A Practical Guide to Quantitative Volatility Trading

Daniel Bloch

15th of January 2016

A Practical Guide to Quantitative Volatility Trading

Daniel BLOCH [1]
QUANT FINANCE LTD
Working Paper

15th of January 2016
Version 1.2.0

[1] db@quantfin.eu

Abstract

Financial time series exhibit multifractal scaling behaviour indicating a complex behaviour with long-range time correlations manifested on different intrinsic time scales. Such a behaviour typically points to the presence of recurrent economic cycles, crises, large fluctuations, and other nonlinear phenomena. We review quantitative volatility trading in classical economics before discussing some necessary modifications needed to account for multifractality in inefficient markets. We then present an arbitrage-free model of implied volatility surface, which is robust, easy to implement and computationally fast, enabling for systematic volatility trading. We consider risk management and discuss some applications on variance swaps and dispersion trading.

Contents

0.1 Introduction

In general, uncertainty related to traditional economic variables (interest rates and foreign exchange), as well as new variables (credit, commodity, weather etc), imposing costs and risks on the society, must be hedged away. Thus, financial markets developed, where derivative products were created to transfer risk between different economic agents. Both parties entering the deal must agree on the price of the contingent claim, even though they have different risks. The decisions for the appropriate pricing of such claims are made contingent on the price behaviour of the underlying securities. The uncertainty affecting the underlying is modelled by considering future trajectories of the risky asset seen as possible scenarios. Thus, it is important to understand the properties of market returns in order to devise their dynamics.

Given a sequence of independent random variables $X_1, X_2, ..., X_n$, and assuming that the random variables are small, Bernoulli [1713] studied the properties of the nomalised sum $\frac{S_n}{n}$ with $S_n = \sum_{i=1}^n X_i$ in the special case where the X_i takes value 1 with probability p and value 0 with probability $1-p$. Computing the probability of the normalised sum in the limit $n \to \infty$, Bernoulli's theorem is a form of the law of large numbers. Later, Laplace [1774] and Gauss [1809] associated the error distribution with the scheme of summation of small and independent random variables. They assumed that the mathematical expectation $a = E[X_i]$ and the variance $\sigma^2 = Var(X_i)$ of these variables are finite, and constructed the corresponding sequence of the normalised sums $Z_1, Z_2, ...$ where $Z_n = \frac{S_n - na}{\sigma\sqrt{n}}$. They computed the probability of the nomalised variable Z_n in the limit $n \to \infty$, obtaining the central limit theorem (CLT). It states that the normalised sum of a number of i.i.d. random variables with finite variances will tend to a normal distribution as the number of variables grows. Dropping the assumption of finite variance, the CLT was generalised to use not only the normal law as limiting approximation, but the stable law (see Gnedenko [1943] and Gnedenko et al. [1954]).

Assuming stock prices to be the sum of many small terms, stable models have been considered to describe financial systems. For instance, Mandelbrot [1963] described the financial market as a system with fat tails, stable distributions and persistence. The argument of large data sets exhibiting heavy tails and skewness combined with the generalised CLT theorem was put together to justify the use of stables models. However, it was largely ignored by Fama [1965a] (Fa65) [1970] and Sharpe [1970] who followed Bachelier's assumption of normally distributed returns (see Bachelier [1900]) to formalise the weak form efficient market where the price changes are independent and may be a random walk. The justification for this assumption is provided by the simplicity of the central limit theorem and the fact that variance and covariance are the only measures of risk. A direct consequence of the efficient market hypothesis (EMH) is that the most important concepts in theoretical and empirical finance developed around the assumption that asset returns follow a normal distribution. This includes the classical portfolio theory, the Black-Scholes-Merton option pricing model and the RiskMetrics variance-covariance approach to Value at Risk (VaR).

Using the concepts of self-similarity, scaling, and fractional processes to assess market returns, various authors found that long-range dependence (LRD) was subject to debate for raw (signed) returns, but was plainly visible in absolute returns, square returns, or any other measure of the extent of fluctuations. Thus, financial models developed to capture the thick tails and long-memory volatility persistence, where returns themselves have no autocorrelation, but their amplitudes have LRD (FIGARCH, MMAR). However, these statistical approaches are based on the moment properties of stochastic processes and must be restricted to second-order stationary processes. In addition, fractal properties are only verified in the infinitesimal limit, making LRD very difficult to be measured on sample data. Further, when the time series of asset returns possess the two features of heavy tails and LRD, most statistical tests fail to work. Yet, stock returns and FX returns are complex systems suffering from systematic effects mainly due to the periodicity of human activities, and can not be considered as processes with stationary increments. Once methods for the multifractal characterisation of non-stationary series developed (MFDFA, WTMM, GMWP), academics found correlations present in systems and highlighted the multifractal nature of financial time series. Further, using the effective local Holder exponents (ELHE) to detect and localise outliers in financial time series, some academics showed that these series exhibit stochastic Hurst exponent with characteristics of abrupt changes in the fractal structure. Such behaviour typically points to the presence of recurrent economic cycles, crises, large fluctuations, and other nonlinear phenomena. As a result, the multifractal nature of financial markets contradict the EMH, and thus the Black-Scholes pricing model and its associated assumptions. Nonetheless, it seems that the academics who developed the option pricing theory (OPT) focused on the empirical results that demonstrated thick tails and long-memory volatility persistence exhibited in the financial time series, but did not account for the evidence of long memory in raw returns. Such models of uncorrelated returns (white spectrum) and semimartingale prices are consistent with economic equilibrium and the EMH of Samuelson [1965]. For example, Engle [1982] proposed to model volatility as conditional upon its previous level, that is, high volatility levels are followed by more high volatility, while low volatility is followed by more low volatility. This led jump-diffusion models to develop in continuous time, such as the Merton model (see Merton [1976]), the Heston model (see Heston [1993]), or a combination of both.

One consequence of the weak efficient market hypothesis (EMH) (Fa65) is that statistical approaches must be restricted to second-order stationary processes. In that setting, variance is finite and measures the risk of a single underlying asset, while the portfolio risk becomes a weighted sum of covariation of all stocks in the portfolio. These statistics are easy to understand and to estimate. As a result, volatility trading via vanilla options developed to hedge away uncertainty related to economic variables. In the 1990s volatility became an asset class of its own, and new products such as volatility swaps and variance swaps were created. When measuring portfolio risk, Value at Risk (VaR) is the risk expressed in dollar terms showing what amount of money your portfolio could lose during a defined interval with a given probability. It is commonly understood that VaR is in essence the volatility of a portfolio expressed in dollar terms. This concept has been applied not only to estimate the risks of a portfolio of assets, but also for trading the market itself, since the market is considered as one large portfolio of various assets each having a weight in the global market place. As a result, relative value trading developed as a way to profit from the mean reversion of two related assets that have diverged. Trades are usually chosen on correlated assets. A pair trade can be carried out via straddles / strangles or variance swaps. As the volatility of indices tends to trade significantly less than its constituents, traders tried to profit from this anomaly by either selling correlation swaps, or through dispersion trading. In dispersion trading, a short index volatility position is traded against a basket of long single stock volatility positions. Such a long dispersion trade is short implied correlation. Considering a probability formula and assuming identical correlation between each stock, the index correlation was estimated as the variance of the index divided by the weighted average single stock variance. Various financial products, developed around the idea of buying or selling correlation, were sold in the market by investment banks, such as worst-of / best-of, Altiplano and other Himalayas products. The vendors, being short implied correlation, put buying pressure to lift the implied correlation above fair value. Since dispersion trading is not a pure correlation trade, correlation swap developed as a swap between the average pairwise correlation of all members of an index and a fixed amount determined at inception. Basket options, with fixed weighting of the members over time, have been proposed, leading to covariance risk. Thus, covariance swaps paying out the correlation multiplied by the volatility of the two assets developed.

Even though the Black-Scholes formula assumes constant volatility, market practice consists in using one BS-model for every pair (K,T), leading to the generation of a non-flat implied volatility (IV) surface. Acknowledging the existence of non-Gaussian market returns and the need to consider higher moments, practitioners had to tweak the Black-Scholes formula. This approach being a proxy for accounting for stochastic variations of the volatility surface, traders use the wrong volatility parameter into the wrong formula. They must therefore consider the sensitivity of option prices to changes in the IV surface. When the volatility σ_t changes with a certain volatility of its own, the sensitivity of the option price with respect to the volatility is called Vega. Hence, the vega risk is associated with unexpected changes in volatility and should be accounted for. Theoretically, the BS-formula does not apply to stochastic volatility, although the explicit BS-vega calculation is a useful approximation to the actual vega over a small time horizon. More importantly, the profit and loss (PnL) of financial products on the variance and covariance of asset returns are fully exposed to moves in volatility of volatility terms, leading to new measures of risk such as the Volga (second derivative of the price with respect to volatility) and the Vanna (cross derivative of the price with respect to the underlying asset and the volatility). Ironically, the only possible forms of market risk within the weak form efficient market, namely the variance and covariance, are themselves stochastic. Going further, we know that standard time series models, including the $GARCH(1,1)$ process and its squares, the stochastic volatility models and solutions to stochastic recurrence equations converge in distribution to an infinite variance stable distribution (see Bartkiewicz et al. [2011]). In the presence of dependence, the conditions for the convergence depend on regular variation of the stationary sequence with index $\alpha \in (0,2)$ plus some other conditions. This would explain the difficulties encountered by practitioners to estimate the volatility of volatility parameters of stochastic volatility models. Since the absence of long range dependence (LRD) in returns is still compatible with its presence in absolute returns, models accounting for LRD in volatility developed. However, recent studies on asset returns highlighted their non-stationary scaling properties (stochastic Hurst exponent) with characteristics of abrupt changes in the fractal structure. Thus, financial time series exhibit multifractal scaling behaviour indicating a complex behaviour with long-range time correlations manifested on different intrinsic time scales. Such a behaviour typically points to the presence of recurrent economic cycles, crises, large fluctuations, and other nonlinear phenomena which must be included in the pricing and risk management of derivative products. These phenomena can easily be observed in the market via the existence of an implied volatility surface for single stock options and an implied correlation matrix for multi-asset options. Moreover, since the variance and covariance of financial time series are not necessarily defined, their historical dynamics are stochastic with erratic jumps. Thus, in order to account for these dynamics when pricing path-dependent options, practitioners and academics focused on modelling the variance and covariace with jump-diffusion processes. Variance and covariance should not be considered as measures of risk and the valuation of variance swaps and other correlation swaps should be revised.

Part I

Quantitative trading in classical economics

Chapter 1

Introducing option pricing and risk management

1.1 Some theory on option pricing

1.1.1 Some terminologies

1.1.1.1 A structured market

In the 1970s, an increase in the uncertainty associated with a number of economical variables resulted in the birth of a structured financial products market, within which new financial instruments were created to transfer risk between different economic agents. Further, some of these agents, such as banks, were ready to take on additional risks in exchange for higher returns. As a results, investors could specialise on specific sectors of the market and diversify their risk. Just as with the uncertainty related to traditional economic variables like interest rates and foreign exchange rates, new variables (credit, commodity, weather etc.) imposed costs and risks on the society which necessitated to be hedged away. Derivatives lead to the immediate disbursement of funds when certain conditions are met as, unlike most insurance schemes, they do not require an assessment of the actual loss incurred. As an example, an organisation such as a bank can play an intermediary role between two parties. The party buying protection against the variations of an economical variable limits its risk, should a relatively adverse outcome materialise, while the seller's risk exposure increases along with the severity of the adverse variable outcome. Before 1973, all option contracts were individually negotiated by a broker on behalf of two clients, one being the buyer and the other one the seller. This transaction is called over-the-counter (OTC). Trading on an official exchange began in 1973 with listed options on the Chicago Board Options Exchange (CBOE). The listing of options on an exchange market dramatically decreased the cost of setting up an option contract due to the increased competition. Options on stocks, indices, futures, government bonds, commodities, currencies are traded on all of the world's major exchanges. Most options are registered and settled via a clearing house, which is a central body responsible for the collection of margin from the writers of options. The margin is a quantity of money held by the clearing house on behalf of the writer to guarantee that he is able to meet his obligations should the asset price move against him.

1.1.1.2 Describing a few market products

We are going to define some of the financial terms used on option pricing theory (for more details see Merton [1973], Hull [2012]). A derivative, or contingent claim, is a financial instrument whose value depends on the values of other more basic underlying variables being the prices of traded securities. One of the simplest and most commonly traded instrument in the market is the forward contract. A forward contract is an agreement to buy or sell an asset at a certain future time for a certain price called the delivery price. At the time the contract is entered into, the delivery price is such that the value of the forward contract is zero to both parties, so that neither party pays any money to obtain the contract. In general the payoff at time T from a long position in a forward contract on one unit of an asset is

$$S_T - K$$

where K is the delivery price and S_T is the spot price of the asset at maturity of the contract. Similarly, the payoff from a short position is $K - S_T$. The main purpose of forward contracts is to reduce risk since the holder can reduce future uncertainty. For example, a company facing the need to make a large fixed payment in foreign currency at a fixed future date can lock in the exchange rate now by buying a forward contract. The seller of such a contract is a speculator who would make a profit if the exchange rate remain below the value fixed at inception. Note, the forward contract entails no cost now, and only potential cost in the future.

An option is a security giving the right to buy or sell an asset, subject to certain conditions, within a specified period of time. Options, or derivatives, on stock prices were first traded on an organised exchange in 1973. Since then there has been a dramatic growth in terms of volumes of options traded in many exchanges throughout the world. Factors affecting option prices are the spot price S, the strike price K, the time to maturity T, the volatility of the underlying σ, and the risk-free interest rate r. The most common underlying assets include stocks, indices, foreign currencies, debt and commodities. More recently a new type of underlying assets was introduced which is link to the realised variance or volatility of the underlying such as the variance swap or the volatility swap.

European, or vanilla, options such as a call option give the holder the right to buy the underlying asset by a certain date called the maturity for a specified price called the exercise price or strike price, while the put option gives him the right to sell it. An American call or put option is as its European counter part except that it can be exercised at any time up to the expiration date, while a Bermudan option is exercisable only on certain specified days during its life. Contrary to the forward contract defined above, the holder of an option has the right to exercise the option but it is not an obligation. However, whereas it costs nothing to enter into a forward contract, there is a cost to entering into an option contract. We let K be the strike price and S_T be the underlying price at time T, then the payoff from a long position in a European call option is

$$\max(S_T - K, 0) = (S_T - K)^+$$

so that the option will be exercised if $S_T > K$. The payoff of a short position in the European call option is $-(S_T - K)^+$. Hence, the stock price is an upper bound to the call price. Also, a lower bound for the call option price on a non-dividend paying stock is $S_t - Ke^{-r(T-t)}$, where r is the risk-free rate. The payoff from a long position in a European put option is

$$\max(K - S_T, 0) = (K - S_T)^+$$

so that the option will be exercised if $S_T < K$ and the option can never be worth more than K while the payoff of a short position in the European put option is $-(K - S_T)^+$. Also, a lower bound for the put option price on a non-dividend paying stock is $Ke^{-r(T-t)} - S_t$. Stoll [1969] first noted the relation between the value of a call option and that of a put option. The put-call parity is a relationship at a given time t between the price of a European call option $C(t, K)$ and a European put option $P(t, K)$ given by

$$C(t, K) - P(t, K) = S_t - Ke^{-r(T-t)}$$

but it does not hold for American options. Hence, if an investor were to buy a call and sell a put option, his returns would be exactly the same as if he bought the stock on margin, borrowing $Ke^{-r(T-t)}$ toward the price of the stock.

