 0.5$, $\lambda_2 = 0.1$ and $\lambda_{12} = 1$, so that $\alpha = 0.667$ and $\beta = 0.909$. These values correspond to a Kendall's tau equal to 0.375.

6.3.5 Estimation

Several approaches have been proposed to estimate the parameters of a copula function. In addition to the standard ML estimation, a two-step estimation procedure is readily available, because the log-likelihood of the model can be written as the sum of two components, the margins and the dependence structure. Therefore, it is natural to estimate the parameters of the margins and the parameters of the copula function separately (Shih and Louis, 1995, Joe and Xu, 1996). The copula parameters can also be semi-parametrically estimated, using the marginal empirical distribution to compute the copula (Genest, Ghoudi, and Rivest, 1995).

Other alternative estimation procedures may be considered as well. For instance, as it has been shown in Section 6.3.4, the parameters of the copula function are in general related to one of the concordance measures (Kendall's

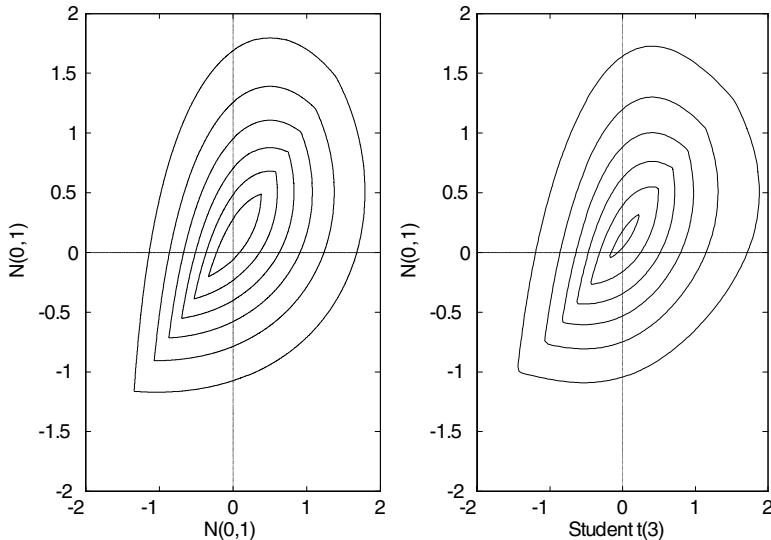


Fig. 6.21. Contour plot of the density of the Marshall-Olkin copula.

tau or Spearman's rho). It is therefore possible to estimate the unknown parameters using the method of moments, in such a way that the Kendall's tau (say) of the theoretical copula fits the empirical Kendall's tau. Genest and Rivest (1993) have developed an estimation procedure adapted for Archimedean copula functions.

ML estimator

We assume now that the unknown parameters associated with the marginal densities f and g are denoted θ_x and θ_y , respectively, and that the unknown parameters associated with the copula function c are denoted θ_γ . We denote the $(K, 1)$ vector $\theta = (\theta_x, \theta_y, \theta_\gamma)$. The maximum likelihood estimate (MLE) of a model is obtained by maximizing the conditional log-likelihood function, which is defined as

$$\begin{aligned}
 L_T(\theta | \underline{x}_t, \underline{y}_t) &= \sum_{t=1}^T \log (c_{\theta_\gamma}(F(x_t, \theta_x), G(y_t, \theta_y)) \times f(x_t, \theta_x) \times g(y_t, \theta_y)) \\
 &= \sum_{t=1}^T \log c_{\theta_\gamma}(F(x_t, \theta_x), G(y_t, \theta_y)) \\
 &\quad + \sum_{t=1}^T (\log [f(x_t, \theta_x)] + \log [g(y_t, \theta_y)]).
 \end{aligned} \tag{6.23}$$

As previously, the ML estimator $\hat{\theta}_{ML}$ is asymptotically normal with asymptotic distribution

$$\sqrt{T} (\hat{\theta}_{ML} - \theta_0) \implies \mathcal{N}(0, A_0^{-1}),$$

where A_0 is the information matrix of Fisher. See Section 4.3.3 for further details on the construction and estimation of A_0 . See also Section 5.1.2 for the computation of the covariance matrix of the QML estimator.

Two-step estimator (or inference functions for margins method)

In practical applications, the ML estimation may be difficult. First, the dimension of the problem can be large, because it requires one to estimate jointly the parameters of the margins and of the copula function. Second, the dependency parameter of the copula function may be a convoluted expression of the margin parameters. Therefore, an analytical expression of the gradient of the likelihood might not exist. Only numerical gradients may be computable, implying a slowing down of the numerical procedure.

In some cases, it is possible to split the vector of parameters into two parts: those associated with the marginal distributions and those associated with the copula function. This is the case, in particular, when there is no cross-restriction between the marginal distributions and the copula function.

A first two-step estimator, initially proposed by Shih and Louis (1995), Joe and Xu (1996), and Joe (1997) and used in a conditional setup by Jondeau and Rockinger (2006b) and Patton (2006), corresponds to the case when the parameters of the two marginal distributions can be estimated separately. The first step is the estimation of the marginal models. In the bivariate setting, we have

$$\tilde{\theta}_x \in \arg \max_{\{\theta_x \in \Theta_x\}} \sum_{t=1}^T \log[f(x_t, \theta_x)], \quad (6.24)$$

$$\tilde{\theta}_y \in \arg \max_{\{\theta_y \in \Theta_y\}} \sum_{t=1}^T \log[g(y_t, \theta_y)]. \quad (6.25)$$

Then, in a second step, the parameters θ_γ of the copula function can be estimated conditionally on the margin parameters

$$\tilde{\theta}_\gamma \in \arg \max_{\{\theta_\gamma \in \Theta_\gamma\}} \sum_{t=1}^T \log \left[c_{\theta_\gamma} \left(F \left(x_t, \tilde{\theta}_x \right), G \left(y_t, \tilde{\theta}_y \right) \right) \right].$$

If the model is correctly specified, then under rather mild assumptions, $\tilde{\theta}_x$, $\tilde{\theta}_y$, and $\tilde{\theta}_\gamma$ are consistent and asymptotically normal estimators (Patton, 2006).

Another estimator corresponds to the case when we cannot separate the parameters of the two marginal distributions. This may be the case, for instance, when the two variables are related by a multivariate GARCH model.

Although the two variables are still assumed to be independent, parameters of the two marginal distributions cannot be estimated separately. In this context, the first-step estimators are obtained by solving

$$(\check{\theta}_x, \check{\theta}_y) \in \arg \max_{\{\theta_x \in \Theta_x, \theta_y \in \Theta_y\}} \sum_{t=1}^T \log[f(x_t, \theta_x, \theta_y)] + \sum_{t=1}^T \log[g(y_t, \theta_x, \theta_y)],$$

while the estimators of the dependence parameters are obtained by solving

$$\check{\theta}_\gamma \in \arg \max_{\{\theta_\gamma \in \Theta_\gamma\}} \sum_{t=1}^T \log [c_{\theta_\gamma}(F(x_t, \check{\theta}_x), G(y_t, \check{\theta}_y))].$$

As above, estimators $\check{\theta}_x$, $\check{\theta}_y$, and $\check{\theta}_\gamma$ are shown to be consistent and asymptotically normal.

Semi-parametric ML

Genest, Ghoudi, and Rivest (1995) have proposed an estimation procedure that avoids specifying the marginal distribution. Instead of using a parametric marginal distribution, they suggest the use of the marginal empirical distribution function. The empirical distribution function of X is

$$\hat{u}_T(\tau) = \hat{F}_T(x_\tau) = \frac{1}{T} \sum_{t=1}^T 1_{\{x_t \leq x_{\tau,T}\}},$$

where $x_{1,T} \leq \dots \leq x_{T,T}$ is the ordered sample of observations. In other words, $\hat{u}_T(\tau)$ represents the frequency of observations below or equal to x_τ in the sample $\{x_t\}_{t=1}^T$. Genest, Ghoudi, and Rivest (1995) suggest that we redefine $\hat{u}_T(\tau)$ as $\frac{T}{T+1}\hat{u}_T(\tau)$ to avoid difficulties arising from the possible unboundedness of the log-likelihood when some \hat{u}_Ts tend to one.

Considering the empirical margins directly avoids assuming any theoretical distribution for margins and therefore avoids the estimation of the parameters of the marginal distributions. Then, the parameters θ_γ of the copula function are estimated by maximizing the pseudo log-likelihood

$$\check{\theta}_\gamma \in \arg \max_{\{\theta_\gamma \in \Theta_\gamma\}} \sum_{t=1}^T \log c_{\theta_\gamma}(\hat{F}_T(x_t), \hat{G}_T(y_t)).$$

Genest, Ghoudi, and Rivest (1995) show that the estimator $\check{\theta}_\gamma$ is asymptotically normal, with a larger asymptotic variance than the ML estimator (obtained assuming that the margins are known).¹⁴

¹⁴ In the case where $\check{\theta}_\gamma$ is multidimensional, this implies that the difference between the asymptotic covariance matrices of $\check{\theta}_\gamma$ and $\hat{\theta}_{\gamma,ML}$ is semi-definite positive.

6.3.6 Adequacy tests

One reason for the success of the copula functions is that they are able to model almost any kind of relationship between time series. But one drawback is that the class of copula functions is now very large, so that it is often difficult to select the “right” one, i.e., the copula function that best fits the data at hand. There is therefore a need for adequacy tests.

Several papers have recently proposed goodness-of-fit tests for copula models. Genest and Rivest (1993) propose a test for the Archimedean family. More recent papers are by Breyman, Dias, and Embrechts (2003), Fermanian (2003), and Malavergne and Sornette (2003). As argued by Fermanian and Scaillet (2004), there are some difficulties in designing a statistical test for the adequacy of copula functions to the data. The reason is that the initial variables have to be transformed in their probability integral transforms through the marginal distribution functions. Since the margins are unknown, they have to be treated as nuisance parameters.

A natural way is to adapt the test proposed in the multivariate distribution context by Diebold, Hahn, and Tay (1999) for copula functions. Such an adaptation has been proposed by Breymann, Dias, and Embrechts (2003). As in Section 6.2.4, we refer to the Rosenblatt (1952) transform and define the probability integral transform in terms of conditional distributions. Consider n random variables X_i whose joint distribution is given by $F_X(x_1, \dots, x_n)$ and marginal distributions by $F_{X_i}(x_i) = \Pr[X_i \leq x_i]$. We define the probability integral transform $U_i = T(X_i)$ where

$$\begin{aligned} T(X_1) &= \Pr[X_1 \leq x_1] = F_{X_1}(x_1), \\ T(X_2) &= \Pr[X_2 \leq x_2 | X_1 = x_1] = F_{X_2|X_1}(x_2|x_1), \\ &\vdots \\ T(X_n) &= \Pr[X_n \leq x_n | X_1 = x_1, \dots, X_{n-1} = x_{n-1}] \\ &= F_{X_n|X_1, \dots, X_{n-1}}(x_n|x_1, \dots, x_{n-1}), \end{aligned}$$

Therefore, the variables U_i for $i = 1, \dots, n$ are *iid* $U(0, 1)$ individually and jointly.

Suppose now that the copula C is such that $C(F_{X_1}(x_1), \dots, F_{X_n}(x_n)) = F_X(x_1, \dots, x_n)$. If we denote $C_i(u_1, \dots, u_i) = C(u_1, \dots, u_i, 1, \dots, 1)$ the joint marginal distribution for the first i variables (U_1, \dots, U_i) , then the conditional distribution of U_i given (U_1, \dots, U_{i-1}) is

$$C_i(u_i | u_1, \dots, u_{i-1}) = \frac{\partial^{i-1} C_i(u_1, \dots, u_i)}{\partial u_1 \cdots \partial u_{i-1}} / \frac{\partial^{i-1} C_{i-1}(u_1, \dots, u_{i-1})}{\partial u_1 \cdots \partial u_{i-1}}.$$

Since we have

$$C_i(F_{X_i}(x_i) | F_{X_1}(x_1), \dots, F_{X_{i-1}}(x_{i-1})) = F_{X_i|X_1, \dots, X_{i-1}}(x_i | x_1, \dots, x_{i-1}),$$

we deduce that

$$U_i = C_i(F_{X_i}(x_i)|F_{X_1}(x_1), \dots, F_{X_{i-1}}(x_{i-1})) \quad i = 2, \dots, n,$$

(with $U_1 = C_1(F_{X_1}(x_1))$). Since the copula is a multivariate distribution function, it follows that the U_i s are *iid* $U(0, 1)$ (individually and jointly) if C is the true distribution.

These results are used by Breymann, Dias, and Embrechts (2003) to construct the following test. They propose to transform the U_i s using the inverse of the univariate normal distribution $\Phi^{-1}(U_i)$, $i = 1, \dots, n$, then the transformed variables $\Phi^{-1}(U_i)$ should be *iid* $\mathcal{N}(0, 1)$. The proposed test statistic is

$$S = \sum_{i=1}^n (\Phi^{-1}(U_i))^2,$$

that is asymptotically distributed, under the null, as a $\chi^2(n)$. This test is in fact an Anderson-Darling test.

A strong limitation of this test is that it is based on the empirical marginal distributions in order to transform the marginal data. Therefore, they should be treated as infinite dimensional nuisance parameters. This obviously affects the critical values of the test. Malevergne and Sornette (2003) suggest the use of bootstrap to compute the empirical critical values of the Anderson-Darling test.

Another difficulty, already mentioned for the test of Diebold, Hahn, and Tay (1999), is that the Rosenblatt transform allows $n!$ ways to factor the joint distribution in terms of conditional distributions.

To deal with the fact that the empirical distribution is unknown and has to be treated as nuisance parameters, we may adopt a nonparametric approach. For instance, Fermanian (2003) considers a kernel estimation of the empirical copula density to circumvent the direct use of the empirical copula process. He proposes a goodness-of-fit test based on the difference between the kernel estimator of the copula and the assumed copula.

6.3.7 Modeling the conditional dependency parameter

For notational convenience, we set $u_t \equiv F(x_t, w_x)$ and $v_t \equiv G(y_t, w_y)$. We denote by ρ the dependency parameter. It may be for instance the correlation parameter in the Gaussian or Student t copulas. The conditioning can be achieved by expressing ρ as a function of explanatory variables, for instance lagged values of u_t and v_t , or some other predetermined variable z_t . A rather general specification for ρ_t is

$$\rho_t = \Gamma(\underline{u}_{t-1}, \underline{v}_{t-1}, \underline{z}_{t-1}, \underline{\gamma}_{t-1}; w_\rho),$$

where Γ is a function depending on the parameter vector w_ρ and \underline{u}_{t-1} denotes $\{u_{t-1}, u_{t-2}, \dots\}$.

Many different specifications of the dependency parameter are possible in this context. As a first approach, we may follow Gouriéroux and Monfort (1992) and adopt a specification in which ρ_t depends on the position of past joint realizations in the unit square. This means that we decompose the unit square of joint past realizations into a grid. The parameter ρ_t will be constant for each element of the grid. More precisely, our basic model is

$$\log(\rho_t) = \sum_{j=1}^{16} d_j \mathbf{1}_{[(u_{t-1}, v_{t-1}) \in \mathcal{A}_j]},$$

where \mathcal{A}_j is the j th element of the unit-square grid. To each parameter d_j , an area \mathcal{A}_j is associated. Figure 6.22 illustrates the position of the areas d_j s. In the figure, we have set equally spaced threshold levels, i.e., p_1 , p_2 , and p_3 take the values 0.25, 0.5, and 0.75. The same for q_1 , q_2 , and q_3 . For instance, $\mathcal{A}_1 = [0, p_1] \times [0, q_1]$ and $\mathcal{A}_2 = [p_1, p_2] \times [0, q_1]$. The choice of 16 subintervals is somewhat arbitrary. This choice of parameterization has the advantage to provide an easy testing of several conjectures concerning the impact of past joint returns on subsequent dependency while still allowing for a large number of observations per area.

This specification does not allow the measurement of persistence in ρ_t , however. The difficulty is to derive an adequate model to capture the dynamic of the dependency parameter. As alternative approaches, we may adopt a specification close to the one proposed by Tse and Tsui (2002) or Engle (2002) in their modeling of the Pearson's correlation in a GARCH context. For instance, in the case of the time-varying conditional correlation model of Tse and Tsui, the dynamic of the Spearman's rho would be given by

$$\rho_t = (1 - \theta_1 - \theta_2) \rho + \theta_1 \rho_{t-1} + \theta_2 \psi_{t-1},$$

where $\psi_t = \left(\sum_{h=0}^{m-1} u_{t-h} v_{t-h} \right) / \left(\sum_{h=0}^{m-1} u_{t-h}^2 \sum_{h=0}^{m-1} v_{t-h}^2 \right)^{1/2}$ represents the correlation between the margins over the recent period. We impose that $0 \leq \theta_1, \theta_2 \leq 1$ and $\theta_1 + \theta_2 \leq 1$. The null hypothesis $\theta_1 = \theta_2 = 0$ can be tested using a standard Wald statistic.

Another alternative approach may be that parameters of the Student t copula, with degree-of-freedom parameter n and correlation ρ , are driven by a Markov-switching model of the type

$$\rho_t = \underline{\rho} S_t + \bar{\rho} (1 - S_t), \quad (6.26)$$

$$n_t = \underline{n} S_t + \bar{n} (1 - S_t), \quad (6.27)$$

where S_t denotes the unobserved regime of the system at time t . S_t is assumed to follow a two-state Markov process, with transition probability matrix given by

$$\begin{pmatrix} p & 1-p \\ 1-q & q \end{pmatrix},$$

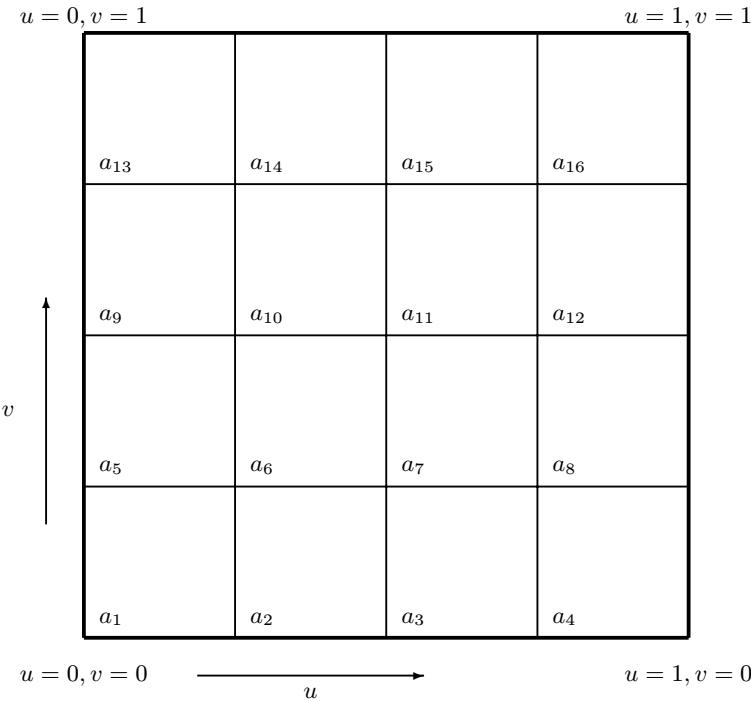


Fig. 6.22. Unit-square grid for the dependency parameter.

with

$$\begin{aligned} p &= \Pr [S_t = 0 | S_{t-1} = 0], \\ q &= \Pr [S_t = 1 | S_{t-1} = 1]. \end{aligned}$$

Note that, in this model, we do not necessarily assume that univariate characteristics of returns also shift. Quasi Maximum-Likelihood estimation of this model can be easily obtained using the approach developed by Hamilton (1989) and Gray (1995). For the degree-of-freedom parameter, we may investigate several hypotheses. For instance, we may test whether it is regime independent ($n = \bar{n}$) or whether it is infinite, so that the Gaussian copula would prevail for a given regime.

6.3.8 Illustration

We now consider the estimation of copula functions for our two pairs of daily returns (SP500 and DAX on one hand and SP500 and FT-SE on the other hand) over the period from January 1980 to August 2004. To compare the

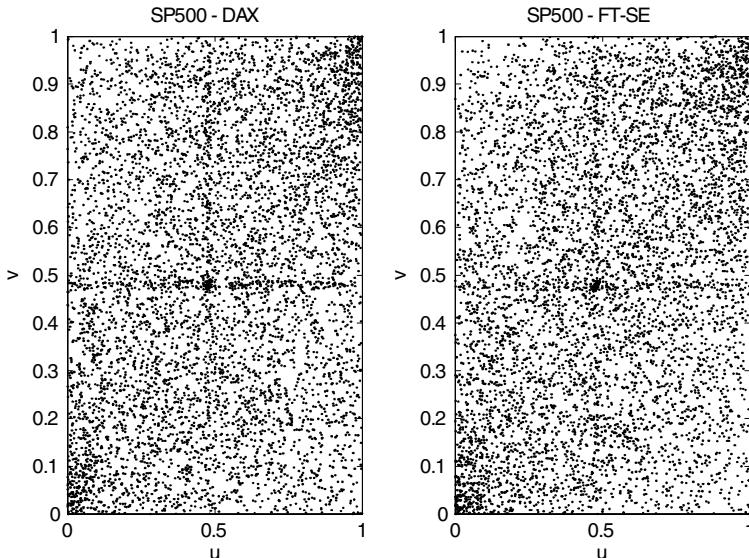


Fig. 6.23. Scatterplot of margins.

various copula functions at hand, we adopt a relatively simple specification for the margins. We assume that the dynamic of each daily return is given by a constant conditional mean and a GARCH(1, 1) conditional volatility. The marginal distribution is assumed to be either the normal or the Student t distribution. The estimation is performed using the two-step procedure described in Section 6.3.5: parameters of the marginal distributions are estimated in the first step (giving $\tilde{\theta}_x$ and $\tilde{\theta}_y$), then the margins $\hat{u}_t = F(x_t, \tilde{\theta}_x)$ and $\hat{v}_t = G(y_t, \tilde{\theta}_y)$ are computed and the parameters θ_γ of the copula $c_{\theta_\gamma}(\hat{u}_t, \hat{v}_t)$ are estimated (giving $\hat{\theta}_\gamma$). Parameter estimates of the marginal distributions are not reported. They are very close to those presented in Chapter 4. We present in Figure 6.23 scatterplots of the marginal cdfs u_t and v_t for the SP500–DAX and for the SP500–FT-SE, respectively, when the margins are supposed to be Student t . We notice that, except for the region where one margin is large and the other small, the unit square is rather uniformly filled with realizations. In both figures, there is a higher concentration in the corners along the diagonal. This clustering corresponds to the observation that correlation may be higher in the tails.

Tables 6.5 and 6.6 report parameter estimates corresponding to these various models. For each copula function, we report the copula component of the log-likelihood ($\log c_{\theta_\gamma}(F(x_t, \theta_x), G(y_t, \theta_y))$), as well as concordance measures (Kendall's tau, Spearman's rho, and the lower-tail and upper-tail dependence measures). We first notice that the estimates of the dependence parameters

are rather similar whatever the marginal distribution and the copula function. The parameters τ and ϱ_S are typically around 0.17 and 0.25 for the pair SP500–DAX and around 0.2 and 0.3 for the pair SP500–FT-SE. We notice that dependence is found to be more pronounced for the symmetric copulas than for the asymmetric ones. We notice that the Student t copula which has both lower-tail and upper-tail dependence is characterized by quite large degree-of-freedom parameters, so that no tail dependence is obtained. For this date, this finding can be confirmed using a test based on the extreme value theory, as illustrated in the next chapter. Finally, we may compare the log-likelihoods obtained with the various copula functions.¹⁵ We observe that the copula that provides the best fit of the data at hand is the Student t copula.

Table 6.5. Estimation of various copula functions (SP500–DAX)

	log-lik.	Parameter estimate	Kendall's tau	Spearman's rho
SP500–DAX (Gaussian margins)				
Gaussian	220.9177	0.2576	0.1659	0.2467
Student t	259.2342	0.2718 16.7154	0.1752	0.2588
Frank	213.7218	1.7728	0.1898	0.2876
Plackett	226.8283	2.4512	0.1971	0.2911
Clayton	109.0424	0.1704	0.0785	0.1174
Rotated Clayton	177.7902	0.2973	0.1294	0.1926
Gumbel	224.3033	1.1914	0.1607	0.2369
Rotated Gumbel	101.5494	1.1617	0.1392	0.2058
SP500–DAX (Student t margins)				
Gaussian	217.5797	0.2594	0.1670	0.2484
Student t	249.0502	0.2592 9.9642	0.1669	0.2464
Frank	196.1103	1.5477	0.1663	0.2551
Plackett	206.1987	2.2003	0.1738	0.2575
Clayton	171.3745	0.2768	0.1216	0.1811
Rotated Clayton	182.6102	0.3185	0.1374	0.2043
Gumbel	221.9475	1.1884	0.1585	0.2338
Rotated Gumbel	203.9667	1.1725	0.1471	0.2174

¹⁵ Since the estimation is performed using the two-step approach, the reported copula component of the log-likelihood can be compared for the various copula functions. However, it is not surprising that there is no systematic pattern when we compare the Gaussian margins with the Student t margins. Inspection of the total log-likelihood (including the margins) clearly indicates that the model with Student t margins performs better than the model with Gaussian margins.

Table 6.6. *Estimation of various copula functions (SP500–FT-SE)*

	log-lik.	Parameter estimate	Kendall's tau	Spearman's rho
SP500–FT-SE (Gaussian margins)				
Gaussian	323.1311	0.3263	0.2116	0.3129
Student <i>t</i>	367.9556	0.3294	0.2137	0.3141
		18.3911		
Frank	323.3922	2.1610	0.2289	0.3421
Plackett	337.0456	2.8816	0.2317	0.3401
Clayton	223.2368	0.3027	0.1314	0.1956
Rotated Clayton	210.9462	0.3479	0.1482	0.2201
Gumbel	284.0590	1.2376	0.1920	0.2818
Rotated Gumbel	255.8926	1.2148	0.1768	0.2602
SP500–FT-SE (Student <i>t</i> margins)				
Gaussian	336.0140	0.3185	0.2063	0.3054
Student <i>t</i>	364.5217	0.3202	0.2075	0.3044
		11.4373		
Frank	305.6729	1.9384	0.2067	0.3112
Plackett	317.3299	2.6161	0.2111	0.3110
Clayton	290.9546	0.3745	0.1577	0.2339
Rotated Clayton	242.3389	0.3848	0.1613	0.2392
Gumbel	303.7355	1.2374	0.1918	0.2816
Rotated Gumbel	330.7893	1.2266	0.1847	0.2714

Extreme Value Theory

In Chapters 4 to 6, we considered the modeling of the entire (univariate or multivariate) distribution of returns. When it is possible as well as effective, the modeling of the entire (univariate or multivariate) distribution of returns is very useful, because it allows a full characterization of returns. However, in some cases, modeling their entire distribution is very tricky (even using a copula approach) because the shape of the distribution may be very complicated. Moreover, for some applications, it is not necessary to model all aspects of the return distribution. For instance, to compute a VaR or to analyze financial crises, it is sufficient to only focus on the tails of the distribution. On one hand, this implies a loss of information because the focus is on a small part of the distribution. On the other hand, however, it allows the use of specific tools that were created to provide a precise description of the tails of distributions.

In this chapter, we turn to the modeling of extreme realizations in both univariate and multivariate frameworks. This is known in the statistics literature as extreme value theory, EVT. A very thorough description of the theory can be found in Leadbetter, Lindgren, and Rootzén (1983) and Embrechts, Klüppelberg, and Mikosch (1997). It should be noticed, from the onset, that many results have been obtained for univariate distributions, whereas only relatively few results are available for multivariate data. The reason for this is that in a univariate setting, an analogous result to the central limit theorem holds. No such result appears to exist in a multivariate setting.

In EVT the focus of interest is the so-called *tail index* ξ that characterizes the shape of the tails of the distribution. There are essentially two parametric approaches to obtain a measure of this index. The *extrema approach* studies the distribution followed by the maximum or minimum of returns over given time horizons. This distribution is shown to be nested in the *generalized extreme value* (gev) distribution, which can be estimated and which provides an estimate of the tail index. The *tail (or peak-over-threshold) approach* considers the exceedances over a high threshold and therefore directly focuses on the tails of the distribution. The distribution of the excesses over a high threshold is shown to be approximately distributed according to a *generalized Pareto*

distribution (gpd), which can be estimated to provide an estimate of the tail index. Both approaches have drawbacks, however. In particular, the two approaches require different assumptions about dependency of returns. The tail approach necessitates that returns are *iid*, whereas the extrema approach only imposes the *iid*-ness of subsamples over which extrema are computed. Therefore, the latter approach has the advantage that subsamples are likely to be *iid*, if we take a sufficiently long time horizon. On the other side, the estimate of the gev distribution presumes that this is the actual distribution of returns, which is generally an overly strong assumption. When the assumptions underlying these approaches are considered as too strong, it is possible to estimate the tail index using semi-parametric methods. For applications that are relevant to finance, the so-called Hill estimate will be such a semi-parametric estimate of the tail index ξ . Another difficulty that we face with financial data is heteroskedasticity, which is known to affect the distribution of extrema or tails. Some solutions have been proposed to deal with this issue.

7.1 Univariate tail estimation

In this section, we consider theoretical and empirical issues concerning the modeling of the tails of a univariate distribution. As already said, there are the two approaches, based on the extreme realizations or on the tails of the distribution. We will describe these two methods in turn. We will also consider some semi-parametric alternative approaches that do not require strong assumptions on the data.

7.1.1 Distribution of extremes

This approach has been developed by Jenkinson (1955), Prescott and Walden (1980), and Hosking, Wallis, and Wood (1985). We adopt notations similar to those of Embrechts, Klüppelberg, and Mikosch (1997).

Standard extreme value distributions

We will be concerned in this section with the distribution of the maximum or the minimum return of a stock. Before presenting the main results, we need some notations. Let X_1, \dots, X_T be a sequence of random variables corresponding to returns. We notice that

$$-\min(-X_1, \dots, -X_T) = \max(X_1, \dots, X_T) \equiv M_T,$$

which shows that without loss of generality, it is enough to develop a theory for the upper tail of the distribution of returns. We next notice that if the X_t are independent and identically distributed, (*iid*), then, if $F_X(\cdot)$ is the cumulative distribution function (*cdf*) of any X_t , it follows that

$$\begin{aligned}\Pr[M_T \leq x] &= \Pr[\max(X_1, \dots, X_T) \leq x] = \Pr[X_1 \leq x, \dots, X_T \leq x] \\ &= \prod_{t=1}^T \Pr[X_t \leq x] = [F_X(x)]^T.\end{aligned}$$

This result shows that under the *iid* assumption, the distribution of the maximum of a finite sample can be easily obtained if $F_X(\cdot)$ is known. If we denote $x_F = \sup\{x \in \mathbb{R}; F_X(x) < 1\} < \infty$ the right endpoint of the sample, we notice that

$$\Pr[M_T \leq x] = \begin{cases} 0 & \text{for all } x < x_F, \\ 1 & \text{for all } x > x_F, \end{cases}$$

and, thus, we obtain a degenerate distribution, which has little practical value. This raises the question whether a scaled version of M_T converges to a finite distribution.¹ For the problem at hand we have the Fisher and Tippett (1928) theorem, which characterizes the limit law for maxima. The first formal proof was provided by Gnedenko (1943).

Theorem 7.1. *Let X_t be a sequence of *iid* random variables. If there exist a location parameter $\mu_T \in \mathbb{R}$, and a scale parameter $\psi_T > 0$, and some non-degenerate cdf H such that the limit distribution of standardized extremes $Y_T = (M_T - \mu_T)/\psi_T$ converge to H*

$$\lim_{T \rightarrow \infty} \Pr\left(\frac{M_T - \mu_T}{\psi_T} \leq y\right) = H(y), \quad \forall y \in \mathbb{R},$$

then H belongs to one of the following three cdfs:

$$\text{Weibull: } \Psi_\alpha(y) = \begin{cases} \exp(-(-y)^\alpha) & \text{for } y \leq 0, \\ 1 & \text{for } y > 0, \end{cases} \quad \alpha > 0,$$

$$\text{Gumbel: } \Lambda_\alpha(y) = \exp(-\exp(-y)) \text{ for } y \in \mathbb{R},$$

$$\text{Fr\'echet: } \Phi_\alpha(y) = \begin{cases} 0 & \text{for } y \leq 0, \\ \exp(-y^{-\alpha}) & \text{for } y > 0, \end{cases} \quad \alpha > 0.$$

F_X is then said to belong to the domain of attraction of H . The three distributions are called standard extreme value distributions.

In Figure 7.1, we represent those various distributions. It is worth emphasizing that the distributions cannot be flipped symmetrically around a vertical axis. This implies intuitively that to get a distribution such as the Weibull's, the support of the distribution underlying the random variables must be finite. We also observe that the Gumbel distribution generates thin tails, whereas

¹ After all, a sum of *iid* random variables also degenerates as the sample size increases whereas a scaled version converges, thanks to the central limit theorem, to a well-defined distribution.

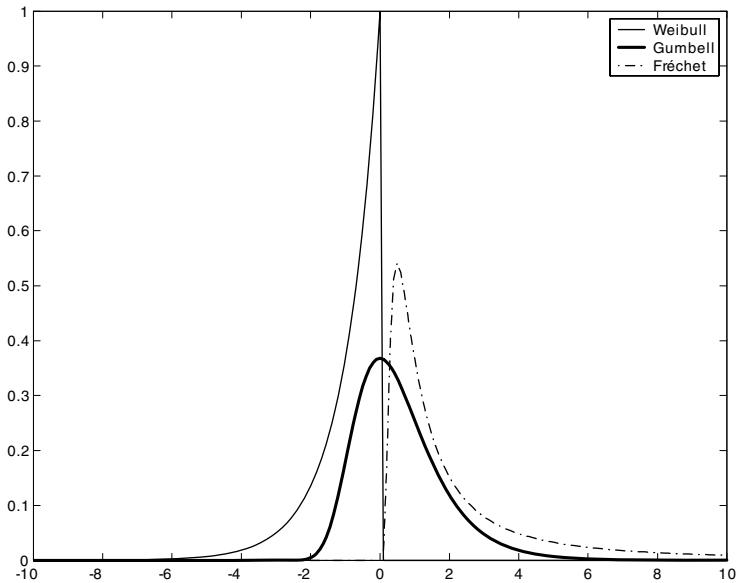


Fig. 7.1. *Pdf of the standard extreme value distributions (for $\alpha = 1$).*

the Fréchet distribution generates fat tails. Empirical evidence suggests that returns are fat-tailed. This indicates that the Fréchet distribution is likely to describe the behavior of extremes of stock market returns better than the other two distributions (for a formal empirical test see Longin, 1996, Loretan and Phillips, 1994, as well as Jondeau and Rockinger, 2003c, for emerging markets).²

Generalized extreme value distribution

Sometimes it is useful to nest the three distributions. This is achieved by the *generalized extreme value distribution* (gev) introduced by Jenkinson (1955). It is defined as

$$H_\xi(y) = \begin{cases} \exp(-(1 + \xi y)^{-1/\xi}) & \text{if } \xi \neq 0 \text{ and } 1 + \xi y > 0, \\ \exp(-\exp(-y)) & \text{if } \xi = 0. \end{cases} \quad (7.1)$$

We then notice that the standard extreme value distributions can be recovered with

² Returns cannot be beyond -100% , which corresponds to bankruptcy. For this reason, the left tail of the return support is bounded. However, as long as on average returns are far away from this boundary, fat tails are a possibility for a given sample.

$$\begin{aligned}\xi = \alpha^{-1} > 0, & \quad \text{for the Fréchet distribution,} \\ \xi = 0, & \quad \text{for the Gumbel distribution,} \\ \xi = -\alpha^{-1} > 0, & \quad \text{for the Weibull distribution.}\end{aligned}$$

The parameter ξ is called the *tail index* and $1/\xi$ is called the *shape index*. Our econometric problem is to decide which is the correct distribution of extremes of returns for the data at hand and to estimate the norming constants μ_T , ψ_T , and ξ . We also define the gev distribution for non-standardized extremes, as (we omit index T on μ and ψ to lighten notations when there is no ambiguity)

$$H_{\xi,\mu,\psi}(m) = \begin{cases} \exp\left(-\left(1 + \xi \frac{m-\mu}{\psi}\right)^{-1/\xi}\right) & \text{if } \xi \neq 0 \text{ and } 1 + \xi \frac{m-\mu}{\psi} > 0, \\ \exp\left(-\exp\left(-\frac{m-\mu}{\psi}\right)\right) & \text{if } \xi = 0. \end{cases}$$

Having obtained the limit distribution, we can reverse the reasoning and ask what type of cdf F_X will have a given limit distribution. Embrechts, Klüppelberg, and Mikosch (1997) characterize the domain of attraction of the various standard extreme value distributions. They also provide several examples. The extremes of normal and log-normal distributions or the mixture of normal (or log-normal) distributions converge to the Gumbel. Cauchy, Pareto, Stable (with exponent smaller than 2) or Student t (with $\nu < \infty$) distributions are in the domain of attraction of the Fréchet distribution. Finally, uniform and beta distributions are in the domain of attraction of the Weibull distribution. For ARCH processes, Kesten (1973) and Jansen and de Vries (1991) have derived the Fréchet distribution as the limit distribution (see also Stărică and Pictet, 1999).

Inspection of the expression $H_{\xi,\mu,\psi}(m)$ shows that a change in the location parameter, μ , shifts the density by a same amount. Next, as the scale parameter increases, the width of the density increases. Thus, this parameter appears to be very similar to a variance. Last, we investigate the impact of a change of the tail index. Figure 7.2 displays several Fréchet distributions. The parameters are indicated in the figure. We notice that as the tail index increases, the probability mass in the right tail increases. This figure shows that ξ describes the behavior of the tails.

To summarize, the tail index focuses on *large* extreme returns. The location parameter indicates where extremes are located on average, i.e., where they cluster. Last, the scale parameter indicates how dispersed most extreme realizations are.

Graphical analysis

Assume that a sample of T observations $\{x_t\}_{t=1}^T$ is available. We construct N -histories (sometimes called *blocks*), that is τ non-overlapping subsamples of length N , with $\tau = \lfloor T/N \rfloor$ (where $\lfloor \cdot \rfloor$ denotes the integer part). Therefore, the initial sample can be rewritten as

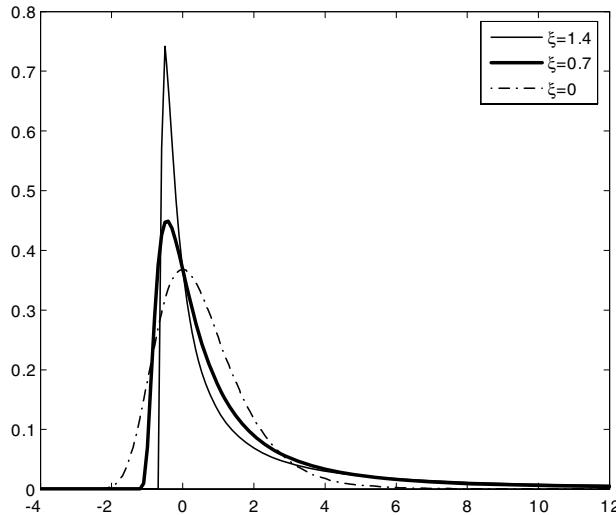


Fig. 7.2. Pdf of various Fréchet distributions.

$$\{x_1, \dots, x_N | x_{N+1}, \dots, x_{2N} | \dots | x_{(\tau-1)N+1}, \dots, x_{\tau N}\}.$$

For instance, the length N may be chosen to cover (approximately) monthly or quarterly data. Now we consider the maximum over each N -history. The maximum of the i th subsample is defined as

$$m_i = \max(x_{(i-1)N+1}, \dots, x_{iN}), \quad \forall i = 1, \dots, \tau.$$

This provides us with a new sample of maxima $\{m_i\}_{i=1}^\tau$.

Figure 7.3 displays the time series of the SP500 daily return and the time series of minima and maxima over subsamples of $N = 60$ days. Using this sample of extrema, the histogram is a simple way to visualize the type of tail behavior. Figure 7.4 displays such histogram for minima and maxima returns. The comparison of this figure with the distributions displayed in Figure 7.1 allows an educated *guess* of the type of extreme value distribution we are dealing with. Gumbel (1958) as well as Embrechts, Klüppelberg, and Mikosch (1997) insist on the importance of such a graphical analysis.

A somewhat more quantitative approach is the quantile plot (or QQ-plot). We define the inverse function of a continuous cdf F as

$$F^{-1}(p) = U_p, \quad \text{for } 0 < p < 1.$$

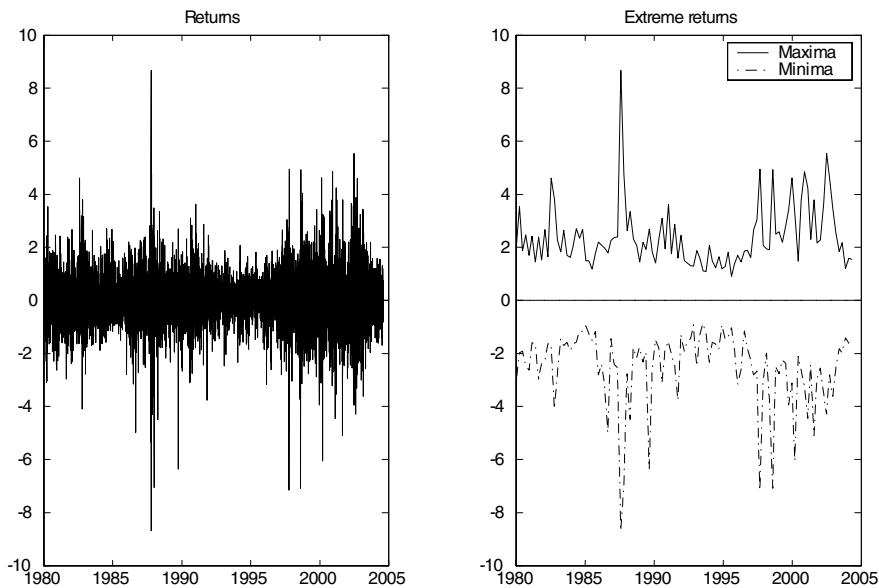


Fig. 7.3. Evolution of SP500 daily returns and quarterly extreme returns.

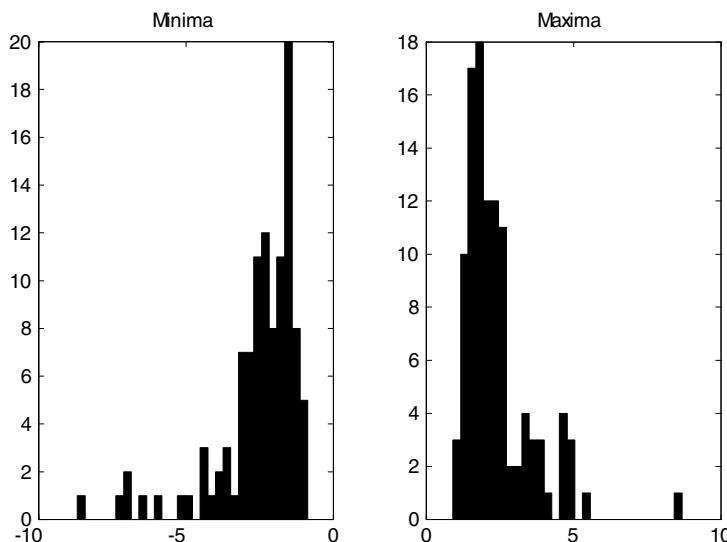


Fig. 7.4. Histogram of SP500 quarterly minima and maxima.

This function is also called the quantile function of the F . It can be interpreted as the value of X such that p percent of the distribution is below this value.

We know that if the distribution function of X is in the domain of attraction of $H_{\xi,\mu,\psi}$, for given ξ , μ and ψ , then the maxima m_i , $i = 1, \dots, \tau$, have $H_{\xi,\mu,\psi}$ as limit distribution, where μ and ψ are normalization constants. If the data is truly generated by a *cdf* that belongs to the domain of attraction of $H_{\xi,\mu,\psi}$, then empirical quantiles of m and theoretical quantiles $H_{\xi,\mu,\psi}^{-1}$ should match approximately. This is the idea of the so-called quantile plot. To implement this idea formally, let $\{m_1, \dots, m_\tau\}$ be the maxima over N -histories and define the vector of ordered maxima $\{m_{1,\tau}, \dots, m_{\tau,\tau}\}$ such that $m_{1,\tau} \leq \dots \leq m_{\tau,\tau}$. The empirical quantiles are given by $\{m_{1,\tau}, \dots, m_{\tau,\tau}\}$. The corresponding theoretical quantiles are obtained as $\left\{H_{\xi,\mu,\psi}^{-1}\left(\frac{1}{\tau+1}\right), \dots, H_{\xi,\mu,\psi}^{-1}\left(\frac{\tau}{\tau+1}\right)\right\}$. The plot $\left\{m_{i,\tau}, H_{\xi,\mu,\psi}^{-1}\left(\frac{i}{\tau+1}\right)\right\}$ is referred to as the *quantile plot*.³

For instance, if we want to test if $\{x_1, \dots, x_T\}$ has a distribution that belongs to the domain of attraction of (say) the Gumbel distribution. In such case, we have

$$H_{\xi,\mu,\psi}(m) = \exp(-\exp(-\frac{m-\mu}{\psi})),$$

so that the quantile of the distribution is

$$H_{\xi,\mu,\psi}^{-1}\left(\frac{i}{\tau+1}\right) = -\log\left(-\log\left(\frac{i}{\tau+1}\right)\right).$$

Finally, if the data is truly generated by a Gumbel distribution, the plot of $m_{i,\tau}$ against $-\log\left(-\log\left(\frac{i}{\tau+1}\right)\right)$ should be roughly linear.

To check for linearity, it can be useful to trace the OLS regression line of the fit of the theoretical on the empirical quantiles. If we take the Gumbel distribution as reference distribution, we have the following conclusions: If the quantile plot is concave, then the limit distribution is Fréchet. If the quantile plot is convex, the limit distribution is Weibull.

Figure 7.5 presents the quantile plot for two polar cases. On the left side, the data is simulated from a Uniform $U(0, 1)$, whereas on the right side it is simulated from a Student t distribution with $\nu = 3$ degrees of freedom. Figure 7.6 presents the same quantile plot for SP500 daily returns. It clearly illustrates that the distributions of minima and maxima probably belong to the domain of attraction of the Fréchet distribution.

³ At this stage, the normalization constants are unknown. Since we are only interested in checking linearity of the relationship, we can standardize m in the usual way as $(m - \bar{m}) / \sigma_m$.

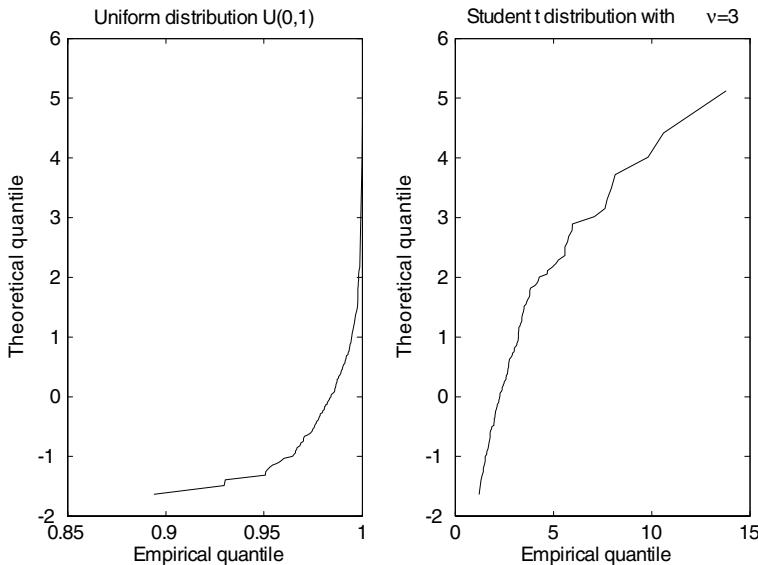


Fig. 7.5. Quantile plot for two samples of simulated data.

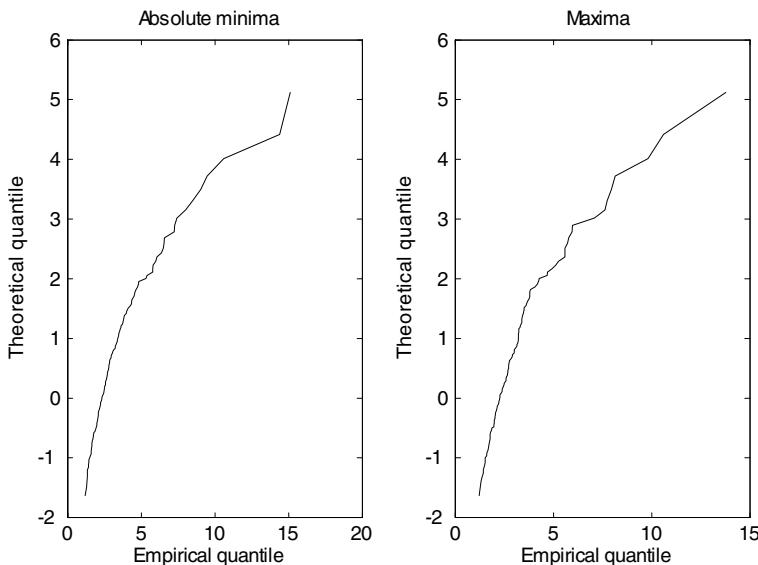


Fig. 7.6. Quantile plot for SP500 daily returns.

Estimation of the gev distribution

Let $\{m_i\}_{i=1}^\tau$ be a vector of extremes of the initial time series $\{x_t\}_{t=1}^T$. To estimate the parameters of the gev distribution, we assume that $\{x_t\}_{t=1}^T$ has the gev $H_{\xi,\mu,\psi}$ as exact distribution and that the maxima $\{m_i\}_{i=1}^\tau$ are *iid*. These two assumptions have important consequences.

The first assumption is required for the estimation of the gev parameters. Often, we may be interested in the estimation of the tail index ξ only. In such case, we do not need that $\{x_t\}_{t=1}^T$ is exactly distributed as a $H_{\xi,\mu,\psi}$, and it is enough to assume that the distribution of $\{x_t\}_{t=1}^T$ belongs to the domain of attraction of $H_{\xi,\mu,\psi}$. We will turn to this issue in the next section.

Second, it is worth emphasizing that it is not required that the original data $\{x_t\}_{t=1}^T$ is itself *iid*. For instance, if we compute maxima over quarterly subsamples, it may well be that returns are dependent during the quarter, but that the quarterly subsamples are *iid*. Obviously, if maxima are computed over *iid* subsamples, they are themselves *iid*. The horizon N may therefore be chosen so that the maxima are broadly *iid*.

We now describe two standard estimation procedures for the gev parameters, the standard ML estimation and the regression-based estimation. Notice that an alternative procedure, based on the method of probability-weighted moments has been proposed by Hosking, Wallis, and Wood (1985).

ML estimation

The ML estimation has been described for instance by Tiago de Oliveira (1973), Smith (1985) or Longin (1996). Let $\theta = (\xi, \mu, \psi)' \in \Theta = \mathbb{R} \times \mathbb{R} \times \mathbb{R}^+$ be the vector of unknown parameters and denote $H_\theta = H_{\xi,\mu,\psi}$. Then, the likelihood of the gev distribution is simply given by the expression

$$L_\tau(\theta|m_i, i = 1, \dots, \tau) = \prod_{i=1}^\tau h_\theta(m_i) \mathbf{1}_{\{1 + \xi \frac{m_i - \mu}{\psi} > 0\}}(m_i),$$

where $h_\theta(m)$ is the *pdf* of the gev distribution, taking three values depending on the situation $\xi \neq 0$ and $1 + \xi \frac{m - \mu}{\psi} > 0$, the situation $\xi = 0$, and to all other situations, as indicated by

$$h_\theta(m) = \begin{cases} \frac{1}{\psi} \left(1 + \xi \frac{m - \mu}{\psi}\right)^{-\frac{1}{\xi}-1} \exp\left(-\left(1 + \xi \frac{m - \mu}{\psi}\right)^{-\frac{1}{\xi}}\right), \\ \frac{1}{\psi} \exp\left(-\frac{m - \mu}{\psi} - e^{-\frac{m - \mu}{\psi}}\right), \\ 0. \end{cases}$$

The ML estimation is obtained by maximizing the log-likelihood

$$L_\tau(\theta|m_i, i = 1, \dots, \tau) = \sum_{i=1}^\tau \ell_i(\theta),$$

where $\ell_i(\theta)$ denotes the log-likelihood of a single observation. In the case $\xi = 0$ for instance, we have

$$\ell_i(\theta) = -\log(\psi) - \frac{m_i - \mu}{\psi} - e^{-\frac{m_i - \mu}{\psi}},$$

while for $\xi \neq 0$ we have

$$\ell_i(\theta) = -\log(\psi) - \left(\frac{1}{\xi} + 1\right) \log\left(1 + \xi \frac{m_i - \mu}{\psi}\right) - \left(1 + \xi \frac{m_i - \mu}{\psi}\right)^{-\frac{1}{\xi}}, \quad (7.2)$$

if $1 + \xi \frac{m_i - \mu}{\psi} > 0$ and 0 otherwise.

The log-likelihood is very nonlinear. Consequently, the estimation has to be performed using standard nonlinear optimization procedure. It turns out (see Smith, 1985) that the usual asymptotic properties hold whenever $\xi > -1/2$. For financial returns, we typically obtain $\xi \geq 0$, so that this limitation should not be a concern in finance.

Expression (7.2) correctly describes the likelihood if the underlying variables are *iid*. In empirical work the use of m -histories with m sufficiently large is likely to yield uncorrelated realizations. As we will see later on, the estimates fluctuate according to the method chosen, suggesting that for financial data the assumption of an identical distribution may not hold.

Regression-based estimation

The method underlying the QQ-plot presented above can be pushed a bit further to yield estimates of gev parameters ξ , μ and ψ (see Gumbel, 1958). This estimation procedure is based on the property

$$E[H_{\xi, \mu, \psi}(m_{i, \tau})] = \frac{i}{\tau + 1}, \text{ for } i = 1, \dots, \tau.$$

This property has been implicitly used for the construction of the QQ-plot. This expression suggests that to estimate the parameters ξ , μ , and ψ one could minimize the distance between the ordered statistics $H_{\xi, \mu, \psi}(m_{i, \tau})$ and their theoretical counterpart $i / (\tau + 1)$.

Since the expression of the gev distribution depends on the value of ξ , we consider the cases $\xi \neq 0$ and $\xi = 0$ in turn. When $\xi \neq 0$, the gev distribution is given by

$$H_{\xi, \mu, \psi}(m) = \exp\left(-\left(1 + \xi \frac{m - \mu}{\psi}\right)^{-1/\xi}\right) \quad \text{if } 1 + \xi \frac{m - \mu}{\psi} > 0.$$

Thus, if we take logs twice and introduce a white noise ε_i , we obtain

$$\log\left(-\log\left(\frac{i}{\tau + 1}\right)\right) = -\frac{1}{\xi} \log\left(1 + \xi \frac{m_{i, \tau} - \mu}{\psi}\right) + \varepsilon_i, \quad i = 1, \dots, \tau.$$

This equation can now be fitted to the sorted sample of maxima with a standard nonlinear optimization procedure. Then, the regression-based estimation is obtained by solving the problem

$$\min_{\theta \in \Theta} \sum_{i=1}^{\tau} \left(\log \left(-\log \left(\frac{i}{\tau+1} \right) \right) + \frac{1}{\xi} \log \left(1 + \xi \frac{m_{i,\tau} - \mu}{\psi} \right) \right)^2.$$

When $\xi = 0$, the gev distribution reduces to

$$H_{\xi,\mu,\psi}(m) = \exp \left(-\exp \left(-\frac{m - \mu}{\psi} \right) \right).$$

If once again we take logs twice and introduce a random noise ε_t , we obtain

$$\log \left(-\log \left(\frac{i}{\tau+1} \right) \right) = -\frac{m_{i,\tau} - \mu}{\psi} + \varepsilon_i \quad i = 1, \dots, \tau.$$

This methodology has been applied by Longin (1996) to show that asset return tails are following a Fréchet distribution.

Illustration

We illustrate some of the results presented above with the daily returns of the SP500, DAX, FT-SE, and Nikkei over the period from 1980 to 2004. For each series, we compute minima and maxima over subsamples of approximately a quarter (60 observations). Then, we estimate the parameters $(\xi, \mu, \psi)'$ of the gev distribution using the ML approach. Notice that we report the estimation of the parameters pertaining to minima with and without the observation of the 1987's crash. This event has been claimed to have very markedly affected the distribution of returns.

Parameter estimates are reported in Table 7.1 (standard errors are in parentheses). We first notice that excluding the 1987's crash only affects the distribution of minima for the SP500, with a decrease of the tail index ξ from 0.2631 to 0.3287. For the other markets, changes are barely noticeable. Second, the estimates obtained for the markets are in general rather close. A notable difference lies in the tail index of the Nikkei that is significantly smaller than the index obtained for other markets. In contrast, its dispersion parameter ψ is larger. This suggests that there are less very large extrema, but that the range of extremes is wider. Third, the differences between the parameters of the left tail and the right tail are rather small. There is no systematic pattern. If we compare this result with Figure 7.4, this result seems reasonable. This type of behavior has also been found by Jondeau and Rockinger (2003c).

7.1.2 Tail distribution

The approach based on the modeling of the tails of the distribution (also called peak-over-threshold approach) was introduced by Balkema and de Haan (1974), Pickands (1975), Davison (1984), and Hosking and Wallis (1987). See also Rootzén and Tajvidi (1996) and McNeil (1997).

Table 7.1. Estimates of the parameter ξ using the gev distribution

	Minima (no 1987 crash)	Minima	Maxima
SP500			
ξ	0.2631 (0.0827)	0.3287 (0.0832)	0.2410 (0.0818)
μ	1.8365 (0.0859)	1.7999 (0.0851)	1.8361 (0.0719)
ψ	0.7560 (0.0703)	0.7565 (0.0725)	0.6406 (0.0582)
DAX			
ξ	0.2641 (0.0827)	0.2822 (0.0839)	0.3289 (0.0915)
μ	2.1658 (0.1109)	2.1473 (0.1117)	2.0718 (0.0956)
ψ	1.0159 (0.0912)	1.0104 (0.0922)	0.8570 (0.0823)
FT-SE			
ξ	0.2835 (0.0818)	0.286 (0.0818)	0.2956 (0.0898)
μ	1.6466 (0.0633)	1.6525 (0.0629)	1.4968 (0.0609)
ψ	0.5627 (0.0530)	0.5564 (0.0528)	0.5356 (0.0510)
Nikkei			
ξ	0.1746 (0.0865)	0.1746 (0.0865)	0.1775 (0.0930)
μ	2.1265 (0.1256)	2.1261 (0.1256)	2.2545 (0.1400)
ψ	1.1351 (0.0986)	1.1351 (0.0986)	1.2007 (0.1113)

Mean excess function

Once again, we consider a sample of *iid* observations $\{x_t\}_{t=1}^T$ with distribution F_X . We are interested in the distribution followed by excesses over a given threshold. Let x_F be the right endpoint of F .

Definition 7.2. Let $u < x_F$ be a fixed real number, the threshold, in the support of X_t . The function

$$F_u(y) = \Pr[X_t - u \leq y | X_t > u] = \frac{F_X(y+u) - F_X(u)}{1 - F_X(u)}, \quad 0 \leq y < x_F - u$$

is called the excess distribution function (edf) of the random variables X_t over the threshold u . The function

$$e(u) = E[X_t - u | X_t > u]$$

is called the mean-excess function (mef).

The excess distribution function measures the probability that the excess realization relative to the threshold ($X_t - u$) is below a certain value y , given that the realization is above the threshold u . The mean-excess function averages those realizations that exceed u and considers the distance between the mean and u . For instance, we have the following mean-excess functions for some standard distributions F_X

$$\begin{aligned} \text{Pareto: } & (k+u)/(\alpha-1), \text{ for } \alpha > 1 \\ \text{Weibull: } & (u^{1-\tau})/\sigma\tau \\ \text{Exponential: } & 1/\lambda \end{aligned}$$

Generalized Pareto distribution

An important result, established by Balkema and de Haan (1974) and Pickands (1975), relates the limit distribution of the scaled excesses over threshold to the tail index ξ .

Theorem 7.3. *If F_X is in the domain of attraction of the extreme value distribution H_ξ , the excess distribution function $F_u(y)$ can be approximated, for u large, by*

$$F_u(y) \approx G_{\xi,\psi}(y), \quad u \rightarrow \infty,$$

where

$$G_{\xi,\psi}(y) = \begin{cases} 1 - \left(1 + \frac{\xi}{\psi}y\right)^{-1/\xi}, & \text{if } \xi \neq 0, \\ 1 - \exp\left(-\frac{y}{\psi}\right), & \text{if } \xi = 0, \end{cases}$$

for $0 \leq y \leq x_F - u$, is the so-called generalized Pareto distribution (gpd).

The support for y is $y \geq 0$ if $\xi \geq 0$, and $0 \leq y \leq -1/\xi$ if $\xi < 0$. This means that the gpd is the limit distribution of the scaled excess function $F_u(y)$. The term ψ is a positive scaling function of the threshold u .

If we define $x = y + u$, we can also express the gpd as a function of x

$$G_{\xi,u,\psi}(x) = \begin{cases} 1 - \left(1 + \frac{\xi}{\psi}(x-u)\right)^{-1/\xi}, & \text{if } \xi \neq 0, \\ 1 - \exp\left(-\frac{(x-u)}{\psi}\right), & \text{if } \xi = 0. \end{cases}$$

Clearly, the tail index ξ from the gev is the same as for the gpd. An illustration of this density can be found in Figure 7.7 for a set of parameters that typically arises in finance ($u = 0$ and $\psi = 1$). An increase of ξ , for a constant level of the scale factor ψ , increases the tail while steepening the slope at the more central part of the density. Those results indicate that if we choose a high enough threshold, then the fit of a gpd to the tail realizations will also yield an estimate of the tail index. If we are not interested in the

central part of the density, estimating the gpd provides an estimation of the behavior of the tails that is not affected by any assumption concerning the entire density.

One of the advantages of using the gpd is that it provides a density of the tail. This may be a useful approach if the focus is on that part of the density rather than on the entire density. Indeed, if we focus on the entire density, for instance by selecting for convenience the one or the other standard density, we may not be able to get the right tail shape. The estimations of the gev distribution in the previous section suggest that the empirical tails probably look like the curve obtained with $\xi = 0.25$.

As an illustration, we plot in Figure 7.8 the histogram of the 500 smallest returns (lower or left tail) and the 500 largest returns (upper or right tail). The histogram can be compared with the distributions displayed in Figure 7.7 to get an insight of the range for ξ . Figure 7.9 displays the mean excess function for the 500 observations in the two tails (we omit the 50 most extreme realizations to get a nice figure). This figure suggests that the tails of the distribution may be drawn from a Pareto distribution.

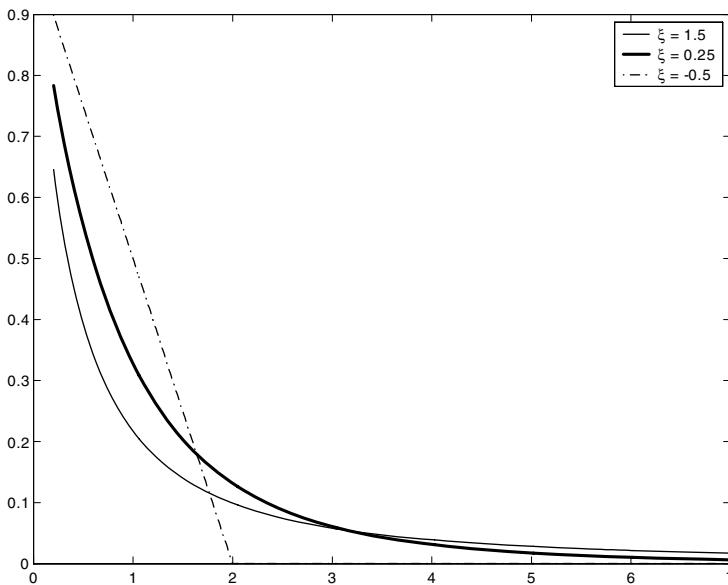


Fig. 7.7. Pdf of the gpd for various values of ξ .

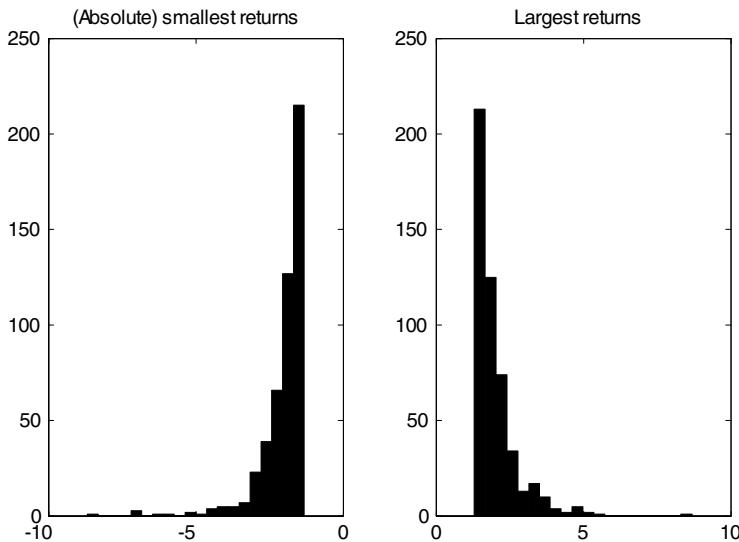


Fig. 7.8. Histogram of the lower and upper tails of SP500 daily returns.

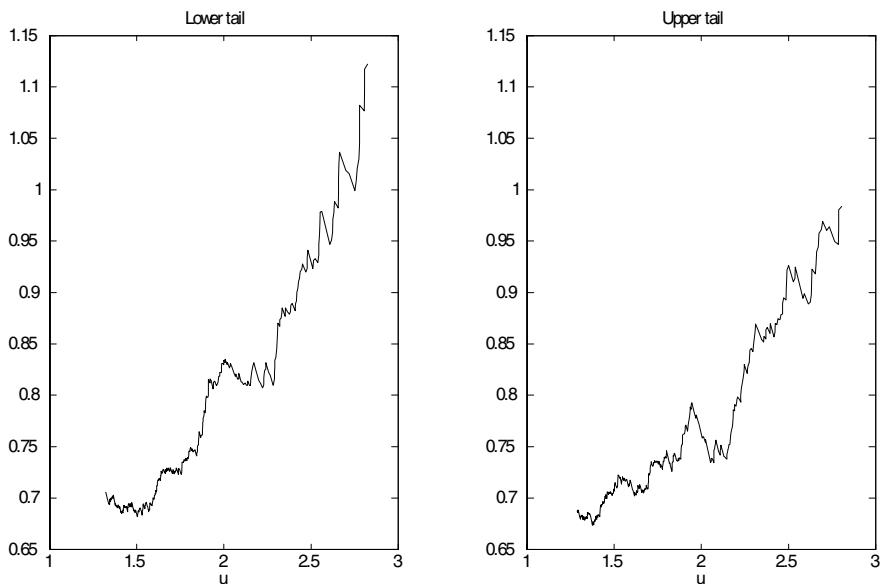


Fig. 7.9. Mean excess function for the lower and upper tails of SP500 daily returns.

Estimation of the gpd

In Section 7.1.1, the estimation of the tail index ξ was based on a sample of extremes, assuming that the exact distribution of returns was the gev distribution H_ξ . Now, we provide an alternative approach to estimating the tail index, only assuming that the distribution of (X_1, \dots, X_T) is in the domain of attraction of H_ξ . It should be noticed however that the estimation procedures described in this section rely on the assumption that (X_1, \dots, X_T) is *iid*. This approach has been developed by Smith (1987), and Davison and Smith (1990).

Selection of the optimal threshold u

The main difficulty in the use of the tail approach is the choice of the threshold u . This threshold has to be high enough as to ensure that the limit distribution of the excess distribution function is actually a gpd. This suggests the use of a rather large threshold. However, as u becomes larger, there are fewer and fewer observations included in the gpd estimation. This makes the estimation of ξ and ψ unstable, with large standard errors. On the other hand, as u decreases, the chance of an observation that does not belong to the tail distribution being included in the gpd estimation increases. Therefore, u should also be selected in such a way that the number of observations above the threshold be sufficient to ensure an accurate estimation of the unknown parameters. Obviously, the properties of the estimators in finite sample depend on the number of observations used in the estimation of the parameters of interest, not on the total number of observations.

Unfortunately, there is no optimal way to select the threshold u , although several tools have been proposed to get an idea of this threshold. A simple graphical method to recover parameters ξ and ψ can be based on the empirical mean excess function $e(u)$. Indeed, as was shown by Daragahi-Noubary (1989), the mean excess function of the gpd with $\xi < 1$ can be written as

$$e(u) = E[X - u | X > u] = \frac{\psi + \xi u}{1 - \xi}, \quad (7.3)$$

so that it is a linear function in u .

An estimate of the mef $e(u)$, denoted $\hat{e}(u)$, can be obtained from a sample (x_1, \dots, x_T) as the sum of the excesses over a threshold divided by the number of realizations exceeding the threshold

$$\hat{e}(u) = \frac{1}{N_u} \sum_{t=1}^T (x_t - u) 1_{\{x_t > u\}}(x_t), \quad u > 0,$$

where $N_u = \sum_{t=1}^T 1_{\{x_t > u\}}(x_t)$. Then, the *mean-excess plot* is defined as the set of pairs

$$\{(x_{t,T}, \hat{e}(x_{t,T})) \text{ where } t = 1, \dots, T\},$$

and where $\{x_{t,T}\}_{t=1}^T$ is the sample of ordered observations. The relation (7.3) suggests that the threshold u should be chosen such that the plot of $\hat{e}(x_{t,T})$ is roughly linear for $x_{t,T} \geq u$.

The mean-excess plot allows a distinction between series with thin-tailed distribution (decreasing mean-excess plot, since the extremes are not very far from the center of the distribution), and series with fat-tailed distribution (increasing mean-excess plot). Moreover, there exists a one-to-one relation between the parameters ξ and ψ and the intercept and slope of the mef in (7.3). A simple regression of $\hat{e}(x_{t,T})$ on $x_{t,T}$ and an intercept provides the parameters ξ and ψ . Once an empirical mef has been obtained, comparison with theoretical shapes allows identifying the type of distribution we are dealing with.

ML estimation

Details on the ML estimation can be found in Smith (1987). The vector of unknown parameters is $\theta = (\xi, \psi) \in \Theta = \mathbb{R} \times \mathbb{R}^+$. Let $\{x_{1,T}, \dots, x_{T,T}\}$ be the vector of returns sorted by increasing order, such that $x_{1,T} \leq \dots \leq x_{T,T}$. We then select the realizations x_t that are above the threshold u . We define $N_u = \sum_{t=1}^T 1_{\{x_t > u\}}(x_t)$ the number of exceedances. For ease of exposition, we define the $(N_u, 1)$ vector of exceedances as $\{v_1, \dots, v_{N_u}\} \equiv \{x_{T-N_u+1,T}, \dots, x_{T,T}\}$. Then, the likelihood of the gpd distribution is given by the expression

$$L_{N_u}(\theta | v_i, i = 1, \dots, N_u) = \prod_{i=1}^{N_u} g_{\xi, u, \psi}(v_i) 1_{\left\{1 + \frac{\xi}{\psi}(v_i - u) > 0\right\}}(v_i),$$

where $g_{\xi, u, \psi}(v)$ is the *pdf* of the gpd distribution, given by

$$g_{\xi, u, \psi}(v) = \begin{cases} \psi^{-1} \left(1 + \frac{\xi}{\psi}(v - u)\right)^{-\frac{1}{\xi}-1} & \xi \neq 0, \\ \psi^{-1} \exp\left(-\frac{v-u}{\psi}\right) & \xi = 0. \end{cases}$$

The ML estimation is obtained by maximizing the log-likelihood

$$L_{N_u}(\theta | v_i, i = 1, \dots, N_u) = \sum_{i=1}^{N_u} \ell_i(\theta),$$

with the log-likelihood of a single observation $\ell_i(\theta)$ defined as

$$\ell_i(\theta) = \begin{cases} -\log(\psi) - \left(\frac{1}{\xi} + 1\right) \log\left(1 + \frac{\xi}{\psi}(v_i - u)\right), \\ -\log(\psi) - \frac{(v_i - u)}{\psi}. \end{cases}$$

Interestingly, Smith (1987) has derived the asymptotic distribution of $\hat{\theta}_{ML}$ provided $\xi > -1/2$ and under the assumption of *iid* observations. We have

$$\sqrt{T} (\hat{\theta}_{ML} - \theta) \Rightarrow \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \frac{1}{1+\xi} \begin{pmatrix} 1+\xi & 1 \\ 1 & 2 \end{pmatrix}^{-1} \right).$$

The advantage of this approach as compared with the extreme one is that we can use all realizations exceeding u and not only the maxima over m -histories. Consequently, this method is likely to yield more efficient estimates. The selection of the threshold u is, unfortunately, a drawback of this approach.

Illustration

Estimates of the parameters of the gpd $(\xi, \psi)'$ for the lower tail and the upper tail are reported in Table 7.2. As for the estimation of the gev distribution, we report the estimation of the parameters of the lower tail with and without the observation of the October 1987 crash. Once again, we obtain that this observation only affects the estimation of the left tail of the SP500 distribution.

Table 7.2. Parameter estimates of the gpd distribution

	Lower tail (no 1987 crash)	Lower tail	Upper tail
SP500			
ξ	0.2032 (0.0629)	0.2759 (0.0641)	0.1151 (0.0669)
ψ	0.5702 (0.0483)	0.5493 (0.0450)	0.6338 (0.0535)
DAX			
ξ	0.1397 (0.0651)	0.1550 (0.0641)	0.2038 (0.0737)
ψ	0.9991 (0.0854)	1.0010 (0.0878)	0.7916 (0.0723)
FT-SE			
ξ	0.2245 (0.0692)	0.2640 (0.0703)	0.1246 (0.0568)
ψ	0.5635 (0.0496)	0.5527 (0.0473)	0.5467 (0.0435)
Nikkei			
ξ	0.1209 (0.0515)	0.1218 (0.0511)	0.1149 (0.0591)
ψ	0.8136 (0.0613)	0.8112 (0.0611)	0.9009 (0.0734)

We also observe more dispersion in the parameter estimates among markets. On the left side, the tail indices for the DAX and the Nikkei returns are found to be significantly lower than those obtained for the SP500 and the FT-SE. In contrast, the dispersion parameters ψ are much larger. This

suggests, once again, that there are less very large extrema but that their range is wider. For the SP500 and the FT-SE, the tail indices ξ on the right tail are found to be smaller than those on the left tail. If we compare with Figure 7.4, this result seems reasonable, reflecting the large asymmetry in the distribution.

Figure 7.10 displays the adjustment of the *cdf* of the estimated gpd to the actual *cdf* of the tails of the SP500 returns. The fit appears broadly correct.

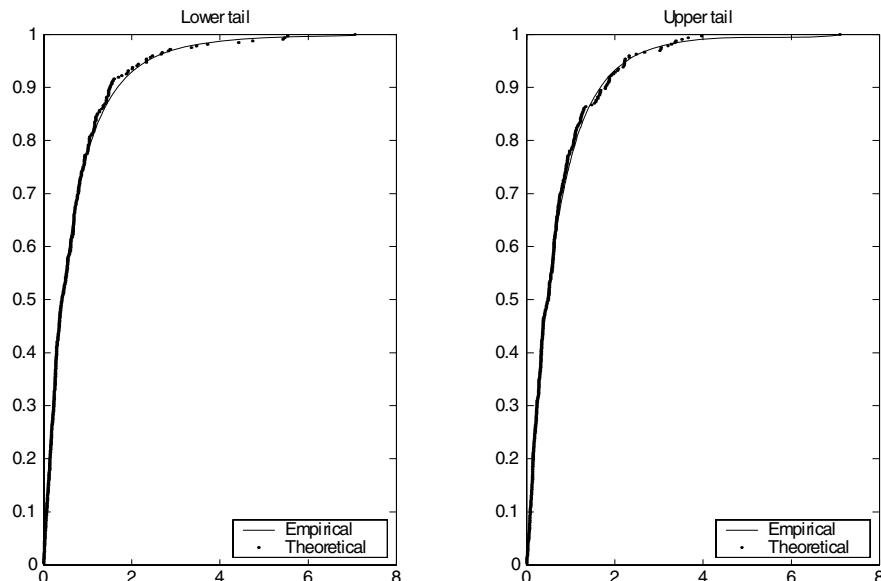


Fig. 7.10. Comparison of the empirical and theoretical *cdf* of the gpd.

Additional results

The fact that returns are leptokurtic has lead to their modeling with fat-tailed distributions. Some of these distributions, such as Mandelbrot's (1963) stable law, do not allow for a finite variance. Given the empirical behavior of the tails of return distribution, the issue of the *existence of moments* can again be addressed within the tail approach.⁴ Embrechts, Klüppelberg, and Mikosch (1997, p. 165) recall that, if X is distributed with a tail index ξ , then for all integers r such that $r < 1/\xi$, the r th moment exists with

⁴ See also Longin (1996, p. 399). He concludes that the U.S. daily returns allow up to the third moment but not beyond.

$$E[X^r] = \frac{\psi^r}{\xi^{r+1}} \frac{\Gamma(\xi^{-1} - r)}{\Gamma(\xi^{-1} + 1)} r!.$$

Once the tail distribution has been characterized, it is also possible to compute the *mean waiting time* between specific extreme events. This result is useful for the practitioner who wishes to estimate the average time before a given extreme value occurs. Clearly, an extreme realization is defined as the exceeding of a return of a given level. Let X_t be *iid* random variables with *cdf* G and u a threshold. The fact that $X_t > u$ or not corresponds to a Bernoulli event with success probability $p = 1 - G(u)$. The probability that X_t exceeds the threshold at time k and not before is given by the geometric distribution $p(1 - p)^{k-1}, k = 1, 2, \dots$. As a consequence, the average waiting time before u is crossed is

$$E[\min(t \geq 1 : X_t > u)] = \sum_{k=1}^{+\infty} k p(1 - p)^{k-1} = 1/p.$$

Semi-parametric estimation of the tail index

An alternative approach to the estimation of the gev distribution or the gpd consists in focusing on the tail index only. Such an approach has been developed by Pickands (1975) and Hill (1975). Hill's estimator is very often applied, because of its easy implementation and asymptotic unbiasedness. Applications and extensions of this estimator to finance are by Jansen and de Vries (1991), Hols and de Vries (1991), as well as Koedijk and Kool (1992).

As we have seen earlier, the asymptotic behavior of extreme values depends on the tail index ξ . Depending on the values taken by this index, we will end up with either a Fréchet, a Gumbel, or a Weibull distribution. It is therefore of great importance to be able to quickly characterize the tail index. In this section, we present two simple ways to estimate the index using a semi-parametric approach, which requires neither that the distribution of the X s be a gev (as for the extreme approach) nor that the X s be *iid* (as for the tail approach).

Pickands estimator

This estimator was introduced by Pickands (1975) using some results on limit distributions. Assume that $\{x_1, \dots, x_T\}$ has a distribution function that belongs to the domain of attraction of H_ξ and define ordered realizations $\{x_{1,T}, \dots, x_{T,T}\}$ with $x_{1,T} \leq \dots \leq x_{T,T}$. Then, the Pickands (1975) estimator is defined by

$$\hat{\xi}_{(q,T)}^P = \frac{1}{\log(2)} \log \left(\frac{x_{T-q+1,T} - x_{T-2q+1,T}}{x_{T-2q+1,T} - x_{T-4q+1,T}} \right), \quad \text{for } 1 \leq q \leq T/4,$$

where q is a number of observations located in the tail of the distribution.

Dekkers and De Haan (1989) have shown that, under some restrictions on the asymptotic behavior of the tail size, Pickands estimator is strongly consistent and asymptotically normal, with

$$\sqrt{q} \left(\hat{\xi}_{(q,T)}^P - \xi \right) \Rightarrow \mathcal{N}(0, \nu(\xi)), \quad (7.4)$$

where $\nu(\xi) = \xi^2 (2^{2\xi+1} + 1) / (2 (2^\xi - 1) \log(2))$. Pickands estimator can be computed for all values of q , so that we can draw a Pickands plot

$$\left\{ \left(q, \hat{\xi}_{(q,T)}^P \right); q = 1, \dots, T \right\}.$$

This plot may help selecting the appropriate value of the tail size q , although there is no optimal selection criterion. We will turn to that issue later on. An advantage of this estimator is that it is defined for all $\xi \in \mathbb{R}$. Therefore, it can be used to discriminate between the three extreme value distributions.

Hill estimator

This method, described in Hill (1975), presumes that $\{x_1, \dots, x_T\}$ has a distribution function that belongs to the domain of attraction of the Fréchet extreme value distribution ($\xi > 0$). We consider again $x_{j,T}$ the j th largest value issued from the raw sample data. Hill's (1975) method estimates ξ using

$$\hat{\xi}_{(q,T)}^H = \frac{1}{q} \sum_{j=1}^q \log \left(\frac{x_{T-j+1,T}}{x_{T-q,T}} \right), \quad \text{for } 1 \leq q < T. \quad (7.5)$$

When the X s are *iid*, it has been shown that $\hat{\xi}_{(q,T)}^H$ is strongly consistent and asymptotically normal with distribution

$$\sqrt{q} \left(\hat{\xi}_{(q,T)}^H - \xi \right) \Rightarrow \mathcal{N}(0, \xi^2).$$

For weakly dependent variables, $\hat{\xi}_{(q,T)}^H$ remains weakly consistent. For empirical purposes, the variance ξ^2 can be estimated using $\hat{\xi}_{(q,T)}^H$. It has been shown that if the X_t are generated by a Fréchet distribution then the Hill estimator will yield a more efficient tail index than the Pickands estimator. As for the Pickands estimator, we can draw a Hill plot, that may help to find an appropriate tail size q .

Deckers-Einmahl-de Haan (DEdH) estimator

A drawback of the Hill estimator is that it presumes that the distribution of the initial variables belongs to the domain of definition of the Fréchet distribution. Deckers, Einmahl, and de Haan (1989) have proposed an extension of the Hill estimator that covers the whole class H_ξ . It is defined as

$$\hat{\xi}_{(q,T)}^{DEH} = 1 + \hat{H}_{(q,T)}^{(1)} + \frac{1}{2} \left(\frac{\left(\hat{H}_{(q,T)}^{(1)} \right)^2}{\hat{H}_{(q,T)}^{(2)}} - 1 \right)^{-1}, \quad \text{for } 1 \leq q < T,$$

where the i -th order logarithmic moments $\hat{H}_{(q,T)}^{(i)}$ is of the form

$$\hat{H}_{(q,T)}^{(i)} = \frac{1}{q} \sum_{j=1}^q \log \left(\frac{x_{T-j+1,T}}{x_{T-q,T}} \right)^i. \quad (7.6)$$

Notice that $\hat{H}_{(q,T)}^{(1)}$ is just the Hill estimator defined above. Under *iid*-ness, this estimator is also shown to be strongly consistent and asymptotically normal

$$\sqrt{q} \left(\hat{\xi}_{(q,T)}^{DEH} - \xi \right) \Rightarrow \mathcal{N}(0, \tilde{\nu}(\xi)),$$

with

$$\tilde{\nu}(\xi) = \begin{cases} 1 + \xi^2, & \text{if } \xi \geq 0, \\ (1 - \xi)^2 (1 - 2\xi) \left(4 - 8 \frac{1-2\xi}{1-3\xi} + \frac{(5-11\xi)(1-2\xi)}{(1-3\xi)(1-4\xi)} \right), & \text{if } \xi < 0. \end{cases}$$

Comparison

In Figures 7.11 to 7.13, we present the estimate of the different semi-parametric estimators of the tail index, for q ranging from 50 to 500. As it appears clearly from these figures, there is a large amount of heterogeneity among the estimations. The Pickands estimator suggests a small, insignificant value for the tail index. The Hill estimator provides a tail index in the range [0.25; 0.4] but does not provide any indication on a possible value for the tail size q . Finally, the DEdH estimator appears as the most stable estimator, with an average value for the tail index around 0.25 for the lower tail and 0.2 for the upper tail. These values are slightly smaller than those obtained with the extrema approach.

Selection of the threshold

Asymptotic results for the semi-parametric estimators described above indicate that consistency and asymptotic normality are reached when the tail size q increases with sample size, but at a slower rate. Of course, it does not provide a criterion for selection of the appropriate tail size. On one hand, if we choose a threshold too much in the tail, the estimate is unbiased because the asymptotic theory applies, but we obtain very inaccurate estimates, because just too few observations are used in the estimation. On the other hand, if we use more observations, the variance of the estimator is reduced, but the bias increases, because tail observations are then contaminated by observations from the central part.

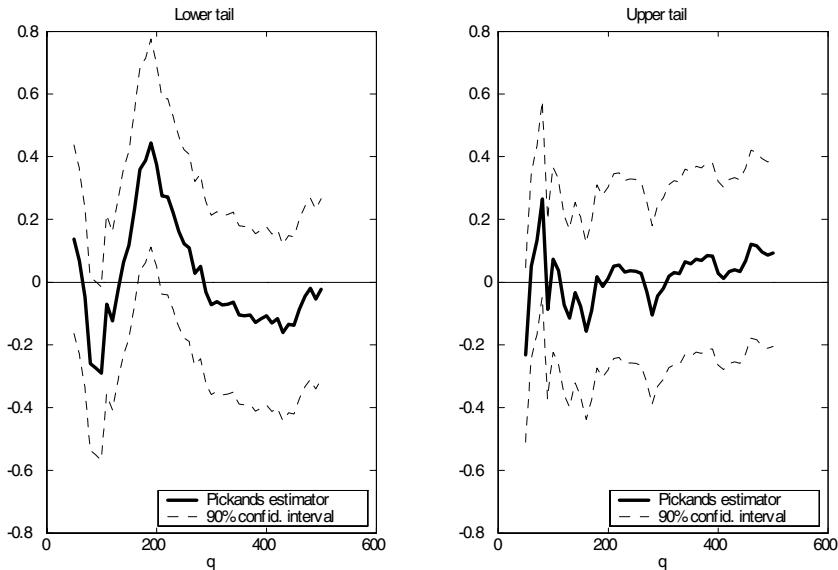


Fig. 7.11. Estimates of the Pickands tail index for various values of u .