The intrinsic value at time t of a call option is the maximum of zero and the value it would have if it was exercised immediately, that is $(S_t - K)^+$. The option is then said to have time value. So, one can identify the total value of the option at time t as the sum of its intrinsic value and its time value. Options are referred to as in-the-money (ITM), at-the-money (ATM), or, out-the-money (OTM). At a given time t, an in-the-money call option would lead to a positive cash flow to the holder if it was exercised immediately, that is $S_t > K$. Similarly, an at-the-money option would lead to zero cash flow if it was exercised, that is $S_t = K$, while an out-the-money call option would lead to negative cash flow, that is $S_t < K$. OTM options have a high risk of expiring worthless, but they tend to be relatively inexpensive. As the time value approaches zero at expiration, OTM options have a greater potential for total loss if the underlying stock moves in an adverse direction. Options that are deep ITM generally trade at or near their actual intrinsic values.

1.1.1.3 Some properties

We now consider the impacts of the factors affecting the call and put option price by changing one of them while keeping the other one fixed. The call option price becomes more valuable as the stock price increases and less valuable as the strike price increases. Similarly, the higher the spot price at the expiration, the lower the payoff on the put option. When the expiration date of the option is far in the future, the price of a bond paying the exercise price on the maturity date will be very low, and the value of the option will be approximately equal to the price of the stock. However, for very near expiration date, the value of the option will be approximately equal to the stock price minus the exercise price, or zero, if the stock price is less than the exercise price. Normally, the value of an option declines as its maturity date approaches, if the value of the stock does not change. Now, we consider the volatility of the stock price as a measure of its uncertainty in the future. As volatility increases the probability of a high or low level of the stock price increases so that the owner of a call option will benefit from a rise in the price, while his risk is limited when prices decrease. Therefore, the value of a call option increases with increasing volatility. To conclude, as the interest rates increase the expected growth rate of the stock price increases but the present value of any future cash flows received by the owner of the option decrease. In general, the market sees the price of calls as increasing with increasing rates, while put prices decrease. In the equity market, the holder of a call option is foregoing the dividends while the holder of a put option is receiving dividends.

1.1.1.4 American options

Merton [1973] showed that the value of a call option is always greater than the value it would have if it were exercised immediately. Note, an in-the-money American option must be worth at least as much as its intrinsic value. This is because when the stock price is much greater than the exercise price, the call option is almost sure to be exercised, and the current value of the option will thus be approximately equal to the price of the stock minus the price of a pure discount bond maturing on the same date as the option, with a face value equal to the strike price of the option. However, if the price of the stock is much less than the exercise price, the option will almost surely expire without being exercised, and its value will be near zero. Therefore, it can be optimal for the holder of such option to wait rather than exercise it at time t. Merton [1973] showed that the value of an American put option will be greater than that of a European put option because it is sometimes advantageous to exercise a put option before maturity, if it is possible to do so. If the stock price falls almost to zero and the probability that the price will exceed the exercise price before the option expires is negligible, then it will pay to exercise the option immediately. Thus, the investor gains the interest on the exercise price for the period up to the time he would otherwise have exercised it. Relaxing the assumption that stocks pays no-dividend, we get into more complicated problems, as under certain conditions it will be profitable to exercise an American call option before maturity. Merton [1973] showed that it was true only just before the stock's ex-dividend date. Although American call option becomes more valuable as the time to expiration increases since the holder can exercise at any time, the European call is not necessarily more valuable as time to expiration increases, since the holder can only exercise at maturity and a dividend between two maturities can make the second option less valuable than the first one.

1.1.1.5 Introducing volatility

Statistical or historical volatility is the volatility shown by the price of the underlying security in the past. The most common way of computing the volatility is through the sample standard deviation over a specified period

$$\widehat{\sigma}_n = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

where $x_i = \log \frac{S_i}{S_{i-1}}$ is the logarithm return over one business day and $\overline{x}$ is the arithmetic average of all log-returns. In the Black-Scholes formula, the returns $R_{t_i} = \frac{S_{t_i}}{S_{t_{i-1}}}$ for $t_i = ih$, $i = 1,..,n$ for some time step $h > 0$ are i.i.d. distributed with $N((\mu - \frac{\sigma^2}{2})h, \sigma^2 h)$. Hence, the sample standard deviation $\widehat{\sigma}_n$ corresponds to the maximum likelihood estimator for the parameter σ. One can then plug in the historical volatility to the BS-formula to compute an estimate of the option price. However, this approach fit very poorly the observed market prices. Since demand and supply determines the prices of liquid market options, one can use these prices to infer an implied volatility. Contrary to the historical volatility, the calculation of the market implied volatility requires an option pricing model. For pricing simple options on a single underlying the financial world uses the Black-Scholes model, which leads to a closed form solution since the stock price at a fixed time follows a lognormal distribution. In the Black-Scholes model all parameters are observed except for the volatility parameter σ. Hence, for practical reasons the BS-formula is used to infer the implied volatility. However, stock prices jump on occasions and do not always move in the smooth manner predicted by the geometric Brownian motion (GBM) model. Further, stock prices tend to have fatter tails than those predicted by the GBM. If the BS-model was correct, we should infer a flat implied volatility surface which is not the case in practice. In addition, the volatility surface varies significantly with time. The volatility skew is a pattern of the implied volatility among different strike prices, while the term structure of volatility is a pattern of the implied volatility with varying time to maturity. The slope of the skew is often regarded as a measure of risk aversion. The implied volatility reflects the market expectation of the stock's future level of volatility.

1.1.2 The pricing assumptions

When risk-managing contingent claims, both parties entering the deal must agree on the price of an option even though they have different risks. For instance, the buyer of a call option has a risk limited to the option premium, while the risk for the seller of the call option increases as the underlying spot increases. Moreover, the maturity risk plays against him and he can be exposed to an upward jump of the spot price just before maturity. As a result, the seller of the claim requires a compensation for the risk he is bearing, and the buyer must pay a premium for the benefit he gets from the contract. The decisions for the appropriate pricing of such claims are made contingent on the price behaviour of the underlying securities. Fortunately, the uncertainty affecting the underlying of the contract at maturity results from observed small daily movements, thus providing information that can be exploited. This information allows a model for the dynamics of the underlying to be defined, and a dynamic adjustment to be made against final risk. Consequently, financial markets model uncertainty by considering future trajectories of the risky asset seen as possible scenarios which are generally assumed to be continuous functions (ω_t) defined on $\mathbb{R}^+$. Different risk arguments can then be applied to determine the option price.

We assume that one unit of the underlying is expressed in dollar and that the spot price, or index, is either quoted in the market or comes from an independent publication representing a consensus between the different parties. The price of liquid options is determined by supply and demand on the market within a bid-ask spread. There exists various pricing methods among which some are based on hedging arguments, on the law of large numbers, or, as actuaries know it, on the standard deviation principle [1] to name a few. In a probabilistic approach, we consider the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ where $\mathcal{F}_t$ is a right continuous filtration including all $\mathbb{P}$ negligible sets in $\mathcal{F}$. Using the concept of absence of arbitrage opportunities (AAO) (see Harrison and Kreps [1979]), asset returns must reflect the fact that the riskier the asset the higher the returns, resulting in an instantaneous drift higher than the risk-free rate. Hence, for the market to be risk neutral all the tradable assets must share the same returns equal to the risk-free rate. As a result, in an arbitrage-free market, the assumption of linearity of prices leads to the existence of a risk-neutral probability measure

[1] $P(X) = E[X] + \lambda\sigma[X]$ where $\lambda\sigma[X]$ corresponds to a risk premium.

$\mathbb{Q}$ equivalent to the historical measure $\mathbb{P}$. Since our underlying is exchanged in a market, its discounted price needs to be a martingale under the risk-neutral measure $\mathbb{Q}$ (see Delbaen et al. [2004]). For an event $A \in \mathcal{F}$, while $\mathbb{P}(A)$ represents its probability of occurrence, $\mathbb{Q}(A)$ represents the value of an option with terminal payoff equal to $\frac{1}{P(t,T)}$ if A occurs and zero otherwise. Therefore, the risk-neutral measure represents a market consensus on the values of derivative instruments rather than an objective description of the evolution of the market (see Cont [2005]).

We let $\mathcal{Q}$ be the set of coexistent equivalent measures $\mathbb{Q}$. In a complete market the risk-neutral probability is unique, but it is not the case in an incomplete market where it must be specified. That is, in an incomplete market, even if $\mathbb{P}$ is known with certainty, we still face uncertainty in the choice of the pricing model $\mathbb{Q}$. In that sense, the pricing of contingent claims is very sensitive to the assumptions made and the choice of a model for the underlying process. Financial markets consists of benchmark instruments with payoff $(H_i)_{i \in I}$, where observed market prices $(P_i^*)_{i \in I}$ are in the range $P_i^* \in [P_i^b, P_i^a]$, corresponding respectively to the market bid and ask prices. Thus, we consider a set of arbitrage free model $\mathcal{Q}$ such that the discounted asset price $(\overline{S}_t)_{t \in [0,T]}$ is a martingale under each $\mathbb{Q} \in \mathcal{Q}$ with respect to its own history $\mathcal{F}_t$ and

$$\forall \mathbb{Q} \in \mathcal{Q} \, , \forall i \in I \, , E^{\mathbb{Q}}[|H_i|] < \infty \, , E^{\mathbb{Q}}[H_i] \in [P_i^b, P_i^a]$$

Hence, multiple risk-neutral distributions can fit the option prices, so that one needs some additional criteria to generate a unique probability distribution function (pdf). For instance, Le [2014] let that extra source of risk to be epistemic, and defined the risk-subjective measure. Alternatively, we can use indifference pricing as a method directly depending on the market counterparts and not on some market probablities (see Barrieu et al. [2009]).

In general, in an incomplete market we can choose any equivalent martingale measure as a self-consistent pricing rule, but option prices no-longer correspond to the cost of a specific hedging strategy. That is, the expected value of the option is purely mathematical and does not relate to its future price in the sense of replication (see Ayache et al. [2004]). One solution is to work in a risk-neutral framework by adding a market price of risk to the underlying dynamics. For simplicity of exposition, given a market price of risk λ, we will assume that there exist an equivalent martingale measure $\mathbb{Q}^\lambda$ denoted by $\mathbb{Q}$. As a result of absence of arbitrage opportunities (AAO), contingent claims can be valued by taking expectation of their discounted payoffs under the risk-neutral measure. Therefore, the price of a European contingent claim $C(t,x)$ on $]0,T] \times]0,+\infty[$ under the risk-neutral measure $\mathbb{Q}$ is

$$C(t, X_t) = E^Q[e^{-\int_t^T r_s ds} h(X_T) | \mathcal{F}_t] \tag{1.1.1}$$

where X_T is a $\mathbb{Q}$-random variable and h is a sufficiently smooth payoff function. In that setting, the difficulty comes from the meaning that one gives to the market price of risk λ as it depends on the preference of investors. Therefore, one need to know where to observe it and how to calibrate it. In the special case where the market price of risk is set to zero we recover the so-called historical option price.

1.1.3 The replication portfolio

In a complete market, every contingent claim can be replicated with a trading strategy, or replicating portfolio, made of the underlying asset X and a risk-free money account. For instance, Black and Scholes derived the price of an option by constructing a replicating portfolio (see details in Section (1.2.2)). As discussed above, this is not the case in an incomplete market. We are briefly going to introduce the theory of portfolio replication in complete and incomplete markets (details can be found in Oksendal [1998], Shreve [2004], Cont et al. [2003]). We let $(X(t))_{t \in [0,T]}$ be a continuous semimartingale on the horizon $[0,T]$, representing the price process of a risky asset. We let a contingent claim be given by $H = h(X(T))$ for some function h, where H is a random variable such that $H \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$. We let $M(t)$ be the risk-free money account, $\alpha(t)$ is the amount of asset held at time t, and $\beta(t)$ is the money account held at time t.

1.1.3.1 Complete market

In a complete market, the value of a portfolio at time t satifies

$$V(t) = \alpha(t)X(t) + \beta(t)M(t) \, , 0 \leq t \leq T$$

For simplicity of exposition we set $M(t) = 1$ for all $0 \leq t \leq T$. The trading strategy (α, β) is admissible, such that the value process $V(t)$ is square-integrable and have right-continuous paths defined by

$$V(t) = V_0 + \int_0^t \alpha(s)dX(s)$$

For $\mathbb{Q}$-almost surely, every contingent claim H is attainable and admits the representation

$$V(T) = H = V_0 + \int_0^T \alpha(s)dX(s)$$

where $V_0 = E^Q[H]$. The strategy is self-financing, meaning the cost of the portfolio is a constant

$$V(t) - \int_0^t \alpha(s)dX(s) = V_0$$

where V_0 is a perfect hedge.

1.1.3.2 Incomplete market

Exposures to uncertain future events constitutes a basis of arbitrage opportunity, called intrinsic risk. Since the complete market theory failed to account for well-known market anomalies (see details in Section (5.1)), an incomplete market theory developed to understand and explain such anomalies. For instance, Follmer et al. [1990] derived a replicating portfolio in incomplete market. Letting Π be the set of all intrinsic risks, $G(\pi)$ is a measure of intrinsic risk $\pi(\omega)$, mapping from Π to R. The measure G depends on the underlying asset as well as the contingent claim, such that $G^H \in \mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$, leading to the new representation of the claim H as

$$H = V_0 + \int_0^T \alpha(s)dX(s) + G^H \tag{1.1.2}$$

Introducing the Kunita-Watanabe decomposition (see Definition (A.4.5)), the measure can be expressed as

$$G^H = G_0 + \int_0^T \alpha^H(s)dX(s) + N(T)$$

where $N(t)$ is a square-integrable martingale orthogonal to X. Thus, we have

$$H = V_0^* + \int_0^T \alpha^*(s)dX(s) + N(T)$$

where $V_0^* = V_0 + G_0$ and $\alpha^* = \alpha + \alpha^H$. Market incompleteness implies the existence of an equivalent measure in the set $\mathcal{Q}$, which is not necessarily a martingale and/or a unique measure.

1.1.4 American options

1.1.4.1 The pricing equations

Unlike European options, which can only be exercised at maturity, American options can be exercised at any time during the lifetime of the option, making them significantly more difficult to analyse than their European counterparts. We consider an American security that can be exercised at any time t up to maturity, T, for some payoff $h(S_t)$, where $(S_t)_{t=0}^T$ represents the underlying process in the range $[0,T]$. Similarly to European options, we wish to create a self-financing strategy, $(X_t)_{t=0}^T$, capable of hedging that option. We know at least that at expiry time T, we must have $X_T \geq h(S_T)$. Then at all subsequent times, t_i, the option holder can either exercise the option or hold on to its position until the next period. Hence, we can use a backward induction argument to deduce that

$$X_{t-1} = \max\{h_{t-1}, P_{t-1}\mathbb{E}^{\mathbb{Q}}[P_t X_t|\mathcal{F}_{t-1}]\}$$

where P_t represents the discount factor at time t (see Bingham et al. [2003]). If we consider instead discounted values, then we have

$$\tilde{X}_{t-1} = \max\{\tilde{h}_{t-1}, \mathbb{E}^{\mathbb{Q}}[\tilde{X}_t|\mathcal{F}_{t-1}]\}$$

Hence, $\tilde{X}_t$ is the Snell envelope of h_t, meaning that $X_t = \sup_{\tau\in\Lambda}\mathbb{E}^{\mathbb{Q}}[\tilde{h_t}|\mathcal{F}_t]$ and the stopping time

$$\tau^* = \{\min s \geq t : \tilde{X}_s = \tilde{h_s}\} \tag{1.1.3}$$

is optimal, where Λ represents the set of all stopping times. In particular, in the case $t=0$, the value of the option is given by

$$x = \tilde{X}_0 = \sup_{\tau\in\Lambda_0}\mathbb{E}^{\mathbb{Q}}[\tilde{h}_\tau] \tag{1.1.4}$$

No analytical solution exists to this problem, and so we must rely on other methods to compute the prices of American options in practice.

1.1.4.2 Optimal exercise of American options

Ddespite the apparent attractiveness of the freedom of exercise, it is rarely in the best of interests of market participants to exercise an American option before its expiry date, especially in the case of American call options. In fact, in a no-dividend environment, it is never optimal to exercise an American call option prior to expiration, rather, they can be viewed as equivalent to their European counterparts as is demonstrated by the following Proposition.