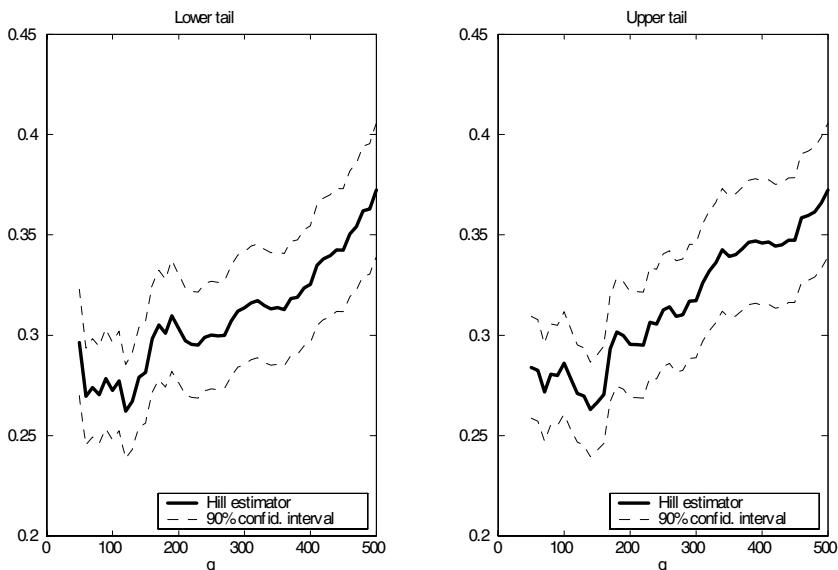


Fig. 7.12. Estimates of the Hill tail index for various values of u .

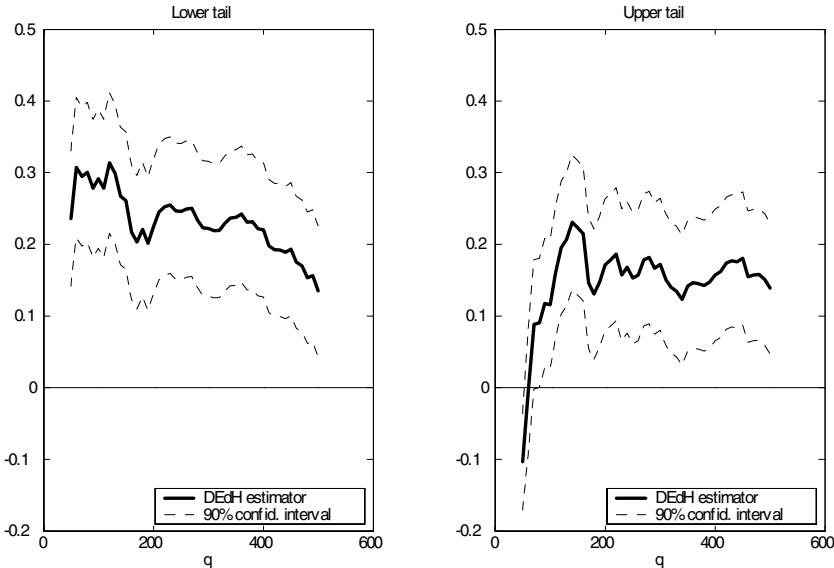


Fig. 7.13. Estimates of the Deckers, Einmahl, and de Haan tail index for various values of u .

Consequently, there is a bias-variance trade-off in the selection of q . As a consequence, there should exist an “optimal” tail size. The first attempts to determine an optimal q^* go back to Hall and Welsh (1985) and Hall (1990). Hall (1990) proposed a subsample bootstrap procedure based on the minimization of the mean squared error of the tail estimate ξ . This approach has been thereafter improved by Danielsson and de Vries (1997). Alternative bootstrap approaches have been developed by Danielsson et al. (2001), Drees and Kaufman (1998), and Beirlant, Vynckier, and Teugels (1996). Lux (2001) compares the different techniques and concludes that the various methods for selecting an optimal q^* provide quite similar outcomes in terms of the resulting ξ , even though the thresholds proposed by each method may not be the same. Hereafter, we detail some of these approaches to give an insight of the way resample bootstrap methods work.

The basic idea of bootstrap procedure for selecting the optimal q^* is the following: the optimal q^* is the number of ordered observations included in the computation of the (say) Hill estimator that minimizes the mean squared error of $\hat{\xi}_{(q,T)}^H$. The optimal q^* is therefore obtained by solving

$$q^* \in \arg \min_q E \left[(\hat{\xi}_{(q,T)}^H - \xi)^2 \right]. \quad (7.7)$$

The reason for the use of the MSE is that it provides, at least in theory, the desired trade-off between bias and variance. The analytical expression for

the MSE can be computed (Hall, 1982) and yields the resulting

$$q^* = \left[\frac{\alpha(\alpha + \beta)^2}{2\beta^3 b^2} \right]^{\frac{\alpha}{\alpha+2\beta}} (aT)^{\frac{2\beta}{\alpha+2\beta}}, \quad (7.8)$$

where $\alpha = 1/\xi$ and a, b and β are additional parameters that are unknown in practice. Since these parameters are unknown, Hall (1990) resorts to the bootstrap technique to minimize the MSE. The idea consists in estimating the empirical MSE by comparing subsample estimates of the tail index with an initial value computed over the full sample. Let $\tilde{\xi}_{(q_0, T)}^H$ denote an initial estimate of the tail index (based on a reasonably chosen q_0) and $T_1 \ll T$ be the subsample size. Then the optimal threshold q_1^* over a subsample of size T_1 is given by

$$q_1^* \in \arg \min_{q_1} E \left[\left(\hat{\xi}_{(q_1, T_1)}^H - \tilde{\xi}_{(q_0, T)}^H \right)^2 \right].$$

In practice, we construct K randomly selected subsamples of size T_1 . Let $k = 1, \dots, K$ be the index of the k th subsample. We compute the Hill estimate $\hat{\xi}_{q_1, T_1}^{H,k}$, $k = 1, \dots, K$ over all simulated subsamples. An approximation of (7.7) is then given by

$$\frac{1}{K} \sum_{k=1}^K \left(\hat{\xi}_{q_1, T_1}^{H,k} - \tilde{\xi}_{(q_0, T)}^H \right)^2.$$

Once the optimal subsample tail size q_1^* is obtained, the full-sample estimate of q^* is deduced by computing q^*/q_1^* , and by using (7.8) to obtain

$$q^* = q_1^* \left(\frac{T}{T_1} \right)^{\frac{2\beta}{\alpha+2\beta}},$$

where we can use the estimate $\hat{\alpha} = 1/\tilde{\xi}_{(q_0, T)}^H$. However, β is still an unknown parameter. Hall (1990) suggested the use of $\hat{\beta} = \hat{\alpha}$ to obtain an estimate of q^* . Danielsson and de Vries (1997) improved this procedure by introducing the following estimator for β . First they construct the following j th empirical log-moments and a variable Δ defined as

$$\begin{aligned} \tilde{m}^{(j)} &= \frac{1}{q_1^*} \sum_{i=1}^{q_1^*} \left[\log \left(\frac{x_{T-i+1, T}}{x_{T-q_1^*, T}} \right) \right]^j, \\ \Delta &= \frac{\tilde{m}^{(1)} - \tilde{m}^{(2)}/2\tilde{m}^{(1)}}{\tilde{m}^{(3)}/3\tilde{m}^{(2)} - \tilde{m}^{(4)}/4\tilde{m}^{(3)}}. \end{aligned}$$

The parameter β is then estimated by

$$\hat{\beta} = \left(\sqrt{\Delta} - 1 \right) / \tilde{\xi}_{(q_0, T)}^H.$$

Finally, the full sample threshold q^* is computed and the optimal Hill tail index is estimated as $\hat{\xi}_{BO}^H = \hat{\xi}_{(q^*, T)}^H$. It should be noticed that in this approach some choices have to be made by the user: the initial tail size q_0 used to estimate $\tilde{\xi}_{(q_0, T)}^H$, the value of T_1 and the number of simulations K . Danielsson and de Vries (1997) chose $q_0 = T/100$, $T_1 = T/10$ and $K = 100$.

Danielsson et al. (2001) describe a slightly different approach for computing the optimal tail size q^* . They observe that one of the drawbacks of minimizing the MSE given by (7.7) is that it requires an initial estimate of ξ . Therefore, they develop an alternative estimator, based on the minimization of the following MSE

$$q_1^* \in \arg \min_{q_1} E \left[\left(\hat{H}_{(q_1, T_1)}^{(2)} - 2 \left(\hat{\xi}_{(q_1, T_1)}^H \right)^2 \right)^2 \right], \quad (7.9)$$

where $\hat{H}_{(q_1, T_1)}^{(2)}$ is the second-order logarithmic moment defined in (7.6). To avoid the additional estimation of the parameter β , they suggest the use of two successive bootstrap exercises. In the first exercise, one draws subsamples of size $T_1 \ll T$ and computes the optimal tail size q_1^* by minimizing the MSE given by (7.9). Then one repeats the same procedure for an even smaller resample size $T_2 = T_1^2/T$ and obtains a second optimal tail size q_2^* . Finally, one computes the optimal tail size for the actual sample

$$q^* = \frac{(q_1^*)^2}{q_2^*} \left(\frac{(\log(q_1^*))^2}{(2 \log(T_1) - \log(q_1^*))^2} \right)^{\frac{\log(T_1) - \log(q_1^*)}{\log(T_1)}},$$

and obtains the optimal Hill tail index as before: $\hat{\xi}_{BO}^H = \hat{\xi}_{(q^*, T)}^H$.

7.1.3 The case of weakly dependent data

Most asymptotic results concerning the estimation of the tail index hold assuming *iid* returns. When returns are still strictly stationary but non-*iid*, these estimators are found to be weakly consistent, at best. Using simulations, Kearns and Pagan (1997) have shown that, with dependent data, the precision of usual tail estimators (such as Hill's) is grossly overstated by conventional significance levels. The case of strictly stationary, non-*iid*, series (such as GARCH processes) has been studied by Leadbetter, Lindgren, and Rootzén (1983), de Haan et al. (1989), or Stărică and Pictet (1999). Embrechts, Klüppelberg, and Mikosch (1997) and McNeil (1998) also provide details and references in the context of the extreme approach applied to financial data.

The extremal index

Assume that X_t is a strictly stationary, non-*iid*, time series, and that \tilde{X}_t is an associated *iid* series with the same marginal distribution F_X . We also de-

fine $M_T = \max(X_1, \dots, X_T)$ and $\tilde{M}_T = \max(\tilde{X}_1, \dots, \tilde{X}_T)$. Then, as shown by Leadbetter, Lindgren, and Rootzén (1983), the standardized extremes of the two series have the same limit distribution H_ξ given in Theorem 7.1 under the two following conditions: (i) the series X has only weak long-range dependence, and (ii) there is no tendency for extremes to cluster.

For financial series, the former condition is rather mild, but the latter is not likely to apply. For instance, GARCH processes are weakly dependent, but have clustering of extremes. When one of the two conditions is not satisfied, the limit distribution of the standardized extremes of X_t is shown to be

$$\lim_{T \rightarrow \infty} \Pr\left(\frac{M_T - \mu_T}{\psi_T} \leq y\right) = H^\theta(y).$$

The parameter θ , called the *extremal index*, measures the relationship between the dependence structure and the extremal behavior of the process (see Hsing, Hüsler, and Leadbetter, 1988, Embrechts, Klüppelberg, and Mikosch, 1997, chap. 8). The extremal index verifies $0 \leq \theta \leq 1$. The case $\theta = 1$ corresponds to weak dependence and independence. Intuitively, the extremes over subsamples of length N from a strictly stationary series with extremal index $\theta < 1$ have the same behavior as the extremes over subsamples of length $N\theta$ from the corresponding *iid* series. We therefore have the following result, established by McNeil (1998):

Theorem 7.4. *Let X_t be a stationary process with extremal index $\theta > 0$. Then,*

$$\lim_{T \rightarrow \infty} \Pr\left(\frac{\tilde{M}_T - \mu_T}{\psi_T} \leq y\right) = H_\xi(y), \quad \forall y \in \mathbb{R},$$

for a non-degenerate cdf H if and only if

$$\lim_{T \rightarrow \infty} \Pr\left(\frac{M_T - \mu_T}{\psi_T} \leq y\right) = H_\xi^\theta(y), \quad \forall y \in \mathbb{R},$$

with H^θ also non-degenerate.

An important consequence of this result is that the theoretical value of the tail index is not affected by the strong dependence of X . Indeed, only the location and dispersion parameters are affected, since we have $H_\xi^\theta(y) = H_\xi(cy + d)$ for some $c \in \mathbb{R}^+$ and $d \in \mathbb{R}$. Consequently, provided we accept a change in the norming constants μ_T and ψ_T , the limit distribution under dependence H_ξ^θ can be chosen identical to the one that would prevail with *iid* data H_ξ .

Hsing, Hüsler, and Leadbetter (1988) and Hsing (1993) indicate how to obtain estimates of the extremal index. Basically, for a given high threshold (say u), the extremal index θ can be estimated by dividing the number of subsamples of size N in which the maximum exceeds the threshold u (denoted K_u) by the total number of exceedances in those subsamples (denoted N_u)

$$\hat{\theta} = \frac{K_u}{N_u} = \frac{\sum_{i=1}^{\tau} 1_{\{m_i > u\}}}{\sum_{t=1}^T 1_{\{x_t > u\}}},$$

where m_i is the maximum over the subsample $\{x_{(i-1)N+1}, \dots, x_{iN}\}$. Since K_u can be viewed as the number of clusters of exceedances, this definition suggests that θ be interpreted as the reciprocal of the mean cluster size. An asymptotically equivalent estimator is given by

$$\tilde{\theta} = \frac{1}{N} \frac{\log(1 - K_u/\tau)}{\log(1 - N_u/(n\tau))}.$$

The asymptotic distribution of the estimator is provided by Hsing (1993). Some simulations reported in Embrechts, Klüppelberg, and Mikosch (1997) suggest that the estimator $\tilde{\theta}$ is less sensible to the size of the sample.

Extremal behavior of an ARCH process

An important result of Theorem 7.4 is that, for non-*iid* returns, the behavior of extremes may be driven by two components, the behavior of extreme innovations and the dependence structure. To make this statement clearer, assume that returns are indeed drawn from the GARCH model: $\varepsilon_t = \sigma_t z_t$, where z_t is an *iid* innovation and σ_t the conditional volatility. When z_t is drawn from a distribution that allows fat tails (say, a Student t distribution), the fat-tailedness of returns comes from the fat-tailedness of innovations, but also from volatility clustering. It is therefore of great importance to measure the extent to which the dependence structure is likely to affect the behavior of extremes. This measure can be obtained by the extremal index.

The first attempt to compute the extremal index of an ARCH process goes back to de Haan et al. (1989). Assume the standard ARCH(1) process $\varepsilon_t = (\omega + \alpha \varepsilon_{t-1}^2)^{1/2} z_t$ with $0 < \alpha < 1$ and z_t an *iid* $\mathcal{N}(0, 1)$ variable. Notice that it can be rewritten in a slightly different form as $\varepsilon_t^2 = (\alpha z_t^2) \varepsilon_{t-1}^2 + (\omega z_t^2)$, so that we have a stochastic difference equation of the form $Y_t = A_t Y_{t-1} + B_t$ for $Y_t = \varepsilon_t^2$. Then, under some regularity conditions, it can be shown that if there exists a $\kappa > 0$ such that $E[\alpha z_1^2]^\kappa = 1$, then the stationary process Y_t has a Pareto-type tail of the form $cy^{-\kappa}$ for large y . In particular, it implies that κ is nothing else than the inverse of the tail index of Y_t . Finally, the ARCH(1) process ε_t has a Pareto-type tail of the form $c\varepsilon^{-2\kappa}$. Therefore, the expression for the tail index of an ARCH(1) process can be computed implicitly as $\xi = 1/(2\kappa)$ where κ is the solution of $E[\alpha z_1^2]^{1/\kappa} = 1$ or, in the case where z_t is $\mathcal{N}(0, 1)$

$$\Gamma\left(\kappa + \frac{1}{2}\right) = \sqrt{\pi} (2\alpha)^{-\kappa}.$$

No analytical expression of the extremal index θ is available in general, but it can be easily estimated using Monte Carlo simulations of an ARCH(1) process and by computing the non-parametric estimators $\hat{\theta}$ or $\tilde{\theta}$ described above.

Some computations, given by de Haan et al. (1989), are reported in Table 7.3. For different values of the parameter α , it gives the corresponding values of the tail index ξ and the extremal index θ . As it appears clearly in the table, when there is a large amount of time dependence in the volatility process (α large), the extremal index is small, indicating that the actual limit distribution of the process is far from the one we would obtain under *iid*-ness. The tail index also shows that the tails of the distribution are very fat, approaching those of a stable distribution with infinite variance (that corresponds to $\xi = 0.5$).

Some of the results presented for ARCH(1) process have been extended for GARCH(1, 1) processes, by Stărică and Pictet (1997). Assume now that the GARCH(1, 1) process $\varepsilon_t = \sigma_t z_t$ and $\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2$ with $0 < \alpha, \beta < 1$ and z_t an *iid* $\mathcal{N}(0, 1)$ variable. Then ε_t has a Pareto-type tail of the form $c\varepsilon^{-2\kappa}$, where κ is given by the relation $E[\alpha z_1^2 + \beta]^\kappa = 1$ for $\kappa > 0$. The tail index is still obtained by the relation $\xi = 1/(2\kappa)$. Stărică and Pictet (1997) also describe how to obtain an estimate and asymptotic confidence intervals for the tail index from the preliminary estimation of the GARCH process.

Table 7.3. Value of the extremal index and the tail index for various value of α

α	θ	ξ
0.1	0.999	0.038
0.5	0.835	0.211
0.9	0.612	0.435
0.99	0.571	0.495

Dealing with strongly dependent data

Up to now, only very few practical recommendations for dealing with strongly dependent data have been provided. Hols and de Vries (1991) suggest a “thinning” procedure, which consists in removing each realization adjacent to an extreme event. This approach appears to be appropriate for series with a low dependence, but probably not for series with a large amount of dependence, such as financial returns.

Using the extrema approach

Another suggestion by Diebold, Schuermann, and Stroughair (1998) and McNeil (1998) involves using series of extremes over subsamples. As we have seen, Theorem (7.4) provides a strong incentive to use the extrema approach

to estimate the tail index. Indeed, this parameter is not affected by the dependence properties of the series. The reason is that this approach asymptotically eliminates the effect of dependence (or clustering) of extremes through a renormalization of the limit distribution.

Notice that, although the tail index is asymptotically unchanged, its finite-sample properties will probably be affected by the dependence in X . The reason is that the extremes over subsamples of length N from the strongly dependent series have the same behavior as the extremes over subsamples of length $N\theta$ from the *iid* series. This implies that the subsamples should be chosen long enough.

Table 7.4 reports some results of a Monte Carlo simulation experiment performed by Jondeau and Rockinger (2003c) on the estimation of the tail index using the extrema approach in the case of a dependent process. They simulate ARCH(1) processes of the form $\varepsilon_t = \sigma_t z_t$ and $\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2$ with z_t an *iid* $\mathcal{N}(0, 1)$ variable for various values of α . Then, they estimate the tail index of ε_t for several subsample sizes. More precisely, they use as parameter set $(\omega, \alpha) = (0.05, 0.70)$ and $(0.05, 0.90)$. If we compute the corresponding tail index, using the formula proposed by de Haan et al. (1989), we see that these parameters correspond to $\xi = 0.315$ and $\xi = 0.434$. We observe that dependence is responsible for a relatively small, and nonsignificant, bias that vanishes as subsamples become larger. For instance, for the case where $\xi = 0.434$, for monthly subsamples, we obtain an estimation of 0.372 that increases up to 0.429 for semi-annual subsamples.

This evidence suggests that using subsamples of quarterly or semi-annual daily returns is in general enough to provide correct estimates of the tail index in the case of ARCH processes.

Table 7.4. Monte Carlo experiment for computing the tail index. Estimates are based on 1,000 replications.

	$T = 2,000$		$T = 10,000$	
	$\hat{\xi}$	Std error	$\hat{\xi}$	Std error
ARCH(0.05,0.70)	True $\xi = 0.315$			
Monthly	0.241	0.095	0.232	0.040
Quarterly	0.307	0.174	0.287	0.069
Semi-annually	0.348	0.281	0.298	0.109
ARCH(0.05,0.90)	True $\xi = 0.434$			
Monthly	0.375	0.105	0.372	0.043
Quarterly	0.442	0.189	0.424	0.078
Semi-annually	0.494	0.317	0.429	0.113

Modeling the tail of the conditional distribution

Another approach, developed in particular by McNeil and Frey (2000), consists in estimating the tail of the conditional, rather than the unconditional,

distribution. For this purpose, one can adopt a two-step strategy. First, we fit a conditional mean and volatility model (such as an AR–GARCH model) to financial return series, in order to account for serial correlation and heteroskedasticity. This step is intended to provide standardized residuals that are approximately *iid* series. Then, we use the tail approach to model the tail distribution of the standardized data.

We assume now that the dynamics of a time series $\{x_1, \dots, x_T\}$ is given by a standard AR(1)–GARCH(1, 1) model of the form

$$x_t = \mu_t + \varepsilon_t, \quad (7.10)$$

$$\varepsilon_t = \sigma_t z_t, \quad (7.11)$$

$$\mu_t = \mu + \varphi x_{t-1}, \quad (7.12)$$

$$\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2, \quad (7.13)$$

with $\omega > 0$, $\alpha > 0$ and $\beta > 0$ to ensure a positive volatility process. Sufficient conditions for ε_t to be strictly stationary are $\alpha + \beta < 1$. We assume $|\varphi| < 1$. The innovation process, z_t , is assumed to be *iid* with zero mean, unit variance, and marginal distribution function $F_Z(z)$.

As suggested by McNeil and Frey (2000), we estimate this model using the quasi maximum-likelihood procedure. Therefore, we estimate the parameter vector $(\mu, \varphi, \omega, \alpha, \beta)'$ assuming normal innovations. It can be shown that the method provides consistent and asymptotically normal estimators (see Gouriéroux, Monfort, and Trognon, 1984). Then, we compute standardized residuals $\hat{z}_t = (x_t - \hat{\mu}_t) / \hat{\sigma}_t$, which should be *iid*, if the model is well adapted to the data. Now, all the tools described above for modeling the tails of the distribution can be applied to the standardized residuals.

This approach has been successfully used for the computation of VaR (see McNeil and Frey, 2000). Its use will be more thoroughly described in Chapter 8.

7.1.4 Estimation of high quantiles

At this stage, we have recalled various theoretical results of how extreme returns or excesses over a threshold may be used to characterize the behavior of the tails of a distribution. Once parameters have been obtained, it is possible to address several useful issues, such as the estimation of high quantiles. This question is central to VaR analysis.

For instance, we may be interested in the type of extreme realization we can expect over a horizon of 5 (or 50) years. If some daily data is sampled over 200 trading days, then, under the assumption that there are 200 trading days in a year, we are looking for the largest among 1,000 (or 10,000) realizations. The issue of computing VaR from financial data will be detailed in Chapter 8. In the present section, we provide some results on the estimation of high quantiles based on the various approaches described earlier in this chapter.

Notice that instead of using the minima or the left tail of the distribution of returns (as it is suggested in VaR computation), we consider maxima and the right tail to remain consistently with the previous analysis. A simple way to view the following is to consider the absolute value of minima or the absolute value of the left tail of the distribution.

Using the extrema approach

The extrema approach to compute high quantiles has been initially proposed by McNeil (1998) and Longin (2000). Let $\{m_i\}_{i=1}^{\tau}$ be a vector of extremes of the initial time series $\{x_t\}_{t=1}^T$ computed over subsamples of size N . Assume that $\{x_t\}_{t=1}^T$ has the gev $H_{\xi, \mu, \psi}$ as exact distribution and the maxima $\{m_i\}_{i=1}^{\tau}$ are *iid* or weakly dependent. Finally, assume that an estimate $(\hat{\xi}, \hat{\mu}, \hat{\psi})'$ of the unknown parameters is available.

Let p^* be the probability that the maximum over a given subsample m_{τ} is above the quantile \tilde{m}_{1-p^*} (\tilde{m}_{1-p^*} is a large positive number, so that p^* is a small probability, such as $p^* = 1\%$).⁵ Then, we have

$$\begin{aligned} p^* &= \Pr[m_i > \tilde{m}_{1-p^*}] = 1 - H_{\xi, \mu, \psi}(\tilde{m}_{1-p^*}) \\ &= 1 - \exp\left(-\left(1 + \xi \frac{m_{1-p^*} - \mu}{\psi}\right)^{-1/\xi}\right), \end{aligned}$$

with $\xi \neq 0$ and $\psi > 0$.

Consequently, the quantile for the distribution of the maxima is simply obtained by inverting this formula

$$\tilde{m}_{1-p^*} = H_{\xi, \mu, \psi}^{-1}(1 - p^*) = \mu + \frac{\psi}{\xi} \left((-\log(1 - p^*))^{-\xi} - 1 \right).$$

The corresponding estimated quantile denoted by \hat{m}_{1-p^*} is obtained by replacing μ , ξ and ψ by their estimates $\hat{\xi}$, $\hat{\mu}$, and $\hat{\psi}$. In the case $\xi = 0$, the estimated quantile is

$$\hat{m}_{1-p^*} = H_{0, \hat{\mu}, \hat{\psi}}^{-1}(1 - p^*) = \hat{\mu} - \hat{\psi} \log(-\log(1 - p^*)).$$

Notice that, since the joint distribution of $(\hat{\xi}, \hat{\mu}, \hat{\psi})'$ is available, if the estimation is performed by ML, we can deduce a confidence interval for the estimate \hat{m}_{1-p^*} , using the delta method.

The quantile \tilde{m}_{1-p^*} in fact corresponds to the distribution of maxima. Therefore, \tilde{m}_{1-p^*} indicates the value that can be exceeded by the maximum

⁵ Notice that we use a slightly different notation for the quantile. Hence the tilde on \tilde{m}_{1-p^*} to avoid confusion.

of subsamples with a probability p^* . For VaR computation, we are actually interested in the quantile of returns, not the maxima. It can be easily computed under the assumption of *iid* maxima. In this case, we have

$$\begin{aligned} 1 - p^* &= \Pr[m_i \leq \tilde{m}_{1-p^*}] = \Pr[x_{(i-1)N+1} \leq \tilde{m}_{1-p^*}, \dots, x_{iN} \leq \tilde{m}_{1-p^*}] \\ &= (\Pr[x_t \leq \tilde{m}_{1-p^*}])^N. \end{aligned}$$

So if we denote the probability $p = \Pr[x_t > \tilde{m}_{1-p^*}]$ that a given return is above the level \tilde{m}_{1-p^*} , we obtain the relation

$$1 - p^* = (\Pr[x_t \leq \tilde{m}_{1-p^*}])^N = (1 - p)^N.$$

It means that \tilde{m}_{1-p^*} is not the $(1 - p^*)$ quantile of the return distribution, but instead its $(1 - p^*)^{1/N}$ quantile, denoted $\tilde{x}_{(1-p^*)^{1/N}}$.⁶ Therefore, if we are actually interested in controlling the probability p that the return exceeds a given quantile,⁷ we redefine $\tilde{m}_{1-p^*} = \tilde{m}_{(1-p)^N}$ and get

$$\tilde{m}_{(1-p)^N} = H_{\xi, \mu, \psi}^{-1}((1 - p)^N) = \mu + \frac{\psi}{\xi} \left((-N \log(1 - p))^{-\xi} - 1 \right).$$

Interestingly, the extrema approach allows correcting the quantile estimation for processes with strong dependence. As we have seen in Section 7.1.3, in such case, the distribution of extremes of $\{x_t\}_{t=1}^T$ is $H_{\xi, \mu, \psi}^\theta(m)$ with θ denoting the extremal index. Then, to be consistent with a probability of $(1 - p)$ for *iid* or weakly dependent returns, the corresponding quantile of the extreme distribution will have to be computed with a probability level equal to $(1 - p)^{N\theta}$

$$\tilde{m}_{(1-p)^{N\theta}} = \mu + \frac{\psi}{\xi} \left((-N\theta \log(1 - p))^{-\xi} - 1 \right).$$

This result is consistent with the previous claim that the extremes over subsamples of length N from strongly dependent data have the same behavior as the extremes over subsamples of length $N\theta$ from the corresponding *iid* or weakly dependent series.

Using the tail approach

The tail approach has been adopted for computing high quantiles by McNeil (1997) and McNeil and Frey (2000). Now, we focus directly on the tails of the distribution and therefore consider the estimation of a high quantile x_{1-p} such that $F(x_{1-p}) = \Pr[X \leq x_{1-p}] = 1 - p$. To solve this issue, Rootzén and

⁶ For instance, assume we compute the 99%-quantile of the distribution of maxima. If maxima are measured over 60-day subsamples, this quantile in fact corresponds to the 99.98%-quantile of the return distribution.

⁷ For instance, the Basle Committee on Banking Supervision (1996) imposes a probability p equal to 0.01, assuming that returns are *iid* or weakly dependent.

Tajvidi (1996) or McNeil (1997) suggest the following approach. They first notice that for $x > u$ we have

$$\begin{aligned} F(x) &= \Pr[X \leq x] = \Pr[X \leq u] + (1 - \Pr[X \leq u]) \Pr[X > u] \\ &= \Pr[X \leq u] + (1 - \Pr[X \leq u]) F_u(x - u) \\ &= F(u) + (1 - F(u)) F_u(x - u). \end{aligned}$$

Hence, it is possible to estimate the probability of being below a certain threshold, $F(x)$, as

$$\hat{F}(x) = F_T(u) + (1 - F_T(u)) G_{\hat{\xi}, u, \hat{\psi}}(x),$$

where $F_T(u)$ is the empirical distribution function. A natural estimate of $F_T(u)$ is given by

$$\hat{F}_T(u) = \frac{1}{T} \sum_{t=1}^T \mathbf{1}_{\{x_t > u\}} = \frac{T - N_u}{T},$$

where N_u is the number of exceedances above the threshold u .

In addition, recall that the gpd is defined as

$$G_{\xi, u, \psi}(x) = \begin{cases} 1 - \left(1 + \frac{\xi}{\psi}(x - u)\right)^{-1/\xi} & \text{if } \xi \neq 0, \\ 1 - \exp\left(-\frac{(x-u)}{\psi}\right) & \text{if } \xi = 0. \end{cases}$$

Therefore, the quantile $x_{1-p} > u$ can be estimated using the following relation

$$\hat{F}(x_{1-p}) = \frac{T - N_u}{T} + \frac{N_u}{T} G_{\hat{\xi}, u, \hat{\psi}}(x_{1-p}) = 1 - p,$$

so that

$$\hat{x}_{1-p} = u + G_{\hat{\xi}, u, \hat{\psi}}^{-1}\left(1 - \frac{T}{N_u}p\right),$$

or⁸

$$\hat{x}_{1-p} = \begin{cases} u + \frac{\hat{\psi}}{\hat{\xi}} \left[\left(\frac{T}{N_u}p\right)^{-\hat{\xi}} - 1 \right] & \text{if } \xi \neq 0, \\ u - \hat{\psi} \log\left(\frac{T}{N_u}p\right) & \text{if } \xi = 0. \end{cases}$$

Notice that the value of the estimated quantile \hat{x}_{1-p} depends on the choice of the threshold u .

When we use the approach advocated by McNeil and Frey (2000), based on the estimation of the gpd for standardized residuals \hat{z}_t of a GARCH model, the procedure described above can be directly applied on \hat{z}_t instead of x_t .

⁸ Notice that this expression can be computed, because by construction $p < N_u/T$.

Using semi-parametric approaches

Details concerning the estimation of high quantiles in a semi-parametric framework are given in Dekkers and de Haan (1989) and Danfelsson and de Vries (1997).

Dekkers and de Haan (1989) describe how the Pickands estimator $\hat{\xi}_{(q,T)}^P$ can be used to derive the expression of the high quantiles. They show that the high quantile can be estimated as

$$\hat{x}_{1-p,q} = \frac{\left(\frac{q}{(1-p)T}\right)^{\hat{\xi}_{(q,T)}^P} - 1}{1 - 2^{-\hat{\xi}_{(q,T)}^P}} (x_{T-q+1,T} - x_{T-2q+1,T}) + x_{T-q+1,T}.$$

In addition, they show the strong consistency of $\hat{x}_{1-p,q}$ and give its asymptotic distribution that turns out to be quite complicated.

Danfelsson and de Vries (1997) suggest the following estimator of the high quantile based on the Hill estimator $\hat{\xi}_{(q,T)}^H$

$$\hat{x}_{1-p,q} = \left(\frac{q}{(1-p)T}\right)^{\hat{\xi}_{(q,T)}^H} x_{T-q+1,T}.$$

Under some restrictive conditions, this estimator is asymptotically normal.

7.2 Multivariate dependence

Much effort has been made to extend univariate extreme value theory to a multivariate context. Such an extension, however, is far from being straightforward. The reason is that there exists no multivariate equivalent to the Fisher-Tippet theorem. Stated differently, there does not appear to exist a limit distribution for scaled maxima. From the point of view of exceedances over high thresholds, there is no multivariate distribution such as the gpd that would naturally fit the tails.

An additional difficulty comes from the notion of dependency. Typically, we may measure dependency between series with the well-established correlation measure, developed by Pearson as

$$\rho[X, Y] = \frac{\sum(x_t - \bar{x})(y_t - \bar{y})}{\sqrt{\sum(x_t - \bar{x})^2 \sum(y_t - \bar{y})^2}},$$

the variables \bar{x} and \bar{y} denoting sample averages. It turns out that this measure of dependency only applies to observations that are not too far out in the tails. As already seen in Chapter 6, this has led researchers to introduce other measures of dependency. Early research was also based on the assumption that the joint tail distribution would be given by some parametric functional

dependency, often captured by a copula function. A set of copula functions has already been discussed in Chapter 6 as a means to join marginal distributions. Here, the copula is playing the role of a parametric representation of some dependence. The use of copula to describe a dependency in the tails may lead to erroneous conclusions unless the adequacy of a given copula has been ascertained.

To gain the intuition of why this may happen, consider fitting a stable distribution to a sample of data generated by some fat-tailed distribution, where the fat-tailed distribution is assumed to have a second and maybe even a third moment. By fitting a stable distribution, we might obtain some parameters, yet without a thorough test of adequacy, we may be led to believe that the data does not allow for a second moment, just because stable distributions do not allow for a finite second moment.

In the statistics literature, Ledford and Tawn (1996, 1997) introduce the notion of asymptotic dependence and asymptotic independence. Most existing multivariate tail estimation models, especially those using ad hoc copula, are inadequate to deal with cases of asymptotic independent data. Coles, Heffernan, and Tawn (1999), therefore, propose a test to check the type of dependency that exists in the data. Poon, Rockinger, and Tawn (2004) apply this methodology to financial data. Only once a certain type of dependency has been identified, it is reasonable to model the joint behavior with a certain copula.

To give a flavor of the ideas behind these concepts of dependency, Figure 7.14 represents a scatterplot of returns involving the CAC and the DAX, and Figure 7.15 traces a scatterplot for CAC and SP500 returns, over the period from 1970 to 2001. The two figures also represent quadrants “along the diagonal” to indicate how the tails become thinner and thinner as we move into the tails. As Figure 7.14 reveals, there are quite a few observations as we move out into the left tail of the joint distribution. As we know, index returns are correlated, here with a coefficient of 0.46. We may, therefore, conclude that the pair CAC-DAX is not only dependent, but also asymptotically dependent. Figure 7.15 indicates that, as we move out into the tails, there are fewer and fewer observations, so that the pair CAC-SP500 may be viewed as asymptotically independent. However, if we compute the correlation, we find a coefficient of 0.23 and, thus, that the data is dependent.⁹ The intuition conveyed by these pictures is that it will be necessary to introduce two measures of dependency: one for the tails and one for the more central part of the distribution. If the test for the more central part concludes that the data may also be asymptotically dependent, then it will be useful to construct a measure of dependency specifically for the tails.

⁹ Since the U.S. market opens during the afternoon of the European markets, we compute correlations using a lag of the European market return. A more precise adjustment to this problem could be made along the work of Martens and Poon (2001).

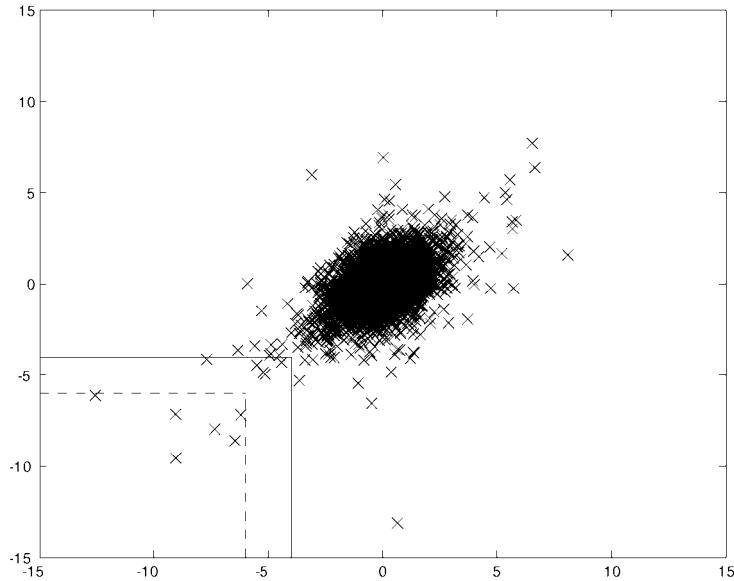


Fig. 7.14. Scatterplot of standardized daily returns of the CAC and the DAX.

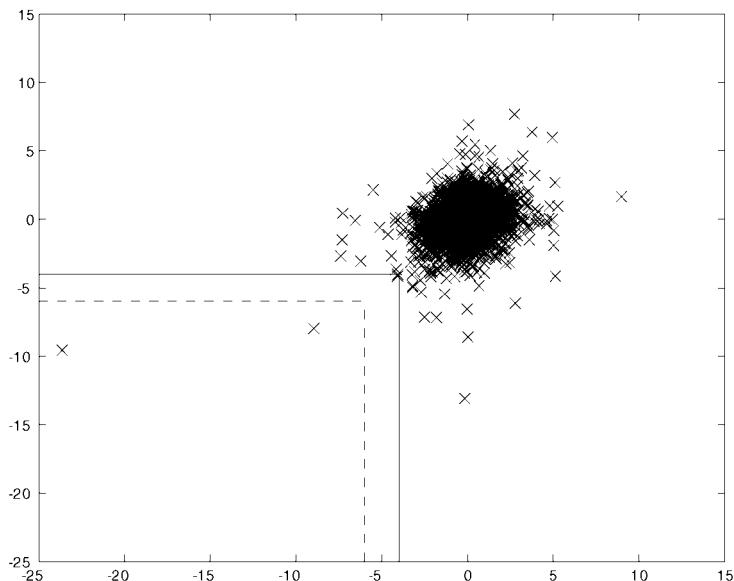


Fig. 7.15. Scatterplot of standardized daily returns of the CAC and the SP500.

7.2.1 Characterizing tail dependency

In almost all multivariate studies, it is helpful to remove the influence of marginal aspects first by transforming the original variables into a common marginal distribution. Indeed, the behavior of the marginal distributions has no effect on the asymptotic dependency.

As such, we may have fat-tailed distributions that turn out to be asymptotically independent and we may have distributions with, say Gaussian tails, that may be asymptotically dependent. After such a transformation, differences in distributions are purely due to dependence aspects. One natural way to remove those aspects that are specific to the marginal distributions of the bivariate returns, written here as (X, Y) , is to work with the distribution functions $U \equiv F_X(X)$ and $V \equiv F_Y(Y)$ that are the respective marginal *cdfs* for X and Y . This means that rather than using the returns, we would use their empirical *cdfs*.

Asymptotic dependence

As the variables U and V are on a common scale, events of the form $\{U > u\}$ and $\{V > u\}$, for large values of u , i.e., values close to one, correspond to equally extreme events for each variable. For large values of u , we may show that $\Pr[U \leq u] \sim u$, a result that will be frequently used below.

In the same way, as it is natural in probability theory to measure dependency via a conditional probability, here it is natural to consider conditional probabilities of one variable given that the other is extreme. Specifically, we consider the behavior of $\Pr[V > u | U > u]$ for high levels of u . The tails will be *perfectly dependent* if $\Pr[V > u | U > u] = 1$. Such a situation could arise in finance for a derivative instrument and its underlying asset if perfect hedging existed. In contrast, if (U, V) are *exactly independent* then $\Pr[V > u | U > u] = \Pr[V > u]$, a situation that should be relatively rare in finance. We observe that these definitions are not only given as limits of $u \rightarrow 1$, but also hold for smaller levels. We may consider the limit case

$$\chi = \lim_{u \rightarrow 1} \Pr[V > u | U > u], \quad (7.14)$$

where $0 \leq \chi \leq 1$. Variables for which $\chi > 0$ are called *asymptotically dependent* and if $\chi = 0$ *asymptotically independent*. Clearly, χ measures the degree of dependence that is persistent in the limit. We define, in terms of uniform margins,

$$\begin{aligned}
\chi(u) &= \Pr[V > u | U > u] = \frac{\Pr[V > u, U > u]}{\Pr[U > u]} \\
&= \frac{1 - \Pr[V \leq u] - \Pr[U \leq u] + \Pr[V \leq u, U \leq u]}{1 - \Pr[U \leq u]} \\
&\approx \frac{1 - 2u + \Pr[V \leq u, U \leq u]}{1 - u} \\
&\approx 2 - \frac{\log(\Pr[V \leq u, U \leq u])}{\log(u)}.
\end{aligned}$$

In the third line, the \approx means that the expression that follows is an approximation, true for u close to 1. The last line follows from a Taylor approximation. For empirical purposes, it may be interesting to investigate the discriminative power of this function. To do so, we assume that the tails of the distribution are given by a Gaussian copula as defined in Chapter 6. It has been shown that, for the Gaussian copula, the upper and lower tails are asymptotically independent. Thus, for this copula, we would expect that $\chi = 0$. We now show that $\chi(u)$ has a low power to detect a situation of asymptotic independence. Indeed, Figure 7.16, inspired by the work of Coles, Heffernan, and Tawn (1999), displays the measure $\chi(u) = \Pr[U > u | V > u]$, for various values of correlation of the Gaussian copula. Considering positive correlations among asset returns as it is the case in practice, we notice that only for very high levels of u , the measure converges towards $\chi = 0$. For instance, if we take $\rho = 0.6$ and $u = 0.95$, a situation implying that we consider that 5% of the sample corresponds to tail observations, we find $\chi(u) = 0.4$. Such a value is sufficiently far away from 0 as to become statistically significant even though it should not be.

The conclusion of this exercise is, therefore, that this measure, will have difficulties to discriminate a situation of asymptotically dependent and independent tails. A corollary of this finding is that if we use a logistic model

$$\Pr[U \leq u, V \leq v] = \exp\left(-\{(-\log u)^{1/\alpha} + (-\log v)^{1/\alpha}\}^\alpha\right), \quad (7.15)$$

then it is difficult to test if the data is asymptotically independent or not. Indeed, for this dependency structure, it has been shown that $\chi = 2 - 2^\alpha$. Thus, we may be tempted to test for asymptotic dependency by using the test $\alpha = 1$, corresponding to $\chi = 0$, against $\alpha < 1$, corresponding to $\chi > 0$. Given this lack of power, we may conclude that there exists asymptotic dependency, when there is none and to overestimate $\Pr[U > u, V > v]$.

So far, we only focused on random variables exceeding a given threshold. For this reason, the measures developed so far are also termed measures of positive dependency. Clearly, we may focus on the lower tail of the distribution. For centered data, this could be done by multiplying the numbers by -1 . In such a case, we would obtain measures for negative dependency.

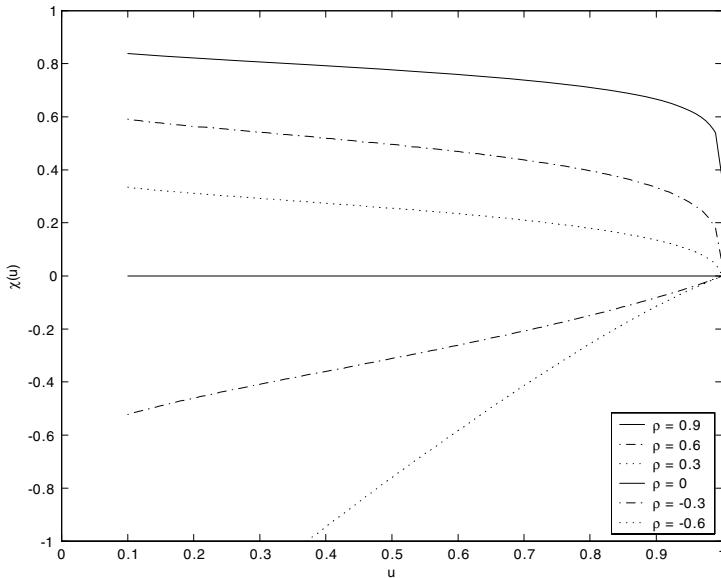


Fig. 7.16. Behavior of $\chi(u)$ for a Gaussian copula as the correlation ρ varies from -0.6 to 0.9 .

An alternative measure of dependence

Facing this difficulty, Coles, Heffernan, and Tawn (1999) created a more powerful statistics, based on the dependency measure $\Pr[V > u | U > u]$. The idea is to transform the various elements appearing in the previous computations, in such a way as to yield a measure that has properties similar to a correlation, i.e., that belongs to $[-1, 1]$ and that is positive in the case of a positive association. Defining

$$\bar{\chi}(u) \equiv \frac{2 \log(\Pr[U > u])}{\log(\Pr[U > u, V > v])} - 1,$$

and

$$\bar{\chi} = \lim_{u \rightarrow 1} \bar{\chi}(u),$$

does the trick, see Coles, Heffernan, and Tawn (1999). Indeed, for asymptotically independent data, $\Pr[U > u, V > u]$ converges to zero rather quickly. Also, $\Pr[U > u]$ will thin out, yet at a smaller rate, given that the margins are similarly scaled, we expect a thinning half as quickly. This implies that for such a situation $\bar{\chi}$ will take the value of 0. Values of $\bar{\chi} > 0$, $\bar{\chi} = 0$ and $\bar{\chi} < 0$ loosely correspond respectively to when (U, V) are positively associated in the extremes, exactly independent, and negatively associated. For the bivariate Gaussian dependence structure, $\bar{\chi}$ is equal to the correlation coefficient,

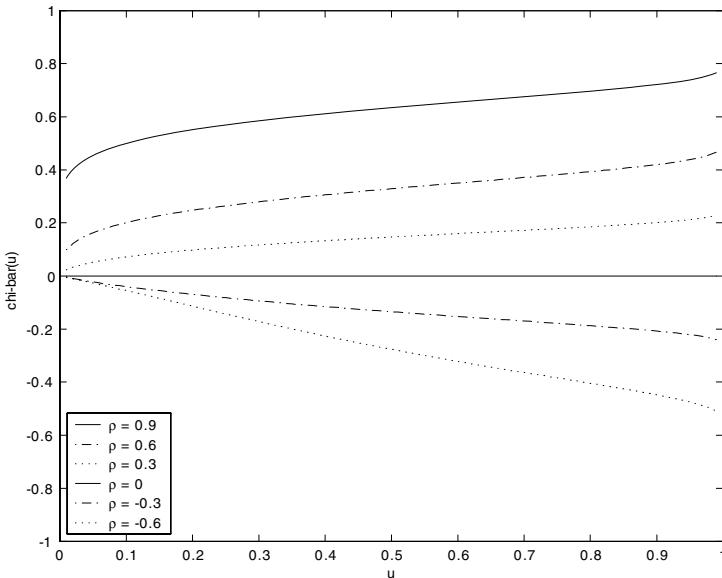


Fig. 7.17. Behavior of $\bar{\chi}(u)$ for a Gaussian copula with correlation ρ ranging from -0.6 to 0.9 .

which aids in interpreting the value of $\bar{\chi}$. For other examples see Heffernan (2000). Figure 7.17 traces the function $\bar{\chi}(u)$ for the Gaussian copula. We notice that as u increases, the pattern of $\bar{\chi}$ disperses. Nonetheless, even for high values of u , such as 0.99 and data generated by highly correlated data, $\rho = 0.9$, the measure $\bar{\chi}$ takes a value of less than 0.8 and is therefore relatively distant from 1, level where one might conclude for asymptotically dependent data. It is only for extremely high levels of u that $\bar{\chi}$ comes close to its theoretical value of 0.9. This illustrates that the new measure is likely to provide a better test for asymptotic dependency than χ .

We may conclude by observing that the pair of dependence measures $(\chi, \bar{\chi})$ together provide all the necessary information to characterize the form and degree of extremal dependence. For asymptotically dependent variables, we have $\bar{\chi} = 1$ with the degree of dependence given by $\chi > 0$. For asymptotically independent variables, we have $\chi = 0$ with the degree of dependence given by $\bar{\chi}$. One key insight of this section is that it is important to test first if $\bar{\chi} = 1$ before drawing conclusions about asymptotic dependence based on estimates of χ . There remains the question of how to estimate these dependency measures. We turn to this issue in the next section.

7.2.2 Estimation and statistical inference on $\bar{\chi}$ and χ

For empirical purposes, it is convenient to transform the bivariate returns (X, Y) via the Fréchet transformation

$$S = -\frac{1}{\log(F_X(X))} \quad \text{and} \quad T = -\frac{1}{\log(F_Y(Y))}, \quad (7.16)$$

where F_X and F_Y are the respective marginal distribution functions for X and Y . In practice, the values of F_X and F_Y that are used in the transformation (7.16) are obtained using the empirical distribution functions of the separate variables. In a finance setting, the Fréchet transformation appears most useful, since return distributions tend to be fat tailed as demonstrated earlier.¹⁰

To estimate $\bar{\chi}$ and χ , one may use the results of Ledford and Tawn (1996, 1997, 1998), where it was established that under weak conditions the probability of joint extreme events is given by

$$\Pr[S > s, T > s] = \mathcal{L}(s)s^{-1/\eta}, \quad \text{as } s \rightarrow \infty, \quad (7.17)$$

where $0 < \eta \leq 1$ is a constant and $\mathcal{L}(s)$ is a slowly varying function.¹¹ For our purposes, it suffices to view a slowly varying function as a constant as long as s is large enough. The result therefore states that the probability of joint extreme events decreases exponentially up to some scaling factor. From this representation, by taking logs, it follows that $\log(\Pr[S > s, T > s]) = \log(\mathcal{L}(s)) - \frac{1}{\eta} \log(s)$. It follows that

$$\bar{\chi} = \frac{-2 \log(s)}{\log(\mathcal{L}(s)) - \frac{1}{\eta} \log(s)} - 1 \xrightarrow{s \rightarrow \infty} 2\eta - 1.$$

This result immediately shows that the tail index for joint extreme realizations will allow us to construct an estimate of $\bar{\chi}$. It suffices to take $Z = \min(S, T)$, and to realize that $\Pr[Z > s] = \Pr[\min(S, T) > s]$. Clearly, if a sample $\{z_1, \dots, z_T\}$ is available, this index may be estimated using the Hill estimator from (7.5). This yields

$$\hat{\bar{\chi}} = \frac{2}{n_s} \sum_{j=1}^{n_s} \log\left(\frac{z_{j,T}}{s}\right) - 1,$$

$$V[\hat{\bar{\chi}}] = (\hat{\bar{\chi}} + 1)^2/n_s,$$

¹⁰ For this type of transform, we may show that $\Pr[S > s] = \Pr[T > s]$ is proportional to s^{-1} for large s . By setting $U = F_X(X)$ and $V = F_Y(Y)$, we recover the notations used in the previous section. The probability $\Pr[S > s]$, for large s , translates into a probability $\Pr[U > u]$ as $u \rightarrow 1$.

¹¹ For a precise definition of what a slowly varying function is, we may consult Embrechts, Klüppelberg, and Mikosch (1997).

where n_s is the amount of observations exceeding s , a high threshold. If $\hat{\chi}$ is significantly less than 1, e.g., if $\hat{\chi} + 1.96\sqrt{V[\hat{\chi}]} < 1$, then we infer the variables to be asymptotically independent and take $\chi = 0$. Only if there is no significant evidence to reject $\bar{\chi} = 1$ do we estimate χ , which we do under the assumption that $\bar{\chi} = \eta = 1$. We then have the definition that, as $s \rightarrow \infty$,

$$\chi = \frac{\Pr[T > s, S > s]}{\Pr[S > s]} = \frac{\mathcal{L}(s)s^{-1}}{s^{-1}} = \mathcal{L}(s).$$

We need empirical counterparts to $\Pr[T > s, S > s]$ and $\Pr[S > s]$ to estimate these measures. An estimate of the first probability is given by the frequency of the joint observation of extreme realizations, i.e., n_s/n . A measure for the second probability is the frequency of those observations that are in the tail. This is by construction given by s^{-1} . The variance may be obtained under the assumption that joint events are Bernoulli trials. This leads us to

$$\hat{\chi} = \frac{sn_s}{n},$$

$$V[\hat{\chi}] = \frac{s^2 n_s(n - n_s)}{n^3}.$$

Furthermore, we can assess whether the variables have a joint tail which decays with the same form as for exact independence by testing if $\hat{\chi}$ is significantly different from 0.

7.2.3 Modeling dependency

In the previous section, we have shown that it is important to test whether data is asymptotically dependent or not. Only if the data is asymptotically dependent does it make sense to model the dependency by a copula that imposes asymptotic tail-dependency. If the data is not found to be asymptotically dependent, we could instead use the Gaussian copula. We remind that for the Gaussian copula, we have

$$\Pr[S \leq s, T \leq t] = \Phi_\rho(\Phi^{-1}(\exp(-1/s)), \Phi^{-1}(\exp(-1/t))),$$

where Φ_ρ is the bivariate normal distribution with mean vector $(0, 0)'$ and covariance matrix containing 1 along the diagonal, and ρ on the off diagonal. As already mentioned in the previous section, for this model $(\chi, \bar{\chi}) = (0, \rho)$. We could, therefore, in practical applications estimate ρ by using $\bar{\chi}$.

For asymptotically dependent data, if explicit modeling dependency of the margins is required, we may use the catalogue of copula, already discussed in Chapter 6. A popular copula is the logistic one

$$\Pr[S \leq s, T \leq t] = \exp\left(-\left(s^{-1/\alpha} + t^{-1/\alpha}\right)^\alpha\right),$$

Table 7.5. Lower and upper dependence measures for various copula. Notice that the parameter φ is a function, whereas all other parameters are scalars.

Name	Parameter	Lower tail dependence	Upper tail dependence
Gaussian	ρ	None	None
Student t	ρ, n	$2\bar{t}_{n+1}(\lambda)$	$2\bar{t}_{n+1}(-\lambda)$
Archimedean	φ	$2 \lim_{s \rightarrow \infty} \frac{\varphi'(2s)}{\varphi^{-1}(2s)}$	$2 - 2 \lim_{s \rightarrow \infty} \frac{\varphi'(2s)}{\varphi^{-1}(2s)}$
Frank	θ	None	None
Clayton	θ	$2^{-1/\theta}$	None
Gumbel	θ	None	$2 - 2^{1/\theta}$
Plackett	θ	θ	θ
Marshall-Olkin	$\alpha > \beta$	None	$\min(\alpha, \beta)$

with $0 < \alpha \leq 1$, and where decreasing α increases dependence. For this model $(\chi, \bar{\chi}) = (2 - 2^\alpha, 1)$. Since, for this copula $\bar{\chi} = 1$, the case of asymptotic independence is automatically precluded.

Other copula may be chosen to describe a given dependence structure. In Table 7.5, we give the values of the upper and lower tail dependence for the various copula introduced in Chapter 6, see also Heffernan (2000). For the Student t copula, we define $t_{n+1} = \Pr[Y > y | X = 0]$ and $\lambda = \sqrt{(1+n)(1-\rho^2)}$. (2000).

7.2.4 An illustration

Poon, Rockinger, and Tawn (2004) consider the SP500 (US), the FT-SE 100 (UK), the DAX 30 (GER), the CAC 40 (FRA), and the Nikkei 225 (JAP) indices, covering the period from 26 December 1968 to 12 November 2001. Splitting the data into three subsamples, they obtain, for the unfiltered left tail of the distribution and bivariate-GARCH (using the ADC model of Kroener and Ng, 1998, see Section 6.1.3) filtered data, the following parameters represented in Table 7.6.

Inspection of these estimates shows that Pearson's correlation coefficient is in general much smaller than $\bar{\chi}$. This observation reinforces our conjecture that Pearson's correlation does not measure the tail behavior well. Also interestingly, most of the $\hat{\chi}$ are statistically smaller than one, implying that there is little evidence for asymptotic dependence among these series and this in particular for the earlier part of the sample. In the later part of the sample, the dependency measure increases. We interpret this finding as a sign that market integration increased over time. In this table, we only focus on the left tail. For the right tail, there exists even less evidence for asymptotic dependency.

Comparison between the original unfiltered data and GARCH-filtered data shows that filtering removes some dependency but not all. Ideally, we would like to know, from a theoretical point of view, what the implications of multivariate GARCH model on the tails are. Thus, useful research would indicate

Table 7.6. Measure $\bar{\chi}$ for the left tail of pairs of returns. Unfiltered data corresponds to raw returns, whereas ADC filtered data involves the residuals of the ADC-GARCH. The parameter ρ stands for Pearson's correlation. Numbers with a * are not significantly different from 1.

	Unfiltered data		ADC-filtered data	
	ρ	$\hat{\chi}$	ρ	$\hat{\chi}$
Subperiod 1: 12/27/1968 to 12/12/1979 (2,859 obs.)				
US-UK	0.218	0.520	0.215	0.416
US-JAP	0.117	0.350	0.115	-0.201
UK-GER	0.105	0.342	0.086	0.182
GER-FRA	0.166	0.434	0.172	0.334
Subperiod 2: 12/13/1979 to 11/27/1990 (2,859 obs.)				
US-UK	0.345	0.794*	0.303	0.440
US-JAP	0.373	0.648	0.292	0.318
UK-GER	0.358	0.641	0.281	0.404
GER-FRA	0.409	1.381*	0.289	0.547
Subperiod 3: 11/28/1990 to 11/12/2001 (2,859 obs.)				
US-UK	0.298	0.737*	0.291	0.534
US-JAP	0.267	0.549	0.265	0.440
UK-GER	0.625	1.043*	0.550	0.779*
GER-FRA	0.695	0.949*	0.648	0.748*

how the measures of asymptotic dependency and independency change if the data is generated by some multivariate GARCH model.

Turning to the measure of asymptotic dependency, Poon, Rockinger, and Tawn (2003) show that for the third subperiod SP500–FT-SE, $\chi = 0.290$, with s.e. of 0.025. Also for the third subperiod, for the pairs FT-SE–DAX $\chi = 0.459$ (s.e. 0.030), and DAX–CAC $\chi = 0.517$ (s.e. 0.037). Since χ estimates the probability of joint occurrence of the most extreme values, these figures suggest that if the CAC has a large crash, in about 50% of the cases, the DAX will also crash. It also means that if the SP500 has a crash, in about 30% of the cases, the FT-SE will also crash.

Once the dependence characteristics have been established, it becomes possible to allocate assets based on their dependency structure. For instance, we have seen that the SP500–FT-SE are asymptotically dependent, whereas SP500–DAX is not. This suggests that extreme portfolio risk can be reduced by using the SP500–DAX pair rather than the SP500–FT-SE pair. For portfolio allocation purposes, it could, therefore, be interesting to select assets for which χ is 0. Then choose those assets where the $\bar{\chi}$ is smallest.

The results presented so far are given in a bivariate setting rather than for a truly multivariate one, which admittedly is a drawback. Nonetheless, the computation of the χ and $\bar{\chi}$ measures is very fast. We should view their role,

at this stage, as a screening tool, useful in the quest for portfolios with good extremal behavior.

Another exercise, proposed by Poon, Rockinger, and Tawn (2004) is the measurement of a mis-specification of the tail dependency. They consider the left tail of returns, by multiplying stock returns by -1 , so that stock market crashes show up as large positive numbers. For a portfolio with weight α , where $0 \leq \alpha \leq 1$,

$$P_\alpha = \alpha X + (1 - \alpha)Y,$$

we will consider the expected shortfall, defined as $ES_\theta = E [P_\alpha | P_\alpha > c_\theta]$ where c_θ is some large number determined by the probability $\Pr [P_\alpha | P_\alpha > c_\theta] = \theta$, and where θ is the probability allocated to tail events. As will be shown in the next chapter, dealing with risk management, the expected shortfall is a measure of risk that has good properties from an economic point of view. Given that the portfolio returns are beyond a certain threshold that will be reached on $\theta\%$ of the time, this measure captures the average losses. In the following, the calibration is done on daily data. The expected shortfall then represents the average loss, given that a crash of a given category took place.

The direct computation of the expected shortfall is not possible. For this reason, simulation-based methods can be used. A description of how this simulation can be done, we refer to Poon, Rockinger, and Tawn (2004). Figure 7.18 represents the expected shortfall for $\theta = 0.01$ in the upper panels and for $\theta = 0.001$ in the lower ones. The left figures correspond to a portfolio with a fraction α of wealth allocated to the SP500 and $(1 - \alpha)$ allocated to the Nikkei. The right figures correspond to a portfolio involving the DAX and the CAC. We measure dependency using a logistic model, a Gaussian model calibrated with $\bar{\chi}$, and a Gaussian model calibrated via Pearson's correlation, with dependence parameter $\bar{\chi}$ set to Pearson's correlation ρ .

Given the asymptotic independency of the SP500–Nikkei, the logistic model will overestimate the dependency of these series. Based on an inspection of the lower left figure, we would conclude according to this model that the expected shortfall is around 7% at the $\alpha = 0.001$ level, whereas for an equally-weighted portfolio it actually is 4.5. Also, if we had erroneously calibrated the model with Pearson's correlation and a Gaussian structure, we would have concluded for an expected shortfall of 2.5%. This time, the tail risks would be underestimated. In the right part of the figure, we represent similar computations, but this time for the DAX–CAC pair for which we have established asymptotic tail dependence. We notice that for such data, the bias is much smaller.

7.2.5 Further investigations

The relevance of the distinction between asymptotic dependency and independency remains an open issue and further studies may be useful. One such study could focus on hedging strategies. When no perfect hedge is available,

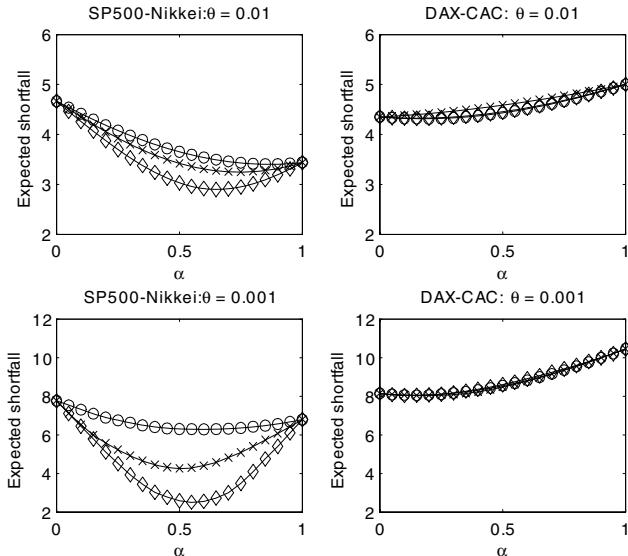


Fig. 7.18. Representation of the expected shortfall. The parameter α indicates the weight allocated to the one or the other country. Curves with x correspond to the bivariate Gaussian copula, calibrated with $\bar{\chi}$ as dependency measure. Curves with o correspond to the bivariate logistic. Last, the \diamond correspond to a bivariate Gaussian with Pearson's ρ as dependency measure.

a typical textbook way to hedge is to consider the portfolio of assets most correlated with the asset to hedge. Typically, the portfolio weights would be estimated by some OLS procedure. It remains an open issue if hedging strategies could be improved by also considering the tail dependency. Indeed, for extreme price variations, as they arise in practice, hedging may no longer perform as expected. To deal with such situations, a hedging strategy that maximizes the tail dependency of an asset and its underlying may be most relevant.

The methodology emphasized so far can also be used to address investors' asset allocation. For instance, Hartmann, Straetmans, and de Vries (2004) focus on stock and bond markets. They show that as stock markets crash, bond markets boom. This suggests that there exists a "flight to quality" phenomenon.

Risk Management and VaR

Until quite recently, the variance was a widely accepted measure of risk. It is very easy to understand and to compute. A shortcoming of this measure is that it is a symmetric one, in the sense that large gains and losses are equally penalized. Financial institutions however are much more concerned by large losses than by large gains.

Financial and regulatory institutions realized that there are many different sources of risk. These sources of risk have been progressively more precisely described and understood. For instance, three major types of risk are nowadays acknowledged; market risk, credit risk, and operational risk. Step by step, the Basel Committee on Banking Supervision (BCBS), at the Bank of International Settlements, imposes new capital requirements to financial institutions to cover these different sources of risk. The first step was the implementation of new standards for computing the exposition to risk and for measuring margin requirements. While the initial *1988 Basel Accord* only covered credit risk, in the *1996 Amendment to the Capital Accord to Incorporate Market Risks*, the BCBS incorporates market risk and explicitly introduces the Value-at-Risk as the main quantitative tool for financial institutions to calculate their capital requirement. VaR then became one of the widely used measures of market risk in the risk-management and fund-management industries.

Let θ denote a small percentage. Then the VaR of a portfolio can be defined as the minimum potential loss that the portfolio can suffer in the $\theta\%$ worst cases, over a given time horizon. Therefore, from a statistical point of view, the VaR is a quantile on the lower tail of the distribution of portfolio returns. Since its naming by the BCBS, the VaR is an actively researched topic, because it raises several interesting theoretical issues and has important implications for financial institutions. Early work on VaR has been done by J.P. Morgan (1996), Jorion (1997), Duffie and Pan (1997). J.P. Morgan, through its RiskMetrics methodology, has played an important role in the increasing popularity of VaR as a risk measure. Then, a huge literature emerged on the practical estimation of the VaR. Several approaches have been proposed to provide a precise evaluation of this measure. These approaches have been

based, for instance, on the use of univariate or multivariate GARCH models, on the modeling of the tails of the distribution, and on the modeling of non-normality. Interestingly, most of the elements needed for such computation have been analyzed in the previous chapters.

More recently, the VaR has been shown to be a potentially misleading measure of risk. A first criticism is that VaR is not a coherent measure because it does not satisfy the sub-additivity property, so that diversification does not necessarily result in a reduction of risk, as measured by VaR (see below for more formal definitions). Perhaps more importantly, it has been argued (BIS Committee on the Global Financial System, 2000) that VaR (as well as variance) misregards the risk of extreme loss. As has been highlighted by Basak and Shapiro (2001), because VaR disregards risk of extreme losses behind the confidence level, it may induce large losses. Consequently, this measure of risk may induce a larger risk exposure than the variance in case of falling markets. To cope with these shortcomings, some authors (Artzner et al., 1999, Basak and Shapiro, 2001) have proposed the use of the so-called Expected Shortfall (ES) as an alternative measure of risk. ES is the expected value of the loss of the portfolio in the $\theta\%$ worst cases over a given time horizon.

This chapter is organized as follows. Section 8.1 defines the two main notions we are going to analyze in this chapter, namely the VaR and the ES of a portfolio. The following sections are devoted to the practical evaluation of VaR as well as ES. There are actually four broad categories of VaR models: historical simulation (Section 8.2); semi-parametric models (Section 8.3), parametric models (Section 8.4); and finally non-linear techniques, which are designed to compute the VaR in presence of derivatives (Section 8.5). In the last section, we provide some tools to compare the performances of the various techniques developed in the previous sections.

8.1 Definitions and measures

8.1.1 Definitions

Let us now formally define the notions of VaR and ES. The VaR at probability $\theta \in (0, 1)$ of a portfolio is defined as the minimum potential loss that the portfolio may suffer in the $\theta\%$ worst cases, over a given time horizon. We define $P_{i,t}$ the price at date t of asset i , so that the return of asset i between date $t - 1$ and date t is $r_{i,t} = (P_{i,t} - P_{i,t-1}) / P_{i,t-1}$. Then the value of the portfolio at date t , for a vector N_t that contains the number of shares in asset i , is simply given by $W_t = \sum_{i=1}^n N_{i,t} P_{i,t} = N'_t P_t$. If we assume that portfolio composition is held constant from t to $t + 1$, the change in the market value of the portfolio is given by $W_{t+1} - W_t = N'_t (P_{t+1} - P_t)$. We deduce that $(W_{t+1} - W_t) / W_t = \alpha'_t r_{t+1} = r_{p,t+1}$, where $\alpha_{i,t} = N_{i,t} P_{i,t} / (\sum_{i=1}^n N_{i,t} P_{i,t})$. We denote the wealth of the investor for a portfolio weight vector α_t as $W_t(\alpha_t)$.

Value at Risk

If we denote $\Delta W_{t+1}(\alpha_t) = W_{t+1}(\alpha_t) - W_t(\alpha_t)$ the VaR of a portfolio is defined by the relation

$$\theta = \Pr [\Delta W_{t+1}(\alpha_t) \leq -\overline{VaR}_{\theta,t} | \mathcal{F}_t],$$

where \mathcal{F}_t denotes the information set at date t . Alternatively, we have¹

$$\theta = \Pr \left[\frac{\Delta W_{t+1}(\alpha_t)}{W_t(\alpha_t)} \leq -\frac{\overline{VaR}_{\theta,t}}{W_t(\alpha_t)} | \mathcal{F}_t \right] = \Pr [r_{p,t+1} \leq -\overline{VaR}_{\theta,t} | \mathcal{F}_t],$$

where $\overline{VaR}_{\theta,t}(r_{p,t+1}) = \overline{VaR}_{\theta,t}/W_t(\alpha_t)$ denotes the VaR for probability θ for \$1 invested. In the following, it will be our definition of VaR. If we also define the conditional cdf $F_{p,t}(x) = \Pr [r_{p,t} \leq x | \mathcal{F}_{t-1}]$, with $F_{p,t}^{-1}$ the inverse of the conditional cdf , we observe that

$$VaR_{\theta,t} = -F_{p,t}^{-1}(\theta).$$

This expression indicates that the VaR of the portfolio at time t for the next period is (minus) the θ -quantile of the conditional cdf of the portfolio return. Evidently, if we assume that returns are *iid*, the VaR is constant over time and is simply given by the inverse of the unconditional cdf : $VaR_\theta = -F_p^{-1}(\theta)$, where $F_p(x) = \Pr [r_{p,t} \leq x]$ is the (assumed to be continuous) cdf of the portfolio return. Computing the VaR therefore “reduces” to estimating a quantile of the conditional distribution of the portfolio return. Notice that this cdf is very likely to be time varying. One reason is that the volatility of returns varies over time, see also Chapter 2.

The distribution of the portfolio return can also be described as depending on the joint distribution of asset returns. This brings about two fashions to consider the VaR and ES. A first one, that we will call portfolio approach will consider the distribution of an aggregate return. The second one will consider the individual assets in an approach that we call the asset approach. In fact, the use of the asset-level approach is worthy to be considered, because it may be used for more in-depth analysis. This point has been outlined by Gouriéroux, Laurent, and Scaillet (2000). For instance, it allows computing the sensitivity of the portfolio VaR to changes in the weights of the portfolio. Ultimately, it can be used to directly optimize the weights of the portfolio under VaR constraints (see for instance, Huisman, Koedijk, and Pownall, 1999, or Krokmal, Palmquist, and Uryasev, 2002). This issue is also addressed in Section 9.2. Notice that a reduction of the dimensionality can be attained by

¹ Rigorously, the VaR is defined as

$$VaR_{\theta,t} = -\sup \{x | \Pr [r_{p,t+1} \leq x] \leq \theta\},$$

because the cdf may be constant over some interval. We use simpler notations for ease of exposition.

defining some factors supposed to capture the main sources of risk affecting the portfolio return.

To estimate the VaR of a large portfolio over time, we can adopt several approaches. The first approach is based on the computation of the VaR of the portfolio. An obvious advantage of such an approach is that it avoids modeling the joint dynamics of asset returns. A drawback is that it will probably miss some important links between asset returns such as the time-varying correlation. The second set of approaches is based on the modeling of the joint distribution of asset returns.

In general, the VaR is computed over a time horizon k (10 day, for instance). In such a case, we have, assuming the position is held constant over the horizon, the following definition for the multi-period VaR

$$\theta = \Pr \left[\frac{\Delta W_{t+k}(\alpha_t)}{W_t(\alpha_t)} \leq -\frac{\overline{VaR}_{\theta,t:t+k}}{W_t(\alpha_t)} \right] = \Pr [r_{p,t}[k] \leq -VaR_{\theta,t:t+k}],$$

where $r_t[k]$ is the cumulative return between t and $t + k$ and we defined the VaR for 1\$ invested as $VaR_{\theta,t:t+k} = \overline{VaR}_{\theta,t:t+k}/W_t(\alpha_t)$. To obtain the multi-period VaR, we therefore need an estimate of the conditional distribution of the multi-period portfolio return $r_{p,t}[k]$.

Coherent measure of risk

Artzner et al. (1997, 1999) proposed a set of conditions that a coherent measure of risk should satisfy:

Definition 8.1. Let V be a set of real-valued random variables (typically, the net final wealth). The function $\rho : V \rightarrow \mathbb{R}$ is a coherent risk measure if it satisfies:

1. *Translation invariance:* $X \in V$, $\alpha \in \mathbb{R}$, then $\rho(X + \alpha) = \rho(X) - \alpha$.
2. *Sub-additivity:* $X, Y \in V$, then $\rho(X + Y) \leq \rho(X) + \rho(Y)$.
3. *Positive homogeneity:* $X \in V$, $\lambda \geq 0$, then $\rho(\lambda X) = \lambda \rho(X)$.
4. *Monotonicity:* $X, Y \in V$, with $X \leq Y$, then $\rho(X) \geq \rho(Y)$.

Translation invariance means that if we add a sure amount α to the position, it will decrease the risk measure by α . Sub-additivity implies that the risk of a portfolio constituted of two sub-portfolios is smaller than the sum of the risk of the two sub-portfolios. Positive homogeneity means that if we increase the size of the portfolio by a factor λ with the same weights, we increase the risk measure by the same factor λ . Monotonicity means that the risk is greater for more negative random outcomes.

They then show that the VaR fails to be a coherent measure because it does not satisfy the sub-additivity property, so that diversification does not necessarily result in a reduction of risk as measured by VaR. Perhaps more importantly, it has been argued (BIS Committee on the Global Financial

System, 2000) that VaR (as well as variance) misregards the risk of extreme loss. As has been highlighted by Basak and Shapiro (2001), because VaR disregards risk of extreme losses behind the confidence level, it may induce large losses. Consequently, this measure may induce a larger risk exposure than the variance in case of down markets.

Expected Shortfall

To cope with these shortcomings, Artzner et al. (1999) and Basak and Shapiro (2001) propose the use of the so-called Expected Shortfall (ES) as an alternative measure of risk. Notice that the terminology is still not clearly established, because Conditional Expected Loss, Conditional VaR, or Tail Conditional Expectations are very closely related notions. ES is the expected value of the loss of the portfolio in the $\theta\%$ worst cases over a given time horizon

$$\overline{ES}_{\theta,t} = -E [\Delta W_{t+1}(\alpha_t) | \Delta W_{t+1}(\alpha_t) \leq -VaR_{\theta,t}] .$$

Alternatively, we have²

$$\begin{aligned} ES_{\theta,t} &= \frac{\overline{ES}_{\theta,t}}{W_t(\alpha_t)} = -E \left[\frac{\Delta W_{t+1}(\alpha_t)}{W_t(\alpha_t)} \middle| \frac{\Delta W_{t+1}(\alpha_t)}{W_t(\alpha_t)} \leq -VaR_{\theta,t} \right] \\ &= -E [r_{p,t} | r_{p,t} \leq -VaR_{\theta,t}], \end{aligned}$$

where $ES_{\theta,t}$ denotes the ES for probability θ for 1\$ invested.

The main advantages of the ES for asset allocation are the following: (i) ES is a coherent measure of risk, because it satisfies the sub-additivity property and consequently can be reduced by diversification; (ii) ES directly controls the risk in the left tail of the distribution, so that extreme losses are explicitly taken into account in the allocation process.

An additional advantage of ES over VaR from an asset allocation point of view is that portfolio optimization is easier to implement with ES objectives than with VaR objectives. The reason is that ES is convex, so that the problem can be solved by linear programming techniques, once the *cdf* has been approximated by its empirical counterpart (see Rockafellar and Uryasev, 2000, Krokhmal, Palmquist, and Uryasev, 2002, Rockafellar and Uryasev, 2002). Fermanian and Scaillet (2005) discuss how the VaR and the ES of a portfolio vary when the weights of the portfolio are slightly altered. See Section 9.2 for additional details on the asset allocation under expected shortfall constraints.

² Rigorously, the ES is defined as (see Acerbi and Tasche, 2001)

$$ES_{\theta,t} = -\frac{1}{\theta} (E [r_{p,t} | r_{p,t} \leq -VaR_{\theta,t}] - VaR_{\theta,t} \Pr [r_{p,t} \leq -VaR_{\theta,t}] - \theta) .$$

8.1.2 Models for portfolio returns

To summarize, computing the VaR or ES of a portfolio requires the following elements:

- the probability θ ,
- the horizon of the investment k ,
- the value $W_t(\alpha_t)$ of the portfolio at date t ,
- the *cdf* of the portfolio return.

The first three elements are given in practice, and the main task consists in estimating the *cdf* of the portfolio return.

As argued before, there is a huge literature on how to compute the VaR of a portfolio. One reason is probably that VaR involves several dimensions that can be dealt with using completely different approaches.

A first issue is the *aggregation level*. For a mere measure of the VaR, using a time series of portfolio returns is in general enough. In contrast, if we are interested in active portfolio management, it is more appropriate to evaluate the VaR for asset returns.³ In such a case, however, we face a problem of dimensionality, because actual portfolios may include several hundreds of assets.

A second issue is the *choice of the model* to estimate the conditional distribution of portfolio returns. For instance, the VaR is a high quantile. So it may be natural to estimate it using an approach that specifically focuses on the tails of the distribution (such as the EVT). But at the same time, it is known that the distribution of returns varies over time, in particular because of changes in volatility. Therefore, it is also of importance to correctly describe how the return distribution evolves through time. There are four broad categories of VaR models that we will discuss below: Historical simulation, semi-parametric models (such as Extreme Value Theory and CAViaR), parametric models (such as RiskMetrics and GARCH models), and finally non-linear techniques that are designed to compute the VaR in presence of derivatives.

Another issue is the fact that the portfolio value may be non-linearly affected by changes in asset prices, for instance when derivative assets are included in the portfolio. The difficulty is that in many cases, historical data is not available. In such cases, non-linear methods, based on Taylor's approximation of the portfolio value or on Monte Carlo simulations, may be used. See Section 8.5.

For large-scale portfolios, such as those managed in financial institutions, simultaneously modeling the joint dynamic of all asset returns may simply be impossible. In such cases, it is preferable to work with a reduced number of base assets (or risk factors) that are thought to drive risks. Adding risk factors

³ The portfolio-level approach is not appropriate for measuring the effect on the VaR of a change in the portfolio weights.

is analytically more demanding, because it requires an additional layer in the modeling step, but it is likely to reduce the eventual computational burden very significantly. Obviously, in such an approach, the main tasks consist in identifying the factors that may capture the various sources of risk and then in modeling their joint dynamics.

8.2 Historical simulation

Historical simulation is probably the simplest and most widely used approach for computing VaR and ES. It is fundamentally non-parametric, in the sense that it does not require any assumption about the distribution of returns. The method works as follows: Assume that a sample of T past realizations $\{r_1, \dots, r_T\}$ is available. We define a window size, N , that is used to construct subsamples of size N . Then, $T - N + 1$ overlapping subsamples $\{r_1, \dots, r_N\}, \dots, \{r_{T-N+1}, \dots, r_T\}$ are available. Each of these subsamples is used to approximate the *cdf* of the series. To do this, take one of these subsamples, for instance the t th one, $\{r_{t-N+1}, \dots, r_t\}$, and sort this subsample in increasing order. Define the sorted data as $\{\tilde{r}_{t-N+1,t}, \dots, \tilde{r}_{t,t}\}$ where $\tilde{r}_{t-N+1,t} \leq \dots \leq \tilde{r}_{t,t}$. Now, the VaR with probability $\theta\%$ is defined as the θ -quantile of the subsample. This is therefore the (θN) th order statistic $\tilde{r}_{\theta N,t}$ if θN is an integer. If θN is not an integer, the quantile is defined using a linear interpolation between $\tilde{r}_{[\theta N],t}$ and $\tilde{r}_{[\theta N]+1,t}$. To simplify, assume that θN is an integer and that $\tilde{r}_{\theta N,t}$ is the θ -quantile of the subsample. Then, the VaR at date t for date $t + 1$ is⁴

$$VaR_{\theta,t} = -\tilde{r}_{\theta N,t}.$$

Finally, the ES is defined as the average of the realizations that are below this level

$$ES_{\theta,t} = -\frac{1}{[\theta N]} \sum_{i=1}^{[\theta N]} \tilde{r}_{i,t}.$$

The method has two main advantages: First, it is very easy to implement. Second, it allows for non-normal returns. Indeed, it does account for fat tails. Since it uses the actual realizations of returns, it is able to capture most of the empirical features of this series.

However, this approach also raises several difficulties. First, although the historical simulation approach is said to be non-parametric, it is based on a strong underlying assumption: the return process is supposed to be *iid*. Empirical evidence clearly suggests that this is not the case. For this reason, the choice of the window size N is crucial in practice. Longer samples increase the accuracy of VaR estimates, but at the same time increase the probability of using irrelevant data, in particular if there are some changes in the underlying

⁴ Given that small quantiles of returns are negative, the VaR measure will be positive.

Table 8.1. Five smallest and largest daily returns

SP500	min	-22.87	-8.68	-7.15	-7.08	-7.04
	max	8.67	5.54	5.23	5.09	4.95
DAX	min	-13.74	-9.90	-9.89	-8.91	-8.41
	max	7.52	7.26	7.14	7.05	6.95
FT-SE	min	-11.95	-10.43	-7.46	-5.55	-5.39
	max	5.66	5.63	5.06	4.68	4.61
Nikkei	min	-16.14	-7.24	-6.87	-6.83	-6.14
	max	12.42	8.88	7.65	7.54	7.27

process. On the other hand, it would be meaningless to use a small window size if we are interested in a very small probability θ , because the historical simulation approach cannot produce a quantile that would be smaller than the minimum return observed within the given sample. If we want to compute the 1%-quantile, we need at least 100 observations in the subsample! Thus there is a trade-off between longer and shorter sample sizes.

Another undesirable feature is that the VaR obtained with this method does not vary often but when it does, it varies sharply. In fact, the only source of variation in this approach is the shift of the window over time. As a consequence, we observe jumps in the reported VaR, when an extreme (negative) return is introduced or dropped from the subsample. This problem is mainly due to the discreteness of the empirical distribution of returns. In particular, in the tails, the interval between adjacent returns can be rather large, as Table 8.1 illustrates. We display the 5 smallest and largest daily returns on the four indices at hand over the period from 1980 to 2004.

The problem is also illustrated in Figures 8.1 and 8.2, where we display the 1% and 5% VaR and ES of a portfolio constituted of the SP500, the DAX, the FT-SE, and the Nikkei (with equal weights). Since the window size is $N = 500$, they correspond to the 5th and the 25th order statistics of each subsample. One explanation for this pattern is that all realizations in the subsample have the same weight. To partially cope with this problem, we may, for instance, introduce declining weights to past observations to smooth computations.

8.3 Semi-parametric approaches

The problem with the historical simulation approach stems from the fact that returns are not *iid* and from the discreteness of the tails of the empirical distribution. The two semi-parametric approaches presented now put some structure on the tails of the distribution by using parametric models but without estimating the complete distribution of returns. To some extent, they are also able to cope with the temporal dependence of returns. The first one is based on the EVT presented in detail in Section 7.1.4. The second approach is

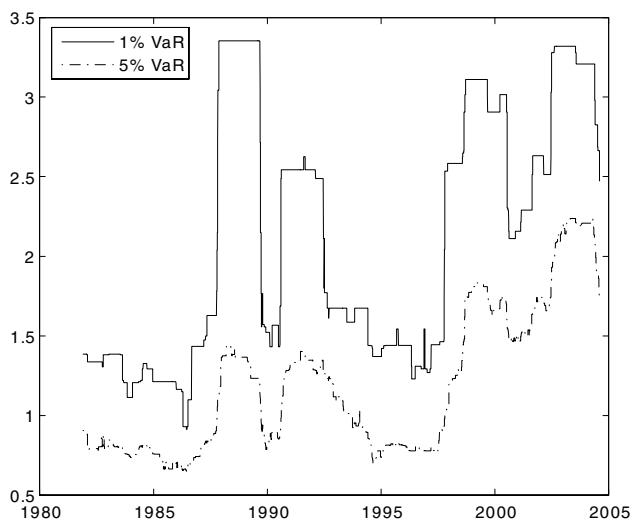


Fig. 8.1. *VaR of portfolio computed with the historical simulation approach.*

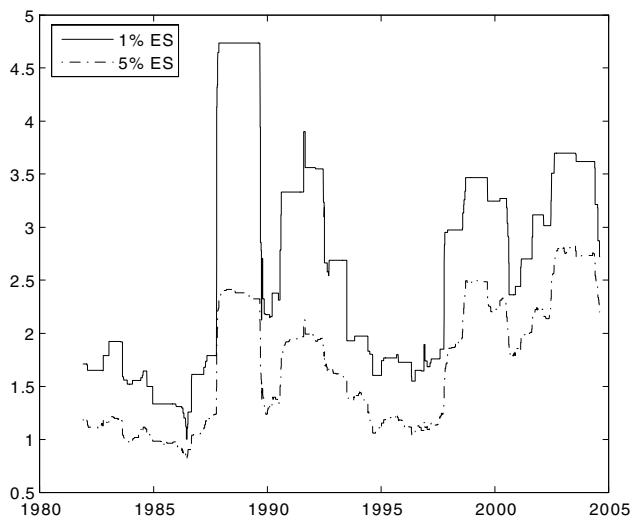


Fig. 8.2. *ES of portfolio computed with the historical simulation approach.*

based on a regression quantile technique as developed by Engle and Manganelli (2004) and Chernozhukov and Umantsev (2001).

8.3.1 Extreme Value Theory (EVT)

The EVT is designed to model the specific behavior of very large (positive or negative) returns. It provides a parametric representation of the distribution of the extremes (or of the tails). This representation is able to give smoothed estimates of the VaR and is designed to produce very high quantiles, possibly smaller than the minimum of the sample distribution. Such an approach has been adopted for VaR evaluation by several authors: McNeil (1997) estimates a gpd to evaluate the tail index; Danielsson and de Vries (1997) uses a semi-parametric estimate of the tail index; McNeil and Frey (2000) introduce a GARCH–EVT model that incorporates the temporal evolution of volatility; Longin (2000) adopts a multivariate approach to capture the joint extreme behavior of risk factors.

In Section 7.1.4, we described how the quantile of a univariate distribution can be computed using the various approaches developed in the context of the EVT: the distribution of extremes, the distribution of the tails or the semi-parametric estimation of the tail index. We briefly recall the estimation of high quantiles in the case of the tail approach.

Unconditional EVT

The main idea of the tail approach to EVT is that the distribution of the lower tail (i.e., the returns that are below a given threshold u) can be approximated, when u is sufficiently large (in absolute value), by the so-called generalized Pareto distribution (gpd). The gpd is defined as

$$G_{\xi,u,\psi}(r) = \begin{cases} 1 - \left(1 + \frac{\xi}{\psi}(r-u)\right)^{-1/\xi}, & \text{if } \xi \neq 0, \\ 1 - \exp\left(-\frac{(r-u)}{\psi}\right), & \text{if } \xi = 0, \end{cases}$$

where the tail index ξ characterizes the shape of the tails of the distribution and ψ is a scaling parameter.

Assume that the estimation of the gpd is performed on the absolute value of the lower tail observations, as detailed in Section 7.1.2. Once the parameters of this gpd are estimated, we can deduce the θ -quantile of the actual distribution as

$$q_\theta = \begin{cases} u + \frac{-\hat{\psi}}{\hat{\xi}} \left(\left(\frac{T}{N_u} \theta \right)^{-\hat{\xi}} - 1 \right), & \text{if } \xi \neq 0, \\ u + \hat{\psi} \log \left(\frac{T}{N_u} \theta \right), & \text{if } \xi = 0, \end{cases}$$

where N_u is the number of exceedances below the threshold. Evidently, since it is computed using results that are valid for extreme returns only, the quantile

q_θ is valid for very small probability θ . Then, the VaR at date t for date $t+1$ is

$$VaR_{\theta,t} = -q_\theta,$$

so that the estimated VaR is actually constant over time.

As outlined by McNeil and Frey (2000), the ES is strongly related to the notion of mean excess function, because we have the relation

$$\begin{aligned} ES_{\theta,t} &= E[-r| -r \geq VaR_{\theta,t}] \\ &= VaR_{\theta,t} + E[-r - VaR_{\theta,t}| -r \geq VaR_{\theta,t}]. \end{aligned}$$

Recall that, if $r - u|r > u \sim G_{\xi,\psi}$, the mean excess function above level u is defined as

$$e(u) = E[-r - u| -r \geq u] = \frac{\psi + \xi u}{1 - \xi}, \quad \psi + \xi u > 0.$$

Therefore, the expected shortfall is equal to the sum of the VaR and the mean excess function above the VaR. Now, since $VaR_{\theta,t} > u$, we may write

$$\begin{aligned} E[-r - VaR_{\theta,t}| -r \geq VaR_{\theta,t}] \\ = E[(-r - u) - (VaR_{\theta,t} - u)| (-r - u) \geq (VaR_{\theta,t} - u)], \end{aligned}$$

with $-r - VaR_{\theta,t} | -r \geq VaR_{\theta,t} \sim G_{\xi,\psi+\xi(VaR_{\theta,t}-u)}$. Finally, we deduce the estimated ES

$$ES_{\theta,t} = VaR_{\theta,t} + \frac{\hat{\psi} + \hat{\xi}(VaR_{\theta,t} - u)}{1 - \hat{\xi}} = \frac{VaR_{\theta,t}}{1 - \hat{\xi}} + \frac{\hat{\psi} - \hat{\xi}u}{1 - \hat{\xi}}.$$

This approach is called unconditional EVT, because it is based on the assumption that returns are *iid*. It therefore produces an unconditional VaR, i.e., pertaining to the unconditional distribution of returns. This implies that the quantile q_θ will not vary over time, even if there is a sudden change in market conditions. Recently, McNeil and Frey (2000) have proposed a model that provides conditional measures of the VaR based on EVT. This approach is described below.

Another drawback of the tail approach is that we have to select a threshold u below which we consider that returns belong to the tail. Although some tools have been proposed to select this threshold (such as bootstrap techniques), it is still an open question how to define the optimal threshold. Figure 8.3 displays the 1%-quantile $q_{0.01}$ as a function of the threshold u for the SP500. This plot suggests that the choice of the threshold does not necessarily play a crucial role for the value of the quantile, given that the VaR varies only between 3.77 and 3.87 as u varies between 0 and 0.1.

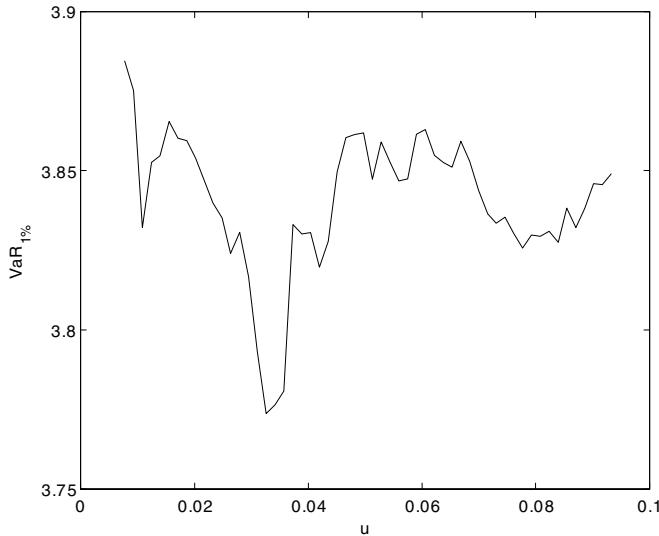


Fig. 8.3. 1%-quantile of the SP500 as a function of the threshold u .

The GARCH–EVT model

The GARCH–EVT model proposed by McNeil and Frey (2000) consists in modeling the conditional volatility and the distribution of the tails separately. In fact, only the left tail is needed, because the right tail is not relevant for VaR computation.

For this purpose, McNeil and Frey (2000) proceed as follows: In the first step, they filter the dependence in the return series by computing the residuals of a GARCH model, which should be *iid* if the GARCH model correctly fits the data. In the second step, they model the extreme behavior of the residual using the tail approach developed in Section 7.1.2. Finally, in order to produce a VaR estimate for the original return, they trace back the steps by first producing the θ -quantile estimate for the GARCH-filtered residuals and convert the θ -quantile estimate to the original return using the conditional volatility forecast for the required horizon.

For a given return series $\{r_1, \dots, r_T\}$, the model adopted to filter out the first- and second-order dynamics is of the form

$$\begin{aligned} r_t &= \mu_t + \varepsilon_t, \\ \varepsilon_t &= \sigma_t z_t, \\ \mu_t &= \mu + \varphi r_{t-1}, \\ \sigma_t^2 &= \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2, \end{aligned}$$

with $\omega > 0$, $\alpha \geq 0$, $\beta \geq 0$, and $\alpha + \beta < 1$ to ensure a positive volatility and a covariance stationary process. To avoid specifying an arbitrary distribution for the innovation process z_t , they resort to the Quasi Maximum Likelihood (QML) estimation, which consists in maximizing the normal log-likelihood of the model even though the true generating process of z_t is not Gaussian. This technique has been shown to provide consistent estimates of the model parameters, provided the conditional mean and variance equations are correctly specified (Gouriéroux, Monfort, and Trognon, 1984). The standardized residual \hat{z}_t is then estimated by $\hat{z}_t = (r_t - \hat{\mu}_t) / \hat{\sigma}_t$, where $\hat{\mu}_t$ and $\hat{\sigma}_t^2$ are the fitted mean and variance, respectively. We also use this model to produce forecasts of the expected return $\mu_t(1) = \hat{\mu} + \hat{\varphi}r_t$ and variance $\sigma_t^2(1) = \hat{\omega} + \hat{\alpha}\hat{\varepsilon}_t^2 + \hat{\beta}\hat{\sigma}_t^2$.

The next step involves estimating the gpd to all exceedances, i.e., all realizations \hat{z}_t that are below a given high threshold u .⁵ Adopting a similar approach to the one in Section 7.1.2, we define $N_u = \sum_{t=1}^T \mathbf{1}_{\{\hat{z}_t < u\}}(\hat{z}_t)$ the number of exceedances and $\{\hat{z}_{1,T}, \dots, \hat{z}_{T,T}\}$ the vector of standardized residuals sorted by increasing order, such that $\hat{z}_{1,T} \leq \dots \leq \hat{z}_{T,T}$. Finally, we define the $(N_u, 1)$ vector of exceedances as $\{\hat{z}_{1,T}, \dots, \hat{z}_{N_u,T}\}$. Then, we estimate the parameters $(\xi, \psi)'$ of the gpd to the exceedances $\{\hat{z}_{i,T}\}_{i=1}^{N_u}$.

Once the parameters ξ and ψ are estimated, the θ -quantile is obtained by inverting the *cdf* of exceedances

$$q_\theta = \begin{cases} u + \frac{-\hat{\psi}}{\hat{\xi}} \left(\left(\frac{T}{N_u} \theta \right)^{-\hat{\xi}} - 1 \right), & \text{if } \xi \neq 0, \\ u + \hat{\psi} \log \left(\frac{T}{N_u} \theta \right), & \text{if } \xi = 0. \end{cases}$$

We are now ready to evaluate the aggregate VaR and ES from t to $t+1$

$$\begin{aligned} VaR_{\theta,t} &= -(\mu_t(1) + q_\theta \sigma_t(1)), \\ ES_{\theta,t} &= \frac{VaR_{\theta,t}}{1 - \hat{\xi}} + \frac{\hat{\psi} - \hat{\xi}u}{1 - \hat{\xi}}. \end{aligned}$$

It is worth emphasizing that the GARCH–EVT approach incorporates the two ingredients required for an accurate evaluation of the conditional VaR, i.e., a model for the dynamics of the first and second moments, and an appropriate model for the conditional distribution. An obvious improvement of this approach as compared to the unconditional EVT is that it incorporates in the VaR changes in expected return and in volatility. For instance, if we assume a change in volatility over the recent period, the GARCH–EVT is able to incorporate this new feature in its VaR evaluation, whereas the unconditional EVT remains stuck at the average level of volatility over the estimation sample.

McNeil and Frey (2000) also provide a back-testing experiment, in which they compare the performances of various methods to correctly reproduce

⁵ Notice that we focus in this section on the lower tail, whereas all developments in Chapter 7 were based on the upper tail.

the quantiles of several asset returns. They show that the GARCH–EVT performs much better than the unconditional EVT, suggesting that the ability to capture changes in volatility is crucial for VaR computation.

Multivariate EVT

The VaR computation described above is useful for a single asset or for cases where there is only one risk factor. The cases for multi-asset and multi-risk-factor are a lot more complex, because they resort to multivariate EVT. We have shown in Section 7.2 how to characterize and measure extremal dependence. The measures introduced are based on the evaluation of the properties of the joint distribution at equal probability marginal quantiles. To use these measures to characterize extremes of a portfolio, i.e., of a linear combination of assets, the full joint distribution needs to be estimated in the tail region. For the case where returns are asymptotically dependent or exactly independent, such methods exist, see Coles and Tawn (1994), and de Haan and de Ronde (1998). Poon, Rockinger, and Tawn (2004) follow the approach of Ledford and Tawn (1997) for handling asymptotic independence via non-parametric and parametric approaches. VaR and ES computations rely on Monte Carlo simulations.

8.3.2 Quantile regression technique

In the quantile regression approach, another route is taken. Instead of focusing on the modeling of the tail distribution as in the EVT, this approach focuses on its dynamic component. The technique of quantile regression has been introduced in the statistical literature by Koenker and Bassett (1978).⁶ It has been applied to VaR computation by Engle and Manganelli (2004) and Chernozhukov and Umantsev (2001). The basic idea consists in modeling a given quantile of the distribution through time. Such an approach is justified by that empirical evidence that volatility tends to cluster, so that the distribution itself is serially correlated. Engle and Manganelli (2004) described a new regression quantile model, called CAViaR, standing for Conditional AutoRegressive Value-at-Risk, which allows modeling the dynamics of the VaR. In this section, we start by reviewing the concept of quantile regression before turning to the one of CAViaR.

Quantile regression

The θ th quantile q_θ of a time series $\{y_t\}_{t=1}^T$ can be defined as the solution of the following minimization problem⁷

⁶ A general presentation can be found in Koenker and Hallock (2001).

⁷ A well-known special case is the median ($q = 1/2$) in which the median β is defined as

$$\min_{q \in \mathbb{R}} \sum_{t \in \{t: y_t \geq q\}} \theta |y_t - q| + \sum_{t \in \{t: y_t < q\}} (1 - \theta) |y_t - q|, \quad 0 < \theta < 1,$$

or, equivalently

$$\min_{q \in \mathbb{R}} \sum_{t=1}^T w_\theta(y_t - q),$$

where

$$w_\theta(z_t) = \begin{cases} \theta z_t, & \text{if } z_t \geq 0, \\ (1 - \theta) z_t, & \text{if } z_t < 0. \end{cases}$$

The estimated q is the unconditional θ -quantile of $\{y_t\}_{t=1}^T$. Now, in a regression context, $y_t = x'_t \beta + u_t$, where x_t is a $(k, 1)$ vector of regressors and u_t is the error term with *cdf* F , the θ th regression quantile of $u_t = y_t - x'_t \beta$ conditional on x_t is obtained as the solution of the minimization problem

$$\min_{\beta \in \mathbb{R}^k} \sum_{t \in \{t: y_t \geq x'_t \beta\}} \theta |y_t - x'_t \beta| + \sum_{t \in \{t: y_t < x'_t \beta\}} (1 - \theta) |y_t - x'_t \beta|, \quad 0 < \theta < 1. \quad (8.1)$$

The purpose of this regression is to find the vector of parameters β that will ensure that the θ -quantile of u_t will be as close to 0 as possible. A well-known quantile regression is the one associated with $\theta = 1/2$. In this case, β is optimized in order to obtain a median equal to 0.

The estimation technique and the asymptotic properties of the estimator of β are developed in Koenker and Bassett (1978). This class of *robust estimators* includes some that have similar efficiency to the least-square estimator under normality of the error term but that out-perform this estimator under non-normal errors.

Koenker and Bassett (1978) also provide some useful asymptotic results on regression quantiles. In particular, the solution $\hat{\beta}_T(\theta)$ of the optimization problem 8.1 for quantile θ is shown to be consistent and asymptotically normal. This result is extended by Engle and Manganelli (2004) to the case of non-linear regressions.

CAViaR

Engle and Manganelli (2004) proposed a model based on quantile regressions, called CAViaR. The idea is to directly model the evolution of the quantile q_θ instead of considering the entire distribution of returns. The general specification they propose has the form

$$q_{\theta,t} = \beta_0 + \beta_1 q_{\theta,t-1} + \sum_{j=1}^r \beta_{j+1} g(x_{t-j}),$$

$$\min_{\beta \in \mathbb{R}} \sum_{t=1}^T |y_t - \beta|.$$

where x_t is a set of regressors and $g(\cdot)$ a possibly non-linear function of the regressors. Some of the suggested specifications are particularly appealing:

Symmetric absolute value model

$$q_{\theta,t} = \beta_0 + \beta_1 q_{\theta,t-1} + \beta_2 |r_{t-1}|.$$

Asymmetric slope model

$$q_{\theta,t} = \beta_0 + \beta_1 q_{\theta,t-1} + \beta_2 1_{\{r_{t-1}>0\}} + \beta_3 1_{\{r_{t-1}<0\}}.$$

Indirect GARCH model

$$q_{\theta,t}^2 = \beta_0 + \beta_1 q_{\theta,t-1}^2 + \beta_2 r_{t-1}^2.$$

The parameters of the quantile regressions are then estimated using non-linear techniques to solve the optimization problem

$$\min_{\beta \in \mathbb{R}^k} \sum_{t \in \{t: y_t \geq q_{\theta,t}(\beta)\}} \theta |y_t - q_{\theta,t}(\beta)| + \sum_{t \in \{t: y_t < q_{\theta,t}(\beta)\}} (1-\theta) |y_t - q_{\theta,t}(\beta)|$$

with $0 < \theta < 1$. This approach is qualified as semi-parametric, because it does not specify the distribution of returns. It can therefore be applied to non-*iid* returns as well as to time-varying volatility. It should be mentioned that the estimation of such a model is far from trivial. Engle and Manganelli (2004) also propose some test procedures for evaluating VaR models. This issue is addressed in detail in Section 8.6.1.

8.4 Parametric approaches

While the historical simulation and the unconditional EVT approaches are by nature unable to capture changes in the behavior of returns, the GARCH–EVT technique and CAViaR incorporates some stylized facts of returns concerning the first- and second-order dynamics. We now turn to parametric models of VaR that hold under a complete set of assumptions concerning the dynamics and the conditional distribution of returns.⁸

We recall here that the main empirical features for asset returns are the following:

1. Returns may be serially correlated, even if this correlation is not very large in practice.
2. Return volatility is serially correlated and possibly asymmetric. These features can be captured by the well-known GARCH family.
3. The conditional distribution of returns is probably non-normal. Typical characteristics of this distribution are skewness and fat-tailedness.

⁸ Evidently, these assumptions can sometimes be excessively strong.

These three components have not been introduced systematically in previous approaches. For instance, the widely used RiskMetrics methodology assumes normality of returns. As long as such an assumption yields correct forecasts of VaR and ES, it may be relevant. A thorough test of the implications of the assumptions and a test for the validity of the model appears as crucial.

8.4.1 RiskMetrics – J.P. Morgan

RiskMetrics is a methodology developed by J.P. Morgan to compute VaR. It has played an important role in the increasing popularity of VaR as a risk measure. See J.P. Morgan's RiskMetrics Technical Document (1996).

The RiskMetrics methodology uses historical return data to forecast future volatility. More precisely, the basic RiskMetrics model is based on the following assumptions:

1. The return r_t is modeled as $r_t = \mu_t + \varepsilon_t$ with $\varepsilon_t = \sigma_t z_t$.
2. Daily log-returns are supposed to be centered (or are preliminary de-meaned), so that $\mu_t = 0$.
3. The dynamic of volatility is modeled using an exponentially weighted moving average (EWMA), with

$$\sigma_t^2 = \lambda \sigma_{t-1}^2 + (1 - \lambda) r_{t-1}^2, \quad \text{for } t = 2, \dots, T, \quad (8.2)$$

with $0 < \lambda < 1$. This model may be viewed as a special case of the Integrated GARCH model. In the first versions of RiskMetrics, the decay factor λ was chosen to be equal to 0.94. The recursion can be initialized by the sample variance ($\sigma_1^2 = \hat{\sigma}^2$) or by the square of the first observation ($\sigma_1^2 = r_1^2$).

4. The innovation z_t is supposed to be distributed as an *iid* $\mathcal{N}(0, 1)$.

Under these various assumptions, the conditional distribution of r_t at date t is $\mathcal{N}(0, \sigma_t^2)$. If we are interested in the one-step-ahead VaR, we need the conditional distribution of r_{t+1} . Conditionally on the information at time t , it is $\mathcal{N}(0, \sigma_t^2(1))$ where $\sigma_t^2(1) = \lambda\sigma_t^2 + (1 - \lambda)r_t^2$. For the $\mathcal{N}(0, 1)$ distribution, we denote $q_\theta = \Phi^{-1}(\theta)$ the quantile for a probability of loss equal to θ . For instance, for $\theta = 1\%$, we have $q_\theta = -2.326$. Now, the daily VaR is given by

$$VaR_{\theta,t} = -q_\theta \times \sigma_t(1). \quad (8.3)$$

In addition, we have the following expression for the ES

$$\begin{aligned} ES_{\theta,t} &= -E[r_t | r_t \leq -VaR_{\theta,t}] \\ &= -E\left[\frac{r_t}{\sigma_t(1)} \mid \frac{r_t}{\sigma_t(1)} \leq \frac{-VaR_{\theta,t}}{\sigma_t(1)}\right] \times \sigma_t(1) \\ &= ES_{\theta,t}^z \times \sigma_t(1) \end{aligned}$$

where

$$ES_{\theta,t}^z = -\frac{1}{\Phi(q_\theta)} \int_{-\infty}^{-q_\theta} z \times \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right) dz = \frac{1}{\theta} \varphi(q_\theta)$$

where $\varphi(z)$ denotes the *pdf* of the $\mathcal{N}(0, 1)$ distribution. Finally, we obtain

$$ES_{\theta,t} = \frac{\varphi(q_\theta)}{\theta} \sigma_t(1).$$

Multi-period VaR

One interesting property of the RiskMetrics approach is that a multi-period VaR can be very easily computed. Assume we have to compute the VaR over the following k periods. We observe that the log-return between t and $t+k$ is simply defined as $r_t[k] = r_{t+1} + \dots + r_{t+k}$. We deduce that the volatility forecast for the k -period log-return is

$$\sigma_t^2[k] = V_t[r_t[k]] = V_t \left[\sum_{i=1}^k r_{t+i} \right].$$

As we have seen in Section 4.3, for an IGARCH model, the κ -step ahead volatility forecast is

$$\sigma_t^2(\kappa) = \dots = \sigma_t^2(2) = \sigma_t^2(1) = \lambda \sigma_t^2 + (1 - \lambda) r_t^2, \quad \text{for } \kappa \geq 1,$$

because the innovation process is *iid*. Therefore, the volatility forecast for the k -period log-return is

$$\sigma_t^2[k] = k \sigma_t^2(1),$$

so that the volatility forecast of $r_t[k]$ is proportional to the horizon k . We finally obtain that the k -day VaR is given by

$$VaR_{\theta,t:t+k} = -q_\theta \times \sqrt{k} \sigma_t(1).$$

This expression is known as the square-root-of-time rule. This rule has been recommended by the Basle Committee's Amendment to the Capital Accord to Incorporate Market Risks (1996) in order to compute the 10-day VaR from daily estimates.

Multiple position

Above, we computed the VaR of a single asset, a case that is not very interesting in practical applications. Notice that we may alternatively view r_t as the return of the portfolio, not as the return of an asset. In such a case, consistently with the portfolio-level approach, the computations correspond to the aggregate VaR. Notice that this interpretation also raises some new

difficulties. Since the portfolios of financial institutions are likely to change daily, it implies that each day, the risk manager would have to compute a historical time series of the new portfolio and to estimate the aggregate VaR using this approach.

Another way, consistent with the asset-level approach, has been suggested for the latter. It is based on the observation that the aggregate VaR can be rewritten as

$$VaR_{\theta,t} = -q_\theta \times \sigma_{p,t}(1) = -q_\theta \times \sqrt{\alpha' \Sigma_t(1) \alpha},$$

where $\Sigma_t(1)$ is the one-step-ahead forecast of the covariance matrix of asset returns. It is computed assuming that all variances and covariances are driven by the same model (8.2)

$$\begin{aligned}\sigma_{i,t}^2 &= \lambda \sigma_{i,t-1}^2 + (1 - \lambda) r_{i,t-1}^2, && \text{for } i = 1, \dots, n, \\ \sigma_{ij,t} &= \lambda \sigma_{ij,t-1} + (1 - \lambda) r_{i,t-1} r_{j,t-1}, && \text{for } i, j = 1, \dots, n,\end{aligned}$$

using the same parameter λ for all assets.

It is clear that such an approach is equivalent to computing the VaR of each asset in the portfolio and then deducing the VaR of the portfolio. To see this, we denote $VaR_{\theta,t}^i$ the VaR of asset $i = 1, \dots, p$. Then, we observe that, for a given asset, we have the relation between variance and VaR

$$\sigma_{i,t}^2(1) = \left(\frac{1}{q_\theta} VaR_{\theta,t}^i \right)^2,$$

so that the VaR of the portfolio is now

$$\begin{aligned}VaR_{\theta,t} &= -q_\theta \sqrt{\alpha' \Sigma_t(1) \alpha} \\ &= -q_\theta \sqrt{\sum_{i=1}^n \alpha_i^2 \sigma_{i,t}^2(1) + 2 \sum_{i=1}^n \sum_{j=i+1}^n \alpha_i \alpha_j \sigma_{ij,t}(1)}.\end{aligned}$$

If we use $\sigma_{ij,t}(1) = \sigma_{i,t}(1) \sigma_{j,t}(1) \rho_{ij,t}(1)$ where $\rho_{ij,t}(1)$ denotes the one-period forecast of the correlation between assets i and j , we obtain after simplification (see, for instance, Longin, 2000)

$$(VaR_{\theta,t})^2 = \sum_{i=1}^n \alpha_i^2 (VaR_{\theta,t}^i)^2 + 2 \sum_{i=1}^n \sum_{j=i+1}^n \rho_{ij,t}(1) \alpha_i \alpha_j VaR_{\theta,t}^i VaR_{\theta,t}^j.$$

The squared VaR of the portfolio is simply the quadratic form of the VaR of returns, weighted by the correlation matrix between returns. Computing the VaR in this way does not even require the knowledge of the portfolio weights.

Advantages and limitations

The main advantage of RiskMetrics is its simplicity of implementation. If we are willing to accept the value of the decay factor adopted by RiskMetrics, no estimation is needed, and the update of the VaR of any portfolio is extremely fast. In addition, computing multi-period VaR or multi-position VaR does not raise any additional difficulty. Obviously, this simplicity has a cost. Some of the underlying assumptions are in fact overly strong.

First, the assumption of normality of innovations is simply untenable. Most asset returns are characterized by a distribution with fat tails and/or asymmetry. Such assumption will lead in general to an underestimation (in absolute value) of the quantile q_θ to be used in the VaR formula.

Second, the dynamics of volatility is too simplistic. Although the IGARCH model has the advantage that the multi-period VaR can be computed using the simple square-root-of-time rule, it also yields some undesirable properties, such as the lack of mean-reversion in the variance process. This issue has been investigated by Diebold et al. (1998). They show that the square-root-of-time rule produces overestimates of the variability of long-horizon volatility.

A last, less stringent, assumption is the absence of dynamics for the expected return. This is probably too strong an assumption, but in practice, the effect of altering the conditional mean equation on the VaR measures is barely noticeable. The reason is that for most asset returns, the VaR computation is largely dominated by volatility, rather than expected return, considerations.

8.4.2 The portfolio-level approach

The variance-covariance method, also known as the correlation method, is essentially a parametric approach in which the VaR is measured from the variances and covariances of the constituents of a portfolio. A simple version of this approach is the RiskMetrics methodology. The main task of this approach is to model the different components of the dynamic of returns that are useful for VaR computation. As seen above, these components are: The dynamic of expected returns, the dynamic of volatility and finally the conditional distribution of the innovation process.

We assume that returns are possibly autocorrelated, that volatility σ_t is driven by a GARCH(1, 1) model and that innovations z_t are distributed as an *iid* $\mathcal{N}(0, 1)$. The conditional distribution of r_t at date t is then $\mathcal{N}(\mu_t, \sigma_t^2)$. The main change as compared with the RiskMetrics method is a conceptual one: the parameters of this model have to be estimated, rather than calibrated. This estimation is more demanding in terms of computational burden, in particular if a set of asset returns must be modeled.

The choice of a GARCH(1, 1) model with parameters α and β estimated without the assumption $\alpha + \beta = 1$ is very likely to provide a more realistic mean-reverting behavior of volatility forecasts. The 1-step ahead forecasts for μ_{t+1} and σ_{t+1}^2 are

$$\begin{aligned}\mu_t(1) &= \mu + \varphi_1 r_t, \\ \sigma_t^2(1) &= \omega + \alpha \varepsilon_t^2 + \beta \sigma_t^2.\end{aligned}$$

Conditionally on the information at time t , the conditional distribution of r_{t+1} is $\mathcal{N}(\mu_t(1), \sigma_t^2(1))$. The daily VaR is given by

$$VaR_{\theta,t} = -(\mu_t(1) + q_\theta \times \sigma_t(1)),$$

and the ES is

$$\begin{aligned}ES_{\theta,t} &= \frac{1}{\theta} \varphi \left(\frac{-VaR_{\theta,t} - \mu_t(1)}{\sigma_t(1)} \right) \sigma_t(1) - \mu_t(1) \\ &= \frac{\varphi(q_\theta)}{\theta} \sigma_t(1) - \mu_t(1).\end{aligned}\tag{8.4}$$

Conditional distribution

Although the model above is able to capture the dynamic in expected returns as well as in volatility, the conditional distribution is still assumed to be normal. Empirical evidence suggests that distributions allowing fat tails and asymmetry should be used for modeling the innovations. In Chapter 5, we described several alternative distributions that may be used in such a context. For instance, the Student t distribution may be very easily used in place of the normal distribution, because procedures to compute the inverse of its cdf are available in most econometric software. The skewed Student t distribution also appears as an obvious alternative.⁹

For instance, assume now that the innovation process is drawn from a standardized Student t distribution with ν degrees of freedom. It is worth emphasizing that in the context of a GARCH model, the innovation process z_t is supposed to have zero mean and unit variance. Consequently, the appropriate distribution is not the usual, but the standardized t , defined as

$$t(z_t|\nu) = c \left(1 + \frac{z_t^2}{\nu - 2} \right)^{-\frac{\nu+1}{2}},$$

with $\nu - 2$ in place of ν and $c = \Gamma(\frac{\nu+1}{2}) / (\sqrt{\pi(\nu-2)} \Gamma(\frac{\nu}{2}))$. Since this is the usual t that is available in most software, the quantiles have to be appropriately corrected.¹⁰ Finally, the θ -VaR for r_{t+1} is given by

$$VaR_{\theta,t} = -(\mu_t(1) + \tilde{q}_\theta \times \sigma_t(1)),$$

⁹ Since the VaR computation only involves the tails of the distribution, the quantiles of the skewed Student t distribution can be computed using the procedure built for the standard Student t , provided the asymmetric component is properly taken into account.

¹⁰ More precisely, if \tilde{z}_q is the q -quantile of the usual t distribution with ν degrees of freedom (in general given by econometric software), then the quantile \tilde{z}_q of

where $\tilde{q}_\theta = t_\nu^{-1}(\theta)$ is the quantile for a probability of loss equal to θ from the standardized t distribution. For the ES, we have to evaluate the same expression as for the normal case

$$ES_{\theta,t} = ES_{\theta,t}^z \times \sigma_t(1) - \mu_t(1),$$

where

$$\begin{aligned} ES_{\theta,t}^z &= -\frac{1}{t_\nu(\tilde{q}_\theta)} \int_{-\infty}^{-\tilde{q}_\theta} z \times c \left(1 + \frac{z^2}{\nu - 2}\right)^{-\frac{\nu+1}{2}} dz \\ &= \frac{c \nu - 2}{\theta \nu - 1} \left(1 + \frac{\tilde{q}_\theta^2}{\nu - 2}\right)^{-\frac{\nu+1}{2}}, \end{aligned}$$

so that we have eventually

$$ES_{\theta,t} = \frac{c \nu - 2}{\theta \nu - 1} \left(1 + \frac{\tilde{q}_\theta^2}{\nu - 2}\right)^{-\frac{\nu+1}{2}} \times \sigma_t(1) - \mu_t(1). \quad (8.5)$$

Multi-period VaR

Computing the VaR over k periods requires the cumulative expected return and volatility forecasts over k -periods. Using, abusively, the definition of the multi-period log-return, the k -period expected return is

$$\mu_t[k] = \mu_{t+1} + \dots + \mu_{t+k} = k\mu + \frac{1 - \varphi_1^k}{1 - \varphi_1} (\mu_t(1) - \mu), \quad \text{for } k > 1.$$

Moreover, the k -step ahead volatility forecast of a GARCH(1, 1) is given by

$$\sigma_t^2(k) = \sigma^2 + (a + b)^{k-1} (\sigma_t^2(1) - \sigma^2), \quad \text{for } k > 1.$$

Summing these volatilities, we obtain the volatility forecast for the k -period log-return

$$\sigma_t^2[k] = k\sigma^2 + \frac{1 - (a + b)^k}{1 - a - b} (\sigma_t^2(1) - \sigma^2), \quad \text{for } k > 1.$$

Contrary to what we obtained with RiskMetrics, the volatility forecast of $r_t[k]$ is now mean-reverting, meaning that the dynamic of volatility estimated over the sample plays a role in the forecast process. Finally, the k -day VaR is given by

the standardized t distribution is deduced using the relation: $\tilde{z}_q = \sqrt{\frac{\nu-2}{\nu}} \check{z}_q$. For instance, for $\nu = 5$, the 1%-quantile of the usual t distribution is $\check{z}_q = -3.3649$, and the corresponding quantile of the standardized t distribution is $\tilde{z}_q = -2.6065$. As it can be noticed, the difference is very significant for small degrees of freedom.

$$VaR_{\theta,t:t+k} = -(\mu_t[k] + q_\theta \times \sigma_t[k]),$$

where

$$\mu_t[k] = k\mu + \frac{1 - \varphi_1^k}{1 - \varphi_1} (\mu_t(1) - \mu), \quad \text{for } k > 1.$$

This expression for the VaR is less intuitive than the square-root-of-time rule. But it is also more consistent with the observed dynamic of volatility, in that it displays mean-reversion.

8.4.3 The asset-level approach

The asset-level estimation of the aggregate VaR has been advocated as allowing for a better control on the VaR estimation. As already argued, it allows measuring the effect on the aggregate VaR of a change in portfolio weights (the portfolio-level approach would require a complete re-estimation of the model). However, this advantage comes at a cost. Since we are interested in the modeling of the joint dynamic of asset returns, we have to turn to a multivariate GARCH-type model. As we have seen in Section 6.1, this approach raises a dimensionality problem, even for a moderate number of assets. Even though we emphasized the importance of non-Gaussian distributions for returns, the multivariate modeling in a non-Gaussian setting still represents a challenge. For this reason, at the asset-level approach, it is customary to assume a multivariate Gaussian distribution.

Once the multivariate GARCH model is estimated, computing the aggregate VaR is easy, since the distribution of the portfolio return is

$$r_{p,t} \sim \mathcal{N}(\mu_{p,t}, \sigma_{p,t}^2),$$

with $\mu_{p,t} = \alpha'_t \mu_t$ and $\sigma_{p,t}^2 = \alpha'_t \Sigma_t \alpha_t$. We deduce the aggregate VaR as

$$VaR_{\theta,t} = -(\mu_{p,t}(1) + q_\theta \times \sigma_{p,t}(1)),$$

where $q_\theta = \Phi^{-1}(\theta)$ is the quantile for a probability of loss equal to θ , from the univariate normal distribution $\mathcal{N}(0, 1)$. The ES is given by expression (8.4).

Conditional distribution

The main limitation of this approach is the maintained assumption of multivariate normality. Although the estimation of the DCC model can be performed under normality with a reasonable computational burden, another distributional assumption would dramatically increase the burden for large-dimension portfolios. The reason is that in such case the log-likelihood cannot be broken in separate components anymore. To be more precise, two cases have to be considered:

- For elliptical distributions (such as the Student t distributions) but the Gaussian distribution, the n univariate GARCH processes cannot be estimated separately anymore, because these different components interact in the log-likelihood. Yet, the DCC part of the model can still be estimated separately. In Section 6.2.4, we have seen that the log-likelihood of elliptical distributions only involves $(r_t - \mu_t)' \Sigma_t (r_t - \mu_t)$, that can be rewritten as $z_t' z_t$, therefore justifying the preliminary estimation of the covariance matrix dynamic.
- For other distributions (including the skewed t distribution), the estimation of the full model has to be performed in one step. Even for moderate-scale portfolios, such an estimation would be simply unmanageable.

An additional issue has to be addressed in the context of non-Gaussian distributions. Once the multivariate GARCH model is estimated, we need to compute the quantile of the distribution of the portfolio. The difficulty is that, in general, the distribution of the aggregate return cannot be deduced from the multivariate distribution of asset returns. The exception is once again the elliptical distribution family. In this case, the aggregate VaR is computed in exactly the same way as for the Gaussian case. Assume for instance that asset returns are distributed as a multivariate Student t distribution (as defined in (6.15) in Section 6.2.1) with ν degrees of freedom. The aggregate VaR is therefore given by

$$VaR_{\theta,t} = -(\mu_{p,t}(1) + \tilde{q}_\theta \times \sigma_{p,t}(1)),$$

where $\tilde{q}_\theta = t_\nu^{-1}(\theta)$ is the quantile for a probability of loss equal to θ from the univariate Student t distribution with ν degrees of freedom. The ES is given by (8.5).

In the non-elliptical cases, no analytical solution for the distribution of the portfolio return is available. Therefore, we have to turn to alternative techniques, such as numerical integration or Monte Carlo simulation.

First, numerical integration would not be possible in most applications of interest in VaR computation. The reason is that, even for moderate-scale portfolios, the computation burden is excessive. The difficulty is accentuated by the fact that the part to be integrated lies in the lower tail of the distribution.

Consequently, Monte Carlo simulation seems to be the only promising way of evaluating the aggregate VaR when asset returns are modeled through a multivariate distribution whose inverse *cdf* is not known analytically. Giot and Laurent (2003) adopted such a simulation-based approach for the estimation of the VaR of a portfolio with asset returns distributed as a multivariate skewed t distribution. To illustrate how this approach works, we assume that the distribution of the innovation process is $F(z|\eta)$ where η is the vector of shape parameters. The general procedure is the following for computing the aggregate VaR between dates t and $t+1$:

1. At simulation j , simulate a sample of innovations $z^j = (z_1^j, \dots, z_n^j)'$, drawn from the multivariate distribution $F(z|\hat{\eta})$. Deduce the sample of

- asset returns $r_{t+1}^j = \mu_t(1) + \Sigma_t(1)^{-1/2} z^j$. Compute the implied portfolio return $r_{p,t+1}^j = \alpha_t' r_{t+1}^j$.
2. Iterate step 1 for $j = 1, \dots, J$, where J should be large enough to provide accurate estimates of the desired quantile.
 3. Sort the sample $\{r_{p,t+1}^j\}_{j=1}^J$ in increasing order and compute the desired θ -quantile as the $[\theta J]$ -th value of the sample.

Needless to say that, for large-scale portfolios, the use of such Monte-Carlo simulations at a daily basis would be quite heavy.

Dealing with large-scale portfolios

Some multivariate GARCH models are well designed for large-scale portfolios. Some of them have been described in Section 6.1.2. They include Factor GARCH models (and their generalizations, such as Orthogonal GARCH models) and the Flexible GARCH model. The former approach relies on reducing the dimensionality of the problem by selecting a reasonably small number of factors to which the multivariate GARCH model gets adjusted. The latter approach decentralizes the estimation task, by estimating the dynamic covariance matrix using univariate and bivariate GARCH models only.

It should be emphasized that these models have been designed in a Gaussian context. For the Flexible GARCH model, it is not clear how it may be extended to a non-Gaussian distribution. Ledoit, Santa-Clara, and Wolf (2003) suggest a trick to circumvent this difficulty. Once the large-dimensional covariance matrix Σ_t is estimated (as described in Section 6.1.2), it is used to estimate the conditional distribution of the portfolio return. They first estimate the variance of the portfolio return as $\hat{\sigma}_{p,t}^2 = \alpha_t' \hat{\Sigma}_t \alpha_t$. Then, they adjust a (standardized) Student t distribution to the standardized innovations evaluated as $\hat{z}_t = \alpha_t' (r_t - \hat{\mu}) / \hat{\sigma}_{p,t}$, where $\hat{\mu}$ is the sample mean of the vector of asset returns. They are then able to compute the quantile of the portfolio return.¹¹

Copula functions

It is clear from the discussion above that the main limitation of the use of the asset-level approach for computing VaR is the difficulty to deal with a multivariate non-elliptical distribution, in particular for the estimation of the complete model. This approach has recently benefited from the development of the copula approach (see Section 6.3). One definite advantage of this approach

¹¹ A comparison of multivariate GARCH models designed for large-scale portfolios has been performed by Ledoit, Santa-Clara, and Wolf (2003). They compare RiskMetrics, the diagonal BEKK model, the DCC model, and the Flexible GARCH models in terms of their ability to estimate the quantile of some empirical portfolios correctly. They do not find significant differences between the various models.

is that the estimation of the marginal distributions (typically, the univariate GARCH models) and the dependence structure can be performed separately. This is true whatever the (possibly different) marginal distributions adopted for the asset returns and whatever the dependence structure. For instance, it is possible to model the univariate distribution of each asset return using a GARCH model with skewed t innovations and then to join these various margins through a Student t copula or any other copula function. Of course, this makes the use of copula much less constraining than the use of multivariate distributions.

The design of a conditional copula in this context can be the following:

1. Each asset return $r_{i,t}$ has its own marginal model given by $r_{i,t} = \mu_{i,t} + \varepsilon_{i,t}$ with $\varepsilon_{i,t} = \sigma_{i,t} z_{i,t}$. Expected returns $\mu_{i,t}$ may be assumed to be constant, so that returns can be demeaned in an initial step. Volatility is modeled as a GARCH(1, 1) model $\sigma_{i,t}^2 = \omega_i + a_i \varepsilon_{i,t-1}^2 + b_i \sigma_{i,t-1}^2$. The standardized innovation $z_{i,t}$ is *iid* with zero mean and unit variance with distribution function $F_i(z_{i,t})$. For instance, it may be a skewed Student t distribution, denoted t_{ν_i, λ_i} , where λ_i denotes the asymmetry parameter.
2. The margin of each univariate distribution is given by $u_{i,t} = t_{\nu_i, \lambda_i}^{-1}(z_{i,t})$. Then, the copula that links the various margins is

$$H(z_{1,t}, \dots, z_{n,t}) = C(F_1(z_{1,t}), \dots, F_n(z_{n,t})).$$

Assuming, for instance, a Student t copula with ν degrees of freedom would yield

$$H(z_{1,t}, \dots, z_{n,t}) = T_{R,\nu} \left(t_{\nu_1, \lambda_1}^{-1}(z_{1,t}), \dots, t_{\nu_n, \lambda_n}^{-1}(z_{n,t}) \right),$$

where $T_{R,\nu}$ is the *cdf* of the multivariate Student t distribution

$$\begin{aligned} & T_{R,\nu}(u_{1,t}, \dots, u_{n,t}) \\ &= \int_{-\infty}^{u_{1,t}} \cdots \int_{-\infty}^{u_{n,t}} \frac{\Gamma(\frac{\nu+n}{2})}{\Gamma(\frac{\nu}{2}) \sqrt{(\pi\nu)^n |R|}} \left(1 + \frac{y' R^{-1} y}{\nu} \right)^{-\frac{\nu+n}{2}} dy, \end{aligned}$$

with R the (n, n) correlation matrix of $u_t = (u_{1,t}, \dots, u_{n,t})'$.

This example illustrates why the copula approach may appear so promising at first sight. The estimation of this model only requires the estimation of $(n+1)$ models (the univariate components and the dependence structure), with each time only a few number of parameters.

It should be noticed however that the great generality allowed by copula functions has also a cost. When we are ultimately interested in VaR computation, we need once again to compute the θ -quantile of the portfolio return distribution. The only available approach appears to be the Monte Carlo simulation. A similar procedure to the one presented in Section 8.4.3 may be adopted. A further cost is the difficulty to use copula functions for the case of more than two or three assets.

8.5 Non-linear models

Other recent research focuses on estimating the VaR of portfolios containing options or other positions with non-linear price behavior. VaR methods employing a linear approximation to the relation between instrument values and the underlying risk factors are unlikely to be robust when applied to non-linear portfolios. Britten-Jones and Schaefer (1999) proposed a VaR framework that is based on a second order “delta-gamma” approximation and recognizes the impact that this will have, not only on variance, but on the form of the distribution.

Let us now describe this approach. As seen in Section 8.1, the change in the value of the portfolio is $\Delta W_{t+1}(\alpha_t) = \alpha'_t \Delta p_t = \sum_{i=1}^n \alpha_{i,t} \Delta p_{i,t}$, where $\Delta p_{i,t}$ denotes the change in the value of asset i . Clearly, the change in the portfolio value is linear in the change of the asset prices. In some cases, however, the portfolio may include some derivatives that would introduce a non-linear relation between asset prices and portfolio value. To cope with this non-linear relation, two main approaches have been proposed: the “delta-only” method uses a linear approximation; the “delta-gamma” method involves a linear-quadratic approximation. We investigate these various approaches in turn. An alternative approach is based on a Monte Carlo simulation of a large number of market scenarios.

8.5.1 The “delta-only” method

We assume that the assets depend on a set of K risk factors. For instance, in a structured product such as an insured portfolio, containing the underlying asset and a put option, the risk factor would be the underlying asset. The first-order approximation of the value of the portfolio is given by

$$\begin{aligned}\Delta^\delta W_{t+1} &= \sum_{i=1}^n \alpha_{i,t} \frac{\partial p_i(f, t)}{\partial t} \Delta t + \sum_{i=1}^n \alpha_{i,t} \sum_{k=1}^K \frac{\partial p_i(f, t)}{\partial f_k} \Delta f_{k,t} \\ &= \mu_\delta + \sum_{k=1}^K \delta_k \Delta f_{k,t} = \mu_\delta + \frac{\partial W'}{\partial f} \Delta f_t,\end{aligned}$$

where μ_δ denotes the change in portfolio value due to time, and δ_k is the aggregate effect on the portfolio value of factor k . The last equation introduces notations to be used below. In some cases, this approach may be viewed as a valid approximation for evaluating the portfolio value. However, in general, as exemplified by Britten-Jones and Schaefer (1999), it may yield large approximation errors.

8.5.2 The “delta-gamma” method

If we incorporate the second-order effects of the risk factors, we then obtain

$$\Delta^\gamma W_{t+1} = \mu_\gamma + \frac{\partial W'}{\partial f} \Delta f_t + \frac{1}{2} \Delta f'_t \frac{\partial^2 W}{\partial f \partial f'} \Delta f_t.$$

If we define $\delta = \partial W / \partial f$ and $\Gamma = \partial^2 W / (\partial f \partial f')$, the $(K, 1)$ vector and (K, K) matrix of aggregate delta and gamma, respectively, we have

$$\Delta^\gamma W_{t+1} = \mu_c + \delta' \Delta f_t + \frac{1}{2} \Delta f'_t \Gamma \Delta f_t. \quad (8.6)$$

Completing the square in (8.6), we obtain

$$\Delta^\gamma W_{t+1} = \mu_c + \frac{1}{2} (\Delta f_t + \Gamma^{-1} \delta)' \Gamma (\Delta f_t + \Gamma^{-1} \delta),$$

with $\mu_c = \mu_\gamma - \frac{1}{2} \delta' \Gamma^{-1} \delta$.

Assume now that the factor vector is distributed as a multivariate normal distribution $\Delta f_t \sim \mathcal{N}(\mu_f, \Sigma_f)$. Then, we have that $(\Delta f_t + \Gamma^{-1} \delta) \sim \mathcal{N}(\mu_f + \Gamma^{-1} \delta, \Sigma_f)$ so that, defining $y_t = \Sigma^{-1/2} (\Delta f_t + \Gamma^{-1} \delta)$, we obtain $y_t \sim \mathcal{N}(\mu_f + \Gamma^{-1} \delta, I_K)$. If we define $A = \Sigma^{-1/2} \Gamma \Sigma^{-1/2}$, we have

$$\Delta^\gamma W_{t+1} = \mu_c + \frac{1}{2} y_t' A y_t,$$

which shows that the change in portfolio value is a linear combination of uncorrelated non-central χ^2_1 variables.

This result is useful for computing the VaR of such a portfolio. It is possible to evaluate the moment generating as well as the characteristic functions of $(\Delta^\gamma W_{t+1} - \mu_c)$. See for instance Johnson, Kotz, and Balakrishnan (1995, vol. 2, p. 447). The characteristic function can be numerically inverted to compute probabilities such as $\Pr[\Delta^\gamma W_{t+1} - \mu_c < x]$ and then to evaluate the VaR of this portfolio.¹²

Glasserman, Heidelberger, and Shahabuddin (2000) propose an extension of the delta-gamma method to the case where risk factors are assumed to be t distributed rather than normally distributed. In such case, the difficulty comes from the fact that, although uncorrelated, risk factors are not independent anymore.

8.6 Comparison of VaR models

Several contributions present and compare the main approaches adopted for computing VaR. They include the work of Hsieh (1993), van den Goorbergh and Vlaar (1999), Christoffersen, Hahn, and Inoue (2001), Giot and Laurent (2003), and Ledoit, Santa-Clara, and Wolf (2002). Many tools have been used to compare these various techniques. Various statistical methods for evaluating VaR models have been suggested by Kupiec (1995), Christoffersen (1998), Lopez (1998), and Engle and Manganelli (2004).

¹² Britten-Jones and Schaefer (1999) propose an alternative approach based on the approximation of the distribution of a sum of independent non-central χ^2 variables.

8.6.1 Evaluation of VaR models

A natural way to evaluate VaR models is the “*hit*” test developed by Christoffersen (1998) already described in Section 5.3.3. The test is designed to evaluate if a given model is able to provide interval forecast that have the same coverage as in the data. Here, the central object in the approach is the Hit_t variable, defined in the case of VaR evaluation as

$$Hit_{t+1} = \begin{cases} 1 & \text{if } r_{p,t+1} < VaR_{\theta,t}, \\ 0 & \text{if } r_{p,t+1} > VaR_{\theta,t}. \end{cases} \quad (8.7)$$

Lopez (1998) proposes a supplementary evaluation based on ad hoc loss functions. The loss function is specified as the cost of the various outcomes

$$C_{t+1} = \begin{cases} f(r_{p,t+1}, VaR_{\theta,t}), & \text{if } r_{p,t+1} < VaR_{\theta,t}, \\ g(r_{p,t+1}, VaR_{\theta,t}), & \text{if } r_{p,t+1} \geq VaR_{\theta,t}, \end{cases}$$

Since this is a cost function and because prevention of VaR exceedances is of paramount importance, $f(x, y) \geq g(x, y)$ for a given y . The best VaR model is the one that minimizes the total cost over the N last days, $\sum_{i=0}^{N-1} C_{t-i}$.

There are many ways to specify f and g depending on the concern of the decision maker. For instance, choosing $f(x, y) = 1$ and $g(x, y) = 0$, we obtain Christoffersen’s hit test. If the exceedances as well as the magnitude of the exceedances are of importance, we may choose

$$C_{t+1} = \begin{cases} 1 + (r_{p,t+1} - VaR_{\theta,t})^2, & \text{if } r_{p,t+1} < VaR_{\theta,t}, \\ 0, & \text{if } r_{p,t+1} \geq VaR_{\theta,t}. \end{cases}$$

For the bank that implements the VaR model and that has to set aside capital reserves, $g = 0$ is not appropriate because liquid assets do not provide good returns. So one cost function that will take into account the opportunity cost of money is

$$C_{t+1} = \begin{cases} |r_{p,t+1} - VaR_{\theta,t}|^\gamma, & \text{if } r_{p,t+1} < VaR_{\theta,t}, \\ |r_{p,t+1} - VaR_{\theta,t}| \times i, & \text{if } r_{p,t+1} \geq VaR_{\theta,t}, \end{cases}$$

where γ reflects the seriousness of large exception and i is a function of interest rate.

8.6.2 Comparison of methods

In this section, we present some guidelines on some empirical studies that appeared in the literature.

Van den Goorbergh and Vlaar (1999) compare several approaches for computing VaR, including historical simulation, unconditional EVT, and various GARCH models with normal (RiskMetrics) and t innovations. Using data

from the Dutch stock index and the Dow Jones, they show that (i) conditional methods (GARCH models) out-perform unconditional ones, suggesting that the main characteristic of returns for evaluating VaR is volatility clustering; (ii) using a conditional t distribution provides a better fit than a normal distribution, implying that capturing distribution fat-tailedness is also crucial for an accurate measure of VaR. The GARCH model with t innovations is the only model found to perform well for all the probabilities considered. Other techniques such as historical simulation and EVT tend to underestimate the actual VaR. Importantly, this empirical evidence suggests that the unconditional EVT approach is unable to capture the consequences of the time-variability of volatility.

McNeil and Frey (2000) compare the conditional EVT method with GARCH models that have either normal or t innovations. They show that the conditional EVT model provides more accurate estimates of the VaR than the GARCH with normal innovations. Their coverage test does not reject the two approaches that are able to capture both the volatility clustering and the fat-tailedness of the distribution, i.e., the GARCH–EVT method and the GARCH model with t innovations.

Giot and Laurent (2003) more specifically investigate the GARCH approach and consider several distributional assumptions. They highlight that, at least for some return series, the asymmetry of the distribution should be taken into account for capturing VaR. Indeed, they obtain for NASDAQ and Nikkei indices, that a GARCH model with t innovations fails to measure the VaR accurately. In contrast, the model with skewed t innovations performs very well. When extended to the multivariate set-up, the model with skewed t innovations provides very accurate measures as well.

Ledoit, Santa-Clara, and Wolf (2002) compare the performances of various techniques in the context of a large-scale portfolio. More precisely, working with a large number of asset returns, they compute the covariance matrix using different approaches and then compute the VaR assuming a t distribution for portfolio returns. They find that the various techniques (including the CCC, BEKK and Flexible GARCH models) perform broadly equally well in terms of unconditional coverage.

8.6.3 10-day VaR and scaling

It is well-known that the variance of a Gaussian variable follows a simple scaling law. Indeed, the Basel Committee, in its 1996 Amendment, states that it will accept a simple \sqrt{T} scaling of 1-day VaR for deriving the 10-day VaR required in calculating market risk related risk capital.

The stylized facts of financial market volatility and research findings have repeatedly shown that a 10-day VaR is not likely to be the same as $\sqrt{10} \times$ 1-day VaR. First, the dynamic of a stationary volatility process suggests that if the current level of volatility is higher than unconditional volatility, the subsequent daily volatility forecasts will decline and converge to unconditional

volatility, and vice versa for the case where the initial volatility is lower than the unconditional one. The rate of convergence depends on the degree of volatility persistence. In the case where initial volatility is higher than unconditional volatility, the scaling factor will be less than $\sqrt{10}$. In the case where initial volatility is lower than unconditional volatility, the scaling factor will be more than $\sqrt{10}$. In practice, due to volatility asymmetry and other predictive variables that might be included in the volatility model, it is always better to calculate $\hat{\sigma}_{t+1}^2, \hat{\sigma}_{t+2}^2, \dots, \hat{\sigma}_{t+10}^2$ separately. The 10-day VaR is then produced using the 10-day volatility estimate computed from the sum $\sum_{i=1}^{10} \hat{\sigma}_{t+i}^2$.

Second, financial asset returns are not normally distributed. Danielsson and de Vries (1997) show that the scaling parameter for quantile derived using the EVT method increases at the approximate rate of T^ξ , which is typically less than the square-root-of-time adjustment. For a typical value of $\xi (= 0.25)$, we have $T^\xi = 1.778$, which is less than $10^{0.5} (= 3.16)$. McNeil and Frey (2000) on the other hand dispute this finding and claim the exponent to be greater than 0.5. The scaling factor of $10^{0.5}$ produced far too many VaR violations in the back-test of five financial series, except for returns on gold. In view of the conflicting empirical findings, one possible solution is to build models using 10-day returns data. This again highlights the difficulty due to the inconsistency in the rule applying to VaR for calculating risk capital and the one applying to VaR for back-testing.

8.6.4 Illustration

We consider once again the four market indices, SP500, DAX, FT-SE, and Nikkei over the period from January 1980 to December 2004. We create a portfolio composed of the four indices with an equal weight of 25%. Then, we compute the 1% and 5% VaR using different approaches developed in the previous sections. We consider the historical simulation (based on subsamples of size $N = 500$), RiskMetrics (with $\lambda = 0.94$), the GARCH–EVT approach of McNeil and Frey (2000) (with u corresponding to the 10% lower tail), a GARCH(1,1) model with t innovations, and finally a GARCH(1,1) model with skewed t innovations.

Notice that the parameter estimates in the gpd for the GARCH–EVT approach are found to be equal to $\hat{\xi} = 0.0504$ (with a standard error of 0.0324) and $\hat{\psi} = 0.5929$ (with a standard error of 0.2993), suggesting that the lower tail of standardized residuals is not very fat and is actually quite close to the attraction domain of the Gumbel distribution. Given that the univariate series produce fat-tailed residuals, this result suggests that extreme risks diversify away in a portfolio. Table 8.2 reports the parameter estimates for the three GARCH(1,1) models estimated (with standard errors in parentheses). The QML estimation of the McNeil and Frey model assumes normality of innovations, whereas the two other GARCH(1,1) models assume t and skewed t innovations, respectively. We notice that estimates of the parameters pertain-

ing to the volatility evolution are not significantly affected by the change of conditional distribution. We also observe that the degree-of-freedom parameter ν is too small to be consistent with the normality assumption.¹³ Finally, the asymmetry parameter λ is strongly significant, suggesting that allowing fat tails without asymmetry would not be sufficient in order to evaluate the VaR accurately.

Table 8.3 reports information concerning the computation of the VaR and ES. First, since we consider conditional VaR and ES, for each day, we have a different estimate of these statistics. For this reason, we therefore present the average of the estimates obtained from the various methods. Second, the number of exceedances is the number of dates t when the observed return exceeds the theoretical VaR. The expected exceedance is given by θT where θ is the confidence level and T the number of observations in the sample. Then, we present the three tests developed by Christoffersen (1998) (see Section 8.6.1). They allow identifying where the possible rejection of the model comes from (p-values are in parentheses).

At the 1% confidence level, two methods perform very well: the GARCH–EVT method and the GARCH model with skewed t innovations. In both cases, the actual number of exceedances is very close to the expected number. The unconditional coverage is not rejected for the two methods, and the conditional coverage is not rejected for the GARCH–EVT method only. At the 5% confidence level, the only method able to satisfy the unconditional coverage test is the GARCH model with skewed t innovations. The independence of *Hits* is rejected for all methods.

To sum up, it appears that the conditional EVT method performs very well for very small confidence levels.¹⁴ In contrast, for larger confidence levels,

Table 8.2. Parameter estimates of GARCH(1, 1) models

	Normal distribution	Student t distribution	Skewed t distribution
ω	0.0122 (0.0017)	0.0084 (0.0017)	0.0085 (0.0017)
α	0.1159 (0.0084)	0.0917 (0.0098)	0.0904 (0.0095)
β	0.8661 (0.0089)	0.8948 (0.0109)	0.8960 (0.0105)
ν	—	8.0622 (0.7245)	8.1545 (0.7463)
λ	—	—	-0.0943 (0.0170)

¹³ Remember that the Student t distribution converges to a normal distribution when $\nu \rightarrow \infty$.

¹⁴ Unreported results indicate that it is the only approach to accurately estimate the VaR at the 0.5% level.

the GARCH model with skewed t innovations provides accurate estimates of the VaR, as well.

Table 8.3. *VaR computation for various methods*

	Var ES	Number of exceed. (avg)	Uncond. coverage	Indep. of <i>Hits</i>	Cond. coverage
1% conf. level	Exp.: 59				
Hist. simulation	2.093 2.609	83 (0.004)	8.462 (0.000)	37.409 (0.000)	45.871 (0.000)
RiskMetrics	1.649 1.952	119 (0.000)	46.857 (0.000)	17.878 (0.000)	64.735 (0.000)
GARCH-EVT	1.897 2.557	55 (0.565)	0.332 (0.109)	2.570 (0.234)	2.902
GARCH- t	1.768 2.256	71 (0.141)	2.170 (0.012)	6.384 (0.014)	8.554
GARCH-skewed t	1.866 2.393	59 (0.963)	0.002 (0.003)	8.941 (0.011)	8.943
5% conf. level	Exp.: 296				
Hist. simulation	1.192 1.689	357 (0.001)	12.105 (0.000)	83.593 (0.000)	95.698 (0.000)
RiskMetrics	1.157 1.518	381 (0.000)	23.166 (0.000)	33.818 (0.000)	56.984 (0.000)
GARCH-EVT	1.156 1.777	340 (0.012)	6.335 (0.005)	7.750 (0.001)	14.085
GARCH- t	1.125 1.590	368 (0.000)	16.761 (0.000)	13.703 (0.000)	30.464 (0.000)
GARCH-skewed t	1.167 1.667	325 (0.098)	2.740 (0.001)	11.921 (0.001)	14.661 (0.001)

Portfolio Allocation

In his seminal work, Markowitz (1952) describes how an investor should allocate her wealth when asset returns are normally distributed. In this context, the optimization problem reduces itself to a mere mean-variance analysis. However, when returns are non-normal, the mean-variance criterion may fail to select the optimum portfolio. Its relevance depends in fact on the preferences of the investor. If she only cares about mean and variance of her portfolio, nothing has to be changed in the non-normal case. In the general case, however (for instance, for a constant relative risk averse investor), higher moments of the return distribution have to be taken into account in the asset allocation problem. This issue has been raised by a number of authors going back to Rubinstein (1973) and Kraus and Litzenberger (1976). In this context, asset allocation is more demanding and sometimes simply intractable.

The consequences of the non-normality of returns have also been investigated in another closely related field, namely asset pricing models. We may mention, among others, Rubinstein (1973), Kraus and Litzenberger (1976), Friend and Westerfield (1980), Barone-Adesi (1985), Harvey and Siddique (2000), and Jurczenko and Maillet (2001, 2006).

In this chapter, we investigate the consequences of non-normality on the portfolio allocation process. In Section 9.1, we present how to adapt the standard mean-variance framework for allocating wealth in the case where asset returns are non-normal. We propose various ways to implement optimal strategies in such a context. In Section 9.2, we investigate the case where alternative measures of risk are considered, such as downside variance, VaR, or ES.

9.1 Portfolio allocation under non-normality

Several recent studies investigate how wealth should be allocated when returns are non-normal. Two approaches have been abundantly studied. The first approach is based on the direct maximization of the expected utility, under alternative distributional assumptions for asset returns. For instance,

Ramchand and Susmel (1998), or Ang and Bekaert (2002) consider the case of regime-switching distribution of returns. An advantage of this approach is that it provides an exact evaluation of the expected utility, so that the optimal portfolio is the actual solution of the initial problem. A drawback is that, in most applications, a cumbersome numerical integration has to be performed to maximize the expected utility. Consequently, most studies focus on a very small number of assets (two or three) to reduce the computational burden.

The second approach is based on an approximation of the optimization problem that does not require any numerical integration. Typically, this approximation involves moments of the portfolio-return distribution. Early work on introducing the effect of higher moments are by Samuelson (1970) and Kraus and Litzenberger (1976). The main difficulty with this approach is to define the way the higher moments affect the expected utility. Although various solutions have been proposed, we focus on the Taylor's expansion of the utility function, which naturally yields an expected utility that depends linearly on the higher moments of the portfolio return.

9.1.1 Direct maximization of expected utility

The optimization problem

We consider an investor who allocates her portfolio by maximizing the expected utility of the next-period wealth W_{t+1} . There are n assets the investor can buy and sell, with vector of gross returns $r_{t+1} = (r_{1,t+1}, \dots, r_{n,t+1})'$. We assume that there is no cost for short-selling. The beginning-of-period wealth is denoted W_t , set arbitrarily equal to one. The next-period wealth is given by $W_{t+1} = (\alpha'_t r_{t+1}) W_t$, where $\alpha_t = (\alpha_{1,t}, \dots, \alpha_{n,t})'$ is the vector of the fractions of wealth allocated to the assets, with the constraint that the portfolio weights, at time t , sum to one, i.e., $\alpha'_t e = 1$, where e is an $(n, 1)$ vector of ones.¹ Therefore, the gross return on the portfolio is given by

$$r_{p,t+1}(\alpha_t) = \alpha'_t r_{t+1}.$$

We assume that the investor has a utility function U that depends on the level of future wealth W_{t+1} . Formally, the optimal portfolio weights are obtained by maximizing the scaled expected utility

$$\begin{aligned} \alpha_t^* &= \arg \max_{\{\alpha\}} E[U(\alpha'_t r_{t+1})] \\ \text{s.t. } \alpha'_t e &= 1. \end{aligned} \tag{9.1}$$

The n first-order conditions (FOCs) of the optimization problem are

$$\frac{\partial E(U(W_{t+1}))}{\partial \alpha_t} = E\left[U^{(1)}(W_{t+1}) r_{t+1}\right] = 0,$$

¹ To simplify the exposition, we assume that there is no risk-free asset and no transaction cost.

where $U^{(j)}$ denotes the j th derivative of the utility function. Since portfolio weights are constrained to sum to one, the FOCs of the investor's problem reduce to the $(n - 1)$ constraints

$$E \left[U^{(1)}(W_{t+1}) \begin{pmatrix} r_{1,t+1} - r_{n,t+1} \\ \vdots \\ r_{n-1,t+1} - r_{n,t+1} \end{pmatrix} \right] = 0,$$

or, alternatively,

$$E \left[U^{(1)}(W_{t+1}) \lambda_{t+1} \right] = 0, \quad (9.2)$$

where λ_{t+1} is the vector of returns of assets 1 to $n - 1$ in excess of asset n . In some cases, the FOCs (9.2) may help obtaining a solution to the optimization problem. But in general it is more convenient to directly maximize the expected utility, rewritten as

$$E[U(W_{t+1})] = \int \cdots \int U(W_{t+1}) f(r_{t+1}) dr_{1,t+1} \cdots dr_{n,t+1}, \quad (9.3)$$

where $f(r_{t+1})$ denotes the joint *pdf* of the vector of returns at time $t + 1$.

The mean-variance case

In some cases, this multiple integration problem reduces to the mere mean-variance problem. Consider for instance the case where the utility function is chosen to be the exponential utility and where returns are assumed to be normal. The exponential utility function (or CARA, for Constant Absolute Risk Aversion) is defined by

$$U(W_{t+1}) = -\exp(-\lambda W_{t+1}), \quad (9.4)$$

where $\lambda \geq 0$ is the coefficient of absolute risk aversion. In addition, the $(n, 1)$ expected returns vector and the (n, n) covariance matrix for the risky assets are denoted $\mu = (\mu_{1,t+1}, \dots, \mu_{n,t+1})'$ and Σ_{t+1} , respectively. As it is well-known, if a random variable X has a normal $\mathcal{N}(a, b^2)$ distribution, then $E[\exp(X)] = \exp(a + \frac{1}{2}b^2)$. Therefore, the problem (9.1) can be rewritten as

$$\begin{aligned} E[-\exp(-\lambda W_{t+1})] &= -E[\exp(-\lambda(\alpha'_t r_{t+1}))] \\ &= -\exp\left(-\lambda\mu_{p,t+1} + \frac{1}{2}\lambda^2\sigma_{p,t+1}^2\right), \end{aligned}$$

where $\mu_{p,t+1} = \alpha'_t \mu_{t+1}$ is the expected gross return of the portfolio, and $\sigma_{p,t+1}^2 = \alpha'_t \Sigma_{t+1} \alpha_t$ is the variance of the portfolio return. Thus, maximizing $E[U(W_{t+1})]$ is equivalent to maximizing the expression $(\mu_{p,t+1} + \frac{1}{2}\lambda\sigma_{p,t+1}^2)$, which is the mean-variance objective function.

Tobin (1958) originally suggested that any distribution defined by two parameters would provide a preference over mean and variance, but counter-examples have been provided (such as the log-normal distribution). This result has been generalized by Owen and Rabinovitch (1983) and Chamberlain (1983), who showed that when returns are drawn from an elliptical distribution, the mean-variance approximation of the expected utility is exact for all utility functions. On the other side, some utility functions imply the mean-variance criterion for arbitrary return distributions. This is the case of the quadratic utility function: $U(W) = a_0 + a_1 W + a_2 W^2$. The reason is simply that the expected utility function $E[U(W)]$ only involves the mean and the variance of the distribution. It should be noticed however that the quadratic utility function has some unappealing properties. In particular, it implies increasing absolute risk aversion and satiation (see Lhabitant, 1998). However, if (i) the distribution of returns to a portfolio is asymmetric, (ii) the investor's utility function is of higher order than the quadratic, and (iii) the mean and variance do not completely determine the distribution, then the third or higher moments and the sign of their coefficients must be considered.

Numerical integration

Tauchen and Hussey (1991) provide a numerical solution to equations such as (9.3) by quadrature. An M -point quadrature rule for the function $h(u)$ where $u \in \mathbb{R}^n$ and *pdf* $f(u)$ is a set of points $\{u_i\}$, $i = 1, \dots, M$, and corresponding weights $\{w_i\}$ such that

$$\int_a^b h(u) f(u) du = \sum_{i=1}^M w_i h(u_i).$$

The choice of the abscissa u_i and the weights w_i , $i = 1, \dots, n$, depend only on the *pdf* f , but not on the function h to be integrated.

For a univariate quadrature, the Gauss rules are discrete approximations to f determined by the method of moments using moments up through $2M-1$. Gauss rules are close to minimum norm rules and possess several optimum properties (Davis and Rabinowitz, 1984). Several alternative rules have been developed depending on the choice of the abscissa and the weights (See Judd, 1998, Chapter 7).

Multi-dimensional quadrature is more demanding, because we have to evaluate the n -dimensional integral

$$\int_a^b h(u) f(u) du = \sum_{i_1=1}^M \cdots \sum_{i_n=1}^M w_{i_1}^1 \cdots w_{i_n}^n h(u_{i_1}^1, \dots, u_{i_n}^n).$$

An exception is the case where the *pdf* f can be factored into the product of n one-dimensional *pdf*, after an affine transformation of variables. A multivariate product rule can then be formed by combining a set of one-dimensional Gauss

rules. A product rule has $N = \prod_{j=1}^M J_j$ points, where J_j is the number of points used along the j th axis. For high-dimensional problems, Stroud (1971) advocates a toolkit of non-product rules that may be applied to a wide class of problems. In particular, the spherical Lobatto rule for integration against the multivariate normal distribution only requires $N = 2^{M+1} - 1$ points, while integrating exactly all polynomials of degree five or less.

In a univariate context, Balduzzi and Lynch (1999) and Campbell and Viceira (1999) note that as few as five quadrature points suffice. As put forward by Ang and Bekaert (2002), using quadrature yields very accurate approximations for *iid* normally distributed returns. They use a Markov-switching model, in which returns are Gaussian within each regime. For some well-known joint distributions (such as the normal or Student t distributions), quadrature points and weights are directly available (see, for instance, Judd, 1998). In other cases, the integration is much more involved (Aït-Sahalia and Brandt, 2001).

9.1.2 An approximate solution based on moments

Instead of solving the general problem, which is sometimes simply intractable, we may focus on an approximation of this problem based on higher moments. Early work on introducing higher moments in the asset-allocation problem is by Samuelson (1970) and Kraus and Litzenberger (1976). Lai (1991), Chunachinda et al. (1997), and Prakash, Chang, and Pactwa (2003) applied the polynomial goal programming (PGP) approach to the portfolio selection with skewness. These studies provide evidence that incorporating skewness into the portfolio decision causes major changes in the optimal portfolio.² An alternative way of dealing with higher moments for asset-allocation purpose is the use of the Taylor series expansion to derive an approximation of the expected utility function. Although this approach has long been used in empirical applications to test the CAPM with higher moments, very few studies have considered the asset allocation problem using Taylor series expansions. Recent contributions are by Harvey et al. (2002) and Guidolin and Timmermann (2003). The first study proposed using Bayesian techniques to determine the optimal asset allocation when returns are driven by a skew normal distribution. The second paper investigated how the approximation of the expected utility by a Taylor series expansion can be implemented in the context of returns driven by a Markov-switching model with conditionally normal innovations. This approach has been also adopted by Jondeau and Rockinger

² However, this approach has a shortcoming; the allocation problem solved in the PGP approach cannot be precisely related to an approximation of the expected utility. In particular, the choice of the parameters used to weigh moment objectives is not related to the utility function parameters. Consequently, no measure of the quality of the approximation can be inferred from such an exercise. Another drawback is that there is no natural extension to an optimization problem that includes moments beyond the third one.

(2006a, 2005). Jondeau and Rockinger (2006a) study, in an unconditional framework, the cost of adopting a mean-variance criterion, while the return series are actually non-normal. Jondeau and Rockinger (2005) extend this analysis to a conditional framework. In both cases, the asset allocation under non-normality is based on a Taylor's approximation of the utility function up to the fourth order.

Approximation of the expected utility

Since we are primarily interested in measuring the effect of higher moments on the asset allocation, we now approximate the expected utility by a Taylor series expansion around the expected wealth.³ In this context, the utility function can be approximated by the following expression, when the expansion is performed up to the fourth order (to lighten notations, we omit the time index)

$$\begin{aligned} U(W) = & U(\bar{W}) + U^{(1)}(\bar{W})(W - \bar{W}) + \frac{1}{2}U^{(2)}(\bar{W})(W - \bar{W})^2 \\ & + \frac{1}{3!}U^{(3)}(\bar{W})(W - \bar{W})^3 + \frac{1}{4!}U^{(4)}(\bar{W})(W - \bar{W})^4 + \varepsilon, \end{aligned}$$

where $\bar{W} = E[W]$ and ε is the Taylor's remainder. Then the expected utility is simply approximated by

$$\begin{aligned} E[U(W)] \approx & U(\bar{W}) + \frac{1}{2}U^{(2)}(\bar{W})\sigma^2(W) \\ & + \frac{1}{3!}U^{(3)}(\bar{W})s^3(W) + \frac{1}{4!}U^{(4)}(\bar{W})k^4(W), \end{aligned} \tag{9.5}$$

where $\sigma^2(W)$, $s^3(W)$, and $k^4(W)$ stand for $E[W - \bar{W}]^j$ for $j = 1, \dots, 4$, respectively.⁴

In the case of the CARA utility function (9.4), the approximation for the expected utility is given by⁵

$$E[U(W)] \approx -\exp(-\lambda\bar{W}) \left[1 + \frac{\lambda^2}{2}\sigma^2[W] - \frac{\lambda^3}{6}s^3[W] + \frac{\lambda^4}{24}k^4[W] \right],$$

or, in terms of moments of the portfolio return

³ For some return distributions (such as the Gaussian one) or for some utility functions (such as polynomial utility), the Taylor series expansion is exact, i.e., the remainder is equal to zero.

⁴ These definitions of skewness and kurtosis as central higher moments differ from the statistical definitions as standardized central higher moments: $E[((W - \bar{W})/\sigma[W])^j]$ for $j = 3$ and 4.

⁵ As shown by Loistl (1976), in the case of the CARA utility function, the infinite Taylor series expansion converges to the expected utility function for all possible values of wealth W .

$$E[U(W)] \approx -\exp(-\lambda\mu_p) \left[1 + \frac{\lambda^2}{2}\sigma_p^2 - \frac{\lambda^3}{6}s_p^3 + \frac{\lambda^4}{24}k_p^4 \right], \quad (9.6)$$

where μ_p , σ_p^2 , s_p^3 , and k_p^4 denote the expected return, variance, skewness, and kurtosis of the portfolio return.

After some obvious simplifications, the FOCs can be defined respectively as

$$(\mu - r_f) = \frac{\frac{\lambda}{2}\frac{\partial\sigma_p^2}{\partial\alpha} - \frac{\lambda^2}{6}\frac{\partial s_p^3}{\partial\alpha} + \frac{\lambda^3}{24}\frac{\partial k_p^4}{\partial\alpha}}{1 + \frac{\lambda^2}{2}\sigma_p^2 - \frac{\lambda^3}{6}s_p^3 + \frac{\lambda^4}{24}k_p^4}. \quad (9.7)$$

Optimal portfolio weights can be obtained alternatively by maximizing expression (9.6) or by solving equalities (9.7). Inspection of relation (9.7) reveals that computing this expression would be rather simple if the variance, skewness, and kurtosis of the portfolio return and the derivatives thereof were known. We will show that these expressions can be computed relatively easily.

We may also consider the power utility function (or CRRA, for Constant Relative Risk Aversion), defined as

$$U(W) = \begin{cases} \frac{W^{1-\gamma}}{1-\gamma}, & \text{if } \gamma > 1, \\ \log(W), & \text{if } \gamma = 1, \end{cases} \quad (9.8)$$

where γ measures the investor's constant relative risk aversion. Contrary to the CARA utility, the CRRA does not converge towards an asymptote for increasing wealth. Consequently, marginal utility of wealth is positive even for large levels of wealth.

In the case of the CRRA utility function, using expression (9.5), we obtain the following approximation for the expected utility⁶

$$\begin{aligned} E[U(W)] &\approx \frac{\bar{W}^{1-\gamma}}{1-\gamma} - \frac{\gamma}{2}\bar{W}^{-\gamma-1}\sigma^2(W) \\ &\quad + \frac{\gamma(\gamma+1)}{6}\bar{W}^{-\gamma-2}s^3(W) \\ &\quad - \frac{\gamma(\gamma+1)(\gamma+2)}{24}\bar{W}^{-\gamma-3}k^4(W), \end{aligned} \quad (9.9)$$

or, in terms of the portfolio return

$$E[U(W)] \approx \frac{\mu_p^{1-\gamma}}{1-\gamma} \left[1 - \frac{\gamma}{2\mu_p^2}\sigma_p^2 + \frac{\gamma(\gamma+1)}{6\mu_p^3}s_p^3 - \frac{\gamma(\gamma+1)(\gamma+2)}{24\mu_p^4}k_p^4 \right]. \quad (9.10)$$

The FOCs for the CRRA function can be rewritten, after simplification

$$(\mu - r_f) = \frac{\frac{\gamma}{2(1+\mu_p)}\frac{\partial\sigma_p^2}{\partial\alpha} - \frac{\gamma(\gamma+1)}{6(1+\mu_p)^2}\frac{\partial s_p^3}{\partial\alpha} + \frac{\gamma(\gamma+1)(\gamma+2)}{24(1+\mu_p)^3}\frac{\partial k_p^4}{\partial\alpha}}{1 + \frac{\gamma(\gamma+1)}{2(1+\mu_p)^2}\sigma_p^2 - \frac{\gamma(\gamma+1)(\gamma+2)}{6(1+\mu_p)^3}s_p^3 + \frac{\gamma(\gamma+1)(\gamma+2)(\gamma+3)}{24(1+\mu_p)^4}k_p^4}. \quad (9.11)$$

⁶ The infinite Taylor series expansion converges to the expected utility function for levels of wealth ranging between 0 and $2\bar{W}$ (Loistl, 1976).

Optimal portfolio weights can be obtained alternatively by maximizing expression (9.9) or by solving equalities (9.11).

Investor's preferences for moments

In the case of the CARA and CRRA utility functions, the weight of skewness and higher moments in the approximated expected utility depends on the risk-aversion parameter. In particular, there is a preference for skewness and an aversion for kurtosis. However, generally speaking, the weight of the j th moment is related primarily to the j th derivative of the utility function.

As outlined by Brockett and Kahane (1992), and Brockett and Garven (1998), we are unlikely to find, on an a priori ground, the sign of sensitivities of the expected utility function to skewness and higher moments. Indeed, when moments are not orthogonal one to the others, the effect of increasing one of them might not be clear.⁷ Therefore, additional utility assessments are required to obtain a consistent preference ordering.

The usual risk averse investor is assumed to have a utility function with the following first two derivatives

$$U^{(1)}(W) > 0, \quad \forall W, \tag{9.12}$$

$$U^{(2)}(W) < 0, \quad \forall W. \tag{9.13}$$

The first condition means that agents have a positive marginal utility for wealth, i.e., non-satiety with respect to wealth. The second condition is that marginal utility decreases with wealth, implying risk aversion.

To investigate preferences with respect to higher moments, the following definition is useful.

Definition 9.1. *An investor who is consistent in preference direction for the n th moment has a utility function for which the n th derivative has the same sign whatever W . The investor is strictly consistent in preference direction for the n th moment if the n th derivative has a strict inequality with respect to zero.*

Then we have the following theorem.

Theorem 9.2 (Scott and Horvath, 1980). *Investors with positive marginal utility of wealth for all wealth levels (condition (9.12)), consistent risk aversion at all wealth levels (condition (9.13)), and strict consistency of moment preference, will have positive preference for positive skewness, i.e.,*

$$U^{(3)}(W) > 0, \quad \forall W. \tag{9.14}$$

⁷ For instance, Hlawitschka (1994) provides examples in which, even if the infinite expansion converges, adding more terms may worsen the approximation at a given truncation level. In contrast, Lhabitant (1998) describes an example in which omitted terms are of importance.

These results are consistent with the conclusions arrived at by Arditti (1967) and Kraus and Litzenberger (1976).

Theorem 9.3 (Scott and Horvath, 1980). *Consistent risk aversion (condition (9.13)), strict consistency of moment preference and positive preference for positive skewness (condition (9.14)) imply negative preference for kurtosis, i.e.,*

$$U^{(4)}(W) < 0, \quad \forall W.$$

More generally, assumptions of positive marginal utility, consistent risk aversion together with strict consistency of moment preference imply

$$\begin{aligned} U^{(n)}(W) &> 0, & \forall W &\quad \text{if } n \text{ is odd and} \\ U^{(n)}(W) &< 0, & \forall W &\quad \text{if } n \text{ is even.} \end{aligned}$$

Further discussion on the conditions that yield such moment preferences or aversions may be found in Pratt and Zeckhauser (1987), Kimball (1993), and Dittmar (2002).

A related issue is the order at which the Taylor series expansion should be truncated. Bansal, Hsieh, and Viswanathan (1993) let the data determine the point of truncation. This approach, however, implies a loss of power and a risk of overfitting the data. A more powerful alternative is to allow preference theory to guide the truncation. In this framework, a preference for positive skewness is characteristic of a decreasing absolute risk aversion, meaning that a non-random reduction in wealth should increase an agent's sensitivity to a wide range of risks. However, Pratt and Zeckhauser (1987) have shown that decreasing absolute risk aversion does not rule out certain counter-intuitive risk-taking behavior. Kimball (1993) thus defines the concept of *standard risk aversion*, which means that any undesirable risk should increase an agent's sensitivity to independent risks whenever a non-random reduction in wealth would. Sufficient conditions for standard risk aversion are decreasing absolute risk aversion and decreasing absolute prudence. The later assumption means that $-U^{(3)}(W)/U^{(2)}(W)$ is decreasing in W , so that $U^{(4)}(W) < 0, \forall W$.

Computation of portfolio moments

As indicated above, in order to implement the FOC 9.7 or 9.11, it is necessary to compute the expression of μ_p , σ_p^2 , and so forth. In this section, we use results put forward by Athayde and Flôres (2004) who define moments as tensors.

The first step is to compute return co-moments. In Chapter 5, we already described some elements of the co-skewness and co-kurtosis matrices for some distribution functions. We do not consider here any particular distribution function. Rather, we assume that we have computed all moments and co-moments useful to the asset-allocation problem.

For an n -variable system, the dimension of the covariance matrix is (n, n) , but only $n(n + 1)/2$ elements have to be computed. Similarly, the co-skewness matrix has dimension (n, n, n) , but only $n(n + 1)(n + 2)/6$ elements have to be computed. Finally, the co-kurtosis matrix has dimension (n, n, n, n) , but only $n(n + 1)(n + 2)(n + 3)/24$ elements have to be computed.⁸ We define co-skewness and co-kurtosis between asset returns as

$$s_{ijk} = E[(r_i - \mu_i)(r_j - \mu_j)(r_k - \mu_k)]^{1/3},$$

and

$$k_{ijkl} = E[(r_i - \mu_i)(r_j - \mu_j)(r_k - \mu_k)(r_l - \mu_l)]^{1/4}.$$

We now wish to adopt the following notations for co-skewness and co-kurtosis matrices. As suggested by Athayde and Flôres (2004), we transform the co-skewness (n, n, n) matrix into a (n, n^2) matrix, simply by slicing each (n, n) layer and pasting them, in the same order, sideways. For instance, in the case of $n = 3$ assets, the resulting $(3, 9)$ co-skewness matrix is

$$\begin{aligned} M_3 &= \left[\begin{array}{ccc|ccc|ccc} s_{111} & s_{112} & s_{113} & s_{211} & s_{212} & s_{213} & s_{311} & s_{312} & s_{313} \\ s_{121} & s_{122} & s_{113} & s_{221} & s_{222} & s_{223} & s_{321} & s_{322} & s_{323} \\ s_{131} & s_{132} & s_{133} & s_{231} & s_{232} & s_{233} & s_{331} & s_{332} & s_{333} \end{array} \right] \\ &= [s_{1jk} \ s_{2jk} \ s_{3jk}]_{j,k=1,2,3}, \end{aligned}$$

where s_{1jk} is a short notation for the (n, n) matrix $(s_{1jk})_{j,k=1,2,3}$. This notation extends the one for the covariance matrix, which is denoted M_2 . Similarly, the $(3, 27)$ co-kurtosis matrix is

$$M_4 = [k_{11kl} \ k_{12kl} \ k_{13kl} \ | \cdots \ | k_{31kl} \ k_{32kl} \ k_{33kl}]_{k,l=1,2,3}.$$

We are now able to define portfolio moments in a very tractable way. For a given portfolio weight vector α , unconditional expected return, variance, skewness, and kurtosis of the portfolio are, respectively

$$\begin{aligned} \mu_p &= \alpha' \mu, \\ \sigma_p^2 &= \alpha' M_2 \alpha, \\ s_p^3 &= \alpha' M_3 (\alpha \otimes \alpha), \\ k_p^4 &= \alpha' M_4 (\alpha \otimes \alpha \otimes \alpha), \end{aligned}$$

where \otimes stands for the Kronecker product. Derivatives of portfolio moments with respect to α are therefore very easy to compute⁹

⁸ For $n = 5$, we have 15 different elements for the covariance matrix, 35 elements for the co-skewness matrix and 70 elements for the co-kurtosis matrix (whereas these matrices have 25, 125, and 625 elements, respectively).

⁹ More generally, if we denote $m_p = (\sigma_p^2, s_p^3, k_p^4, \dots)$ the vector of central portfolio moments, we have $m_{p,i}^i = \alpha' M_i (\alpha \otimes^i \alpha)$, where $\alpha \otimes^i \alpha = \alpha \otimes (\alpha \otimes^{i-1} \alpha)$ and $\alpha \otimes^1 \alpha = \alpha \otimes \alpha$. Subsequently, we obtain $\partial m_{p,i}^i / \partial \alpha = i M_i (\alpha \otimes^i \alpha)$.

$$\begin{aligned}\frac{\partial \sigma_p^2}{\partial \alpha} &= 2M_2\alpha, \\ \frac{\partial s_p^3}{\partial \alpha} &= 3M_3(\alpha \otimes \alpha), \\ \frac{\partial k_p^4}{\partial \alpha} &= 4M_4(\alpha \otimes \alpha \otimes \alpha).\end{aligned}$$

The optimal portfolio allocation

We now address the issue of how to solve the asset allocation, when higher moments are introduced in the optimization problem. Different approaches have been proposed in the literature. Some authors suggest a numerical solution to this problem, whereas others are able to propose an analytical solution. On one hand, Lai (1991) as well as Chunhachinda et al. (1997) use polynomial goal programming (PGP) to solve the portfolio selection with skewness. To solve this problem, one has to specify the investor's subjective judgments and relative preferences on objectives. On the other hand, Athayde and Flôres (2004) provide a quasi-analytic solution to the efficient portfolio problem by defining moments as tensors and then solving the optimization problem. The solution is highly non-linear in the vector of weights, but it can be very easily solved, using non-linear optimization routines.