***Proposition* 1** *Given an American call option with payoff function $h : [0,\infty) \to \mathbb{R}$ given by $h(S_t) = (S_t - K)^+$, for asset price S_t at time t and strike price K, the value of this security is the same as the value of the equivalent European option with payoff at maturity $h(S_T)$. That is,*

$$\mathbb{E}^{\mathbb{Q}}[\tilde{h}(S_T)] = \sup_{\tau\in\Lambda}\mathbb{E}^{\mathbb{Q}}[\tilde{h}(S_\tau)]$$

and the optimal exercise time $\tau^* = T$.

This result follows from the convexity of the payoff function and Jensen's inequality. We provide a proof similar to that given by Shreve [2003].

■ Since the payoff function h is convex for some $\lambda \in [0,1]$ we have $h(\lambda x) \leq \lambda h(x)$ for all $x \in [0,\infty)$. Hence for times $\{t_i\}_{i=1}^n$ and rate $r \geq 0$, we have

$$
\begin{aligned}
\mathbb{E}^{\mathbb{Q}}[(1+r)^{-t_{i+1}}h(S_{t_{i+1}})|\mathcal{F}_{t_i}] &= (1+r)^{-t_i}\mathbb{E}^{\mathbb{Q}}[(1+r)^{-(t_{i+1}-t_i)}h(S_{t_{i+1}})|\mathcal{F}_{t_i}] \\
&\geq (1+r)^{-t_i}\mathbb{E}^{\mathbb{Q}}[h((1+r)^{-(t_{i+1}-t_i)}S_{t_{i+1}})|\mathcal{F}_{t_i}] \\
&\geq (1+r)^{-t_i}h\left(\mathbb{E}^{\mathbb{Q}}[(1+r)^{-(t_{i+1}-t_i)}S_{t_{i+1}}|\mathcal{F}_{t_i}]\right) \\
&= (1+r)^{-t_i}h(S_{t_i})
\end{aligned}
$$

where we have also used Jensen's inequality. Hence, we can see that the process $\{(1+r)^{-t_i}h(S_{t_i})\}_{i=0}^{n}$ is a sub-martingale, and so by the Optional Sampling Theorem (Appendix A.5) we have

$$
(1+r)^{-\tau}h(S_\tau) \leq \mathbb{E}^{\mathbb{Q}}[(1+r)^{T}h(S_T)|\mathcal{F}_\tau]
$$

Taking expectations, and the supremum over all stopping times yields

$$
\begin{aligned}
\sup_{\tau\in\Lambda}\mathbb{E}^{\mathbb{Q}}[(1+r)^{-\tau}h(S_\tau)] &\leq \mathbb{E}^{\mathbb{Q}}\left(\mathbb{E}^{\mathbb{Q}}[(1+r)^{-T}h(S_T)|\mathcal{F}_\tau]\right) \\
&= \mathbb{E}^{\mathbb{Q}}[(1+r)^{-T}h(S_T)]
\end{aligned}
$$

Since the option holder will always have the possibility of exercising at expiration time T, it must be the case that the strict inequality in the above can not hold. Hence, we must conclude

$$
\sup_{\tau\in\Lambda}\mathbb{E}^{\mathbb{Q}}[(1+r)^{-\tau}h(S_\tau)] = \mathbb{E}^{\mathbb{Q}}[(1+r)^{-T}h(S_T)]
$$

In fact, the proof of Proposition (1) will hold for any positive convex function h. However, the result no longer holds when the underlying asset pays dividends. It can, on occasion, be optimal for the owner of an American call option to exercise immediately prior to an ex dividend date. For example, as discussed by Hull [2012], let us assume N ex-dividend dates $t_{d_1} < ... < t_{d_N}$ during the lifetime of an American call option, paying dividends $d_1, ..., d_N$ respectively. Approximating in the region around the ex-dividend date t_{d_N}, if the call option is exercised immediately prior to the payment the holder will receive $S(t_{d_N}) - K$, where as the stock price will fall to $S(t_{d_N}) - d_N$ if the option is not exercised, and the value of the option will be $S(t_{d_N}) - d_N - Ke^{-r(T-t_{d_N})}$. Hence, if

$$
\begin{aligned}
S(t_{d_N}) - d_N - Ke^{-r(T-t_{d_N})} &\geq S(t_{d_N}) - K \\
&\iff d_N \leq K[1 - e^{r(T-t_{d_N})}]
\end{aligned}
$$

then it can not be optimal to exercise the American call option at prior to time t_{d_N}. Conversely, if

$$
d_N > K[1 - e^{r(T-t_{d_N})}]
$$

then it may be optimal to do so, given reasonable assumptions about the dynamics of the underlying asset. Clearly, this condition is most likely to be satisfied for a large dividend, d_N, paid close to maturity (when $T - t_{d_N}$ is small). Likewise, by backward recursion, it may only be optimal to exercise at some ex-dividend t_{d_i}, $i < N$ given that

$$
d_i > K[1 - e^{r(t_{d_{i+1}} - t_{d_i})}]
$$

It appears that an American call option will most likely be exercised early, if done so immediately prior to the final ex-dividend date, and even then under some conditions. However, this analysis does show the possibility for early exercise and gives us some insight into how European and American call options can differ, as will become increasingly important as we attempt to build a model capable of capturing a volatility surface in both cases. For American put options the situation is somewhat different. Not only is early exercise a possibility when dividends are present, but even without dividends, the holder of an American put option may benefit from exercising their right to sell prior to maturity. In fact, it is regarded that the put option should always be exercised if at any point it is sufficiently deep in the money during the lifetime of the option. This is related to the lower bound on the stock price and therefore upper bound on the put option payoff, that does not exist in the case of a call option. With a stock price very close to zero, and a strike price such that the put option is deep in the money, the holder of a put option can do little better than to exercise the option and take the current value. Since the stock price could only rise, the payoff of the put option could only fall. Moreover, exercising immediately is preferable to waiting, even with a constant stock price, due to the time value of money and the fact that the strike value you receive, will be worth less in the future. Hull [2012] notes that both call and put options provide insurance against movements in the value of the underlying asset. A put option held with stock insures against the risk of the underlying value becoming too low. However, due to the zero bound on stock prices, such insurance becomes irrelevant when the bound is hit, making it optimal to forgo the insurance and receive the payoff immediately.

1.2 The Black-Scholes framework

1.2.1 The assumption of lognormality

In the absence of complete information about the time evolution of the risky asset $(S_t)_{t \geq 0}$, it is natural to model its value at a future date as a random variable defined on some probability space. Bachelier [1900] first modelled the dynamics of financial assets with a Brownian motion and a drift. As prices can become negative in that setting, Samuelson [1960] proposed to directly model the returns of the underlying assets. Assuming that the returns between two periods are measured by the difference between the logarithms of the asset prices, and that returns are modelled with a Brownian motion with volatility σ and a drift $\mu - \frac{1}{2}\sigma^2$, then for $\{S_t; t \in [0,T]\}$ the returns $\log S_t - \log S_s$ are normally distributed with mean $(\mu - \frac{1}{2}\sigma^2)(t-s)$ and variance $\sigma^2(t-s)$. Further, given $0 < t_1 < .. < t_n$, then the relative prices $\{\frac{S_{t_{i+1}}}{S_i}; 0 \leq i \leq n-1\}$ are independent and follow the same law. That is, there exists a Brownian motion $\widehat{W}$ such that

$$S_t = f(t, \widehat{W}_t) = xe^{\mu t + \sigma \widehat{W}_t - \frac{1}{2}\sigma^2 t}$$

with initial condition $S_0 = x$. Applying Ito's lemma to the function $f(t,z) = xe^{\mu t + \sigma z - \frac{1}{2}\sigma^2 t}$, we compute the time and space derivatives

$$f_t^{'}(t,z) = f(t,z)(\mu - \frac{1}{2}\sigma^2) \, , \, f_z^{'}(t,z) = f(t,z)\sigma \, , \, f_{zz}^{''}(t,z) = f(t,z)\sigma^2$$

and we get the dynamics of the stock price as

$$\frac{dS_t}{S_t} = \mu dt + \sigma d\widehat{W}_t$$

Given the initial condition $S_0 = x$, the first two moments of the stock price are given by

$$\begin{aligned} E[S_t] &= xe^{\mu t} \, , \, E[S_t^2] = x^2 e^{(2\mu + \sigma^2)t} \\ Var(S_t) &= x^2 e^{2\mu t}(e^{\sigma^2 t} - 1) \end{aligned}$$

and the Sharpe ratio becomes

$$R_S = \frac{E[S_t] - x}{\sqrt{Var(S_t)}} = \frac{e^{(2\mu + \sigma^2)t}}{\sqrt{e^{2\mu t}(e^{\sigma^2 t} - 1)}}$$

which is independent from the initial value x. Further, given a bounded positive function h, the option price in Equation (1.1.1) simplifies to

$$E[h(S_t^x)] = \int h(y)\phi(t,x,y)dy = \int h(xe^{(\mu-\frac{1}{2}\sigma^2)t+\sigma\sqrt{t}u})\psi(u)du$$

where $\psi(z) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}$ is the Gaussian density. In the special case where the volatility is set to zero ($\sigma = 0$) the rate of return of the asset should equate the risk-free rate r

$$\frac{dS_t^0}{S_t^0} = rdt$$

In all other cases ($\sigma \neq 0$) the market compares the rate of return per unit of time μ to the return a risk-free asset, so that the market reference is $\mu - r$. In that setting the Sharpe ratio is assumed to be the market price of risk, λ, that the market assign to the noise $\widehat{W}$ as

$$R_S = \lambda = \frac{\frac{1}{dt}E[\frac{dS_t}{S_t}] - r}{\sqrt{\frac{1}{dt}Var(\frac{dS_t}{S_t})}} = \frac{\mu - r}{\sigma}$$

We can then rewrite the dynamics of the stock price as

$$\frac{dS_t}{S_t} = rdt + \sigma\big(d\widehat{W}_t + \lambda dt\big)$$

As a result, the market price of risk is not specific to the asset price S_t, but to the noise $\widehat{W}_t$.

1.2.2 Defining the replicating portfolio

Initial work on the valuation of options has been expressed in terms of warrants (see Sprenkle [1961], Samuelson [1965] and Samuelson et al. [1969]), all providing valuation formulas involving one or more arbitrary parameters. Thorp et al. [1967] obtained an empirical valuation formula for warrants which they used to calculate the ratio of shares of stock to options needed to create a hedge position by going long in one security and short in the other. However, they failed to see that in equilibrium, the expected return on such a hedge position must be equal to the return on a riskless asset. Using this equilibrium condition, Black et al. [1973] derived a theoretical valuation formula. To do so, they assumed the following ideal conditions in the market

1. The short-term interest rate is known and constant through time.
2. The stock price follows a random walk in continuous time with a variance rate proportional to the square of the stock price.
3. The stock pays no dividends or other distributions.
4. The option is European.
5. There are no transaction costs in buying or selling the stock or the option.
6. It is possible to borrow any fraction of the price of a security to buy it or to hold it, at the short-term interest rate.
7. There are no penalties to short selling.

Under these assumptions, the value of the option will only depend on the price of the stock and time, so that it is possible to create a hedged position, consisting of a long position in the stock and a short position in the option. In a complete market, every terminal payoff can be replicated by a portfolio, so that the problem of defining the price of an option is linked with that of finding a replicating portfolio. That is, using the option premium, one can buy and sell the underlying to hedge risk against the derivative. Following that argument, Black and Scholes defined the price of a derivative option as the price of its hedge. Knowing the dynamics of the underlying price we need to formalise the evolution of the self-financing replicating portfolio dynamically managed (without adding or removing capital during the trading period). Following the approach in Section (1.1.3.1), we put ourself in the Black-Scholes world and we assume a market made of a risk-free asset S_t^0 and a risky asset with price S_t. We model the dynamics of the future stock prices under the historical measure $\mathbb{P}$ as

$$\begin{aligned} dS_t &= S_t\big(\mu_t dt + \sigma_t d\widehat{W}_t\big) \\ \mu_t &= r_t + \sigma_t \lambda_t \end{aligned}$$

We denote the quantity of cash that one can hold as $\beta(t)S_t^0$ with dynamics given by

$$dS_t^0 = S_t^0 r dt$$

We build a self-financing portfolio V made of cash and stocks with value at time t being

$$V_t = \alpha(t)S_t + \beta(t)S_t^0$$

and we denote the value invested in stocks as $\pi_t = \alpha(t)S_t$. Over a very short period of time the variations of the portfolio are only due to the variations of the asset price and the interest rate rewarding the cash held at the bank. Hence, the portfolio dynamics are given by

$$dV_t = \alpha(t)dS_t + \beta(t)dS_t^0$$

Replacing the respective dynamics in the above equation, the self-financing portfolio satisfies the stochastic differential equation (SDE)

$$dV_t = r_t V_t dt + \pi_t \sigma_t (d\widehat{W}_t + \lambda_t dt)$$

Market completeness is equivalent to assuming that markets are efficient, that is, the price of an asset at a given time contains all past information as well as the market expectation of its future value. This means that there exists a strategy leading to a final P/L of null value in all possible future configuration. As a result, we can conclude that the option price is equal to the value of the replicating portfolio at the initial time. A direct application of the above assumption is that we can express the Black-Scholes price dynamics as

$$dC_{BS}(t,S_t) = r_t C_{BS}(t,S_t)dt + \pi_t \sigma_t (d\widehat{W}_t + \lambda_t dt)$$

Using Girsanov's theorem (see Theorem (A.5.3)), we can re-express that price dynamics under the risk-neutral measure $\mathbb{Q}$ by applying the change of measure

$$W_t = \widehat{W}_t + \int_0^t \lambda_s ds$$

where W_t is a $\mathbb{Q}$-Brownian motion. Therefore, the Black-Scholes price at all future time is a solution to the following stochastic differential equation with terminal condition

$$\begin{aligned} dC_{BS}(t,S_t) &= r_t C_{BS}(t,S_t)dt + \pi_t \sigma_t dW_t \\ C_{BS}(T,S_T) &= h(S_T) \end{aligned} \tag{1.2.5}$$

where the pair $(\alpha(t), \beta(t))$ allows replication of the contingent claim paying $h(S_T)$ at maturity. Note, a consequence of the option price being expressed under the risk-neutral measure is that it neither depend on the return μ of the risky asset nor the market price of risk λ. However, the risk due to the variations of the risky asset is still present and significantly impact the option price via the volatility parameter σ. Thus, to avoid arbitrage, a market price of risk is computed such that the discounted spot price is a martingale with respect to the equivalent probability measure. In the risk-neutral measure, the drift of the spot price is equal to the risk-free rate so that the key parameter becomes the volatility of the dynamics. Volatility is a key parameter when risk managing option prices as it allows the portfolio manager, or trader, to define his hedging strategies by computing the necessary Greeks (see details in Section (1.6.2)). The Vega is a risk measure quantifying exposure to a misspecified volatility.