Equation (9.11) can, thus, be rewritten as

$$\begin{aligned}(\mu - r_f) - \delta_1(\alpha)[M_2\alpha] + \delta_2(\alpha)[M_3(\alpha \otimes \alpha)] \\ - \delta_3(\alpha)[M_4(\alpha \otimes \alpha \otimes \alpha)] = 0,\end{aligned}\tag{9.15}$$

where δ_1 , δ_2 , and δ_3 are non-linear functions of α . These n equations can be easily solved numerically, using a standard optimization package. Such a specification has been explored by Jondeau and Rockinger (2006a, 2005). The difficulty in solving this problem is not of the same order as compared with problems involving numerical integration as in Aït-Sahalia and Brandt (2001) or Ang and Bekaert (2002). Here, a very accurate solution is obtained in a very few seconds, even in the case of a large number of assets.

It is also worth noting that we have described how to solve the asset-allocation problem when the four first moments are to be incorporated. At the same time, we may consider as special cases the situations where only mean and variance or mean, variance, and skewness are assumed to be important.

9.2 Portfolio allocation under downside risk

To allocate wealth under non-normal returns, another path can be taken. Instead of focusing on the whole distribution, which requires estimating a large number of moments, we can focus on the part that we are interested in. In asset allocation, we are concerned with the possibility of a large decrease in

asset prices. Such concern gave rise to the notion of *downside risk*. Several examples of downside risk can be presented, such as downside variance or ES. Then for an investor who is concerned about downside risk, the optimization can no longer be based on the usual mean-variance criterion. We thus have to design a portfolio allocation model that comes up with unidirectional (leftward) concern.

Essentially, two types of solution have been advocated. The first one still considers the mean-variance criterion but assumes an additional constraint supposed to account for downside risk, whereas the second approach relies on a direct minimization of the latter.

9.2.1 Definition

A first important issue, before entering the portfolio allocation under downside risk, is to define the class of downside risk measures. A very clear definition can be found in Berkelaar and Kouwenberg (2000). Downside risk is directly related to the so-called lower-partial moments (LPM) introduced by Bawa and Lindenberg (1977). The LPM of order γ for a given target θ is given by

$$R_\gamma(\theta) = E[(\max(\theta - r_p, 0))^\gamma] = \int_{-\infty}^{\theta} (\theta - r)^{\gamma} dF_r(x), \quad \gamma \geq 0,$$

where $F_r(x)$ is the *cdf* of the portfolio return r_p . Interestingly, this class of moments covers some well-known risk measures. When $\gamma = 0$, we observe that $R_0(\theta) = \int_{-\infty}^{\theta} dF_r(x)$, so that, if we denote by $\theta(q)$ the VaR for probability q , we have that $R_0(\theta(q)) = q$. In this case, θ represents the q th percentile. If $\gamma = 1$, $R_1(\theta)$ can be interpreted as the expected shortfall, since $ES_\theta = E[r_p - \theta | r_p \leq \theta]$. Also, when $\gamma = 2$, we obtain the so-called downside variance $DV_\theta = E[(r_p - \theta)^2 | r_p \leq \theta]$.

Now, two types of optimization programs are able to take downside risk into account.

9.2.2 Downside risk as an additional constraint

A first approach consists in maintaining the standard mean-variance criterion, but with an additional constraint such that the downside risk of the portfolio should not exceed a given level. For instance, Alexander and Baptista (2004) focus on the following optimization program

$$\begin{aligned} \max_{\{\alpha\}} \quad & r_p - \lambda \sigma_p^2, \\ \text{s.t.} \quad & ES_\theta \leq L, \end{aligned}$$

where λ denotes the risk aversion parameter and L a bound on the ES. They adopt a parametric approach to solve this program, assuming a specific distribution for asset returns. In particular, they show that for a multivariate normal (or Student t) distribution, the equality $ES_\theta = L$ is given by a line in the plane (σ_p, μ_p) . Consequently, the efficient frontier is likely to be restricted by this additional constraint, depending on the probability (or confidence level) θ and the bound L . In fact, the effect of such a constraint is ambiguous as compared with the baseline unconstrained mean-variance criterion, because the standard deviation of the portfolio may actually increase under some ES constraints. This may be the case in particular when the investor is highly risk-averse and the selected bound L is too small.

9.2.3 Downside risk as an optimization criterion

In the second approach, the utility function is directly defined in terms of downside risk, instead of variance. For instance, Berkelaar and Kouwenberg (2000) consider the case

$$\max_{\{\alpha\}} \quad r_p - \lambda R_\gamma(\theta).$$

When $\theta \rightarrow \infty$ and $\gamma = 2$, we get the mean-variance criterion. When, $\gamma = 1$, we get the mean-ES criterion, and so forth. This specification is close to the one adopted by Ahn et al. (1999) and Krokmal, Palmquist, and Uryasev (2002).

This approach has been analytically solved by Berkelaar and Kouwenberg (2000) under some rather restrictive assumptions concerning the dynamic of asset returns. Asset prices are supposed to be driven by a geometric Brownian motion with constant interest rate and volatility.

Krokmal, Palmquist, and Uryasev (2002) follow another route.¹⁰ First, they redefine the ES in terms of portfolio price as follows

$$\begin{aligned} \overline{ES}_{\theta,t} &= -E [W_{t+1}(w_t) | W_{t+1}(w_t) \leq -\overline{VaR}_\theta] = -\frac{1}{\theta} \int_{-\infty}^{-\overline{VaR}_\theta} x f_P(x) dx \\ &= \overline{VaR}_\theta - \frac{1}{\theta} \int_{-\infty}^{-\overline{VaR}_\theta} (x + \overline{VaR}_\theta) f_P(x) dx, \end{aligned}$$

where the VaR is, as before, assumed to be a positive number. Then, using scenarios (i.e., historical simulations), they obtain the following linear estimation

¹⁰ The brief description of their approach we present here is slightly different from theirs. In particular, they allow several additional constraints in their optimization program that we disregard to simplify the exposition.

$$\widehat{ES}_{\theta,t} = \overline{VaR}_\theta - \frac{1}{\theta J} \sum_{j=1}^J \pi_j (x_j + \overline{VaR}_\theta),$$

where π_j , $j = 1, \dots, J$, is the probability of scenario j and $\{x_j\}_{j=1}^J$ is a set of simulated prices drawn from the empirical distribution. Then, they define the optimization problem as

$$\begin{aligned} & \max_{\{\alpha, \overline{VaR}_\theta\}} \quad \sum_{i=1}^n \alpha_i E_t [p_{i,t+1}], \\ \text{s.t.} \quad & \overline{VaR}_\theta - \frac{1}{\theta J} \sum_{j=1}^J \pi_j (x_j + \overline{VaR}_\theta) \leq \theta \sum_{i=1}^n \alpha_i p_{i,t}^0, \\ & \sum_{i=1}^n \alpha_i p_{i,t}^0 = \sum_{i=1}^n \alpha_i p_{i,t} + \text{transaction costs}. \end{aligned}$$

This problem is clearly linear programming. It provides us with the optimal portfolio weight vector α_t^* and the optimal VaR (\overline{VaR}_θ^*). The main advantage of this approach is that it is non-parametric, so that the ES is computed without relying on any distributional assumption.

It should be noticed that this approach also suffers from the same drawback as historical simulations do, i.e., it is fundamentally unconditional. Since the ES is computed using historical simulations, the estimated weights are obtained assuming that the empirical distribution will be still valid in the future. Evidently, if we think that volatility may vary over the investment period, this approach would probably be inappropriate.

A final approach that can be suggested in such context is based on a conditional (parametric) modeling of asset returns. Assume for instance that the portfolio return is drawn from a Student t distribution with degree-of-freedom parameter ν . Then, as seen in Section 8.4, the ES is given by

$$ES_{\theta,t} = \frac{1}{\theta} \frac{\nu - 2}{\nu - 1} \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi(\nu-2)} \Gamma(\frac{\nu}{2})} \left(1 + \frac{\tilde{q}_\theta^2}{\nu-2}\right)^{-\frac{\nu+1}{2}} \times \sigma_t(1) - \mu_t(1),$$

where $\tilde{q}_\theta = t_\nu^{-1}(\theta)$ is the quantile for a probability of loss equal to θ from the standardized t distribution, and $\mu_t(1)$ and $\sigma_t(1)$ are the expected return and volatility for date $t+1$. Then, the optimal portfolio under ES objective is obtained by solving the following maximization problem

$$\begin{aligned} & \max_{\{\alpha\}} \quad \sum_{i=1}^n \alpha_i E_t [r_{i,t+1}], \\ \text{s.t.} \quad & ES_{\theta,t} \leq L. \end{aligned}$$

Notice that this approach requires re-computing at each step j of the optimization algorithm the portfolio return as a function of the new weight vector $\alpha_t^{(j)}$. This can be done using the approach developed in Section 9.1.2.

Fundamentals of Option Pricing

In this part, we illustrate the techniques for pricing options and extracting information from traded option prices. We also describe various ways in which this information has been used in a number of applications. When dealing with options, we inevitably encounter the Black-Scholes-Merton option pricing formula, which has revolutionized the way in which options are priced in modern time. In Chapter 10, we describe in detail the seminal work of Black and Scholes (1973) and Merton (1973) (BSM, thereafter) on pricing European style options. BSM assumes that stock price follows a geometric Brownian motion, which implies that the terminal stock price has a lognormal distribution. Through hedging arguments, BSM shows that the terminal stock price distribution needed for pricing option can be stated without reference to the preference parameter and to the growth rate of the stock. This is now known as the risk-neutral approach to option pricing. The terminal stock price distribution, for the purpose of pricing options, is now known as the state-price density or the risk-neutral density in contrast to the actual stock price distribution, which is sometimes referred to as the physical, objective, or historical distribution.

In the BSM model, the objective and the risk-neutral densities are both lognormal. The risk-neutral density involves only one parameter that requires estimation, namely the so-called implied volatility σ . In Chapter 11, we describe how σ may be extracted from option price data. There exists huge empirical evidence on volatility smile and surface that suggests that σ is neither constant across exercise prices nor across option maturities. This evidence implies that the shape of the BSM assumption of a lognormal stock price distribution does not hold in practice. The BSM anomaly has led to voluminous research. This kind of work typically bypasses the objective distribution and pricing/hedging mechanism and could be described as an empirical data mining process for finding the risk-neutral distribution that best fits the option prices. This type of “non-structural” work is described in this chapter as well. We provide illustrations that some of the non-structural approaches may work rather well in capturing the actual risk-neutral density.

In Chapter 12, we turn to more structural approaches of option pricing. A clear limitation of the BSM model is the assumption that volatility is constant across exercise prices and maturities. Extension to a time-varying volatility is the next step on the research agenda. Models with stochastic volatility have been proposed by Hull and White (1987, 1988) and Heston (1994). Then, models with jumps have been developed by Merton (1976) and Ball and Torous (1983). Further generalization has been proposed with the development of the so-called Lévy processes. Such a process allows for an infinite amount of jumps of very small size over finite time intervals and, thus, represents an interesting alternative to the usual processes.

10.1 Notations

This chapter goes through the detailed proofs of the option pricing model developed by Black and Scholes (1973) and Merton (1973). The BSM model has already been described in many textbooks. The reason for going over it here is two fold. First, we want to present the model to allow for a fluid and, in terms of notation, coherent presentation of the subject. Also, many textbooks derive the so-called fundamental partial differential equations but do not solve this equation. This equation may, however, be solved without exaggerated effort by students and researchers who are not mathematicians. Such a reader should get familiar with the content of Chapter 13 before reading this one. The second purpose of this chapter is to use the BSM model, a well-known example, in order to introduce some important mathematical concepts crucial for later understanding more advanced topics, such as stochastic volatility and Lévy option pricing models. Both models are important developments in addressing the non-Gaussian feature of the underlying asset distribution.

There are several ways of proving the BSM model. Here, we focus on the no-arbitrage approach and the martingale measure approach, as they are useful in understanding the more advanced option pricing models. Other approaches such as the binomial and the pricing kernel approach are pedagogical and useful for understanding the fundamentals of option pricing. As they are not directly related to the later chapters, we would refer the readers to other textbooks on option pricing for these methods.

Options are financial contracts that depend on certain events. Options involve a buyer and a seller, i.e., two *counterparts*. A trader who buys an option is said to have a *long position*; symmetrically, if he sells an option he is said to have a *short position*. Options are different because of the rights that they confer to their owner. Let us start with some general definitions, then we will comment on these definitions.

Definition 10.1. *The purchase of a European call option gives its owner the right to purchase at a given date, but not before, a given asset for a pre-determined price. The purchase of a European put option gives its owner the*

right to sell at a given date, but not before, a given asset for a pre-determined price. The date, when the option can be exercised, is called exercise date. The pre-determined price is called exercise price.

Various words in this definition are worth emphasizing. For instance, the owner of the option has all the initiatives. She will obviously exercise the option if it is in her interest. Symmetrically, the seller of the option does not have the possibility of refusing the exercise of an option. A *European option*, as stated in the definition, allows only exercise of the option at a given moment in time. The so-called *American option* allows exercise at any time between the purchase of the option and a given date. This date will be called *expiration date*. Options such as *Bermudian options* and *mid-Atlantic options* can be exercised at several moments between purchase of the option and a given expiration date. The asset that may be purchased or sold at exercise is called *underlying asset*. For the option valuation models considered in this book, it is assumed that the underlying asset may always get traded, i.e., purchased or short sold, without transaction costs, in a market called the *spot market*. Additional terms are sometimes used. For instance, the exercise price is also called *strike price*. The time before an option expires is called *tenor*.

To understand when an option gets exercised or not, let us introduce some notations. Let S_t denote the price of the underlying asset at time t , traded in the spot market. We denote by T the expiration date of the option, and $\tau = T - t$ the tenor of the contract. Also, let K denote the strike price. Consider now a European call option. If the price of the underlying asset at expiration, S_T , is smaller than K , it means that the owner of the option can purchase the underlying asset in the spot market for less than the strike price, K . In such a case, it would not be optimal for her to exercise the option. On the other hand, if the price of the underlying asset is higher than K , it is rational to exercise the option. The option owner only needs to pay K for something worth more. Notice that in many markets, the option owner does not actually need to pay K . She will receive a compensation payment of $S_T - K$ called cash-settlement. If the option owner wishes to actually acquire the underlying asset, she can purchase it in the spot market. Her actual cost at T will then be $S_T - (S_T - K) = K$.

To summarize. If $S_T \leq K$, the option owner does not exercise, thus at T the value of the option is 0. If $S_T > K$, the option is worth $S_T - K$. The value of the option at time T is called the *pay-off*. The mathematical formula for a European call is $\max(S_T - K, 0)$.¹

We may now turn to the European put option. Again, denote by K the strike price of the option, and by S_T the value of the underlying at time T . The owner of the put option will not exercise the option if S_T is larger than K . Indeed, if the owner did exercise, she would receive only K for the underlying, whereas she could get S_T for a sale in the spot market. Symmetrically, if S_T is smaller than K , the owner of the option should exercise, because she can

¹ We also find the notation $(S_T - K)^+$ for $\max(S_T - K, 0)$.

then sell the underlying asset for K , which is more than she would get in the spot market. In case the owner of the option does not own the underlying asset, she may have to purchase it first. This is the case for markets where actual delivery is expected. In many markets, she would get a compensation payment in cash.

Again, we may summarize. If $S_T < K$, the option is worth $K - S_T$. If $S_T \geq K$, then the option will expire valueless. Therefore, the *pay-off* of the European put option is given by the formula $\max(0, K - S_T)$.

In Figure 10.1, we represent the pay-off diagrams of the put and call options. Looking at these figures we understand why the seller of the option requires a compensation payment at the *writing of the option*, i.e., when the seller and buyer establish a contract. Indeed, in the best case for the seller, the option buyer lets the option expire. In the worse case, the seller has to make a compensation payment.

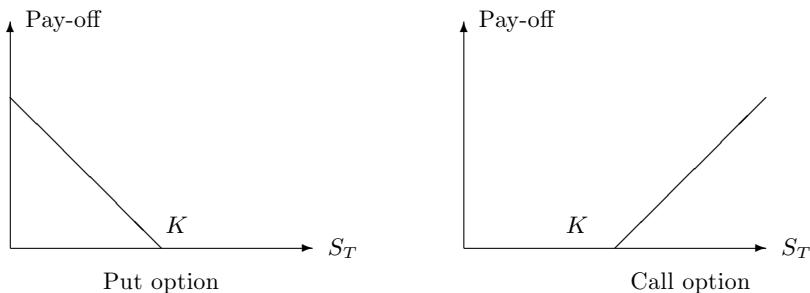


Fig. 10.1. Payoffs of put and call options.

An option for which the pay-off, at time T , is nil is said to expire *out-of-the-money*. An option for which the pay-off is positive is said to expire *in-the-money*. It is also possible to extend these notions to a moment where the option has not yet expired. For a European call (put) option, the ratio $S_t/(Ke^{-rt})$ where r is the risk-free interest rate, is called the *moneyness*. If this ratio is larger (smaller) than one, the option is said to be *in-the-money*. For the situation where it is smaller (larger) than one, the option is said to be *out-of-the-money*. Option traders also often use the words *at-the-money* for an option where the ratio is close to one, the notion of closeness being interpreted in a liberal manner.

So far, we spent much time on the European option. *American options* embed the right of *early exercise*. This means that they can be exercised between the purchase instant and some expiration date. Given that American options confer to their buyer an additional right, one may expect that the option price will be higher than the one of the European counterpart, i.e., of same strike and tenor. We may show that interesting moments, when an

American option might get exercised, are instants when a payment occurs. For a dividend paying asset this implies, that early exercise should be considered before the dividend payment. If the dividend payment is larger than the drop in the value of the option as the dividend gets paid, the option should be exercised.

Before concluding this introductory section, it should be noticed that for European options, the put and the call prices, denoted by C and P , where both options have same strike and tenor, and where the underlying does not pay dividends, are related by the so-called *put-call parity*

$$C - P = S_t - Ke^{-r\tau}.$$

The proof of this equality follows by contemplating the payoffs corresponding to the various symbols. The position $C - P$ corresponds to a purchase of a call and a sale of a put. The pay-off is $\max(S_T - K, 0) - \max(0, K - S_T)$. This equals $S_T - K$. If we contemplate the right-hand side of the equality, we see that S_t corresponds to the purchase price of the underlying asset. Purchase of the underlying will yield a value of S_T at T (that is a random variable, considered from time t). Also, $Ke^{-r\tau}$ is the discounted value of K . The right-hand side of the equality then represents the purchase of the underlying and a credit. At time T , this portfolio will be worth $S_T - K$. Given that the pay-off associated with the symbols on the left side corresponds to the one of the right-hand side, it follows that the value of the assets yielding the pay-off is also equal. Otherwise, an arbitrage opportunity would exist.

10.2 The no-arbitrage approach to option pricing

The no-arbitrage approach starts with a statistical model or stochastic process for the underlying asset price. From there, the dynamic of the derivative asset price is obtained, after which a risk-free portfolio involving the derivative asset and the underlying asset is created. As the portfolio is risk free, its instantaneous return must be equal to the risk-free rate of interest, otherwise an arbitrage opportunity would exist. From this condition, a fundamental partial differential equation (FPDE) is established for pricing all derivatives. Solving the FPDE under the given boundary conditions provides a solution, which may be closed form or numerical, for pricing options.

10.2.1 Choice of a stock price process

Figure 10.2 represents the evolution of the SP500 between January 1, 1980, and August 31, 2004. As mentioned before, the first step in the no-arbitrage approach to option pricing is to select an appropriate stock price dynamic in the form of a stochastic differential equation (or SDE, see Section 13.6) that yields trajectories similar to the one represented in Figure 10.2. We assume

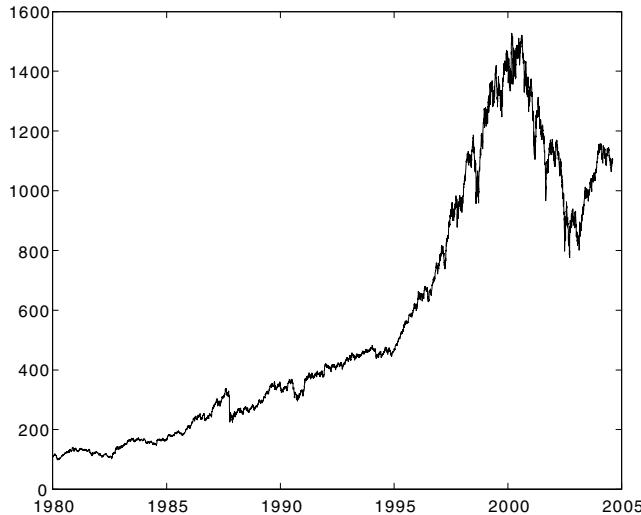


Fig. 10.2. Evolution of the SP500 daily return between January 1, 1980, and August 31, 2004.

for the moment that the stock price S_t behaves according to the following diffusion process

$$dS_t = \mu dt + \sigma dW_t, \quad (10.1)$$

where dS_t denotes the instantaneous price change, μ is the expected return, σ is the volatility of the price process, and W_t is an arithmetic Brownian motion such that $dW_t \sim \mathcal{N}(0, dt)$. The parameters μ and σ are assumed to be constant over time.

Is this a good assumption? The answer is no because integrating (10.1) shows that the price change between dates t and T is given by

$$S_T = S_t + \mu(T - t) + \sigma(W_T - W_t),$$

so that

$$S_T \sim \mathcal{N}(S_t + \mu(T - t), \sigma^2(T - t)).$$

This expression means that stock price can become negative and that the volatility is constant disregarding the level of the stock price. Both features are counterfactual. Companies have limited liability, so that shareholders can lose at most the value of the share. Constant volatility means that cheaper stock varies more relatively to expensive stock. To elaborate the second point, we simulated two trajectories of S_t with $\mu = 0$ and $\sigma = 1$. One graph starts at $S_0 = 30$, whereas the other starts at $S_0 = 2000$. We notice that the trajectory

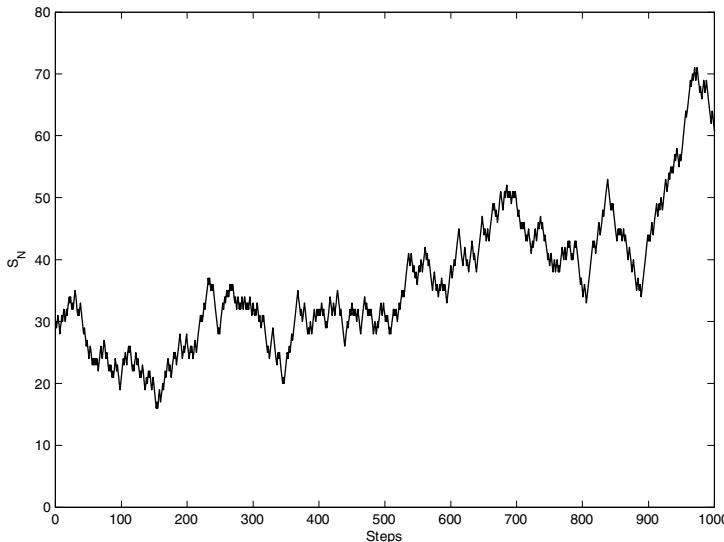


Fig. 10.3. Simulated prices – Model starts at $S_0 = 30$.

for the lower stock price has relatively more variations (Figure 10.3), whereas the one with the higher stock price is almost flat (Figure 10.4).

These observations suggest that the stock price dynamic in (10.1) is not appropriate. One alternative is to define $\log(S_t)$ as an arithmetic Brownian motion

$$d\log(S_t) = adt + \sigma dW_t.$$

Then, the dynamics of S_t is given by Ito's lemma. We obtain, setting $\mu = a - \frac{1}{2}\sigma^2$, that the stock price S_t follows a *geometric Brownian motion* with stochastic differential equation

$$dS_t = S_t \mu dt + S_t \sigma dW_t. \quad (10.2)$$

The advantage of such a representation is that the stock price, because of the log operator, must be greater than zero. In addition, stock returns, contrary to stock prices, have a constant volatility σ .

10.2.2 The fundamental partial differential equation

Now that we have a pricing dynamic for the underlying asset, the second step is to infer one for a derivative asset. We denote by $f(S_t, t)$ the price of a derivative asset and we introduce the notations

$$f_{SS} = \frac{\partial^2 f}{\partial S^2}, \quad f_S = \frac{\partial f}{\partial S}, \quad f_t = \frac{\partial f}{\partial t}.$$

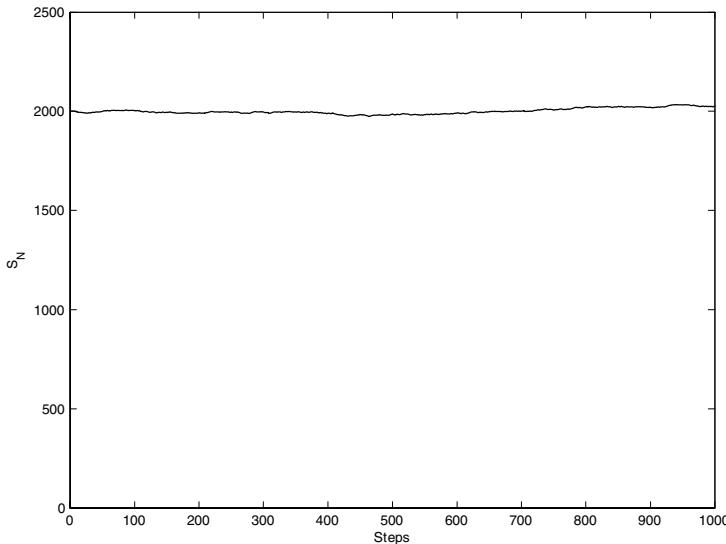


Fig. 10.4. Simulated prices – Model starts at $S_0 = 2000$.

To obtain the pricing dynamic of f , we apply Ito's lemma (see Section 13.7) on (10.2), yielding

$$df = \left(\frac{1}{2}\sigma^2 S_t^2 f_{SS} + \mu S_t f_S + f_t \right) dt + \sigma S_t f_S dW_t. \quad (10.3)$$

Next, we create a portfolio consisting of 1 unit of the derivative asset and a short position of f_S units in the underlying asset. The portfolio value V_t is

$$V_t = f - f_S S_t,$$

with pricing dynamic

$$dV_t = df - f_S dS_t.$$

Now substitute df from (10.3) and dS_t from (10.2) to get

$$\begin{aligned} dV_t &= \left(\frac{1}{2}\sigma^2 S_t^2 f_{SS} + \mu S_t f_S + f_t - \mu S_t f_S \right) dt + \sigma S_t f_S dW_t - \sigma S_t f_S dW_t \\ &= \left(\frac{1}{2}\sigma^2 S_t^2 f_{SS} + f_t \right) dt + 0dW_t. \end{aligned} \quad (10.4)$$

The dynamic of this portfolio is without risk because the term dW_t has a coefficient of 0. To avoid arbitrage, the instantaneous return of this portfolio must be the same as the risk-free rate of interest. Hence, the change in portfolio value is

$$dV_t = rV_t dt = r(f - f_S S_t) dt. \quad (10.5)$$

Equating (10.4) and (10.5), we get

$$\frac{1}{2}\sigma^2 S_t^2 f_{SS} + rS_t f_S + f_t = rf.$$

This is the Black-Scholes-Merton *fundamental partial differential equation*, FPDE. It governs the pricing of all derivatives written on S_t that has a pricing dynamic given by (10.2). The explicit solution of this fundamental PDE depends on the boundary conditions related to the terms of the derivative contract. For instance, consider a European call option with strike price K and time to maturity T . The terminal pay-off of this option is $\max(S_T - K, 0)$, as described in Figure 10.1. The terminal pay-off defines the boundary condition for solving the FPDE. Let $C(S_t, t)$ be the price of the call option. We can then write the FPDE for the European call option as

$$\frac{1}{2}\sigma^2 S_t^2 \frac{\partial^2 C}{\partial S^2} + rS_t \frac{\partial C}{\partial S} + \frac{\partial C}{\partial t} = rC. \quad (10.6)$$

with terminal condition $C(S_T, T) = \max(S_T - K, 0)$.

10.2.3 Solving the fundamental PDE

There are many fast and accurate numerical methods for solving PDEs. To explain them all is beyond the scope of this book. Here, we demonstrate the use of the Feynman-Kac formula to solve (10.6). A detailed exposition of this topic is provided in Sections 13.10 to 13.12. To work with the Feynman-Kac formula, we first need to obtain the transition probability of S_t given the stock price dynamic in (10.2). The SDE in (10.2) is that of a geometric Brownian motion, which by Ito's lemma (see Section 13.7) gives

$$d \log(S_t) = \left(\mu - \frac{1}{2}\sigma^2 \right) dt + \sigma dW_t. \quad (10.7)$$

Integrating (10.7) from t to T gives

$$\log(S_T) = \log(S_t) + \left(\mu - \frac{1}{2}\sigma^2 \right) (T - t) + \sigma (W_T - W_t).$$

We denote the horizon of the option $\tau = T - t$ to keep notations simple. Clearly, S_T has a *log-normal distribution*, which means that we can derive the conditional density for S_T from the normal density noting that $\log(S_T) \sim \mathcal{N}(\zeta, \sigma)$ where $\zeta = \log(S_t) + (\mu - \frac{1}{2}\sigma^2)\tau$. To see this, we start with the *cdf* $\Pr[S_T \leq x]$. Since the log transformation is a strictly increasing function, we have

$$\Pr[S_T \leq x] = \Pr \left[\frac{\log(S_T) - \zeta}{\sigma} < \frac{\log(x) - \zeta}{\sigma} \right],$$

and given that $(\log(S_T) - \zeta)/\sigma \sim \mathcal{N}(0, 1)$, we have

$$\begin{aligned}\Pr[S_T < x] &= \int_{-\infty}^{\frac{\log(x) - \zeta}{\sigma}} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(\log(S_T))^2\right] d\log(S_T) \\ &= \int_0^x \frac{1}{\sqrt{2\pi}\sigma S_T} \exp\left[-\frac{1}{2}\left(\frac{\log(S_T) - \zeta}{\sigma}\right)^2\right] dS_T.\end{aligned}\quad (10.8)$$

The conditional density of S_T is obtained by replacing the volatility parameter σ in (10.8) by $\sigma\sqrt{\tau}$, the volatility over the transition period from t to T

$$p(S_T, T | S_t, t) = \frac{1}{\sqrt{2\pi}\sigma\sqrt{\tau}S_T} \exp\left[-\frac{1}{2}\left(\frac{\log(S_T) - \zeta}{\sigma\sqrt{\tau}}\right)^2\right], \quad (10.9)$$

where

$$\zeta = \log(S_t) + \left(\mu - \frac{1}{2}\sigma^2\right)\tau.$$

From (10.9), we can compute partial differentials $\partial p/\partial S_T, \partial^2 p/\partial S_T^2$, and $\partial p/\partial t$, and prove that the transition probability of a geometric Brownian motion satisfies the PDE

$$\frac{1}{2}\sigma^2 S^2 \frac{\partial^2 p}{\partial S^2} + \mu S \frac{\partial p}{\partial S} + \frac{\partial p}{\partial t} = 0.$$

For a more elegant derivation, see Sections 13.10–13.12. Next we define $f(S_t, t)$ as the expected value of a smooth function $h(\cdot)$ defined on S_T as follows

$$f(S_t, t) \equiv E[h(S_T) | S_t, t] = \int_0^\infty h(S_T) p(S_T, T | S_t, t) dS_T. \quad (10.10)$$

Then we see that the Kolmogorov backward equation

$$\begin{aligned}\frac{1}{2}\sigma^2 S^2 \frac{\partial^2 f}{\partial S^2} + \mu S \frac{\partial f}{\partial S} + \frac{\partial f}{\partial t} &= \int_0^\infty h(S_T) \left(\frac{1}{2}\sigma^2 S^2 \frac{\partial^2 p}{\partial S^2} + \mu S \frac{\partial p}{\partial S} + \frac{\partial p}{\partial t}\right) dS_T \\ &= 0,\end{aligned}\quad (10.11)$$

is satisfied. Given that all the partial derivatives are smooth and the square variation is bounded, this result means that (10.10) is a unique solution for (10.11) under the boundary condition

$$\lim_{t \rightarrow T} f(S_t, t) = h(S_T),$$

because the transition probability collapses to a Dirac measure (see Section 13.11).

However, the FPDE in (10.6) that we want to solve comes in a slightly different form. First, it has an r instead of a mean drift μ as the coefficient for

C_S . Moreover, the PDE equals to rC instead of zero. The mean drift difference can be dealt with by selecting a slightly different SDE

$$dS_t = rS_t dt + \sigma S_t dW_t^*. \quad (10.12)$$

We put a * on the Brownian motion to indicate that it is a different SDE than before. The term rC can be dealt with by defining the present value of the call option contract as $C(S_t, t) \equiv e^{-r\tau} f(S_t, t)$ with $f(S_t, t) \equiv E^*[h(S_T) | S_t, t]$. Again we put * on the expectation operator, E , to remind ourselves that the expectation is computed under the new SDE in (10.12). From the definition of $C(S_t, t)$, we get

$$\begin{aligned} \frac{\partial C}{\partial t} &= re^{-r\tau} f(S_t, t) + e^{-r\tau} \frac{\partial f}{\partial t} = rC(S_t, t) + e^{-r\tau} \frac{\partial f}{\partial t} \\ \Rightarrow \frac{\partial f}{\partial t} &= e^{-r\tau} \left(\frac{\partial C}{\partial t} - rC(S_t, t) \right). \end{aligned}$$

Similarly, writing $f(S_t, t) = e^{r\tau} C(S_t, t)$ we get

$$\frac{\partial f}{\partial S_t} = e^{r\tau} C_S, \quad \text{and} \quad \frac{\partial^2 f}{\partial S_t^2} = e^{r\tau} C_{SS}.$$

If we substitute all the partial derivatives into the LHS of (10.11), we obtain

$$e^{r\tau} \left(\frac{1}{2} \sigma^2 S^2 C_{SS} + rSC_S + C_t - rC \right) = 0,$$

or

$$\frac{1}{2} \sigma^2 S^2 C_{SS} + rSC_S + C_t = rC.$$

It follows that $C(S_t, t) \equiv e^{-r\tau} E^*[h(S_T) | S_t, t]$ is a solution for the FPDE (10.6) under the boundary condition. To finish, it is necessary to compute the conditional expectation. We obtain

$$\begin{aligned} C(S_t, t) &= e^{-r\tau} E^* [\max(S_T - K, 0) | S_t, t] \\ &= e^{-r\tau} \int_K^\infty (S_T - K) \frac{1}{\sqrt{2\pi}\sigma\sqrt{\tau}S_T} \exp \left[-\frac{1}{2} \left(\frac{\log(S_T) - \zeta}{\sigma\sqrt{\tau}} \right)^2 \right] dS_T, \end{aligned} \quad (10.13)$$

where

$$\zeta = \log(S_t) + \left(r - \frac{1}{2}\sigma^2 \right) \tau.$$

10.2.4 The Black-Scholes-Merton formula

In this section, we derive the BSM formula from the European call expression given in (10.13) in the previous section. First, we separate the integral in $C(S_t, t)$ in (10.13) into $I_1 + I_2$, where

$$I_1 = e^{-r\tau} \int_K^\infty \frac{S_T}{\sqrt{2\pi}\sigma\sqrt{\tau}S_T} \exp\left[-\frac{1}{2}\left(\frac{\log(S_T) - \zeta}{\sigma\sqrt{\tau}}\right)^2\right] dS_T,$$

and

$$I_2 = -Ke^{-r\tau} \int_K^\infty \frac{1}{\sqrt{2\pi}\sigma\sqrt{\tau}S_T} \exp\left[-\frac{1}{2}\left(\frac{\log(S_T) - \zeta}{\sigma\sqrt{\tau}}\right)^2\right] dS_T.$$

Apply a change of variable with $y = \log(S_T)$, $dy = \frac{1}{S_T}dS_T$ and $e^y = S_T$, we get for I_1 , focusing on just a few steps

$$\begin{aligned} I_1 &= e^{-r\tau} \int_{\log(K)}^\infty \frac{e^y}{\sqrt{2\pi}\sigma\sqrt{\tau}} \exp\left[-\frac{(y - \zeta)^2}{2\sigma^2\tau}\right] dy \\ &= e^{-r\tau} \int_{\log(K)}^\infty \frac{e^{\zeta + \frac{1}{2}\sigma^2\tau}}{\sqrt{2\pi}\sigma\sqrt{\tau}} \exp\left[-\frac{[y - (\sigma^2\tau + \zeta)]^2}{2\sigma^2\tau}\right] dy \\ &= e^{-r\tau} e^{\zeta + \frac{1}{2}\sigma^2\tau} \Phi(d_1), \end{aligned}$$

where

$$e^{\zeta + \frac{1}{2}\sigma^2\tau} = \exp\left[\log(S_t) + \left(r - \frac{1}{2}\sigma^2\right)\tau + \frac{1}{2}\sigma^2\tau\right] = S_t e^{r\tau},$$

and

$$d_1 = \frac{\log(S_t/K) + (r + \frac{1}{2}\sigma^2)\tau}{\sigma\sqrt{\tau}}.$$

The second component is easier to solve. With the same change in variable, we get

$$\begin{aligned} I_2 &= -Ke^{-r\tau} \int_{\log(K)}^\infty \frac{1}{\sqrt{2\pi}\sigma\sqrt{\tau}} \exp\left(-\frac{1}{2}\left(\frac{y - \zeta}{\sigma\sqrt{\tau}}\right)^2\right) dy \\ &= -Ke^{-r\tau} \Phi(d_2), \end{aligned}$$

where

$$d_2 = \frac{\log(S_t/K) + (r - \frac{1}{2}\sigma^2)\tau}{\sigma\sqrt{\tau}}.$$

Hence, for the European call option, we have

$$C(S_t, t) = S_t \Phi(d_1) - Ke^{-r\tau} \Phi(d_2)$$

with

$$d_1 = \frac{\log(S_t/K) + (r + \frac{1}{2}\sigma^2)\tau}{\sigma\sqrt{\tau}},$$

$$d_2 = \frac{\log(S_t/K) + (r - \frac{1}{2}\sigma^2)\tau}{\sigma\sqrt{\tau}},$$

a formula that was rewarded by the Nobel prize in Economics.

10.3 Martingale measure and BSM formula

In Chapter 14, we provide details on martingale and the techniques to change probability measure. The Holy Grail in this changing measure approach to option pricing is to convert the asset price dynamic and, hence, the derivative price dynamic into the so-called *martingale measure*. Once this is achieved, today's option price can be calculated as the expected option pay-off under the risk-neutral measure and discounted at the risk-free rate. We present only the basic workings here. The fundamental theorems on asset pricing were given in two papers by Harrison and Pliska. In Harrison and Pliska (1981), it is shown that if a market has a martingale measure, there is no arbitrage opportunity. In Harrison and Pliska (1983), it is also shown that the martingale measure is unique if and only if every claim can be hedged, that is, exclusively if the market is complete. Recent contributions on this subject include Delbaen and Schachermayer (1994).

10.3.1 Self-financing strategies and portfolio construction

Assume that the risky as well as riskless assets have the following pricing dynamics

$$\begin{aligned} dS_t &= \mu S_t dt + \sigma S_t dW_t, \\ dB_t &= r B_t dt. \end{aligned}$$

After integration, we obtain

$$\begin{aligned} S_t &= S_0 \exp \left[\left(\mu t - \frac{1}{2} \sigma^2 t \right) + \sigma W_t \right], \\ B_t &= B_0 e^{rt}. \end{aligned} \tag{10.14}$$

Assume also that we can construct a portfolio by holding, at time t , α_t and θ_t units of the riskless and the risky assets, respectively. The portfolio weights, α_t and θ_t , change through time and they are often referred to as trading strategies. We may now write the portfolio value at time t as

$$V_t = \alpha_t B_t + \theta_t S_t. \tag{10.15}$$

Provided that the trading strategies are self-financing, i.e., there is no cash injected or withdrawn in the interim periods, then the portfolio value at time t can also be expressed as

$$V_t = V_0 + \int_0^t \alpha_s dB_s + \int_0^t \theta_s dS_s,$$

i.e., the portfolio value changes are entirely due to movements in S_t and B_t .

10.3.2 Change of numeraire

The first step in the martingale approach is to perform a change of numeraire by expressing the risky element in terms of the risk-free instrument, i.e., $S_t^\dagger = S_t/B_t$. We introduce a superscript \dagger to denote a variable rescaled by the numeraire. Hence, (10.15) becomes

$$V_t^\dagger = \alpha_t + \theta_t S_t^\dagger, \quad (10.16)$$

and similarly (10.14) becomes

$$S_t^\dagger = S_0^\dagger \exp \left[\left(\mu - r - \frac{1}{2} \sigma^2 \right) t + \sigma W_t \right],$$

which implies a transformed SDE for the risky asset as

$$dS_t^\dagger = (\mu - r) S_t^\dagger dt + \sigma S_t^\dagger dW_t. \quad (10.17)$$

10.3.3 Change of Brownian motion

The second step is to apply a change of probability measure (see Section 14.4) so that the SDE in (10.17) becomes a martingale with zero drift. By the transformation $dW_t = dW_t^* + \beta dt$, we get

$$dS_t^{\dagger*} = [(\mu - r) + \sigma\beta] S_t^{\dagger*} dt + \sigma S_t^{\dagger*} dW_t^*,$$

and for $dS_t^{\dagger*}$ to have zero drift,

$$(\mu - r) + \sigma\beta = 0,$$

or

$$\beta = -(\mu - r)/\sigma.$$

In this setting, the rescaled and transformed process $dS_t^{\dagger*} = \sigma S_t^{\dagger*} dW_t^*$, and $S_t^{\dagger*}$ is a martingale under the new Brownian motion, i.e.,

$$E_Q[S_s^{\dagger*} | S_t^{\dagger*}] = S_t^{\dagger*}, \text{ for } s > t,$$

where $E_Q[\cdot]$ is an alternative notation for $E^*[\cdot]$. From (10.16), we have

$$dV_t^{\dagger*} = \theta_t dS_t^{\dagger*} = \theta_t \sigma S_t^{\dagger*} dW_t^*.$$

It is clear that the rescaled and transformed process $dV_t^{\dagger*}$ is also a martingale under the new Brownian motion dW_t^* .

10.3.4 Evolution of S_t under Q

From (10.14), we note that

$$S_t = S_0 \exp \left(\mu t - \frac{1}{2} \sigma^2 t + \sigma W_t \right),$$

under the original P probability measure. Replacing W_t by $W_t^* + \beta t$ yields

$$\begin{aligned} S_t^* &= S_0 \exp \left[\left(\mu - \frac{1}{2} \sigma^2 \right) t + \sigma \left(W_t^* - \frac{\mu - r}{\sigma} t \right) \right] \\ &= S_0 \exp \left[\left(r - \frac{1}{2} \sigma^2 \right) t + \sigma W_t^* \right], \end{aligned} \quad (10.18)$$

under the new Q probability measure, and in terms of SDE

$$dS_t^* = r S_t^* dt + \sigma S_t^* dW_t^*.$$

The change of probability has changed the drift of the geometric Brownian motion and hence the SDE.

10.3.5 The expected pay-off as a martingale

Let us define a contingent claim whose pay-off is a smooth function $h(\cdot)$ of S_T . At time T , the portfolio value V_T is exactly the same as $h(S_T)$ by carefully applying a trading strategy. Since $V_t^{†*}$ is a martingale under dW_t^* , it means that

$$V_t^{†*} = E_Q \left[V_T^{†*} \middle| W_t^* \right] = E_Q \left[\frac{h(S_T)}{B_T} \middle| W_t^* \right],$$

and with constant interest rate,

$$V_t^* = e^{-r\tau} E_Q [h(S_T) | W_t^*]. \quad (10.19)$$

By substituting the diffusion parameters from (10.18) into (10.19), we get

$$V_t^* = e^{-r\tau} \int_{-\infty}^{\infty} \frac{h(S_T)}{\sqrt{2\pi}\sigma\sqrt{\tau}S_T} \exp \left[-\frac{1}{2} \left(\frac{\log(S_T) - \zeta}{\sigma\sqrt{\tau}} \right)^2 \right] dS_T, \quad (10.20)$$

where $\zeta = \log(S_T) + (r - \frac{1}{2}\sigma^2)\tau$.

In the case of a European call option, the pay-off function is

$$h(S_T) = \max(S_T - K, 0),$$

and the value of the European call option can be computed as

$$C(S_t, t) = e^{-r\tau} \int_K^{\infty} \frac{S_T - K}{\sqrt{2\pi}\sigma\sqrt{\tau}S_t} \exp \left[-\frac{1}{2} \left(\frac{\log(S_T) - \zeta}{\sigma\sqrt{\tau}} \right)^2 \right] dS_T, \quad (10.21)$$

with ζ defined as above. A detailed solution of (10.21) has already been provided in the previous section.

10.3.6 The trading strategies

We may invoke the martingale representation theorem, which states that there exists a process θ_t such that $V_t^{\dagger*}$ may be written as

$$V_t^{\dagger*} = V_0^{\dagger*} + \int_0^t \theta_t \sigma S_t^* dW_t^*,$$

or

$$dV_t^{\dagger*} = \theta_t \sigma S_t^* dW_t^*. \quad (10.22)$$

This result means that we can reach $V_t^{\dagger*}$ by affecting the right weights to the Brownian motion. For finance, given that $V_t^{\dagger*}$ represents the value of the derivative asset after a change of numeraire, this means that there exists θ_t , the weight to affect to the risky asset, so that the value of the duplication strategy equals the value of the derivative asset. Unfortunately, the martingale representation theorem is not a constructive proof and does not give us θ_t directly. For cases where $V_t^{\dagger*}$ is known, such as in (10.21), we may obtain θ_t explicitly.

To see this, we recognize that the information generated by S_t and S_t^* is equivalent. We consider therefore the dynamics of $dV_t^{\dagger*}(S_t, t)$. By applying Ito's lemma, we obtain under Q

$$dV_t^{\dagger*} = \left(\frac{1}{2} \sigma^2 S_t^2 \frac{\partial^2 V_t^{\dagger*}}{\partial S_t^2} + r S_t \frac{\partial V_t^{\dagger*}}{\partial S_t} + \frac{\partial V_t^{\dagger*}}{\partial t} \right) dt + \sigma S_t \frac{\partial V_t^{\dagger*}}{\partial S_t} dW_t^*. \quad (10.23)$$

Given that $V_t^{\dagger*}$ is a martingale, its drift must be zero, yielding the FPDE

$$\frac{1}{2} \sigma^2 S_t^2 \frac{\partial^2 V_t^{\dagger*}}{\partial S_t^2} + r S_t \frac{\partial V_t^{\dagger*}}{\partial S_t} + \frac{\partial V_t^{\dagger*}}{\partial t} = 0, \quad (10.24)$$

and, also,

$$dV_t^{\dagger*} = \sigma S_t \frac{\partial V_t^{\dagger*}}{\partial S_t} dW_t^*. \quad (10.25)$$

Equating (10.22) and (10.25), we get

$$\theta_t \sigma S_t^{\dagger*} = \sigma S_t^* \frac{\partial V_t^{\dagger*}}{\partial S_t} = \sigma S_t^{\dagger*} \frac{\partial V_t^*}{\partial S_t},$$

which means

$$\theta_t = \frac{\partial V_t^*}{\partial S_t}.$$

That is, the trading strategy involves investing $\partial V_t^*/\partial S_t$ amount of the risky asset at time t .

The amount to be invested in the risk-free asset, α_t , follows from the self-financing aspect of the trading strategy. Indeed, if the derivative asset is initially worth $V_0(S_0, 0)$, then we obtain by setting $B_0 = 1$

$$V_0(S_0, 0) = \alpha_0 + \theta_0 S_0,$$

or

$$\alpha_0 = V_0(S_0, 0) - \theta_0 S_0.$$

In practice, continuous trading is not possible, hence let t_i be times where the portfolio gets rebalanced. We have

$$\alpha_{t_{i-1}} B_{t_i} + \theta_{t_{i-1}} S_{t_i} = \alpha_{t_i} B_{t_i} + \theta_{t_i} S_{t_i},$$

or

$$\alpha_{t_i} = \alpha_{t_{i-1}} + (\theta_{t_{i-1}} - \theta_{t_i}) \frac{S_{t_i}}{B_{t_i}}.$$

We also notice that (10.24) corresponds exactly to the FPDE of (10.6). The martingale approach to option pricing therefore yields all the same results as in the previous section. The steps are, however, not in the same order as before. We first obtain the option price from which we derive the fundamental PDE and the hedging ration θ_t . Also, in this method, the trading strategy yielding the pay-off plays an important role.

10.3.7 Equivalent martingale measure

We have seen that by construction

$$E_Q \left[S_T^{\dagger*} | S_t^{\dagger*} \right] = S_t^{\dagger*}, \quad \text{for } T \geq t.$$

Given that the information contained in $S_t^{\dagger*}$ and S_t is the same, we may write without the change of numeraire that

$$e^{-r\tau} E_Q [S_T | S_t] = S_t. \quad (10.26)$$

As a consequence, we notice that the expected future value of the asset, discounted at the risk-free rate equals the current value of S_t . This type of discounting is characteristic for risk-neutral investors. For this reason, the probability measure Q is called a risk-neutral probability. Also (10.26) is important and will be referred to in the following as the martingale condition. Often, as we will see in the next Chapter, we impose this condition to derive a risk-neutral probability.

Another interesting issue has to do with the passage from P to Q . As shown in Chapter (14), there exists a function, say M , so that

$$M(S_T) = \frac{dQ(S_T)}{dP(S_T)}.$$

Therefore, (10.19) can also be written formally as

$$V_t^* = e^{-r\tau} \int_{-\infty}^{\infty} h(S_T) dQ(S_T) = e^{-r\tau} \int_{-\infty}^{\infty} h(S_T) M(S_T) dP(S_T).$$

Similar expressions exist in macroeconomics, where such equations are called Euler equations. There, $M(S_T)$ is called a *pricing kernel* or intertemporal marginal rate of substitution. We see that dQ , M , and dP are closely linked. If we have two of the three elements, the third can be deduced. This is a task that we will explore in the next section.

Non-structural Option Pricing

The Black-Scholes-Merton option pricing model, described in Chapter 10, critically relies on the assumption that asset price follows a log-normal diffusion process. A direct consequence is that volatility deduced from the BSM formula should be constant across exercise prices and across maturities. In Section 11.1, we illustrate that the log-normal assumption does not hold in practice. Implied volatilities of options of different strike prices are not the same, and in addition the volatility smiles are different from one maturity to another. These two features contradict the log-normal diffusion assumption and suggest that option pricing models more general than the BSM model are required. We then explain in Section 11.2 that state prices may be inferred from option prices. In a continuous setting, prices of future states will be given by a state price density. This state price density is often referred to in the literature as a *risk-neutral densitie* (RND).

The literature on estimating RNDs is extensive. The models described in the literature could be broadly divided into two main categories, structural versus non-structural approaches. We qualify a model as “structural” if it proposes a full description for the stock price dynamics and in some cases for the volatility process as well. In contrast, a model is said “non-structural” if it yields a description of the RND without completely describing the dynamic of price. The non-structural approaches can themselves be divided into three types of models, namely parametric, semi-nonparametric, or non-parametric models. The rest of the chapter describes some non-structural models developed to extract state prices from option data. *Parametric models* propose a direct expression for the RND, without referring to any price dynamic. They include a mixture of distributions and a wide collection of distributions allowing for higher moments. *Semi- and non-parametric models* propose some approximation of the true RND. For instance, Jarrow and Rudd (1982) show how the RND can be obtained as an Edgeworth expansion around a log-normal density. Also, Madan and Milne (1994) use a Hermite polynomial approximation. *Non-parametric models* do not try to give an explicit form

for the RND. They include spline methods and kernel estimation, where no assumption about the RND is made and “the data is left to speak for itself”.

It is worth emphasizing that in this chapter we do not discuss structural models, based for instance on stochastic volatility (Heston, 1993) or jump diffusion (Bates, 1996) models. These models, which provide a full description of the price dynamic, are described in the next chapter. Due to the large number of unknown parameters involved, they are rarely used for estimating RNDs, but rather for calibrating option pricing models.

11.1 Difficulties with the standard BSM model

In this first section, we illustrate some of the anomalies that are typically encountered when the standard BSM model is used to infer the RND from option prices. As already argued, under the log-normal assumption, the volatility should be constant across exercise prices and across horizons. We provide two examples in which this assumption is clearly rejected by the data.

The first illustration is based on French Franc/Deutsche Mark options. Such data has been used in Jondeau and Rockinger (2000). For foreign exchange, the option pricing model based on the log-normal assumption of the exchange rate has been developed by Garman and Kohlhagen (1983), following the methodology outlined by Black and Scholes (1973) and Merton (1973) for stock prices. The call option price formula has to be slightly modified, because the two currencies typically yield a risk-free interest rate

$$C_t = e^{-r^* \tau} S_t \Phi(d_1) - e^{-r \tau} K \Phi(d_2),$$

where $d_1 = (\log(S_t/K) + (r - r^* + \frac{1}{2}\sigma^2)\tau) / (\sigma\sqrt{\tau})$ and $d_2 = d_1 - \sigma\sqrt{\tau}$, and S_t denotes the exchange rate, r the domestic rate, and r^* the foreign exchange rate. Another difference with the framework of Chapter 10 is that OTC options’ quotation is not done in terms of prices for a set of exercise prices, but in terms of volatilities for options of various *deltas*. There is a one to one relation between deltas and the strike price. The delta of an option is defined as the derivative of the price with respect to the underlying value. Hence, for a call option, we have $\delta = \frac{\partial C_t}{\partial S_t} = e^{-r^* \tau} \Phi(d_1)$. Since δ is a strictly decreasing function of K , for each δ there corresponds a unique strike price that can be extracted numerically.

Figures 11.1 and 11.2 trace the volatility smile and the term structure of volatilities for two dates, May 17, 1996, and June 27, 1997, as functions of the delta and maturity. The first date is chosen at random, whereas the second date corresponds to the first trading day after President Chirac called for snap elections. If log-normality held, we should observe a straight line independent of maturity. For a given maturity, the deviation from the straight line is called the *volatility smile*. The shift across maturities is the *term structure of volatilities*. Here, options with low δ (high strike prices) are highly valued, meaning

that the market expected an increase in the exchange rate (a FRF depreciation). We notice for May 17, 1996, the normal day, that as the expiration of the options is more remote in time, volatility is higher. For the day after the announcement, the volatility smile translates into a steeper object, called a *smirk*. Also the term structure of volatilities is inverted, the shortest maturity having the highest volatility.

The second illustration is based on CAC 40 options. We consider end-of-the-month options for November 7, 2000. For this day, there were 24 quoted strikes. The implied volatility smile has the pattern presented in Figure 11.3. As it appears clearly, the smile is not flat but instead is decreasing. Once again, it suggests strong rejection of the log-normal assumption.

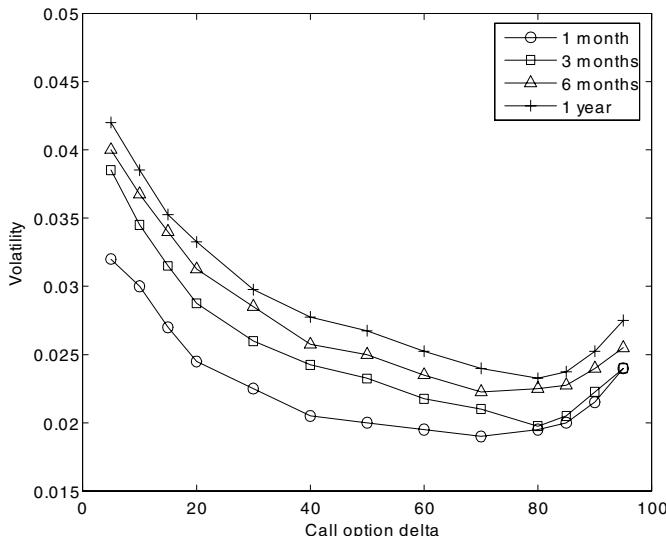


Fig. 11.1. Volatility smile for May 17, 1996.

11.2 Direct estimation of the risk-neutral density

11.2.1 Expression for the RND

We recall that the price of a call option is given by the expression

$$\begin{aligned} C(S_t, t) &= e^{-r\tau} E^* [\max(S_T - K, 0) | S_t, t] \\ &= e^{-r\tau} \int_0^\infty \max(S_T - K, 0) q(S_T | S_t, t) dS_T. \end{aligned}$$

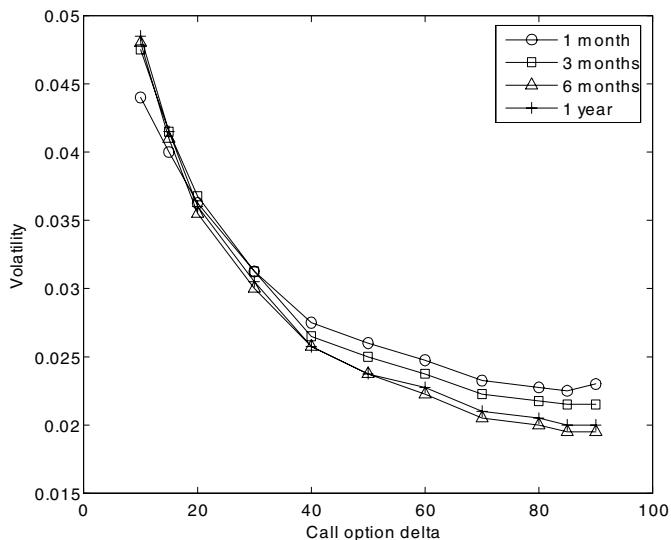


Fig. 11.2. Volatility smile for April 25, 1997.

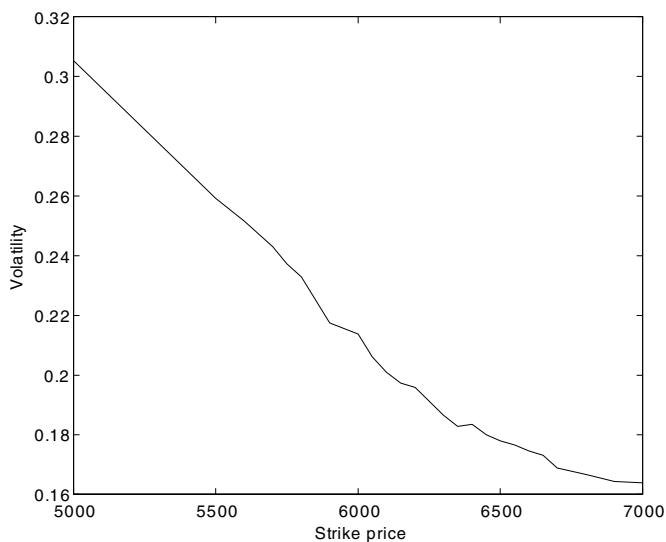


Fig. 11.3. Volatility smile for the CAC 40 for November 7, 2000.

The first attempt to derive the RND using this expression is by Breeden and Litzenberger (1978). They realized that differentiating this equation with respect to the exercise price K yields the discounted *cdf*

$$\frac{\partial C}{\partial K} = -e^{-r\tau} \int_K^\infty q(S_T) dS_T,$$

and that differentiating twice yields the discounted *pdf*

$$\left. \frac{\partial^2 C}{\partial K^2} \right|_{K=S_T} = e^{-r\tau} q(S_T). \quad (11.1)$$

These computations show that the second derivative of the call price yields the discounted RND.¹ This suggests that a first method to extract RND is to approximate it numerically applying the finite difference approach to (11.1). Breeden and Litzenberger (1978) actually estimate a RND in this way. Unfortunately, this method relies on the assumption that there exist traded option prices for many strikes. This is not likely to be the case in practice. Also, it has been shown that RNDs estimated in this manner are very unstable. The reason is that taking differences twice exacerbates even tiny errors in the prices. To illustrate how the second-order derivatives of a call option yield a density but that numerical computations may lead to instability, we proceed with a simulation. We create a set of artificial call option prices under the assumption that the BSM model holds and the parameters are $S_0 = 100$, $r = 5\%$, $\sigma = 20\%$, $T = 60$ days. Figure 11.4 displays with dashes the RND obtained from unperturbed option prices and with dots the RND computed from perturbed option prices.²

This figure illustrates that the computation of an RND from option prices via second-order differentiation may be difficult. For instance, if option prices suffer from non-synchronicity bias (that is, the underlying asset price is not observed at the same time as the option price), if the option price is fudged because of some microstructure reason (for instance, due to the bid-ask spread), then the RND may be very imprecisely estimated. This evidence suggests that we should extract the RND using alternative methods that put more structure on the option prices. Before turning to the discussion of various such methods, we briefly describe how the parameters of these models are estimated.

11.2.2 Estimating the parameters of the RND

Assume that we have to estimate a given model with parameter set θ . The simplest model we may think of is BSM in which the volatility parameter σ

¹ It should be mentioned that $q(\cdot)$ is the undiscounted RND on which we focus in this chapter, whereas $e^{-r\tau}q(\cdot)$ represents an Arrow-Debreu state price, which is referred to as the risk-neutral density (RND) or state price density (SPD).

² Formally, if C is an option price, then the random perturbation is obtained as $C + \varepsilon$ where $\varepsilon \sim N(0, (0.01 \times \bar{C})^2)$ and where \bar{C} is the average call price over all options.

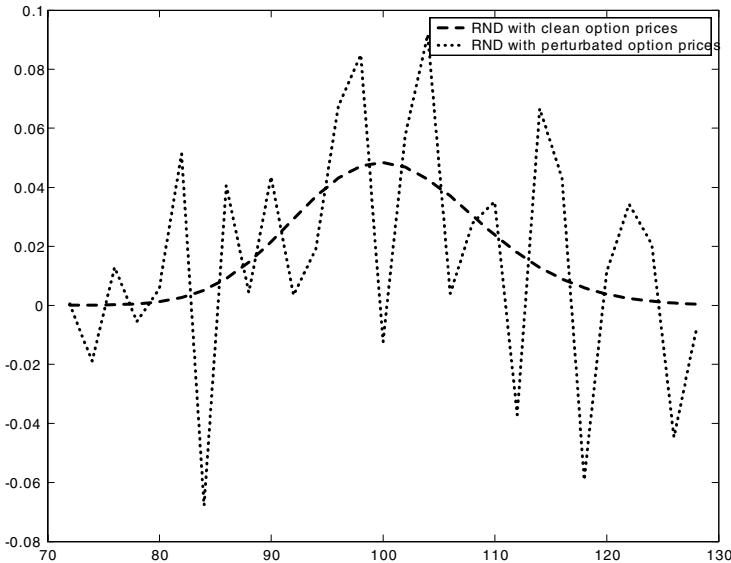


Fig. 11.4. *RND obtained with a second-order numerical differentiation. The RND with the dotted line stems from perturbed option prices. This object does not have a probability mass of one, nor are all the values positive.*

is assumed to be constant for all strike prices. In such a case, assume that for horizon τ , we have N_τ^c strike prices for which we have call options and N_τ^p strike prices for which we have put options. For date t and horizon τ , observed call and put option prices are denoted $C_{t,\tau,i}$, $i = 1, \dots, N_\tau^c$ and $P_{t,\tau,i}$, $i = 1, \dots, N_\tau^p$. Theoretical call and put option prices, implied by the assumed model, are denoted $C_t(K, \tau, \theta)$ and $P_t(K, \tau, \theta)$, respectively, for strike price K and horizon τ . Then, the parameter vector $\theta \in \Theta$ is typically estimated by non-linear least squares, by minimizing for each day and each maturity

$$\min_{\theta \in \Theta} \sum_{i=1}^{N_\tau^c} \omega_i^c (C_{t,\tau,i} - C_t(K_i, \tau, \theta))^2 + \sum_{i=1}^{N_\tau^p} \omega_i^p (P_{t,\tau,i} - P_t(K_i, \tau, \theta))^2, \quad (11.2)$$

where ω_i^c and ω_i^p are weights associated with option i and Θ is the domain of θ . These weights could, for instance, be given by a measure of liquidity of a given option. Examples of measures of liquidity could be the number of trades for a given day or the width of the bid-ask spread.³

³ Instead of using call and put option price formulae, we may first transform put option prices into call option prices using the put-call parity. When a put and a call option exist for the same strike, we generally use the out-of-the-money option, because of their greater liquidity.

There are several practical issues in the estimation of RNDs. A first issue is the type of option price available. In empirical work, we generally use compensation prices, i.e., prices that are collected at the end of the trading day when all market participants give a representative quote. Another possibility would be to use the average of the bid and ask price. We should use only the out-of-the-money options, because they are more liquid. Also we may use intraday information when it exists.

Another important issue is the construction of RNDs with constant time to maturity. A difficulty encountered in the comparison of RNDs across time is that the time of maturity is often constant, so that the horizon varies every day. Butler and Davies (1998) suggest linear interpolation as the preferred way to obtain information for options with a constant time to maturity.⁴ Formally, let q_{t,T_1} and q_{t,T_2} be the RNDs, extracted at day t , for the maturities T_1 and T_2 . They suggest to construct a RND with maturity date T , ($T_1 < T < T_2$) using a convex combination of the extracted RNDs where the weighting depends on the relative position of T between T_1 and T_2 . Hence, we have

$$q_{t,T}(S_T) = \frac{T_2 - T}{T_2 - T_1} q_{t,T_1}(S_T) + \frac{T - T_1}{T_2 - T_1} q_{t,T_2}(S_T), \quad \forall S_T.$$

In theory, more complex combinations can be considered such as a quadratic interpolation. It should be noticed that this method must be used with care for extrapolations, because it may yield negative densities. Abken, Madan, and Ramamurtie (1996) also derive a relationship that links the parameters for different maturities within their Hermite polynomial approach.

11.3 Parametric methods

11.3.1 Mixture of log-normal distributions

Bahra (1996), Melick and Thomas (1997), and Söderlind and Svensson (1997) are among the first ones to describe the RND as a mixture of distributions. It should be noticed that the description of a density as a mixture of other densities has been a very convenient way to improve the fit in many statistical problems. Such an approach allows capturing very flexible distributions, yet at the price of some technical difficulties for the estimation. For option pricing, the most well-known distribution studied in the literature is the mixture of log-normal densities. The reason is that it appears as a trivial extension of the BSM model that involves the single log-normal density.

The model

We remind that the log-normal *pdf* with mean μ and volatility s is defined by

⁴ The technique proposed by Butler and Davies (1998) to construct RNDs with constant horizon can be used for any model.

$$l(S_T, \mu, \sigma) = \frac{1}{S_T \sqrt{2\pi s^2}} \exp \left(-\frac{1}{2} \left(\frac{\log(S_T) - \mu}{s} \right)^2 \right).$$

A mixture of such densities yields

$$q(S_T; \theta) = \sum_{i=1}^M \alpha_i l(S_T; \mu_i, s_i),$$

where θ regroups all the unknown parameters α_i, μ_i, s_i for $i = 1, \dots, M$, where M denotes the number of mixtures describing the RND. Obviously, to guarantee that q is a density, it must be that $\alpha_i \geq 0$ for all $i = 1, \dots, M$, and $\alpha_1 + \dots + \alpha_M = 1$. In other words, q should be a convex combination of the various log-normal densities.

The option price for such a mixture of log-normal distributions is, for a given strike K and time to maturity $\tau = T - t$

$$\begin{aligned} C^{LN}(K; \theta) &= e^{-r\tau} \int_K^{+\infty} (S_T - K) q(S_T; \theta) dS_T \\ &= e^{-r\tau} \int_K^{+\infty} (S_T - K) \sum_{i=1}^M \alpha_i l(S_T; \mu_i, s_i) dS_T \\ &= e^{-r\tau} \sum_{i=1}^M \alpha_i \int_K^{+\infty} (S_T - K) l(S_T; \mu_i, s_i) dS_T, \end{aligned} \quad (11.3)$$

where we define the volatility over the horizon of the option as $s_i = \sigma_i \sqrt{\tau}$, to simplify notation. The last equality is obtained by simply inverting the sum and integral operators. There are various ways to evaluate the integral. For instance, we have⁵

$$\begin{aligned} \int_K^{+\infty} (S_T - K) l(S_T; \mu, s) dS_T &= (E[S_T | S_T > K] - K) \Pr[S_T | S_T > K] \\ &= \exp \left(\mu + \frac{1}{2} s^2 \right) \times \\ &\quad \left[1 - \Phi \left(\frac{\log(K) - \mu - s^2}{s} \right) \right] \\ &\quad - K \left[1 - \Phi \left(\frac{\log(K) - \mu}{s} \right) \right]. \end{aligned} \quad (11.4)$$

⁵ This result builds on Johnson, Kotz, and Balakrishnan (1994, p. 241) who indicate that if S follows a log-normal distribution with mean μ and variance σ^2 , then

$$E[S | S > K] = \exp \left(\mu + \frac{1}{2} \sigma^2 \right) \frac{1 - \Phi(U - \sigma)}{1 - \Phi(U)},$$

where $U = (\log(K) - \mu) / \sigma$.

Finally, the option price is given by

$$\begin{aligned} C^{LN}(K; \theta) &= e^{-r\tau} \sum_{i=1}^M \alpha_i \left\{ \exp \left(\mu_i + \frac{1}{2} s_i^2 \right) \left[1 - \Phi \left(\frac{\log(K) - \mu_i - s_i^2}{s_i} \right) \right] \right. \\ &\quad \left. - K \left[1 - \Phi \left(\frac{\log(K) - \mu_i}{s_i} \right) \right] \right\}. \\ &= e^{-r\tau} \sum_{i=1}^M \alpha_i \exp \left(\mu_i + \frac{1}{2} s_i^2 \right) \Phi \left(\frac{-\log(K) + \mu_i + s_i^2}{s_i} \right) \\ &\quad - e^{-r\tau} K \sum_{i=1}^M \alpha_i \Phi \left(\frac{-\log(K) + \mu_i}{s_i} \right), \end{aligned}$$

where the last expression is obtained by rearranging terms.

Under the risk-neutral probability, we have to impose the martingale condition that states that the current price S_t under the RND is equal to the expected discounted price of the underlying asset $e^{-r\tau} E[S_T]$, so that

$$S_t = e^{-r\tau} E[S_T] = e^{-r\tau} \sum_{j=1}^M \alpha_j \exp \left(\mu_j + \frac{1}{2} s_j^2 \right).$$

As an alternative, we can directly use the BSM formula and set, in a similar spirit to what precedes,

$$C^{LN}(K, \theta) = \sum_{i=1}^M \alpha_i [S_t \Phi(d_{1,i}) - K e^{-r\tau} \Phi(d_{2,i})], \quad (11.5)$$

where $d_{1,i} = [\log(S_t/K) + (\mu_i \tau + \frac{1}{2} s_i^2)]/s_i$, and $d_{2,i} = d_{1,i} - s_i$. The martingale condition may be imposed as follows

$$S_t = C^{LN}(0, \theta).$$

This condition just means that, for a strike price of 0, the option will always get exercised and, hence, at maturity we will always get the underlying asset. In practice, imposing $K = 0$ in the BSM formula is not a good idea because $\log(0)$ is not defined. It is possible to approach this limit case by setting K equal to some very small value.

The earlier studies used the explicit description of the option pricing formula (11.3) or (11.4), whereas the latter ones built on a direct use of the BSM formula (11.5). There are several advantages of using the latter technique. First, the implementation is straightforward, because it suffices to slightly modify the usual BSM formula by introducing the additional parameter μ . Second, the implementation via the BSM formula introduces the current price of the underlying asset, S_t . The parameter μ_i will, therefore, be of the magnitude of an annualized asset return. This in turn means that its value can

be easily bounded, which is a welcome feature for numerical purposes. One further advantage is that the time to maturity τ is explicitly taken into account. This means that μ_i will be an annualized number which implies that parameters extracted for different maturities are directly comparable.

Estimation issues

The mixture-of-distribution approach comes, however, at a cost. A first drawback in fitting a mixture of distributions is the symmetry between the densities. To illustrate this point, assume that a given RND can be correctly described by a mixture involving exactly two normal densities. Using obvious notations, we can write

$$q(S_T, \alpha, \mu_1, \mu_2, s_1, s_2) = \alpha \times l(S_T; \mu_1, s_1) + (1 - \alpha) \times l(S_T; \mu_2, s_2).$$

Obviously, the same density is obtained if we invert the two log-normal distributions, i.e., $q(S_T, 1 - \alpha, \mu_2, \mu_1, s_2, s_1)$. Clearly, the order of the parameters plays a critical role here. For an optimization program, this means, however, that several parameter vectors are associated with a same density. This in turn could yield numerically unstable programs where the optimizer cycles in an infinite loop.

Another difficulty is testing the number of distributions that are involved in the RND. One may be tempted to use a likelihood-ratio type test to check if the i th density should be included in the mixture by testing if $\alpha_i = 0$. However, if $\alpha_i = 0$, the parameters μ_i and s_i associated with the i th density are unidentified. Stated differently, μ_i and s_i could take any value, because they would not play a role in the density q . Sometimes, such parameters are called *nuisance* parameters. In such a situation, the usual statistical theory breaks down and specific tests need to be developed. This type of problem appears also in switching regressions. There is no obvious solution to it. Critical values may be obtained by simulations. In practice, one may add distributions up to the point where adding them yields to no further improvement.

There exist various solutions to help the optimizer converge in terms of MSE. We will discuss the situation in which we estimate a mixture of two or three densities. It seems to provide a good fit in most cases. In the event of two densities, $M = 2$, we start with a grid for α . Since it is known that $0 \leq \alpha \leq 1$, one may subdivide the interval $[0, 1]$ into equally spaced points, where $\alpha_i = i\Delta$, for $i = 0, \dots, N$. In practice, $\Delta = 0.1$ often yields very good results. Furthermore, to avoid the problem of symmetry mentioned above, it is possible to impose that the densities remain in a given order. One possibility that appears to give satisfactory results is to impose that $s_1 > s_2$. The first density will then have a larger standard deviation than the second one.

Similar extensions may be given for the case where $M = 3$. The α parameters may be taken over a simplex, as indicated in Figure 11.5. In this figure, α_3 takes various values, ranging between 0 and 1. For each level of α_3 ,

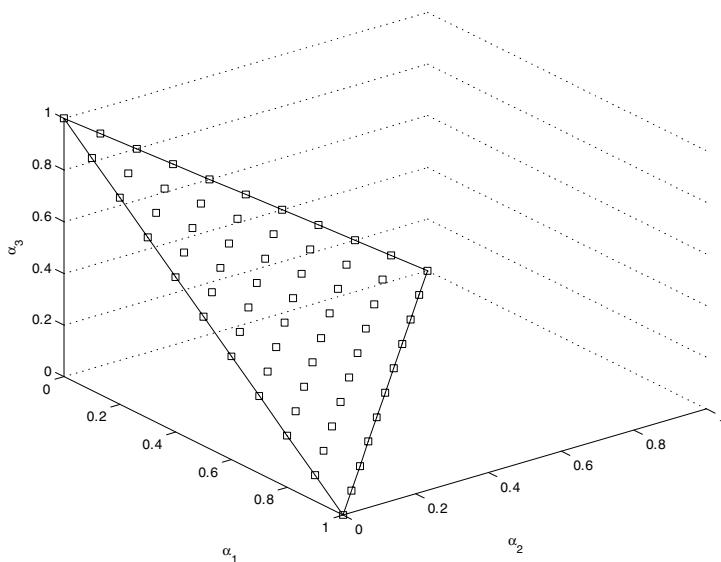


Fig. 11.5. Grid for α_1 , α_2 , and α_3 yielding the weights for the mixture of three densities.

the parameters α_1 and α_2 must satisfy $\alpha_1 + \alpha_2 = 1 - \alpha_3$. It is very easy to construct such a grid. Whereas this evaluation over a grid is useful for numerical purposes, imposing the martingale condition is necessary for the density to be risk neutral. The martingale restriction represents an equality constraint. Even though it is possible in many optimizers to explicitly impose this equality constraint, in practice it is often enough to impose it as a Lagrangean penalty. This means that if θ denotes the parameter set and if NC_τ , NP_τ are the number of calls and puts for which prices are available, denoted by $C_{\tau,i}$, $i = 1, \dots, NC_\tau$ and $P_{\tau,i}$, $i = 1, \dots, NP_\tau$, the optimization will be

$$\sum_{i=1}^{NC_\tau} \omega_i^c (C_{\tau,i} - C(\tau, K_i, \theta))^2 + \sum_{i=1}^{NP_\tau} \omega_i^p (P_{\tau,i} - P(\tau, K_i, \theta))^2 \quad (11.6)$$

$$+ w \left(S_t - e^{-r\tau} \sum_{j=1}^M \alpha_j \exp \left(\mu_j + \frac{1}{2} \sigma_j^2 \tau \right) \right)^2,$$

where ω_i^c and ω_i^p are weights associated with certain options. The parameter w corresponds to a penalty parameter on the martingale constraint. The larger the value of w , the more important the role played by the martingale restriction. For practical purposes, in a first run w may be set equal to 1 and in a second run equal to 100. Some optimizers allow for an explicit implementation of the martingale condition as an equality constraint.

11.3.2 Mixtures of hypergeometric functions

Abadir and Rockinger (2003) observe that option prices involve an expectation whereas the RND involves second-order derivatives. They recognize that if the density could be represented as a polynomial, then integration thereof would easily yield the *cdf*. Furthermore, truncated expectations could also be easily computed. One convenient way to represent densities is by using the so-called generalized hypergeometric functions

$${}_pF_q \left(\begin{matrix} \alpha_1, \dots, \alpha_p \\ \beta_1, \dots, \beta_q \end{matrix}; s \right) \equiv \sum_{j=0}^{\infty} \frac{\prod_{k=1}^p (\alpha_k)_j}{\prod_{k=1}^q (\beta_k)_j} \frac{s^j}{j!}, \quad (11.7)$$

where

$$(a)_j \equiv (a)(a+1)\cdots(a+j-1) = \frac{\Gamma(a+j)}{\Gamma(a)},$$

and where the gamma function is denoted by $\Gamma(\nu)$ for all ν . It is well-known that many densities can be conveniently written in terms of the generalized hypergeometric function. For instance, the gamma *cdf* can be written as

$$\gamma(\nu, s) \equiv \int_0^s e^{-x} x^{\nu-1} dx \equiv \frac{s^\nu}{\nu} {}_1F_1(\nu; \nu+1; -s), \quad -\nu \notin \mathbb{N} \cup \{0\}, \quad (11.8)$$

and the normal one as

$$\begin{aligned} \Phi(s) &\equiv \int_{-\infty}^s e^{-x^2/2} \frac{dx}{\sqrt{2\pi}} \equiv \frac{1}{2} + \frac{s}{\sqrt{2\pi}} {}_1F_1\left(\frac{1}{2}; \frac{3}{2}; -\frac{s^2}{2}\right) \\ &\equiv \frac{1}{2} + \frac{\text{sign}(s)}{2\sqrt{\pi}} \gamma\left(\frac{1}{2}, \frac{s^2}{2}\right), \end{aligned} \quad (11.9)$$

where $\gamma(\cdot, \cdot)$ is the incomplete-gamma function, $\Phi(s)$ is the standard normal *cdf*, and $\text{sign}(s)$ is the sign function.

For double integrals of densities, a mixture that extends the cases seen above is given by

$$\begin{aligned} C(s) &\equiv c_1 + c_2 s + 1_{s>m_1} a_1 (s-m_1)^{b_1} {}_1F_1\left(a_2; a_3; b_2 (s-m_1)^{b_3}\right) \\ &\quad + (a_4) {}_1F_1\left(a_5; a_6; b_4 (s-m_2)^2\right), \end{aligned} \quad (11.10)$$

where $-a_3, -a_6 \notin \mathbb{N} \cup \{0\}$, $b_2, b_4 \in \mathbb{R}_-$. The indicator function is required to represent a component of the density with bounded support. It is also enough for keeping the function real-valued for general values of b_1 and b_3 .

Not all the parameters in $C(\cdot)$ are free to vary unrelatedly, because some restrictions are needed for the function to be the integral of a *pdf*. Again, the parameters must be numerically estimated. The RND is very easily obtained by derivation of a polynomial.

11.3.3 Generalized beta distribution

Bookstaber and McDonald (1987) proposed the generalized beta distribution to model asset returns. Recently, Liu et al. (2003) use this distribution as a description of stock prices. It turns out that the passage from the risk-neutral to the objective probability can be made via a rather simple modification of the parameters. This is what renders this distribution interesting. This density is called GB2. It involves four positive parameters that we group in a vector $\theta = (a, b, p, q)$. The density of the GB2 is given by

$$f(s; \theta) = \frac{a}{b^{ap} B(p, q)} s^{ap-1} [1 + (\frac{s}{b})^a]^{-(p+q)} \quad \text{for } s > 0.$$

We remind that $B(p, q) = \Gamma(p)\Gamma(q)/\Gamma(p+q)$. If s is a random variable distributed as a GB2, then

$$E[s^n] = \frac{b^n B(p + \frac{n}{a}, q - \frac{n}{a})}{B(p, q)} \quad \text{for } n < aq.$$

Risk neutrality can be imposed as suggested by Gerber and Shiu (1994) by imposing the following martingale condition

$$S_t = e^{-r\tau} \frac{b B(p + \frac{1}{a}, q - \frac{1}{a})}{B(p, q)}.$$

As shown by Liu et al. (2003), the price of a European call option may be written as

$$\begin{aligned} C(K; \theta) &= e^{-r\tau} \int_0^\infty (S_T - K)^+ f(S_T; \theta) dS_T \\ &= S_t [1 - F(K|a, b, p + \frac{1}{a}, q - \frac{1}{a})] - K e^{-r\tau} [1 - F(K|a, b, p, q)] \end{aligned}$$

where F is the *cdf* of the GB2 density. It is related to the one of the beta distribution, written $F_\beta(\cdot)$, using

$$F(s; a, b, p, q) = F_\beta(z(s; a, b); p, q),$$

where $z(s; a, b) = (s/b)^a/[1 + (s/b)^a]$. The *cdf* of the beta distribution is implemented in modern matrix languages. In the estimation of the RND from actual option prices, it is important to impose the martingale restriction.

11.4 Semi-parametric methods

11.4.1 Edgeworth expansions

In this section, we outline the method developed by Jarrow and Rudd (1982) for which a numerical application can be found in Corrado and Su (1996). The

idea of Jarrow and Rudd (1982) is to capture deviations from log-normality by an Edgeworth expansion of the RND $q(S_T|\theta)$ around the log-normal density.⁶ The construction of an Edgeworth expansion is conceptually similar to computing a Taylor expansion but applies to functions in general and to densities in particular. In a conventional Taylor expansion, the function is approximated by a simple polynomial around a given point. Here, the actual (unknown) RND is approximated by an expansion around a log-normal distribution. A clear advantage of the expansion approach is that the approximation, by involving parameters that can be varied, allows generating more general functions. Other studies using the Edgeworth expansions are Flamouris and Giamouridis (2002), and Beber and Brandt (2003).

In the next section, we will present an alternative semi-parametric approach proposed by Madan and Milne (1994) based on the Hermite approximation of the Gaussian density. Although the two resulting RNDs are conceptually different, because they are based on different approximation methods, they are usually numerically very close one to the other.

First we sketch how Edgeworth expansions can be obtained. Let Q be the *cdf* of a random variable S_T and q its density. Define the characteristic function of S as $\phi_Q(u) \equiv \int e^{isu} q(s) ds$. If moments of S_T exist up to order n , then there exist cumulants of the distribution Q , denoted $\kappa_{Q,j}$ and implicitly defined by the expansion

$$\log(\phi_Q(u)) = \sum_{j=1}^{n-1} \kappa_{Q,j} \frac{(iu)^j}{j!} + o(u^{n-1}).$$

Thus, if the characteristic function $\phi_Q(.)$ is known, by taking an expansion of its logarithm around $u = 0$, it is possible to obtain the cumulants. We have the following relationships between cumulants and moments up to the fourth order: $\kappa_{Q,1} = E[S_T]$, $\kappa_{Q,2} = V[S_T]$, $\kappa_{Q,3} = E[(S_T - E[S_T])^3]$, $\kappa_{Q,4} = E[(S_T - E[S_T])^4] - 3V[S_T]^2$. The knowledge of the first four cumulants is, therefore, equivalent to knowing the mean, variance, skewness, and kurtosis.

Jarrow and Rudd show that an Edgeworth expansion of the fourth order for the true probability distribution Q around the log-normal *cdf* L can be written, after imposing that the first moment of the approximating density and the true density are equal, ($\kappa_{Q,1} = \kappa_{L,1}$) and by denoting densities with small letters

$$\begin{aligned} q(s) &= l(s) + \frac{\kappa_{Q,2} - \kappa_{L,2}}{2!} \frac{d^2 l(s)}{ds^2} - \frac{(\kappa_{Q,3} - \kappa_{L,3})}{3!} \frac{d^3 l(s)}{ds^3} \\ &\quad + \frac{(\kappa_{Q,4} - \kappa_{L,4}) + 3(\kappa_{Q,2} - \kappa_{L,2})^2}{4!} \frac{d^4 l(s)}{ds^4} + \epsilon(s), \end{aligned}$$

⁶ Edgeworth expansions are frequently used in statistical theory to obtain distributions that deviate from the normal one. A related approach has been followed in Chapter 5 to describe the non-normal conditional distribution of returns.

where $\epsilon(s)$ captures terms neglected in the expansion. The various terms in the expansion correspond to adjustments of the variance, skewness, and kurtosis. As argued before, the interpretation of this expression is similar to a Taylor expansion.

Jarrow and Rudd further show that, under this approximated density, the price of a European call option with strike K can be approximated as

$$\begin{aligned} C(Q) &= e^{-r\tau} \int_K^\infty (S_T - K) q(S_T) dS_T \\ &\approx e^{-r\tau} \int_K^\infty (S_T - K) l(S_T) dS_T \\ &\quad + e^{-r\tau} \frac{\kappa_{Q,2} - \kappa_{L,2}}{2!} \int_K^\infty (S_T - K) \frac{d^2 l(S_T)}{dS_T^2} dS_T \\ &\quad - e^{-r\tau} \frac{\kappa_{Q,3} - \kappa_{L,3}}{3!} \int_K^\infty (S_T - K) \frac{d^3 l(S_T)}{dS_T^3} dS_T \\ &\quad + e^{-r\tau} \frac{(\kappa_{Q,4} - \kappa_{L,4}) + 3(\kappa_{Q,2} - \kappa_{L,2})^2}{4!} \int_K^\infty (S_T - K) \frac{d^4 l(S_T)}{dS_T^4} dS_T. \end{aligned}$$

We notice that the first term is simply the Black-Scholes formula. In addition, the log-normal distribution has the following property

$$\int_K^\infty (S_T - K) \frac{d^j l(S_T)}{dS_T^j} dS_T = \left. \frac{d^{j-2} l(S_T)}{dS_T^{j-2}} \right|_{S=K}, \quad \text{for } j \geq 2.$$

We deduce for the call option price

$$\begin{aligned} C(Q) &\approx C(L) + e^{-r\tau} \frac{\kappa_{Q,2} - \kappa_{L,2}}{2!} l(K) - e^{-rT} \frac{\kappa_{Q,3} - \kappa_{L,3}}{3!} \frac{dl(K)}{dS_T} \\ &\quad + e^{-r\tau} \frac{(\kappa_{Q,4} - \kappa_{L,4}) + 3(\kappa_{Q,2} - \kappa_{L,2})^2}{4!} \frac{d^2 l(K)}{dS_T^2}. \end{aligned} \quad (11.11)$$

For the log-normal density, the first cumulants are given by

$$\begin{aligned} \kappa_{L,1} &= S_t e^{r\tau}, \\ \kappa_{L,2} &= [\kappa_{L,1} \vartheta]^2, \\ \kappa_{L,3} &= [\kappa_{L,1} \vartheta]^3 (3\vartheta + \vartheta^3), \\ \kappa_{L,4} &= [\kappa_{L,1} \vartheta]^4 (16\vartheta^2 + 15\vartheta^4 + 6\vartheta^6 + \vartheta^8), \end{aligned}$$

where $\vartheta = (e^{\sigma^2 \tau} - 1)^{1/2}$ and where the first relation follows from risk-neutral valuation.

Jarrow and Rudd suggest identifying the second moment by imposing $\kappa_{L,2} = \kappa_{Q,2}$. This argument is also justified on numerical grounds by Corrado and Su (1996) who notice that without this condition there exists a problem of multicollinearity between the second and the fourth moment. Rather than

estimating the remaining cumulants, $\kappa_{Q,3}$ and $\kappa_{Q,4}$, Corrado and Su (1996) estimate standardized skewness and kurtosis (written respectively $\gamma_{Q,1}$ and $\gamma_{Q,2}$), which are defined through the relationships

$$\begin{aligned}\gamma_{Q,1} &= \frac{\kappa_{Q,3}}{(\kappa_{Q,2})^{3/2}} = 3\vartheta + \vartheta^3, \\ \gamma_{Q,2} &= \frac{\kappa_{Q,4}}{(\kappa_{Q,2})^2} = 16\vartheta^2 + 15\vartheta^4 + 6\vartheta^6 + \vartheta^8.\end{aligned}$$

Clearly, these expressions also hold for the log-normal density. The skewness and kurtosis of the log-normal density can therefore be derived easily from the cumulants above.

With the assumption of equality of the second cumulants for the approximating and the true distribution, it follows that

$$\begin{aligned}C(Q) \approx C(L) - e^{-r\tau}(\gamma_{Q,1} - \gamma_{L,1}) \frac{\kappa_{L,2}^{3/2}}{3!} \frac{dl(K)}{dS_T} \\ + e^{-r\tau}(\gamma_{Q,2} - \gamma_{L,2}) \frac{\kappa_{L,2}^2}{4!} \frac{d^2l(K)}{dS_T^2}. \quad (11.12)\end{aligned}$$

Using this expression, it is easy to estimate with nonlinear least squares the implied volatility, σ^2 , skewness, $\gamma_{Q,1}$, and kurtosis, $\gamma_{Q,2}$.

The expression of the RND can be obtained after twice differentiating (11.12) with respect to K and then evaluation over S_T

$$q(S_T) = l(S_T) - (\gamma_{Q,1} - \gamma_{L,1}) \frac{\kappa_{L,2}^{3/2}}{6} \frac{d^3l(S_T)}{dS_T^3} + (\gamma_{Q,2} - \gamma_{L,2}) \frac{\kappa_{L,2}^2}{24} \frac{d^4l(S_T)}{dS_T^4},$$

where the partial derivatives can be computed iteratively using

$$\begin{aligned}\frac{dl(S_T)}{dS_T} &= - \left(1 + \frac{\log(S_T) - m}{\sigma^2 \tau} \right) \frac{l(S_T)}{S_T}, \\ \frac{d^2l(S_T)}{dS_T^2} &= - \left(2 + \frac{\log(S_T) - m}{\sigma^2 \tau} \right) \frac{1}{S_T} \frac{dl(S_T)}{dS_T} - \frac{1}{S_T^2 \sigma^2} l(S_T), \\ \frac{d^3l(S_T)}{dS_T^3} &= - \left(3 + \frac{\log(S_T) - m}{\sigma^2 \tau} \right) \frac{1}{S_T} \frac{d^2l(S_T)}{d^2S_T} - \frac{2}{S_T^2 \sigma^2} \frac{dl(S_T)}{dS_T} \\ &\quad + \frac{1}{S_T^3 \sigma^2} l(S_T), \\ \frac{d^4l(S_T)}{dS_T^4} &= - \left(4 + \frac{\log(S_T) - m}{\sigma^2 \tau} \right) \frac{1}{S_T} \frac{d^3l(S_T)}{d^3S_T} - \frac{3}{S_T^2 \sigma^2} \frac{d^2l(S_T)}{dS_T^2} \\ &\quad + \frac{3}{S_T^3 \sigma^2} \frac{dl(S_T)}{dS_T} - \frac{2}{S_T^4 \sigma^2} l(S_T),\end{aligned}$$

and where $m = \log(S_t) + (r - \sigma^2/2)\tau$. Those computations indicate that the RND in the Edgeworth case will be a polynomial whose coefficients directly

command the skewness and kurtosis of the RND. We also notice that the RND involves rather complicated terms with derivatives of the log-normal density. The expression of these derivatives can be obtained using straightforward computations.

11.4.2 Hermite polynomials

An alternative, yet similar, semi-parametric approach relies on an approximation of the Gaussian density based on Hermite polynomials. The theoretical foundation of this method is elaborated in Madan and Milne (1994) and applied in Abken, Madan, and Ramamurtie (1996) and Coutant, Jondeau, and Rockinger (2001).⁷

The model described by Madan and Milne works as follows. First, for numerical reasons, they consider the map from the actual prices into the space generated by standardized log-return z , denoting once again the volatility over the horizon of the option as $s = \sigma\sqrt{\tau}$,

$$S_T = S_t \exp \left(\mu\tau - \frac{1}{2}s^2 + sz \right) \implies z = \frac{\log(S_T/S_t) - (\mu\tau - s^2/2)}{s}.$$

The standardized log-return z has zero mean and unit variance.

If we focus on a call option, the payoff of such an option, as a function of z , is

$$c(z, S_t, K, \mu, s, \tau) = \max \left(S_t \exp \left(\mu\tau - \frac{1}{2}s^2 + sz \right) - K, 0 \right) = g(z),$$

so that the price of a call option may be written as

$$\begin{aligned} C(S_t, K, \mu, s, \tau, r) &= e^{-r\tau} \int_0^{+\infty} c(z, S_t, K, \mu, s, \tau) q_z(z) dz \\ &= e^{-r\tau} \int_0^{+\infty} g(z) q_z(z) dz, \end{aligned}$$

where we have introduced the notation $g(z)$ to illustrate that the result holds for all sorts of payoffs. $q_z(\cdot)$ denotes the risk-neutral density of z .

Notice that we can go from the RND defined in z space to the RND defined in S_T space using the following change of variable

$$q(S_T) dS_T = q_z \left(\frac{\log(S_T/S_t) - (\mu\tau - s^2/2)}{s} \right) \times \frac{1}{s} \times dS_T.$$

⁷ It may be argued that, in these papers, the authors adopt a structural approach, because they fully describe the dynamic of the model. However, in almost all applications of the method, this approach has been viewed as a way to approximate the actual RND. We therefore decided, rather arbitrarily, to present this approach as a non-structural one.

Now, we assume that any payoff $g(z)$ may be expressed as a function of basis elements. It turns out that a basis for the Gaussian space is given by Hermite polynomials. In other words, there exist real numbers a_k such that any payoff can be written as

$$g(z) = \sum_{k=0}^{\infty} a_k h_k(z),$$

with $a_k = \int_z g(z) h_k(z) \phi(z) dz$, where $\phi(z)$ is the Gaussian distribution and $h_k(z)$ denotes Hermite polynomials normalized to unit variance, defined as

$$\begin{aligned} h_0(z) &= 1, \\ h_1(z) &= z, \\ h_2(z) &= \frac{1}{\sqrt{2}} (z^2 - 1), \\ h_3(z) &= \frac{1}{\sqrt{6}} (z^3 - 3z), \\ h_4(z) &= \frac{1}{\sqrt{24}} (z^4 - 6z^2 + 3). \end{aligned}$$

The coefficients a_k can be interpreted as the covariance between the option payoff and the k th Hermite polynomial risk. Consequently, the price of the call option becomes

$$C(S_t, K, \mu, s, \tau, r) = e^{-r\tau} \sum_{k=0}^{\infty} a_k \int_z h_k(z) q_z(z) dz.$$

Finally, Madan and Milne (1994) assume that the RND $q_z(z)$ can be represented as the product of a change of measure density and a reference measure density

$$q_z(z) = \lambda(z) \phi(z).$$

Here, the reference measure density $\phi(z)$ is simply the Gaussian one and the risk-neutral change of measure density $\lambda(z)$ is approximated by a Hermite polynomial expansion

$$\lambda(z) = e^{r\tau} \sum_{l=0}^{\infty} \pi_l h_l(z).$$

Under the reference measure, z is by construction normally distributed with zero mean and unit variance. But, more generally, the RND $q_z(z)$ may incorporate some departure from normality. The RND can be rewritten as

$$q_z(z) = \phi(z) \left(\sum_{l=0}^{\infty} \pi_l h_l(z) \right).$$

The π_l are interpreted as the implicit price of polynomial risk $h_l(z)$. Obviously, since the polynomial components are not traded, these risks cannot be traded either. For practical purposes, the infinite sum can be truncated up to the fourth order. Since the Hermite polynomial of order l will depend on the l th moment, we will also refer to π_3 and π_4 as the price of skewness and kurtosis, respectively. It can be easily shown that for $q_z(z)$ to be a density, we need $\pi_0 = e^{-r\tau}$. In addition, since z is the standardized log-return (with zero mean and unit variance by construction), it follows that the shifts for the mean (π_1) and the variance (π_2) of $q_z(z)$ relative to the reference measure can be set equal to 0. Consequently, we can estimate the mean μ and the variance σ^2 of log-returns and set $\pi_1 = \pi_2 = 0$. Therefore, we obtain

$$\begin{aligned} q_z(z) &= \phi(z) \left(e^{r\tau} \sum_{l=0}^4 \pi_l h_l(z) \right) = \phi(z) e^{r\tau} (e^{-r\tau} + \pi_3 h_3(z) + \pi_4 h_4(z)) \\ &= \phi(z) \left(1 + \frac{b_3}{\sqrt{6}}(z^3 - 3z) + \frac{b_4}{\sqrt{24}}(z^4 - 6z^2 + 3) \right), \end{aligned}$$

where the $b_i = e^{r\tau} \pi_i$, $i = 3, 4$, are the future value of the i th price of risk coefficient (to be estimated). The parameters b_3 and b_4 correspond to the skewness and kurtosis if the reference measure density of z is chosen to be the normal distribution. It is important to emphasize that unlike the Edgeworth case, since a further change of variable from z to S_T has to be made, b_3 and b_4 will not correspond to the skewness and kurtosis of the underlying price process S_T but to the skewness and kurtosis of log-returns z . Finally, the skewness and kurtosis of the expansion are given by

$$\begin{aligned} Sk[z] &= \sqrt{6}b_3, \\ Ku[z] &= 3 + \sqrt{24}b_4. \end{aligned}$$

The estimation of the parameters μ , σ , b_3 , and b_4 can be easily performed using a nonlinear estimation as indicated in (11.6).

The general expression for the price of a call option is given by

$$C(S_t, K, \mu, s, \tau, r) = \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} a_k \pi_l \int_z h_l(z) h_k(z) \phi(z) dz = \sum_{k=0}^{\infty} a_k \pi_k,$$

where the second equality holds because Hermite polynomials form an orthonormal system. When the infinite sum is truncated up to the fourth order, we obtain

$$C(S_t, K, \mu, s, \tau, r) = e^{-r\tau} a_0 + \pi_3 a_3 + \pi_4 a_4.$$

Therefore, it remains to be shown how to obtain the a_k coefficients. Abken, Madan, and Ramamurtie (1998) introduce the call option generating function

$$G(u, S_t, K, \mu, s, \tau) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} c(z, S_t, K, \mu, s, \tau) \exp\left(-\frac{1}{2}(z-u)^2\right) dz, \quad (11.13)$$

where u is a dummy variable. The coefficients a_k can then be obtained from the generating function as

$$a_k(S_t, K, \mu, s, \tau) = \frac{\partial^k G(u, S_t, K, \mu, s, \tau, r)}{\partial u^k} \Big|_{u=0} \frac{1}{\sqrt{k!}}.$$

The evaluation of (11.13) is shown to be

$$G(u, S_t, K, \mu, s, \tau) = S_t \exp(\mu\tau + su) \Phi(d_1(u)) - K \Phi(d_2(u)),$$

where $d_1(u) = \log(S_t/K)/s + (\mu\tau/s + s/2) + u$, and $d_2(u) = d_1(u) - s$. From there on, it is possible to compute the coefficients a_k . To do so, we introduce the notation $d_j = d_j(0)$, that is we evaluate the $d_j(0)$ at zero. We obtain after tedious computations, and introducing the notations ϕ' , ϕ'' , and ϕ''' , for the first, second, and third derivatives of a standard normal density with respect to its argument

$$\begin{aligned} a_0 &= S_t e^{\mu\tau} \Phi(d_1) - K \Phi(d_2), \\ a_1 &= s S_t e^{\mu\tau} \Phi(d_1) + S_t e^{\mu\tau} \phi(d_1) - K \phi(d_2), \\ a_2 &= \frac{1}{\sqrt{2}} [s^2 S_t e^{\mu\tau} \Phi(d_1) + 2s S_t e^{\mu\tau} \phi(d_1) + S_t e^{\mu\tau} \phi'(d_1) - K \phi'(d_2)], \\ a_3 &= \frac{1}{\sqrt{6}} [s^3 S_t e^{\mu\tau} \Phi(d_1) + 3s^2 S_t e^{\mu\tau} \phi(d_1) + 3s S_t e^{\mu\tau} \phi'(d_1) \\ &\quad + S_t e^{\mu\tau} \phi''(d_1) - K \phi''(d_2)], \\ a_4 &= \frac{1}{\sqrt{24}} [s^4 S_t e^{\mu\tau} \Phi(d_1) + 4s^3 S_t e^{\mu\tau} \phi(d_1) + 6s^2 S_t e^{\mu\tau} \phi'(d_1) \\ &\quad + 4s S_t e^{\mu\tau} \phi''(d_1) + S_t e^{\mu\tau} \phi'''(d_1) - K \phi'''(d_2)]. \end{aligned}$$

This method has been applied in Coutant, Jondeau, and Rockinger (2001) to interest-rate options. Jondeau and Rockinger (2001) also show how conditions on b_3 and b_4 can be imposed so that the polynomial approximation remains positive.

11.5 Non-parametric methods

11.5.1 Spline methods

Shimko (1993) proposes to directly implement the results of Breeden and Litzenberger (1978), but after a preliminary smoothing of the volatility smile (see also Malz, 1997). As we have shown in Section 11.2, a direct estimation leads however to numerically very unstable results. The idea is therefore to summarize the information contained in the volatility smile via a more or less sophisticated polynomial, say $\sigma(K)$, a function of the strike price, K , and

then to use this expression to evaluate the density. In other words, we fit the function $\sigma(K)$ to the various volatilities and then apply Leibnitz rule as indicated in the Section 11.2. Outside the range of quoted strikes, the volatility is supposed to be constant. Formally, we can write, focusing on a call option that

$$C(S_t, K, \tau, r, \sigma) = C(S_t, K, \tau, r, \sigma(K)).$$

A first idea, which is actually used by Shimko, is to use a quadratic polynomial for $\sigma(K)$. Formally, he supposes that

$$\sigma_i = a_0 + a_1 K_i + a_2 K_i^2, \quad \text{for } i = 1, \dots, N,$$

where N represents the number of observed prices. The parameters of this polynomial can be easily estimated using an nonlinear least square regression as discussed in (11.6).

Later contributions assume a cubic spline interpolation for volatility.⁸ See for instance Bliss and Panigirtzoglou (2002) or Weinberg (2001). It may be useful at this stage, before returning to the extraction of the RND, to recall how a cubic spline interpolation operates. First of all, a cubic spline is a third-order polynomial in the strike price,

$$\sigma(K) = a_0 + a_1 K + a_2 K^2 + a_3 K^3,$$

so that for each quoted strike K_i , the function exactly takes the value of the known volatilities, i.e., $\sigma(K_i) = \sigma_i$. To obtain a cubic spline, we assume for the moment that the second-order derivatives of $\sigma(\cdot)$ are also known for the strike K_i . Later on, we will show how these second-order derivatives may be eliminated. Cubic splines assume that the second-order derivative is given as a linear interpolation between the observed points. Clearly, we have

$$\sigma''(K_i) = 2a_2 + 6a_3 K_i,$$

so that the linear interpolation yields, for any $K \in [K_{i-1}, K_i]$, after a slight change of notation, the expression

$$\sigma''(K) = -\sigma''(K_{i-1}) \frac{K - K_i}{K_i - K_{i-1}} + \sigma''(K_i) \frac{K - K_{i-1}}{K_i - K_{i-1}}.$$

Integrating this expression twice with respect to K gives

$$\sigma(K) = -\sigma''(K_{i-1}) \frac{(K - K_i)^3}{6(K_i - K_{i-1})} + \sigma''(K_i) \frac{(K - K_{i-1})^3}{6(K_i - K_{i-1})} + c_0 + c_1 K.$$

Remember that this spline function must take the values $\sigma(K_i) = \sigma_i$ at K_i so that we may solve for a_0 and a_1 . We obtain after some tedious computations that

⁸ Within the context of obtaining term structure of interest rate curves, Waggoner (1997) introduces the idea of using splines. In that contribution, there is also the idea of weighting the spline according to its curvature.

$$\begin{aligned}\sigma(K) = & \sigma''(K_{i-1}) \frac{(K - K_i)^3}{6(K_i - K_{i-1})} + \sigma''(K_i) \frac{(K - K_{i-1})^3}{6(K_i - K_{i-1})} \\ & + \left(\frac{\sigma_i}{K_i - K_{i-1}} - \frac{\sigma''(K_i)(K_i - K_{i-1})}{6} \right) (K - K_{i-1}) \\ & - \left(\frac{\sigma_{i-1}}{K_i - K_{i-1}} - \frac{\sigma''(K_{i-1})(K_i - K_{i-1})}{6} \right) (K - K_i).\end{aligned}$$

By differentiating this expression and taking into account the fact that the derivative should be continuous at the nodes, we obtain finally the system of equations for $i = 2, \dots, N - 1$

$$\begin{aligned}\sigma''(K_{i-1}) \frac{K_i - K_{i-1}}{6} + \sigma''(K_i) \frac{K_{i+1} - K_{i-1}}{3} + \sigma''(K_{i-1}) \frac{K_{i+1} - K_i}{6} \\ = \frac{\sigma_{i+1} - \sigma_i}{K_{i+1} - K_i} - \frac{\sigma_i - \sigma_{i-1}}{K_i - K_{i-1}}.\end{aligned}$$

In other words, we have a system of $N - 2$ equations in the N unknown $\sigma''(K_i)$. To solve this system, additional assumptions need to be made. One of them concerns the boundaries. For instance, we may assume that there is no convexity at the boundaries. This is equivalent to assuming that $\sigma''(K_1) = \sigma''(K_N) = 0$. It is worth noting that each of the parameters that we try to compute is only linked to its nearest neighbors. In a matrix representation, we would obtain a so-called tridiagonal system, which can be very efficiently solved using the algorithm described in Press, Teukolsky, Vetterling, and Flannery (1999).

If we go back now to the estimation of the RND, we notice that the implied volatility – and therefore the price – of option i is depending on the strike price K_i , $i = 1, \dots, N$

$$C^{\text{SH}}(t, S, K, T) = S\Phi(d_1(\sigma(K))) - e^{-r\tau} K\Phi(d_2(\sigma(K))).$$

Consequently, the RND takes the expression

$$\begin{aligned}q^{\text{SH}}(K) &= e^{r\tau} \frac{\partial^2 C(t, S, \sigma(K), T)}{\partial K^2} \\ &= e^{r\tau} S \left(d_1''\phi(d_1(K)) - (d_1')^2 d_1\phi(d_1(K)) \right) - d_2'\phi(d_2(K)) \\ &\quad - K \left(d_2''\phi(d_2(K)) - (d_2')^2 d_2\phi(d_2(K)) \right).\end{aligned}$$

For the quadratic case considered initially by Shimko (1993), we have

$$\begin{aligned}d_1(K) &= \frac{1}{\sigma(K)\sqrt{\tau}} \log \left(\frac{S}{Ke^{r\tau}} \right) + \frac{1}{2}\sigma(K)\sqrt{\tau}, \\ d_2(K) &= d_1(K) - \sigma(K)\sqrt{\tau}, \\ \sigma(K) &= (a_0 + a_1 K + a_2 K^2) \mathbf{1}_{(K > \min(K_i) \text{ and } K < \max(K_i))} \\ &\quad + \sigma_1 \mathbf{1}_{(K \leq \min(K_i))} + \sigma_M \mathbf{1}_{(K \geq \max(K_i))},\end{aligned}$$

where the function $1_{(A)}$ is the indicator function taking the value 1 if A is true. The first- and second-order derivatives of d_1 and d_2 are given by

$$\begin{aligned} d'_1(K) &= -\frac{\sigma'(K)\sqrt{\tau}}{\sigma^2(K)\tau} \log\left(\frac{S}{Ke^{r\tau}}\right) - \frac{1}{K\sigma(K)\sqrt{\tau}} + \frac{1}{2}\sigma'(K)\sqrt{\tau}, \\ d'_2(K) &= d'_1(K) - \sigma'(K)\sqrt{\tau}, \\ d''_1(K) &= -\frac{\sigma''(K)\sigma(K)\tau - 2\sigma'(K)^2\tau}{(\sigma(K)\sqrt{\tau})^3} \log\left(\frac{S}{Ke^{r\tau}}\right) + \frac{1}{K}\frac{\sigma'(K)\sqrt{\tau}}{\sigma^2(K)\tau} \\ &\quad + \frac{\sigma(K)\sqrt{\tau} + K\sigma'(K)\sqrt{\tau}}{K^2\sigma^2(K)\tau} + \frac{1}{2}\sigma''(K)\sqrt{\tau}, \\ d''_2(K) &= d''_1(K) - \sigma''(K)\sqrt{\tau}, \\ \sigma'(K) &= (a_1 + 2a_2 K) \mathbf{1}_{(K > \min(K_i) \text{ and } K < \max(K_i))}, \\ \sigma''(K) &= 2a_2 \mathbf{1}_{(K > \min(K_i) \text{ and } K < \max(K_i))}. \end{aligned}$$

One of the difficulties when we fit a cubic spline to actual data is that the interpolated curve tends to oscillate strongly. Bliss and Panigirtzoglou (2002) suggest to use *smoothed splines*. These authors therefore recommend the following procedure. As a first step, they propose to seek a spline satisfying

$$\min_{\theta} \sum_{j=1}^N w_j (\dot{\sigma}_j - \sigma(K_j; \theta))^2 + \lambda \int_0^{+\infty} (\sigma''(K; \theta))^2 dK, \quad (11.14)$$

where N is the number of quoted strikes, $\dot{\sigma}_j$ is the actual implied volatility corresponding to the strike K_j , and θ is the vector of parameters involved in the spline. In the traditional spline computation, we would obtain an exact fit for the first part of the expression since the spline would, by definition, match the implied volatility at each strike. The first term therefore measures the goodness of fit. If we performed a linear interpolation $\sigma(K) = a_0 + a_1 K$, the second term would be equal to zero. For a highly oscillating function, we expect large levels of the curvature, i.e., of the second moment, and therefore a rather large integral. The second term of the expression therefore measures the smoothness of the spline. The solution will be a trade-off between goodness of fit and smoothness. The parameter λ is a penalty for the roughness of the function.

Weinberg (2001) suggests to allow the smoothing parameter λ to vary over time, so that the amount of smoothness may vary depending on the data. The approach is based on the method of generalized cross-validation, as described in Wahba (1990). The idea consists, for a given value of λ , in solving the spline problem (11.14) by omitting each observation j in turn. Let $\sigma_{\lambda}^{[j]}(K_j; \theta)$ denote the estimated smile minimizing (11.14) for a given λ with observation j omitted. Then, the optimal λ is the value that minimizes the error

$$\sum_{j=1}^N w_j \left(\dot{\sigma}_j - \sigma_{\lambda}^{[j]}(K_j; \theta) \right)^2.$$

This approach reduces the effect on the estimated volatility smile of possible outliers in the data.

In RND applications, it is also necessary to extrapolate the spline beyond the range of available strikes. Without any control, the extrapolation could lead to negative values of the implied volatility. Bliss and Panigirtzoglou (2002), therefore, suggest to extrapolate smoothly in a horizontal manner. To do this, they introduce two pseudo-data points spaced three strike intervals above and below the range of strike prices and setting the implied volatilities equal to the ones of the extreme-strike options.

As a second step, the spline is evaluated at a huge number of points, say 5,000. These points in strike-volatility space get converted into price-strike values. Once these points are obtained, the call prices are numerically differentiated twice to yield the estimated *pdf*.

Dumas, Fleming, and Whaley (1998) investigate several specifications of such deterministic volatility functions. In a pricing context, they obtain that these models do not perform significantly better than ad hoc procedures that smooth Black-Scholes implied volatilities across exercise prices and horizons.

11.5.2 Tree-based methods

A tree-based method to extract RNDs out of financial options has been presented by Rubinstein (1994). This method is further extended in Jackwerth and Rubinstein (1996) and Jackwerth (1999). Jackwerth (1997) gives a numerical illustration of the method. The idea of these methods is based on the following steps. First, we take existing options nearest to the money then, after computing their Black-Scholes implied volatility, we compute the average of these implied volatilities. The second step consists in computing the RNDs that would be associated to the binomial tree of Cox, Ross, and Rubinstein (1979) and that would be comparable with these implied volatilities. Formally, start by assuming that the tree has m steps. A terminal node, j , of the binomial tree has a probability to realize of P'_j . If p' is the risk-neutral probability of an up movement on each node of the tree, then

$$P'_j = \frac{m!}{j!(m-j)!} p'^j (1-p')^{m-j}.$$

In that formula, it is assumed that the nodes are denoted in such a way that a small j corresponds to a low value of the support S_T . Let us denote these values as S_j , $j = 1, \dots, m$. Given the current value of the underlying asset, it is also possible to construct the nodal values of the underlying asset in a risk-neutral environment. Even though by working backwards, we could use the binomial tree to value options, this tree would not be compatible with the observed option smile. The third possibility is, therefore, to seek terminal probabilities that are closest to the tree-based probabilities, yet that also correspond to the given options. Formally, let the risk-neutral probabilities, which are compatible with actual options, be denoted by P_j . These risk-neutral probabilities

may then be obtained by solving an optimization such as

$$\min_{P_j} \sum_{j=1}^m (P_j - P'_j)^2,$$

subject to

$$\begin{cases} \sum_{j=1}^m P_j = 1, & P_j \geq 0, \quad j = 0, \dots, m, \\ S_t = e^{-r\tau} \sum_{j=1}^m S_j P_j, \\ C_{\tau,i} = e^{-r\tau} \sum_{j=1}^m P_j (S_j - K_i)^+, \quad i = 1, \dots, N. \end{cases}$$

The first constraint provides conditions so that the values P_j define probabilities. The second condition corresponds to the martingale condition. Finally, the last condition states that the probabilities P'_j should be compatible with existing option prices. Various metrics can be used to define the distance between the probabilities P_j and P'_j . The idea of using other metrics is expressed in Rubinstein (1994) and tested in Jackwerth and Rubinstein (1996). For instance we could choose the following distances

$$\begin{aligned} \min_{P_j} \sum_{j=1}^m \frac{(P_j - P'_j)^2}{P'_j}, \\ \min_{P_j} \sum_{j=1}^m |P'_j - P_j|, \end{aligned}$$

or

$$\min_{P_j} - \sum_{j=1}^m P_j \log \left(\frac{P_j}{P'_j} \right).$$

The first measure is a metric based on percentage deviations. The second measure is based on absolute deviations. The third measure corresponds to seeking the maximum entropy. We discuss this type of measure in the following subsection.

11.5.3 Maximum entropy principle

A method based on the principle of maximum entropy, ME, was introduced to finance by Buchen and Kelly (1996) and Stutzer (1996). A first application to derivative valuation is by Stutzer (1996). See also Chapter 5 for an application of entropy techniques for modeling the distribution of returns.

The general definition of entropy is

$$E = - \int_0^{+\infty} q(x) \log(q(x)) dx, \quad (11.15)$$

where as before $q(\cdot)$ represents the RND. The entropy can be viewed as a metric, whose maximization will give the RND with the *maximum* possible information content.⁹

The constraints can be written as

$$\int_0^{+\infty} q(x)dx = 1, \quad (11.16)$$

$$\int_0^{+\infty} c_i(x)q(x)dx = C_i/e^{-r\tau}, \quad i = 1, \dots, N, \quad (11.17)$$

$$\int_0^{+\infty} xq(x)dx = S_t. \quad (11.18)$$

Equation (11.16) insures that q is a density. Equation (11.17) relates the RND to the i th call option. The last equation is the martingale restriction.

Maximization of the entropy (11.15) subject to the constraints (11.16)–(11.18) is done by maximization of a Hamiltonian where the multipliers will be written as λ_i . Buchen and Kelly (1996) show that the RND is equal to

$$q(x) = \frac{\exp\left(\sum_{i=0}^N \lambda_i c_i(x)\right)}{\int_0^{+\infty} \exp\left(\sum_{i=0}^N \lambda_i c_i(x)\right) dx}.$$

Clearly, the multipliers must be estimated numerically. This may be elegantly achieved following Agmon, Alhassid, and Levine (1979a, 1979b). See also Chapter 5 for more explanations on how to perform this estimation.

11.5.4 Kernel regression

Aït-Sahalia and Lo (1998) propose a further generalization of the approach developed by Shimko (1993). They propose to solve

$$\min_{C_i \in \Gamma} \sum_{i=1}^N (C_i - C(Z_i))^2,$$

where C_i is the option price, Z_i is the set of characteristics of the option, $Z_i = (S_{i,t}, K_i, \tau_i, r_{i,t}^\tau)'$, and $C(Z_i)$ is a twice-continuously differentiable function of Z_i . The statistical procedure adopted by Aït-Sahalia and Lo (1998) is non-parametric kernel regression. It consists in estimating the conditional expectation of $C(\cdot)$ conditioned on Z without parameterizing the function $C(\cdot)$. The strength of this approach lies in the fact that it incorporates the information contained in the characteristics Z_i without specifying any functional form.

⁹ Entropy is an intuitively appealing method because its maximization yields the uniform density if the only thing that we know is that the support is finite. If the mean is known, the method yields the exponential and if, furthermore, the variance is known, it gives the normal density.

The conditional expectation of C is obtained using the Nadaraya-Watson kernel estimator

$$\hat{C}(Z) = E[C|Z] = \frac{\sum_{i=1}^N h\left(\frac{Z-Z_i}{u}\right) C_i}{\sum_{i=1}^N h\left(\frac{Z-Z_i}{u}\right)},$$

where u is the bandwidth and h the kernel to be specified. The idea of this estimator is that the price of an option with the set of characteristics Z is given by a weighted average of the observed prices C_i s with more weight given to the options whose characteristics Z_i s are closer to the characteristics Z of the option to be priced.

Several technical issues have to be addressed for implementing this approach. Since the problem, as specified here, is four-dimensional, we use a multivariate kernel as the product of four univariate kernels, as in

$$\hat{C}(S_t, K, \tau, r_t^\tau) = \frac{\sum_{i=1}^N h_S\left(\frac{S_i - S_{i,t}}{u_S}\right) h_K\left(\frac{K - K_i}{u_K}\right) h_\tau\left(\frac{\tau - \tau_i}{u_\tau}\right) h_r\left(\frac{r_t^\tau - r_{i,t}^\tau}{u_r}\right) C_i}{\sum_{i=1}^N h_S\left(\frac{S_i - S_{i,t}}{u_S}\right) h_K\left(\frac{K - K_i}{u_K}\right) h_\tau\left(\frac{\tau - \tau_i}{u_\tau}\right) h_r\left(\frac{r_t^\tau - r_{i,t}^\tau}{u_r}\right)}.$$

The choice of the kernels $h(\cdot)$ and of the bandwidth u depends on the degree of smoothness required for estimating the RND. Then, the RND is given by

$$q(S_T) = e^{-r\tau} \left. \frac{\partial^2 \hat{C}(S_t, K, \tau, r_t^\tau)}{\partial K^2} \right|_{K=S_T}.$$

There is a considerable difference between this approach and the ones previously developed in this section. Most prior approaches are based on the estimation of the RND for each cross-section of options. Consequently, they provide RNDs that are consistent with option prices at each date, but they are not necessarily consistent over time. In contrast, the nonparametric kernel regression provides an estimator that is based on both cross-sectional and time-series option prices. Therefore, this is a fixed function of certain variables (the characteristics in Z). As a consequence, this approach is consistent over time, but may provide a poor fit for certain dates.

11.6 Comparison of various methods

Several papers compare the performance of various methods to extract RND. Each one has advantages and drawbacks. Jondeau and Rockinger (2000) use exchange rates and compare the mixture of log-normals, the Edgeworth and Gram-Charlier approximation, as well as the single-jump and stochastic volatility model of Heston. They find that, up to a certain extent, all the methods provide similar RNDs. They conclude that during normal periods, a mixture of log-normals provides rather good results but that this should, however, be complemented by the estimation of a model that allows for jumps. During a troubled period, the jump model performs better.

Coutant, Jondeau, and Rockinger (2001) use interest rates. In that paper, they compare the mixture of log-normals, the Gram-Charlier expansion and the entropy-based method. They compare the various methods on their speed, robustness of estimation, and ease of implementation. Again they find that all methods provide similar densities that are all rather distant from the benchmark that would hold under the assumption of log-normality. For interest rates, the method based on Gram-Charlier expansion appears to provide rather stable results.

Bliss and Panigrahi (2002) compare the robustness of two models, the mixture of two log-normal distributions and a smoothed implied volatility smile model, in the spirit of Shimko (1993). They investigate the consequences of different potential sources of error in the option prices – such as non-synchronicity, lack of liquidity or price discreteness – on the estimation of the RND. They find that the estimations of the RNDs are very unstable in the tails and that great care should be taken if one computes higher moments, such as skewness and kurtosis, from RNDs. This type of finding also appears to highlight the limitations of RND to provide tail probabilities as one may wish to use for value at risk purposes. They also perform a comparison between their smoothed spline method and the mixture of log-normal distributions. They find that their smoothed spline method performs better than the mixture of log-normal distributions.

We now compare several techniques to estimate RNDs, using the same data on CAC 40 options as in Section 11.1. The first technique is the numerical approximation of the RND based on the second derivative of option prices with respect to the strike price, as suggested by Breeden and Litzenberger (1978). The second RND is obtained using the Generalized Beta distribution of Bookstaber and McDonald (1978). The third RND is obtained using a mixture of two log-normal distributions, following Melick and Thomas (1997). Finally, we consider the Edgeworth expansion around the log-normal distribution of Jarrow and Rudd (1982). The various resulting RNDs are plotted in Figures 11.6 to 11.9, with a comparison with the BSM log-normal distribution.

The first striking result is that the numerical approximation of the RND (along the lines of Breeden and Litzenberger) yields very unstable results, with some negative probabilities. This kind of very disappointing results has been the basis for the development of nonparametric approaches described above. The RNDs based on the generalized beta distribution, the mixture of log-normals and on Edgeworth expansion are more in line. They are both consistent with a negative skewness and with tails that are fatter than the plain vanilla log-normal distribution.

To further compare the different models, we check the statistical properties of the various RNDs, displayed in Table 11.1. First, we check how the constraints imposed by log-normality on the third and fourth moments can bias the variance estimates. Second, it is tempting to compare the estimates of the skewness and the kurtosis obtained under the different RNDs.

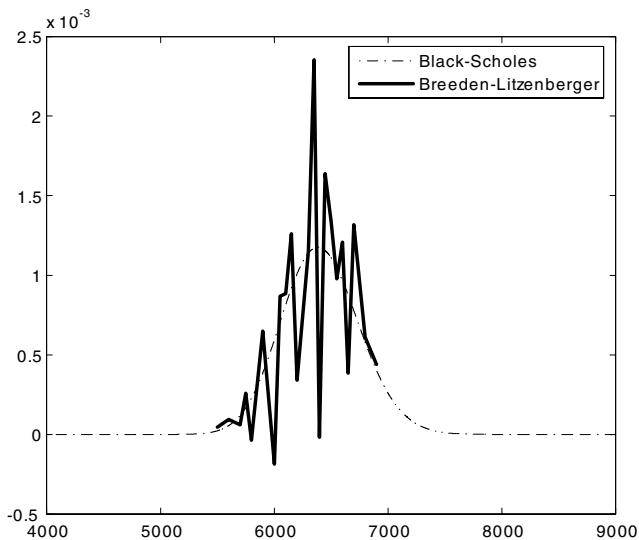


Fig. 11.6. Estimation of the RND using the Breeden and Litzenberger approach.

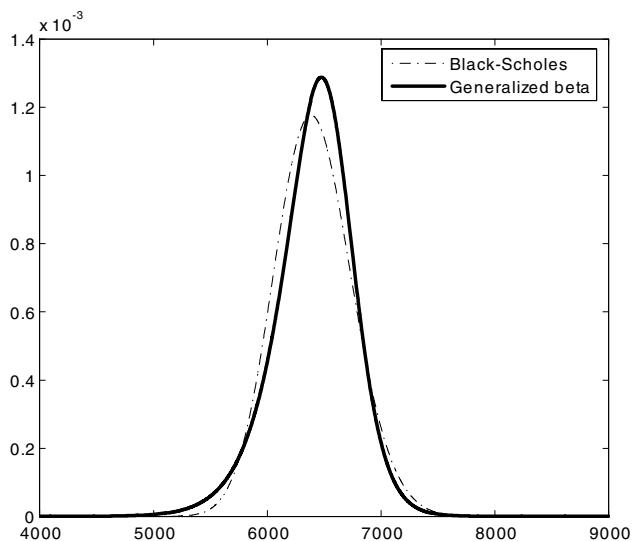


Fig. 11.7. Estimation of the RND using the Generalized Beta distribution.

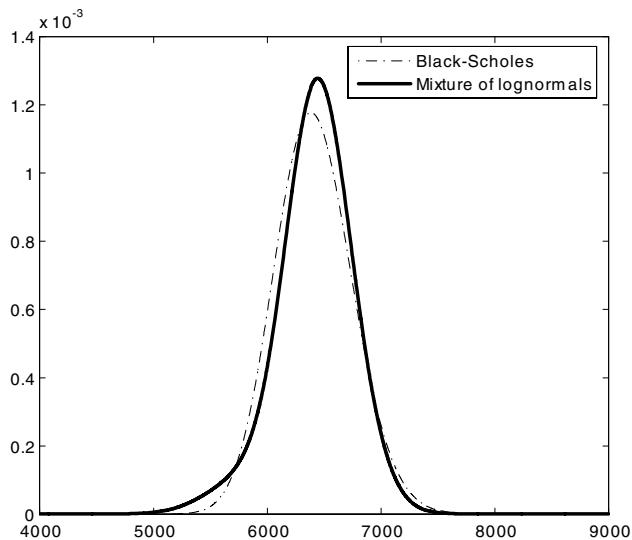


Fig. 11.8. Estimation of the RND using the Mixture-of-Lognormal approach.

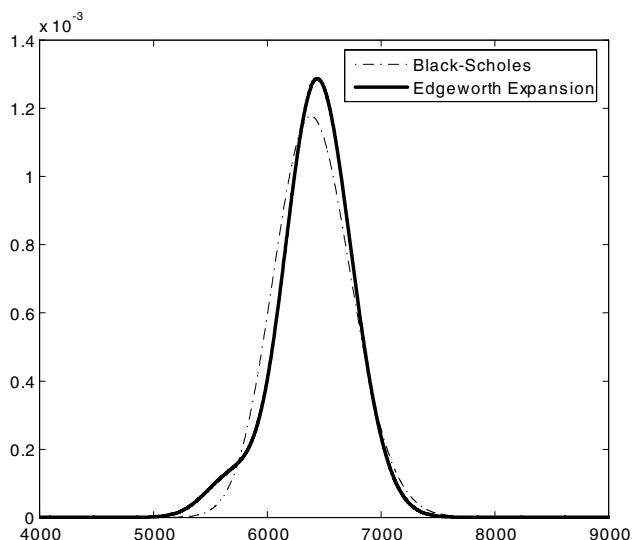


Fig. 11.9. Estimation of the RND using the Edgeworth Expansion approach.

Table 11.1. Moments of the RNDs obtained with various estimation methods

	Std dev.	Skewness	Kurtosis	MSE	ARE
Black and Scholes	340.1	0.159	3.045	3991.42	0.114
Generalized beta dist.	348.1	-0.468	4.079	107.87	0.024
Mixture of log-normal dist.	347.1	-0.474	3.907	61.63	0.018
Edgeworth Expansion	346.9	-0.438	3.650	91.70	0.016

As far as volatilities are concerned, we see the bias implied by the log-normality assumption: The volatility induced by the BSM log-normal model appears systematically smaller than the one we obtain with the other approaches. Otherwise, we observe substantial homogeneity in the volatilities given by the other models. The estimates of skewness and kurtosis are much more divergent, because at this level the specifics of the different models matter. The log-normal model is less interesting from this point of view, because on theoretical grounds it does not allow for asymmetry or fat tails. First, we observe that skewness as well as kurtosis are generally far from what is obtained under log-normality: Skewness is strongly negative and excess kurtosis is rather large. The skewness obtained from semi-nonparametric models is systematically lower than the one obtained with other models, although this difference is small. The graphs of the RND corroborate these findings.

Table 11.1 also reports the mean squared errors and the absolute relative errors for the various models. All more sophisticated models provide a much better fit than the BSM model. Yet, we notice that mixture of log-normals provides the best fit of the data at hand. Since it allows a larger number of parameters, it is able to capture the feature of the actual distribution more precisely than the other ones. Clearly, this result may change for other periods and other data.

Once such RNDs are estimated, it is possible to deduce a lot of different information from them. For instance, we can compute confidence intervals around the expected value, the evolution of which is indicative of market participants' perceptions of the future. Therefore, RNDs are useful tools for an investor who needs to measure how markets are thought to evolve through time. RNDs can also provide an objective measure of expected extreme variations in the underlying asset's price, which is useful for risk management. There are, however, two important caveats. First, as it has been highlighted in several empirical papers, the use of option prices for inferring the RND of the underlying asset on a day-to-day basis is somewhat heroic. For many markets, the number of observations is excessively close to the number of unknown parameters. In those cases, the reliability of the estimates is questionable. Second, because of a possibly time-varying risk premium, RNDs differ from real-life probability distributions. Since the RND is evaluated under the risk-neutral probability, the actual distribution also incorporates information concerning the risk aversion of investors. We will elaborate on this point in the Section 11.7. As a general conclusion to this section, we may notice, as in

Jackwerth (1999), that “there is a danger in reading too much information in the graphs of risk-neutral distributions.”

11.7 Relationship with real probability

In Chapter 10, by using dynamic hedging arguments and by solving the Black-Scholes partial differential equations (PDE), we derive the risk-neutral distribution for options pricing. In this particular well-known case, the real and the risk-neutral distributions are both log-normal and they share the same shape parameter. They only differ in terms of location. In other cases, when markets are incomplete, the transformation between real and risk-neutral distributions is nontrivial. There are various techniques for performing such a transformation, of which the PDE approach is an example. Other techniques include the changing measure approach, which is mathematically demanding, and the general equilibrium approach, which is more popular among economists. It is this general equilibrium approach that is adopted by the second generation of RND with a focus on relating the real as well as the risk-neutral distributions.

We now provide a simple introduction to this general equilibrium approach to asset pricing. The RND and the real distributions are shown to be linked via the pricing kernel. The pricing kernel is a function of utility, and the utility theory permeates classical finance models. The BSM model in Chapter 10 could have been derived along the steps of Brennan (1979). He shows, in an equilibrium setting, that the risk-neutral density is related to the pricing kernel and to the objective probability. Hence, the knowledge of two of the three components yields the third one.

11.7.1 The link between RNDs and objective densities

Assume a single period economy where there is only one representative investor, whose objective is to maximize utility derived from consumptions. Let C_0 and C_1 denote the current and terminal consumptions, S as the amount of saving or investment in risky assets, and W_0 denote the current wealth. In a single period model, all the savings will be consumed at the end of the period. In this framework, we have

$$\begin{aligned} C_0 &= W_0 - S, \\ C_1 &= S(1 + R), \\ R &= \frac{X - P_0}{P_0}, \end{aligned}$$

where R is the risky asset return, and X is the uncertain payoff of asset whose current price is P_0 . The investor’s decision is to determine the optimum level of S so that the total utility to be derived from current and expected terminal consumptions are maximized

$$\max_S U(C_0) + \beta E[U(C_1)].$$

The parameter β represents the discount factor, with $\beta < 1$.

The first-order conditions of the optimization problem state that

$$\begin{aligned} -U'(W_0 - S) + \beta E[(1+R)U'(S(1+R))] &= 0, \\ \beta E\left[(1+R)\frac{U'(S(1+R))}{U'(W_0 - S)}\right] &= 1, \\ \beta E\left[X\frac{U'(C_1)}{U'(C_0)}\right] &= P_0. \end{aligned} \quad (11.19)$$

Equation (11.19) must hold for any asset, including that of a discount bond paying \$1 whatever the state of the nature tomorrow. Hence, its price must also satisfy condition (11.19)

$$\beta E\left[\frac{U'(C_1)}{U'(C_0)}\right] = e^{-r}. \quad (11.20)$$

Dividing (11.19) by (11.20), we get

$$P_0 = e^{-r} \int X \frac{U'(C_1(X))/U'(C_0)}{E[U'(C_1)/U'(C_0)]} f(X) dX = e^{-r} \int X \psi(X) f(X) dX.$$

The quantity $\psi(X)$ is known as the *intertemporal marginal rate of substitution*, the *stochastic discount factor* or the *pricing kernel*. Now, we introduce

$$q(X) = \psi(X) f(X), \quad (11.21)$$

and note that $q(X) > 0$ and $\int q(X) dX = 1$. Hence, $q(X)$ may be interpreted as a probability measure. More precisely, $q(X)$ is the RND and $f(X)$ is the objective distribution. Since we may write $P_0 = e^{-r} \int X q(X) dX$, we discount an expected pay-off with the risk-free rate and thus $q(X)$ has the interpretation of a RND. It follows that if we have two of the three elements, the third element may be deduced from the other two. This opens the door for much research.

Now, if we rewrite (11.21) as

$$q(X) = \lambda \times U'(C_1(X)) \times f(X),$$

take log on both sides

$$\log(q(X)) = \log(\lambda) + \log(U'(C_1(X))) + \log(f(X)), \quad (11.22)$$

and differentiate this equation, we get

$$\frac{q'(X)}{q(X)} = \frac{U''(C_1(X))}{U'(C_1(X))} + \frac{f'(X)}{f(X)}.$$

These computations show that the magnitude of absolute and relative risk aversions, defined respectively as

$$A = -\frac{U''(C_1(X))}{U'(C_1(X))},$$

$$R = -X \frac{U''(C_1(X))}{U'(C_1(X))},$$

can easily be derived once $q(X)$ and $f(X)$ are known.

11.7.2 Empirical findings

A few papers have extracted the pricing kernel by matching the real and the risk-neutral densities. Among these papers, we may mention Aït-Sahalia and Lo (2000), Jackwerth (2000), and Rosenberg and Engle (2002). Their common objective is to investigate the empirical properties of the pricing kernel and/or the risk-aversion function.

The two earlier papers, Aït-Sahalia and Lo (2000) and Jackwerth (2000), assume that the real distribution is somewhat static, estimating it using kernel estimation for SP500 data for 1993. These papers do not impose any structure on the pricing kernel and find that it is not a monotonic function of the asset price. Indeed, there is a hump in the mid-range of strike prices. They also obtain a U-shape for the risk-aversion parameter as a function of the strike prices. It means that investors are more risk averse when the index return takes large (either positive or negative) values. Another important result reported in Aït-Sahalia and Lo (2000) is that the risk aversion varies a lot depending on the level of the SP500, with an average level close to 12.7. This evidence suggests that the often adopted CRRA function for preferences is mis-specified.

Rosenberg and Engle (2002) use a version of asymmetric GARCH to estimate the real distribution for each day of the sample period. Using SP500 index options for the period between 1991 to 1995, they find a similar hump in the pricing kernel. The average level of the risk aversion (7.4) is smaller than the one reported by Aït-Sahalia and Lo (2000) with a range between 2.3 and 12.6. Then, they relate the risk aversion parameter (of the power pricing kernel function) to business conditions, such as the change in the credit spread or in the term spread. They find the empirical risk aversion to be counter-cyclical, supporting the habit persistence model of Campbell and Cochrane (1999). It is worth noting that Rosenberg and Engle (2002) do not produce the RND from option data but derive the RND that best fits option prices from the product of pricing kernel and the simulated underlying asset distribution.

The relationship between the risk aversion and macroeconomic variables is further investigated by Beber and Brandt (2003). In particular, they consider the effect of macroeconomic news to risk aversion. They show that good news generally leads investors to be less risk averse.

Structural Option Pricing

In option pricing, the first sign that asset returns may not have a normal distribution surfaced as Black-Scholes implied volatility smile discussed in the previous chapter. The volatility smile is clear evidence that options are priced as if the tails of the distribution are much fatter than those of the normal distribution. There have been many extensions to the Black-Scholes model to tackle the smile problem. Among these efforts, the most widely studied is stochastic volatility option pricing reflecting a wide recognition that volatility is changing through time and may attract a risk premium. The negative relationship between the price and volatility processes has also been frequently used as an explanation for volatility skew. The classical stochastic volatility option pricing model is Heston's (1993), where characteristic function, a special case of the Fourier transform, was used for the first time in the option pricing literature to produce a semi-closed form solution. Development in option pricing theory did not stop here. Later, Bates (1996, 2000) includes normal distributed jumps in a model with stochastic volatility. Kou (2002) uses double-exponential jumps but does not permit stochastic volatility. Duffie, Pan, and Singleton (2000) produces the most general affine diffusion model where jumps are permitted at both mean and volatility levels.

12.1 Stochastic volatility model

A widely used stochastic volatility model is due to Heston (1993). It is based on the following stock price and variance dynamics

$$dS_t = \mu S_t dt + S_t \sqrt{v_t} dZ_{1,t}, \quad (12.1)$$

$$dv_t = \kappa (\theta - v_t) dt + \sigma \sqrt{v_t} dZ_{2,t}, \quad (12.2)$$

where $\kappa, \theta, \sigma > 0$ are constant parameters. The two Brownian motions, $Z_{1,t}$ and $Z_{2,t}$, are correlated, i.e., $\text{Corr}[dZ_{1,t}, dZ_{2,t}] = \rho dt$. The dynamics of the stock price S_t in (12.1) is a geometric Brownian motion with time varying

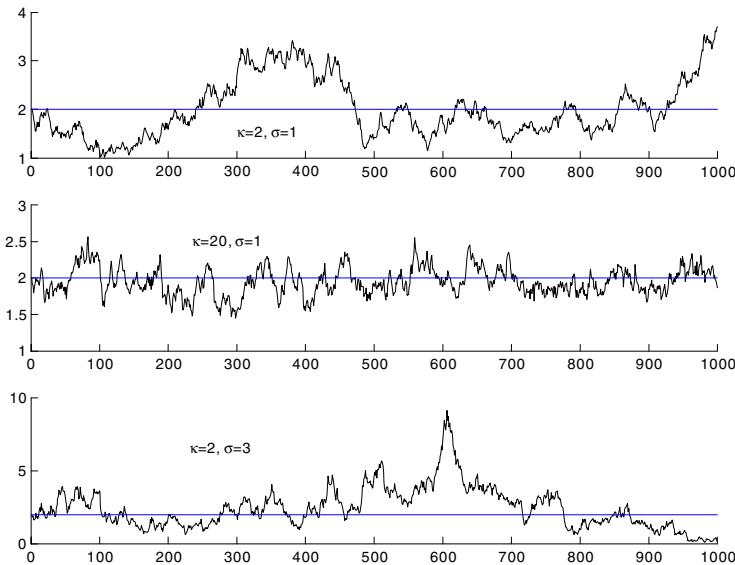


Fig. 12.1. Simulated square root process with $\Delta t = 1/365$, $v_0 = 2$, and $\theta = 2$. The other parameters are indicated in the panels.

volatility. The variance v_t in (12.2) follows a square root process (also known as the Feller process or the Cox-Ingersoll-Ross process).¹ As we will show below, the parameter θ corresponds to the long-run average of v_t , and κ controls the speed by which v_t returns to its long-run mean.

12.1.1 The square root process

To gain an understanding of the square root process, we simulated some trajectories of this process. For the simulations, we assume a finite time step, Δ , and consider the discretization

$$v_{t+\Delta} = (1 - \kappa\Delta)v_t + \kappa\theta\Delta + \sigma\sqrt{v_t}\sqrt{\Delta}\varepsilon_t,$$

where $\varepsilon_t \sim \mathcal{N}(0, 1)$. We select $\Delta = 1/365$, $v_0 = 2$, and $\theta = 2$, and produced Figure 12.1 for various values of κ and σ . The first trajectory was obtained with $\kappa = 2$ and $\sigma = 1$. The second trajectory corresponds to $\kappa = 20$ and $\sigma = 1$. We notice that both trajectories oscillate around the long-run mean

¹ It seems that it is Feller (1950) who first investigated this process, in a study of parabolic second-order partial differential equations. Feller uses Laplace transform techniques, very much related to characteristic function techniques, to derive the conditional and unconditional densities. Cox, Ingersoll, and Ross (1985) introduced this process to finance as a model for interest rates.

at $\theta = 2$. We also notice that the second trajectory crosses the central value more frequently than the first trajectory. The third trajectory is obtained with $\kappa = 2$ and $\sigma = 3$. In this case, the amount of crossings of the mean level are about the same as in the first trajectory, but the variability of the third trajectory is more pronounced. We conclude from these simulations that the square root process v_t is a continuous time first-order autoregressive process where v_t is constantly pulled toward a mean level, θ , at a speed controlled by κ , and has its variability controlled by σ .

12.1.2 Solving the PDE based on characteristic function

From the dynamic of the square root process in (12.2), how do we derive the probability that v_T will reach some domain B , given that it is currently at v_t ? In other words, how do we evaluate the following integral

$$\int_{v_T \in B} p(v_T, T | v_t, t) dv_T, \quad (12.3)$$

given (12.2)? As shown in Chapter 13, the transition probability must satisfy

$$\frac{1}{2} \sigma^2 v_t \frac{\partial^2 p}{\partial v_t^2} + \kappa(\theta - v_t) \frac{\partial p}{\partial v_t} + \frac{\partial p}{\partial t} = 0. \quad (12.4)$$

Note that the parameters of this PDE are affine functions, i.e., of the type “ $a + bv_t$ ”. This linearity suggests that a solution exists. The main objective now is to find a characteristic function (CF) that will satisfy (12.4) and the boundary conditions. Once we have the CF, following Chapter 15 we can work out the *pdf* or the *cdf* for solving the integral in (12.3). Let us now introduce the CF²

$$\phi_{v_t, t}(u) \equiv E[e^{iuv_T} | v_t, t] = \int_{v_T} e^{iuv_T} p(v_T, T | v_t, t) dv_T. \quad (12.5)$$

Given the linearity of the problem, we could guess that the solution to (12.5) is also a linear function of the form

$$\phi_{v_t, t}(u) = \exp [A(\tau, u) + B(\tau, u) v_t], \quad (12.6)$$

where A and B are both functions of $\tau = T - t$ and u , and must satisfy the boundary condition

$$\phi_{v_t, t}(u) \xrightarrow[t \rightarrow T]{} \exp(iuv_T),$$

implying that as $\tau \rightarrow 0$, $A(0, u) = 0$, and $B(0, u) = iu$. Then from (12.6), we have

² We make a slight abuse of notation using the symbol v_t as a random variable or a realization.

$$\begin{aligned}\frac{\partial \phi_{v_t,t}}{\partial v_t} &= B\phi_{v_t,t}, \\ \frac{\partial^2 \phi_{v_t,t}}{\partial v_t^2} &= B^2\phi_{v_t,t}, \\ \frac{\partial \phi_{v_t,t}}{\partial t} &= -\frac{\partial A}{\partial t}\phi_{v_t,t} - \frac{\partial B}{\partial t}v_t\phi_{v_t,t}.\end{aligned}$$

Substitute these into (12.4). Then noticing that $\phi_{v_t,t} \neq 0$, this term may be simplified. Eventually, we obtain the PDE

$$\frac{1}{2}\sigma^2 v_t B^2 + \kappa(\theta - v_t)B - \frac{\partial A}{\partial t} - \frac{\partial B}{\partial t}v_t = 0. \quad (12.7)$$

Equation (12.7) must be true for all v_t . So by first setting $v_t = 0$ and then $v_t = 1$, we see that the following equations must be true

$$\frac{\partial A}{\partial t} = \kappa\theta B, \quad (12.8)$$

$$\frac{\partial B}{\partial t} = \frac{1}{2}\sigma^2 B^2 - \kappa B, \quad (12.9)$$

again with boundary conditions $A(0) = 0$ and $B(0) = iu$.

Equations (12.8) and (12.9) are called the *Riccati equations*, and are the key feature of problems involving PDE. Inspection of (12.9) shows that it is a second-order ordinary differential equation, ODE, that may be solved directly. Once a solution to (12.9) is obtained, a solution to (12.8) may be found. Solving the pair of Riccati equations (12.8) and (12.9) manually could be non-trivial, and the use of a symbolic solver is highly recommended.³ Substituting the solutions for A and B into (12.6) yields,

$$\phi_{v_t,t}(u) = \left(1 - \frac{iu}{c}\right)^{-\frac{2\kappa\theta}{\sigma^2}} \exp\left(i\frac{ue^{-\kappa\tau}}{1 - iu/c}v_t\right), \quad (12.10)$$

where

$$c = \frac{2\kappa}{\sigma^2 [1 - e^{-\kappa\tau}]} \cdot \quad (12.11)$$

Finally, by invoking the Fourier inversion theorem, described in Chapter 15, we get the transition probability (or the conditional probability)⁴

$$p(v_T|v_t) = ce^{-\alpha-\nu} \left(\frac{\nu}{\alpha}\right)^{q/2} I_q\left[2(\alpha\nu)^{1/2}\right], \quad (12.12)$$

where c is defined as in (12.11), $I_q(\cdot)$ is the modified Bessel function of the first kind of order q and

³ Examples of popular symbolic solvers are Mathematica or Maple.

⁴ All the formulae presented here are also in Cox, Ingersoll, and Ross (1985).

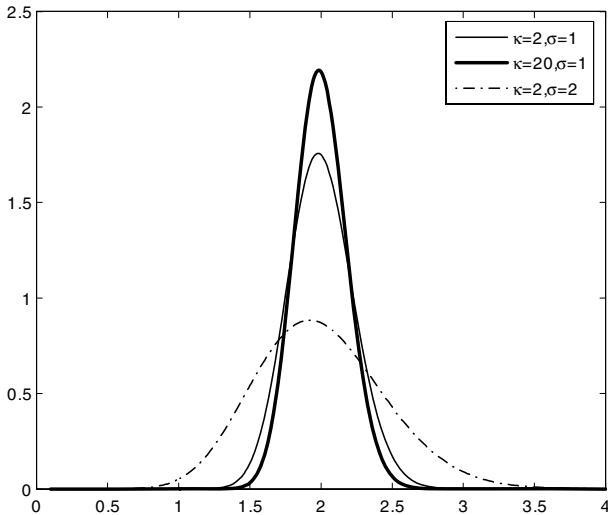


Fig. 12.2. Conditional probability at a 10-day horizon for v_T with parameters corresponding to the simulated trajectories of Figure 12.1. The current value, v_t , is taken as the steady state level, 2.

$$\alpha \equiv cv_t e^{-\kappa\tau}, \quad \nu \equiv cv_t, \quad q \equiv \frac{2\kappa\theta}{\sigma^2} - 1.$$

Inspection of the transition probability in (12.12) reveals that the distribution function is a non-central χ^2 , evaluated at $2cv_t$, with $2q + 2$ degrees of freedom and asymmetry parameter 2α . Next, by differentiating the characteristic function and evaluating it at $u = 0$ yields

$$\begin{aligned} E[v_T|v_t] &= v_t e^{-\kappa\tau} + \theta [1 - e^{-\kappa\tau}], \\ V[v_T|v_t] &= \frac{v_t \sigma^2}{\kappa} [e^{-\kappa\tau} - e^{-2\kappa\tau}] + \frac{\theta \sigma^2}{2\kappa} [1 - e^{-\kappa\tau}]^2. \end{aligned}$$

Cox, Ingersoll, and Ross (1985) note that as $\kappa \rightarrow \infty$, the expected value converges to θ and the variance to 0. This result confirms that κ controls the speed by which v_t returns to its mean value. As $\kappa \rightarrow 0$, the conditional mean is equal to the current value v_t , the variance converges to $\sigma^2 v_t \tau$.

Figure 12.2 presents the density function for $\tau = T - t = 10$. All the other parameters are the same as those used for the simulations of the trajectories in Figure 12.1. Contemplating (12.10), it is easy to see that for $\kappa > 0$, and $T \rightarrow \infty$, the characteristic function becomes

$$\phi_{v_t, t}^\infty(u) = \left(1 - \frac{iu\sigma^2}{2\kappa}\right)^{-\frac{2\kappa\theta}{\sigma^2}}.$$

This characteristic function $\phi_{v_t,t}^\infty(u)$ corresponds to a steady-state distribution. By invoking the Fourier inversion theorem on $\phi_{v_t,t}^\infty(u)$, we obtain the unconditional density

$$f_{v_T}(v_T) = \frac{w^\eta}{\Gamma(\eta)} v_T^{\eta-1} e^{-wv_T},$$

where

$$w = \frac{2\kappa}{\sigma^2} \text{ and } \eta = \frac{2\kappa\theta}{\sigma^2}.$$

We recognize that $f_{v_T}(v_T)$ is the density of a gamma distribution with mean θ and variance $\sigma^2/2\kappa$.⁵

In this section, we have studied the property of a mean reverting square root process that frequently appears in the stochastic volatility literature. We are now ready to tackle the full Heston model as given by (12.1) and (12.2). His model involves two Brownian motions, hence, a multivariate extension of Ito's lemma is required. This lemma is explained in Chapter 13.

12.1.3 A new partial differential equation

Given that the Heston model has two sources of randomness, the bivariate Ito's lemma is used to derive the fundamental partial differential equation. The steps involved in the derivation of the Heston option pricing formula are the same as those in the no-arbitrage derivation for the Black-Scholes formula. Because of the multidimensionality of the problem, two derivative assets are required to obtain a risk-neutral portfolio. Similar to the case of Black-Scholes in Chapter 10, we could consider the sale of a call option C and cover the position with the purchase of δ units of the underlying asset and γ units of a second derivative C_1 written on the same underlying. C_1 differs from C by its maturity, strike and so forth. Let us rewrite, using shorter notations, the system (12.1) and (12.2) as

$$\begin{aligned} dS_t &= \mu_S dt + \sigma_S dZ_{1,t}, \\ dv_t &= \mu_v dt + \sigma_v dZ_{2,t}. \end{aligned}$$

Let $C(S_t, v_t, t)$ denote the price of a call option. We obtain from the bivariate Ito's lemma that the dynamics C may be written as

$$\begin{aligned} dC &= \left[\frac{1}{2} \sigma_S^2 \frac{\partial^2 C}{\partial S_t^2} + \rho \sigma_S \sigma_v \frac{\partial^2 C}{\partial S_t \partial v_t} + \frac{1}{2} \sigma_v^2 \frac{\partial^2 C}{\partial v_t^2} \right. \\ &\quad \left. + \mu_S \frac{\partial C}{\partial S_t} + \mu_v \frac{\partial C}{\partial v_t} + \frac{\partial C}{\partial t} \right] dt + \sigma_S \frac{\partial C}{\partial S_t} dZ_{1,t} + \sigma_v \frac{\partial C}{\partial v_t} dZ_{2,t}. \end{aligned}$$

⁵ More details on the non-central chi-square and the gamma density may be found in Johnson, Kotz, and Balakrishnan (1994).

The value of the portfolio, W_t , consisting of the sale of the call option and the purchase of δ units of the underlying and of γ units of C_1 is $W_t = C - \delta S_t - \gamma C_1$. The dynamic of W_t is

$$\begin{aligned} dW_t &= dC - \delta dS_t - \gamma dC_1 \\ &= \left[\frac{1}{2} \sigma_S^2 \frac{\partial^2 C}{\partial S_t^2} + \rho \sigma_S \sigma_v \frac{\partial^2 C}{\partial S_t \partial v_t} + \frac{1}{2} \sigma_v^2 \frac{\partial^2 C}{\partial v_t^2} \right. \\ &\quad \left. + \mu_S \frac{\partial C}{\partial S_t} + \mu_v \frac{\partial C}{\partial v_t} + \frac{\partial C}{\partial t} - \delta \mu_S \right] dt \\ &\quad - \gamma \left[\frac{1}{2} \sigma_S^2 \frac{\partial^2 C_1}{\partial S_t^2} + \rho \sigma_S \sigma_v \frac{\partial^2 C_1}{\partial S_t \partial v_t} + \frac{1}{2} \sigma_v^2 \frac{\partial^2 C_1}{\partial v_t^2} \right. \\ &\quad \left. + \mu_S \frac{\partial C_1}{\partial S_t} + \mu_v \frac{\partial C_1}{\partial v_t} + \frac{\partial C_1}{\partial t} \right] dt \\ &\quad + \left[\sigma_S \frac{\partial C}{\partial S_t} - \delta \sigma_S - \gamma \sigma_S \frac{\partial C_1}{\partial S_t} \right] dZ_{1,t} \\ &\quad + \left[\sigma_v \frac{\partial C}{\partial v_t} - \gamma \sigma_v \frac{\partial C_1}{\partial v_t} \right] dZ_{2,t}. \end{aligned} \tag{12.13}$$

To obtain risk neutrality, the coefficients of $dZ_{1,t}$ and $dZ_{2,t}$ must be zero, so that the two sources of uncertainty no longer play a role in the portfolio value dynamic. This means that

$$\frac{\partial C}{\partial S_t} = \delta + \gamma \frac{\partial C_1}{\partial S_t}, \tag{12.14}$$

$$\frac{\partial C}{\partial v_t} = \gamma \frac{\partial C_1}{\partial v_t}. \tag{12.15}$$

With these two conditions, the instantaneous change of value of the fully hedged portfolio must be equal to the return on a risk-free investment. Otherwise there will be an arbitrage opportunity. Hence,

$$dW_t = r [C - \delta S_t - \gamma C_1] dt. \tag{12.16}$$

If we equate (12.13) and (12.16), and substitute the values of δ and γ from (12.14) and (12.15), we obtain

$$\begin{aligned} &\left[\frac{1}{2} \sigma_S^2 \frac{\partial^2 C}{\partial S_t^2} + \rho \sigma_S \sigma_v \frac{\partial^2 C}{\partial S_t \partial v_t} + \frac{1}{2} \sigma_v^2 \frac{\partial^2 C}{\partial v_t^2} + r S_t \frac{\partial C}{\partial S_t} \right. \\ &\quad \left. + \mu_v \frac{\partial C}{\partial v_t} + \frac{\partial C}{\partial t} - r C \right] / \frac{\partial C}{\partial v_t} \\ &= \left[\frac{1}{2} \sigma_S^2 \frac{\partial^2 C_1}{\partial S_t^2} + \rho \sigma_S \sigma_v \frac{\partial^2 C_1}{\partial S_t \partial v_t} + \frac{1}{2} \sigma_v^2 \frac{\partial^2 C_1}{\partial v_t^2} + r S_t \frac{\partial C_1}{\partial S_t} \right. \\ &\quad \left. + \mu_v \frac{\partial C_1}{\partial v_t} + \frac{\partial C_1}{\partial t} - r C_1 \right] / \frac{\partial C_1}{\partial v_t}. \end{aligned}$$

Interestingly, the two sides of the equality sign are the same in their expression, differing only in the option to which they apply. Given that the same equation must hold for any type of call option of any maturity and strike price, we notice that each side of the equality must be independent from the type of option that one considers. This suggests that each side will be equal to some function, say $\lambda(S_t, v_t, t)$ that depends on S_t and v_t . This function may be interpreted as a volatility risk premium. Replacing the parameters by their actual units, we see that the fundamental partial differential equation is now

$$0 = \frac{1}{2} S_t^2 v_t \frac{\partial^2 C}{\partial S_t^2} + \rho S_t \sigma v_t \frac{\partial^2 C}{\partial S_t \partial v_t} + \frac{1}{2} \sigma^2 v_t \frac{\partial^2 C}{\partial v_t^2} \\ + [\kappa(\theta - v_t) - \lambda(S_t, v_t, t)] \frac{\partial C}{\partial v} + r S \frac{\partial C}{\partial S} + \frac{\partial C}{\partial t}.$$

Introducing $x = \log(S_t)$ and considering a call, $C(e^x, t)$, we notice that $\frac{\partial C}{\partial x} = \frac{\partial C}{\partial S} \frac{\partial S}{\partial x} = \frac{\partial C}{\partial S} S$ and $\frac{\partial^2 C}{\partial x^2} = S^2 \frac{\partial^2 C}{\partial S^2}$. It follows that we may rewrite the partial differential equation as

$$0 = \frac{1}{2} v_t \frac{\partial^2 C}{\partial x^2} + \rho \sigma v_t \frac{\partial^2 C}{\partial x \partial v_t} + \frac{1}{2} \sigma^2 v_t \frac{\partial^2 C}{\partial v_t^2} \\ + [\kappa(\theta - v_t) - \lambda(x_t, v_t, t)] \frac{\partial C}{\partial v} + r \frac{\partial C}{\partial x} + \frac{\partial C}{\partial t}. \quad (12.17)$$

The boundary conditions that must be satisfied, in the case of a European call option with strike price K are

$$C(S_T, v_t, r, K, T, t) = \max(S_T - K, 0), \quad (12.18)$$

$$C(0, v_t, r, K, T, t) = 0, \quad (12.19)$$

$$\frac{\partial C}{\partial S_t}(\infty, v_t, r, K, T, t) = 1. \quad (12.20)$$

Condition (12.18) determines what the payoff at T should be. Condition (12.19) means that if the value of the underlying becomes worthless, then the option value becomes nil. If the underlying is an equity, we could interpret this as meaning that if the company becomes bankrupt, and ceases to exist, then an option written on it also becomes worthless. Condition (12.20) states that an option that is very much in the money, given that it will be exercised, will have a value that changes exactly by the same amount as the underlying.

In order to go further, a hypothesis concerning the shape of the risk premium $\lambda(x_t, v_t, t)$ needs to be formulated. Depending on the assumptions made, we will get the option pricing formulae in the context of Hull and White (1987, 1988) or Heston (1993).

12.2 Option pricing with stochastic volatility

Several models have been proposed to price options in the context of stochastic volatility. We may mention Hull and White (1987, 1988), Wiggins (1987), Melino and Turnbull (1990), Stein and Stein (1991) and Heston (1993). We now describe two specific models, developed by Hull and White (1987, 1988) and Heston (1993).

12.2.1 Hull and White (1987, 1988)

Hull and White (1987) and many others made a crucial assumption that volatility risk is uncorrelated with aggregate consumption, and hence has zero systematic risk. They then argue that risk that is not systematic will not be priced, i.e., $\lambda = 0$, by invoking an equilibrium model such as Capital Asset Pricing Model (CAPM). From this it follows that $\rho = 0$. With this crucial assumption, they then argued that options can be priced under the risk-neutral valuation principle even if volatility is stochastic. In fact, these assumptions are equivalent to decoupling the volatility equation from the mean equation.

We then have to solve the following problem

$$C(S_t, v_t, t) = e^{-r\tau} \int_{S_T} \max(S_T - K, 0) p(S_T | S_t, v_t) dS_T.$$

In the work of Merton, it is shown that if volatility is constant or time deterministic, then the Black-Scholes formula still holds. The call price is then the same as the Black-Scholes price, using the average variance as volatility input. We may write

$$C(S_t, v_t, t) = C^{BS}(S_t, t, T, \bar{v}),$$

with $\bar{v} = \frac{1}{\tau} \int_t^T v_u du$. This formula cannot be used here since volatility is stochastic. The idea of Hull and White (1987) is to condition the average future volatility at the current level using the Chapman-Kolmogorov identity valid for Markov processes,⁶

$$p(x|y) = \int_z p(x|z) p(z|y) dz.$$

The price of the call option may then be written as

$$\begin{aligned} C(S_t, v_t, t) &= e^{-r\tau} \int_{S_T} \int_{\bar{v}} \max(S_T - K, 0) p(S_T | \bar{v}) p(\bar{v} | v_t) dS_T d\bar{v} \\ &= \int_{\bar{v}} e^{-r\tau} \left[\int_K^\infty (S_T - K) p(S_T | \bar{v}) dS_T \right] p(\bar{v} | v_t) d\bar{v}. \end{aligned} \quad (12.21)$$

⁶ This relation means that the probability of going from y to x equals the sum of all the probabilities of going through some intermediate point z .

The expression in $[.]$ in (12.21) is in fact the Black-Scholes price with the average variance \bar{v} from t to T for one particular sample path of v_t .⁷ Thus, the price of the call option is

$$C(S_t, v_t, t) = \int C^{BS}(S_t, t, T, \bar{v}) p(\bar{v}|v_t) d\bar{v}.$$

The problem with (12.21) is the difficulty in deriving all possible sample paths of v_t and to get the *pdf* $p(\bar{v}|v_t)$. Even without further computation, we can see that the problem becomes more complex when $\rho \neq 0$, which means that when we think of the evolution of v_t we must also keep track of where S_t has evolved. In Hull and White (1987), it is assumed that $\rho = 0$. The computation of (12.21) is then based on a Taylor series expansion around the Black-Scholes price.

Clearly, these computations are a useful step towards obtaining a general solution in which the Brownian motions are correlated and in which agents price volatility risk. This leads us to Heston's model presented in the following section.

12.2.2 Heston (1993)

Heston solved the partial differential (12.17) in its full generality. His solution is based on several innovating insights. To begin, first let θ denote the parameter vector, then the price of a European call option with terminal payoff function $\max(S_T - K, 0)$ is

$$C(S_t, v_t, K, T, t, \theta) = e^{-r\tau} \int_0^\infty \max(S_T - K, 0) p(S_T | S_t, v_t) dS_T$$

and with a change in variable⁸

$$\begin{aligned} C(S_t, v_t, K, T, t, \theta) &= e^{-r\tau} \int_{-\infty}^{\infty} \max(e^{x_T} - K, 0) p(x_T | x_t, v_t) dx_T \\ &= e^{-r\tau} \int_{\log(K)}^{\infty} e^{x_T} p(x_T | x_t, v_t) dx_T \\ &\quad - e^{-r\tau} K \int_{\log(K)}^{\infty} p(x_T | x_t, v_t) dx_T. \end{aligned} \tag{12.22}$$

Heston then uses the martingale relation

⁷ In this section, we make a slight abuse of notation. In the Black-Scholes formula, the parameter is volatility, measured as a standard deviation. To be precise, the square root of the averaged variance should be used in the evalution of the formula.

⁸ We have kept the notation p for the transition probability even though a change of variable has taken place. This is appropriate because the two variables are exactly one-to-one.

$$S_t = e^{x_t} = e^{-r\tau} \int_{-\infty}^{\infty} e^{x_T} p(x_T | x_t, v_t) dx_T, \quad (12.23)$$

to rewrite the call option price as

$$\begin{aligned} C(S_t, v_t, K, T, t, \theta) &= S_t \int_{\log(K)}^{\infty} \frac{e^{x_T} p(x_T | x_t, v_t)}{\int_{y_T=-\infty}^{\infty} e^{y_T} p(y_T | x_t, v_t) dy_T} dx_T \\ &\quad - e^{-r\tau} K \int_{\log(K)}^{\infty} p(x_T | x_t, v_t) dx_T, \end{aligned} \quad (12.24)$$

with the first integrand in (12.24) being positive and integrating up to one. The first integrand therefore defines a new probability measure that we denote by q below

$$\begin{aligned} C(S_t, v_t, K, T, t, \theta) &= S_t \int_{\log(K)}^{\infty} q(x_T | x_t, v_t) dx_T \\ &\quad - e^{-r\tau} K \int_{\log(K)}^{\infty} p(x_T | x_t, v_t) dx_T \\ &= S_t P_1(X_T > \log(K) | x_t, v_t) \\ &\quad - e^{-r\tau} K P_2(X_T > \log(K) | x_t, v_t). \end{aligned} \quad (12.25)$$

If ϕ is the characteristic function of some density, then from the Fourier inversion theorem in Chapter 15, we have

$$\begin{aligned} F(x) &= \frac{1}{2} - \frac{1}{\pi} \int_0^{\infty} \text{Re} \left[\frac{\exp(-iux)\phi(u)}{iu} \right] du, \\ \Pr[X > x] &= 1 - F(x) = \frac{1}{2} + \frac{1}{\pi} \int_0^{\infty} \text{Re} \left[\frac{\exp(-iux)\phi(u)}{iu} \right] du. \end{aligned}$$

Hence, we can write P_1 and P_2 in (12.25) as

$$P_j = \frac{1}{2} + \frac{1}{\pi} \int_0^{\infty} \text{Re} \left[\frac{\exp(-iu \log(K))\phi_{j;x_t,v_t,t}(u)}{iu} \right] du,$$

with $j = 1, 2$.⁹

The characteristic function $\phi_{j;x_t,v_t,t}$ for $j = 1$ and 2, is associated with the transition probabilities q and p , respectively. If the characteristic function associated with the two transition probabilities were known, the price of the derivative asset could be numerically computed in semi-closed form. We turn to this issue now.

⁹ See Chapter 15 for a detailed discussion of the inverse Fourier transform.

12.2.3 Characteristic function of the SV model

In order to compute the characteristic function corresponding to the two probabilities p and q , we again use the approach of Feynman-Kac used in Chapter 10 to obtain the solution to the Black-Scholes model. Clearly, for p the partial differential (12.17) provides all the necessary information. Unfortunately, the partial differential equation corresponding to q is not known at this stage. We, therefore, need first to derive the partial differential equation corresponding to q .

To do so, we inject the general solution $C = S_t P_1(x_t, v_t) - e^{-r\tau} K P_2(x_t, v_t)$ from (12.25) into the partial differential (12.17) and then regroup terms in P_1 and P_2 , respectively. Furthermore, Heston (1993) makes the important assumption that the volatility risk premium is a linear function of v_t such that $\lambda(S_t, v_t, t) = \lambda v_t$. With this assumption in place, substitute all the partial derivatives into (12.17) and introduce the notations $a = \kappa\theta$, $b_1 = \kappa + \lambda - \sigma\rho$, $b_2 = \kappa + \lambda$, $u_1 = +1/2$, $u_2 = -1/2$ to produce two partial differential equations of the form

$$\begin{aligned} & (r + u_j v_t) \frac{\partial P_j}{\partial x} + \frac{1}{2} v_t \frac{\partial^2 P_j}{\partial x^2} + \sigma \rho v_t \frac{\partial^2 P_j}{\partial x \partial v_t} \\ & + (a - v_t b_j) \frac{\partial P_j}{\partial v_t} + \frac{1}{2} \sigma^2 v_t \frac{\partial^2 P_j}{\partial v_t^2} + \frac{\partial P_j}{\partial t} = 0, \end{aligned} \quad (12.26)$$

again indexed by $j = 1, 2$ corresponding to the two cdfs P_1 and P_2 .

From here on, the usual Feynman-Kac technique applies. Recall our objective is to obtain the characteristic function $\phi_{j;x_t,v_t,t}(u)$ that satisfies (12.26).

Given that the coefficients involved in the partial differential equation (12.26) are linear, it is possible to guess a solution of the form

$$\phi_{j;x_t,v_t,t}(u) = \exp [A_j(\tau, u) + B_j(\tau, u)v_t + C_j(\tau, u)x_t], \quad (12.27)$$

where A_j , B_j , and C_j are functions of τ and u . Since we want $\phi_{j;x_t,v_t,t}(u)$ to be a characteristic function, the boundary condition is

$$\phi_{j;x_t,v_t,t}(u) \xrightarrow[t \rightarrow T]{} \exp(iux_T).$$

Hence, $A_j(0, u) = 0$, $B_j(0, u) = 0$ and $C_j(0, u) = iu$. Given that $\phi_{j;x_t,v_t,t}(u)$ must satisfy the PDE (12.26), we substitute all the appropriate partial derivatives of (12.27) into (12.26) to produce

$$\begin{aligned} & (r + u_j v_t) C_j + \frac{1}{2} v_t C_j^2 + \rho \sigma B_j C_j v_t + (a - b_j v_t) B_j + \frac{1}{2} \sigma^2 v_t B_j^2 \\ & = \frac{\partial A_j}{\partial t} + \frac{\partial B_j}{\partial t} v_t + \frac{\partial C_j}{\partial t} x_t. \end{aligned}$$

Regrouping the terms and observing that the relation must hold for all v_t , and x_t , we get, by setting v_t and x_t equal to the three pairs of values $(v_t, x_t) =$

$(0, 0)$, $(v_t, x_t) = (1, 0)$, and $(v_t, x_t) = (0, 1)$, three ODEs. For C_j , $\partial C_j / \partial t = 0$ and by using the boundary condition this gives the solution $C_j(\tau, u) = iu$. A system of two ODEs, or Riccati equations, remains

$$\begin{aligned}\frac{\partial B_j}{\partial t} &= -\frac{1}{2}u^2 + \rho\sigma iuB_j + \frac{1}{2}\sigma^2 B_j^2 - b_j B_j + iu^2, \\ \frac{\partial A_j}{\partial t} &= aB_j + riu.\end{aligned}$$

Heston (1993) is the first to provide a solution to these Riccati equations. He sets $x_t = \log(S_t)$ and obtains for $j = 1, 2$

$$\begin{aligned}C(S_t, v_t, K, T, t, \theta) &= S_t P_1 - Ke^{-r\tau} P_2, \\ P_j &= \frac{1}{2} + \frac{1}{\pi} \int_0^\infty \operatorname{Re} \left[\frac{\exp(-iu \log(K)) \phi_{j;x_t,v_t,t}(u)}{iu} \right] du, \\ \phi_{j;x_t,v_t,t}(u) &= \exp[A_j(\tau, u) + B_j(\tau, u)v_t + iux_t], \\ A_j(\tau, u) &= rui\tau + \frac{\kappa\theta}{\sigma^2} \left[(b_j - \rho\sigma ui + d_j)\tau - 2 \log \left(\frac{1 - g_j e^{d_j \tau}}{1 - g_j} \right) \right], \\ B_j(\tau, u) &= \frac{b_j - \rho\sigma ui + d_j}{\sigma^2} \times \frac{1 - e^{d_j \tau}}{1 - g_j e^{d_j \tau}},\end{aligned}$$

with

$$\begin{aligned}g_j &= \frac{b_j - \rho\sigma ui + d_j}{b_j - \rho\sigma ui - d_j}, \\ d_j &= \sqrt{(\rho\sigma ui - b_j)^2 - \sigma^2(2u_j ui - u^2)}, \\ u_1 &= 1/2, \text{ and } u_2 = -1/2, \\ b_1 &= \kappa + \lambda - \rho\sigma, \quad b_2 = \kappa + \lambda.\end{aligned}$$

Even though these formulae look complicated, it is relatively easy to implement them numerically. In particular, as indicated by Carr and Madan (1999) and Bakshi and Madan (2000), see also Chapter 15, the fast Fourier method may be used to compute the simultaneous value of many options.

12.2.4 Further insights

Simulation of the SV dynamics

In the previous section, we considered the simpler and straightforward case of a European call option. Invoking the put-call parity, the European put price can be deduced. There are situations, e.g., pricing path dependent options, where we have to simulate trajectories either under the risk-neutral probability or under the objective probability. In this case, we will need the system of stochastic differential equations and the associated PDE. This can be done by analogy. Indeed, for

$$dS_t = \mu_S dt + \sigma_S dZ_{1,t}, \quad (12.28)$$

$$dv_t = \mu_v dt + \sigma_v dZ_{2,t}, \quad (12.29)$$

and instantaneous correlation $\text{Corr}[dZ_{1,t}, dZ_{2,t}] = \rho dt$, the transition probability $p(S_T, v_T | S_t, v_t)$ must satisfy

$$\frac{1}{2} \sigma_S^2 \frac{\partial^2 p}{\partial S_t^2} + \rho \sigma_S \sigma_v \frac{\partial^2 p}{\partial S_t \partial v_t} + \frac{1}{2} \sigma_v^2 \frac{\partial^2 p}{\partial v_t^2} + \mu_S \frac{\partial p}{\partial S_t} + \mu_v \frac{\partial p}{\partial v_t} + \frac{\partial p}{\partial t} = 0.$$

Inspection of the PDE (12.26) for the case $j = 2$, where the transition probability is risk neutral, shows that for $x_t = \log(S_t)$ it must be that

$$\mu_S = r - \frac{1}{2} v_t, \quad \sigma_S = \sqrt{v_t}, \quad \mu_v = \kappa(\theta - v_t) - \lambda v_t, \quad \sigma_v = \sigma \sqrt{v_t}.$$

If we also decompose $dZ_{1,t}$ into

$$dZ_{1,t} = \rho dZ_{2,t} + \sqrt{1 - \rho^2} d\tilde{Z}_{1,t}, \quad (12.30)$$

where $\tilde{Z}_{1,t}$ is a Brownian motion and where $dZ_{2,t}$ and $d\tilde{Z}_{1,t}$ are uncorrelated, then we get

$$dx_t = \left(r - \frac{1}{2} v_t \right) dt + \sqrt{v_t} \left[\rho dZ_{2,t} + \sqrt{1 - \rho^2} d\tilde{Z}_{1,t} \right], \quad (12.31)$$

$$dv_t = [\kappa(\theta - v_t) - \lambda v_t] dt + \sigma \sqrt{v_t} dZ_{2,t}. \quad (12.32)$$

Since $\text{Corr}[d\tilde{Z}_{1,t}, dZ_{2,t}] = 0$, $d\tilde{Z}_{1,t}$ and $dZ_{2,t}$ can be simulated separately and then linked via (12.31) to yield a correlated process.