1.2.3 The Black-Scholes formula

1.2.3.1 The closed-form solution

In the special case where we assume that the stock prices $(S_t^x)_{t\geq 0}$ are lognormally distributed and pay a continuous dividend, we can derive most of the contract market values in closed-form solution. For example, Black and Scholes [1973] derived the price of a call option seen at time t with strike K and maturity T as

$$C_{BS}(t,x,K,T) = xe^{-q(T-t)}N(d_1(T-t,F(t,T),K)) - Ke^{-r(T-t)}N(d_2(T-t,F(t,T),K)) \tag{1.2.6}$$

where $F(t,T) = xe^{(r-q)(T-t)}$ is the forward price and $N(\bullet)$ is the normal cumulative distribution function (see details in Appendix (B.5.1)). Further, we have

$$d_2(t,x,y) = \frac{1}{\sigma\sqrt{t}}\log\frac{x}{y} - \frac{1}{2}\sigma\sqrt{t} \text{ and } d_1(t,x,y) = d_2(t,x,y) + \sigma\sqrt{t}$$

with the following properties

$$d_2(t,y,x) = -d_1(t,x,y) \text{ and } d_1(t,y,x) = -d_2(t,x,y)$$

and

$$d_1(t,x,y)d_2(t,x,y) = \frac{1}{\sigma^2 t}(\log\frac{x}{y})^2 - \frac{1}{4}\sigma^2 t$$

Computing the limits when the volatility goes to zero and to infinity, we get

$$\lim_{\sigma\to 0} d_i = \infty \text{ for } i = 1,2$$

and

$$\lim_{\sigma\to\infty} d_1 = +\infty \text{ and } \lim_{\sigma\to\infty} d_2 = -\infty$$

From the limits of the CDF, $\lim_{x\to-\infty} N(x) = 0$ and $\lim_{x\to\infty} N(x) = 1$, we get

$$\lim_{\sigma\to 0} N(d_i) = 1 \text{ for } i = 1,2$$

and

$$\lim_{\sigma\to\infty} N(d_1) = 1 \text{ and } \lim_{\sigma\to\infty} N(d_2) = 0$$

Note, we can compute the limit of a function at a point when that function is continuous, which is not the case with a call option. In the following, we use a heuristic to take into consideration the discountinuity. In a call option we get the limit cases

$$\lim_{\sigma \to 0} C_{BS}(t,x,K,T) = xe^{-q(T-t)}N(\infty) - Ke^{-r(T-t)}N(\infty) = \left(xe^{-q(T-t)} - Ke^{-r(T-t)}\right)^+$$

and

$$\lim_{\sigma \to \infty} C_{BS}(t,x,K,T) = xe^{-q(T-t)}N(\infty) - Ke^{-r(T-t)}N(-\infty) = xe^{-q(T-t)}$$

Similarly, the price of a put option is given by

$$P_{BS}(t,x,K,T) = Ke^{-r(T-t)}N(d_1(T-t,K,xe^{(r-q)(T-t)})) - xe^{-q(T-t)}N(d_2(T-t,K,xe^{(r-q)(T-t)})) \quad (1.2.7)$$

and the limit cases of a put option satisfy

$$\lim_{\sigma \to 0} P_{BS}(t,x,K,T) = Ke^{-r(T-t)}N(\infty) - xe^{-q(T-t)}N(\infty) = \left(Ke^{-r(T-t)} - xe^{-q(T-t)}\right)^+$$

and

$$\lim_{\sigma \to \infty} P_{BS}(t,x,K,T) = Ke^{-r(T-t)}N(\infty) - xe^{-q(T-t)}N(-\infty) = Ke^{-r(T-t)}$$

Being long a call option and short a put option, when the volatility tends to infinity, we get

$$\lim_{\sigma \to \infty} C_{BS}(t,x,K,T) - \lim_{\sigma \to \infty} P_{BS}(t,x,K,T) = xe^{-q(T-t)} - Ke^{-r(T-t)}$$

and we recover the call-put parity. Hence, we see that for $\sigma \to 0$ both the call and put price tend to their intrinsic values. Further, option price is postively related to volatility, no matter if the option is a call or a put. For a call option, the higher the volatility is, the greater the chance that the underlying value raises above strike price. For a put option, the higher the volatility is, the greater the chance that the underlying value falls below strike price.

1.2.3.2 The total variance representation

For notational purpose, we let $P(t,T)$ be the discount factor, $Re(t,T)$ be the repo factor and we define

$$\eta = \frac{K}{F(t,T)} = \frac{KP(t,T)}{xRe(t,T)}$$

to be the forward moneyness of the option. It leads to the limit case

$$\begin{aligned} \lim_{\eta \to 1} d_2(.) &= -\frac{1}{2}\sigma\sqrt{t} \\ \lim_{\eta \to 1} d_1(.) &= \frac{1}{2}\sigma\sqrt{t} \end{aligned}$$

It is well known that when the spot rate, repo rate and volatility are time-dependent, we can still use the Black-Scholes formula (1.2.3) with the model parameters expressed as

$$r = \frac{1}{T-t}\int_t^T r(s)ds \; , \; q = \frac{1}{T-t}\int_t^T q(s)ds \; , \; \sigma^2 = \frac{1}{T-t}\int_t^T \sigma^2(s)ds$$

Further, we let the Black-Scholes total variance be given by $\omega(t) = \sigma^2 t$, and rewrite the BS-formula in terms of the total variance, denoted $C_{TV}(t,x,K,T)$, where

$$d_2(t,x,y) = \frac{1}{\sqrt{\omega(t)}} \log \frac{x}{y} - \frac{1}{2}\sqrt{\omega(t)} \text{ and } d_1(t,x,y) = d_2(t,x,y) + \sqrt{\omega(t)} \quad (1.2.8)$$

Expressing the strike in terms of the forward price $K = \eta F(t,T)$, the call price in Equation (1.2.3) becomes

$$C_{TV}(t,x,K,T)\big|_{K=\eta F(t,T)} = xe^{-q(T-t)}\big(N(d_1(\eta,\omega(T-t)) - \eta N(d_2(\eta,\omega(T-t)))\big) \tag{1.2.9}$$

where $d_2(\eta,\omega(t)) = \frac{1}{\sqrt{\omega(t)}}\log\frac{1}{\eta} - \frac{1}{2}\sqrt{\omega(t)}$, or equivalently

$$d_2(\eta,\omega(t)) = -\frac{1}{\sqrt{\omega(t)}}\log\eta - \frac{1}{2}\sqrt{\omega(t)} \text{ and } d_1(\eta,\omega(t)) = d_2(\eta,\omega(t)) + \sqrt{\omega(t)} \tag{1.2.10}$$

which only depends on the forward moneyness and the total variance. This is the scaled Black-Scholes function discussed by Durrleman [2003].

1.2.3.3 Some properties

In the special case where the initial spot price $S_t = x$ is close to the at-the-money forward strike, $K \approx xe^{(r-q)(T-t)}$, and the call price in Equation (1.2.3) can be approximated with

$$C_{BS}(t,x,K,T) \approx xe^{-q(T-t)}\big(\frac{1}{2} + \frac{1}{5}\sigma\sqrt{T-t}\big) - Ke^{-r(T-t)}\big(\frac{1}{2} - \frac{1}{5}\sigma\sqrt{T-t}\big) \tag{1.2.11}$$

which is linear in the spot price and the volatility. Similarly, from the formula of the put option in Equation (1.2.7), the put price can be approximated as

$$P_{BS}(t,x,K,T) \approx Ke^{-r(T-t)}\big(\frac{1}{2} + \frac{1}{5}\sigma\sqrt{T-t}\big) - xe^{-q(T-t)}\big(\frac{1}{2} - \frac{1}{5}\sigma\sqrt{T-t}\big)$$

Moreover, when the initial spot price $S_t = x$ is exactly at-the-money forward $K = F(t,T)$ with $\eta = 1$, the call price can be approximated with

$$C_{TV}(t,x,K,T)\big|_{K=F(t,T)} \approx xe^{-q(T-t)}\big(N(\frac{1}{2}\sqrt{\omega(T-t)}) - N(-\frac{1}{2}\sqrt{\omega(T-t)})\big) = 0.4xe^{-q(T-t)}\sqrt{\omega(T-t)} \tag{1.2.12}$$

which is linear in the spot price and the square root of the total variance. It comes from the fact that

$$N(x) - N(-x) \approx \frac{1}{\sqrt{2\pi}}(2x - \frac{1}{3}x^3)$$

For $x = \frac{1}{2}\sigma\sqrt{T-t}$, which is small, so is the quantity $\frac{1}{3}x^3$ and we get

$$N(x) - N(-x) \approx \frac{1}{\sqrt{2\pi}}2x = \frac{1}{\sqrt{2\pi}}\sigma\sqrt{T-t}$$

where $\frac{1}{\sqrt{2\pi}} \approx \frac{2}{5}$. One of the property of the call and put prices in the Black-Scholes model is that they are homogeneous functions

$$C_{BS}(t,\lambda x,\lambda K,T) = \lambda C_{BS}(t,x,K,T) \tag{1.2.13}$$

which is why strikes are expressed as percentage of the spot price.

1.3 Introduction to binomial trees

We presented in Section (1.1.1.2) European and American options and argued in Section (1.1.2) that their prices were contingent on the dynamics of their underlying asset. Using elementary mathematics, the uncertainty of the underlying stock price reduce to discrete binomial movements, leading to a binomial option pricing formula for both European and American options. The binomial model was introduced into option pricing theory by Cox et al. [1979], and since then, it has been extensively covered in many textbooks on mathematical finance (see Shreve [2003], London [2005]). Binomial models highlights the economic intuition behind option pricing theory in a simple way. In the following sections, we are going to briefly describe different binomial trees and introduce American pricing with discrete dividends.

1.3.1 Describing a few binomial trees

1.3.1.1 The CRR and JRR models

We assume a discrete stock price $S_i = S(i\Delta t)$, $i = 0, .., N$ going up with probability p and down with probability $1-p$, such that over one time period, Δt, if the stock price moves up its value is Su with $u = e^{\sigma \Delta t}$, and if it moves down its value is Sd with $d = \frac{1}{u}$. Since in a risk-neutral world the investor expect to earn the risk-free rate on riskless portfolios, we get

$$pSu + (1-p)Sd = Se^{r\Delta t}$$

Solving for p, we get

$$p = \frac{e^{r\Delta t} - d}{u - d} \text{ and } 1 - p = \frac{u - e^{r\Delta t}}{u - d}$$

We let $\tau = T - t = N\Delta t$ where N is the number of time periods, such that there are 2^N possible price paths from (t, S) to (T, S_T). Further, we let (i, j) denotes the jth node at the ith time step, with price S_{ij} given by

$$S_{ij} = S_0 u^j d^{i-j} \; , i = 0, 1, ..., N \; , j = 0, ..., i$$

with $S_0 d^i < S_0 u d^{i-1} < S_0 u^2 d^{i-2} < ... < S_0 u^{i-1} d < S_0 u^i$. Since $d = \frac{1}{u}$, we can rewrite the asset price as $S_{ij} = S_0 u^{2j-i}$ for $j = 0, ..., i$, leading to $S_0 u^{-i} < S_0 u^{2-i} < S_0 u^{4-i} < ... < S_0 u^{2(i-1)-i} < S_0 u^i$. The payoff after N periods of time is

$$f_{Nj} = F(S_0 u^j d^{i-j})$$

Using backward induction, the price of an option at node (i, j) is found by taking the expected valus of the option at time $i + 1$

$$f_{ij} = e^{-r\Delta t}\big(p f_{i+1,j+1} + (1-p) f_{i+1,j}\big)$$

The multiperiod binomial valuation formula for a European option is

$$f = e^{-r\tau} \sum_{j=1}^{N} p_N f_{N,j}$$

where $p_{N,j}$ is the risk-neutral binomial probability to end up in state (N, j), given by

$$p_{N,j} = \binom{N}{j} p^j (1-p)^{N-j}$$

with $\binom{n}{j} = \frac{n!}{j!(n-j)!}$. We need to relate the discrete-time multiplicative binomial model with parameters (u, d, p, r) to the continuous time BS-model with parameters (σ, r). In the limit of infinitesimally small time steps, $N \to \infty$ and $\Delta t \to 0$, the binomial option pricing formula must converge to the BS-formula

$$f = e^{-r\tau} \int_0^\infty F(S_T) p^Q(S_T, T|S, t) dS_T$$

for the two formulations to be consistent.

Proposition 2 *The distribution of the terminal stock price in the binomial model with parameters (u, d, p, r) converges to the Black-Scholes lognormal distribution with parameters (σ, r) as $\Delta t \to 0$ if and only if*

$$p \ln \frac{u}{d} + \ln d = \mu \Delta t + o(\Delta t)$$

and

$$p(1-p)(\ln \frac{u}{d})^2 = \sigma^2 \Delta t + o(\Delta t)$$

Proof can be found in the book by London [2005]. Since there are two equations for three unknowns (u, d, p) relating them to (σ, r) which must hold only in the first order in Δt, one can find an infinite number of binomial models consistent with the BS-model. One way forward is for the mean and variance of the stock price distribution at the end of the period Δt to match exactly the mean and variance of the lognormal distribution. That is,

$$\begin{aligned} pu + (1-p)d &= e^{r\Delta t} \\ pu^2 + (1-p)d^2 &= e^{(2r+\sigma^2)\Delta t} \end{aligned}$$

and we therefore get

$$\begin{aligned} p &= \frac{e^{r\Delta t} - d}{u - d} \\ e^{r\Delta t}(u + d) - du &= e^{(2r+\sigma^2)\Delta t} \end{aligned}$$

However, we still need an additional equation to solve for u and d in terms of r and σ.
Assuming $u = \frac{1}{d}$, Cox et al. [1979] obtained a solution to the above system of equations

$$\begin{aligned} u &= A + \sqrt{A^2 - 1} \\ d &= A - \sqrt{A^2 - 1} \end{aligned}$$

where $A = \frac{1}{2}\left(e^{-r\Delta t} + e^{(r+\sigma^2)\Delta t}\right)$, which can be approximated as

$$\begin{aligned} u &= e^{\sigma\sqrt{\Delta t}} \\ d &= e^{-\sigma\sqrt{\Delta t}} \end{aligned}$$

In the limit $\Delta t \to 0$, both the binomial and BS-model converge to the same limit.
Jarrow et al. [1983] considered equal probabilities for up and down price movements, $p = 1 - p = \frac{1}{2}$, leading to

$$\begin{aligned} u &= e^{\mu\Delta t + \sigma\sqrt{\Delta t}} \\ d &= e^{\mu\Delta t - \sigma\sqrt{\Delta t}} \end{aligned}$$

with $\mu = r - \frac{1}{2}\sigma^2$. Note, the CRR tree is symmetric since $ud = 1$, but the up and down probabilities are not equal, while in the JR tree the probabilities are equal, but the tree is skewed since $ud = e^{2\mu\Delta t}$. We can combine the two approaches by setting $ud = e^{2\nu\Delta t}$ for some scalar ν. We then get

$$\begin{aligned} u &= e^{\nu\Delta t + \sigma\sqrt{\Delta t}} \\ d &= e^{\nu\Delta t - \sigma\sqrt{\Delta t}} \end{aligned}$$

with probabilities

$$p = \frac{1}{2} + \frac{1}{2}\frac{\mu - \nu}{\sigma}\sqrt{\Delta t} \text{ and } 1 - p = \frac{1}{2} - \frac{1}{2}\frac{\mu - \nu}{\sigma}\sqrt{\Delta t}$$

where $\nu = 0$ is the CRR and $\nu = \mu$ is the JR model.
Trigeorgis [1991] proposed to relax the constraint of small time intervals by considering the natural logarithm of the asset price $X = \ln S$ with drift $\mu = r - \frac{1}{2}\sigma^2$ going up to $\Delta X_u = X + \Delta X$ with probability p_u, or down to $\Delta X_d = X - \Delta X$ with probability $p_d = 1 - p_u$. This is the additive binomial model, as opposed to the multiplicative one. Equating the mean and variance of that model to that of BS-formula, we get

$$\begin{aligned} \Delta X_u &= \frac{1}{2}\mu\Delta t + \frac{1}{2}\sqrt{4\sigma^2\Delta t - 3\mu^2(\Delta t)^2} \\ \Delta X_d &= \frac{3}{2}\mu\Delta t - \frac{1}{2}\sqrt{4\sigma^2\Delta t - 3\mu^2(\Delta t)^2} \end{aligned}$$

Assuming equal jump size, we get the system

$$\begin{aligned} &p_u(\Delta X) + p_d(\Delta X) = \mu\Delta t \\ &p_u(\Delta X)^2 + p_d(\Delta X)^2 = \mu^2(\Delta t)^2 + \sigma^2\Delta t \end{aligned}$$

which we solve as

$$\begin{aligned} \Delta X &= \sqrt{\mu^2(\Delta t)^2 + \sigma^2} \\ p_u &= \frac{1}{2} + \frac{1}{2}\frac{\mu\Delta t}{\Delta X} \end{aligned}$$

1.3.1.2 American-style options

In the case of American-style options with early exercise, we need to evaluate whether early exercise is optimal at each node. Thus, we can modify the binomial trees as follow

$$f_{ij} = \max\big(F(S_0 u^j d^{i-j}), e^{-r\Delta t}\big(pf_{i+1,j+1} + (1-p)f_{i+1,j}\big)\big)$$

We can also extend the model to price American-style options on foreign currencies by setting $r \to r - r_f$ where r_f is the foreign risk-free rate. Further, we can obtain options on stock indexes by setting $r \to r - q$ where q is the dividend yield on the index. When computing the hedge statistics of an option, we can read the delta and gamma off the tree. The delta can be approximated by finite difference as follow

$$\Delta = \frac{\partial C}{\partial S} \approx \frac{\Delta C}{\Delta S} = \frac{C_{1,1} - C_{1,0}}{S_{1,1} - S_{1,0}}$$

and the gamma is given by

$$\Gamma = \frac{\partial^2 C}{\partial S^2} \approx \frac{\frac{C_{2,2} - C_{2,1}}{S_{2,2} - S_{2,1}} - \frac{C_{2,1} - C_{2,0}}{S_{2,1} - S_{2,0}}}{\frac{1}{2}(S_{2,2} - S_{2,0})}$$

Unfortunately, the vega, rho, and theta are estimated by recomputation of the option price for small changes in volatility, risk-free rate, and time to maturity, respectively.