If we introduce the notations $\kappa^* = \kappa + \lambda$ and $\theta^* = \kappa\theta / (\kappa + \lambda)$, (12.30) can also be written as

$$dv_t = \kappa^*(\theta^* - v_t) dt + \sigma \sqrt{v_t} dZ_{2,t}. \quad (12.33)$$

Using the system (12.31) with either (12.32) or (12.33), it is possible to simulate risk-neutral trajectories using a first or second-order discretization scheme. The difference between (12.28) and (12.31) is the replacement of μ by r . The difference between (12.32) and (12.33) is the replacement of κ , θ and λ by the risk-neutral equivalent κ^* and θ^* . That is the risk-neutral parameters κ^* and θ^* have absorbed λ , the volatility risk premium. Both (12.32) and (12.33) are risk-neutral stochastic differential equations (SDEs). Clearly, to produce stock price trajectories under the objective measure (or actual measure) from the risk-neutral measure, it suffices to replace r by μ . But, to produce trajectories under the risk-neutral measure, we will need an estimate for λ , which is harder and more controversial to estimate.

The bivariate characteristic function

A number of studies have estimated the parameters of the stochastic volatility model via the so-called characteristic function methods.¹⁰ For these techniques, we need the bivariate characteristic function below

$$\phi_{x_t, v_t, t}(u_1, u_2) \equiv E[\exp(iu_1 x_T + iu_2 v_T)].$$

To obtain a solution for this characteristic function, the same insights used in the previous section still apply. In particular, for the objective probabilities, with transition probability p , the PDE

$$\begin{aligned} & \left(\mu - \frac{1}{2}v_t \right) \frac{\partial p}{\partial x} + \frac{1}{2}v_t \frac{\partial^2 p}{\partial x^2} + \rho \sigma v_t \frac{\partial^2 p}{\partial x \partial v_t} \\ & + \kappa(\theta - v_t) \frac{\partial p}{\partial v_t} + \frac{1}{2}\sigma^2 v_t \frac{\partial^2 p}{\partial v_t^2} + \frac{\partial p}{\partial t} = 0 \end{aligned} \quad (12.34)$$

must hold. The next step is to guess a solution to the PDE

$$\phi_{x_t, v_t, t}(u_1, u_2) = \exp[A(\tau, u_1, u_2) + B(\tau, u_1, u_2)v_t + C(\tau, u_1, u_2)x_t].$$

Calculation of the appropriate partial derivatives from the suggested solution and their substitution into (12.34) gives rise to the Riccati equations

$$\begin{aligned} \frac{\partial A}{\partial t} &= \mu C + \kappa \theta B, \\ \frac{\partial B}{\partial t} &= -\frac{1}{2}C + \frac{1}{2}C^2 + \rho \sigma BC - \kappa B + \frac{1}{2}\sigma^2 B^2, \\ \frac{\partial C}{\partial t} &= 0, \end{aligned}$$

with boundary conditions

$$\begin{aligned} A(0, u_1, u_2) &= 0, \\ B(0, u_1, u_2) &= iu_1, \\ C(0, u_1, u_2) &= iu_2. \end{aligned}$$

We immediately get $C(\tau, u_1, u_2) = iu_2$, and the Riccati equations are reduced to

$$\begin{aligned} \frac{\partial A}{\partial t} &= i\mu u_2 + \kappa \theta B, \\ \frac{\partial B}{\partial t} &= -\frac{1}{2}iu_2 - \frac{1}{2}u_2^2 + \rho \sigma iu_2 B - \kappa B + \frac{1}{2}\sigma^2 B^2. \end{aligned}$$

Using a program that allows for symbolic computations, it is possible to evaluate this system of equations without too much effort.

¹⁰ See, for instance, Singleton (2001), Jiang and Knight (2002), and Chacko and Viceira (2003).

12.3 Models with jumps

In this section, we indicate how to price an option under the hypothesis that the underlying asset contains a jump component. Various contributions considered such models. The seminal work in this field is by Merton (1976) who considered a geometric Brownian motion with Gaussian jumps as underlying process. Ball and Torous (1983) introduced the single jump process. This setting is very simple and the resulting option pricing formula is just a mixture. Such a model may be used for an option with a short remaining life. Another contribution is by Bates (1996, 2000) who allows for stochastic volatility as well as normally distributed jumps. For an even more recent contribution, we may mention Duffie, Pan, and Singleton (2000) and Pan (2002) whose stochastic volatility model allows for jumps in the mean as well as in the variance. Kou (2002) deviates slightly from this strand, because he shuts the stochastic volatility part down and generalizes the jump component by introducing the asymmetric double exponential.

Much of the details and fundamentals of jump processes are provided in Chapter 16. Since options depend on the underlying asset, in order to price options with a jump component, it is necessary to express the dynamics of a function of the underlying asset. In other words, we require an extension of Ito's lemma to jump diffusions, an issue to which we will turn now.

12.3.1 Stochastic process with jumps

We consider Z_t , a Brownian motion and N_t , a pure Poisson process, independent of Z_t . Let J_t be the random variable corresponding to a jump realization, i.e., a marker. The Ito process that we consider is X_t where

$$X_t = X_0 + \int_0^t a_s ds + \int_0^t b_s dZ_s + \int_0^t c_s (J_s + 1) dN_s. \quad (12.35)$$

Such a process is also called a jump diffusion if all the parameters a_s , b_s , c_s only depend on X_t and t . In the expression (12.35), all the integrals are assumed to exist. The first integral is interpreted as a Riemann integral. The second integral is an Ito integral. The last one is notation in that

$$\int_0^t c_s (J_s - 1) dN_s = \sum_{j=1}^{N_t} c_{\tau_j} (J_{\tau_j} + 1),$$

where N_t measures the number of jumps that took place before time t and where the τ_j , for $j = 1, \dots, N_t$, are the times when a jump takes place. Some care needs to be taken concerning the timing. Indeed, if a jump takes place exactly at time τ , then the question is if the coefficient c adjusted before time τ or during τ . This means that the same difficulty as with the definition of an Ito integral occurs again. From a finance point of view, if the X_t is

interpreted as a price process, the question is if it is possible to adjust the weights allocated to an asset during a crash. The interpretation that we follow here is that it is not possible to adjust the portfolio during a crash and that it suffers the crash. This means that we have to take the weight right before the crash. To emphasize this, we denote

$$c_{\tau-} = \lim_{s \rightarrow \tau, s < \tau} c_s.$$

It is also possible to write the jump diffusion (12.35) as a stochastic differential equation

$$dX_t = a_t dt + b_t dZ_t + c_t(J_t - 1) dN_t.$$

If X_t is a jump-diffusion process where the diffusion part is a geometric Brownian motion, we obtain

$$dX_t = a X_{t-} dt + \sigma X_{t-} dZ_t + X_{t-}(J_t - 1) dN_t. \quad (12.36)$$

Following the logic of Ito integration, the term premultiplying dN_t must be a càd-làg process. Hence, whenever there is a jump, the total jump size is given by

$$X_{t-}(J_t - 1).$$

In this case, the marker J_t is interpreted as a percentage shock.

Ito's formula yields the expression of $Y_t = f(X_t, t)$. Clearly, concerning the Brownian motion part, $a_t dt + b_t dZ_t$, Ito's lemma yields the same expression as before. The new part concerns the jumps. Intuitively, if X_t jumps from some level X_{t-} to X_t , the function $Y_t = f(X_t, t)$ will jump from $Y_t = f(X_{t-}, t-)$ to $Y_t = f(X_t, t)$. Hence, Ito's lemma is given by

$$\begin{aligned} dY_t = & \left(\frac{1}{2} b_t^2 f_{xx} + a_t f_x + f_t \right) dt + b_t f_x dZ_t \\ & + [f(X_{t-} + c_t(J_t + 1), t) - f(X_{t-}, t-)] dN_t. \end{aligned}$$

As an illustration, for the geometric Brownian motion in (12.36), and using the simple function $Y_t = \log(S_t)$, we obtain

$$dY_t = \left(\mu - \frac{1}{2} \sigma^2 \right) dt + \sigma dZ_t + \log(J_t) dN_t. \quad (12.37)$$

The first part of (12.37) follows from Ito's lemma for a geometric Brownian motion. The jump part is given by the term

$$\log(S_{t-} + S_{t-}(J_t - 1)) - \log(S_{t-}) = \log(J_t).$$

that matters only if a jump occurs, meaning that $dN_t = 1$. Assume now that the jump intensity is λdt . Then, if we assume that $\log(J) \sim \mathcal{N}(\gamma, \delta^2)$, we obtain that the instantaneous mean in (12.37) is $E[dY_t] = (\mu - \frac{1}{2} \sigma^2 + \gamma \lambda) dt$.

In this case, the jump contributes to the mean. With this so-called *uncompensated process*, we are not able to distinguish between the effect of μ and $\gamma\lambda$. As we will see in Chapter 16, it is possible to choose a specification of the model where the jump contribution cancels out. With such a *compensated process*, we have no identification problem.

Given this general jump diffusion process, we can now turn to option pricing. First, we will consider the Kou (2002) model that has a Brownian motion part and some exponential jumps. Then, we turn to the Bates (1996) model with stochastic volatility and Gaussian jumps. In the following subsections, we will follow the steps from the previous subsection when we derived the Heston model, i.e., we first derive the fundamental PDE, create the risk-neutral portfolio, then find the right characteristic function and option pricing formulae.

12.3.2 Diffusion with double exponential jumps

We present here Kou's (2002) model, who adopts the SDE in (12.37) with jumps distributed as double exponentials. The jump process J_t is *iid* and $Y = \log(J)$ is distributed as an asymmetric double exponential distribution with density

$$f_Y(y) = p\eta_1 e^{-\eta_1 y} I_{\{y \geq 0\}} + q\eta_2 e^{+\eta_2 y} I_{\{y < 0\}},$$

for $\eta_1 > 1$, $\eta_2 > 0$, $p \geq 0$, $q \geq 0$ and $p + q = 1$. The first part of this density defines positive jumps with average jump size $1/\eta_1$. The second part corresponds to the negative jumps with average jump size $-1/\eta_2$. These jumps are easy to simulate. When there is a jump, we draw y from

$$y = \begin{cases} \xi^+ & \text{with } \xi^+ \sim \exp(\eta_1), \text{ (with probability } p), \\ -\xi^- & \text{with } \xi^- \sim \exp(\eta_2), \text{ (with probability } 1 - p). \end{cases}$$

Kou (2002) shows that the price of a European call option, with time to maturity $\tau = T - t$, exercise price K and price of underlying S_t , is

$$\begin{aligned} C &= S_t \Upsilon \left(r + \frac{1}{2}\sigma^2 - \lambda\theta, \sigma, \tilde{\lambda}, \tilde{p}, \tilde{\eta}_1, \tilde{\eta}_2; \log(K/S_t), \tau \right) \\ &\quad - Ke^{-r\tau} \Upsilon \left(r - \frac{1}{2}\sigma^2 - \lambda\zeta, \sigma, \lambda, p, \eta_1, \eta_2; \log(K/S_t), \tau \right), \end{aligned}$$

where

$$\begin{aligned} \zeta &= \frac{p\eta_1}{\eta_1 - 1} + \frac{q\eta_2}{\eta_2 + 1} - 1, \\ \tilde{\lambda} &= \lambda(\zeta + 1), \\ \tilde{\eta}_2 &= \eta_2 + 1, \\ \tilde{\eta}_1 &= \eta_1 - 1, \\ \tilde{p} &= \frac{p}{1 + \zeta} \frac{\eta_1}{\eta_1 - 1}. \end{aligned}$$

The function Υ is in turn a function of several other functions. Specifically, we have

$$\begin{aligned} \Upsilon(\gamma, \sigma, \lambda, p, \eta_1, \eta_2; a, \tau) &= \frac{e^{(\sigma\eta_1)^2\tau/2}}{\sigma\sqrt{2\pi\tau}} \sum_{n=1}^{\infty} \pi_n \sum_{k=1}^n P_{n,k} (\sigma\sqrt{\tau}\eta_1)^k \\ &\quad \times I_{k-1} \left(a - \mu\tau; -\eta_1, -\frac{1}{\sigma\sqrt{\tau}}, -\sigma\eta_1\sqrt{\tau} \right) \\ &\quad + \frac{e^{(\sigma\eta_2)^2\tau/2}}{\sigma\sqrt{2\pi\tau}} \sum_{n=1}^{\infty} \pi_n \sum_{k=1}^n Q_{n,k} (\sigma\sqrt{\tau}\eta_2)^k \\ &\quad \times I_{k-1} \left(a - \eta\tau; \eta_2, \frac{1}{\sigma\sqrt{\tau}}, -\sigma\eta_2\sqrt{\tau} \right) \\ &\quad + \pi_0 \Phi \left(-\frac{a - \mu\tau}{\sigma\sqrt{T}} \right), \end{aligned} \quad (12.38)$$

where Φ is the *cdf* of the standard normal distribution. The parameter $\pi_n \equiv e^{-\lambda\tau}(\lambda\tau)^n/n!$ represents the probability of a Poisson random variable taking the value n when its instantaneous jump intensity is λ and jumps take place over the time interval t to T , with $\tau = T - t$.

For $1 \leq k \leq n - 1$, we have

$$\begin{aligned} P_{n,k} &\equiv \sum_{i=k}^{n-1} \binom{n-k-1}{i-k} \binom{n}{i} \left(\frac{\eta_1}{\eta_1 + \eta_2} \right)^{i-k} \left(\frac{\eta_2}{\eta_1 + \eta_2} \right)^{n-i} p^i q^{n-i}, \\ Q_{n,k} &\equiv \sum_{i=k}^{n-1} \binom{n-k-1}{i-k} \binom{n}{i} \left(\frac{\eta_1}{\eta_1 + \eta_2} \right)^{n-i} \left(\frac{\eta_2}{\eta_1 + \eta_2} \right)^{i-k} p^{n-i} q^i. \end{aligned}$$

The last values are $P_{n,n} = p^n$ and $Q_{n,n} = q^n$. The expression $\binom{0}{0}$ is defined to be one.

The sequence of functions I_k is defined as follows. If $\beta > 0$ and $\alpha \neq 0$, then for all $n \geq -1$,

$$\begin{aligned} I_n(c; \alpha, \beta, \delta) &\equiv -\frac{e^{\alpha c}}{\alpha} \sum_{i=0}^n \left(\frac{\beta}{\alpha} \right)^{n-i} H h_i(\beta c - \delta) \\ &\quad + \left(\frac{\beta}{\alpha} \right)^{n+1} \frac{\sqrt{2\pi}}{\beta} \exp \left(\frac{\alpha\delta}{\beta} + \frac{\alpha^2}{2\beta^2} \right) \Phi(-\beta c + \delta + \frac{\alpha}{\beta}), \end{aligned} \quad (12.39)$$

and for $\beta < 0$ and $\alpha < 0$, we have for all $n \geq 1$,

$$\begin{aligned} I_n(c; \alpha, \beta, \delta) &\equiv -\frac{e^{\alpha c}}{\alpha} \sum_{i=0}^n \left(\frac{\beta}{\alpha} \right)^{n-i} H h_i(\beta c - \delta) \\ &\quad - \left(\frac{\beta}{\alpha} \right)^{n+1} \frac{\sqrt{2\pi}}{\beta} \exp \left(\frac{\alpha\delta}{\beta} + \frac{\alpha^2}{2\beta^2} \right) \Phi(\beta c - \delta - \frac{\alpha}{\beta}), \end{aligned} \quad (12.40)$$

and where $Hh(\cdot)$ is defined as

$$Hh_{-1}(x) = e^{-x^2/2} = \sqrt{2\pi}\varphi(x), \quad (12.41)$$

$$Hh_0(x) = \sqrt{2\pi}\Phi(-x), \quad (12.42)$$

$$Hh_n(x) = \frac{1}{n}[Hh_{n-2}(x) - xHh_{n-1}(x)], \quad n \geq 1. \quad (12.43)$$

Here, φ represents the *pdf* of a $\mathcal{N}(0, 1)$. The three equations (12.41)–(12.43) show that the Hh function can be computed iteratively. These formulae look worse than they actually are. The infinite sum in (12.38) can be truncated and Kou (2002) shows that the approximation works well with only a few terms, because π_n will quickly converge towards zero for finance-oriented applications. The Hh function can also be evaluated using the ${}_1F_1$ confluent hypergeometric function

$$Hh_n(x) = 2^{-n/2}\sqrt{\pi}e^{-\frac{x^2}{2}} \times \left[\frac{{}_1F_1(\frac{1}{2}n+1, \frac{1}{2}, \frac{1}{2}x^2)}{\sqrt{2}\Gamma(1+\frac{1}{2}n)} - x \frac{{}_1F_1(\frac{1}{2}n+1, \frac{3}{2}, \frac{1}{2}x^2)}{\Gamma(\frac{1}{2}+\frac{1}{2}n)} \right].$$

Numerical procedures for calculating ${}_1F_1$ are available in many popular software.

Kou's (2002) model is interesting in that it provides an option pricing formula that is nearly in closed form. The implementation is tedious. But, once this done, the execution is fast. Kou thus provides an alternative to models that only allow for log-normal jumps. Unfortunately, it does not take into account the fact that volatility comes in clusters. Hence, combining Kou's model with a stochastic volatility one would be a useful extension.

12.3.3 Combining stochastic volatility with jumps

Bates (1996) considers a compensated jump-diffusion process with variance following a square root process as in Heston (1993). The full model is

$$dS_t = (\mu - \lambda\bar{J})S_t dt + \sqrt{v_t}S_t dZ_{1,t} + S_t J_t dN_t, \quad (12.44)$$

$$dv_t = (\alpha - \beta v_t)dt + \sigma_v \sqrt{v_t} dZ_{2,t}. \quad (12.45)$$

The term $\lambda\bar{J}$ is the expected instantaneous jump size. As the price process is defined, the process is compensated, i.e., the jump contribution in the drift will be zero. Moreover,

$$\text{Cov}[dZ_{1,t}, dZ_{2,t}] = \rho dt, \quad (12.46)$$

$$\Pr[dN_t = 1] = \lambda dt, \quad (12.47)$$

$$\log(1 + J_t) \sim \mathcal{N}\left(\log(1 + \bar{J}) - \frac{1}{2}\delta^2, \delta^2\right). \quad (12.48)$$

Clearly, J_t is the stochastic jump, expressed as a percentage, whenever a jump takes place. The last equation, (12.48), defines $1 + J_t$ as a log-normal jump. Clearly, we have

$$E[1 + J_t] = \exp\left(\log(1 + \bar{J}) - \frac{1}{2}\delta^2 + \frac{1}{2}\delta^2\right) = 1 + \bar{J},$$

and thus

$$E[J] = \bar{J}.$$

Furthermore

$$E[J_t dN_t] = E[J \mid \text{a jump occurs}] + 0 \cdot E[J_t \mid \text{a jump occurs}] = \bar{J} \lambda dt.$$

As was claimed, in the first equation (12.44), the jump component has been compensated by the introduction of $-\lambda \bar{J} dt$ in the mean. In other words, the expected jump contribution is zero per time unit.

The first step to solve this problem is to associate with (12.44)–(12.48) the partial differential equation. The usual steps are then the same as in Heston. The question that opens is what is the infinitesimal generator for a jump diffusion. To answer this question, let us start with the univariate jump diffusion given by

$$dS_t = a_t dt + b_t dZ_t + c_t J_t dN_t.$$

Let $h(S_T)$ be some integrable function and define

$$\psi(S_t, t) \equiv E[h(S_T) | S_t, t] = \int_{s_T} h(s_T) p(s_T, T | S_t, t) ds_T,$$

the latter integral involving the transition probability. Given that the problem is Markovian, the Chapman-Kolmogorov equation still holds and using the argument of iterated expectations as we did for Heston's model, it is possible to show that ψ is a martingale. It follows that its drift must be zero, or in other terms that the expected instantaneous variation must be zero. We may use the Ito formula for diffusions involving jump-processes to obtain¹¹

$$0 = E[d\psi] = \int_{s_T} h(s_T) \left[\left(\frac{1}{2} b_t^2 \frac{\partial^2 p}{\partial s^2} + a_t \frac{\partial p}{\partial s} + \frac{\partial p}{\partial t} \right) dt + b_t \frac{\partial p}{\partial s} dZ_t + (p(s_{t-} + c_{t-} J_t, t) - p(s_{t-}, t-)) dN_t \right] ds_T,$$

where we suppressed the arguments (s_T, T) figuring in the transition probabilities. We notice that there are two components to the drift term. One due to the diffusion part and the other due to the jump part. The expectation of the Brownian motion part, dZ_t is nil and thus it must be that

¹¹ This extension of Ito's lemma to processes with discrete jumps is immediate.

$$\int_{s_T} h(s_T) \left[\left(\frac{1}{2} b_t^2 \frac{\partial^2 p}{\partial s^2} + a_t \frac{\partial p}{\partial s} + \frac{\partial p}{\partial t} \right) dt + E[(p(s_{t-} + c_{t-} J_t, t) - p(s_{t-}, t-)) dN_t] \right] ds_T = 0.$$

Given that this equation must hold for all $h(\cdot)$ functions, it must also be true for Dirac functions. It follows, using the fact that the jump size J_t and the Poisson process are independent, that

$$\frac{1}{2} b_t^2 \frac{\partial^2 p}{\partial s^2} + a_t \frac{\partial p}{\partial s} + \frac{\partial p}{\partial t} + \lambda E[p(s_{t-} + c_{t-} J_t, t) - p(s_{t-}, t-)] = 0,$$

an expression called the infinitesimal generator of the jump diffusion. Alternatively, this is the PDE that the transition probability of a jump diffusion must obey.

We notice that the only difference with what has been achieved before is the expectation part in the formula. If we are dealing with a more general jump diffusion such as a stochastic volatility model, the same logic applies and the infinitesimal generator follows easily. From there on, the different steps of creating risk-neutral portfolios are similar to Heston's. Eventually, Bates obtains the following PDEs for the transition probability, expanding the notations for a_t , b_t and c_t

$$\begin{aligned} & \frac{\partial p}{\partial t} + \left(r - \lambda^* \bar{J} - \frac{1}{2} v_t \right) \frac{\partial p}{\partial s} + (\alpha - \beta^* v_t) \frac{\partial p}{\partial v} + \\ & \frac{1}{2} v_t \left(\frac{\partial^2 p}{\partial s^2} + 2\rho \sigma_v \frac{\partial^2 p}{\partial s \partial v} + \sigma_v^2 \frac{\partial^2 p}{\partial v^2} \right) + \lambda^* E[p(s_t + \gamma^*, t) - p(s_t, t)] = 0, \end{aligned} \tag{12.49}$$

with

$$\gamma^* \equiv \log(1 + J^*) \sim \mathcal{N} \left(\log(1 + \bar{J}^*) - \frac{1}{2} \delta^2, \delta^2 \right).$$

The parameters λ^* , β^* , and \bar{J}^* are risk neutral. In the case of a European call option with maturity date T , strike price K and parameter vector θ , the pricing formula is

$$C(S_t, T, t, K, \theta) = S_t P_1(S_T > \log(K)) - K e^{-r\tau} P_2(S_T > \log(K)),$$

where P_1 and P_2 are survival functions. The differential equations associated with P_2 follows, as in Heston, from the construction of a non-arbitrage portfolio. The partial differential equation associated with P_1 follows by recognizing that C must follow the same partial differential equation as P_2 .¹²

A characteristic function that will solve the PDE in (12.49) can be guessed in the usual way given the linearity of the problem

¹² The appendix in Bakshi, Cao, and Chen (1997) also provides details on the various PDEs and guidance on how to solve them.

$$\varphi(u; S_t, v_t, t) = \exp [A(u, \tau) + B(u, \tau)v_t + iuS_t].$$

Bates derives

$$C(S_t, T, t, K, \zeta) = S_t P_1 - K e^{-r\tau} P_2,$$

with, for $j = 1, 2$,

$$P_j = \frac{1}{2} + \frac{1}{\pi} \int_0^\infty \operatorname{Re} \left[\frac{\varphi_{j;x_t, v_t, t}(u) \exp(-iu \log(K))}{iu} \right] du,$$

where

$$\begin{aligned} \varphi_{j;x_t, v_t, t}(u) &\equiv E[\exp(iuX_T)] \\ &= \exp \left(A_j(\tau, u) + B_j(\tau, u)v_t + iuS_t + \lambda^* \tau (1 + \bar{J}^*)^{\mu_j + \frac{1}{2}} \right. \\ &\quad \times \left. [(1 + \bar{J}^*)^{iu} \exp \left(\delta^2 (\mu_j iu - \frac{1}{2} u^2) - 1 \right)] \right), \end{aligned}$$

and

$$\begin{aligned} B_j(\tau, u) &= (r - \lambda^* \bar{J}^*) iu\tau - \frac{\alpha\tau}{\sigma_v^2} (\rho\sigma_v iu - \beta_j - \gamma_j) \\ &\quad - \frac{2\alpha}{\sigma_V^2} \log \left[1 + \frac{1}{2} (\rho\sigma_v iu - \beta_j - \gamma_j) \frac{1 - e^{\gamma_j \tau}}{\gamma_j} \right]. \end{aligned}$$

$$A_j(\tau; u) = -2 \frac{\mu_j iu - \frac{1}{2} u^2}{\rho\sigma_v iu - \beta_j + \gamma_j (1 + e^{\gamma_j \tau}) / (1 - e^{\gamma_j \tau})},$$

$$\gamma_j = \sqrt{(\rho\sigma_v iu - \beta_j)^2 - 2\sigma_v^2 \left(\mu_j iu - \frac{1}{2} u^2 \right)},$$

$$u_1 = +1/2, \quad u = -1/2, \quad \beta_1 = \beta^* - \rho\sigma_v, \quad \text{and} \quad \beta_2 = \beta^*.$$

This formula is again relatively straightforward to implement. As before, some care is needed in the computation of the inverse Fourier transform because the integrand is an oscillating function.

Bates also shows that the risk-neutral counterparts to equations (12.44)–(12.48) are

$$dS_t/S_{t-} = (r - \lambda^* \bar{J}^*) dt + \sqrt{V_t} dZ_{1,t}^* + J_t^* dN_t,$$

$$dV_t = [\alpha - \beta^* V_t] dt + \sigma_v \sqrt{V_t} dZ_{2,t}^*,$$

$$\operatorname{Cov}[dZ_{1,t}^*, dZ_{2,t}^*] = \rho dt,$$

$$\Pr[dN^* = 1] = \lambda^* dt,$$

$$E^*[J^*] = \bar{J}^*.$$

All starred variables represent risk-neutral versions of the actual variables. This system of equations may be discretized with a fine first or second-order scheme as the basis for simulations. The model of Bakshi, Cao, and Chen (1997) is similar to Bates's, but with the extension to include a square root process for interest rate as well.

12.3.4 Jumpy affine models

Duffie, Pan, and Singleton (2000, hereafter DPS) extend the Heston and Bates models by allowing jumps in the volatility processes. DPS also propose a very general setting whereby the jump size may depend on some of the variables of the model. Formally, the DPS model is expressed in a multivariate form

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t + dZ_t,$$

where W_t is a Brownian motion in \mathbb{R}^n . The functions μ and σ map the n -dimensional vector X_t into \mathbb{R}^n and $\mathbb{R}^{n \times n}$ respectively. The process Z is a compound Poisson process with jumps distributed along a distribution function ν . The jump intensity $\lambda(X_t)$ could depend on the vector X_t . It is assumed that the distribution of the jump sizes, i.e., the markers, is independent of X_t . The intensity of the jump is supposed to be the same for all the components of X_t . It is also assumed that there exists a discount function written R that can serve as the numeraire. DPS consider affine models of the form

$$\begin{aligned} \text{Drift} : \mu(x) &= K_0 + K_1 x, \\ \text{Jump intensity} : \lambda(x) &= l_0 + l_1 x, \\ \text{Discount factor} : R(x) &= \rho_0 + \rho_1 x, \end{aligned}$$

where l_0 and ρ_0 are real scalars, K_0 , l_1 and ρ_1 are $(n, 1)$ real vectors, and K_1 is an (n, n) matrix. Define $\Sigma = \sigma(X_t)\sigma(X_t)'$, then DPS assume that $\Sigma_{ij} = H_{0,ij} + H_{1,ij}x$, with $H_{0,ij} \in \mathbb{R}$ and $H_{1,ij}$ an (n, n) matrix. As a consequence, the specification is a linear one for all the parameters. DPS characterize the jumps via their moment generating function. To do so, they define c a vector of n complex numbers and introduce the moment generating function as $\theta(c) = \int_{\mathbb{R}^n} \exp(c \cdot z) d\nu(z)$. Given the very general multivariate affine nature of the problem, DPS propose a discounted version of the characteristic function as a solution to the PDE associated with the problem

$$\begin{aligned} \varphi(u, X_{t+\tau}, \tau | X_t) &\equiv E \left[\exp \left(- \int_t^T R(X_s) ds \right) \exp(iuX_{t+\tau} | X_t) \right] \\ &= \exp [A(\tau, u) + B(\tau, u)X_t], \end{aligned}$$

where $A(\cdot)$ and $B(\cdot)$ are solutions to the following Riccati equations

$$\begin{aligned} \frac{\partial B}{\partial \tau}(\tau, u) &= \rho_1 - K'_1 B(\tau, u) - \frac{1}{2} B(\tau, u)' H_1 B(\tau, u) - l_1(\theta(B(\tau, u)) - 1), \\ \frac{\partial A}{\partial \tau}(\tau, u) &= \rho_0 - K'_0 B(\tau, u) - \frac{1}{2} B(\tau, u)' H_0 B(\tau, u) - l_0(\theta(B(\tau, u)) - 1), \end{aligned}$$

with boundary conditions

$$B(0, u) = iu, \quad \text{and} \quad A(0, u) = 0.$$

As DPS mention, in certain situations, depending on the structure of the model, solutions may be found analytically. In some other circumstances, it may be necessary to solve the equations numerically.¹³

DPS show how to obtain for specific settings risk-neutral dynamics and they derive, using characteristic function type techniques expressions for a large set of financial instruments such as bond derivatives, quantos, foreign bond options, chooser options, as well as Asian options.

12.4 Models with even wilder jumps: Lévy option pricing

More recently, an almost completely different approach is adopted for dealing with time-varying volatility and fat-tailed distributions. Rather than assuming that the variance changes through time as, for example, a square root process, the new approach assumes stochastic time changes. As more activities take place in the markets, time changes faster and generates greater volatility. In a certain way, these models are not dissimilar to the Mixture of Distributions Hypothesis of Clark (1973) described in Chapter 3. Indeed, the MDH is a special case of the models discussed below when there is no conditional time change.

The time changed process is used in modeling the underlying asset price and for pricing options. The use of time changed Lévy processes to price options is now known as Lévy option pricing. The Lévy option pricing model relies heavily on characteristic functions as well as the representation of Lévy processes. These two topics are discussed in Chapters 15 and 17. Readers are encouraged to refer to these chapters before proceeding to Lévy option pricing below.

Let us denote the log of asset prices as $s_t = \log(S_t)$. Without loss of generality, assume that the process s_t starts at time $t = 0$, from 0. The statement that time changes in a stochastic manner means that there exists an increasing process, written T_t , so that

$$s_t = \mu t + \sigma X_{T_t},$$

where T_t is an increasing process such that for $t_1 < t_2$, $T_{t_1} < T_{t_2}$. The index t corresponds to actual time (or calendar time). The process T_t must be increasing since it represents time. The increasing process T_t may be constructed by integrating some positive quantities that represent quantities of information. We will study later how to construct such processes. Let the random variable X_t denote some Lévy process. Here, it is assumed that T_t and X_t are both independent. Notice that we do not assume a change of time for the drift.

Let $\phi_{X_t}(u)$ denote the characteristic function of X_t . For a Lévy process, we know that the argument of the characteristic function is linear in time

¹³ Study of this system of differential equations reveals that, because of the quadratic term, the system is not linear, hence it will be difficult to solve numerically.

$$\phi_{X_t}(u) = \exp(t \cdot \psi(u)),$$

where the function $\psi(u)$ is the Lévy exponent. Next, denote by $\phi_{T_t}(u)$ the characteristic function of T_t . Typically, $\phi_{T_t}(u)$ is conditioned on the information available at time $t = 0$, such as the level of volatility or the level of the process directing time.

Now, we are ready to explore the relationship between the characteristic function of the time-changed Lévy process X_{T_t} , $\phi_{X_{T_t}}(u)$, the characteristic function of the increasing process, $\phi_{T_t}(u)$, and the characteristic function of the original process, $\phi_{X_t}(u)$, the latter of which is now expressed in terms of $\psi(u)$. First, invoke the definition of characteristic function of the time changed process

$$\begin{aligned}\phi_{X_{T_t}}(u) &= E[e^{iuX_{T_t}}] \\ &= E[E[e^{iuX_{T_t}} | T_t]] = E[e^{T_t\psi(u)}] \\ &= E[e^{i(-i\psi(u))T_t}] = \phi_{T_t}(-i\psi(u)).\end{aligned}\quad (12.50)$$

As these equations show, if the characteristic function $\psi(u)$ of the Lévy process is known and an increasing process for the time change has been chosen via ϕ_{T_t} , then it is relatively easy to obtain the characteristic function of the time changed process $\phi_{X_{T_t}}(u)$ and hence $\phi_{s_t}(u)$ for $s_t = \mu t + \sigma X_{T_t}$, as

$$\phi_{s_t}(u) = e^{i\mu tu} \phi_{T_t}(-i\psi(\sigma u)).$$

By combining various Lévy processes with selected T_t processes, a large selection of models may be generated. The idea of working with characteristic functions and Lévy processes, goes back to work by Madan and Seneta (1990), Madan and Milne (1991), Barndorff-Nielsen (1998), Madan, Carr, and Chang (1998), Barndorff-Nielsen and Shephard (2001), Carr and Wu (2002), and culminating with the works of Geman, Madan, and Yor (2001, 2002) and Carr et al. (2002, 2003).¹⁴

Since Lévy processes are represented by characteristic functions, to solve option pricing problems using Lévy processes, we must first identify a characteristic function that represents the selected Lévy process. The immediate questions that follow are (i) Which types of Lévy processes are useful in finance? In other words, how should we choose $\psi(u)$? (ii) How should we choose the increasing process T_t ? That is, what is the best way to change time? (iii) How can we price an option with such a model? We will now address these questions in turn. The structure of this section follows that of Carr et al. (2003).

¹⁴ For recent textbooks that give excellent treatment of these topics, we ought to mention Schoutens (2003) and Cont and Tankov (2004).

Table 12.1. Various candidate Lévy processes, described by their characteristic function. The $\Gamma(\cdot)$ corresponds to the gamma function. NIG means Normal Inverse Gaussian, GH means Generalized Hyperbolic, and CGMY means Carr, Geman, Madan, and Yor

Name	Characteristic function	Parameters
Gaussian	$\exp(iu\mu - \frac{1}{2}u^2\sigma^2)$	μ, σ
Compound Poisson	$\exp(\lambda[\phi_{X_{T_t}}(u) - 1])$	$\lambda, \phi_{X_{T_t}}(\cdot)$
Variance Gamma	$(1 - i\mu u + \sigma^2\nu \frac{u^2}{2})^{-1/\nu}$	μ, σ, ν
NIG	$\exp(-\delta(\sqrt{\alpha^2 - (\beta + iu)^2} - \sqrt{\alpha^2 - \beta^2}))$	α, β, δ
GH	$e^{i\mu u} \left(\frac{\alpha^2 - \beta^2}{\alpha^2 - (\beta + iu)^2} \right)^{\nu/2} \frac{K_\nu(\delta(\sqrt{\lambda^2 - (\beta + iu)^2}))}{K_\nu(\delta\sqrt{\alpha^2 - \beta^2})}$	$\alpha, \beta, \delta, \nu$
CGMY	$\begin{aligned} & \exp \left[C_p \Gamma(-Y_p) ((M - iu)^{Y_p} - M^{Y_p}) \right. \\ & \quad \left. + C_u \Gamma(-Y_u) ((G + iu)^{Y_u} - G^{Y_u}) \right] \end{aligned}$	$C_p, C_u, G,$ M, Y_p, Y_u

12.4.1 Commonly used Lévy processes

As shown in Chapter 17, there are many ways to represent Lévy processes. In Table 12.1, we provide a selection of commonly used Lévy processes, their characteristic functions, and the parameters involved. CGMY has been developed by Carr, Geman, Madan, and Yor (2003).

Note that the Lévy processes include both continuous and discontinuous processes. The Gaussian model in Table 12.1 is the simplest example of a Lévy process, taken for a unit time increment, corresponding to the Black-Scholes model. The compound Poisson process described next will be discussed extensively in Chapter 16. It depends on the jump-intensity parameter λ and on the characteristic function of the marker process. The compound Poisson process is characterized by periods of no activity, alternated by jumps whose size is dictated by the characteristic function $\phi_{X_{T_t}}(\cdot)$. Thus, the compound Poisson process alone cannot capture the properties of financial returns. But, when it is combined with a Brownian motion, it produces the jump-diffusion model in Merton (1976). Clearly, the parameters μ and σ control the mean and the standard deviation.

The Variance Gamma model was introduced by Madan and Seneta (1990) in the direct spirit of the mixture of distribution hypothesis. Madan and Seneta (1990) obtain this model by subordinating a Gaussian process with a gamma process that interacts through the variance of the Gaussian process. The transformed Gaussian process thus has mean μ and variance $\sigma^2\nu$, where ν , representing the information arrival rate, is a gamma process. For identification purposes, the average information arrival is scaled as to have a mean of one. Madan, Carr, and Chang (1998) extend this model allowing for asymme-

tries in the arrival of good news (positive increments) and bad news (negative increments).

The Normal Inverse Gaussian (NIG) process of Barndorff-Nielsen (1998) and its generalization by Eberlein, Keller, and Prause (1998) are designed to gain a better understanding of the distribution of returns generated by a stochastic volatility model.

The CGMY model, introduced by Carr et al. (2003), allows capturing the asymmetry of the left and right tails of the distribution. It can also be viewed as a generalization of the Variance Gamma process.

These various characteristic functions can be used as candidate distributions of asset returns. As mentioned earlier, to introduce temporal dependency and to explain volatility clustering, we need an increasing process for the time change. We turn to this issue in the next subsection.

12.4.2 Choice of the time-changing process

Many different increasing processes could be used. Here, we will use the square root process in Cox, Ingersoll, and Ross (CIR, 1985) as an example. CIR use this process to model interest rates. With an appropriate choice of parameter values, the model generates interest rate levels that are always positive. After integration, this yields an increasing process. CIR show that if

$$dy_t = \kappa(\theta - y_t)dt + \lambda\sqrt{y_t}dW_t,$$

and

$$T_t = \int_{s=0}^t y_s ds,$$

then the characteristic function may be expressed as

$$\phi_{T_t}(u) = E[e^{iuT_t}] = A(t, u) \exp(B(t, u)y_0), \quad (12.51)$$

where

$$\begin{aligned} A(t, u) &= \frac{\exp(\kappa^2\theta t/\lambda^2)}{[\cosh(\gamma t/2) + \kappa/\gamma \sinh(\gamma t/2)]^{2\kappa\eta/\lambda^2}}, \\ B(t, u) &= \frac{2iu}{\kappa + \gamma \coth(\gamma t/2)}, \\ \gamma &= \sqrt{\kappa^2 - 2\lambda^2 iu}. \end{aligned}$$

These computations show that, from the CIR model, it is possible to obtain the characteristic function of the increasing process, T_t . As (12.51) shows, the characteristic function is conditional on the current level of the activity process y_0 . Intuitively, if this parameter is very high, then for some time the activity process will remain high. Only after some time, determined by the speed of mean reversion κ , will it revert to the mean θ .

12.4.3 Option pricing

Assume that the dynamic of the asset price has been established by the choice of a Lévy process and of a time-changing process. What we need now is a risk-neutral process that we can use to price derivative assets. This is a complicated issue, because asset prices are now driven by a Lévy process. With the exception of the Brownian motion, the Lévy process family contains processes with jumps of all sorts. This means that, in general, there may not be a unique martingale measure. One possible way to proceed is to follow Madan and Milne (1991) and to perform a so-called *convexity correction*. We may view this as a change of drift so that the resulting process is a discounted martingale.¹⁵ The trick is to replace

$$S_t = S_0 \exp(\mu t + \sigma X_{T_t})$$

with

$$S_t = S_0 \exp(\mu t + mt + \sigma X_{T_t}) \quad (12.52)$$

such that

$$e^{-rt} E[S_t] = S_0. \quad (12.53)$$

The term in m represents a convexity correction.¹⁶ Substituting (12.52) into (12.53), and using expression (12.50), we get

$$\begin{aligned} e^{\mu t + mt} E[e^{\sigma X_{T_t}}] &= e^{rt} \\ \implies e^{\mu t + mt} \phi_{T_t}(-i\psi(-i\sigma)) &= e^{rt} \\ \implies e^{mt} &= e^{(r-\mu)t} [\phi_{T_t}(-i\psi(-i\sigma))]^{-1} \\ \implies mt &= (r - \mu)t - \log(\phi_{T_t}(-i\psi(-i\sigma))). \end{aligned}$$

The risk-neutral process for s_t now becomes

$$s_t = s_0 + rt + \sigma X_{T_t} - \log(\phi_{T_t}(-i\psi(-i\sigma))).$$

As a next step, we need the risk-neutral characteristic function. This follows easily. If $\phi_{s_t}(u)$ is the characteristic function of s_t , then

$$\phi_{s_t}(u) = E[e^{ius_t}] = \exp(iu(s_0 + rt)) \frac{\phi_{T_t}(-i\psi(\sigma u))}{\phi_{T_t}(-i\psi(-i\sigma))^{iu}}. \quad (12.54)$$

Now that we have the risk-neutral characteristic function, we are ready to price options.

¹⁵ This approach is also used in the actuaries literature where it is called the Esscher transform. See also Gerber and Shiu (1994).

¹⁶ For certain Lévy processes, this change of drift is not enough to obtain risk neutrality and a change of time needs also to be made.

12.4.4 Pricing options with risk-neutral characteristic function

As in the previous section and following the approach in Heston (1993), Bates (1997), Bakshi and Madan (1998), the option is priced by inverting the characteristic function of some complicated risk-neutral process. Here, we follow Carr and Madan (1999) who show how to relate the call option with its characteristic function and show how to invert the problem as to obtain the option price. The inversion may be efficiently implemented using the fast Fourier transform.

Let S_T be the risk-neutral stock price and $s_T \equiv \log(S_T)$ its logarithm. In many applications the characteristic function of s_T is available but not its density $q(s_T)$. The characteristic function of s_T is

$$\phi_{s_T}(u) = \int_{-\infty}^{\infty} e^{ius_T} q(s_T) ds_T.$$

If the risk-neutral density, $q(\cdot)$, was known, we would write the price of a European call option as

$$C_T(k) = e^{-rT} \int_{s=k}^{+\infty} (e^s - e^k) q(s) ds. \quad (12.55)$$

A direct attempt first by computing the density $q(s)$ via Fourier transform and then by computing the integral 12.55 does not seem to yield a simple expression. The idea of Carr and Madan (1999) is to consider the Fourier transform of the European call option. As we will see, this computation will, in the end, give a relatively simple expression. Let us therefore consider the Fourier transform of $C_T(k)$. Unfortunately, $C_T(k)$ tends to S_0 as $k \rightarrow -\infty$. As a consequence, $C_T(k)$, not converging to zero, is not square integrable and its Fourier transform cannot be computed. To overcome this difficulty, Carr and Madan (1999) introduce a *dampening factor* $e^{\alpha k}$ that converges to zero as $k \rightarrow -\infty$. Considering

$$c_T(k) = e^{\alpha k} C_T(k) \iff C_T(k) = e^{-\alpha k} c_T(k), \quad (12.56)$$

we obtain a square integrable function that retains all the information contained in $C_T(k)$. Consider now the Fourier transform

$$\psi(v) = \int_{-\infty}^{\infty} \exp(ivk) c_T(k) dk. \quad (12.57)$$

The objective is to link $\psi(v)$ with $\phi_{s_T}(u)$. Following Carr and Madan (1999), we substitute (12.56) into (12.57) and get

$$\begin{aligned}
\psi(v) &= \int_{k=-\infty}^{\infty} e^{ivk} \int_{s=k}^{+\infty} e^{\alpha k} e^{-rT} (e^s - e^k) q(s) ds dk \\
&= e^{-rT} \int_{s=-\infty}^{\infty} q(s) \int_{k=-\infty}^s \left[e^{(\alpha+iv)k+s} - e^{(1+\alpha)k+ivk} \right] dk ds \\
&= e^{-rT} \int_{s=-\infty}^{\infty} q(s) \left[\frac{e^{(\alpha+iv)k+s}}{(\alpha+iv)} - \frac{e^{(1+\alpha)k+ivk}}{1+\alpha+iv} \right] \Big|_{k=-\infty}^s ds \\
&= e^{-rT} \int_{s=-\infty}^{\infty} q(s) e^{(1+\alpha)s+ivs} \left[\frac{1}{(\alpha+iv)} - \frac{1}{1+\alpha+iv} \right] ds \\
&= e^{-rT} \phi_{s_T}(-(1+\alpha)i+v) \left[\frac{1}{(\alpha+iv)} - \frac{1}{1+\alpha+iv} \right].
\end{aligned}$$

This result shows that the Fourier transform of $c_T(k)$ may be expressed in terms of the characteristic function (12.54) for which a closed-form solution exists. Using the inverse Fourier transform, we obtain that

$$C_T(k) = e^{-\alpha k} \frac{1}{\pi} \int_0^\infty e^{-ivk} \psi(v) dv. \quad (12.58)$$

Hence, if a risk-neutral characteristic function is available for s_T , then the option price may be easily computed using a Fourier inversion via (12.58). Efficient codes to implement this problem are available. Carr and Madan (1999) indicate how to choose an upper bound for the integral. Heuristically, the upper and lower bound of the integral depend on the density q of the price at time T . Often it is possible to get an idea of a reasonable range over which to integrate by considering the available strike prices. We may expect that the distribution of future option price to roughly match the range of strikes.

12.4.5 Empirical results

There are aspects of empirical findings that are of relevance in deciding if a model is useful or not. The first aspect concerns the model performance used for modeling returns. The second concerns the fit to option prices. Regarding the first question, the time-changed process covers a large class of distributions with numerous time-changing processes. Carr et al. (2002) show that Lévy processes with infinite activity provide a better fit to asset returns than the continuous Brownian motion. This result is important as most financial models are based on Brownian motion or on a combination of Brownian motion and jumps of finite size.

At the time of writing, there is not much work on testing the performances of time-changed Lévy processes to price options. Carr et al. (2003) consider various time-changed Lévy processes and two approaches to obtain risk-neutral prices. Of the two approaches tested, the one described earlier in this section, i.e., the convexity correction method, appears to yield a better fit to option prices. In terms of percentage pricing errors, the CGMY model produces a very good fit, comparable to that of the Normal Inverse Gaussian.

This section has presented techniques on how to price options based on Lévy processes. The theoretical architecture appears very impressive and allows for many different types of models. It would be interesting for further research to provide more in-depth econometric investigations of the fit of these promising and elegant models for both returns and option prices.

13

Brownian Motion and Stochastic Calculus

This chapter is about stochastic calculus, i.e., calculus that involves random variables and Brownian motions in particular. The original Brownian motion refers to the trajectory of pollen moving around in a dish of water. The trajectory of such a particle is very random in the sense that its future position is not deterministic, moreover the particle makes countless zig-zags even over an infinitesimally small time interval. This description turns out to fit stock prices very well, as stock price movements also tend to be rather unpredictable over a short horizon. Brownian motion plays a crucial role in the derivation of option prices. This appendix provides the reader with some elementary tools in probability crucial to understanding option pricing studies. Here, we aim to provide an intuitive discussion leaving it to the more mathematically inclined readers to refer to more advanced texts in mathematical finance.¹ There are also probability related books that the interested reader may wish to consult such as Williams (1991). For an alternative soft introduction to the probabilistic tools presented here, we may consult Neftci (1996).²

In the following section, we construct Brownian motion following the approach by Donsker (1951), which is based on the central limit theorem and a re-scaled random walk. With the Brownian motion in place, we present Ito's lemma, Kolmogorov equations, and Feynman-Kac formula. These tools are essential for working on option pricing models in continuous time.

13.1 Law of large numbers and the central limit theorem

The *law of large numbers* and the *central limit theorem* play a very important role in the construction of the Brownian motion. Let X_i be independent ran-

¹ The interested reader may consult among others Dana and Jeanblanc-Piqu  (2003), Duffie (1988), Karatzats and Shreve (1988), and Shreve (2004).

² Some of the material presented here and in the following chapter also stems from the lecture notes of Robert Merton and of Chi-fu Huang, whose courses one of the authors attended in 1989.

dom variables with $i = 1, \dots, N$, and x_i their realizations. If the r th moment of these random variables exists, then the law of large numbers states that the average of the r th power of a large number of realizations of that random variable, x_i , converges to their theoretical moment. Formally, this may be written as

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N x_i^r = E[X^r]. \quad (13.1)$$

In addition, if all X_i have the same distribution with mean μ and variance σ^2 , then the central limit theorem indicates that

$$\frac{\frac{1}{N} \sum_{i=1}^N X_i - \mu}{\sigma/\sqrt{N}} \Rightarrow \mathcal{N}(0, 1), \quad (13.2)$$

where $\mathcal{N}(0, 1)$ denotes the standard normal distribution with zero mean and unit variance, and ' \Rightarrow ' means convergence in distribution. To understand the concept of convergence in distribution, we should imagine simulating realizations of $\sum_{i=1}^N X_i$ for large N many times, and scaling this sum as indicated in formula (13.2). Then, by constructing the histogram of these simulated ratios, we would observe that it is indistinguishable from the one obtained for draws from the $\mathcal{N}(0, 1)$ distribution.³

To understand where this formula intuitively comes from, imagine that the X_i are normal. Then given that a sum of normal random variables is normal, to characterize the distribution of $\sum_{i=1}^N X_i$, it suffices to compute the mean and the variance

$$\begin{aligned} E \left[\sum_{i=1}^N X_i \right] &= \sum_{i=1}^N \mu = N\mu, \\ V \left[\sum_{i=1}^N X_i \right] &= \sum_{i=1}^N \sigma^2 = N\sigma^2. \end{aligned}$$

From there on, rescaling the sum of the X_i yields

$$\sum_{i=1}^N X_i \sim \mathcal{N}(N\mu, N\sigma^2) \quad \Rightarrow \quad \frac{\frac{1}{N} \sum_{i=1}^N X_i - \mu}{\sigma/\sqrt{N}} \sim \mathcal{N}(0, 1).$$

In this formula the tilde \sim means “distributed as”. Unlike convergence in distribution, this convergence means that even for small N the histogram of simulated ratios would not be distinguishable from the one of the normal distribution. The central limit theorem does not require the random variables X_i to have a normal distribution. It will be in large samples that the distribution will be normal. Thus, while the sum of normal variates is exactly normal, the

³ For formal definitions of convergence and limit theorems, see Billingsley (1968) or Gnedenko and Kolmogorov (1954).

sum of scaled exponentials behaves only asymptotically as the normal distribution. The key criterion here is the existence of second and higher moments. For instance, the sum of stable random variables does not converge to a normal distribution. The central limit theorem also holds for variables that are somehow dependent (but not too strongly).

13.2 Random walks

From the central limit theorem, we infer that, if all X_i have zero mean, then

$$\frac{\frac{1}{N} \sum X_i}{\sigma/\sqrt{N}} = \frac{1}{\sigma\sqrt{N}} \sum_{i=1}^N X_i \Rightarrow \mathcal{N}(0, 1).$$

Suppose now that X_i takes the values $+1$ and -1 only, with $\Pr(X_i = +1) = \Pr(X_i = -1) = 1/2$. Then the sequence

$$S_N \equiv \sum_{i=1}^N X_i$$

defines a random walk. Figure 13.1 represents the trajectory of a random walk involving 101 steps with initial step starting at zero. Simple computation shows that this random walk has $E[X_i] = 0$ and $V[X_i] = E[X_i^2] = 1$.

13.3 Construction of the Brownian motion

The procedures for constructing Brownian motion described here follow those of Donsker (1951). The aim is to construct the evolution of a Brownian motion over the time interval $[0, T]$. Let us start with X_i defined, as before, as a random variable taking values $+1$ and -1 at each step with equal probability, and S be the aggregate value of X_i . What could be the value of S at time T given that there could potentially be an infinite number of steps between 0 and T ? Clearly, this requires some scaling, since, if each step had unit length, then an infinite number of steps would aggregate to a length that is also infinite. Our task here is to scale each step in such a manner that the problem does not explode. To do this, let us divide $[0, T]$ into N intervals, each of length Δ . In other words, we have $N\Delta = T$ which implies that $\Delta = T/N$. Next, define a “rescaled” random walk

$$U_{k\frac{T}{N}} \equiv \sqrt{\frac{T}{N}} S_k = \sqrt{\frac{T}{N}} \sum_{i=1}^k X_i.$$

Clearly, $U_{k\frac{T}{N}}$ is a random variable. The scaling factor $\sqrt{T/N}$ (or $\sqrt{\Delta}$) is chosen following the “square-root-of-time” rule, and we shall see the benefit

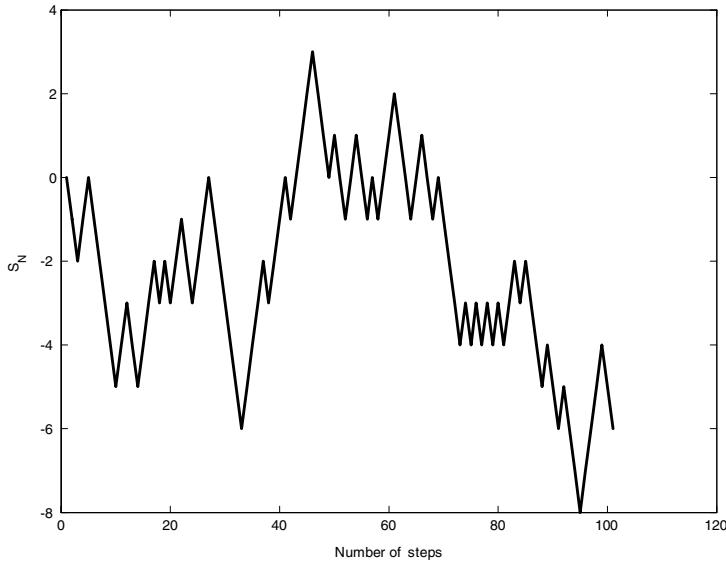


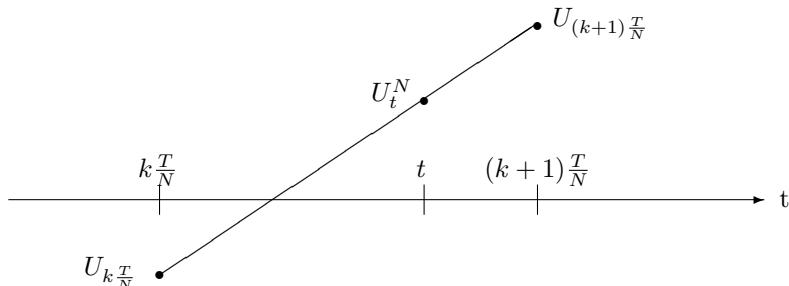
Fig. 13.1. Simulated trajectory of a random walk.

of such a scaling factor soon. What we want to do now is increase the steps N while adjusting k so that $k\Delta$ remains roughly at the same place. We know that for all $t \in [0, T]$, there exists a k such that $k\frac{T}{N} \leq t < (k+1)\frac{T}{N}$ as shown below:

$$\xrightarrow{\quad k\frac{T}{N} \quad} \xrightarrow{\quad t \quad} \xrightarrow{\quad (k+1)\frac{T}{N} \quad}$$

So as N increases, we also have to increase k .

The next step in the construction of the Brownian motion is to note that $U_{k\frac{T}{N}}$ is defined for discrete points $k = 0, \dots, N$. To create a continuous trajectory, we construct U_t^N as a linear interpolation between $U_{k\frac{T}{N}}$ and $U_{(k+1)\frac{T}{N}}$ as shown graphically below.



The expression of U_t^N is trivial to write. To do so does not add insights for what follows. We therefore leave the computation as an exercise.

Now, we are ready to study some of the properties of this “rescaled” random walk $U_{k\frac{T}{N}}$:

- (a) How does U_t^N behave as $N \rightarrow \infty$?

We have

$$U_{k\frac{T}{N}} = \sqrt{\frac{T}{N}} S_k = \sqrt{\frac{Tk}{N}} \frac{1}{\sqrt{k}} S_k \Rightarrow \mathcal{N}(0, t).$$

Also, as $N \rightarrow \infty$, the distance between $k\frac{T}{N}$ and $(k+1)\frac{T}{N}$ becomes very short, and $\frac{T}{N}k$ comes arbitrarily close to t , a fact we write as $\frac{T}{N}k \approx t$. It results that the trajectory, at the distance t , behaves as a $\mathcal{N}(0, t)$ random variable. This means that, U_t^N is confined within a region described by the distribution $\mathcal{N}(0, t)$.

- (b) For $t > s$, how does $U_t^N - U_s^N$ behave as $N \rightarrow \infty$?

We first select k and l such that $k\Delta \leq t \leq (k+1)\Delta$, and $l\Delta \leq s \leq (l+1)\Delta$. As before, as $N \rightarrow \infty$, $t \approx k\frac{T}{N}$, $s \approx l\frac{T}{N}$ and

$$(U_t^N - U_s^N) \approx \left(U_{k\frac{T}{N}} - U_{l\frac{T}{N}} \right) = \sqrt{\frac{T}{N}} \sum_{i=l+1}^k X_i. \quad (13.3)$$

As all the X_i are independent with the same distribution, we could shift all the X_i in (13.3), so that the index i starts at 1. Since between $l+1$ and k there are $k-l$ random variables X_i , it follows that

$$(U_t^N - U_s^N) \approx \sqrt{\frac{T}{N}} \sum_{i=1}^{k-l} X_i \Rightarrow \mathcal{N}(0, t-s).$$

This means that increments of the trajectory that we have just constructed behave like a normal random variable with variance equal to $t-s$, the time increase over which the increments are measured. Moreover, the property of the increments does not depend on t , the time over which we consider the increments. We say that the increments are stationary.

- (c) For $t_1 < t_2 < t_3 < t_4$, what is the relationship between $U_{t_2}^N - U_{t_1}^N$ and $U_{t_4}^N - U_{t_3}^N$?

As before, we choose some l_i so that for $N \rightarrow \infty$, $l_i\Delta \approx t_i$. If N is sufficiently large, then X_i for $i = l_1 + 1, \dots, l_2$ and X_j for $j = l_3 + 1, \dots, l_4$ are independent. The consequence is that $(U_{t_2}^N - U_{t_1}^N) \approx \sqrt{\frac{T}{N}} \sum_{i=l_1+1}^{l_2} X_i$ and $(U_{t_4}^N - U_{t_3}^N) \approx \sqrt{\frac{T}{N}} \sum_{i=l_3+1}^{l_4} X_i$ are independent. This follows from the fact that each increment is a normal random variable and

$$\text{Cov} [U_{t_2}^N - U_{t_1}^N, U_{t_4}^N - U_{t_3}^N] = 0.$$

Hence, for a rescaled random walk, U_t^N , defined on $[0, T]$, we can conclude that:

- (i) it starts, without loss of generality, at 0;
- (ii) it is distributed at t as an $\mathcal{N}(0, t)$;
- (iii) it has independent and stationary increments; and
- (iv) it is continuous.

The limit process, i.e., as $N \rightarrow \infty$, satisfying (i), (ii), (iii), and (iv) is called a Brownian motion or a Wiener process written as W_t . A process that only satisfies (i) and (iii) is called a Lévy process, which may include jumps. Lévy processes play an important role in more advanced work on derivative pricing and the properties of Lévy processes are discussed in detail in Chapter 17.

13.4 Properties of the Brownian motion

Consider a Brownian motion defined on the time interval $[0, T]$. Then, place ourselves at time t , for $t < T$, and ask the question, What will be the best forecast of the process for time s given that the current position of the Brownian motion is at W_t ? The answer to this question is, in fact, the current position of the Brownian motion at time t . This statement can be expressed mathematically as

$$E[W_s | W_t] = W_t \quad \text{for } s > t. \quad (13.4)$$

That is, the best predictor of the future position of a Brownian motion is its current position.

Proof of the claim: We know from the previous section that the Brownian motion W_s is the limit of an interpolation of a rescaled random walk, but with a magnitude much smaller than the random walk. Heuristically, we can replace the Brownian motion by the rescaled random walk. Select l and k such that $l\Delta \approx s$ and $k\Delta \approx t$. As $s > t$, clearly $l > k$. We end up with an approximation for (13.4) as

$$\begin{aligned} & E \left[\sqrt{\frac{T}{N}} \sum_{i=1}^l X_i \middle| \sqrt{\frac{T}{N}} \sum_{i=1}^k X_i \right] \\ &= E \left[\sqrt{\frac{T}{N}} \sum_{i=k+1}^l X_i \middle| \sqrt{\frac{T}{N}} \sum_{i=1}^k X_i \right] + \sqrt{\frac{T}{N}} \sum_{i=1}^k X_i = \sqrt{\frac{T}{N}} \sum_{i=1}^k X_i. \end{aligned}$$

The first equality follows from the fact X_i for $i = 1, \dots, k$ are known at time k . The second equality follows from the fact that X_i for $i = 1, \dots, k$ contains no information concerning future values of X_i for $i = k+1, \dots, l$.

Note that if the future of a process only depends on the past through its current value then such a process is called a *Markovian process*. Hence,

a Brownian motion is Markovian. Furthermore, we have just shown that the expectation of future values X_i is given by their current values. An object with such a property is called a *martingale*.

The following remark is useful for later on: for sufficiently small time interval ε , the *quadratic variation* $(W_{t+\varepsilon} - W_t)^2$ is of the same magnitude as ε . In other words, the quadratic variation is equal to the length of time over which it is measured. To prove this claim, let us assume that $t = k \frac{T}{N}$ for simplicity, and that $t + \varepsilon = (k + 1) \frac{T}{N}$. This means that $\varepsilon = \frac{T}{N}$. The quadratic variation

$$(W_{t+\varepsilon} - W_t)^2 \approx \left(\sqrt{\frac{T}{N}} S_{k+1} - \sqrt{\frac{T}{N}} S_k \right)^2 = \frac{T}{N} X_{k+1}^2 = \frac{T}{N} = \varepsilon, \quad (13.5)$$

because $E(X_{k+1}^2) = 1$. Note that while the increment of a Brownian motion may be positive or negative when associated with a given probability, the quadratic variation is equal to the size of each step.

13.5 Stochastic integration

Let the Brownian motion W_t represent the price of an asset and α_t the number of units, or portfolio weight, invested on that asset. In practice α_t is held constant over investment periods. After a time interval ε , the wealth change is $\alpha_t (W_{t+\varepsilon} - W_t)$, and over a longer time period T , the aggregate change in wealth is $\int_0^T \alpha_t dW_t$. Hence, when we use Brownian motion to model asset prices, it is important to evaluate product such as $\alpha_t (W_{t+\varepsilon} - W_t)$, and integral such as $\int_0^T \alpha_t dW_t$. Unfortunately, classical integration does not work in continuous time involving Brownian motion. Indeed, the classic calculus suggests that

$$\int_0^T W_t dW_t = \int_0^T d\frac{1}{2} W_t^2 = \frac{1}{2} (W_T^2 - W_0^2). \quad (13.6)$$

We will show that this result only holds if the integration is defined in a very peculiar way.⁴ To see this, define $t_0 = 0 < t_1 < t_2 < \dots < t_N = T$ and $\tau_i \in [t_{i-1}, t_i]$, then the integral $\int_0^T W_t dW_t$ can be approximated by the sum

$$\begin{aligned} I_N &= \sum_{i=1}^N W_{\tau_i} (W_{t_i} - W_{t_{i-1}}) \\ &= \frac{W_T^2}{2} - \frac{W_0^2}{2} - \frac{1}{2} \sum_{i=1}^N (W_{t_i} - W_{t_{i-1}})^2 + \sum_{i=1}^N (W_{\tau_i} - W_{t_{i-1}})^2 \\ &\quad + \sum_{i=1}^N (W_{t_i} - W_{\tau_i}) (W_{\tau_i} - W_{t_{i-1}}). \end{aligned} \quad (13.7)$$

⁴ The following has been drawn from Arnold (1974).

The definition of a conventional Riemann integral is given as $\lim_{N \rightarrow \infty} I_N$. The formal proof of (13.7) involves tedious computations. Here, we just want to show that, in stochastic calculus, the value taken by I_N depends crucially on the position of τ_i in the interval $[t_{i-1}, t_i]$. First write the last three components of (13.7) as a , b and c . So $I_N = (W_T^2 - W_0^2)/2 + a + b + c$.

For component a : From the property of quadratic variation, stated in (13.5), we know that $(W_{t_i} - W_{t_{i-1}})^2$ approaches $t_i - t_{i-1}$. Hence, for large N we have

$$\sum_{i=1}^N (W_{t_i} - W_{t_{i-1}})^2 \approx \sum_{i=1}^N (t_i - t_{i-1}) = t_N - t_0 = T.$$

For component b : Again $(W_{\tau_i} - W_{t_{i-1}})^2$, from the property of quadratic variation, approaches $\tau_i - t_{i-1}$, and we get

$$\sum_{i=1}^N (W_{\tau_i} - W_{t_{i-1}})^2 \approx \sum_{i=1}^N (\tau_i - t_{i-1}).$$

For component c : We have $\lim_{N \rightarrow \infty} \sum_{i=1}^N (W_{t_i} - W_{\tau_i})(W_{\tau_i} - W_{t_{i-1}}) = 0$. To see this, let us assume that t_{i-1} , τ_i , and t_i are each separated by exactly one X_i . In that case, we have $W_{t_i} - W_{\tau_i} \approx \sqrt{\frac{T}{N}} X_i$ and $W_{\tau_i} - W_{t_{i-1}} \approx \sqrt{\frac{T}{N}} X_{i-1}$. It follows that

$$\begin{aligned} \lim_{N \rightarrow \infty} \sum_{i=1}^N (W_{t_i} - W_{\tau_i})(W_{\tau_i} - W_{t_{i-1}}) &= \lim_{N \rightarrow \infty} \frac{T}{N} \sum_{i=1}^N X_i X_{i-1} \\ &= T \times E[X_i X_{i-1}] = 0. \end{aligned} \quad (13.8)$$

In (13.8), the second equality is justified by invoking the law of large numbers. The third and last equality comes from the independence between X_i . At this point, we have

$$I_N \sim \frac{W_T^2}{2} - \frac{W_0^2}{2} - \frac{T}{2} + \sum_{i=1}^N (\tau_i - t_{i-1}). \quad (13.9)$$

where τ_i belongs to the interval $[t_{i-1}, t_i]$. If τ_i is always at the same position in each interval, we write $\tau_i = t_{i-1} + \lambda[t_i - t_{i-1}]$ for some fixed $\lambda \in [0, 1]$. If we substitute this expression into (13.9), we get

$$I_N \sim \frac{W_T^2}{2} - \frac{W_0^2}{2} - \frac{T}{2} + \lambda T. \quad (13.10)$$

We can see from (13.10) that the integral I_N may take different values according to the value of λ . For $\lambda = 1/2$, we obtain the classic formula of standard

calculus in (13.6), and the integral is called a *Stratanovitch integral*. However, if $\lambda = 0$, the standard calculus formula no longer holds and we have an *Ito integral*

$$\int_0^T W_t dW_t = \lim_{N \rightarrow \infty} \sum_{i=1}^N W_{t_{i-1}} (W_{t_i} - W_{t_{i-1}}),$$

where $\max_{i=1, \dots, N} |t_i - t_{i-1}| \rightarrow 0$. Each type of integration has its advantages and disadvantages. In finance, we typically form portfolio at the beginning of the time period and wait for the prices to evolve. A natural choice of integration will therefore be an Ito integration. Hence, we always interpret the stochastic integral $\int_0^T \alpha_t dW_t$ as the limit $\lim_{N \rightarrow \infty} \sum_{i=1}^N \alpha_{t_{i-1}} (W_{t_i} - W_{t_{i-1}})$, where the t_i s satisfy $t_0 = 0 < t_1 < t_2 < \dots < t_N = T$. We have just seen that standard calculus will give us a wrong answer for this type of integration, and indeed any integration with $\lambda \neq 1/2$. Therefore, we need stochastic differential equations and tools such as the Ito lemma to deal with stochastic calculus involving Brownian motion. These are the topics which we turn to in the next two sections.

13.6 Stochastic differential equations

Let us consider the object

$$X_t = X_0 + \int_0^t \mu(\text{terms}, s) ds + \int_0^t \sigma(\text{terms}, s) dW_s, \quad (13.11)$$

where $\mu(\text{terms}, s)$ means that μ is a function of *terms* and s . *Terms* could be X_t or events other than X_t . If *terms* also include events that are not X_t , then X_t is called an *Ito process*. If *terms* consist of only X_t and t , then X_t is called a *diffusion process*.

In the case where μ and σ are constants, we can write

$$\begin{aligned} X_t &= X_0 + \int_0^t \mu ds + \int_0^t \sigma dW_s \\ &= X_0 + \mu t + \sigma \lim_{N \rightarrow \infty} \sum_{i=1}^N (W_{t_i} - W_{t_{i-1}}) \\ &= X_0 + \mu t + \sigma W_t. \end{aligned}$$

Clearly, $E[X_t] = X_0 + \mu t$, and $V[X_t] = \sigma^2 t$. Given that W_t is normal, it follows that $X_t \sim \mathcal{N}(\mu t, \sigma^2 t)$.

Once the meaning of the various integrals defining X_t has been understood, we often only write

$$dX_t = \mu_t dt + \sigma_t dW_t. \quad (13.12)$$

Such an expression is called the *stochastic differential equation* (SDE) in contrast to the *stochastic integral* in (13.11). The two time-varying parameters μ_t and σ_t are functions of terms that are yet to be defined. Implicit in (13.12) is also the fact that at time 0, the diffusion starts at some value X_0 . If μ_t and σ_t are constants, then (13.12) is reduced to

$$dX_t = \mu dt + \sigma dW_t,$$

which is called the *arithmetic Brownian motion*.

13.7 Ito's lemma

A differential expresses the variation of some function $f(x)$, where $x \in \mathbb{R}$, when x varies by an infinitesimal amount. It is written as

$$df(x) = f'(x)dx,$$

if f admits a derivative. If we consider a discrete step, we could have

$$\Delta f(x) = f(x + \Delta) - f(x) = f'(x)\Delta + O(\Delta),$$

where $O(\Delta)$ is a remainder such that

$$\lim_{\Delta \rightarrow 0} \frac{O(\Delta)}{\Delta} = 0.$$

In other words, the rest in the approximation contains terms that are smaller than Δ .

We are going to apply this idea to some function f , where f depends on an Ito process, X_t , and time, t . We assume that f is smooth in the sense that all the required derivatives exist. Next, we introduce the Ito process

$$X_t = X_0 + \int_0^t \mu_s ds + \int_0^t \sigma_s dW_s.$$

The stochastic differential equation is

$$dX_t = \mu_t dt + \sigma_t dW_t.$$

The discretized version is

$$X_{t+\Delta} = X_t + \mu_t \Delta + \sigma_t (W_{t+\Delta} - W_t).$$

We are now interested in the dynamic of $f(X_t, t)$. For this purpose, we take a second-order Taylor's expansion. In the following, we suppress the argument (X_t, t) associated with the partial derivatives of f . We have

$$\begin{aligned} f(X_{t+\Delta}, t + \Delta) &= f(X_t, t) + \frac{\partial f}{\partial X}(X_{t+\Delta} - X_t) + \frac{\partial f}{\partial t}\Delta \\ &\quad + \frac{1}{2} \left[\frac{\partial^2 f}{\partial X^2}(X_{t+\Delta} - X_t)^2 + 2 \frac{\partial^2 f}{\partial X \partial t}(X_{t+\Delta} - X_t)\Delta + \frac{\partial^2 f}{\partial t^2}\Delta^2 \right]. \end{aligned} \quad (13.13)$$

Now, the idea is to keep those terms that are of the same order of magnitude as Δ and $(W_{t+\Delta} - W_t)$, and drop all the other terms that are smaller. Exactly as we do in a regular differentiation, we get

$$\begin{aligned} X_{t+\Delta} - X_t &= \mu_t \Delta + \sigma_t(W_{t+\Delta} - W_t), \\ (X_{t+\Delta} - X_t)^2 &= \mu_t^2 \Delta^2 + \sigma_t^2(W_{t+\Delta} - W_t)^2 + 2\mu_t \sigma_t(W_{t+\Delta} - W_t)\Delta \approx \sigma_t^2 \Delta, \\ (X_{t+\Delta} - X_t)\Delta &= \mu_t \Delta^2 + \sigma_t(W_{t+\Delta} - W_t)\Delta \approx 0. \end{aligned}$$

Now substitute the non-zero terms into the Taylor expansion in (13.13), and $(W_{t+\Delta} - W_t)^2 = \Delta$ from the property of quadratic variation in (13.5), to get

$$f(X_{t+\Delta}, t + \Delta) - f(X_t, t) = \frac{\partial f}{\partial X} [\mu_t \Delta + \sigma_t(W_{t+\Delta} - W_t)] + \frac{\partial f}{\partial t} \Delta + \frac{1}{2} \frac{\partial^2 f}{\partial X^2} \sigma_t^2 \Delta.$$

Taking the limit as $\Delta \rightarrow 0$ gives us the formula of Ito's lemma

$$df = \left[\frac{1}{2} \frac{\partial^2 f}{\partial X^2} \sigma_t^2 + \frac{\partial f}{\partial X} \mu_t + \frac{\partial f}{\partial t} \right] dt + \frac{\partial f}{\partial X} \sigma_t dW_t.$$

To remember Ito's lemma, it may be useful to keep in mind the following table of multiplications

		dt	dW
dt	0	0	
dW	0	dt	

As an illustration, let $S_t = f(X_t, t) = \exp(X_t)$ with X_t an arithmetic Brownian motion such that $dX_t = \mu dt + \sigma dW_t$. We have

$$\frac{\partial f}{\partial t} = 0, \quad \frac{\partial f}{\partial X_t} = \exp(X_t) = S_t, \quad \frac{\partial^2 f}{\partial X_t^2} = \exp(X_t) = S_t.$$

Ito's lemma gives $dS_t = (\frac{1}{2}S_t\sigma^2 + S_t\mu) dt + \sigma S_t dW_t$. If we define $\mu_S \equiv \mu + \frac{1}{2}\sigma^2$, we have

$$dS_t = \mu_S S_t dt + S_t \sigma dW_t,$$

and S_t is said to follow a *geometric Brownian motion*. We can easily see that S_t has some nice properties, good for describing the behavior of stock prices. First, since by construction $S_t \equiv \exp(X_t)$, S_t cannot be negative. Second, if we take as a crude approximation

$$\frac{S_t - S_{t-1}}{S_{t-1}} = \mu_S dt + \sigma dW_t \sim \mathcal{N}(\mu_S, \sigma^2),$$

we see that returns defined in this way have a constant variance. Hence, independently from the level, volatility will be constant. This is a desirable feature, at least as a first step.

The process S_t with dynamics

$$dS_t = \mu_S S_t dt + S_t \sigma dW_t,$$

has frequently been used as a price process of the underlying assets. Taking again Ito's lemma, we obtain

$$d\log(S_t) = \left(\mu_S - \frac{1}{2}\sigma^2 \right) dt + \sigma dW_t.$$

Integrating, we get

$$\log(S_T) - \log(S_0) = \left(\mu_S - \frac{1}{2}\sigma^2 \right) T + \sigma (W_T - W_0).$$

Given that $W_T - W_0$ is $\mathcal{N}(0, T)$, it follows that

$$\log(S_T) \sim \mathcal{N} \left(\log(S_0) + \left(\mu_S - \frac{1}{2}\sigma^2 \right) T, \sigma^2 T \right).$$

Thus, the price has a log-normal distribution and returns are normal random variables. In practice, however, most financial asset returns are not normal random variables. Hence, the attempt in this book to cover the non-Gaussian features and the associated modeling aspects.

13.8 Multivariate extension of Ito's lemma

We consider in this section an extension of Ito's lemma to the case where there are two sources of randomness

$$\begin{aligned} dX_{1,t} &= \mu_1 dt + \sigma_1 dW_{1,t}, \\ dX_{2,t} &= \mu_2 dt + \sigma_2 dW_{2,t}, \\ y &= f(X_{1,t}, X_{2,t}, t). \end{aligned}$$

The first two equations specify the dynamics of some processes, $X_{1,t}$ and $X_{2,t}$. We assume that the two Brownian motions are autocorrelated with instantaneous correlation ρdt . The last equation states that y is a function of the two underlying processes as well as of time. The parameters μ and σ may be functions of $X_{1,t}$ and $X_{2,t}$. It is assumed that these parameters follow certain smoothness conditions so that all the following formulae hold. We get, after differentiation,

$$\begin{aligned} dy = & \frac{\partial f}{\partial X_1} dX_{1,t} + \frac{\partial f}{\partial X_2} dX_{2,t} + \frac{\partial f}{\partial t} dt + \frac{1}{2} \frac{\partial^2 f}{\partial X_1^2} (dX_{1,t})^2 \\ & + \frac{1}{2} \frac{\partial^2 f}{\partial X_2^2} (dX_{2,t})^2 + \frac{1}{2} \frac{\partial^2 f}{\partial t^2} (dt)^2 + \frac{\partial^2 f}{\partial X_1 \partial X_2} dX_{1,t} dX_{2,t} \\ & + \frac{\partial^2 f}{\partial X_1 \partial t} dX_{1,t} dt + \frac{\partial^2 f}{\partial X_2 \partial t} dX_{2,t} dt. \end{aligned}$$

As in the case of a single source of uncertainty, we have $(dX_{1,t})^2 = \sigma_1^2 dt$; $(dX_{2,t})^2 = \sigma_2^2 dt$; and $dX_{1,t} dX_{2,t} = \sigma_1 \sigma_2 dW_{1,t} dW_{2,t}$. Furthermore, using the same heuristic arguments as before, we get $dW_{1,t} dW_{2,t} = \sqrt{dt} \varepsilon_{1,t} \sqrt{dt} \varepsilon_{2,t} = dt \varepsilon_{1,t} \varepsilon_{2,t}$. Hence, the product $dW_{1,t} dW_{2,t}$ behaves, on average, as ρdt . Regrouping all the terms, we obtain the following bivariate extension of Ito's lemma

$$\begin{aligned} dy = & \left[\frac{1}{2} \sigma_1^2 \frac{\partial^2 f}{\partial X_1^2} + \rho \sigma_1 \sigma_2 \frac{\partial^2 f}{\partial X_1 \partial X_2} + \frac{1}{2} \sigma_2^2 \frac{\partial^2 f}{\partial X_2^2} \right. \\ & \left. + \mu_1 \frac{\partial f}{\partial X_1} + \mu_2 \frac{\partial f}{\partial X_2} + \frac{\partial f}{\partial t} \right] dt \\ & + \sigma_1 \frac{\partial f}{\partial X_1} dW_{1,t} + \sigma_2 \frac{\partial f}{\partial X_2} dW_{2,t}. \end{aligned}$$

13.9 Transition probabilities and partial differential equations

Various option pricing models encountered in this book involve partial differential equations, PDEs. Since certain PDEs may be solved using probabilistic arguments, it is worth investigating how this can be done. Much research was devoted to the general understanding of the link between PDEs and their probabilistic solutions. Again, our aim is neither to provide a description how this can be done in its generality, nor to provide a mathematically rigorous description. Our emphasis will be on the intuitive aspects. In what follows, our objective is

- To show that a PDE is associated to the Brownian motion. To do so we will use arguments of calculus and introduce the notion of transition probability.
- To derive the PDE associated with the transition probability of a general diffusion process. We will perform this derivation by using martingale properties.
- To provide the solution to certain PDEs under boundary conditions. We will refer to this solution technique as the Feynman-Kac theorem.

A transition probability says where a process will be at some point in the future given its present position. What about a Brownian motion? Where can

it be after some time, given that currently it is at x ? Figure 13.2 shows a possible path for a Brownian motion passing through x at time t . At time $t = 0$, $W_0 = 0$. At time t the Brownian motion is located at x . From there on, the Brownian motion diffuses again and it may take all sorts of values. How can we describe the probability of W_T , at time T , conditional on it being located at x at time t ? Since $(W_T - x) \sim \mathcal{N}(0, T - t)$, we have $W_T \sim \mathcal{N}(x, T - t)$. Therefore, the location of W_T is confined by $\mathcal{N}(x, T - t)$. In other words, the Brownian motion at time T cannot simply be just anywhere.

Now, we ask a further question, “what is the probability of having $W_T \in B$ where $B = \{z : b_l < z < b_u\}$ as in Figure 13.3?” To compute this probability, we need to integrate over the *pdf* of W_T . For this purpose, we need an expression for the density of W_T given that the Brownian motion is located at x at time t . This is easy in this context, because we have seen that W_T is Gaussian. Thus we have

$$p(W_T, T|x, t) \equiv \frac{1}{\sqrt{2\pi}\sqrt{\tau}} \exp\left[-\frac{1}{2} \frac{(W_T - x)^2}{\tau}\right],$$

with $\tau = T - t$. We derive the required conditional probability as

$$\begin{aligned} \Pr[W_T \in B \mid W_t = x, t] &= \Pr[b_l < W_T < b_u \mid W_t = x, t] \\ &= \Pr\left[\frac{b_l - x}{\sqrt{\tau}} < \frac{W_T - x}{\sqrt{\tau}} < \frac{b_u - x}{\sqrt{\tau}} \mid W_t = x, t\right] \\ &= \int_{\frac{b_l - x}{\sqrt{\tau}}}^{\frac{b_u - x}{\sqrt{\tau}}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u^2\right) du \\ &= \Phi\left(\frac{b_u - x}{\sqrt{\tau}}\right) - \Phi\left(\frac{b_l - x}{\sqrt{\tau}}\right), \end{aligned}$$

where Φ is the *cdf* of a standard normal distribution. As these computations indicate, it is simple to measure the probability of finding the Brownian motion in a given interval at some future date.

13.10 Kolmogorov backward and forward equations

To simplify the notation, we write in this section W_T as y . Let $p(y, T|x, t)$ be the transition probability of obtaining y at time T when the process is currently at x at time t . In this section, we will derive the so called *Kolmogorov backward equation*, which states that the transition probability associated with the stochastic differential equation

$$dx = \mu(x, t) dt + \sigma(x, t) dW_t \tag{13.14}$$

is

$$\frac{1}{2}\sigma^2(x, t) \frac{\partial^2 p}{\partial x^2} + \mu(x, t) \frac{\partial p}{\partial x} + \frac{\partial p}{\partial t} = 0. \tag{13.15}$$

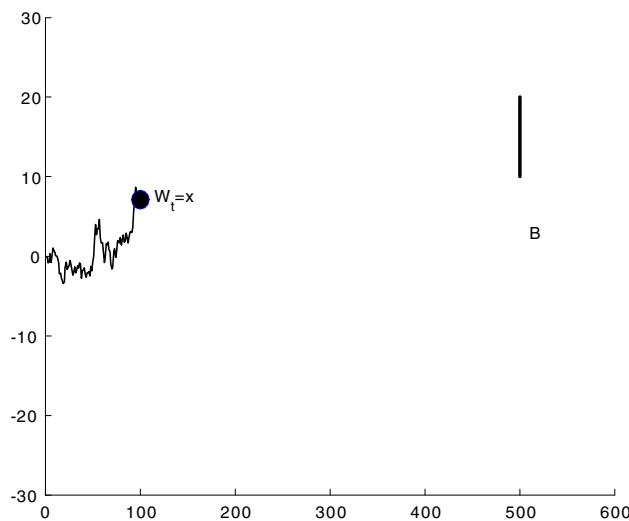


Fig. 13.2. Path of a Brownian motion.

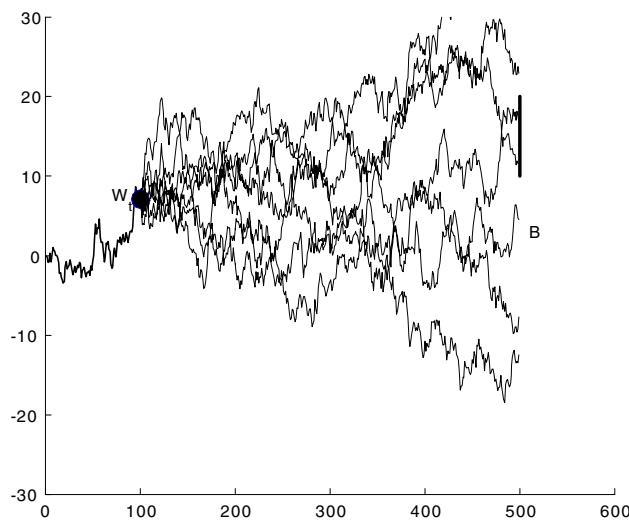


Fig. 13.3. Bundle of Brownian motions.

The *Kolmogorov forward equation*, also known as the *Fokker-Plank equation*, is given by

$$\frac{1}{2} \frac{\partial^2 \sigma^2(y, T) p}{\partial y^2} - \frac{\partial \mu(y, T) p}{\partial y} - \frac{\partial p}{\partial T} = 0. \quad (13.16)$$

The forward equation indicates what will happen in the future, whereas the backward equation gives the initial conditions that will lead to a certain state. In the finance literature, the backward equation is most useful.

At this stage, we provide a heuristic construction of the transition probability associated with the Brownian motion. Later on, based on the notion of martingale, we provide a construction for the general diffusion.