1.3.1.3 The BBSR model

The binomial-Black-Scholes-Richardson (BBSR) model was proposed by Broadie et al. [1996] to enhance the CRR tree approach for pricing American or European options. The option price tree is modified by replacing the continuation values at the nodes just prior to expiration with values from the BS-formula. This is the binomial-Black-Scholes (BBS) algorithm. Then the BBSR price is computed by using the Richardson extrapolation, which sets the BBSR price equal to twice the BBS price estimated using the desired number of time-steps minus the BBS price using half the number of time-steps. The authors found that the BBSR algorithm offered significant improvements in pricing accuracy over alternative algorithms.

1.3.2 Introducing discrete dividends

We discussed discrete dividends in Appendix (B.2), where we presented a few modelling assumptions and showed that when considering the escrowed model and the forward model for American options, it was easy to construct recombing binomial trees. In the case of the escrowed model, the CRR tree for European options is modified by reducing the initial asset price with the present value of expected dividends during the remaining life of the option. However, for American options the exercise decision must be evaluated at each node, and the present value of future dividends during the remaining life of the option must be added back to the asset price on the tree before evaluating the exercise decision. Further, in the case of American options with a single dividend, Black [1975] gave an approximation as well as Geske [1979] and Whaley [1981]. However, these methods can lead to arbitrage opportunities. Even though it is easy to extend binomial trees to include cash dividends, a direct approach such as the piecewise lognormal model, leads to non-recombining trees. For instance, in the CRR tree, if a dividend D is paid at time step m, we get

$$\big(S_0 u^j d^{m-j} - D\big)d \neq \big(S_0 u^{j-1} d^{m-(j-1)} - D\big)u$$

and one would have to build a new tree from every node after each dividend date (see Hull [2012]). Vellekoop et al. [2006] proposed a way around to get a recombining tree by letting the dividends be stock-price dependent and using an interpolation technique after each dividend date. This approach, which we briefly describe, was introduced by Wilmott et al. [1993] in the case of a partial differential equation (PDE). Assuming a single dividend at time t_D, we define the set of nodes at that date as

$$A^n = \{S : \mathbb{Q}^n(S_{m(n)\Delta_n} = S) > 0\}$$

where the time step $m(n) \in \mathbb{N}$ satisfies $m(n)\Delta_n \leq t_D < (m(n)+1)\Delta_n$. Then the procedure is as follow:

- Build a no-dividend binomial tree with n time steps, compute the payoff at maturity and work backwards until we reach the dividend date at time step $m(n)$. We then obtain the values $f_{m(n)} = F(S)$ for the option contract in all points S of the binomial tree, that is, for all $S \in A^n$. These values $F(S)$ approximate the option values at time just after the dividend has been paid, given the stock price at that time is S.
- Since the stock price jumps down with the amount $D(S_{t_-})$ when it goes ex-dividend, the option value for a stock price S just before the dividend date equals the option value for the stock price $S - D(S)$ just after the dividend date. Hence, we need the option values $f_{m(n)} = F(\bullet)$ in all the points $S - D(S)$. Since we have calculated the values of $f_{m(n)} = F(S)$ for $S \in A^n$, we can devise a function approximating $f_{m(n)}$ on the whole of $\mathbb{R}^+$, based on the values of $f_{m(n)}$ on A^n. We let $\mathcal{B}_F$ be such an interpolation function, and we compute $\mathcal{B}_F(S - D(S))$ as an approximation to the values of $F(S - D(S))$ needed to continue the backward propagation.

- We then proceed backward as in the no-dividend binomial tree until we reach time zero.

We can then extend this approach to the case of multiple discrete dividends. Posing some explicit conditions on the interpolation procedure, the authors prooved convergence of the method.

1.3.3 The implied binomial tree

The seminal work of Breeden et al. [1978] on option-implied risk-neutral distribution is at the origin of the implied trees. An implied binomial tree (IBT) is a generalisation of the CRR tree for option pricing where the probabilities attached to outcomes in the tree are inferred from a collection of actual option prices for a given maturity, rather than simply deduced from the behaviour of the underlying asset. Some well known implementation of IBT are due to Derman et al. [1994] and Rubinstein [1994]. The latter being easier to implement and more stable than the former, we focuss on Rubinstein's tree.

1.3.3.1 Rubinstein's tree

Given option prices, the initial underlying price, and the price of the underlying at ending nodes of the CRR tree, we solve a quadratic program to infer the posterior ending node risk-neutral probabilities p_j given a prior $p_j^{'}$ based on a CRR tree. Assuming that we have m European options on the same underlying and of the same maturity, we use the BS-formula to infer the implied volatility σ of the nearest to-the-money option. Assuming n time steps of length Δt years, we get

$$u = e^{\sigma \sqrt{\Delta t}} \text{ and } d = e^{-\sigma \sqrt{\Delta t}}$$

and fixed risk-neutral probability of an up move over each step as

$$p^{'} = \frac{r - d}{u - d}$$

where $r = e^{R_t \Delta t}$ is the riskless compounding factor per time step. Then, the ending nodal risk-neutral probabilities $(i = N)$ are given by

$$p_j^{'} = \frac{n!}{j!(n-j)!} (p^{'})^j (1 - p^{'})^{n-j}$$

where $\binom{n}{j} = \frac{n!}{j!(n-j)!}$ is the standard binomial coefficient. We let C_i^b, C_i^a, and K_i be the bid, ask, and strike prices for $i = 1, .., m$ with n much larger than m. Knowing the prior probabilities $p_j^{'}$ associated with the ending nodes of a CRR tree, we can minimise the following objective function

$$\min_{p_j} \sum_{j=1}^{n} (p_j - p_j^{'})^2$$

subject to

$$\begin{aligned}
&\sum_{j=0}^{n} p_j = 1 \text{ and } p_n \geq 0 \\
&S_0 = \frac{1}{r^n} \sum_{j=1}^{n} p_j S(j) \\
&C_i^b < C_i < C_i^a \text{ where } C_i = \frac{1}{r^n} \sum_{j=0}^{n} p_j F(S(j), K_i)
\end{aligned}$$

We then use a recursive algorithm to go backwards through the tree to build the IBT tree. For each node on the tree we need to find the path probability Q at that node, the cumulative return R, and the probability of an up movement q. The sum of the path's probabilities at a node is the total (nodal) probability that the price will arrive at this node at that time step in the tree. Both CRR and Rubinstein's tree assume that each path leading to a node is of equal probability, so that the nodal probability is the path probability multiplied by the number of paths leading to that node. Hence, at the nth time step and the jth node there are $\frac{n!}{j!(n-j)!}$ paths through the tree leading to this node, so that the path probability at this node is the nodal probability divided by the number of path, that is, $p_j \frac{j!(n-j)!}{n!}$. For example, at a particular node we consider (Q, R, q) where R denotes one plus the cumulative risk-neutral return through the tree to this node, Q is the path probability of arriving at this node, and q is the probability of an up move from this node. Then, if the one-step ahead probabilities Q^+ and Q^- correspond to the ending nodes of the tree, we calculate them using $p_j \frac{j!(n-j)!}{n!}$. The path probabilities are additive backwards $Q = Q^+ + Q^-$, and the path probabilities determine the up probabilities $q = \frac{Q^+}{Q^-}$. At last, the returns cumulate probabilistically as

$$R = \frac{1}{r}\big(qR^+ + (1-q)R^-\big)$$

1.3.3.2 Accounting for incomplete markets

Given an n-period IB tree, there are $(n+1)$ ending nodes to calibrate to $(n-1)$ observed option prices and two constraints on probabilities and the underlying asset price. Hence, to get a unique set of ending nodal probabilities, we need to solve a system of $(n+1)$ linear equations. However, in market practice, we do not have enough listed strike prices to extract a unique set of ending nodal probabilities, and one must distinguish among these feasible sets the most appropriate one. Jackwerth et al. [1996] considered the set of ending nodal probabilities leading to a smoothed implied risk-neutral density by solving the following optimisation problem for the ending nodal probabilities

$$\min_{p_j} \sum_{j=1}^{n-1} \big(p_{j-1} - 2p_j + p_{j+1}\big)^2 + \alpha\big(V_i^{model} - V_i^{mkt}\big)^2$$

subject to

$$p_j \geq 0 \text{ for all } j \ (0 \leq j \leq n)$$
$$\sum_{j=0}^{n} p_j = 1$$
$$S_0 = e^{-(r-q)T} \sum_{j=0}^{n} p_j S(j)$$

where V_i^{model} is the model price and V_i^{mkt} the market price, T is the common maturity of all options, q is the dividend yield of the underlying asset, and $\alpha > 0$ is a parameter specifying the penalty for not matching the option price exactly. In the case of European options, the model prices are linear function of the ending nodal probabilities given by

$$V_i^{model} = e^{-rT} \sum_{j=0}^{n} p_j F(S(j), K_i)$$

where $F(S(j), K_i)$ is the payoff at maturity. While European options are linear function of the ending nodal probabilities, when considering American options the problem is no-longer a simple quadratic program, but a general nonlinear optimisation problem due to the early exercise feature.

1.4 Presenting the implied volatility

1.4.1 From market prices to implied volatility

1.4.1.1 The implied volatility as a predictor

We saw in Section (1.2) that the Black-Scholes model [1973] for pricing European options assumes a continuous-time economy where trading can take place continuously with no differences between lending and borrowing rates, no taxes and short-sale constraints. Investors require no compensation for taking risk, and can construct a self-financing riskless hedge which must be continuously adjusted as the asset price changes over time. In that model, the volatility is a parameter quantifying the risk associated to the returns of the underlying asset, and it is the only unknown variable. In principle, one should use the future volatility in the Black-Scholes formula, but its value is not known and needs to be estimated. As a result, practitioners use the implied volatility (IV) when managing their books, assuming that the IV bears valuable information on the asset price process and its dynamics. However, since the market crash of October 1987, options with different strikes and expirations exhibit different Black-Scholes implied volatilities. Hence, the Black-Scholes formula can be used as a mapping device from the space of option prices to a set of single real number called the implied volatilities. Hence, acknowledging the limitations of the Black-Scholes model, traders keep having to change the volatility assumption in order to match market prices. Therefore, as explained by Fengler [2005], the information content of the IV and its capability of being a predictor for future asset price volatility is of particular importance on trading markets. It is assumed that the IV bears valuable information on the asset price process and its dynamics which can be exploited in models for the pricing and hedging of other complex derivatives. One of the reason being the existence of highly liquid option and futures markets dated back from the early nineteen-nineties. In an efficient market, options instantaneously adjust to new information such that the IV predictions do not depend on the historical price or volatility series. The overall consensus of the literature is that IV based predictors do contain a substantial amount of information on future volatility and are better than only time series based methods. Nonetheless most authors conclude that IV is a biased predictor.

1.4.1.2 The implied volatility in the Black-Scholes formula

We consider non-negative price processes, and assume a perfectly liquid market for European calls. That is, call option prices for all strikes $K > 0$, and maturities $T \geq 0$ are known in the market. We define a price surface as in Roper [2010].

Definition 1.4.1 *A call price surface parameterised by s is a function*

$$C : [0,\infty) \times [0,\infty) \to \mathbb{R}$$
$$(K,T) \to C(K,T)$$

along with a real number $s > 0$.

However, market practice is to use the implied volatility when calculating the Greeks of European options. Thus, we let the implied volatility (IV) be a mapping from time, spot prices, strike prices and expiry days to $\mathbb{R}^+$

$$\Sigma : (t, S_t, K, T) \to \Sigma(t, S_t; K, T)$$

Hence, we define the implied volatility as follow:

Definition 1.4.2 *Given the option price $C(t, S_t, K, T)$ at time t for a strike K and a maturity T, the market implied volatility $\Sigma(t, S_t; K, T)$ satisfies*

$$C(t, S_t, K, T) = C_{BS}(t, S_t, K, T; \Sigma(K, T)) \tag{1.4.14}$$

where $C_{BS}(t, S_t, K, T; \sigma)$ is the Black-Scholes formula in Equation (1.2.3) for a call option with volatility σ.

Given (S, K, T) and the price of an option, there is a unique implied volatility associated to that price, since

$$\partial_\sigma C_{BS}(t, S_t, K, T; \sigma) > 0$$

Tuus, the implied volatility is obtained by inverting the Black-Scholes formula $C_{BS}^{-1}(C(t, S_t, K, T); K, T)$. Consequently, we refer to the two-dimensional map

$$(K, T) \to \Sigma(K, T)$$

as the implied volatility surface. Note, we will some time denote $\Sigma_{BS}(K, T)$ the Black-Scholes implied volatility. A representation of the volatility surface is given in Figure (1.1). When visualising the IVS, it makes more sense to consider the two-dimensional map $(\eta, T) \to \Sigma(\eta, T)$ [2] where η is the forward moneyness (corresponding to the strike $K = \eta F(t, T)$). We can also use the forward log-moneyness $\overline{\eta}$ such that the two-dimensional map becomes $(\overline{\eta}, T) \to \Sigma(\overline{\eta}, T)$ [3]. Further, we let the total variance be given by

$$\omega(\eta, T) = \nu^2(\eta, T) = \Sigma^2(\eta, T)(T - t)$$

and let the implied total variance satisfies

$$C(t, S_t, K, T) = C_{TV}(t, S_t, \eta F(t, T), T; \omega(\eta, T)) \tag{1.4.15}$$

where the two-dimensional map $(\eta, T) \to \omega(\eta, T)$ is the total variance surface. Note, $\sqrt{\omega(\overline{\eta}, T)}$ corresponds to the time-scaled implied volatility in forward log-moneyness form discussed by Roper [2010].

1.4.1.3 The implied volatility as a measure of risk

In the Black-Scholes formula, the option price is equivalent to the cost of continuously hedging the option, but in practice dynamic hedging is not a risk-free proposition. Additional risks must be taken into consideration such as changes in volatility, changes in interest rates, changes in dividends, trading costs and liquidity. If the Black-Scholes assumptions described in Section (1.2.2) were satisfied, then the IV surface would be flat, which is clearly not the case in reality. In practice, option prices with lower strikes tend to have higher implied volatilities. For a fixed maturity T, one can observe the volatility being a U-shaped function of the strike, called the volatility smile. Further, in some markets the smile is not symmetric but skewed in the direction of large strikes. We tend to see a volatility smile in markets such as the foreign exchange market, where both out-of-the-money put and call options are priced above the level assumed by the log-normal distribution of the Black-Scholes model, whilst less mass is concentrated at the center of the implied risk-neutral distribution. Hence, we view a volatility contour in which the most out-of-the-money options yield a higher implied volatility compared with the flat level of the Black-Scholes model, whilst the most at-the-money options yield lower implied volatility, generating a smile shape. In equity markets, we see more commonly a shape known as a skew. Here, out-of-the-money put options yield a higher implied volatility than that of out-of-the-money call options, creating a downward sloping contour. Again, this is a result of the fatter tail of the implied risk-neutral distribution on on the left hand side, where as the log-normal distribution possesses a fatter tail on the side of out-of-the-money call options. If we instead hold the strike, or moneyness, fixed at some level and move along the time axis, we observe the mapping $(T) \to \Sigma(T)$, known as the term structure of volatility. For a given strike K, the IV can be either increasing or decreasing with time-to-maturity, but it will converge to a constant as $T \to \infty$. For T small, we observe an inverted volatility surface with short-term options having much higher volatilities than longer-term options. This can be used to infer the markets views of future volatility, and is often used as a precursor to future realised volatiltiy, allowing traders to identify profitable strategies through discrepancies between implied and realised volatility.

[2] Here $\Sigma(\eta, T)$ must be understood as $\Theta(\eta, T) \to \Sigma(F(t, T)\eta, T)$
[3] Here $\Sigma(\overline{\eta}, T)$ must be understood as $\Theta(\overline{\eta}, T) \to \Sigma(F(t, T)e^{\overline{\eta}}, T)$

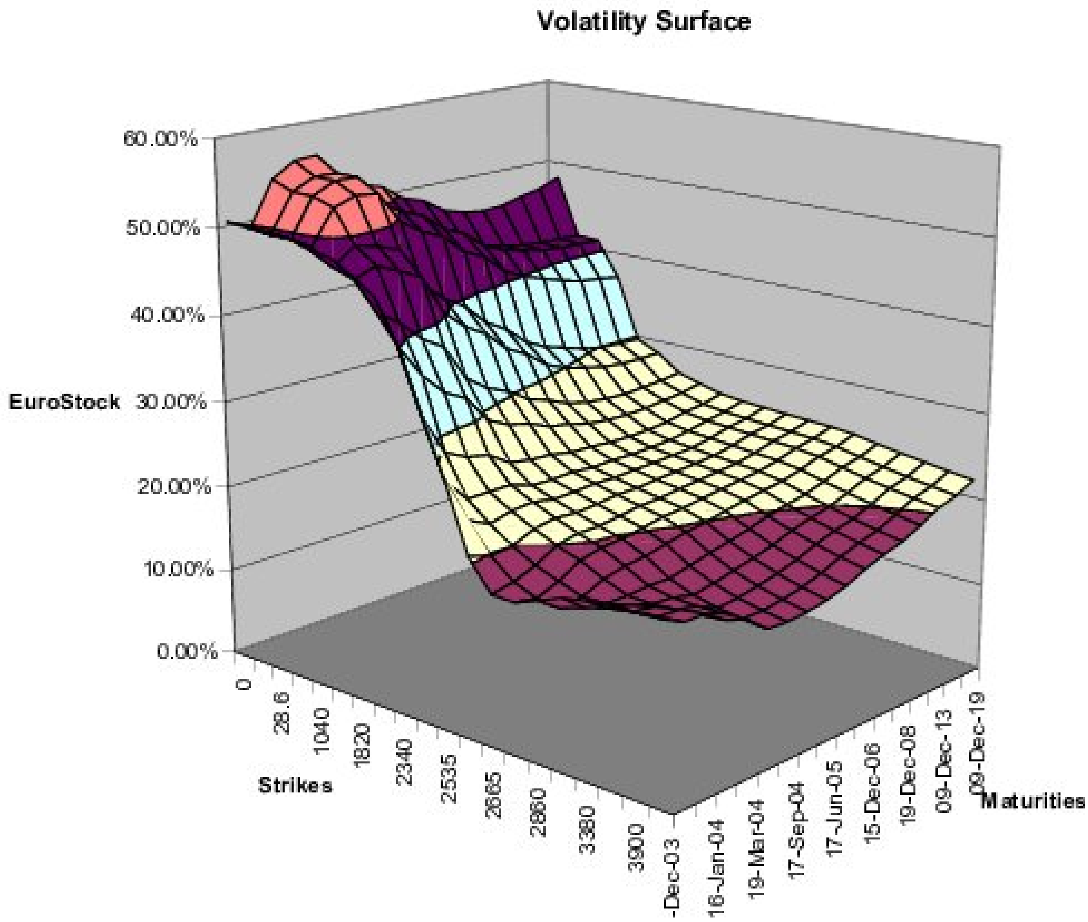


Figure 1.1: EuroStock Volatility Surface on the $8/12/2003$.

As discussed by Figlewski [1989], the IV is a free parameter containing expected volatility and everything else that affects option demand and supply, but it is not the model, making it very difficult to disentangle the different risk factors. As a results, all these risks are incorporated into the implied volatility in such a way that the skew can be seen as the view from the market that options with different strikes and different expiries have different risks and should be valued accordingly. That is, one expect a normal behaviour of the stock prices near ATM options which can be reasonably hedged, but when the stock prices exhibit large downward movements, the fear of non-headgeable jumps dominate. Bollen et al. [2004] showed that changes in the level of an option implied volatility were positively related to variation in demand for the option, and then argued that demand for out-of-the-money puts to hedge against stock market declines pushes up implied volatilities on low strike options in the stock index options market. Hence, the out-of-the-money (OTM) put prices have been viewed as an insurance product against substantial downward movements of the stock price and have been overpriced relative to OTM calls that will pay off only if the market rises substantially. For example, in the case of short maturity options where OTM put prices should have a zero market value, they actually exhibit positive values representing exclusively a market risk premium. As a result, the implicit distribution inferred from option prices is substantially negatively skewed compared to the lognormal distribution inferred from the Black-Scholes model. That is, given the Black-Scholes assumptions of lognormally distributed returns, the market assumes a higher return than the risk-free rate in the tails of the distributions.

1.4.2 Checking for no-arbitrage

The shape of the implied volatility surface is constrained by absence of arbitrage which are as follow

- we must have $\Sigma(K,T) \geq 0$ for all strikes K and maturity T.
- at any given maturity T, the skew can not be too steep (to avoid butterfly arbitrage).
- the term structure of implied volatility can not be too inverted (to avoid calendar spread arbitrage).

In this section we are going to introduce formally the no arbitrage conditions. Further, understanding the behaviour of the implied volatility (IV) surface for far expiries and far strikes is fundamental for extrapolation problems. Thus, various necessary conditions for an implied volatility surface to be properly defined have been proposed (see Durrleman [2003], Lee [2005]). We are going to briefly describe some well known results on the no-arbitrage constraints imposed on the IV surface and refer the readers to Fengler [2005] for more detailed results. Note, this section assumes deterministic interest rates and the underlying is not paying discrete dividends.

1.4.2.1 Some well known results

Given the underlying process $\{S_t; t \in [0,T]\}$, both American and European option prices can never be worth more than the spot price. Hence

$$\begin{aligned} C_A &\leq S_0 \\ C &\leq S_0 \end{aligned}$$

where C_A is an American call price and C is a European call price. Similarly, for American put price P_A and European put price, we get the upper bound

$$\begin{aligned} P_A &\leq K \\ P &\leq K \end{aligned}$$

More precisely, for European put option we get

$$P(t;K,T) \leq P(t,T)K$$

where $P(t,T)$ is a zero-coupon bond price. We get the following lower bound for European call options

$$C(t;K,T) \geq \big(S_t - P(t,T)K\big)^+$$

which also applies to American call options as they are more flexible than the European ones and are worth more. In the case of a European put option on non-dividend paying stock, we get the lower bound

$$P(t;K,T) \geq \big(P(t,T)K - S_t\big)^+$$

which also applies to American put options. As European call prices are monotonically decreasing functions of strike prices and puts are monotonically increasing functions of strike prices, we can get broad no-arbitrage bounds on the slope of the smile. First, given $K_1 \leq K_2$ for any expiry date T, we have

$$C(t;K_1,T) \geq C(t;K_2,T) \text{ and } P(t;K_1,T) \leq P(t;K_2,T)$$

or

$$\frac{P(t;K_1,T)}{K_1} \leq \frac{P(t;K_2,T)}{K_2}$$

Moreover, the call price must be a decreasing and convex function with respect to the strike K

$$-P(t,T) \leq \partial_K C(t;K,T) \leq 0 \text{ and } \partial_{KK} C(t;K,T) \geq 0$$

and similarly the put price must satisfies $\partial_K P(t;K,T) \geq 0$ and $\partial_{KK} P(t;K,T) \geq 0$. We now concentrate on the constraints imposed on the IV surface. Following the definition of the implied volatility in Section (1.4.1.2), we remove its dependency on the spot price and get

$$\Sigma : (t,K,T) \to \Sigma(t;K,T)$$

Further, we define the forward moneyness as $\eta = \frac{K}{F(t,T)}$, where $F(t,T)$ is the forward price seen at time t for the maturity T, and let the total variance be given by $\nu^2(\eta,T) = \Sigma^2(\eta,T)(T-t)$.

Proposition 3 *Assuming deterministic rates and dividend yield, if $\nu^2(\eta,T_i)$ is a strictly increasing function for $i = 1,2$ with $T_1 < T_2$, then there is no calendar arbitrage.*

Assuming the explicit dependence of volatility on strikes, we differentiate the market prices with respect to the strike getting

$$\partial_K C(t;K,T) = \partial_K C_{BS}(t;K,T) + \partial_\Sigma C_{BS}(t;K,T) \partial_K \Sigma(t;K,T) \leq 0$$

which gives

$$\partial_K \Sigma(t;K,T) \leq -\frac{\partial_K C_{BS}(t;K,T)}{\partial_\Sigma C_{BS}(t;K,T)}$$

Since the Vega is given by $\partial_\Sigma C_{BS}(t;K,T) = S_t \sqrt{\tau} n(d_1) = K e^{-r\tau} \sqrt{\tau} n(d_2)$, we write the upper bound as

$$\partial_K \Sigma(t;K,T) \leq \frac{N(d_2)}{K\sqrt{\tau} n(d_2)}$$

In the special case where $\eta = 1$, as $\tau \to 0$ then $N(d_2) \to \frac{1}{2}$ and $n(d_2) \to \frac{1}{\sqrt{2\pi}}$, such that

$$\lim_{\tau \to 0} \partial_K \Sigma(t;K,T)\big|_{K=F(t,T)} \leq \mathcal{O}(\tau^{-\frac{1}{2}})$$

but when the time to maturity goes to infinity, we get

$$\lim_{\tau \to \infty} \partial_K \Sigma(t;K,T)\big|_{K=F(t,T)} \leq \mathcal{O}(\tau^{-1})$$

Similarly, to get a lower bound we differentiate the put price and use the call-put parity to get

$$\partial_K \Sigma(t;K,T) \geq \frac{P_{BS}(t;K,T)/K - \partial_K P_{BS}(t;K,T)}{\partial_\Sigma P_{BS}(t;K,T)}$$

Using

$$e^{-r\tau} K N[d_2] = e^{-q\tau} S_t N[d_1]$$

where $\tau = T - t$, it simplifies to

$$-\frac{N[-d_1]}{\sqrt{\tau} K N[d_1]} \leq \partial_K \Sigma(t;K,T) \leq \frac{N[-d_1]}{\sqrt{\tau} K N[d_1]}$$

which can be rewritten in terms of the forward moneyness η as

$$-\frac{N[-d_1]}{\sqrt{\tau}\eta N[d_1]} = -\frac{1-N[d_2]}{\sqrt{\tau}\eta N[d_1]} \leq \partial_\eta \Sigma(t;K,T) \leq \frac{N[-d_1]}{\sqrt{\tau}\eta N[d_1]}$$

since $\partial_K \Sigma(t;K,T) = \frac{1}{F(t,T)}\partial_\eta \Sigma(t;\eta,T)$.

1.4.2.2 Limits for far strikes

Lee [2004] established the behaviour of the IVS as strikes tend to infinity. It has vital implications for the extrapolation of the IVS for far strikes. We quickly describe the results found by Lee for the large and small strike behaviour of the smile function. Considering the forward log-moneyness $x = \ln(\eta)$ (with no repo rate), Lee showed that

$$\Sigma(t;x,T) < \sqrt{\frac{2|x|}{T}} \tag{1.4.16}$$

for some sufficiently large $|x| > x^*$. To prove this result in the case of large strikes $x > x^*$, one must rely on the monotonicity of the BS formula in volatility. Hence, we get

$$C_{BS}(t;x,T,\Sigma(x,T)) < C_{BS}(t;x,T,\sqrt{\frac{2|x|}{T}})$$

For any call price function, we have

$$\lim_{x\to\infty} C(t;x,T) = \lim_{K\to\infty} e^{-r\tau} E_t[(S_T-K)^+] = 0$$

since given $E_t[S_T] < \infty$ we can interchange the limit and the expectation by the dominated convergence theorem. Similarly, applying L'Hopital's rule to the right hand side we get

$$\lim_{x\to\infty} C_{BS}(t;x,T,\sqrt{\frac{2|x|}{T}}) = e^{-r\tau}F(t,T)\big(N[0] - \lim_{x\to\infty} e^x N[-\sqrt{2|x|}]\big) = \frac{e^{-r\tau}}{2}F(t,T)$$

Further, Lee [2004] showed that there is a precise one-to-one correspondence between the asymptotic behaviour of the smile function and the number of finite moments of the distribution of the underlying S_T and its inverse $\frac{1}{S_T}$. The idea is to find a coefficient that can replace the 2 in Equation (1.4.16) with the number of finite moments in the underlying distribution of S_T and $\frac{1}{S_T}$. Hence, define

$$\tilde{p} = \sup\{p : E[S_T^{1+p}] < \infty\}$$
$$\tilde{q} = \sup\{q : E[S_T^{-q}] < \infty\}$$

and

$$\beta_R = \lim_{x\to\infty} \sup\{\frac{\Sigma^2(x,T)}{|x|/T}\}$$
$$\beta_L = \lim_{x\to-\infty} \sup\{\frac{\Sigma^2(x,T)}{|x|/T}\}$$

Here, the coefficients β_R and β_L can be interpreted as the slope coefficients of the asymptotes of the implied variance function. Lee showed that their values are $\beta_R, \beta_L \in [0,2]$ and

$$\tilde{q} = \frac{1}{2\beta_L} + \frac{\beta_L}{8} - \frac{1}{2}$$

That is, the IV smile must carry the same information as the underlying risk-neutral transition density. Hence, the asymptotic behaviour of the smile is shaped by the tail behaviour of the risk-neutral transition density and vice versa. Further, as options are bounded by moments, which can be interpreted as exotic options with payoffs, and moments are bounded by mixtures of a strike continuum of plain vanilla options, then the tail decay of the risk-neutral density determines the number of finite moments in the distribution. This has implications in the extrapolation of the IVS since the prices of exotic options depend significantly on the specific extrapolation. The results show that linear or convex skews for far strikes are wrong by the $\mathcal{O}(|x|^{\frac{1}{2}})$ behaviour. So, the smile (IV wings) should not grow faster than $\sqrt{x}$ and it should not grow slower than $\sqrt{x}$ unless we assume that S_T has finite moments of all orders.