13.11 PDE associated with diffusions

The standard Brownian motion has transition density

$$p(y, T|x, t) = \frac{1}{\sqrt{2\pi}\sqrt{\tau}} \exp\left[-\frac{1}{2} \frac{(y-x)^2}{\tau}\right],$$

and partial derivatives

$$\begin{aligned} \frac{\partial p}{\partial x} &= \frac{\tau^{-\frac{3}{2}}}{\sqrt{2\pi}} (y-x) \exp\left[-\frac{1}{2} \frac{(y-x)^2}{\tau}\right], \\ \frac{\partial^2 p}{\partial x^2} &= \frac{\tau^{-\frac{3}{2}}}{\sqrt{2\pi}} \exp\left[-\frac{1}{2} \frac{(y-x)^2}{\tau}\right] \left[\frac{(y-x)^2}{\tau} - 1\right], \\ \frac{\partial p}{\partial t} &= -\frac{1}{2} \frac{\tau^{-\frac{3}{2}}}{\sqrt{2\pi}} \exp\left[-\frac{1}{2} \frac{(y-x)^2}{\tau}\right] \left[\frac{(y-x)^2}{\tau} - 1\right]. \end{aligned}$$

Hence, the backward equation is

$$\frac{\partial p}{\partial t} + \frac{1}{2} \frac{\partial^2 p}{\partial x^2} = 0.$$

Since $\frac{\partial p}{\partial t} = -\frac{\partial p}{\partial T}$, $\frac{\partial p}{\partial x} = \frac{\partial p}{\partial y}$ and $\frac{\partial^2 p}{\partial x^2} = \frac{\partial^2 p}{\partial y^2}$, we obtain easily the following forward equation

$$\frac{\partial p}{\partial T} - \frac{1}{2} \frac{\partial^2 p}{\partial y^2} = 0.$$

So far, we derived (13.15) in a particular case. An alternative, more general, proof of (13.15), which is not based on calculus, involves Ito's lemma. We consider not only the Brownian motion but the general diffusion (13.14). Consider the conditional expectation

$$f(y, t) = E[h(X_T)|X_t = y, t] = \int_z h(z)p(z, T|y, t)dz. \quad (13.17)$$

where $h(\cdot)$ is a smooth integrable function. Next, we take the expectation of $f(y, t)$ with respect to X_s for $s < t$. A remark on notation: values of random variables at time T are denoted by z , values at time t are denoted by y , and those at time s as x . We obtain

$$\begin{aligned} f(x, s) &= \int_z h(z)p(z, T|x, s)dz \\ &= \int_z h(z) \int_y p(z, T|y, t)p(y, t|x, s)dy \\ &= \int_y \left[\int_z h(z)p(z, T|y, t)dz \right] p(y, t|x, s)dy \\ &= E[E[h(X_T)|X_t = y, t]|X_s = x, s] \\ &= E[f(y, t)|X_s = x, s]. \end{aligned}$$

The first equality in this sequence is just the definition. The second line states that as one goes from $X_s = x$ to $X_T = z$, one must transit through all possible $X_T = y$. The way we have decomposed the problem assumes implicitly that the process is a Markovian process. The third equation regroups the different terms slightly differently. The next equation introduces the expectations operator to show that here we have implicitly derived a condition on conditional expectations. In the last equation, we replace the inner expectation by its value. This sequence of equations, starting from the end and proceeding upwards, shows that

$$E[f(y, t)|X_s = x, s] = f(x, s).$$

Thus, the best guess for the future value of $f(X_T, T)$ conditional on all available information at time s is $f(x, s)$. This result means that $f(X_t, t)$ is a martingale. It must, therefore, be that its drift is nil. From Ito's lemma, we may write

$$df = \left(\frac{1}{2}\sigma^2 \frac{\partial^2 f}{\partial x^2} + \mu \frac{\partial f}{\partial x} + \frac{\partial f}{\partial t} \right) dt + \sigma \frac{\partial f}{\partial x} dX_t.$$

Given that $f(x, t)$ has zero drift, it must be that

$$\frac{1}{2}\sigma^2 \frac{\partial^2 f}{\partial x^2} + \mu \frac{\partial f}{\partial x} + \frac{\partial f}{\partial t} = 0.$$

Using the definition of f , $f(x, t) = \int_z h(z)p(z, T|x, t)dz$, we get

$$\frac{\partial f}{\partial x} = \int_z h(z) \frac{\partial p}{\partial x} dz, \quad \frac{\partial^2 f}{\partial x^2} = \int_z h(z) \frac{\partial^2 p}{\partial x^2} dz, \quad \frac{\partial f}{\partial t} = \int_z h(z) \frac{\partial p}{\partial t} dz,$$

and we conclude that

$$\int_z h(z) \left(\frac{1}{2}\sigma^2 \frac{\partial^2 p}{\partial x^2} + \mu \frac{\partial p}{\partial x} + \frac{\partial p}{\partial t} \right) dz = 0. \quad (13.18)$$

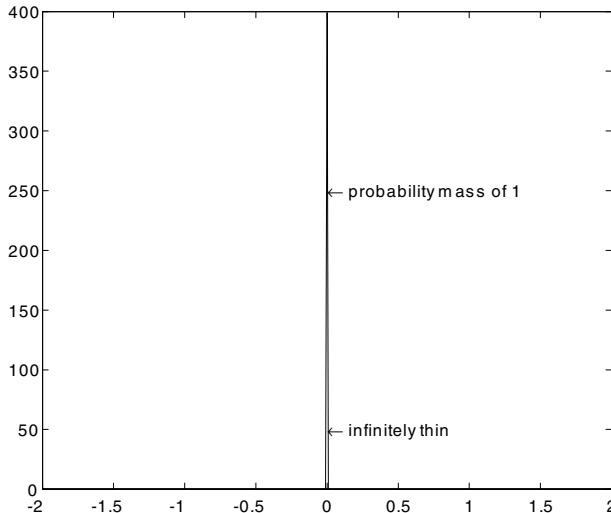


Fig. 13.4. Representation of a Dirac measure.

Equation (13.18) must hold for all values of h . Consider as a particular choice of h function the Dirac function denoted by $\delta_a(x)$. The Dirac function has been introduced by this physicist to characterize the probability function according to which some particle changes instantaneously state. Figure 13.4 gives a graphical representation of such a function. The function is actually a density. It has infinite height, no width, and a probability mass equal to 1.

With the choice of $h(z) = \delta_y(z)$ function we obtain that the integral (13.18) yields

$$\frac{1}{2}\sigma^2(x,t)\frac{\partial^2 p(y,T|x,t)}{\partial x^2} + \mu(x,t)\frac{\partial p(y,T|x,t)}{\partial x} + \frac{\partial p(y,T|x,t)}{\partial t} = 0. \quad (13.19)$$

For the Brownian motion, trivially $\mu = 0$ and $\sigma = 1$. This shows that with a somewhat more advanced approach, we recover the PDE of the Brownian motion.

13.12 Feynman-Kac formula

At this stage, we have shown that a diffusion written

$$dX_t = \mu(X_t, t) dt + \sigma(X_t, t) dW_t \quad (13.20)$$

is associated with a transition probability $p(y, T|x, t)$, that obeys the PDE given by (13.19). From there on, it is tempting to seek solutions to partial

differential equations that are characterized by transition probabilities, diffusions, and, as we will see, expectations. Typically, in finance, we are dealing with price processes $f(x, t)$ where the associated PDE may be written, suppressing the arguments, as

$$\frac{1}{2}\sigma^2 \frac{\partial^2 f}{\partial x^2} + \mu \frac{\partial f}{\partial x} + \frac{\partial f}{\partial t} = rf, \quad (13.21)$$

under a boundary condition

$$f(z, T) = h(z). \quad (13.22)$$

Using the insights of the proof leading to the PDE of transition probabilities, we may consider the function

$$f(x, t) = e^{-r\tau} g(x, t), \quad (13.23)$$

$$g(x, t) = \int_z h(z) p(z, T|x, t) dz. \quad (13.24)$$

Differentiation with respect to t yields

$$\frac{\partial f}{\partial t} = re^{-r\tau} g + e^{-r\tau} \frac{\partial g}{\partial x}.$$

This computation shows that if we substitute f and all its partial differentials by g and the corresponding partial differentials, we are led, after a simplification by $e^{-r\tau}$, which is never negative, to the new partial differential equation

$$\frac{1}{2}\sigma^2 \frac{\partial^2 g}{\partial x^2} + \mu \frac{\partial g}{\partial x} + \frac{\partial g}{\partial t} = 0.$$

At this stage, we have slightly simplified the problem in that the right-hand side of the differential equation is zero. Using analogous arguments as in the previous section, we notice that also the transition probability p must satisfy this transition probability. Retracing all the steps from the previous section, the stochastic process X_t with SDE given by (13.20) is compatible with this PDE. It remains to be shown that equations (13.23)–(13.24) are a solution compatible with the boundary condition (13.22). This is achieved by noticing that p is a transition probability. As t converges to T , the probability of movement will decrease dramatically. Actually, in the limit this probability melts down to a Dirac measure. We have

$$\lim_{t \rightarrow T} p(z, T|x, t) = \delta_x(z).$$

It follows that

$$\lim_{t \rightarrow T} e^{-r\tau} \int_z h(z) p(z, T|x, t) dz = h(x).$$

Henceforth, in the limit, as t converges to T , the solution f converges to the boundary condition. This shows that f verifies the PDE and is compatible with the boundary. Inspection of equations (13.23)–(13.24) shows that the solution to the problem defined by (13.21) and (13.22) is nothing else but a discounted conditional expectation $f(x, t) = e^{-r\tau} E[h(X_T)|X_t = x, t]$, where the transition probability is determined by $dX_t = \mu X_t dt + \sigma X_t dW_t$, that is

$$p(x_T, T|x_t, t) = \frac{1}{\sqrt{2\pi}\sigma\sqrt{\tau}x_T} \exp \left[-\frac{1}{2} \left(\frac{\log(x_T/x_t) - (\mu - \frac{1}{2}\sigma^2)\tau}{\sigma\sqrt{\tau}} \right)^2 \right].$$

In practical applications, as the reader may see in the main part of the book, quite some guessing may be required to go from the PDE, to the SDE, and eventually to the transition probability that is ultimately needed to compute the conditional expectation.

Martingale and Changing Measure

In this chapter, we provide some technical details on the fundamentals regarding martingale and the changing measure approach to option pricing. This approach is very popular among some mathematicians because for martingales, many results have already been established. We perform a change of measure to convert the price process into a martingale. Based on this example, we explain the concepts Radon-Nikodym derivative, Girsanov's theorem, and the Kunita and Watanabe representation theorem. Further references to the issues treated in this chapter are Karatzas and Shreve (1988) or Shreve (2004).

14.1 Martingales

Let M_s be a stochastic process. M_s is a martingale if

$$E[M_s | \mathcal{F}_t] = M_t, \text{ for } s > t,$$

where \mathcal{F}_t represents “information available at time t ”. Note that in general, \mathcal{F}_t should be defined meticulously; what information means and how it is generated, etc. For our purpose here, information means that \mathcal{F}_t must include all time t information on M_t , although it may contain other information also.

Example: A Wiener process is a martingale, since

$$E[W_s | W_t] = W_t, \text{ for } s > t.$$

There are several extensions to this result. For example, the aggregated increment of a Wiener process

$$W_s = \int_0^s dW_u,$$

is also a martingale since

$$\begin{aligned} E \left[\int_0^s dW_u \middle| \int_0^t dW_u \right] &= E \left[\int_t^s dW_u \middle| \int_0^t dW_u \right] + \int_0^t dW_u, \\ &= \int_0^t dW_u. \end{aligned}$$

The second equality above relies on the fact that the expectation of the future evolution of the Wiener process based on current information is nil. Furthermore, if h_u is some well-behaved function, then

$$E \left[\int_0^s h_u dW_u \middle| \int_0^t h_u dW_u \right] = \int_0^t h_u dW_u.$$

It is easy to prove this result by invoking the definition of a stochastic integral

$$\int_0^s h_u dW_u = \lim_{N \rightarrow \infty} \sum h_{t_{i-1}} (W_{t_i} - W_{t_{i-1}}),$$

for $t_0 = 0 < t_1 < \dots < t_N < 1$ and $\max_{i=1, \dots, N} t_i - t_{i-1}$ converges to 0 as N converges to infinity. At time t_{i-1} , $h_{t_{i-1}}$ is known, but as W_{t_i} is a martingale, it follows that each element of the sum must obey

$$E [h_{t_{i-1}} (W_{t_i} - W_{t_{i-1}})] = 0,$$

by invoking conditional expectations.

14.2 Changing probability of a normal distribution

Now we introduce the concept called *change of time*. This is done by changing the drift of the Wiener process. Consider the moment generating function of a normal variate $N(\mu, \sigma^2)$

$$E_P[e^{\lambda x}] = \int e^{\lambda x} p(x) dx = \int e^{\lambda x} \frac{1}{\sqrt{2\pi}\sigma} \exp \left[-\frac{1}{2} \left(\frac{x-\mu}{\sigma} \right)^2 \right] dx, \quad (14.1)$$

for $\lambda \in \mathbb{R}$. The symbol E with a subscript P means the expectation is taken under the original P -measure, here a $N(\mu, \sigma^2)$.

From (14.1), we have

$$\begin{aligned} E_P[e^{\lambda x}] &= \int \frac{1}{\sqrt{2\pi}\sigma} \exp \left[-\frac{1}{2\sigma^2} (-2\sigma^2\lambda x + x^2 - 2\mu x + \mu^2) \right] dx \\ &= \int \frac{1}{\sqrt{2\pi}\sigma} \exp \left[-\frac{1}{2\sigma^2} \left[(x - (\mu + \sigma^2\lambda))^2 - \sigma^4\lambda^2 - 2\mu\sigma^2\lambda \right] \right] dx \\ &= \int e^{\frac{1}{2}\sigma^2\lambda^2 + \mu\lambda} \frac{1}{\sqrt{2\pi}\sigma} \exp \left[-\frac{1}{2\sigma^2} [x - (\mu + \sigma^2\lambda)]^2 \right] dx \\ &= \exp \left(\lambda\mu + \frac{1}{2}\lambda^2\sigma^2 \right). \end{aligned} \quad (14.2)$$

Rearranging (14.2) yields

$$\begin{aligned} E_P \left[\exp \left(\lambda(x - \mu) - \frac{1}{2}\lambda^2\sigma^2 \right) \right] &= 1, \\ \int \exp \left(\lambda(x - \mu) - \frac{1}{2}\lambda^2\sigma^2 \right) p(x) dx &= 1. \end{aligned} \quad (14.3)$$

Note that the expression in the integral in (14.3) defines a function that is positive and integrates up to 1. We conclude that we have just constructed a new probability measure, which we define at this stage as

$$q(x) \equiv \exp \left[\lambda(x - \mu) - \frac{1}{2}\lambda^2\sigma^2 \right] p(x). \quad (14.4)$$

Next, we expand all the terms in (14.4) and get

$$\begin{aligned} q(x) &= \exp \left(\lambda(x - \mu) - \frac{1}{2}\lambda^2\sigma^2 \right) \frac{1}{\sqrt{2\pi}\sigma} \exp \left[-\frac{1}{2} \left(\frac{x - \mu}{\sigma} \right)^2 \right] \\ &= \frac{1}{\sqrt{2\pi}\sigma} \exp \left[-\frac{1}{2\sigma^2} \left[(x - \mu)^2 - 2\sigma^2\lambda(x - \mu) + \sigma^4\lambda^2 \right] \right] \\ &= \frac{1}{\sqrt{2\pi}\sigma} \exp \left[-\frac{1}{2\sigma^2} \left[x^2 - 2x\mu + \mu^2 - 2\sigma^2\lambda(x - \mu) + \sigma^4\lambda^2 \right] \right] \\ &= \frac{1}{\sqrt{2\pi}\sigma} \exp \left[-\frac{1}{2\sigma^2} \left[x - (\mu + \lambda\sigma^2) \right]^2 \right]. \end{aligned} \quad (14.5)$$

We can see from (14.5) that $q(x)$ is the density of a new normal random variable with mean $\mu + \lambda\sigma^2$ and variance σ^2 . In other words, under the initial probability P , $X \sim N(\mu, \sigma^2)$, but under the new probability Q , $X \sim N(\mu + \lambda\sigma^2, \sigma^2)$. That is,

$$E_P [X] = \mu, \quad \text{and} \quad E_Q [X] = \mu + \lambda\sigma^2.$$

It follows that we can, by changing the probability measure, adjust the mean to create a new normal variate under the Q measure. If the desired mean is a , then one could set

$$a = \mu + \lambda\sigma^2 \Rightarrow \lambda = (a - \mu)/\sigma^2.$$

In general, we could change the drift in many ways; in particular, it is useful to be able to change probability in such a way as to achieve a targeted mean under the new probability measure. This is an important first step in the study of derivative pricing under the martingale measure.

14.3 Radon-Nikodym derivative

Rearranging the terms in (14.4), we get the ratio

$$\frac{q(x)}{p(x)} = \exp\left(\lambda(x - \mu) - \frac{1}{2}\lambda^2\sigma^2\right). \quad (14.6)$$

This expression is an example of the *Radon-Nikodym derivative*. The reason why this ratio is called a derivative comes from the fact that any smooth density can be expressed as the derivative of a *cdf*. For example, if $Q(x) = \int_{-\infty}^x q(u)du$, then $dQ(x) = q(x)$. Hence, the Radon-Nikodym derivative in (14.6) can also be written as $dQ(x)/dP(x)$.

Formally, if P is finite (i.e., $P(x) < \infty$) and Q is absolutely continuous with respect to P (i.e., if $p(x) = 0 \Rightarrow q(x) = 0$), then there exists a nonnegative function

$$M(x) = \frac{q(x)}{p(x)},$$

called the *Radon-Nikodym derivative* of q with respect to p , and it is unique.

The Radon-Nikodym derivative is a strictly positive random variable with $E[M(x)] = 1$ and the expectations of the random variable X under the two probability measures are related by

$$\int x dQ(x) = \int x M(x) dP(x).$$

If P is also absolutely continuous with respect to Q , then P and Q are called *equivalent probability measures*.

14.4 Girsanov's theorem

We can also apply the changing probability procedures to stochastic processes. We consider the arithmetic Brownian motion

$$\begin{aligned} dX_t &= \mu dt + \sigma dW_t, \text{ with } X_0 = 0, \\ X_t &= \mu t + \sigma W_t \sim N(\mu t, \sigma^2 t), \end{aligned}$$

with $X_0 = 0$. Under the initial probability P , the mean of the process is μt . We could change the mean of X_t to 0 under the Q probability measure by setting

$$0 = \mu t + \lambda \sigma^2 t \Rightarrow \lambda = -\mu/\sigma^2. \quad (14.7)$$

If we substitute the value of λ in (14.7) into (14.5), we get

$$q(x) = \frac{1}{\sqrt{2\pi}\sigma\sqrt{t}} \exp\left(-\frac{x^2}{2\sigma^2 t}\right).$$

Then, we have

$$E_Q[X_t] = \int x q(x) dx = \int x \frac{1}{\sqrt{2\pi}\sigma\sqrt{t}} \exp\left(-\frac{x^2}{2\sigma^2 t}\right) dx. \quad (14.8)$$

Equation (14.8) is the evaluation of the expected value of a normal variate $N(0, \sigma^2 t)$. These computations show that by changing the drift of the normal random variable, we obtain a centered normal random variable under a new probability measure Q . What we did here for X_t can also be done for the Brownian motion W_t . One may change its drift while retaining the fact that under a new measure it will be with mean zero. This changing of probability measure is usually presented under the heading *Girsanov's theorem*. It describes how Brownian motion transforms under different probability measures. It provides the conditions under which the Radon-Nikodym derivative exists for continuous stochastic process, and it is used for converting a non-standard Brownian motion into a *standard Brownian motion* that has the martingale property, i.e., has zero drift.

Assume that W_t is a Brownian motion under some probability P , the original measure. Then $W_t \sim N(0, t)$. Consider the transformation

$$W_t^* = W_t + \beta t,$$

where β is a constant. Now W_t^* has a drift of βt under the probability P . Setting as above $\lambda = -\beta$ one can undo the drift. Under a new probability induced by λ , W_t^* will become $N(0, t)$ a new standard Brownian motion.¹ The Radon-Nikodym derivative is, adapting (14.6)

$$M_t = \exp \left(-\beta W_t - \frac{1}{2} \beta^2 t \right).$$

The relation $dQ = M_t dP$ defines a probability measure Q which is equivalent to P . Taking Ito's lemma, we also obtain that M_t is a martingale since its drift is nil.

In the main part of the book, we use this technique to change the drift of a Brownian motion yet retaining the property that under a new probability, the drift changed Brownian motion is still a standard Brownian motion.

14.5 Martingale representation theorem

There is another useful theorem, called Kunita and Watanabe representation theorem, which relates a martingale with respect to W_t to some well-specified function η_t .

Theorem 14.1. *A martingale M_t , where the information is solely generated by the Brownian motion W_t , can be written as*

$$M_t = M_0 + \int_0^t \eta_s dW_s$$

where η_s is a well-behaved function.

¹ Now t takes the role of σ^2 .

The formal proof of this theorem can be found in many probability textbooks.² The intuition behind it is that M_t is a martingale, and hence, it must have a drift that is equal to zero. Also, given that W_t generates the information driving the martingale, it must be that M_t is a function of W_t . Assume therefore that $M_t = f(W_t, \text{ terms})$. Using Ito's lemma and the fact that the drift of the resulting dynamics must be nil, one obtains that

$$dM_t = f_W(W_t, \text{ terms})dW_t.$$

A trivial change of notation then gives the announced result. Unfortunately, the theorem has nothing to say in terms of the expression taken by η_t . It just says that the function η_t must exist and that it will be well behaved. In the case that the expression of the martingale M_t is known, then it becomes possible to infer η_t .

² See for instance Karatzas and Shreve (1988).

15

Characteristic Functions and Fourier Transforms

In this chapter, we introduce the characteristic function and some of its properties (Section 15.1). We also illustrate here various results of Fourier analysis, which is related to the inversion and integration of characteristic function (Section 15.2). The intention here is not to provide an exhaustive treatment of these topics but to provide sufficient basic concepts for studying option pricing. For more advanced references, see Ushakov (1999) or Kendall and Stuart (1977) for characteristic functions and Brigham (1988) for fast Fourier transform.

15.1 Characteristic functions

Let X be a random variable. The *characteristic function* of X is defined as

$$\phi_X(u) = E[e^{iuX}] = \int_{-\infty}^{\infty} e^{iux} dF_X(x), \quad (15.1)$$

$$= \int_{-\infty}^{\infty} e^{iux} f_X(x) dx, \quad (15.2)$$

where $F_X(x)$ is the cumulative distribution function (*cdf*), $f_X(x)$ is a probability distribution function (*pdf*), and u is a real number. The last expression (15.2) is valid only when $f_X(x)$ exists, whereas (15.1) is valid regardless of whether a *pdf* exists or not. For instance, if X is a discrete random variable, then (15.1) exists but not (15.2). In the following, we will assume that (15.2) holds, as it is the case in most finance related applications.

Example: For a normal random variate $X \sim \mathcal{N}(\mu, \sigma^2)$, we get the CF of X as

$$\begin{aligned}
\phi_X(u) &= \int_R \exp(iux) \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] dx \\
&= \int_R \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2\sigma^2}(x^2 - 2x\mu + \mu^2 - 2\sigma^2 i xu)\right] dx \\
&= \int_R \exp\left(\mu i u - \frac{1}{2}\sigma^2 u^2\right) \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}[x - (\mu + \sigma^2 i u)]^2\right) dx \\
&= \exp\left(\mu i u - \frac{1}{2}\sigma^2 u^2\right). \tag{15.3}
\end{aligned}$$

The concept of characteristic function is important because it allows us to characterize certain distributions that cannot be explicitly described by a *pdf*. One such example is the stable distribution. Also, the class of distributions known as Lévy processes is defined only via their characteristic function.

15.1.1 Basic properties

All CFs have the following properties

- (i) $\phi_X(0) = 1$, since $\exp(0) = 1$.
- (ii) $|\phi_X(u)| \leq 1$, for all real u , since

$$|\phi_X(u)| = \left| \int_{-\infty}^{\infty} e^{iux} f_X(x) dx \right| \leq \int_{-\infty}^{\infty} |e^{iux}| f_X(x) dx \leq 1.$$

- (iii) $\overline{\phi_X(u)} = \phi_X(-u)$, since¹

$$\overline{\phi_X(u)} = \int_{-\infty}^{\infty} \overline{e^{iux}} f_X(x) dx = \int_{-\infty}^{\infty} e^{-iux} f_X(x) dx = \phi_X(-u).$$

- (iv) $\phi_X(u)$ is continuous for all real u . This follows directly from the fact that the exponential function is a continuous function.
- (v) $\phi_{aX+b}(u) = e^{iub} \phi_X(au)$. This useful result follows immediately from $\phi_{aX+b}(u) = E[e^{iu(aX+b)}]$.

15.1.2 Moments and the characteristic function

A very useful property of CF is that derivatives with respect to u evaluated at 0 yield the moments of X

¹ We use the horizontal bar atop of $\phi_X(t)$ to denote the complex conjugate of $\phi_X(t)$. For $z = a + ib$, the complex conjugate of z is \bar{z} , where $\bar{z} = a - ib$.

$$\begin{aligned}\frac{d\phi_X(u)}{du}\Big|_{u=0} &= i \int x f_X(x) dx = i E[X], \\ \frac{d^2\phi_X(u)}{du^2}\Big|_{u=0} &= - \int x^2 f_X(x) dx = -E[X^2], \\ &\vdots \\ \frac{d^n\phi_X(u)}{du^n}\Big|_{u=0} &= i^n \int x^n f_X(x) dx = i^n E[X^n].\end{aligned}$$

Hence, the characteristic function could be used to compute the moments of a random variable provided that the moment exists.

15.1.3 Convolution theorem

The characteristic function is also useful in the construction and the study of the density of a sum of random variables. If X_1 and X_2 are two independent random variables and $X = X_1 + X_2$, then $\phi_X(u) = \phi_{X_1}(u)\phi_{X_2}(u)$. The proof follows immediately from the properties of the expectations operator, for which it is known that the expectation of two independent random variables equals the product of the expectations.

$$\phi_X(u) = E[e^{iu(X_1+X_2)}] = E[e^{iuX_1}e^{iuX_2}] = \phi_{X_1}(u)\phi_{X_2}(u).$$

It follows that the product of two characteristic functions is a characteristic function, and if $\phi_X(u)$ is a characteristic function, then $(\phi_X(u))^2$ is also a characteristic function. Obviously, this theorem extends to the sum of even more independent random variables as we will see in the following proposition.

Proposition 15.1. *Let X_1, \dots, X_n be a set of n independent random variables, and let c_1, \dots, c_n be n real number constants. If all the random variables have the same density with characteristic function $\phi_X(u)$, then the following relation holds:*

$$\phi_{c_1X_1+\dots+c_nX_n}(t) = \phi_X(c_1u) \times \dots \times \phi_X(c_nu).$$

Proof. We have

$$\begin{aligned}\phi_{c_1X_1+\dots+c_nX_n}(u) &= E[\exp(i(c_1X_1 + \dots + c_nX_n)u)] \\ &= E[\exp(ic_1uX_1) \times \dots \times \exp(ic_nuX_n)] \\ &= E[\exp(ic_1uX_1)] \times \dots \times E[\exp(ic_nuX_n)] \\ &= \phi_X(c_1u) \times \dots \times \phi_X(c_nu).\end{aligned}$$

The first equality follows from the definition of the characteristic function. The third equality follows from the assumption of independence among the random variables. Furthermore, if $c_1 = c_2 = \dots = c_n = 1$, then

$$\phi_{X_1+\dots+X_n}(u) = [\phi_X(u)]^n.$$

15.1.4 Uniqueness

Two *cdfs* $F_1(x)$ and $F_2(x)$ are identical if and only if their characteristic functions $\phi_1(u)$ and $\phi_2(u)$ are identical, i.e., $\phi_1(u)$ and $\phi_2(u)$ are equal for all values of u . In other words, the characteristic function characterizes completely the underlying *pdf* and *cdf* of X . This property is particularly useful when working on the inverse Fourier transform later. In some cases, this property is sufficient for “guessing” the solution, as shown in the example below.

Suppose that (X_1, \dots, X_n) is a set of normal random variables, each with mean μ_i and variance σ_i^2 , for $i = 1, \dots, n$. Also let c_1, \dots, c_n be a set of real constants, then we have

$$\begin{aligned} & \phi_{c_1 X_1 + \dots + c_n X_n}(u) \\ &= \exp\left(\mu_1 c_1 i u - \frac{1}{2} \sigma_1^2 c_1^2 u^2\right) \times \dots \times \exp\left(\mu_n c_n i u - \frac{1}{2} \sigma_n^2 c_n^2 u^2\right) \\ &= \exp\left[\left(\mu_1 c_1 + \dots + \mu_n c_n\right) i u - \frac{1}{2} \left(\sigma_1^2 c_1^2 + \dots + \sigma_n^2 c_n^2\right) u^2\right]. \end{aligned}$$

We recognize that this is the characteristic function of a new normal random variable with mean $\mu_1 c_1 + \dots + \mu_n c_n$ and variance $\sigma_1^2 c_1^2 + \dots + \sigma_n^2 c_n^2$. This result is due to the well-known property that a linear combination of independent normal random variables yields a new normal random variable.

15.1.5 Inversion theorem

It is often impossible to just “guess” the distribution function from the characteristic function. In such cases, it is necessary to have an explicit formula for inverting the distribution from the characteristic function. The inversion theorem is just such a formula. Many modern option pricing models, such as the stochastic volatility option pricing model, use the inversion of characteristic function to provide semi-closed form option pricing formulae.

Proposition 15.2. (*Inversion theorem*) If $\phi_X(t)$ is the characteristic function of the *cdf* $F_X(x)$, then

$$F_X(x) = \frac{1}{2} - \frac{1}{\pi} \int_0^\infty \frac{\text{Im}[e^{-itx}\phi_X(t)]}{t} dt. \quad (15.4)$$

A detailed proof for the inversion theorem may be found in Cramér (1946) or Kendall and Stuart (1977, chapter 4).

Lemma 15.3. Let ξ and $t \neq 0$ be real numbers, then

$$\frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{\sin t\xi}{t} dt = \text{sgn}(\xi),$$

where $\text{sgn}(\xi)$ is the sign function, taking the values -1 , 0 , and $+1$, respectively, for $\xi < 0$, $\xi = 0$, and $\xi > 0$.

Proof. We admit this lemma. A formal proof is based on an integration in the complex domain. One may consult the above mentioned references for a proof.

Proof. (Inversion theorem) From the definition of $\text{sgn}(\cdot)$, it follows that

$$\begin{aligned}\int_{-\infty}^{\infty} \text{sgn}(x-u)f_X(u)du &= \int_{-\infty}^x f_X(u)du + \int_x^{\infty} (-1)f_X(u)du \\ &= 2F_X(x) - 1.\end{aligned}\quad (15.5)$$

Now, for $z = a + bi$, $z - \bar{z} = 2bi$ and $\text{Im}(z) = \frac{1}{2i}(z - \bar{z})$, we get

$$\text{Im}[e^{-itx}\phi_X(t)] = \frac{1}{2i} [e^{-itx}\phi_X(t) - \overline{e^{-itx}\phi_X(t)}].$$

Then from property (iii) of CF in Section 15.1.1, we get

$$\text{Im}[e^{-itx}\phi_X(t)] = -\frac{1}{2i} [e^{itx}\phi_X(-t) - e^{-itx}\phi_X(t)].$$

If we substitute this result into the RHS of (15.4), we get

$$\frac{1}{2} + \frac{1}{2\pi} \int_0^{\infty} \frac{[e^{itx}\phi_X(-t) - e^{-itx}\phi_X(t)]}{it} dt \quad (15.6)$$

$$\begin{aligned}&= \frac{1}{2} + \frac{1}{2\pi} \int_0^{\infty} \left[\frac{e^{itx}}{it} \int_{-\infty}^{\infty} e^{-itu} f_X(u) du - \frac{e^{-itx}}{it} \int_{-\infty}^{\infty} e^{itu} f_X(u) du \right] dt \\ &= \frac{1}{2} + \frac{1}{2\pi} \int_0^{\infty} \left[\int_{-\infty}^{\infty} \frac{e^{it(x-u)} - e^{-it(x-u)}}{it} f_X(u) du \right] dt.\end{aligned}\quad (15.7)$$

Since the complex number $z = a + bi$ can also be written as $|z|(\cos \theta + i \sin \theta)$ where $\theta = \arctan(b/a)$, we have, from the results by Euler on complex exponential,

$$e^{i\theta} = \cos \theta + i \sin \theta,$$

and

$$\sin \theta = (e^{i\theta} - e^{-i\theta})/2i.$$

Substituting this result into (15.7) gives

$$\begin{aligned}&\frac{1}{2} + \frac{1}{2\pi} \int_{-\infty}^{\infty} \left[\int_0^{\infty} \frac{2 \sin t(x-u)}{t} dt \right] f_X(u) du \\ &= \frac{1}{2} + \frac{1}{2\pi} \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \frac{\sin t(x-u)}{t} dt \right] f_X(u) du,\end{aligned}$$

and from Lemma 15.3, it becomes

$$\frac{1}{2} + \frac{1}{2\pi} \int_{-\infty}^{\infty} \pi \text{sgn}(x-u) f_X(u) du.$$

Finally from (15.5), we get

$$\frac{1}{2} + \frac{1}{2}(2F_X(x) - 1) = F_X(x).$$

Next, we derive alternative expressions of the inversion theorem.

Proposition 15.4. (*Cdf and pdf*) *We have the following equivalent expressions*

$$F_X(x) = \frac{1}{2} - \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-itx}\phi_X(t)}{it} dt \quad (15.8)$$

$$= F_X(0) - \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-itx} - 1}{it} \phi_X(t) dt \quad (15.9)$$

and

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \phi_X(t) dt \quad (15.10)$$

Proof. From (15.6), we could write

$$\begin{aligned} F_X(x) &= \frac{1}{2} + \frac{1}{2\pi} \int_0^{-\infty} \frac{e^{-itx}\phi_X(t)}{it} dt - \frac{1}{2\pi} \int_0^{\infty} \frac{e^{-itx}\phi_X(t)}{it} dt \\ &= \frac{1}{2} - \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{e^{-itx}\phi_X(t)}{it} dt. \end{aligned} \quad (15.11)$$

Next, we have

$$\begin{aligned} F_X(0) &= \frac{1}{2} - \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\phi_X(t)}{it} dt, \\ \frac{1}{2} &= F_X(0) + \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\phi_X(t)}{it} dt. \end{aligned}$$

Substitute this for $\frac{1}{2}$ in (15.11) gives (15.9).

This formula has the advantage that it expresses $F_X(x)$ with respect to some central value, $F_X(0)$. For this reason, this alternative expression may produce more accurate results when the $F_X(x)$ is evaluated numerically. Differentiating (15.8) with respect to x gives the density function in (15.10).

Inspection of the proof of the inversion theorem shows that the theorem holds even when $f_X(x)$ is not defined. In this case, the inversion holds for

$$F_X(x) = \int_{-\infty}^x f_X(u) du,$$

under the condition that $F_X(+\infty) = K$, where K is a real constant. This insight shows that the inversion formula (15.8) holds for a much larger class

of distributions. If a characteristic function $\phi_X(u)$ is absolutely integrable over $(-\infty, +\infty)$ then the corresponding *cdf* $F_X(x)$ and the *pdf* $f_X(x)$ are also continuous. There are various efficient numerical methods for evaluating the integral in (15.10), one of which is the fast Fourier transform described now.

Turning to the density, the integral over which we need to perform our computation may be somewhat simplified. To do so, define $h(u) = e^{-iux}\phi_X(u)$ so that

$$h(-u) = e^{iux}\phi_X(-u) = \overline{e^{iux}\phi_X(u)}.$$

It results that

$$h(u) + h(-u) = 2 \operatorname{Re} [e^{-iux}\phi_X(u)],$$

and that eventually

$$f_X(x) = \frac{1}{\pi} \int_0^\infty \operatorname{Re} [e^{-iux}\phi_X(u)] du. \quad (15.12)$$

This formula yields in practice good results. There are also integrals that are not defined at 0. Since here this bound appears explicitly, it is easier to control the integrand around a neighborhood of 0.

15.2 Fourier transform and characteristic function

The formulae that we have obtained so far are very useful in that they show that for a density it is possible to obtain the associated characteristic function, and vice versa, for a given characteristic function, we may obtain the density. As an example, consider the density of the $\mathcal{N}(0, 1)$,

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{1}{2} \left(\frac{x - \mu}{\sigma} \right)^2 \right),$$

then, a simple computation seen as an example at the beginning of this chapter, shows that

$$\phi_X(u) = \exp \left(i\mu u - \frac{1}{2}\sigma^2 u^2 \right).$$

Now, applying (15.10) gives us the integral

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-iux} e^{\mu iu - \frac{1}{2}\sigma^2 u^2} du. \quad (15.13)$$

The reader may verify, using a direct computation, that this integral gives the density $f_X(x)$. This result is often useful in its own because it allows us in many cases to obtain the explicit density associated with a difficult problem. One such instance is the derivation of the conditional density of

a random variable following a square root process, as seen in Chapter 12. Sometimes, this type of computation yields a semi-closed form solution as to Heston (1994).

Inspection of the computations involved in (15.13) shows, however, that for each value of x a relatively long computation needs to be performed. Even if one uses advanced numerical integration rules, if one evaluates each integral on n_1 points and the function f over n_2 points, there still remain $n_1 n_2$ computations. For large n_1 and n_2 , these computations may be too time consuming for an actual implementation. This is where certain techniques developed for Fourier transforms are useful. As a matter of fact, the type of computations indicated above are often called Fourier transform and inverse Fourier transform. Indeed, the characteristic function may be viewed as a special case of the Fourier transform in the case that the function used in the integration is a density. The reason for considering the Fourier transform is that efficient numerical techniques haven been devised by Cooley and Tukey to compute the discrete Fourier transform, DFT, and its inverse, IDFT. For the case where $n_1 = n_2 = n$, the computation mentioned above would involve n^2 operations. The DFT and IDEF can be computed in only $n \log(n)$ operations using the so-called fast Fourier transform (FFT) algorithm. The reader does not need to understand the smart evaluation procedure that are involved in the FFT and indeed, we will not go into the details of the algorithm. Instead, we explain how the FFT routine can be executed from programs. Typically, these programs dispose of two functions named **fft** and **ifft** that implement the transform and inverse transform using the Cooley-Tukey algorithm.² For a given vector of numbers x_n , where $n = 1, \dots, N$, these functions compute

$$\begin{aligned} X_k &= \sum_{n=1}^N x_n e^{-i2\pi(k-1)(n-1)/N}, & \text{for } k = 1, \dots, N, \\ x_n &= \frac{1}{N} \sum_{k=1}^N X_k e^{i2\pi(k-1)(n-1)/N}, & \text{for } n = 1, \dots, N, \end{aligned} \quad (15.14)$$

where N is typically a power of 2. The issue is how our integrals can be brought into this discrete form. To relate characteristic function with continuous Fourier transform, we apply to (15.10) a change of variable $u/2\pi = w$ and $du = 2\pi dw$. It follows that

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} e^{-i2\pi wx} \phi_X(w) dw, \\ \phi_X(w) &= \int_{-\infty}^{\infty} e^{i2\pi w} f_X(x) dx. \end{aligned}$$

Similar expressions can be obtained for (15.4), indicating that the type of computation can also be useful in a general option valuation context, where

² A program that contains such commands, **fft** and **ifft**, is MATLAB.

these formula are useful. We will presently show how the first integral may be discretized into an expression such as (15.14). We have, in the general case, by approximating the integral by Simpson's rule

$$f_X(x) \approx \frac{1}{\pi} \sum_{n=1}^N e^{-iu_n x} \phi_X(u_n) \Delta_u, \quad (15.15)$$

where $u_n = \Delta_u(n - 1)$. Notice that we will explain later how Δ_u is chosen. In (15.15) the sign \approx means “approximately equal to”. Now, let $[x_L, x_U]$ be the support of x and define

$$\Delta_x = \frac{x_U - x_L}{N - 1},$$

so that the following gives a discretization of the x range

$$x_k = x_L + \Delta_x(k - 1), \quad \text{for } k = 1, \dots, N.$$

With this notation, (15.15) becomes

$$f_X(x_k) = \operatorname{Re} \left[\frac{1}{\pi} e^{-i\Delta_u(n-1)x_L} \sum_{n=1}^N e^{-i\Delta_u \Delta_x(k-1)(n-1)} \phi_X(u_n) \Delta_u \right].$$

The first term involving x_L may be computed once and for all and does not provide difficulties. To cast this formula into the expression of the FFT, set

$$\begin{aligned} \Delta_u \Delta_x &= \frac{2\pi}{N}. \\ \Rightarrow \Delta_u &= \frac{2\pi}{N} \frac{N-1}{x_U - x_L} \approx \frac{2\pi}{x_U - x_L}, \quad \text{for large } N, \end{aligned}$$

giving the expression for Δ_u . We notice that the use of the FFT imposes strong conditions on the grid over which the u_n are defined. If the range over which the x are defined is large, then Δ_u will be small and vice versa.

As an example, consider the case of the normal density for which the characteristic function can be written as

$$\phi_X(u_n) = e^{\mu i \Delta_u(n-1) - \frac{1}{2} \sigma^2 [\Delta_u(n-1)]^2}.$$

As a numerical exercise, we use a normal distribution $\mathcal{N}(4, 1)$. A plot of the real part of the CF would reveal its oscillatory character. In Figure 15.1, the dashed curve displays the *pdf* obtained after running the `fft` command in MATLAB. We notice that the shape of the “*pdf*” is correct, but the level experienced a vertical shift. This phenomenon is known as *discretization bias*. For the case at hand, a quick fix would be to compute (15.15) over a fine grid for a wide range of u at a given point x , and to use the result to correct for the bias. The short dash line in Figure 15.1 shows the density after this bias correction.

We may conclude that characteristic function methods are most powerful techniques that the empirically interested finance researcher should know. This section demonstrates that the FFT is indeed a very efficient way for computing the Fourier integral, but it also reminds the readers that the technique should be used carefully.

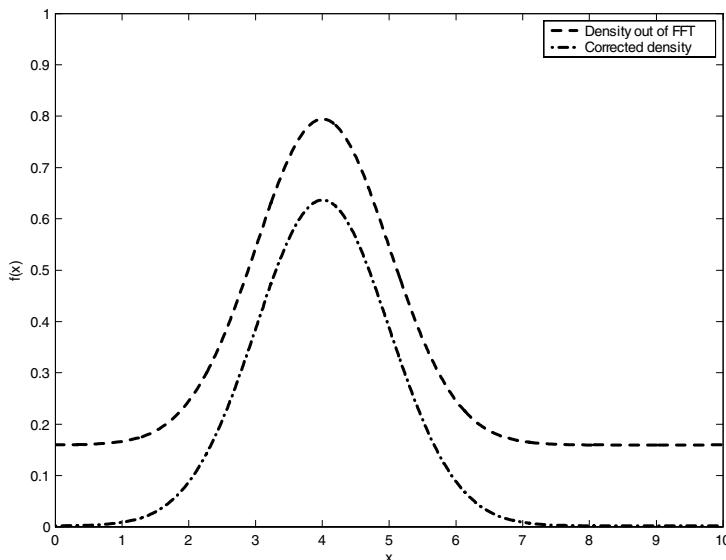


Fig. 15.1. Pdf result from discrete FFT inversion of CF of a normal variate $\mathcal{N}(4, 1)$.

Jump Processes

This chapter features jump, which is a key element in Chapters 4, 7, and 12. Jump processes are useful at modeling temporal behavior of asset prices such as crashes or the arrival of orders in a microstructure context. Jumps are also the key ingredients in Lévy processes, the latter of which will be discussed in detail in Chapter 17. For the reader interested in pursuing the study of jump processes, one may refer to Brémaud (1981). Elements on jumps may also be found in recent books dealing with the default of credits, see Schönbucher (2003). For a textbook with emphasis on jump processes, see Cont and Tankov (2004).

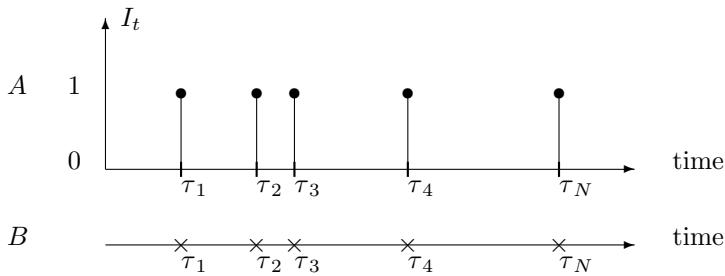
16.1 Counting and marked point process

An *indicator process*, I_t , is a process that takes the value of 1 when a particular event occurs and 0 otherwise. The moments in time when the event occurs are random times τ_i . Since at time t , we know whether or not the event has taken place (i.e., $I_t = 1$ or 0), these random times τ_i are *stopping times*.¹ The following picture displays in the upper part, Part A, a possible realization of an indicator process. In the lower part, Part B, an equivalent representation is given.

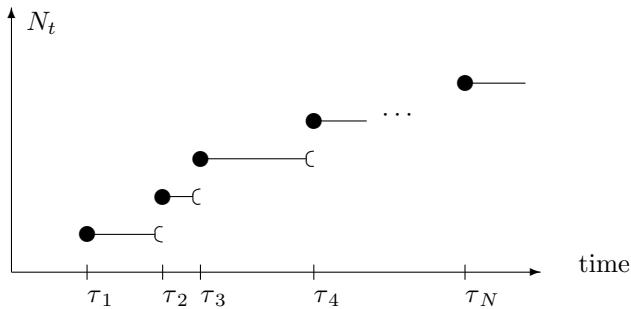
A *point process*, sometimes called a *counting process*, N_t indicates how many times the event has occurred at and before time t . N_t is an integer valued stochastic process with

$$N_t = \sum_i 1_{\{\tau_i \leq t\}}.$$

¹ A counter example of a random time that is not a stopping time is $\tilde{T} = \max_{t \in [0, T]} W_t$ for Brownian motion W_t . At time t , we do not know if $\tilde{T} = t$. Only at time T , looking backwards, can one learn \tilde{T} . Hence, \tilde{T} is not a stopping time.



Moreover, N_t is càd-làg (continue à droite, limite à gauche), which literally means right continuous with left limit. It is perhaps easier to appreciate what càd-làg means by looking at the graphical presentation of N_t below:



In many situations, the interest is not only on the event occurring but also on the magnitude of the realization. For example, when the stock return exceeds a threshold, we would want to know the exact amount the threshold is exceeded by. This type of phenomenon can be modeled with what is called a *marked point process*. A marked point process is an indicator process I_t coupled with a sequence of realization Y_{τ_i} , where τ_i denote the times when the event occurs. The realizations Y_{τ_i} are called marks, which explains why the point process is sometimes called a *marked point process*. Figure 16.1 presents a possible realization of Y_{τ_i} assuming that there are three jumps per unit time on average distributed as a Poisson process and a jump intensity $Y_{\tau_i} \sim \mathcal{N}(-1, 2)$.

If we use X_t to denote a marked point process, then

$$X_t = \sum_{i=0}^{N_t} Y_{\tau_i},$$

and Figure 16.2 is a picture of X_t , where the jump events Y_{τ_i} are the same as in Figure 16.1. We notice some of the positive jumps such as the one at time 7.8. At about time 1, there is a particularly negative jump.

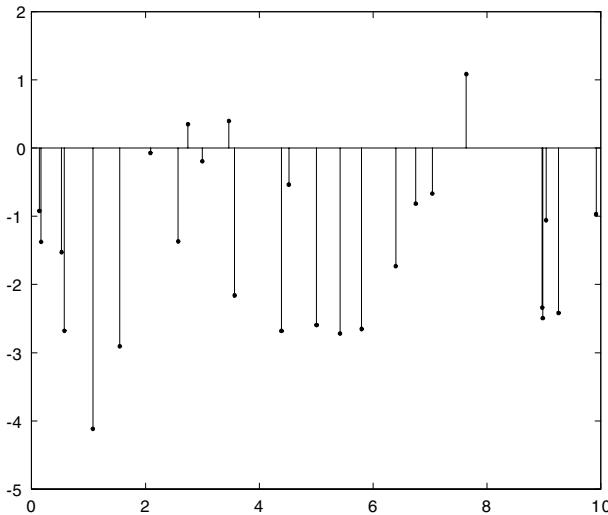


Fig. 16.1. Simulation of the marks associated with a point process.

16.2 The Poisson process

In this section, we provide a more formal description of the Poisson process.

16.2.1 Construction of the Poisson distribution

The construction of the Poisson distribution presented here follows that of Mood, Graybill, and Boes (1974). Let λ denote the *intensity* measure of a jump event, i.e., the probability of occurrence of the jump event over a unit time interval. The Poisson distribution is obtained under the following assumptions:

- (i) The probability of having one jump during a time interval is proportional to the length of that time interval, i.e., $\Pr[\text{one event occurring between } t \text{ and } t + \Delta t] = \lambda\Delta t + o(\Delta t)$, where $o(\Delta t)$ denotes elements that are smaller than Δt .
- (ii) The probability of having two events in an infinitely small time interval is negligible. Formally, this can be written as $\Pr[\text{two events occurring between } t \text{ and } t + \Delta t] = o(\Delta t)$.
- (iii) The numbers of occurrence in non-overlapping time intervals are independent.

With these assumptions, we can write, neglecting the $o(\Delta t)$,

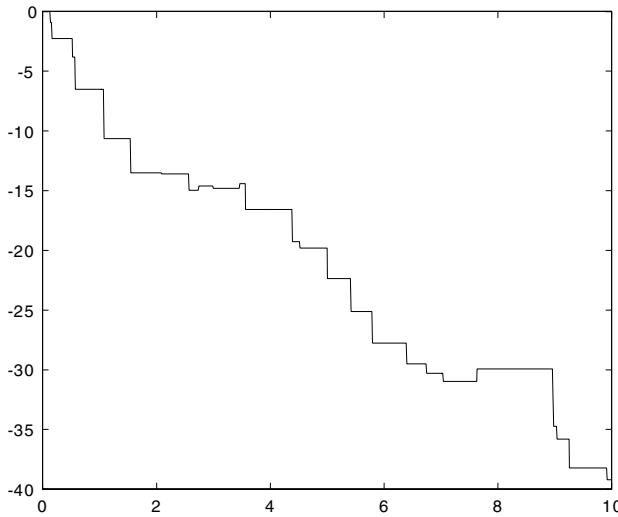


Fig. 16.2. Simulation of a marked point process.

$$\begin{aligned}\Pr[N_{t+\Delta t} - N_t = 1] &= \lambda \Delta t, \\ \Pr[N_{t+\Delta t} - N_t = 0] &= 1 - \lambda \Delta t, \\ \Pr[N_{t+2\Delta t} - N_t = 0] &= (1 - \lambda \Delta t)^2 = \Pr[N_{2\Delta t} - N_0 = 0].\end{aligned}$$

The first and the second lines define, respectively, the probability that a jump and no jump occurs in the time interval $[t, t + \Delta t]$. The last equality follows from the fact that the jump intensity λ does not depend on time.

Consider now a longer time interval $[0, T]$ that is divided into n subintervals with $\Delta t = T/n$. From the results above, it follows that the probability of no jump is

$$\Pr[N_{n\Delta t} - N_0 = 0] = (1 - \lambda \Delta t)^n = \left(1 - \lambda \frac{T}{n}\right)^n,$$

and

$$\lim_{n \rightarrow \infty} \Pr[N_{n\Delta t} - N_0 = 0] = e^{-\lambda T}. \quad (16.1)$$

What about the probability of having only one jump? There are n choices of only 1 jump among n intervals

$$\begin{aligned}\Pr[N_{n\Delta t} - N_0 = 1] &= n \lambda \Delta t (1 - \lambda \Delta t)^{n-1} \\ &= \frac{\lambda T}{\left(1 - \lambda \frac{T}{n}\right)} \left(1 - \lambda \frac{T}{n}\right)^n \rightarrow \lambda T e^{-\lambda T}.\end{aligned} \quad (16.2)$$

The last convergence follows for $n \rightarrow \infty$. What about the probability of having k jumps in $[0, T]$? We have

$$\begin{aligned}\Pr[N_{n\Delta t} - N_0 = k] &= \binom{n}{k} (\lambda \Delta t)^k (1 - \lambda \Delta t)^{n-k} \\ &= \frac{n!}{k!(n-k)!} \frac{(\lambda \frac{T}{n})^k}{(1 - \lambda \frac{T}{n})^k} (1 - \lambda \Delta t)^n \\ &= \frac{1}{k!} \frac{n \times \cdots \times (n-k+1)}{n \times n \times \cdots \times n} \frac{(\lambda T)^k}{(1 - \lambda \frac{T}{n})^k} \left(1 - \lambda \frac{T}{n}\right)^n,\end{aligned}$$

and in the limit

$$\lim_{n \rightarrow \infty} \Pr[N_{n\Delta t} - N_0 = k] = \frac{(\lambda T)^k}{k!} e^{-\lambda T}. \quad (16.3)$$

The result also holds for any selected time interval $[t, T]$

$$\Pr[N_T - N_t = k] = \frac{[\lambda(T-t)]^k}{k!} e^{-\lambda(T-t)}.$$

In the case of a Poisson process, an indicator process (resp. marked point process) is called pure Poisson process (resp. compound Poisson process).

16.2.2 Properties of the Poisson distribution

We assume now that the time interval is standardized such that $T - t = 1$. As is customary, we set $N_0 = 0$. Then, omitting the time index, from (16.3), we get

$$\Pr[N = k] = e^{-\lambda} \frac{\lambda^k}{k!}. \quad (16.4)$$

From (16.4), we can derive the expected number of jumps and the characteristic function of the jump as a pure Poisson process. But, to do so we will need the following lemma:

Lemma 16.1.

$$ae^a = \sum_{k=0}^{\infty} k \frac{a^k}{k!}.$$

Proof. Define $f(x) = e^{ax}$. From Taylor's series expansion, we get

$$\begin{aligned}f(x) &= e^{ax} = 1 + ax + \frac{a^2 x^2}{2!} + \frac{a^3 x^3}{3!} + \cdots = \sum_{k=0}^{\infty} \frac{a^k x^k}{k!}, \\ f'(x) &= ae^{ax} = a + 2 \frac{a^2 x}{2!} + 3 \frac{a^3 x^2}{3!} + \cdots = \sum_{k=0}^{\infty} k \frac{a^k x^{k-1}}{k!}.\end{aligned}$$

If we substitute $x = 1$ into $f(x)$ and $f'(x)$, we get

$$e^a = \sum_{k=0}^{\infty} \frac{a^k}{k!}, \quad \text{and} \quad ae^a = \sum_{k=0}^{\infty} k \frac{a^k}{k!}.$$

Now, we can compute the expected number of jumps as follows

$$E[N] = \sum_{k=0}^{\infty} k \Pr[N = k] = \sum_{k=0}^{\infty} k e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!},$$

and, using Lemma 16.1, we get

$$E[N] = e^{-\lambda} \lambda e^{\lambda} = \lambda.$$

Similarly, we can derive the characteristic function of the pure jump process

$$\phi_J(u) = E[e^{iuN}] = \sum_{k=0}^{\infty} e^{iuk} e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{e^{iuk+k \log(\lambda)}}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{m^k}{k!},$$

where $m = e^{iu+\log(\lambda)}$. Lemma 16.1 yields

$$\phi_J(u) = e^{-\lambda} e^m = \exp[\lambda(e^{iu} - 1)]. \quad (16.5)$$

This is an important characteristic function also playing a role in the construction of the Lévy process. It is also easy to see that the characteristic function of N_t is $\exp[\lambda t(e^{iu} - 1)]$.

16.2.3 Moments of pure Poisson process

As shown in Chapter 15, the k th moment, if it exists, can be computed as the k th derivative of the characteristic function $\phi_J(u)$ and evaluated at $u = 0$. From (16.5), we have

$$\begin{aligned} \phi'_J(u) &= \lambda i e^{iu} \phi_J(u), \\ \phi''_J(u) &= (-\lambda e^{iu} - \lambda^2 e^{2iu}) \phi_J(u), \\ \phi'''_J(u) &= (-\lambda i e^{iu} - 2i\lambda^2 e^{2iu}) \phi_J(u) + (-\lambda e^{iu} - \lambda^2 e^{2iu}) \lambda i e^{iu} \phi_J(u). \end{aligned}$$

It follows that

$$\begin{aligned} E[N] &= \lambda, \\ E[N^2] &= \lambda + \lambda^2, \quad \text{and thus} \quad V[N] = \lambda, \\ E[N^3] &= \lambda + 3\lambda^2 + \lambda^3. \end{aligned}$$

Hence, the third moment about the mean is

$$E[(N - E[N])^3] = E[N^3] - 3E[N^2]E[N] + 2(E[N])^3 = \lambda.$$

We can also show that the fourth moment about the mean is

$$E[(N - E[N])^4] = 3\lambda^2 + \lambda$$

so that standardized skewness and excess kurtosis are, respectively,

$$\begin{aligned}s &= Sk[N] = \frac{1}{\sqrt{\lambda}}, \\ \kappa &= Ku[N] = \frac{1}{\lambda}.\end{aligned}$$

The expression of these moments is useful in the context of method of moment estimation involving the Poisson process. Such a situation arises in the context of tests of the Mixture of Distribution Hypothesis.

16.2.4 Compound Poisson process

The compound Poisson process is defined as $X_t = \sum_{i=1}^{N_t} Y_{\tau_i}$, where N_t is a pure Poisson process and where Y_{τ_i} represent the actual jumps at time τ_i , the moments when a jump takes place. The characteristic function of the compound Poisson process plays a crucial role in the construction of the Lévy process. Let

$$\begin{aligned}\phi_{X_t}(u) &= E[e^{iuX_t}] = E\left[\exp iu \sum_{k=1}^{N_t} Y_{\tau_k}\right] \\ &= E\left[E\left[\exp iu \sum_{k=1}^{N_t} Y_{\tau_k} \mid N_t\right]\right] = E[\phi_Y(u)^{N_t}] \\ &= \sum_{k=0}^{\infty} \frac{(\phi_Y(u)\lambda t)^k}{k!} e^{-\lambda t} = \sum_{k=0}^{\infty} \frac{(\phi_Y(u)\lambda t)^k}{k!} e^{-\phi_Y(u)\lambda t} e^{-\lambda t + \phi_Y(u)\lambda t} \\ &= \exp(\lambda t(\phi_Y(u) - 1)).\end{aligned}$$

The second line conditions on the number of jumps. The third line follows from the independence of the various Y and from the definition of an expectation. We recognize in this formula a term similar to $x^k/k!$. This term will converge to e^x . We multiply by such a term and also divide.

These computations show that, in most cases, even though the characteristic function is easy to determine, the density of a compound Poisson process may not exist in explicit form. However, given the characteristic function, using the inversion theorem presented in Chapter 15, we could, at least numerically obtain the associated density. In certain cases, such as when the jump sizes are Gaussian, then conditional on N_t , the distribution of X_t is $\mathcal{N}(\mu N_t, \sigma^2 N_t)$. For this particular case, the density may be obtained as a series. Indeed, we have

$$\begin{aligned}
\Pr[X_t \leq x] &= \Pr \left[\sum_{k=1}^{N_t} Y_{\tau_k} \leq x \mid N_t \in \{1, 2, \dots\} \right] \\
&= \sum_{k=1}^{\infty} \Pr \left[\sum_{l=1}^{N_t} Y_{\tau_l} \leq x \mid N_t = k \right] \Pr[N_t = k] \\
&= \sum_{k=1}^{\infty} \Pr \left[\sum_{l=1}^k Y_{\tau_l} \leq x \right] \frac{(\lambda t)^k}{k!} e^{-\lambda t}.
\end{aligned}$$

The first equality introduces the set all of possible values that N_t can take. In the second line, we use the assumption of independency of the jumps and the properties of conditional probabilities.

In terms of density, we obtain, if the markers are $\mathcal{N}(\mu, \sigma^2)$, that

$$f_{X_t}(x) = \sum_{k=1}^{\infty} \frac{1}{\sqrt{2\pi k \sigma^2}} \exp \left[-\frac{1}{2} \frac{(x - k\mu)^2}{k\sigma^2} \right] \frac{(\lambda t)^k}{k!} e^{-\lambda t}.$$

This expression involves an infinite sum. However, it is often possible to truncate the infinite sum to just a few terms because in most cases, the probability $\Pr[N_t = k]$ converges to 0 as k increases. Also, all the terms of the sum are positive, this implies numerical stability.

16.3 The exponential distribution

16.3.1 Definition and properties

An exponential random variable X has

$$\begin{aligned}
f_X(x) &= \lambda e^{-\lambda x} 1_{\{x>0\}}(x), \\
F_X(x) &= (1 - e^{-\lambda x}) 1_{\{x>0\}}(x), \text{ and} \\
\phi_X(u) &= E[e^{iuX}] = \int_0^{+\infty} e^{iux} \lambda e^{-\lambda x} dx = \frac{\lambda}{\lambda - iu},
\end{aligned} \tag{16.6}$$

where $f_X(x)$, $F_X(x)$, and $\phi_X(u)$ are, respectively, the density, distribution, and characteristics functions. Figure 16.3 traces the shape of $f_X(x)$ and $F_X(x)$ for an exponential variable with parameter $\lambda = 1/3$. Note that, due to the exponential operator, the exponential variable is always positive.

The exponential distribution was first used in the study of waiting time. Formally, if X denotes the time between events, then for the exponential variable X , we have

$$\Pr[X \leq x_0 + x \mid X > x_0] = \Pr[X \leq x].$$

That is, knowing that $X > x_0$ does not help in predicting if $X \leq x_0 + x$. Put this into the context of waiting time, it means that the probability of a bus

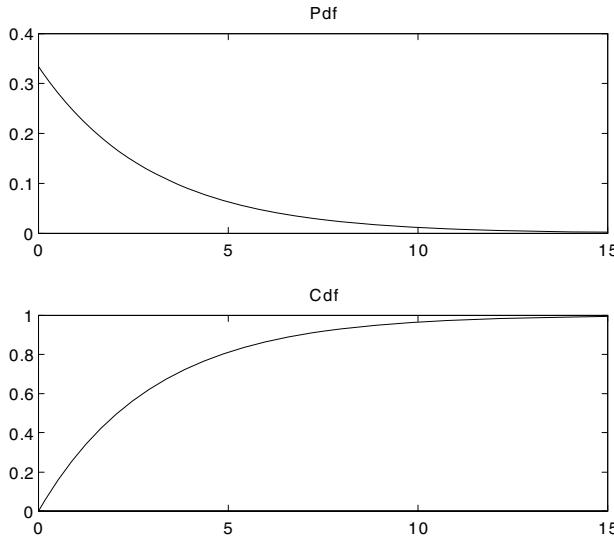


Fig. 16.3. Pdf and cdf of the exponential distribution with $\lambda = 1/3$.

arriving between 3:00 and 3:30 is the same whether we have already waited 30 minutes or 3 hours! Formally, we have

$$\begin{aligned} \Pr[X \leq x_0 + x | X > x_0] &= \frac{\Pr[X > x_0, X \leq x_0 + x]}{\Pr[X > x_0]} \\ &= \frac{\int_{x_0}^{x_0+x} \lambda e^{-\lambda u} du}{e^{-\lambda x_0}} \\ &= \frac{[1 - e^{-\lambda(x+x_0)} - (1 - e^{-\lambda x_0})]}{e^{-\lambda x_0}} \\ &= 1 - e^{-\lambda x} = F_X(x). \end{aligned}$$

This shows that the exponential variable has the property of “no memory”.

16.3.2 Moments of the exponential variable

Again from Chapter 15, moments, if they exist, can be derived from the characteristic function. From the characteristic function of the exponential variable in (16.6), we get

$$\begin{aligned} \phi'_X(u) &= \frac{i\lambda}{(\lambda - iu)^2}, \\ \phi''_X(u) &= \frac{2i^2\lambda}{(\lambda - iu)^3}, \end{aligned}$$

and, in general,

$$\phi_X^{(n)}(u) = \frac{n!i^n \lambda}{(\lambda - iu)^{(n+1)}}.$$

For $u = 0$, we have

$$\phi_X^{(n)}(0) = n!i^n \lambda^{-n},$$

and therefore

$$E[X^n] = n! \lambda^{-n}.$$

Thus

$$\begin{aligned} E[X] &= \lambda^{-1}, \\ E[X^2] &= 2\lambda^{-2}, \quad \text{and} \quad V[X] = \lambda^{-1}, \\ E[X^3] &= 6\lambda^{-3}, \\ E[X^4] &= 24\lambda^{-4}. \end{aligned}$$

16.3.3 Hazard and survivor functions

The probability, $\theta(t)$, of having an event in the next unit of time after having “survived” up to time t is called the *hazard function*. Formally, we have

$$\theta(t) = \lim_{dt \rightarrow 0} \frac{1}{dt} \Pr[t \leq T < t + dt \mid T \geq t].$$

Since

$$\begin{aligned} \frac{1}{dt} \frac{\Pr[t \leq T < t + dt, T \geq t]}{\Pr[T \geq t]} &= \frac{1}{dt} \frac{\Pr[t \leq T < t + dt]}{\Pr[T \geq t]} \\ &= \frac{1}{dt} \frac{F(t + dt) - F(t)}{1 - F(t)}, \end{aligned}$$

we obtain that, as $dt \rightarrow 0$,

$$\theta(t) = \frac{f(t)}{1 - F(t)}.$$

The expression, $1 - F(t)$, in the denominator, is called the *survivor function* $\bar{F}(t)$. For the exponential distribution, the hazard function is constant with $\theta(t) = \lambda$. Next, we note that

$$d\bar{F}(t) = d(1 - F(t)) = -f(t)dt.$$

Hence, since

$$\theta(t) = \frac{-d\bar{F}(t)}{dt} \frac{1}{\bar{F}(t)} = -\frac{d \log(\bar{F}(t))}{dt},$$

we deduce

$$\log(\bar{F}(t)) - \log(\bar{F}(0)) = - \int_0^t \theta(s) ds,$$

and

$$\bar{F}(t) = \exp\left(- \int_0^t \theta(s) ds\right)$$

because $\bar{F}(0) = 1$ and $\log(\bar{F}(0)) = 0$. The associated *pdf* is

$$f(t) = - \frac{d\bar{F}(t)}{dt} = \theta(t) \exp\left(- \int_0^t \theta(s) ds\right).$$

Finally, we obtain $f(t) = \lambda e^{-\lambda t}$, because $\theta(t) = \lambda$ is a constant and $\theta(0) = 0$. We conclude from this that if we seek a distribution that has a constant hazard function, we automatically obtain the exponential one.

16.4 Duration between Poisson jumps

Here, we show that the time between Poisson jumps follows an exponential distribution. This is done by proving that the *pdf* is the one of the exponential distribution. We have

$$\begin{aligned} f(\tau)\Delta &= \Pr[\text{next jump occurs between } \tau \text{ and } \tau + \Delta] \\ &= \Pr[\text{no jump till } \tau, \text{ jump in } (\tau, \tau + \Delta)] \\ &= \Pr[N(\tau) - N(0) = 0, N(\tau + \Delta) - N(\tau) = 1] \\ &= \Pr[N(\tau) - N(0) = 0] \times \Pr[N(\tau + \Delta) - N(\tau) = 1]. \end{aligned}$$

From (16.1) and (16.2), we get

$$\begin{aligned} f(\tau)\Delta &= e^{-\lambda\tau} \times \lambda\Delta e^{-\lambda\Delta} \\ f(\tau) &= \lambda e^{-\lambda\tau} e^{-\lambda\Delta}. \end{aligned}$$

Finally, we have

$$\lim_{\Delta \rightarrow 0} f(\tau) = \lambda e^{-\lambda\tau},$$

since $e^{-\lambda\Delta} \rightarrow 1$ as $\Delta \rightarrow 0$. The usefulness of this proposition becomes apparent when simulating a Poisson process.

From the previous results, there are two possible algorithms for simulating a Poisson process with parameter λ over the time interval $[0, T]$:

- (i) Given that the waiting time between events follows an exponential distribution, we could draw exponential durations till their sum exceeds T . Exponential draws could be obtained using the inverse *cdf*.
- (ii) Given that $\Pr[N_T - N_0 = k]$ follows a Poisson distribution with parameter λT , we could simulate N_T using the Poisson distribution with parameter λT . Once, the number of jumps is obtained, the N_T jump events can be distributed over $[0, T]$ by using uniform draws scaled to the interval $[0, T]$.

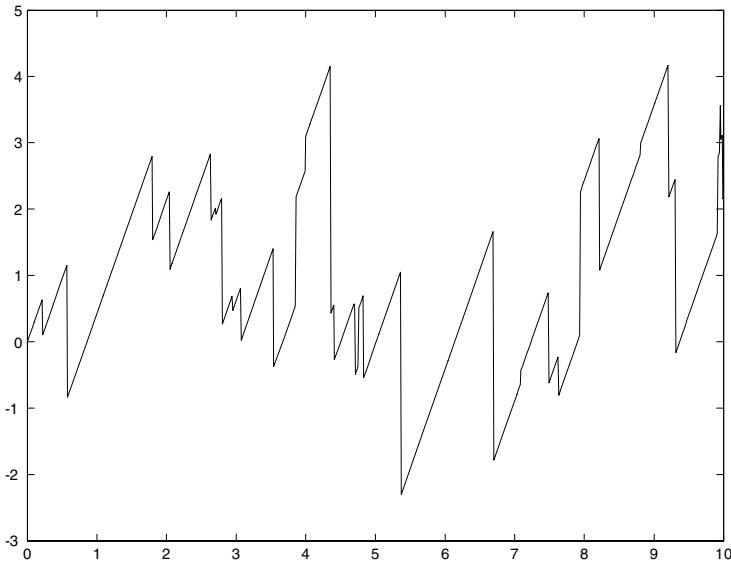


Fig. 16.4. Simulation of a compensated compound Poisson process.

16.5 Compensated Poisson processes

The Poisson counting process N_t has an upward drift. Hence, its mean is greater than zero. In marked Poisson process, the mean of Y_{τ_i} need not be greater than zero. In some applications, e.g., option pricing through martingale measure, it is necessary to ensure that the adjusted process is a martingale with zero drift. This is often done through *compensation*,

$$M_t \equiv N_t - A_t,$$

where A_t is called a compensator.

In the case of the (pure) Poisson counting process, we have $E[N_t] = \lambda t$. Hence, the corresponding compensated Poisson process is

$$M_t^{PP} = N_t - \lambda t.$$

For a compound Poisson process, where each marker Y_{τ_i} is independently drawn from the same distribution with mean $E[Y_{\tau_i}]$, we have

$$M_t^{CP} = N_t - \lambda t E[Y_{\tau_i}].$$

Figure 16.4 illustrates the simulation of a compensated Poisson process.

So far, we assumed the parameter λ to be a constant. In this case, the Poisson process is called *homogenous*. If λ_t is time varying, the Poisson process is called *inhomogeneous*.

It is useful to derive the characteristic function of a compensated compound Poisson process. This is easy. Let us illustrate the working with the example of the compensated pure Poisson process

$$\begin{aligned}\phi_{M_t^{PP}}(u) &= \exp(\lambda t(e^{iu} - 1)) \exp(-\lambda t i u) \\ &= \exp(\lambda t(e^{iu} - 1 - i u)).\end{aligned}$$

Also, we obtain immediately for the compensated compound Poisson process

$$\phi_{M_t^{CP}}(u) = \exp(\lambda t (\phi_Y(u) - 1 - i u E[Y_{\tau_k}])), \quad (16.7)$$

Sometimes it is useful to express this characteristic function in terms of the *cdf* F . Obviously, we have

$$\begin{aligned}\phi_Y(u) &= \int_{z=-\infty}^{\infty} e^{iuz} dF(z), \\ 1 &= \int_{z=-\infty}^{\infty} dF(z), \\ E[Y_{\tau_k}] &= \int_{z=-\infty}^{\infty} z dF(z).\end{aligned}$$

Plugging these terms into (16.7), we obtain

$$\phi_{M_t^{CP}}(u) = \exp\left(\lambda t \int_{z=-\infty}^{\infty} (e^{iuz} - 1 - iuz) dF(z)\right).$$

This expression of the compound Poisson process will play an important role in the next chapter.

Lévy Processes

Paul Lévy was a French mathematician, maybe the best of his generation. His idea was to combine the Brownian motion and the compound Poisson process. This led him to study processes described by the so-called Lévy-Khintchine formula that encompasses a large class of continuous and discontinuous stochastic processes. This chapter provides the fundamentals of Lévy process that are useful for understanding Lévy option pricing models. Besides the book by Lévy (1954), which contains lots of insights, unfortunately only available in French, other useful readings on Lévy processes are Feller (1971) (see chapter *VI* and *XVIII*), Bertoin (1996), and Sato (1999). The notations in this chapter follows the ones in Chapter 16. Reading of that chapter before this one is recommended.

17.1 Construction of the Lévy process

The original construction of the Lévy process was rather heuristic. Lévy had the idea to combine a Brownian motion and a compound Poisson process. This resulted in an expression, given as a characteristic function. Lévy then generalized this characteristic function essentially by discussing the existence of an integral that figures in the expression of the characteristic function. Let a process X_t consist of a Brownian motion, X_t^{BM} , and a compensated compound Poisson process, X_t^{CP} , as follows

$$\begin{aligned} X_t &= X_t^{BM} + X_t^{CP}, \\ X_t^{BM} &= \mu t + \sigma W_t, \\ X_t^{CP} &= \sum_{k=0}^{N_t} Y_{\tau_k} - \lambda t E[Y_{\tau_k}]. \end{aligned}$$

In this definition, all three sources of randomness, the Brownian motion, W_t , the pure jump process, N_t , and the jump sizes Y_{τ_k} are independent. The parameters μ , σ , and λ , are real numbers, with $\sigma \geq 0$ and $\lambda \geq 0$. The characteristic function of X_t^{BM} is

$$\phi_{BM}(u) = \exp \left[t \left(i\mu u - \frac{1}{2} \sigma^2 u^2 \right) \right].$$

The characteristic function of X_t^{CP} is

$$\phi_{CP}(u) = \exp \left[\lambda t \int_{z=-\infty}^{\infty} (e^{iuz} - 1 - iuz) dF(z) \right].$$

Provided that the Brownian motion and the compensated Poisson process are independent, combining their characteristic functions yields the so-called *de Finetti characteristic function* (or expression)

$$\phi_F(u) = \exp \left[t \left\{ i\mu u - \frac{1}{2} \sigma^2 u^2 + \lambda \int_{z=-\infty}^{\infty} (e^{iuz} - 1 - iuz) dF(z) \right\} \right].$$

We will call the term enclosed by the curly brackets the exponent.

So far we combined a Brownian motion with a compound jump process. We can ask what happens if we superpose additional components. Concerning the continuous Brownian motion part, the answer is easy. Since a Brownian motion is Gaussian, the sum of Brownian motions will result in another Brownian motion. Such a result does not hold for the jump part. To understand the issue at hand, let us focus only on the jump part. We get, by mixing two jumps, each with intensity λ_i and markers distributed as F_i , an exponent of the form

$$\begin{aligned} I &= \lambda_1 \int_{z=-\infty}^{\infty} (e^{iuz} - 1 - iuz) dF_1(z) \\ &\quad + \lambda_2 \int_{z=-\infty}^{\infty} (e^{iuz} - 1 - iuz) dF_2(z) \\ &= \int_{z=-\infty}^{\infty} (e^{iuz} - 1 - iuz) \nu(dz), \end{aligned}$$

where we define in the last equation $\nu(dz) = \lambda_1 dF_1(z) + \lambda_2 dF_2(z)$. Inspection of the expression $\nu(dz)$ shows that for given λ_1 and λ_2 there is no reason that the probability mass integrates up to 1. On the other hand, λ_1 , λ_2 , dF_1 , and dF_2 are all positive, thus, we expect the measure $\nu(dz)$ to be positive. The measure $\nu(z)$ will be interpreted as measuring jumps of size z . Given its importance, this expression will be called the *Lévy measure*. Once we introduce a general measure, this opens the question of existence of the integral. At the point 0, the measure may have a discontinuity. For this reason, the integral should always be considered as being $\int_{-\infty}^0 + \int_0^{\infty}$.

The key question is under what conditions is the function I well defined. Lévy's contribution is to spell out a set of sufficient conditions that ensures the existence of the integral. Two reasonable assumptions to start with are

- (i) ν is positive;

- (ii) $\int_z \nu(dz) < \infty$, which basically means the mass associated with each z is finite;

If we closely study this formula, we see that nothing prevents jumps from being negative, hence, we will take $z \in \mathbb{R}$. Because of the possibility of jumps becoming infinitely active with ever decreasing jump size, the integral needs to be interpreted with some caution as

$$\begin{aligned} \int_{z=-\infty}^{\infty} (e^{iuz} - 1 - iuz)\nu(dz) &= \lim_{a \rightarrow 0, a > 0} \int_{z=-\infty}^{-a} (e^{iuz} - 1 - iuz)\nu(dz) \\ &\quad + \lim_{a \rightarrow 0, a > 0} \int_{z=a}^{+\infty} (e^{iuz} - 1 - iuz)\nu(dz). \end{aligned}$$

To discover further conditions, let us start focusing on jumps with very small size, yet that may have a great intensity. We may imagine, in the limit, jumps of ever smaller size with greater and greater intensity. Under which condition will the integral remain finite? In order to answer this issue, we will consider a neighborhood of 0, say the interval $[-1, 1]$. Let us also consider the Taylor series expansion of the integrand around $z = 0$,

$$\begin{aligned} e^{iuz} - 1 - iuz &= 1 - \frac{z^2}{2} + i \left[uz - \frac{(uz)^3}{3!} \right] - 1 - iuz + o(z^3) \\ &= -\frac{z^2}{2} + o(z^2). \end{aligned} \tag{17.1}$$

The term $o(z)$ means terms smaller than z . Inspection of (17.1) reveals that I is integrable as long as $\int z^2 \nu(dz)$ exists. When this condition is satisfied, and $\nu(z) \rightarrow \infty$ as $z \rightarrow 0$, the process is said to be one of *infinite activity*. Intuitively, it means that it is possible to have many jumps so long as these jumps become smaller and smaller. Often this assumption of existence and (ii) are regrouped by imposing the following assumption:

- (iii) $\int_z \min(1, z^2)\nu(dz) < \infty$.

In the construction, we have used compensated Poisson processes. We may ask if it is necessary to impose compensation. To investigate this issue, we will now inspect the speed by which $\nu(z)$ may grow for a compensated and uncompensated process. Let us assume that $\nu(z) = z^{-\alpha}$ for z small, so that $\nu(dz) = -\alpha z^{-\alpha-1} dz$. Then the integral

$$\int_{-1}^1 z^2 \nu(dz) = \lim_{a \rightarrow 0, a > 0} \int_{-1}^{-a} -\alpha z^{1-\alpha} dz + \lim_{a \rightarrow 0, a > 0} \int_a^1 -\alpha z^{1-\alpha} dz.$$

The right-most integral equals

$$\int_a^1 -\alpha z^{1-\alpha} dz = \frac{-\alpha}{2-\alpha} [1 - a^{2-\alpha}].$$

Obviously, this integral is well defined as a converges to 0, as long as $\alpha < 2$. On the other hand, if the jumps have not been compensated, then (17.1) becomes

$$e^{iuz} - 1 - iuz = iuz + o(z).$$

In this case, similar computations as previously indicate that the integrand is finite only if $\alpha < 1$. Clearly, as z will converge to zero, $z^{-\alpha}$ will become larger, the larger the power. This implies that the compensated Poisson process allows for a larger jump intensity than the uncompensated Poisson process. This result is also intuitive. Indeed, if we contemplate Figure 16.4 representing a compensated Poisson process, we notice that the process tends to revert towards zero. In other words, for such a process, the consequence of jumps remains controlled. Given that compensated process allows for more activity in terms of small jumps, they incorporate a larger class of processes. For this reason, we always work with compensated compound Poisson processes.

To complete our investigation of the integral, we also need to examine its behavior for large jumps. We will focus on the upper tail. The study of the lower tail would proceed in analogous manner. The integral will, by definition, be finite if

$$\lim_{b \rightarrow +\infty} \int_a^b (e^{iuz} - 1 - iuz)\nu(dz) < \infty \quad \text{for some } a > 0.$$

To investigate the finiteness of this integral, let us examine the various terms involved in the integral. First, as long as $\int_a^b \nu(dz)$ exists, which holds if (iii) is true, meaning that $\nu(z)$ will converge toward 0 for large z , since $e^{iuz} - 1$ remains bounded, the first part of the integral will exist. However, if $\nu(dz)$ has no finite moments, as would be the case of the Cauchy and the Pareto density, then the integral $\int_a^b z\nu(dz)$ ceases to exist, for large b . Given the importance of distributions such as the Cauchy and the Pareto distributions, it makes sense to modify the definition of the exponent in such a way as to guarantee that the integral will also remain finite for large jumps. Lévy's excellent idea to solve this problem is to modify the compensator by restricting the "average jump size". He introduced a function $c(z) = 1/(1+z^2)$ to dampen out the tails. By setting

$$\int_a^{+\infty} zc(z)\nu(dz) = \int_a^{+\infty} \frac{z}{1+z^2}\nu(dz) < \infty,$$

as long as (iii) is satisfied, the integral is now well defined.

Clearly, many other choices of dampening function are possible. Sometimes we encounter the function $c(z) = I_{\{|z|<1\}}(z)$ taking the value 1 for $|z| < 1$ and 0 otherwise. This would yield an alternative expression for the class of processes under investigation. As soon as we touch that part of the integral corresponding to the compensator, the mean μ of the Lévy process changes. Hence, if we introduce or change the dampening function, to obtain a similar Lévy process, it is necessary to modify the mean.

Finally, combining all the terms, we obtain the Lévy-Khintchine formula of the characteristic function for Lévy process as

$$\phi(u) = \exp \left[t \left\{ i\mu u - \frac{1}{2} \sigma^2 u^2 + \int_{z=-\infty}^{\infty} [e^{izu} - 1 - iuzc(z)] \nu(dz) \right\} \right]. \quad (17.2)$$

We remind that the conditions under which this formula is valid are

$$\begin{aligned} \nu &\text{ is positive,} \\ \int_z \min(1, z^2) \nu(dz) &< \infty, \\ c(z) &= 1/1 + z^2. \end{aligned}$$

The term between the curly brackets in (17.2) is the Lévy exponent. Inspection of this formula also indicates that a Lévy process is fully described by the three parameters (μ, σ, ν) , the so-called Lévy or characteristic triplet.

17.2 Properties of Lévy processes

Consider a time interval $[0, T]$ and $\{X_t\}_{t=0}^T$ a Lévy process starting at $X_0 = 0$. The characteristic function of the Lévy process of X_T will be of the form

$$\phi_{0:T}(u) = \exp(T\psi(u)),$$

where $\psi(u)$ is the Lévy exponent and the index $0 : T$ emphasizes the increment over which we consider the process. Consider now a partition $0 = t_0 < t_1 < \dots < t_n = T$ of $[0, T]$. It is always true that

$$T = (t_n - t_{n-1}) + (t_{n-1} - t_{n-2}) + \dots + (t_1 - t_0).$$

Hence,

$$\phi_{0:T}(u) = \prod_{j=1}^n \phi_{t_{j-1}:t_j}(u), \quad (17.3)$$

where each of the $\phi_{t_{j-1}:t_j}(u) = \exp((t_j - t_{j-1})\psi(u))$ is distributed as a Lévy process. For the particular case, where $t_j - t_{j-1} = T/n$ for all j , all the characteristic functions under the product sign will be equal and

$$\phi_{0:T}(u) = \left[\phi_{0:T/n}(u) \right]^n.$$

In other words, the distribution of X_T is the distribution of the sum of n independent random variables, each distributed with the same distribution. If this property of summability holds, then the process is said to be *infinitely divisible*. Obviously, since the construction evolved around the Lévy process, the Lévy process is infinitely divisible.

This property is related to having independent and stationary increments. Indeed, inspection of (17.3) shows exactly this. Each of the characteristic functions corresponds to the increment of a Lévy process. Given that the characteristic function factorizes, the random increments will also be independent. The stationarity follows from the fact that none of the characteristic functions depends on the time when it is evaluated, only on time increments.

We may now consider stable distributions. This type of distribution was introduced to finance by Mandelbrot (1963) in an attempt to model the fat-tailedness of asset returns. By definition, a distribution is said to be stable if the sum of two independent random variables, say X_1 and X_2 , is equal to a new random variable X with the same distribution. This in turn implies that $X_1 + X_2 = aX$, for a a constant. Also

$$X = \frac{1}{a}(X_1 + X_2).$$

Proceeding iteratively, we understand that a random variable distributed as a stable distribution, can be expressed as a sum of identically distributed random variables. We also have

$$\phi_X(au) = [\phi_X(u)]^2 \Rightarrow \phi_X(u) = \left[\phi_{\frac{1}{a}X}(u) \right]^2.$$

This shows that stable distributions are infinitely divisible. The class of Lévy processes therefore contains the class of stable distributions. Clearly, considering the example of the pure Poisson process, which is a Lévy process, we understand that there are Lévy processes that are not stable distributions. Also, whereas for stable distributions with the exception of the Gaussian, the second moment does not exist. For a Lévy process, the second moment and higher may exist.

What the construction of the Lévy process tells us is that, besides the continuous Brownian motion component, there are jumps of finite measure that occur occasionally and there are jumps of very small size that are infinitely active. Interestingly, the literature dealing with realized volatility only distinguishes the continuous part from the jumps of finite size. This literature does not deal with the jumps of infinite activity, perhaps an area for future research.

References

- Abadir K., Rockinger M. (2003) Density embedding functions. *Econometric Theory* 19:778–811.
- Abken P., Madan D.B., Ramamurtie S. (1996) Estimation of risk-neutral and statistical densities by Hermite polynomial approximation: With an application to eurodollar futures options. *Federal Reserve Bank of Atlanta, Working Paper*.
- Abramowitz M., Stegun I.A. (1970) *Handbook of mathematical functions with formulas, graphs, and mathematical tables*. Dover Publications, New York.
- Acerbi C., Tasche D. (2002) Expected shortfall: A natural coherent alternative to value-at-risk. *Economic Notes by Banca Monte dei Paschi di Siena* 31:379–388.
- Agmon N., Alhassid Y., Levine R.D. (1979a) An algorithm for finding the distribution of maximal entropy. *Journal of Computational Physics* 30:250–259.
- Agmon N., Alhassid Y., Levine R.D. (1979b) The maximum entropy formalism. In: Levine R.D., Tribus M. (eds) *The maximum entropy formalism*. MIT Press, Cambridge, Massachusetts, 207–209.
- Ahn D.-H., Boudoukh J., Richardson M., Whitelaw R.F. (1999) Optimal risk management using options. *Journal of Finance* 54:359–375.
- Ait-Sahalia Y. (2004) Disentangling diffusion from jumps. *Journal of Financial Economics* 74:487–528.
- Ait-Sahalia Y., Brandt M.W. (2001) Variable selection for portfolio choice. *Journal of Finance* 56:1297–1351.
- Ait-Sahalia Y., Lo A.W. (1998) Nonparametric estimation of state-price densities implicit in financial asset prices. *Journal of Finance* 53:499–547.
- Ait-Sahalia Y., Lo A.W. (2000) Nonparametric risk management and implied risk aversion. *Journal of Econometrics* 94:9–51.
- Ait-Sahalia Y., Mykland P., Zhang, L. (2005a) How often to sample a continuous-time process in the presence of market microstructure noise. *Review of Financial Studies* 18:351–416.
- Ait-Sahalia Y., Mykland P., Zhang, L. (2005b) Ultra-high frequency volatility estimation with dependent microstructure noise. *NBER Working Paper* 11380.