1.4.2.3 Limits for far expiries

In order to generate meaningful stress scenarios for risk management purpose, we need to understand the dynamics of the IV over time. For long maturities, Tehranchi [2009] showed that the IV surface flattens for infinitely large expiries. Given the forward log-moneyness $\overline{\eta} = \ln \eta$ (with no repo rate), for any $M > 0$ he showed that

$$\lim_{T \to \infty} \sup_{\overline{\eta}_1, \overline{\eta}_2 \in [-M,M]} |\Sigma(\overline{\eta}_2, T) - \Sigma(\overline{\eta}_1, T)| = 0$$

This is to relate to the result found by Gatheral [1999] where the gradient of the implied volatility, if it exists, decays pointwise like $\frac{1}{T}$. More formally, Tehranchi's results made precise the rate of flattening of the IV skew from a theoretical stand point. They emphasised that the flattening of the implied volatility smile is a universal property of all martingale models. For example, one of the inequality states that there exists a martingale $(S_t)_{t \geq 0}$ with $S_t \to 0$ as $t \to \infty$ in probability, such that

$$T \partial_{\overline{\eta}} \Sigma^2(\overline{\eta}, T) \to -4$$

as $T \to \infty$ uniformly for $\overline{\eta} \in [-M, M]$. Hence, given the total variance $\nu^2(\overline{\eta}, T) = \Sigma^2(\overline{\eta}, T)T$, the implied variance skew must be bounded by $|\partial_{\overline{\eta}} \nu^2| \leq 4$ and it should decay at a rate of $\frac{1}{T}$ between expiries.
Further, the literature on the modelling of the stochastic dynamics of the Black-Scholes IV being flourishing, to prove that the IV surface can not move by parallel shifts, as the shape must also change, Rogers et al. [2010] derivied certain model-independent properties of the implied volatility surface. It was originally a conjecture established by Ross [2006]. It represents an important result showing that one can not blindly impose dynamics on the IV surface, such as moving up or down the IVS by parallel shifts, as it may leads to inconsistency. Given the forward moneyness $\eta = \frac{K}{x} e^{-(r-q)(T-t)}$ and taking its logarithm $\overline{\eta} = \log \eta$, they showed that if there exists a process $(\xi_t)_{t \geq 0}$ such that for all $t \geq 0$, $(T-t) > 0$

$$\Sigma(t; \overline{\eta}, T) = \Sigma(0; \overline{\eta}, T) + \xi_t \tag{1.4.17}$$

then $\xi_t = 0$ almost surely for all $t \geq 0$. To do so, they showed that the long IV can not fall, such that if the conjecture holds then ξ is non-decreasing. Then, they proved that if the IV surface moves in parallel shifts, the surface must be constant.

1.4.3 Defining no-arbitrage in complete markets

1.4.3.1 Arbitrage-free conditions: The call price surface

In Section (1.4.2), we discussed the no-arbitrage conditions and introduced constraints imposed on the implied volatility surface (IVS), but we did not consider necessary and sufficient conditions for an IVS to be free from static arbitrage. There are several ways of defining static arbitrage, all very close from one another. Durrleman [2003] presented a number of necessary restrictions for call options, establishing sufficiency under very strong conditions. Davis et al. [2007] showed that absence of arbitrage is normally defined in relation to a specific model of market prices. A set of option prices is consistent with absence of arbitrage if there exist a model M and an equivalent martingale measure $\mathbb{Q}$

such that the price of an option is the discounted expected value of the payoff under the measure $\mathbb{Q}$. We first follow the definition given by Roper [2010].

Definition 1.4.3 *Arbitrage-Free Call Price Surface*
A call option price surface $(K,T) \to C(K,T)$ is free from static arbitrage if and only if there exists a non-negative martingale, say S, such that

$$C(K,T) = E[e^{-rT}(S_T - K)^+]$$

for every K, and $T \geq 0$.

We can then give necessary and sufficient conditions for a call price surface to be free of static arbitrage (see Follmer et al. [2004] and Roper [2010]).

Theorem 1.4.1 *Let $s > 0$ be a constant, T be the maturity, and let $C : (0,\infty) \times [0,\infty) \to \mathbb{R}$ satisfy the following conditions:*

(A1) (Convexity in K)
$C(\cdot,T)$ *is a convex function,* $\forall T \geq 0$

(A2) (Monotonicity in T)
$C(K,\cdot)$ *is non-decreasing,* $\forall K \geq 0$

(A3) (Large Strike Limit)
$\lim_{K\to\infty} C(K,T) = 0$, $\forall T \geq 0$

(A4) (Bounds)
$(s - e^{-rT}K)^+ \leq C(K,T) \leq s$, $\forall K > 0, T \geq 0$

(A5) (Expiry Value)
$C(K,0) = (s-K)^+$, $\forall K > 0$

Then, there exists a non-negative Markov martingale X, with the property that

$$C(K,T) = \mathbb{E}[e^{-rT}(X_T - K)^+] \, , \forall K \text{ and } \forall T \geq 0$$

As a result, Roper argued that an implied volatility surface is free from static arbitrage if the call price surface defined by Equation (1.4.14) is free from static arbitrage.

Remark 1.4.1 *The implied volatility surface being a mapping from Black-Scholes prices, necessary and sufficient conditions for the surface to be free from static arbitrage must be defined in terms of the properties and limits of the Black-Scholes formula.*

Remark 1.4.2 *To infer static arbitrage from implied volatility surface, one must first establish necessary and sufficient conditions on the call price surface for it to be free of static arbitrage, and then translate these conditions into conditions on the implied volatility surface.*

We state the theorem given by Roper [2010] that is sufficient to ensure that the implied volatility surface remains free from static arbitrage. In that setting, we get $x = \ln\left(\frac{K}{F(t,T)}\right)$ and we define the time scaled implied volatility as

$$\begin{aligned} \Xi : \mathbb{R} \quad \times \quad & [0,\infty) \to [0,\infty] \\ & (x,T) \to \sqrt{T-t}\,\Sigma(F(t,T)e^x,T) \end{aligned} \tag{1.4.18}$$

where $F(t,T)$ is the forward price. That is, $\Xi(x,T) = \sqrt{\omega(\overline{\eta},T)}$. Then, the theorem is as follow:

Theorem 1.4.2 *Let $F(t,T) > 0$, T be the maturity, and let $\Xi : \mathbb{R} \times [0,\infty) \to \mathbb{R}$. Let Σ satisfies the following conditions:*

(IV1) (Smoothness)
$\forall T > 0$, $\Xi(\bullet, T)$ *is twice differentiable.*

(IV2) (Positivity)
$\forall x \in \mathbb{R}$ *and* $T > 0$, $\Xi(x,T) > 0$.

(IV3) (Durrleman Condition)
$\forall x \in \mathbb{R}$ *and* $T > 0$,

$$\big(1 - \frac{x\partial_x \Xi}{\Xi}\big)^2 - \frac{1}{4}\Xi^2(\partial_x \Xi)^2 + \Xi \partial_{xx}\Xi \geq 0 \quad (1.4.19)$$

(IV4) (Monotonicity in T)
$\forall x \in \mathbb{R}$, $\Xi(x, \bullet)$ *is non-decreasing.*

(IV5) (Large-Moneyness Behaviour)
$\forall T > 0$, $\lim_{x \to \infty} d_1(x, \Xi(x,T)) = -\infty$.

(IV6) (Value at Maturity)
$\forall x \in \mathbb{R}$, $\Xi(x,0) = 0$.

Then the call price surface parameterised by $F(t,T)$ is free from static arbitrage. In particular, there exists a non-negative Markov martingale X, with the property that

$$C(K,T) = \mathbb{E}[e^{-rT}(X_T - K)^+ | X_0 = F(0,T)] \text{ , } \forall K \text{ and } \forall T \geq 0$$

1.4.3.2 From price arbitrage to implied volatility arbitrage

Originally, the implied volatility surface was defined as free from static arbitrage if it is void from both calendar-spread and butterfly-spread artibitrage (see Carr et al. [2005b], Cox et al. [2005]). It motivated the following definition:

Definition 1.4.4 *An implied volatility surface is free from static arbitrage if and only if the following conditions are satisfied.*

(V1) For a given forward moneyness, η, the surface is free from calendar-spread arbitrage.

(V2) For each time-to-maturity, the surface is free from butterfly-spread arbitrage.

The butterfly-spread arbitrage ensures the existence of a (non-negative) implied risk-neutral density for the underlying, whilst calendar-spread arbitrage implies the monotonicity of option prices with respect to maturity. The latter motivates the following definition:

Proposition 4 *Assuming deterministic rates and dividend yield, if $\omega(\eta, T_i)$ is a strictly increasing function for $i = 1,2$ with $T_1 < T_2$, then there is no calendar arbitrage.*

or quivalently,

Definition 1.4.5 *Assuming deterministic rates and proportional dividend yield, the IVS is free of calendar spread arbitrage when*

$$\forall x \text{ and } T > 0 \text{ , } \partial_T \omega(x,T) \geq 0$$

Note, this is the numerator of the local volatility expressed in terms of the implied volatility. In the special case where we assume that the implied volatility neither depends on the underlying S_t nor on the strike, we get

$$\omega(T) = \Sigma^2(T)T = \int_0^T \sigma^2(s)ds$$

so that the implied volatility is equal to the quadratic mean of the local volatility during the life of the option. In that case, the local volatility does not depend on the strike price, and after differentiating with respect to maturity, it must satisfy

$$\sigma^2(T) = \Sigma^2(T) + 2\Sigma(T)T\partial_T\Sigma(T) \tag{1.4.20}$$

Thus, from the definition of static arbitrage (see Cox et al. [2005]), call prices must be decreasing and convex functions of the strike leading to non-negative density (absence of butterfly arbitrage). In addition, calendar spread arbitrage is expressed as the monotonicity of European call option prices with respect to the maturity. Hence, Gatheral et al. [2012] defined the absence of calendar spread arbitrage and that of butterfly arbitrage by the non-negativity of the numerator and denominator of Dupire's equation in implied volatility (see Equation (1.7.42)). However, these properties are necessary but not sufficient for no-arbitrage. The right statement is that there exists an inhomogeneous diffusion process associated to the implied volatility through Dupire's equation. To this end, the local volatility defined through Dupire's equation should have adequate continuity and growth properties. One way forward is to consider additional conditions such as the Large-Moneyness Behaviour (LMB) defined by Roper [2010] as

$$\lim_{\overline{\eta}\to\infty} d_1(\overline{\eta}, \omega(\overline{\eta}, t)) = -\infty \tag{1.4.21}$$

where $\overline{\eta} = \log \eta$ and $d_1(\eta, \omega(t))$ is given in Equation (1.2.10). This condition is equivalent to call option prices tending to zero as the strike tends to infinity, which is the Large Strike Limit in Theorem (1.4.1). Note, the price formulation of arbitrage freeness given by Roper [2010] in Theorem (1.4.1) is minimal in the sense that the regularity conditions on the Call option prices are necessary and sufficient: to be convex in the strike direction and non-decreasing in the maturity direction. However, Roper assumed that the total variance was twice differentiable in the strike direction. This regularity is certainly not required. Guo et al. [2012] showed that by starting from option prices convex in K, their first derivative are defined almost everywhere, and so are those of the total variance (in K or η), since the Black-Scholes mapping in total variance is smooth. As a result, one can assume that for any t, the function $k \to w(\overline{\eta}, t)$ is continuous and almost everywhere differentiable.

In addition, some other asymptotic behaviours of d_1 and d_2 hold in great generality. For instance, the Small-Moneyness Behaviour (SMB) condition, proved by Rogers et al. [2010], states

$$\lim_{\overline{\eta}\to-\infty} d_2(\overline{\eta}, \omega(\overline{\eta}, t)) = \infty \tag{1.4.22}$$

Further, Rogers et al. [2010] imposed two further constraints on the LMB and SMB, obtaining the following Lemma:

Lemma 1.4.1 *Let ω be any positive real function. Then*

$$\lim_{\overline{\eta}\to\infty} d_2(\overline{\eta}, \omega(\overline{\eta})) = -\infty$$
$$\lim_{\overline{\eta}\to-\infty} d_1(\overline{\eta}, \omega(\overline{\eta})) = \infty$$

Equivalently, using Equation (1.4.22), Guo et al. [2012] showed that the marginal law of the stock price at some fixed time $t > 0$ has no mass at zero if and only if

$$\lim_{K\to 0} \partial_K C(K, t) = -P(0, t) \tag{1.4.23}$$

which is a statement about a Small-Moneyness Behaviour. These conditions were then recasted in terms of implied volatility providing a complete characterisation of an IVS free from static arbitrage.

Definition 1.4.6 *Arbitrage-Free Implied Volatility Surface*
Assuming deterministic interest and repo rates, an implied volatility surface is free from static arbitrage if and only if the following conditions are satisfied

(A1) *(Monotonicity of total variance)*
$\forall x$ *and* $T > 0$, $\partial_T \omega(x, T) \geq 0$

(A2) *(Convexity in K)*
$\forall T > 0$, $\partial_{KK} C(K, T) \geq 0$

(A3) *(Large-Moneyness Behaviour)*
$\lim_{\overline{\eta} \to \infty} d_1(\overline{\eta}, \omega(\overline{\eta}, t)) = -\infty$ *and* $\lim_{\overline{\eta} \to \infty} d_2(\overline{\eta}, \omega(\overline{\eta})) = -\infty$

(A4) *(Small-Moneyness Behaviour)*
$\lim_{\overline{\eta} \to -\infty} d_2(\overline{\eta}, \omega(\overline{\eta}, t)) = \infty$ *and* $\lim_{\overline{\eta} \to -\infty} d_1(\overline{\eta}, \omega(\overline{\eta})) = \infty$

1.4.4 Computing the implied volatility surface

Rather than estimating the historical volatility to plug in the BS-formula, Latane et al. [1976] showed that we could invert the process and infer an implied volatility from observed market prices. This is the volatility of the underlying that is being used by market participants to price and trade options. As the option market developed, the Black-Scholes model implied multiple volatilities for the same underlying asset, producing a smile. Various explanations have been discussed extensively in the literature. Assuming a model for the dynamics of the prices, in order to calibrate the model parameters to market prices we need to estimate the Black-Scholes volatility a large number of time, quickly and precisely. To do so we need to solve the nonlinear Equation (1.2.6) with respect to σ given the market price. However, there is no closed-form solution for the implied volatility, and Gerhol [2012] formally showed that the BS-model did not belong to a class of functions for which solutions could easily be found. Numerical solutions can be obtained by iterative algorithms using the Newton approximation (see Section (1.4.4.2). However, iterative algorithms having many shortages such as explosion when the gradient tends to zero, it lead researchers to develop formulas to find an approximated implied volatility. We are going to discuss both closed-form approximations and iterative methods.