- Alexander C.O. (2001) Orthogonal GARCH. In: Alexander C.O. (ed.) Mastering risk, Volume 2: application. Financial Times Prentice Hall, London.
- Alexander C.O., Chibumba A.M. (1997) Multivariate orthogonal factor GARCH. University of Sussex, Working Paper.
- Alexander G.J., Baptista A.M. (2004) A comparison of VaR and CVaR constraints on portfolio selection with the mean-variance model. *Management Science* 50:1261–1273.
- Almgren R., Chriss N. (2000) Optimal execution of portfolio transactions. *Journal of Risk* 3:5–39.
- Andersen T.G., Benzoni L., Lund J. (2002) An empirical investigation of continuous-time equity return models. *Journal of Finance* 57:1239–1284.
- Andersen T.G., Bollerslev T., Diebold F.X. (2003) Some like it smooth, and some like it rough: Untangling continuous and jump components in measuring, modeling, and forecasting asset return volatility. PIER Working Paper Archive 03-025, Penn Institute for Economic Research, Department of Economics, University of Pennsylvania.
- Andersen T.G., Bollerslev T., Diebold F.X., Labys P. (2003) Modeling and forecasting realized volatility. *Econometrica* 71:579–625.
- Andersen T.G., Bollerslev T., Diebold F.X., Ebens H. (2001) The distribution of realized stock return volatility. *Journal of Financial Economics* 61:43–76.
- Andersen T.G., Bollerslev T., Meddahi N. (2004) Correcting the errors: Volatility forecast evaluation using high-frequency data and realized volatilities. *Econometrica* 73:279–296.
- Ang A., Bekaert G. (2002) International asset allocation with time-varying correlations. *Review of Financial Studies* 15:1137–1187.
- Ang A., Chen J. (2002) Asymmetric correlations of equity portfolios. *Journal of Financial Economics* 63:443–494.
- Arditti F.D. (1967) Risk and the required return on equity. *Journal of Finance* 22:19–36.
- Arditti F.D. (1971) Another look at mutual fund performance. *Journal of Financial and Quantitative Analysis* 10:909–912.
- Arnold L. (1974) Stochastic differential equations (theory and applications). Wiley, New York.
- Arnold B.C., R.J. Beaver R.J. (2000) Skewed multivariate models related to hidden truncation and/or selective reporting. *Test* 11:7–54.
- Artzner P., Delbaen F., Eber J.-M., Heath D. (1997) Thinking coherently. *Risk* 10:33–49.
- Artzner P., Delbaen F., Eber J.-M., Heath D. (1999) Coherent measures of risk. *Mathematical Finance* 9:203–228.
- Athayde G.M., Flôres Jr R.G. (2004) Finding a maximum skewness portfolio – A general solution to three-moments portfolio choice. *Journal of Economic Dynamics and Control* 28:1335–1352.
- Azzalini A. (1985) A class of distribution which includes the normal ones. *Scandinavian Journal of Statistics* 12:171–178.

- Azzalini A., Capitanio A. (1999) Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistical Association B* 61:579–602.
- Azzalini A., Capitanio A. (2003) Distributions generated by perturbation of symmetry with emphasis on a multivariate skew-t distribution. *Journal of the Royal Statistical Society B* 65:367–389.
- Azzalini A., Dala Valle A. (1996) The multivariate skew normal distribution. *Biometrika* 83:715–726.
- Bahra B. (1996) Probability distributions of future asset prices implied by option prices, *Bank of England Quarterly Bulletin*: August 1996, 299–311.
- Bai J., Ng S.N. (2005) Testing skewness, kurtosis, and normality for time series. *Journal of Business and Economic Statistics* 23:49–61.
- Bai X., Russell J.R., Tiao G.C. (2003) Kurtosis of GARCH and stochastic volatility models with non-normal innovations. *Journal of Econometrics* 114:349–360.
- Baillie R.T., Bollerslev T. (1990) A multivariate generalized ARCH approach to modeling risk premia in forward foreign exchange rate markets. *Journal of International Money and Finance* 9, 309–324.
- Bakshi G., Cao C., Chen Z. (1997) Empirical performance of alternative option pricing models. *Journal of Finance* 52:2003–2049.
- Bakshi G., Madan D. (2000) Spanning and derivative-security valuation. *Journal of Financial Economics* 55:205–238.
- Baldazzi P., Lynch A. (1999) Transaction costs and predictability: some utility cost calculations. *Journal of Financial Economics* 52:47–78.
- Balkema A., De Haan L. (1974) Residual life at great age. *Annals of Probability* 2:792–804.
- Ball C.A., Torous W.N. (1983) A simplified jump process for common stock returns. *Journal of Financial and Quantitative Analysis* 18(1):53–65
- Bandi F., Russell J. (2004) Separating market microstructure noise from volatility, University of Chicago, Working Paper.
- Bansal R., Hsieh D., Viswanathan S. (1993) A new approach to international arbitrage pricing. *Journal of Finance* 48:1719–1747.
- Barndorff-Nielsen O.E. (1998) Processes of normal inverse gaussian type. *Finance and Stochastics* 2:41–68.
- Barndorff-Nielsen O.E., Shephard N. (2001) Non-gaussian Ornstein-Uhlenbeck based models and some of their uses in financial economics. *Journal of the Royal Statistical Society B* 63:167–241.
- Barndorff-Nielsen O.E., Shephard N. (2004a) Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics* 2:1–48.
- Barndorff-Nielsen O.E., Shephard N. (2004b) Econometric analysis of realised covariation: High frequency based covariance, regression, and correlation in financial economics. *Econometrica* 72:885–925.
- Barone-Adesi G. (1985) Arbitrage equilibrium with skewed asset returns. *Journal of Financial and Quantitative Analysis* 20:299–313.

- Barone-Adesi G., Whaley R.E. (1987) Efficient analytic approximation of American option values. *Journal of Finance* 42:301–320.
- Barton D.E., Dennis K.E.R. (1952) The conditions under which Gram-Charlier and Edgeworth curves are positive definite and unimodal. *Biometrika* 39:425–427.
- Basak S., Shapiro A. (2001) Value-at-risk based risk management: optimal policies and asset prices. *Review of Financial Studies* 14:371–405.
- Basle Committee on Banking Supervision (1996): Amendment to the capital accord to incorporate market risks, Bank for International Settlement, Basel Committee on Banking Supervision.
- Bates D.S. (1991) The crash of '87: was it expected? The evidence from options markets. *Journal of Finance* 46:1009–1044.
- Bates D.S. (1996) Jumps and stochastic volatility: exchange rate processes implicit in Deutsche Mark options. *Review of Financial Studies* 9:69–107.
- Bates D.S. (2000) Post-'87 crash fears in S&P 500 futures options. *Journal of Econometrics* 94:181–238.
- Bauwens L., Laurent S. (2005) A new class of multivariate skew densities, with application to GARCH model. *Journal of Business and Economic Statistics* 23:346–354.
- Bauwens L., Laurent S., Rombouts J. (2005) Multivariate GARCH models: A survey. *Journal of Applied Econometrics*, 21:79–109.
- Bawa V., Lindenberg E. (1977) Capital market equilibrium in a mean-lower partial moment framework. *Journal of Financial Economics* 5:189–200.
- Beber A., Brandt M.W. (2003) The effect of macroeconomic news on beliefs and preferences: Evidence from the options market. NBER Working Paper 9914.
- Beirlant J., Vynckier P., Teugels J.L. (1996) Tail index estimation, Pareto quantile plots, and regression diagnostics. *Journal of the American Statistical Association* 91:1659–1667.
- Bekaert G., Harvey C.L. (1995) Time-varying world market integration. *Journal of Finance* 50:403–444.
- Bera A.K., Higgins M.L. (1993) ARCH models: properties, estimation and testing. *Journal of Economic Surveys* 7:305–362.
- Bera A.K., Jarque C.M. (1981) Efficient tests for normality, heteroskedasticity, and serial independence of regression residuals: Monte Carlo evidence. *Economics Letters* 7:313–318.
- Bera A.K., Kim S. (2002) Testing constancy of correlation and other specifications of the BGARCH model with an application to international equity returns. *Journal of Empirical Finance* 9:171–195.
- Berkelaar A., Kouwenberg R. (2000) Dynamic asset allocation and downside-risk aversion. *Econometric Institute Report* 2000-12/a, Erasmus University Rotterdam.
- Bertismas D., Lo A.W. (1998) Optimal control of execution costs. *Journal of Financial Markets* 1:1–50.

- Biais B., Glosten L.R., Spatt C.S. (2002) The microstructure of stock markets. CEPR Discussion Paper 3288.
- Biais B., Hillion P., Spatt C. (1995) An empirical analysis of the limit order book and order flow in the Paris bourse. *Journal of Finance* 50:1655–1689.
- Billingsley P. (1968) Convergence of probability measures. Wiley, New York.
- BIS Committee on the Global Financial System (2000) Stress testing by large financial institutions: current practice and aggregation issues. CGFS Publication 14.
- Black F. (1976) Studies in stock price volatility changes. Proceedings of the 1976 Meeting of the Business and Economic Statistics Section, American Statistical Association 177–181.
- Black F., Scholes M. (1973) The pricing of options and corporate liabilities. *Journal of Political Economy* 81:637–654.
- Blattberg R., Gonedes N. (1974) A comparison of stable and Student distributions as statistical models for stock prices. *Journal of Business* 47:244–280.
- Bliss R.R., Panigirtzoglou N. (2002) Testing the stability of implied probability density functions. *Journal of Banking and Finance* 26:381–422.
- Bollerslev T. (1986) Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31:307–327.
- Bollerslev T. (1987) A conditional heteroskedastic time series model for speculative prices and rates of return. *Review of Economics and Statistics* 69:542–547.
- Bollerslev T. (1988) On the correlation structure for the generalized autoregressive conditional heteroskedastic process. *Journal of Time Series Analysis* 9:121–131.
- Bollerslev T. (1990) Modeling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH approach. *Review of Economics and Statistics* 72:498–505.
- Bollerslev T., Chou R., Kroner K. (1992) ARCH modeling in finance: a review of the theory and empirical evidence. *Journal of Econometrics* 52:5–59.
- Bollerslev T., Engle R.F. (1993) Common persistence in conditional variance. *Econometrica* 61:167–186.
- Bollerslev T., Engle R.F., Nelson D.B. (1994) ARCH models, in Engle R.F., McFadden D.L. (eds) *Handbook of econometrics*, Chapter 49, Vol. IV, Elsevier Science, Amsterdam, 2959–3038.
- Bollerslev T., Engle R.F., Wooldridge J.M. (1988) A capital asset pricing model with time-varying covariance. *Journal of Political Economy* 96:116–131.
- Bollerslev T., Jubinski D. (1999) Equity trading volume and volatility: Latent information arrivals and common long-run dependencies. *Journal of Business and Economic Statistics* 17:9–21.
- Bollerslev T., Wooldridge J.M. (1992) Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric Reviews* 11:143–172.
- Bookstaber R.M., McDonald J.B. (1987) A general distribution for describing

- security price returns. *Journal of Business* 60:401–424.
- Bougerol P., Picard N. (1992) Stationarity of GARCH processes and of some nonnegative time series. *Journal of Econometrics* 52:115–127.
- Boyer B.H., Gibson M.S., Loretan M. (1999) Pitfalls in tests for changes in correlations, Board of Governors of the Federal Reserve System, International Finance Discussion Paper 597.
- Branco M., Dey D. (2001) A class of multivariate skew-elliptical distributions. *Journal of Multivariate Analysis* 79:99–113.
- Brandt M.W., Santa-Clara P. (2002) Simulated likelihood estimation of diffusions with an application to exchange rate dynamics in incomplete markets. *Journal of Financial Economics*, 63:161–210.
- Brännäs K., Nordman N. (2003) Conditional skewness modelling for stock returns. *Applied Economics Letters* 10:725–728.
- Breeden D., Litzenberger R. (1978) Prices of state-contingent claims implicit in option prices. *Journal of Business* 51:621–651.
- Breidt F.J., Carriquiry A.L. (1996) Improved quasi-maximum likelihood estimation for stochastic volatility models. In: Lee J.C., Johnson W.O., Zellner A. (eds) *Modelling and prediction: Honoring Seymour Geisser*. Springer-Verlag, New York.
- Brémaud P. (1981) Point processes and queues (martingale dynamics). Springer-Verlag, New York.
- Brennan M.J. (1979) The pricing of contingent claims in discrete time models. *Journal of Finance* 24:53–68.
- Breyman W., Dias A., Embrechts P. (2003) Dependence structures for multivariate high frequency data in finance. *Quantitative Finance* 3:1–16.
- Brigham E.O. (1988) The fast Fourier transform and its applications. Prentice-Hall, Englewood Cliffs, New Jersey.
- Britten-Jones M., Schaeffer S.M. (1999) Non-linear value-at-risk. *European Finance Review* 2:161–187.
- Brockett P.L., Garven J.R. (1998) A reexamination of the relationship between preferences and moment orderings by rational risk averse investors. *Geneva Papers on Risk and Insurance Theory* 23:127–137.
- Brockett P., Kahane Y. (1992) Risk, return, skewness and preference. *Management Science* 38:851–866.
- Brooks C., Burke S.P., Persand G. (2005) Autoregressive conditional kurtosis, ISMA Centre Discussion Paper in Finance 2002–05. *Journal of Financial Econometrics* 3:399–421.
- Buchen P., Kelly M. (1996) The maximum entropy distribution of an asset inferred from option prices. *Journal of Financial and Quantitative Analysis* 31:143–159.
- Butler C., Davies H. (1998) Assessing market views on monetary policy: The use of implied risk neutral probability distributions. Bank of England Monetary Instruments and Markets Division, Working Paper.
- Campbell J., Cochrane J. (1999) By force of habit: A consumption based explanation of aggregate stock market behavior. *Journal of Political Economy*

- 107:205–251.
- Campbell J.Y., Hentschel L. (1992) No news is good news: An asymmetric model of changing volatility. *Journal of Financial Economics* 31:281–318.
- Campbell J.Y., Lo A.W., MacKinlay A.C. (1997) *The econometrics of financial markets*. Princeton University Press, Princeton, New Jersey.
- Campbell J.Y., Viceira L.M. (1999) Consumption and portfolio decisions when expected returns are time varying. *Quarterly Journal of Economics* 114:433–495.
- Cappiello L., Engle R.F., Sheppard K. (2003) Asymmetric dynamics in the correlations of global equity and bond returns. *European Central Bank Working Paper* 204.
- Carr P., Geman H., Madan D., Yor M. (2002) The fine structure of asset returns: An empirical investigation. *Journal of Business* 75:305–332.
- Carr P., Geman H., Madan D., Yor M. (2003) Stochastic volatility for Lévy processes. *Mathematical Finance* 13:345–382.
- Carr P., Madan D. (1999) Option pricing and the fast Fourier transform. *Journal of Computational Finance* 61–73.
- Carr P., Wu L. (2004) Time-changed Lévy processes and option pricing. *Journal of Financial Economics* 71:113–141.
- Chacko G., Viceira L.M. (2003) Spectral GMM estimation of continuous-time processes. *Journal of Econometrics* 116:259–292.
- Chamberlain G. (1983) A characterization of the distributions that imply mean-variance utility functions. *Journal of Economic Theory* 29:185–201.
- Chan K., Fong W-M. (2000) Trade size, order imbalance, and the volatility-volume relation. *Journal of Financial Economics* 57:247–273.
- Chan L.K.C., Lakonishok J. (1993) Institutional trades and intraday stock price behavior. *Journal of Financial Economics* 33:173–199.
- Chan L.K.C., Lakonishok J. (1995) The behavior of stock prices around institutional trades. *Journal of Finance* 50:1147–1174.
- Chan W.H., Maheu J.M. (2002) Conditional jump dynamics in stock market returns. *Journal of Business and Economic Statistics* 20:377–389.
- Chaudhury M., Wei J. (1994) Upper bounds for american futures options: A note. *Journal of Futures Markets* 14:111–116.
- Chernozhukov V., Umantsev L. (2001) Conditional value-at-risk: Aspects of modeling and estimation. *Empirical Economics* 26:271–292.
- Christie A.A. (1982) The stochastic behavior of common stock variances: Value, leverage, and interest rate effects. *Journal of Financial Economics* 10:407–432.
- Christoffersen P. (1998) Evaluating interval forecasts. *International Economic Review* 39:841–862.
- Christoffersen P., Hahn J., Inoue A. (2001) Testing and comparing value-at-risk measures. *Journal of Empirical Finance* 8:325–342.
- Chunhachinda P., Dandapani K., Hamid S., Prakash A.J. (1997) Portfolio selection and skewness: Evidence from international stock markets. *Journal of Banking and Finance* 21:143–167.

- Clark P.K. (1973) A subordinated stochastic process model with finite variance for speculative prices. *Econometrica* 41:135–155.
- Clements M.P. (2002) An evaluation of the survey of professional forecasters probability distributions of expected inflation and output growth. Warwick University Discussion Paper.
- Coles S.G., Heffernan J., Tawn J.A. (1999) Dependence measures for extreme value analyses. *Extremes* 3:5–38.
- Coles S.G., Tawn J.A. (1991) Modelling extreme multivariate events. *Journal of the Royal Statistical Society* 53:377–392.
- Coles S.G., Tawn J.A. (1994) Statistical methods for multivariate extremes: An application to structural design. *Applied Statistics* 43:1–48.
- Cont R. (2001) Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance* 1:223–236.
- Cont R., Tankov P. (2004) Financial modelling with jump processes. Chapman and Hall/CRC.
- Corrado C.J., Su T. (1996) S&P 500 index option tests of Jarrow and Rudd's approximate option valuation formula. *Journal of Futures Markets* 6:611–629.
- Coutant S., Jondeau E., Rockinger M. (2001) Reading PIBOR futures options smiles: The 1997 snap election. *Journal of Banking and Finance* 25:1957–1987.
- Cox J.C., Ingersoll J.I., Ross S.A. (1985) A theory of the term structure of interest rates. *Econometrica* 53:385–407.
- Cox J.C., Ross S., Rubinstein M. (1979) Option pricing: A simplified approach. *Journal of Financial Economics* 7:229–264.
- Cox J.C., Rubinstein M. (1985) Options markets. Prentice-Hall, Englewood Cliffs, New Jersey.
- Cramér H. (1946) Mathematical methods of statistics. Princeton University Press, Princeton, New Jersey.
- D'Agostino R.B. (1970) Transformation to normality of the null distribution of g_1 . *Biometrika* 57:679–681.
- Dana R.-A., Jeanblanc-Piqué M. (2002) Financial markets in continuous time. Springer-Verlag, Berlin.
- Danielsson J., de Haan L., Peng L., de Vries C.G. (2001) Using a bootstrap method to choose the sample fraction in tail index estimation. *Journal of Multivariate Analysis* 76:226–248
- Danielsson J., de Vries C.G. (1997) Tail index and quantile estimation with very high frequency data. *Journal of Empirical Finance* 4:241–257.
- Daragahi-Noubary G.R. (1989) On tail estimation: An improved method. *Mathematical Geology* 21:829–842.
- Darolles S., Gouriéroux C., Jasik J. (2001) Compound autoregressive models. CREST and York University, Working paper.
- Das S.R., Sundaram R.K. (1997) Taming the skew: Higher-order moments in modeling asset price processes in finance. NBER Working Paper 5976.
- Davis P.J., Polonsky I. (1970) Numerical interpolation, differentiation and integration. In: Abramowitz M., Stegun I.A. (eds) *Handbook of mathematical*

- functions. Dover Publications, New York, 875–924.
- Davis P.J., Rabinowitz P. (1984) Methods of numerical integration. Academic Press, New York.
- Davison A.C. (1984) Modeling excesses over high thresholds with an application. In: Tiago de Oliveria J. (ed.) Statistical extremes and applications. Reidel, Dordrecht, 461–482.
- Davison A.C., Smith R.L. (1990) Models for exceedances over high thresholds. *Journal of the Royal Statistical Society* 52:393–442.
- De Haan L. (1985) Extremes in higher dimensions. Proceedings of the 45th Session of the International Statistics Institute, paper 26.3.
- De Haan L., Resnick S.I. (1977) Limit theory for multivariate sample extremes. *Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete* 40:317–337.
- De Haan L., Resnick S.I., Rootzén H., de Vries C.G. (1989) Extremal behavior of solutions to a stochastic difference equation with applications to ARCH processes. *Stochastic Processes and Their Applications* 32:213–224.
- De Haan L., de Ronde J. (1998) Sea and wind: Multivariate extremes at work. *Extremes* 1:7–45.
- Deheuvels P. (1979) La fonction de dépendance empirique et ses propriétés – Un test non paramétrique d'indépendance. *Académie Royale de Belgique – Bulletin de la Classe des Sciences 5e Série* 65:274–292.
- Deheuvels P. (1981) A nonparametric test for independence. *Publication de l'Institut Statistique Universitaire de Paris* 26:29–50.
- Dekkers A.L.M., Einmahl J.H.J., de Haan L. (1989) A moment estimator for the index of an extreme-value distribution. *Annals of Statistics* 17:1833–1855.
- Dekkers A.L.M., de Haan L. (1989) On the estimation of the extreme value index and large quantile estimation. *Annals of Statistics* 17:1795–1832.
- Delbaen F., Schachermayer W. (1994) A general version of the fundamental theorem of asset pricing, *Mathematische Annalen* 300:463–520.
- Dellacherie C., Meyer P.A. (1982) Probabilities and potential B: Theory of martingales. North-Holland, Amsterdam.
- Demarchi M., Foucault T. (2000) Equity trading systems in Europe. *Annales d'Economie et de Statistique* 60:73–115.
- Demos A., Sentana E. (1998) Testing for GARCH effects: a one-sided approach. *Journal of Econometrics* 86:97–127.
- Diebold F.X. (1988) Empirical modelling of exchange rates. Springer-Verlag, New York.
- Diebold F.X., Gunther T.A., Tay A.S. (1998) Evaluating density forecasts with applications to financial risk management. *International Economic Review* 39:863–883.
- Diebold F.X., Hahn J.Y., Tay A.S. (1999) Multivariate density forecast evaluation and calibration in financial risk management: High frequency returns on foreign exchange. *Review of Economics and Statistics* 81:661–673.
- Diebold F.X., Hickman A., Inoue A., Schuermann T. (1998) Converting 1-day volatility to h-day volatility: Scaling by root-h is worse than you think. Wharton Financial Institutions Center, Working paper 97–34.

- Diebold F.X., Schuermann T., Stroughair J.D. (1998) Pitfalls and opportunities in the use of extreme value theory in risk management. Wharton School, University of Pennsylvania, Working paper.
- Dittmar R.F. (2002) Nonlinear pricing kernels, kurtosis preference, and evidence from the cross section of equity returns, *Journal of Finance* 57:369–403.
- Donsker M.D. (1951) An invariance principle for certain probability limit theorems, *Memoirs of the American Mathematical Society* 6.
- Doornik J.A., Hansen H. (1994) An omnibus test for univariate and multivariate normality, Nuffield College, Oxford, Working paper.
- Drees H., Kaufman E. (1998) Selecting the optimal sample fraction in univariate extreme value estimation. *Stochastic Processes and Their Applications* 75:149–172.
- Drost F.C., Klassen C.A.J. (1997) Efficient estimation in semiparametric GARCH models. *Journal of Econometrics* 81:193–221.
- Drost F.C., Nijman T.E. (1993) Temporal aggregation of GARCH processes. *Econometrica* 61:909–927.
- Drost F.C., Werker B.J.M. (1996) Closing the GARCH gap: Continuous time GARCH modeling. *Journal of Econometrics* 74:31–57.
- Dubil R. (2002) How quickly should you liquidate your vested stock? University of Connecticut, Working Paper.
- Duffie D. (1988) Security markets: Stochastic models. Academic Press, Boston.
- Duffie D., Pan J. (1997) An overview of value at risk. *Journal of Derivatives* 4:7–49.
- Duffie D., Pan J., Singleton K.J. (2000) Transform analysis and asset pricing for affine jump-diffusions. *Econometrica* 68:1343–1376.
- Duffie D., Singleton K.J. (1993) Simulated moments estimation of Markov models of asset prices. *Econometrica* 61:929–952.
- Dumas B., Fleming J., Whaley R.E. (1998) Implied volatility functions: empirical tests. *Journal of Finance* 53:2059–2106.
- DuMouchel W.H. (1973) On the asymptotic normality of the maximum-likelihood estimate when sampling from a stable distribution. *Annals of Statistics* 1:948–957.
- Durham G., Gallant A.R. (2002) Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes. *Journal of Business and Economic Statistics* 20:297–317.
- Easley D., Kiefer N.M., O’Hara M. (1997) One day in the life of a very common stock. *Review of Financial Studies* 10:805–835.
- Eberlein E., Keller U., Prause K. (1998) New insights into smile, mispricing and value at risk. *Journal of Business* 71:371–406.
- Embrechts P., Klüppelberg C., Mikosch T. (1997) Modelling extremal events for insurance and finance. Springer-Verlag, Berlin.
- Embrechts P., Lindskog F., McNeil A. (2003) Modelling dependence with copulas and applications to risk management. In: Rachev S.T. (ed.) *Handbook of heavy tailed distributions in finance*. Elsevier/North-Holland, Amsterdam.

- Engle R.F. (1982) Auto-regressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50:987–1007.
- Engle R.F. (2002) Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business and Economic Statistics* 20:339–350.
- Engle R.F., Bollerslev T. (1986) Modeling the persistence of conditional variances. *Econometric Review* 5:1–50.
- Engle R.F., González-Rivera G. (1991) Semi-parametric ARCH models. *Journal of Business and Economic Statistics* 9:345–359.
- Engle R.F., Kroner K. (1995) Multivariate simultaneous GARCH. *Econometric Theory* 11:122–150.
- Engle R.F., Lilien D.M., Robbins R.P. (1987) Estimating time varying risk premia in the term structure: the ARCH-M model. *Econometrica* 55:391–407.
- Engle R.F., Manganelli S. (2004) CAViaR: Conditional value at risk by regression quantiles. *Journal of Business and Economic Statistics* 22:367–381.
- Engle R.F., Mezrich J. (1996) GARCH for groups. *Risk* 9:36–40.
- Engle R.F., Ng V.K. (1993) Measuring and testing the impact of news on volatility. *Journal of Finance* 48:1749–1778.
- Engle R.F., Ng V.K., Rothschild M. (1990) Asset pricing with a factor ARCH covariance structure: Empirical estimates for Treasury Bills. *Journal of Econometrics* 45:213–237.
- Engle R.F., Ng V.K., Rothschild M. (1992) A multi-dynamic factor model for stock returns. *Journal of Econometrics* 52:245–266.
- Engle R.F., Russel J.R. (1998) Autoregressive conditional duration: A new model for irregularly spaced transaction data. *Econometrica* 66:1127–1162.
- Engle R.F., Sheppard K. (2001) Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH, NBER Working Paper 8554.
- Eraker B., Johannes M., Polson N. (2003) The impact of jumps in volatility and returns. *Journal of Finance* 58:1269–1300.
- Fama E. (1963) Mandelbrot and the stable paretian hypothesis. *Journal of Business* 36:420–429.
- Fama E. (1965) The behavior of stock market prices. *Journal of Business* 38:34–105.
- Fang K.-T., Kotz S., Ng K. (1990) Symmetric multivariate and related distributions, Chapman and Hall/CRC, London.
- Feller W. (1950) Two singular diffusion problems. *Annals of Mathematics* 54:173–182.
- Fermanian J.D. (2003) Goodness-of-fit tests for copulas, CREST Working Paper 2003-34.
- Fermanian J.D., Scaillet O. (2003) Nonparametric estimation of copulas for time series. *Journal of Risk* 5:25–54.
- Fermanian J.D., Scaillet O. (2005) Sensitivity analysis of VaR and expected

- shortfall for portfolios under netting agreements. *Journal of Banking and Finance* 29:927–958.
- Fernández C., Steel M.F.J. (1998) On Bayesian modelling of fat tails and skewness. *Journal of the American Statistical Association* 93:359–371.
- Ferreira J.T.A.S., Steel M.F.J. (2004) A constructive representation of univariate skewed distributions. *Econometrics* 0403002, Economics Working paper Archive at WUSTL.
- Fisher R.A. (1915) Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika* 10:507–521.
- Fisher R.A., Tippett L.H.C. (1928) Limiting forms of the frequency distributions of the largest or smallest member of a sample. *Proceedings of the Cambridge Philosophical Society* 24:180–190.
- Flamouris D., Giamouridis D. (2002) Estimating implied PDFs from American options on futures: A new semi-parametric approach. *Journal of Futures Markets* 22:1–30.
- Fletcher R. (1994) An overview of unconstrained optimization. In: Spedicato E. (ed.) *Algorithms for continuous optimization: The state of the art*. Kluwer Academic Publishers, 109–143.
- Forbes K., Rigobon R. (2002) No contagion, only interdependence: measuring stock-market comovements. *Journal of Finance* 52:2223–2261.
- Foster F.D., Viswanathan S. (1995) Can speculative trading explain the volume-volatility relation? *Journal of Business and Economic Statistics* 13:379–396.
- Francq C. and Zakoian J.M. (2000) Estimating weak GARCH representations, *Econometric Theory* 16:692–728.
- French K.R., Schwert G.W., Stambaugh R.F. (1987) Expected stock returns and volatility. *Journal of Financial Economics* 19:3–29.
- Friend I., Westerfield R. (1980) Co-skewness and capital asset pricing. *Journal of Finance* 35:897–913.
- Gagliardini P., Gourieroux C. (2006) Efficient nonparametric estimation of models with nonlinear dependence. Forthcoming in *Journal of Econometrics*.
- Gallant A.R., Hsieh D., Tauchen G. (1991) On fitting a recalcitrant series: The Pound/Dollar exchange rate, 1974–1983. In: Barnett W.A., Powell J., Tauchen G. (eds) *Nonparametric and semiparametric methods in econometrics and statistics*. Cambridge University Press, Cambridge.
- Gallant A.R., Rossi P.E., Tauchen G. (1992) Stock prices and volume. *Review of Financial Studies* 5:199–242.
- Gallant A.R., Tauchen G. (1989) Semi-nonparametric estimation of conditionally constrained heterogeneous processes: Asset pricing applications. *Econometrica* 57:1091–1120.
- Gallant A.R., Tauchen G. (1996) Which moments to match? *Econometric Theory* 12:657–681.
- Garman M., Klass M. (1980) On the estimation of security price volatilities from historical data. *Journal of Business* 53:67–78.

- Garman M., Kohlhagen S. (1983) Foreign currency option values. *Journal of International Money and Finance* 2:231–238.
- Geman H., Madan D., Yor M. (2001) Time changes for Lévy processes. *Mathematical Finance* 11:79–96.
- Geman H., Madan D., Yor M. (2002) Stochastic volatility, jumps and hidden time changes. *Finance and Stochastics* 6:63–90.
- Genest C., Ghoudi K., Rivest L.-P. (1995) A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika* 82:543–552.
- Genest C., Rivest L.-P. (1993) Statistical inference procedures for bivariate archimedean copulas. *Journal of the American Statistical Association* 88:1034–1043.
- Ghysels E., Harvey A.C., Renault E. (1996) Stochastic volatility. In: Maddala G.S., Rao C.R. (eds.) *Handbook of statistics*, Volume 14, Statistical methods in finance. North-Holland, Amsterdam, 119–191.
- Gill P.E., Murray W., Saunders M.A. (1999) User's guide for SNOPT 5.3: A Fortran package for large-scale nonlinear programming, Working Paper, Department of Mathematics, UCSD, California.
- Gill P.E., Murray W., Saunders M.A. (1997) SNOPT: An SQP Algorithm of large-scale constrained optimization, Report NA 97-2, Department of Mathematics, UCSD, California.
- Giot P., Laurent S. (2003) Value-at-risk for long and short trading positions. *Journal of Applied Econometrics* 18:641–663.
- Glasserman P., Heidelberger P., Shahabuddin P. (2002) Portfolio value-at-risk with heavy-tailed risk factors. *Mathematical Finance* 12:239–270.
- Glosten L.R., Jagannathan R., Runkle D.E. (1993) On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance* 48:1779–1801.
- Gnedenko B.V. (1943) Sur la distribution limite du terme maximum d'une série aléatoire. *Annals of Mathematics* 44:423–453.
- Gnedenko B.V., Kolmogorov A.N. (1954) Limit distributions for sums of independent random variables. Addison-Wesley, Cambridge, Massachusetts.
- Goettler R., Parlour C., Rajan U. (2005) Equilibrium in a dynamic limit order market. *Journal of Finance* 60:2149–2192.
- Golan A., Judge G.G., Miller D. (1996) Maximum entropy econometrics: Robust estimation with limited data, Wiley, New York.
- Gouriéroux C., Jasiak J. (2001a) Financial econometrics: Problems, models, and methods. Princeton University Press, Princeton.
- Gouriéroux C., Jasiak J. (2001b) Memory and infrequent breaks. *Economics Letters* 70:29–41.
- Gouriéroux C., Jasiak J. (2003) Autoregressive gamma models. CREST and York University, Working Paper.
- Gouriéroux C., Laurent J.-P., Scaillet O. (2000) Sensitivity analysis of values at risk. *Journal of Empirical Finance* 7:225–245.

- Gouriéroux C., Monfort A. (1992) Qualitative threshold ARCH models. *Journal of Econometrics* 52:159–199.
- Gouriéroux C., Monfort A., Renault E. (1996) Simulation based econometric methods, Core Lectures, Oxford University Press.
- Gouriéroux C., Monfort A., Trognon A. (1984) Pseudo maximum likelihood methods: Theory. *Econometrica* 52:680–700.
- Goyal A., Santa-Clara P. (2003) Idiosyncratic risk matters. *Journal of Finance* 58:975–1007.
- Gradshteyn I.S., Ryzhik I. M. (1994) Table of integrals, series, and products. Academic Press, San Diego, 5th edition.
- Gray S.F. (1995) An analysis of conditional regime-switching models, Fuqua School of Business, Duke University, Working Paper.
- Guidolin M., Timmermann A. (2003) Optimal portfolio choice under regime switching, skew and kurtosis preferences. University of California San Diego, Working Paper.
- Gumbel E.J. (1958) Statistics of extremes. Columbia University Press, New York.
- Hafner C., Rombouts J. (2003) Semiparametric multivariate GARCH models. CORE DP 2003/03.
- Hall P. (1990) Using the bootstrap to estimate mean square error and select smoothing parameter in nonparametric problem. *Journal of Multivariate Analysis* 32:177–203.
- Hall P., Welsh A.H. (1985) Adaptive estimates of parameters of regular variation. *Annals of Statistics* 13:331–341.
- Hamao Y., Masulis R.W., Ng V.K. (1990) Correlations in price changes and volatility across international stock markets. *Review of Financial Studies* 3:281–307.
- Hamburger H. (1920) Über eine Erweiterung des Stieltjesschen Momentproblems. *Mathematische Zeitschrift* 7:235–319.
- Hamilton J.D. (1994) Time series analysis. Princeton University Press, Princeton, New Jersey.
- Hamilton J.D. (1989) A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* 57:357–384.
- Hansen B. E. (1994) Autoregressive conditional density estimation. *International Economic Review* 35:705–730.
- Hansen L. (1982) Large sample properties of generalized method of moments estimator. *Econometrica* 50:1029–1054.
- Härdle W. (1990) Applied nonparametric regression, Econometric Society Monographs 19, Cambridge University Press, Cambridge.
- Harris L. (1986) Cross-security tests of the mixture of distributions hypothesis. *Journal of Financial and Quantitative Analysis* 21:39–46.
- Harris L. (1987) Transactions data tests of the mixture of distributions hypothesis. *Journal of Financial and Quantitative Analysis* 22:127–141.

- Harrison J.M., Kreps D. (1979) Martingales and arbitrage in multiperiod security markets. *Journal of Economic Theory* 20:381–408.
- Harrison J.M., Pliska S.R. (1981) Martingales and stochastic integrals in the theory of continuous trading. *Stochastic Processes and Their Applications* 11:215–260.
- Harrison J.M., Pliska S.R. (1983) A stochastic calculus model of continuous trading: Complete markets. *Stochastic Processes and Their Applications* 15:313–316.
- Hartmann P., Straetmans S., de Vries C.G. (2004) Asset market linkages in crisis periods. *Review of Economics and Statistics* 86:313–326.
- Harvey A.C., Ruiz E., Shephard N. (1994) Multivariate stochastic variance models. *Review of Economic Studies* 61:247–264.
- Harvey C.R. (1995) Predictable risk and returns in emerging markets. *Review of Financial Studies* 8:773–816.
- Harvey C.R., Liechty J.C., Liechty M.W., Müller P. (2002) Portfolio selection with higher moments. Duke University Durham, Working Paper.
- Harvey C.R., Siddique A. (1999) Autoregressive conditional skewness. *Journal of Financial and Quantitative Analysis* 34:465–487.
- Harvey C.R., Siddique A. (2000) Conditional skewness in asset pricing tests. *Journal of Finance* 55:1263–1295.
- Heffernan J.E. (2000) A directory of coefficients of tail dependence. *Extremes* 3:279–290.
- Hentschel L. (1995) All in the family: nesting symmetric and asymmetric GARCH models. *Journal of Financial Economics* 39:71–104.
- Heston S. (1993) A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies* 6:327–343.
- Hill B.M. (1975) A simple general approach to inference about the tail of a distribution. *Annals of Statistics* 3:1163–1173.
- Higgins M.L., Bera A.K. (1992) A class of nonlinear ARCH models. *International Economic Review* 33:137–158.
- Hlawitschka W. (1994) The empirical nature of Taylor series approximations to expected utility. *American Economic Review* 84:713–719.
- Hols M.C.A.B., de Vries C.G. (1991) The limiting distribution of extremal exchange rate returns. *Journal of Applied Econometrics* 6:287–302.
- Holthausen R.W., Leftwich R.W., Mayers D. (1987) The effect of large block transactions on security prices: A cross-sectional analysis. *Journal of Financial Economics* 19:237–267.
- Holthausen R.W., Leftwich R.W., Mayers D. (1990) Large-block transactions, the speed of response, and temporary and permanent stock-price effects. *Journal of Financial Economics* 26:71–95.
- Hosking J.R.M. (1980) The multivariate portmanteau statistic. *Journal of the American Statistical Association* 75:602–608.
- Hosking J.R.M., Wallis J.R. (1987) Parameter and quantile estimation for the

- generalized Pareto distribution. *Technometrics* 29:339–349.
- Hosking J.R.M., Wallis J.R., Wood E.F. (1985) Estimation of the generalized extreme value distribution by the method of probability-weighted moments. *Technometrics* 27:251–261.
- Hsieh D. (1993) Implications of nonlinear dynamics for financial risk management. *Journal of Financial and Quantitative Analysis* 28:41–64.
- Hsing T. (1993) Extremal index estimation for a weakly dependent stationary sequence. *Annals of Statistics* 2:2043–2071.
- Hsing T., Hüsler J., Leadbetter M.R. (1988) On the exceedance point process for a stationary sequence. *Probability Theory and Related Fields* 78:97–112.
- Huisman R., Koedijk K.G., Pownall R.A.J. (1999) Asset allocation in a value-at-risk framework. Erasmus University, Rotterdam, Faculty of Business Administration, Working Paper.
- Hull J. (1993) Options, futures and other derivative securities. Prentice-Hall, Englewood Cliffs, New Jersey, 2nd edition.
- Hull J., White A. (1987) The pricing of options on assets with stochastic volatilities. *Journal of Finance* 42:281–300.
- Hull J., White A. (1988) An analysis of the bias in option pricing caused by a stochastic volatility. *Advances in Futures and Options Research* 3:27–61.
- Hwang S., Satchell S.E. (1999) Modelling emerging market risk premia using higher moments. *Journal of Finance and Economics* 4:271–296.
- Ingersoll J.E. (1987) Theory of financial decision making. Rowman and Littlefield Publishers, Totowa, New Jersey.
- Jackwerth J.C. (1997) Generalized binomial trees. *Journal of Derivatives* 5:7–17.
- Jackwerth J.C. (1999) Option implied risk-neutral distributions and implied binomial trees: A literature review. *Journal of Derivatives* 7:66–82.
- Jackwerth J.C. (2000) Recovering risk aversion from option prices and realized returns. *Review of Financial Studies* 13:433–451.
- Jackwerth J.C., Rubinstein M. (1996) Recovering probability distributions from option prices. *Journal of Finance* 51:1611–1631.
- Jacod J., Shiryaev A.N. (2002) Limit theorems for stochastic processes. Springer-Verlag, Berlin.
- Jacquier E., Polson N.G., Rossi P.E. (1994) Bayesian analysis of stochastic volatility models. *Journal of Business and Economic Statistics* 12:371–389.
- Jansen D.W., de Vries C.G. (1991) On the frequency of large stock returns: putting booms and busts into perspective. *Review of Economics and Statistics* 73:18–24.
- Jarque C.M., Bera A.K. (1980) Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economic Letters* 6:255–259.
- Jarrow R., Rudd A. (1982) Approximate valuation for arbitrary stochastic processes. *Journal of Financial Economics* 10:347–369.
- Jaynes E.T. (1957) Information theory and statistical mechanics. *Physical Review* 106:620–630.

- Jaynes E.T. (1982) On the rationale of maximum-entropy methods. *Proceedings of the IEEE* 70:939–952.
- Jeantheau T. (1998) Strong consistency of estimators for multivariate ARCH models. *Econometric Theory* 14:70–86.
- Jenkinson A.F. (1955) The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorology Society* 87:145–158.
- Jennrich R.J. (1970) An asymptotic chi-square test for the equality of two correlation matrices. *Journal of the American Statistical Association* 65:904–912.
- Jiang G.J., Knight J.L. (2002) Estimation of continuous-time processes via the empirical characteristic function. *Journal of Business and Economic Statistics* 20:198–212.
- Joe H. (1997) Multivariate models and dependence concepts. Chapman and Hall, London.
- Joe H., Xu J.J. (1996) The estimation method of inference functions for margins for multivariate models. Technical Report, Department of Statistics, University of British Columbia, 166.
- Johnson N.L., Kotz S., Balakrishnan N. (1994) Continuous univariate distribution, Vol. 1, Wiley, New York, 2nd ed.
- Jondeau E., Rockinger M. (2000) Reading the smile: The message conveyed by methods which infer risk neutral densities. *Journal of International Money and Finance* 19:885–915.
- Jondeau E., Rockinger M. (2001) Gram-Charlier densities. *Journal of Economic Dynamics and Control* 25:1457–1483.
- Jondeau E., Rockinger M. (2003a) Conditional volatility, skewness, and kurtosis: Existence, persistence, and comovements. *Journal of Economic Dynamic and Control* 27:1699–1737.
- Jondeau E., Rockinger M. (2003b) User's guide. *Journal of Economic Dynamic and Control* 27:1739–1742.
- Jondeau E., Rockinger M. (2003c) Testing for differences in the tails of stock-market returns. *Journal of Empirical Finance* 10:559–581.
- Jondeau E., Rockinger M. (2005) Conditional asset allocation under non normality, FAME and HEC Lausanne, Working Paper.
- Jondeau E., Rockinger M. (2006a) Optimal portfolio allocation under higher moments. *European Financial Management* 12:29–55.
- Jondeau E., Rockinger M. (2006b) The copula-GARCH model of conditional dependencies: An international stock-market application. Forthcoming in *Journal of International Money and Finance*.
- Jones C.M., Kaul G., Lipson M.L. (1994) Transactions, volume and volatility. *Review of Financial Studies* 7:631–651.
- Jones M.C., Faddy, M.J. (2003) A skew extension of the t distribution, with applications. *Journal of the Royal Statistical Society B* 65:159–174.
- Jorion P. (1989) On jump processes in the foreign exchange and stock markets. *Review of Financial Studies* 1:431–445.

- Jorion P. (1997) Value at risk: The new benchmark for controlling market risk. Irwin Professional Publishing, Chicago, Illinois.
- Judd K.L. (1998) Numerical methods in economics. MIT Press, Cambridge, Massachusetts.
- Jurczenko E., Maillet B. (2001) The 3-CAPM: Theoretical foundations and an asset pricing model comparison in a unified framework. In: Dunis C., Timmermann A., Moody J. (eds.) Developments in forecast combination and portfolio choice. Wiley, New York.
- Jurczenko E., Maillet B. (2006) The 4-CAPM: In between asset pricing and asset allocation. In: Adcock C., Jurczenko E., Maillet B. (eds) Multi-moment capital asset pricing models and related topics. Springer. Forthcoming.
- Kaplanis E. C. (1988) Stability and forecasting of the co-movement measures of international stock market return. *Journal of International Money and Finance* 8:63–95.
- Karatzas I., Shreve S.E. (1988) Brownian motion and stochastic calculus. Springer-Verlag, New York.
- Karpoff J.M. (1987) The relation between price changes and trading volume: A survey. *Journal of Financial and Quantitative Analysis* 22:109–126.
- Kearns P., A. (1977) Estimating the density tail index for financial time series. *Review of Economics and Statistics* 79:171–175.
- Kendall M., Stuart A. (1977) The advanced theory of statistics, Vol. 1. Macmillan, New York.
- Kim S., Shephard N., Chib S. (1998) Stochastic volatility: likelihood inference and comparison with ARCH models. *Review of Economic Studies* 65:361–393.
- Kimball M.S. (1993) Standard risk aversion. *Econometrica* 61:589–611.
- King M., Sentana E., Wadhwani S. (1994) Volatility and links between national stock markets. *Econometrica* 62:901–933.
- Knight J., Satchell S. (2001) Return distributions in finance. Butterworth-Heinemann, Oxford.
- Koch P.D., Koch T.W. (1991) Evolution in dynamic linkages across national stock indexes. *Journal of International Money and Finance* 10:231–251.
- Koedijk K.G., Kool C.J.M. (1992) Tail estimates of East European exchange rates. *Journal of Business and Economic Statistics* 10:83–96.
- Koedijk K.G., Schafgans M.M.A., de Vries C.G. (1990) The tail index of exchange rate returns. *Journal of International Economics* 29:93–108.
- Koenker R., Bassett G.W. (1978) Regression quantiles. *Econometrica* 46:33–50.
- Koenker R., Hallock K.F. (2001) Quantile regression. *Journal of Economic Perspectives* 15:143–156.
- Kolmogorov A.N. (1933) Grundbegriffe der Wahrscheinlichkeitsrechnung. Springer, Berlin. English translation (1950): Foundations of the theory of probability. Chelsea, New York.
- Kou S.G. (2002) A jump-diffusion model for option pricing. *Management Science* 48:1086–1101.

- Koutrouvelis I.A. (1980) Regression-type estimation of the parameters of stable laws. *Journal of the American Statistical Association* 75:918–928.
- Kraft D.F., Engle R.F. (1982) Autoregressive conditional heteroskedasticity in multiple time series, UCSD Working Paper.
- Kraus A., Litzenberger R.H. (1976) Skewness preference and the valuation of risk assets. *Journal of Finance* 31:1085–1100.
- Kraus A., Stoll H. (1972) Price impacts of block trading on the New York stock exchange. *Journal of Finance* 27:569–588.
- Krokhmal P., Palmquist J., Uryasev S. (2002) Portfolio optimization with conditional value-at-risk objective and constraints. *Journal of Risk* 4:11–27.
- Kroner K.F., Ng V.K. (1998) Modeling asymmetric comovements of asset returns. *Review of Financial Studies* 11:817–844.
- Kupiec P.H. (1995) Techniques for verifying the accuracy of risk measurement models. Board of Governors of the Federal Reserve System, Finance and Economics Discussion Series 95-24.
- Lai T.Y. (1991) Portfolio selection with skewness: A multiple-objective approach. *Review of Quantitative Finance and Accounting* 1:293–305.
- Lambert P., Laurent S. (2002) Modeling skewness dynamics in series of financial data using skewed location-scale distributions. Université Catholique de Louvain and Université de Liège, Working Paper.
- Lamoureux, C.G., Lastrapes W.D. (1990) Heteroscedasticity in stock return data: Volume versus GARCH effects. *Journal of Finance* 45:221–230.
- Lamoureux, C.G., Lastrapes W.D. (1994) Endogenous trading volume and momentum in stock-return volatility. *Journal of Business & Economic Statistics* 12:253–260.
- Lanne M., Saikkonen P. (2005) A multivariate generalized orthogonal factor GARCH model. Discussion Paper 63, Helsinki Center of Economic Research.
- Laurent S., Peters J.P. (2005) G@RCH 4.0, estimating and forecasting ARCH models. Timberlake Consultants.
- Leadbetter M.R., Lindgren G., Rootzén H. (1983) Extremes and related properties of random sequences and processes. Springer-Verlag, New York.
- Ledford A.W., Tawn J.A. (1996) Statistics for near independence in multivariate extreme values. *Biometrika* 83:169–187.
- Ledford A.W., Tawn J.A. (1997) Modeling dependence within joint tail regions. *Journal of the Royal Statistical Society B* 49:475–499.
- Ledford A.W., Tawn J.A. (1998) Concomitant tail behaviour for extremes. *Advances in Applied Probability* 30:197–215.
- Ledoit O., Santa-Clara P., Wolf M. (2003) Flexible multivariate GARCH modeling with an application to international stock markets. *Review of Economics and Statistics* 85:735–747.
- Lee J.H.H. (1991) A Lagrange multiplier test for GARCH models. *Economics Letters* 37:265–271.
- Lee Y., Tse T. (1991) Term structure of interest rates in the Singapore Asian Dollar market. *Journal of Applied Econometrics* 6:143–152.

- Lepingle D. (1976) La variation d'ordre p des semi-martingales. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 36:295–316.
- Lévy P. (1954) Théorie de l'addition des variables aléatoires. Gauthier-Villars, Paris.
- Lhabitant F.S. (1998) On the (ab)use of Taylor series approximations for portfolio selection, portfolio performance and risk management. University of Lausanne, Working Paper.
- Li W.K., Ling S., McAleer M. (2002) A survey of recent theoretical results for time series models with GARCH errors. *Journal of Economic Survey* 16:245–269.
- Liesenfeld R. (1998) Dynamic bivariate mixture models: Modeling the behavior of prices and trading volume. *Journal of Business and Economic Statistics* 16:101–109.
- Liesenfeld R. (2001) A generalized bivariate mixture model for stock price volatility and trading volume. *Journal of Econometrics* 104:141–178.
- Liu, X., Shackleton M.B., Taylor S.J., Xu X. (2003) Closed-form transformations from risk-neutral to real-world distributions. Dept. of Accounting and Finance Working Paper, Lancaster University.
- Ljung G.M., Box G.E.P. (1978) On a measure of lack of fit in time series models. *Biometrika* 65:297–303.
- Lo A.W., MacKinlay A.C., Zhang J. (2002) Econometric models of limit-order executions. *Journal of Financial Economics* 65:31–71.
- Lo A.W., Wang J. (2000) Trading volume: Definitions, data analysis, and implications of portfolio theory. *Review of Financial Studies* 13:257–300.
- Loistl O. (1976) The erroneous approximation of expected utility by means of a Taylor's series expansion: analytic and computational results. *American Economic Review* 66:904–910.
- Longin F.M. (1996) The asymptotic distribution of extreme stock market returns. *Journal of Business* 69:383–408.
- Longin F.M. (2000) From VaR to stress testing: The extreme value approach. *Journal of Banking and Finance* 24:1097–1130.
- Longin F.M., Solnik B. (1995) Is the correlation in international equity returns constant: 1960–1990? *Journal of International Money and Finance* 14:3–26.
- Longin F.M., Solnik B. (2001) Extreme correlation of international equity markets. *Journal of Finance* 56:649–676.
- Lopez J. (1998) Methods for evaluating value-at-risk estimates. *Economic Policy Review, Federal Reserve Bank of New York October*:119–124.
- Loretan M., English W.B. (1999) Evaluating “correlation breakdowns” during periods of market volatility. Board of Governors of the Federal Reserve System, Working Paper.
- Loretan M., Phillips P.C.B. (1994) Testing the covariance stationarity of heavy-tailed time series. *Journal of Empirical Finance* 1:211–248.
- Lumsdaine R. (1996) Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH(1,1) and covariance stationary

- GARCH(1,1) models. *Econometrica* 64:575–596.
- Lux T. (2001) The limiting extremal behavior of speculative returns: an analysis of intra-daily data from the Frankfurt stock exchange. *Applied Financial Economics* 11:299–315.
- Madan, D.B., Carr P., Chang E. (1998) The variance gamma process and option pricing. *European Finance Review* 2:79–105.
- Madan D.B., Milne F. (1991) Option pricing with VG martingale components. *Mathematical Finance* 1:39–56
- Madan D.B., Milne F. (1994) Contingent claims valued and hedged by pricing and investing in a basis. *Mathematical Finance* 4:223–245.
- Madan D., Seneta E. (1990) The VG model for share market returns. *Journal of Business* 63:511–524.
- Maheu J.M., McCurdy T.H. (2004) News arrival, jump dynamics, and volatility components for individual stock returns. *Journal of Finance* 59:755–791.
- Malevergne Y., Sornette D. (2003) Testing the gaussian copula hypothesis for financial assets dependences. *Quantitative Finance* 3:231–250.
- Malz A.M. (1996) Using option prices to estimate realignment probabilities in the European monetary system. *Journal of International Money and Finance* 15:717–748.
- Malz A.M. (1997) Estimating the probability distribution of the future exchange rate from option prices. *Journal of Derivatives* 5:18–36.
- Mandelbrot B. (1963) The variation of certain speculative prices. *Journal of Business* 35:394–419.
- Markowitz H.M. (1952) Portfolio selection. *Journal of Finance* 7:77–91.
- Martens M., Poon S. (2001) Returns synchronization and daily correlation dynamics between international stock markets. *Journal of Banking and Finance* 25:1805–1827.
- McNeil A.J. (1997) Estimating the tails of loss severity distributions using extreme value theory. *ASTIN Bulletin* 27:117–137.
- McNeil A.J. (1998) Calculating quantile risk measures for financial time series using extreme value theory. ETH E-Collection (<http://e-collection.ethbib.ethz.ch/cgi-bin/show.pl?type=bericht&nrt=85>).
- McNeil A.J., Frey R (2000) Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance* 7:271–300.
- Mead L.R., Papanicolaou N. (1984) Maximum entropy in the problem of moments. *Journal of Mathematical Physics* 25:2404–2417.
- Meddahi N., Renault E. (2004) Temporal aggregation of volatility models. *Journal of Econometrics* 119:355–379.
- Melick W.R., Thomas C.P. (1997) Recovering an asset's implied PDF from options prices: An application to crude oil during the Gulf crisis. *Journal of Financial and Quantitative Analysis* 32:91–116.
- Melino A., Turnbull S. (1990) Pricing foreign currency options with stochastic volatility. *Journal of Econometrics* 45:239–265.

- Merton R. (1973) Theory of rational option pricing. *Bell Journal of Economics and Management Science* 4:141–183.
- Merton R. (1976) Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics* 3:125–144.
- Monroe I. (1978) Processes that can be embedded in brownian motion. *Annals of Probability* 6:42–56.
- Mood A.M., Graybill F.A., Boes D.C. (1974) *Introduction to the theory of statistics*. McGraw-Hill, New York.
- Morgan J.P. (1996) RiskMetrics – Technical document, New York.
- Neftci S.N. (1996) *An introduction to the mathematics of financial derivatives*. Academic Press, Boston.
- Nelsen R.B. (1999) *An introduction to copulas*. Springer-Verlag, New York.
- Nelson D.B. (1990) Stationarity and persistence in the GARCH(1,1) model. *Econometric Theory* 6:318–334.
- Nelson D.B. (1991) Conditional heteroskedasticity in asset returns: A new approach. *Econometrica* 59:347–370.
- Newey W.K. (1985) Generalized method of moments specification testing. *Journal of Econometrics* 29:229–256.
- Newey W.K., Steigerwald D.G. (1997) Asymptotic bias for quasi-maximum likelihood in conditional heteroskedasticity models. *Econometrica* 65:587–599.
- Nijman T.E., Palm F.C. (1990) Predictive accuracy gain from disaggregate sampling in ARIMA models. *Journal of Business and Economic Statistics* 8:405–415.
- Nijman T.E., Sentana E. (1996) Marginalization and contemporaneous aggregation in multivariate GARCH processes. *Journal of Econometrics* 71:71–87.
- O'Hara M. (1995) *Market microstructure theory*. Blackwell Publishers, Cambridge, Massachusetts.
- Ormoneit D., White H. (1999) An efficient algorithm to compute maximum entropy densities. *Econometric Reviews* 18:127–140.
- Owen J., Rabinovitch R. (1983) On the class of elliptical distributions and their applications to the theory of portfolio choice. *Journal of Finance* 38:745–752.
- Pagan A.R. (1986) Two stage and related estimators and their applications. *Review of Economic Studies* 53:517–538.
- Pagan A.R. (1996) The econometrics of financial markets. *Journal of Empirical Finance* 3:15–102.
- Pagan A.R., Schwert G.W. (1990) Alternative models for conditional stock volatility. *Journal of Econometrics* 45:267–290.
- Palm F.C. (1996) GARCH models of volatility. In: Maddala G.S., Rao C.R. (eds) *Handbook of statistics Vol. 14: Statistical methods in finance*. Elsevier Science, Amsterdam, 209–240.
- Pan J. (2002) The jump-risk premia implicit in options: Evidence from an integrated time-series study. *Journal of Financial Economics* 63:3–50.
- Parkinson, M. (1980) The extreme value method for estimating the variance

- of the rate of return. *Journal of Business* 53:61–65.
- Patton A.J. (2006) Estimation of copula models for time series of possibly different lengths. *Journal of Applied Econometrics* 21:147–173.
- Perry P. (1982) The time-variance relationship of security returns: Implications for the return-generating stochastic process. *Journal of Finance* 37:857–870.
- Pickands J. (1975) Statistical inference using extreme order statistics. *Annals of Statistics* 3:119–131.
- Plackett R.L. (1965) A class of bivariate distributions. *Journal of the American Statistical Association* 60:516–522.
- Poon S.H., Rockinger M., Tawn J. (2004) Extreme-value dependence in financial markets: Diagnostics, models and financial implications. *Review of Financial Studies* 17:581–610.
- Prakash A., Chang C., Pactwa T. (2003) Selecting a portfolio with skewness: Recent evidence from US, European, and Latin American equity markets. *Journal of Banking and Finance* 27:1375–1390.
- Pratt J.W., Zeckhauser R.J. (1987) Proper risk aversion. *Econometrica* 55:143–154.
- Premaratne G., Bera A.K. (2000) Modeling asymmetry and excess kurtosis in stock return data. University of Illinois, Illinois Research and Reference Working Paper 00-123.
- Prescott P., Walden A.T. (1980) Maximum-likelihood estimation of the parameters of the three parameter generalized extreme-value distribution. *Biometrika* 67:723–724.
- Press J. (1967) A compound events model for security prices. *Journal of Business* 40:317–335.
- Press W., Teukolsky S., Vetterling W., Flannery B. (1999) Numerical recipes in C, the art of scientific computation. Cambridge University Press, Cambridge, Massachusetts.
- Protter P. (2004) Stochastic integration and differential equations. Springer-Verlag, Berlin.
- Ramchand L., Susmel R. (1998) Volatility and cross correlation across major stock markets. *Journal of Empirical Finance* 5:397–416.
- Ratner M. (1992) Portfolio diversification and the inter-temporal stability of international indices. *Global Finance Journal* 3:67–78.
- Riboulet G., Roncalli T., Bouyé E. (2000) Copulas for finance. A reading guide and some applications. Crédit Lyonnais, Working Paper.
- Richardson M., Smith T. (1993) A test for multivariate normality in stock returns. *Journal of Business* 66:295–321.
- Richardson M., Smith T. (1994) A direct test of the mixture of distributions hypothesis: Measuring the daily flow of information. *Journal of Financial and Quantitative Analysis* 29:101–116.
- Rockafellar R.T., Uryasev S. (2000) Optimization of conditional value-at-risk. *Journal of Risk* 2:21–41.
- Rockafellar R.T., Uryasev S. (2002) Conditional value-at-risk for general loss

- distributions. *Journal of Banking and Finance* 26:1443–1471.
- Rockinger M., Jondeau E. (2002) Entropy densities with an application to autoregressive conditional skewness and kurtosis. *Journal of Econometrics* 102:119–142.
- Rockinger M., Semenova M. (2005) Estimation of jump-diffusion processes via empirical characteristic functions. FAME and HEC Lausanne, Working Paper.
- Rogers L.C.G., Williams D. (2000) Diffusions, Markov processes and martingales, Volumes 1 and 2, Cambridge University Press, Cambridge, Massachusetts.
- Roll R. (1984) A simple implicit measure of the effective bid-ask spread. *Journal of Finance* 39:1127–1139.
- Rootzén H., Tajvidi N. (1996) Extreme value statistics and wind storm losses: A case study. *Scandinavian Actuarial Journal* 1:70–94.
- Rosenberg J.V., Engle R.F. (2002) Empirical pricing kernel. *Journal of Financial Economics* 64:341–372.
- Rosenblatt M. (1952) Remarks on a multivariate transformation. *Annals of Mathematical Statistics* 23:470–472.
- Rubinstein M.E. (1973) The fundamental theorem of parameter-preference security valuation. *Journal of Financial and Quantitative Analysis* 8:61–69.
- Rubinstein M.E. (1994) Implied binomial trees. *Journal of Finance* 49:771–818.
- Ruiz E. (1994) Quasi-maximum likelihood estimation of stochastic volatility models. *Journal of Econometrics* 63:289–306.
- Sahu S.K., Dey D.K., Branco M.D. (2003) A new class of multivariate skew distributions with applications to bayesian regression models. *Canadian Journal of Statistics* 31:129–150.
- Samorodnitsky G., Taqqu M.S. (1994) Stable non-gaussian random processes: Stochastic models with infinite variance. Chapman and Hall, New York.
- Samuelson P. (1970) The fundamental approximation theorem of portfolio analysis in terms of means, variances and higher moments. *Review of Economic Studies* 37:537–542.
- Schönbucher P.J. (2003) Credit derivatives pricing models: Models, pricing and implementation. John Wiley, Chichester.
- Schoutens W. (2003) Levy processes in finance: Pricing financial derivatives. John Wiley, Chichester.
- Schweizer B., Sklar A. (1974) Operations on distribution functions not derivable from operations on random variables. *Studia Mathematica* 52:43–52.
- Schweizer B., Wolff E. (1981) On nonparametric measures of dependence for random variables. *Annals of Statistics* 9:879–885.
- Scott D. (1992) Multivariate density estimation. Theory, practice and visualization, Wiley, New York.
- Scott R.C., Horvath P.A. (1980) On the direction of preference for moments of higher order than the variance. *Journal of Finance* 35:915–919.
- Sentana E. (1995) Quadratic ARCH models. *Review of Economic Studies* 62:639–661.

- Sentana E., Fiorentini G. (2001) Identification, estimation and testing of conditionally heteroskedastic factor models. *Journal of Econometrics* 102:143–164.
- Shannon C.E. (1948) The mathematical theory of communication. *Bell System Technical Journal* 27:379–423 and 623–656.
- Shephard N. (1996) Statistical aspects of ARCH and stochastic volatility. In: Cox D.R., Hinkley D.V., Barndorff-Nielsen O.E. (eds) *Time series models in econometrics, finance and other fields*. Chapman & Hall, London, 1–67.
- Shih J.H., Louis T.A. (1995) Inferences on the association parameter in copula models for bivariate survival data. *Biometrics* 51:1384–1399.
- Shimko D. (1993) Bounds of probability. *RISK* 6:33–47.
- Shreve S.E. (2004) Stochastic calculus for finance II (continuous-time models). Springer-Verlag, New York.
- Silverman B.W. (1986) Density estimation for statistics and data analysis. Monographs on Statistics and Applied Probability 26, Chapman and Hall, New York.
- Simaan Y. (1993) Portfolio selection and asset pricing – Three-parameter framework. *Management Science* 39:568–577.
- Simkowitz M.A., Beedles W.L. (1978) Diversification in a three-moment world. *Journal of Financial and Quantitative Analysis* 13:927–941.
- Singleton K.J. (2001) Estimation of affine asset pricing models using the empirical characteristic function. *Journal of Econometrics* 102:111–141.
- Singleton J.C., Wingender J. (1986) Skewness persistence in common stock returns. *Journal of Financial and Quantitative Analysis* 21:335–341.
- Sklar A. (1959) Fonctions de répartition à n dimensions et leurs marges. *Publications de l’Institut Statistique de l’Université de Paris* 8:229–231.
- Smith R.L. (1985) Maximum likelihood estimation in a class of non-regular cases. *Biometrika* 72:67–90.
- Smith R.L. (1987) Estimating tails of probability distributions. *Annals of Statistics* 15:1174–1207.
- Söderlind P., Svensson L.E.O. (1997) New techniques to extract market expectations from financial instruments. *Journal of Monetary Economics* 40:383–429.
- Stărică C., Pictet O. (1999) The tales the tails of GARCH(1,1) processes tell. Dept. Statistics, University of Pennsylvania, Working Paper.
- Stein E., Stein J. (1991) Stock price distributions with stochastic volatility: An analytic approach. *Review of Financial Studies* 4:727–752.
- Stieltjes T.J. (1894) Recherches sur les fractions continues. *Annales de la Faculté des Sciences de Toulouse* 8:1–22.
- Stroud A.H. (1971) Approximate calculation of multiple integrals. Prentice-Hall, Englewood Cliffs, New Jersey.
- Stutzer M. (1996) A simple nonparametric approach to derivative security valuation. *Journal of Finance* 51:1633–1652.
- Susmel R., Engle R.F. (1994) Hourly volatility spillovers between international equity markets. *Journal of International Money and Finance* 13:3–25.

- Tapia R.A., Thomson J.R. (1978) Nonparametric probability density estimation. Johns Hopkins University Press, Baltimore.
- Tauchen G., Hussey R. (1991) Quadrature-based methods for obtaining approximate solutions to nonlinear asset pricing models. *Econometrica* 59:371–396.
- Tauchen G., Pitts M. (1983) The price variability-volume relationship on speculative markets. *Econometrica* 51:485–505.
- Taylor S.J. (1982) Financial returns modelled by the product of two stochastic processes, a study of daily sugar prices, 1961–1979. In: Anderson O.E. (ed.) *Time series analysis: Theory and practice*. North-Holland, Amsterdam, 203–226.
- Taylor S.J. (1986) Modelling financial time series. Wiley, New York.
- Theodossiou P. (1998) Financial data and the skewed generalized t distribution. *Management Science* 44:1650–1661.
- Tiago de Oliveira J. (1973) Statistical extremes – A survey. Center of Applied Mathematics, Lisbon.
- Timmermann A. (2000) Moments of Markov switching models. *Journal of Econometrics* 96:75–111.
- Tobin J. (1958) Liquidity preference as behavior towards risk. *Review of Economic Studies* 25:68–86.
- Tsay R.S. (2002) Analysis of financial time series, John Wiley, New York.
- Tse Y.K. (2000) A test for constant correlations in a multivariate GARCH model. *Journal of Econometrics* 98:107–127.
- Tse Y.K. (2002) Residual-based diagnostics for conditional heteroscedasticity models. *Econometrics Journal* 5:358–374.
- Tse Y.K., Tsui A.K.C. (2002) A multivariate generalized autoregressive conditional heteroscedasticity model with time-varying correlations. *Journal of Business and Economics Statistics* 20:351–362.
- Ushakov N.G. (1999) Selected topics in characteristic functions. Modern Probability and Statistics, VSP VB, Utrecht, The Netherlands.
- Van den Goorbergh R.W.J., Vlaar P.J.G. (1999) Value-at-risk analysis of stock returns: Historical simulation, variance techniques or tail index estimation? WO Research Memoranda 578, Netherlands Central Bank, Research Department.
- Van der Weide R. (2002) GO-GARCH: A multivariate generalized orthogonal GARCH model. *Journal of Applied Econometrics* 17:549–564.
- Vlaar P., Palm F. (1993) The message in weekly exchange rates in the European monetary system: Mean reversion, conditional heteroskedasticity and jumps. *Journal of Business and Economic Statistics* 11:351–360.
- Vrontos I.D., Dellaportas P., Politis D.N. (2003) A full-factor multivariate GARCH model. *Econometrics Journal* 6:312–334.
- Waggoner D.F. (1997) Spline methods for extracting interest rate curves from coupon bond prices. Federal Reserve Bank of Atlanta, Working Paper 97-10.
- Wahba G. (1990) Spline models for observational data. SIAM: Philadelphia.

- Wallis K.F. (2003) Chi-squared tests of interval and density forecasts, and the Bank of England's fan charts. *International Journal of Forecasting* 19:165–175.
- Weinberg S.A. (2001) Interpreting the volatility smile: An examination of the information content of option prices. Board of Governors of the Federal Reserve System, *International Finance Discussion Paper* 706.
- Weiss A.A. (1986) Asymptotic theory for ARCH models: Estimation and testing. *Econometric Theory* 2:107–131.
- Wheeler J.C., Gordon R.G. (1969) Rigorous bounds for thermodynamic properties of harmonic solids. *Journal of Chemical Physics* 51:5566–5583.
- White H. (1982) Maximum likelihood estimation of misspecified models. *Econometrica* 50:1–25.
- Widder D.V. (1946) The Laplace transform. Princeton University Press, Princeton.
- Wiggins J.B. (1987) Option values under stochastic volatility: Theory and empirical estimates. *Journal of Financial Economics* 19:351–372.
- Williams D. (1991) Probability with martingales. Cambridge University Press.
- Zakoïan J.-M. (1994) Threshold heteroskedasticity models. *Journal of Economic Dynamics and Control* 18:931–955.
- Zellner A., Highfield R.A. (1988) Calculation of maximum entropy distributions and approximation of marginal posterior distributions. *Journal of Econometrics* 37:195–209.
- Zivot E., Wang J. (2003) Modeling financial time series with S-Plus. Springer-Verlag, New York.

Index

- Adequacy test
 - Density forecast, 179, 239, 258
 - Interval forecast, 180
- Aggregation of GARCH model
 - Cross-sectional, 114
 - Temporal, 109
- ARCH model, 79
 - ARCH-in-mean model, 83
 - Definition, 81
 - Forecasting, 81
 - Kurtosis, 82
 - Stationarity conditions, 81
 - Tail index, 293
 - Test, 82
- Asymmetric distribution, 9, 16, 166
 - Hidden truncation, 166
 - Inverse scale factor, 167
- Asymptotic dependence, 303, 309
- AutoRegressive Conditional Density, 191
- Bayes' rule, 68, 103
- Bid-ask spread, 38, 39
- Bipower variation, 126
- Black-Scholes-Merton model, 366, 375, 384, 397, 425
- Broker, 35
- Brownian motion, 45, 451, 463, 501
 - Arithmetic, 370, 460
 - Definition, 453
 - Geometric, 371, 374, 433, 461
 - Properties, 456
 - Risk neutral, 378
- Calendar time, 45
- CAPM, 349, 353
- Cauchy distribution, 42, 269
- CAViaR, 329
- Central limit theorem, 40, 451
- Characteristic function, 396
 - Definition, 477
 - Exponential, 494
 - Jump process, 492, 493
 - Lévy process, 441, 444, 445, 501, 505
 - Model with jumps, 438, 440
 - Normal, 483
 - Properties, 478
 - Stable, 41
 - Stochastic volatility, 419, 427, 430
- Copula, 240
 - Adequacy test, 258
 - Archimedean, 248
 - Clayton, 250
 - Elliptical, 245
 - Estimation, 254
 - Extremes, 300
 - Families, 245
 - Frank, 249
 - Gaussian, 246, 304, 308
 - Gumbel, 252
 - Logistic, 308
 - Marshall-Olkin, 254
 - Non-parametric, 244
 - Plackett, 252
 - Student t , 248
 - Time-varying dependency, 259
 - VaR, 339

- Correlation
 - Conditional correlation, 206
 - Constant correlation, 206
 - Constant correlation test, 28, 30, 214
 - Dynamic correlation, 207
 - Jennrich test, 30
 - Kendall's tau, 243
 - Linear, 27
 - Measure of concordance, 243
 - Mixture of distributions hypothesis, 55
 - Pearson's rho, 27, 195, 243, 244, 300
 - Spearman's rho, 243
 - Stylized facts, 26
 - Time-variability, 259
 - Time-varying correlation, 208
 - z -transformation, 30
- Cross-correlation, 10
- Cumulants, 397
- Cumulative distribution function, 10
- Dealer, 34
- Density forecast, 179, 239, 258
- Dependence
 - Extreme, 300
 - Perfect dependence, 303
 - Perfect independence, 303
- Depth (market), 38
- Diffusion process, 459
- Dirac measure, 374, 468
- Directing process, 45, 51
- Discretization bias, 45
- Distribution, 10
- Domain of attraction, 267, 272
- Downside risk, 360
- Duration
 - Between jumps, 497
 - Between trades, 62
- Dynamic conditional correlation
 - Estimation, 210
 - Model, 207
- EGARCH model, 94, 96
- Elliptical distribution
 - Estimation, 238
 - Multivariate, 229
- Emerging markets, 268
- Entropy distribution, 172, 407
 - Domain of definition, 175
- Estimation, 173
- Estimation
 - Copula, 254
 - Entropy distribution, 173
 - GARCH model, 89
 - Generalized extreme value distribution, 274
 - Generalized Pareto distribution, 281
 - GMM, 19, 53
 - Maximum Likelihood, 89, 210, 236, 255, 274, 282
 - Multivariate GARCH model, 210, 236
 - Quasi Maximum Likelihood, 114, 117, 148, 154
 - Regression-based estimation, 275
 - Semi-parametric Maximum Likelihood, 257
 - Stochastic volatility, 117
 - Tail index, 281
 - Two-step, 210, 238, 256
- Event time, 45
- Excess distribution function, 278
- Existence of moments, 284
- Expected shortfall
 - Definition, 319
 - Extreme value theory, 324, 328
 - Historical simulation, 321
 - Portfolio-level approach, 335
 - RiskMetrics, 331
 - Skewed t distribution, 336
- Exponential distribution, 278, 496, 497
- Extrema approach
 - Generalized extreme value distribution, 268
 - Shape index, 269
 - Tail index, 269
- Extremal index, 291
- Extreme value theory
 - Expected shortfall, 324
 - Extrema approach, 266
 - Iid data, 266, 300
 - Multivariate, 300
 - Tail approach, 276
 - Univariate, 266
 - VaR, 324
 - Weakly dependent data, 291
- Factor GARCH model, 201
- Fast Fourier transform, 477, 484

- Fat tails, 9, 12, 16
- Financial markets, 34
- Fisher-Tippet theorem
 - Iid data, 267
 - Multivariate, 300
 - Weakly dependent data, 292
- Flexible GARCH model
 - Model, 204
 - VaR, 339
- Fréchet distribution, 267, 269
- GARCH model, 61, 79, 84
 - Asymmetric model, 94, 95, 209
 - Cox-Box transform, 95
 - Cross-sectional aggregation, 114
 - Definition, 84
 - EGARCH model, 94
 - Estimation, 89
 - Exponential GARCH model, 96
 - Extrema behavior, 269
 - Forecasting, 88
 - GJR model, 95, 96
 - Higher moments, 146, 147
 - Integrated GARCH model, 89
 - Jumps, 101
 - Kurtosis, 145
 - Multivariate model, 197
 - News impact curve, 94, 96
 - Non-parametric model, 96
 - Semi-parametric, 153
 - Semi-strong GARCH, 109
 - Skewness, 145
 - Stationarity conditions, 85
 - Strong GARCH, 109
 - Structure of model, 80
 - Tail index, 294
 - Temporal aggregation, 109
 - Test for asymmetry, 97
 - Test for GARCH effects, 92
 - TGARCH model, 95
 - VaR, 334
 - Weak GARCH, 109, 114
- Gaussian distribution
 - Multivariate, 223
- General error distribution, 148
- Generalized beta distribution, 395
- Generalized extreme value distribution, 265
 - Definition, 268
- Maximum Likelihood, 274
- Parameter estimates, 276
- Regression-based estimation, 275
- Generalized Pareto distribution, 266
 - Definition, 278
 - Graphical identification, 279
 - High threshold selection, 281
 - Maximum Likelihood, 282
 - Parameter estimates, 283
 - VaR, 324
- GJR model, 95, 96
- GMM, 19, 53, 58, 59
- Gram-Charlier expansion, 155
- Gumbel distribution, 267, 269
- Hazard function, 496
- Heterogeneity of beliefs, 51
- High frequency data, 118, 128
- High quantiles
 - Estimation, 296
 - Extrema approach, 297
 - Generalized Pareto distribution, 324
 - Semi-parametric approach, 300
 - Tail approach, 298
- Higher moments
 - Co-kurtosis, 31
 - Co-skewness, 31
 - Distributions, 153
 - Edgeworth expansion, 397
 - Existence, 151
 - GARCH model, 145
 - Hermite polynomials, 401
 - Modelling, 189
 - Multivariate, 31
 - Portfolio allocation, 353
 - Portfolio return, 32, 357
 - Skewed t distribution, 162, 233
 - Test, 189
 - Time variability, 26, 188
- Increasing process, 45
- Information arrival, 44, 47, 48, 61
- Integral, 457, 459, 460
- Integrated GARCH model, 89, 331
- Integrated quarticity, 141
- Interquartile range, 15
- Interval forecast
 - Adequacy test, 180
 - VaR, 343

- Ito process, 459, 460
- Ito's lemma, 372
 - Multivariate, 462
- Jumps, 119, 123, 497
 - Characteristic function, 438
 - GARCH model, 101
 - Higher moments, 104
 - Option pricing, 434, 438
- Kernel regression, 408
- Kolmogorov backward equation, 374, 464
- Kurtosis, 12, 15
 - ARCH model, 82
 - Edgeworth expansion, 397
 - Excess kurtosis, 14
 - GARCH model, 145–147
 - Hermite polynomials, 401
 - Jump model, 104
 - Mixture of distributions hypothesis, 54
 - Modeling, 189
 - Stochastic volatility, 116
- Lévy process, 478
 - Construction, 501
 - Model, 441
 - Option pricing, 445
 - Properties, 505
- Law of large numbers, 451
- Leverage effect, 25, 46, 94, 210
- Ljung-Box test, 23
- Log-normal distribution, 373
- Marginal distribution function, 303
- Market maker, 62
- Market microstructure, 33
 - Order-driven market, 34, 35
 - Quote-driven market, 34, 62
- Martingale
 - Local martingale, 124
 - Measure, 377
 - Semi-martingale, 125
 - SVSM, 125
- Maximum Likelihood
 - Copula, 255
 - Efficiency, 223
 - GARCH model, 89
- Generalized extreme value distribution, 274
- Generalized Pareto distribution, 282
- Multivariate GARCH model, 210, 236
- Mean, 14
- Mean absolute deviation, 15
- Mean excess function, 278, 281, 325
- Mean excess plot, 281
- Mean waiting time, 285
- Measures of dispersion, 14
- Measures of location, 14
- Median, 14
- Microstructure noise, 141
- Mixture of distribution hypothesis, 41, 45, 48
 - GMM, 53, 58
 - Moments of densities, 53
 - Testing, 58, 60
- Mixture variable, 51
- Moments, 10, 12
- Multi-period VaR
 - Portfolio-level approach, 336
 - RiskMetrics, 332
- Multivariate distribution, 195, 223
 - Copula function, 240
 - Elliptical distribution, 229
 - Gaussian distribution, 223
 - Skewed elliptical distribution, 230, 232
 - Skewed Gaussian distribution, 230
 - Skewed t distribution, 233
 - Student t distribution, 225
 - VaR, 337
- Multivariate GARCH model, 197
 - BEKK model, 199, 209
 - Conditional correlation, 206
 - Diagonal model, 198
 - Estimation, 210, 236
 - Factor GARCH model, 201
 - Flexible model, 204
 - GO-GARCH model, 203
 - O-GARCH model, 202
 - Semi-parametric estimation, 239
 - Specification test, 212
 - VaR, 337
 - Vectorial model, 198
- News impact curve, 94, 96
- Non-normal distribution, 153

- Asymmetric distribution, 166
- Edgeworth expansions, 395
- Entropy distribution, 172, 191, 407
- Generalized beta, 395
- Gram-Charlier expansion, 155
- Hermite polynomials, 399
- Higher moments, 147
- Hypergeometric function, 394
- Mixture of log-normals, 389
- Pearson IV distribution, 169
- Semi-parametric, 153
- Skewed t distribution, 160, 190, 233
- Stable, 41, 506
- Non-normal multivariate model
 - Adequacy test, 239
 - Estimation, 236
- Non-normality, 191
 - Distributions, 153
 - General problem, 144
 - Portfolio allocation, 349
 - Time-varying higher moments, 188
- Normal distribution, 10, 14, 40
 - Higher moments, 146
 - Test, 16
- Normality, 16
 - Aggregate returns, 9
 - Jarque-Bera test, 16
 - Kolmogorov-Smirnov test, 20
 - Lilliefors test, 20
 - Tests for normality, 16
- Numerical integration, 42, 47, 352, 485
- Option, 366
 - American, 367
 - At-the-money, 368
 - Boundary condition, 374
 - European, 367, 379
 - In-the-money, 368
 - Jump process, 438
 - Lévy process, 445
 - Moneyness, 368
 - Out-of-the-money, 368
- Partial differential equation, 371
- Pay-off, 367
- Put-call parity, 369
- Stochastic Volatility, 425, 426
- Strike price, 367
- Term structure of volatilities, 384
- Underlying asset, 367
- Volatility smile, 384, 402
- Order, 35, 36
- Order flow
 - Estimating parameters, 67
 - Modeling quotes, 68
 - Simulation, 71
- Order-driven market, 34, 35
- Orthogonal GARCH model, 202
- Pareto distribution, 269, 278
- Partial differential equation, 369, 371, 419, 422, 463, 466, 469
- Pay-off, 367, 379
- Pearson IV distribution, 169
 - Domain of definition, 169
 - Higher moments, 169
- Poisson process, 497
 - Compensated, 498
 - Compound, 488, 493, 501
 - Construction, 489
 - Higher moments, 492
 - Marked, 488
 - Order arrival, 73
 - Properties, 491
 - Pure, 487
- Portfolio allocation
 - Downside risk, 360
 - Expected shortfall, 319, 360
 - Expected utility, 350
 - Higher moments, 31, 350, 353
 - Mean-variance criterion, 349, 351
 - Optimal portfolio, 359
 - Under non-normality, 349
 - VaR, 317
- Portfolio liquidation, 73
- Power variation, 124
- Pricing kernel, 382, 414, 415
- Probability distribution function, 10
- Probability integral transform, 179, 239, 258
- Quadratic variation, 123
- Quantile plot, 20, 270
- Quantile regression, 328
- Quasi Maximum Likelihood, 114, 117, 148, 154
- Quote, 35
- Quote-driven market, 34, 62

- Random walk, 453, 455
- Realized covariance, 135
- Realized quarticity, 141
- Realized variance, 128, 134
- Realized volatility, 118, 128
- Return - definitions, 7
- Risk aversion, 351, 355, 356, 416
- Risk measure
 - Coherent measure, 318
 - Downside risk, 360
 - Expected shortfall, 319
 - VaR, 317
- Risk-neutral density, 383
- Edgeworth expansions, 395
- Estimation, 387
- Generalized beta, 395
- Hermite polynomials, 399
- Kernel regression, 408
- Maximum entropy principle, 407
- Mixture of hypergeometric functions, 394
- Mixture of log-normals, 389
- Non-parametric model, 384, 402
- Parametric model, 383, 389
- Semi-parametric model, 383, 395
- Spline methods, 402
- Tree-based methods, 406
- RiskMetrics, 315, 331
 - Multi-period VaR, 332
 - Multiple position, 332
- Semi-martingale, 124
- Serial correlation, 9
 - Ljung-Box test, 22, 213
- Mixture of distributions hypothesis, 56
- Skewed elliptical distribution
 - Gaussian, 230
 - Multivariate, 230, 232
- Skewed t distribution, 160, 191
 - Alternative specifications, 164
 - Domain of definition, 163
 - Expected shortfall, 336
 - Higher moments, 161
 - Multivariate, 233
 - VaR, 335
- Skewness, 12, 15
 - Edgeworth expansion, 397
 - GARCH model, 145–147
- Hermite polynomials, 401
- Jump model, 104
- Mixture of distributions hypothesis, 54
- Modelling, 189
- Stochastic volatility, 116
- Software, 92
- Specification test
 - Multivariate model, 212
 - Portmanteau test, 213
 - Residual-based test, 212
 - Univariate model, 177
- Spline method, 402
- Stable distribution, 39–41, 478, 506
 - Cauchy distribution, 42
- Standard deviation, 15
- Stochastic differential equation, 369, 374, 460
- Stochastic discount factor, 415
- Stochastic volatility, 46, 115, 417
 - Estimation, 117
 - Kurtosis, 116
 - Option pricing, 425, 438, 445
 - Quadratic variation, 123
 - Skewness, 116
- Student t distribution, 47, 269, 362
 - Higher moments, 147
 - Multivariate, 225
- Stylized facts of asset returns, 9
- Subordinated process, 45, 46, 51
- Tail index, 265
- ARCH model, 293
- Deckers, Einmahl, and de Haan estimator, 286
- Estimation, 281
 - Extrema approach, 269
 - for GARCH residuals, 296
 - GARCH model, 293
 - High threshold selection, 281, 287
 - Hill estimator, 286
 - Maximum Likelihood, 282
 - Multivariate, 307
 - Pickands estimator, 285
 - Semi-parametric estimation, 285
 - Weakly dependent data, 293
- Taylor series expansion, 354, 460
- Term structure of volatilities, 384
- Test

- Adequacy test, 179, 180, 239, 258
- ARCH effects, 82
- Asymmetric GARCH effects, 97
- Autoregressive higher moments, 189
- Constant correlation test, 214
- GARCH effects, 92
- Information matrix test, 214
- Jarque-Bera test, 16
- Kolmogorov-Smirnov test, 20
- LM test, 98, 216
- Specification test, 177, 212
- TGARCH model, 95
- Threshold selection
 - Bootstrap, 289
 - Generalized Pareto distribution, 281
 - Semi-parametric approach, 287
- Tick size, 38
- Time, 45
- Time dependency
 - Serial correlation, 9, 21, 22
 - Volatility clustering, 10, 23
- Time-varying conditional correlation, 208
- Time-varying higher moments, 189
- Trade flow, 63
- Trader, 39, 63
- Tree-based methods, 406
- Two-step estimation
 - Copula, 256
 - Dynamic conditional correlation, 210
- Utility function, 350, 354
 - CARA utility, 351, 354
 - CRRA utility, 355
 - Preferences, 356
- Value at Risk
 - Asset-level approach, 337
 - CAViaR, 329
 - Copula, 339
 - Definition, 317
 - Delta-only method, 341
 - Extreme value theory, 324, 326, 328
 - Historical simulation, 321
 - Interval forecast, 343
 - Multi-period, 318
 - Multiple position, 333
 - Multivariate distribution, 337
 - Non-linear approach, 341
 - Portfolio-level approach, 334
 - RiskMetrics, 331, 333
 - Skewed t distribution, 335
- Variance, 14
- Volatility, 15, 23
 - Asymmetry, 25
 - Leverage effect, 25
 - Long memory, 24
 - Unit root, 24
- Volatility clustering, 10, 23
 - GARCH model, 84
 - Mixture of distributions hypothesis, 57
 - Test for asymmetry, 97
- Volatility smile, 384
- Volume, 49, 52, 53, 134
 - Alternative measures, 62
 - Filtering for seasonality, 57
- Weibull distribution, 267, 269, 278
- Wiener process (see Brownian motion), 456