1.4.4.1 Closed-form approximations

While the BS-formula has been considered in the commodity market, in the equity market, to account for dividend yield and repo rate, we let the stock price $\{S_t^x\}$ always be multiplied by the repo factor $Re(t, T)$. Curtis et al. [1988] considered the special case where the discounted strike price $\overline{K} = KP(t, T)$ is exactly equal to the present stock price times the repo bond $Re(t, T)$. In that setting the BS-formula becomes

$$C_{BS}(t, x, \frac{xRe(t, T)}{P(t, T)}, T) = xRe(t, T)\big(2N(\frac{\sigma\sqrt{T - t}}{2}) - 1\big)$$

which can be solved for the volatility as

$$\sigma = \frac{2}{\sqrt{T - t}} N^{-1}(\frac{C_{BS} + xRe(t, T)}{2xRe(t, T)})$$

where $N^{-1}(\bullet)$ the inverse standard normal distribution. In that setting, Brenner et al. [1988] used a first order Taylor polynomial approximation for the inverse standard normal distribution around $\frac{1}{2}$, and obtained the volatility approximation

$$\sigma \approx \sqrt{\frac{2\pi}{T-t}} \frac{C_{BS}}{xRe(t,T)}$$

Later, considering a quadratic approximation to the implied volatility and using a higher order Taylor expansion on the normal distribution function, Corrado et al. [1996b] extended that result to cases where stock prices deviate from discounted strike prices. The quadratic equation is

$$q_1 \sigma^2 + q_2 \sigma + q_3 = 0$$

where

$$\begin{aligned} q_1 &= (T-t)(xRe(t,T) + \overline{K}) \\ q_2 &= \big(\frac{xRe(t,T) - \overline{K}}{2} - C_{BS}\big)\sqrt{8\pi(T-t)} \\ q_3 &= 2\big(xRe(t,T) - \overline{K}\big) \ln \frac{xRe(t,T)}{\overline{K}} \end{aligned}$$

with solutions given by

$$\sigma = \frac{-q_2 \pm \sqrt{q_2^2 - 4q_1 q_3}}{2q_1} = \frac{-q_2}{2q_1} \pm \sqrt{\frac{q_2^2}{4q_1^2} - \frac{q_3}{q_1}}$$

The largest root is

$$\begin{aligned} \sigma &\approx \sqrt{\frac{2\pi}{T-t}} \frac{1}{xRe(t,T) + \overline{K}} \Big[C_{BS} - \frac{xRe(t,T) - \overline{K}}{2} \\ &+ \sqrt{\big(C_{BS} - \frac{xRe(t,T) - \overline{K}}{2}\big)^2 - 2(xRe(t,T) + \overline{K})(xRe(t,T) - \overline{K}) \ln \frac{xRe(t,T)}{\overline{K}}} \Big] \end{aligned}$$

Setting $\overline{\eta} = \ln \frac{xRe(t,T)}{\overline{K}} = 2\frac{xRe(t,T) - \overline{K}}{xRe(t,T) + \overline{K}}$, the accuracy of the quadratic formula is significantly improved by minimising its concavity. The resulting approximated volatility is

$$\sigma \approx \sqrt{\frac{2\pi}{T-t}} \frac{1}{xRe(t,T) + \overline{K}} \Big[C_{BS} - \frac{xRe(t,T) - \overline{K}}{2} + \sqrt{\big(C_{BS} - \frac{xRe(t,T) - \overline{K}}{2}\big)^2 - \frac{(xRe(t,T) - \overline{K})^2}{\pi}} \Big]$$

For short maturities and strikes away from the money, the terms inside the square root may become negative. The approximation is accurate enough in the domain $\eta \in [0.9, 1, 1]$ for maturities greater than three months, and it is accurate in the domain $\eta \in [0.95, 1, 05]$ if the maturity is longer than one month. Chance et al. [2014] considered the same quadratic equation as above with

$$\begin{aligned} q_1 &= \frac{xRe(t,T) + \overline{K}}{2} \\ q_2 &= \big(\frac{xRe(t,T) + \overline{K}}{2} - C_{BS}\big)\sqrt{2\pi(T-t)} \\ q_3 &= \big(xRe(t,T) + \overline{K}\big) \ln \frac{xRe(t,T)}{\overline{K}} \end{aligned}$$

and showed that the larger root σ^+ corresponds to the solution proposed by Corrado et al. [1996b]. The smaller root can be positive when $q_2 < 0$, which is equivalent to $\frac{xRe(t,T)-\overline{K}}{2} < C_{BS}$. Isengildina-Massa et al. [2007] studied various implied volatility approximations and concluded that the method of Corrado et al. yields relatively more accurate results than other methods when call premiums are used. Increasing the range of the domain, Liang et al. [2009] gave under certain conditions an approximation of the implied volatility with an error estimate for it. They first averaged the Black-Scholes terms d_i for $i = 1, 2$, to d_3, and performed a Taylor expansions of $N(d_1)$ and $N(d_2)$ around d_3. That is, assuming d_3 close to zero and $\frac{\xi}{|d_3|} \leq 1$, or equivalently $\eta \in [0.1, 2]$, they expanded the pricing formula given in Equation (B.1.2), getting rid of the factor $\frac{1}{1-\eta}$ and decreasing the approximation error. Then, they performed a Taylor expansions of $N(d_3)$ and $N'(d_3)$ around zero and approximated the pricing formula in Equation (B.1.2) with $\widehat{C}_{BS}(t, x, K, T) = g_3(d_3(T-t, xC(t,T), K))$ where

$$g_3(d_3(T-t, xC(t,T), K)) = \frac{1}{2} + \frac{4-m}{4\sqrt{2\pi}} d_3 + \frac{m}{2\sqrt{2\pi}} \frac{1}{d_3}$$

The function $g_3(\bullet)$ being an approximation of $g_1(\bullet)$, its solution $\tilde{d}_3$, obtained by solving a quadratic equation, is itself an approximation to d_3 given by

$$\tilde{d}_3 = \begin{cases} \frac{-\sqrt{2\pi}(1-2\widehat{C}_{BS}(t,x,K,T)) - \sqrt{2\pi(1-2\widehat{C}_{BS}(t,x,K,T))^2 - 2m(4-m)}}{4-m} & \text{if } 0.1 < \eta < 1 \\ \frac{-\sqrt{2\pi}(1-2\widehat{C}_{BS}(t,x,K,T)) + \sqrt{2\pi(1-2\widehat{C}_{BS}(t,x,K,T))^2 - 2m(4-m)}}{4-m} & \text{if } 1 < \eta < 2 \end{cases}$$

Since $\sigma = -\frac{\log \eta}{\sqrt{T-t}} \frac{1}{d_3}$ the approximated implied volatility $\tilde{\sigma}(T, K)$ is given by

$$\tilde{\sigma}(T, K) = -\frac{\log \eta}{\sqrt{T-t}} \frac{1}{\tilde{d}_3(T-t, xC(t,T), K)} \tag{1.4.24}$$

This formula being an approximation to the implied volatility, it may not be accurate enough for certain strikes and maturities. Therefore we choose to use the iterative algorithm discussed in Section (1.4.4.2), but with the first guess given by Equation (1.4.24).

1.4.4.2 Iterative methods

Even though there exists different iterative methods taking an initial estimate, or interval of estimates, to infer the implied volatility, the financial industry favours the Newton-Raphson method (see Chriss [1997]). Chance et al. [2014] showed that computational considerations, such as the tolerence level and the initial input, have a significant influence on the existence and shape of the volatility smile. When we let the Black-Scholes volatility in Equation (1.2.6) tends respectively to zero and infinity, we get the limits

$$\begin{aligned} \lim_{\sigma \to 0} C_{BS}(\sigma) &= (xRe(t,T) - KP(t,T))^+ \\ \lim_{\sigma \to \infty} C_{BS}(\sigma) &= xRe(t,T) \end{aligned}$$

Therefore, the Black-Scholes formula is strictly increasing in the volatility. Hence, if an approximated solution exists, it is unique by the strictly positive vega. Moreover, the Volga in Equation (B.1.8) shows that the function $\sigma \to C_{BS}(\sigma)$ is convex on the interval $[0, \sigma_{rf}]$ and concave on the interval $[\sigma_{rf}, \infty]$, where

$$\sigma_{rf} = \sqrt{2\frac{|-\log \eta|}{T-t}}$$

is the reflection point of $f(\sigma)$ derived by Manaster et al. [1982]. Therefore, the equation $C_{BS}(\sigma) = C_M$, where C_M the market call price in the range

$$(xRe(t,T) - KP(t,T))^+ < C_M(t,T) < xRe(t,T)$$

can be solved iteratively with the Newton algorithm (see Press et al. [1992]). Setting $f(\sigma) = C_M - C_{BS}(\sigma)$ and noting that $f^{'}(\sigma) = -\partial_\sigma C_{BS}(\sigma)$, we get

$$\begin{aligned}
\sigma_0 &= \sigma_{rf} \\
\sigma_{n+1} &= \sigma_n + (C_M - C_{BS}(\sigma_n)) \max\big(\frac{1}{Vega}, 100\big)
\end{aligned}$$

where $Vega = \partial_\sigma C_{BS}(\sigma)$, and $\max(.,.)$ is used to avoid explosion when the vega gets too small, and σ_0 is the first guess of the algorithm since the series $(\sigma_n) \geq 0$ is monotonic. The NR algorithm terminates when the iterations yield an estimated price that is within a preset tolerance of the market price. With initial guess σ_{rf}, one can show that if the NR algorithm converges, it must converge to the correct solution. However, the algorithm itself can not guarantee convergence (see Chance et al. [2014]). We can also calculate the second order derivative terms and use a higher order iterative root finding algorithm such as the Halley algorithm. One can show that $\frac{\partial^2 f(\sigma)}{\partial \sigma^2} > 0$ provided that $0 < \sigma < \sigma_{rf}$ and that $\frac{\partial^2 f(\sigma)}{\partial \sigma^2} < 0$ provided that $\sigma > \sigma_{rf}$. Note, even if $f(\sigma)$ is differentiable everywhere, the NR method may not lead to a root. When the strike prices are far from the money or the price is close to the no-arbitrage bounds, the vega becomes insignificant and the above ratio explode. In that case one must switch to the Bisection method. Also, as discussed by Jackel [2006], this algorithm is very sensitive to its initial guess and the volatility can easily be in the range $[10^{-4}\%, 1000\%]$. Recognising the near flat shape of the normalised Black function for small volatilities, to enlarge the space domain in that region $(\sigma < \sigma_0)$, he performed a change of variable expressing the prices in the logarithm space. In that setting the initial guess must be modified and asymptotic expansion is performed on the prices around $\sigma \to 0$ to calculate the new initial guess. A similar approach is used for $\sigma \to \infty$ and interpolation is performed between the two initial guesses. Alternatively, as explained in Section (1.4.4.1), we can estimate the initial guess by Taylor expanding the normal cumulative integral in the Black-Scholes formula.

1.5 Mathematical representation of some option strategies

1.5.1 The Digital Bond and Digital Stock

The payoff of a call option is a function given by

$$C(S_t) = (S_t - K)^+ = (S_t - K) I_{\{S_t \geq K\}}$$

where $C(S_t)$ is a function of the stock price S_t. Differentiating that payoff with respect to the stock price, we get

$$\begin{aligned}
\frac{\partial C(S_t)}{\partial S_t} &= I_{\{S_t \geq K\}} \\
\frac{\partial^2 C(S_t)}{\partial S_t^2} &= \frac{\partial \mathcal{H}}{\partial S_t} = \delta(S_t - K)
\end{aligned}$$

where $I_{\{S_t \geq K\}} = \mathcal{H}(S_t - K)$ is the Heaviside function and $\mathcal{H}^{'} = \delta(S_t - K)$ is its derivative the Dirac function. We can therefore calculate the prices of some special contingent claims via the derivation of the call price $C(t, S_t, T, K)$ as follow

$$\begin{aligned}
\text{Digital} &= -\frac{\partial C(t,S_t,T,K)}{\partial K} = P(t,T) E_t[I_{\{S_T \geq K\}}] \\
\text{Density} &= \frac{\partial^2 C(t,S_t,T,K)}{\partial K^2} = P(t,T) E_t[\delta(S_T - K)]
\end{aligned}$$

From the definition of a digital option we can deduce that

$$P(S_T \geq K) = -\frac{1}{P(t,t)} \frac{\partial C(t,S_t,T,K)}{\partial K}$$

and the cumulative distribution becomes

$$P(S_T < K) = 1 + \frac{1}{P(t,t)} \frac{\partial C(t,S_t,T,K)}{\partial K} \tag{1.5.25}$$

As a result, the price of a call option can be expressed in terms of those quantities as

$$\begin{aligned} C(t,S_t,T,K) &= P(t,T)E_t[(S_T-K)^+] = P(t,T)E_t[(S_T-K)I_{\{S_T \geq K\}}] \\ &= P(t,T)\big(E_t[S_T I_{\{S_T \geq K\}}] - KE_t[I_{\{S_T \geq K\}}]\big) \end{aligned}$$

which gives

$$\frac{C(t,S_t,T,K)}{P(t,T)} = E_t[S_T I_{\{S_T \geq K\}}] - KE_t[I_{\{S_T \geq K\}}] = \Delta_S + K\Delta_K \tag{1.5.26}$$

where Δ_K is the probability that the stock price end up higher or equal to the strike price at maturity, and Δ_S is a modified probability that the stock price end up higher or equal to the strike price at maturity. Following Ingersoll [1998] [2000], we call $D_B(S,t,T;\xi)$ the Digital Bond and $D_S(S,t,T;\xi)$ the Digital Stock for the event ξ. The former is the value at time t of receiving one dollar at the maturity T if and only if a probabilist event ξ occurs, while the latter is the value at time t of receiving one share of the stock at the maturity T (excluding any intervening dividends) if and only if a probabilist event ξ occurs. In the special case where $\xi = \{S_T > K\}$, we have $D_B(S,t,T;\xi) = P(t,T)\Delta_K$ and $D_S(S,t,T;\xi) = P(t,T)\Delta_S$. Geman et al. [1995] showed that $\Delta_K = P(S_T \geq K)$ and $\Delta_S = P^S(S_T \geq K)$, where the latter is a probability under the spot measure with the stock price as numeraire. We can rewrite Δ_S as

$$E_t[S_T I_{\{S_T \geq K\}}] = \frac{C(t,S_t,T,K)}{P(t,T)} + KE_t[I_{\{S_T \geq K\}}]$$

From the definition of the digital option, the above expression becomes

$$E_t[S_T I_{\{S_T \geq K\}}] = \frac{C(t,S_t,T,K)}{P(t,T)} - K\frac{1}{P(t,T)} \frac{\partial C(t,S_t,T,K)}{\partial K}$$

and we see that the Digital Stock satisfies

$$\Delta_S = \frac{1}{P(t,T)} \frac{C(t,S_t,T,K)}{S_t} - \frac{K}{S_t} \frac{1}{P(t,T)} \frac{\partial C(t,S_t,T,K)}{\partial K}$$

which can be statically replicated with a call option and a call-spread. Therefore, the pricing of other European derivatives with piecewise linear and path-independent payoffs only requires valuing Digital Bond and Digital Share with event $\xi = \{L < S_T < H\}$ for some constants L and H. For example, the call option price is

$$C(t,T,K) = D_S(S,t,T;S_T > K) - KD_B(S,t,T;S_T > K) \tag{1.5.27}$$

while the put option price is

$$P(t,T,K) = KD_B(S,t,T;S_T < K) - D_S(S,t,T;S_T < K)$$

In the special case where the rate, repo and volatility are constants, Cox and Ross [1976] showed that the Digital Bond and Digital Share with event $\xi = \{S_T > K\}$, where K is the strike price, could be valued under the risk-neutral measure $\mathbb{Q}$ with the Black-Scholes formula, that is

$$
\begin{aligned}
D_B(S,t,T;\xi) &= E^Q[e^{-\int_t^T r_s ds} I_{S_T>K}|\mathcal{F}_t] = e^{-r(T-t)} N(d_2(T-t, S_t e^{(r-q)(T-t)}, K)) \\
D_S(S,t,T;\xi) &= E^Q[e^{-\int_t^T r_s ds} S_T I_{S_T>K}|\mathcal{F}_t] = S_t e^{-q(T-t)} N(d_1(T-t, S_t e^{(r-q)(T-t)}, K))
\end{aligned}
$$

where

$$
d_2(t,x,y) = \frac{1}{\sigma\sqrt{t}} \log \frac{x}{y} - \frac{1}{2}\sigma\sqrt{t} \text{ and } d_1(t,x,y) = d_2(t,x,y) + \sigma\sqrt{t}
$$

However, when the volatility of the stock price is stochastic, and more generally when the instantaneous volatility of the stock price, the spot rate and repo rate are stochastic, one can no-longer use the Black-Sholes formula. Under general Markov processes for the model parameters, the conditional probabilities of the Digital Bond and Digital Share are difficult to solve analytically under any probability measure, and numerical tools must be used.

1.5.2 Some properties of market prices

1.5.2.1 The convexity of market prices

Since the financial markets use the Black-Scholes formula for pricing standard option contracts, a simple way of understanding the properities of linear combination of European options in the BS world is to assume that option prices are close to at-the-money (ATM) forward, and to linearise them by using Equation (1.2.11). To illustrate this approach, we consider an initial spot price $S_0 = 100$, a maturity $T = 0.5$, an interest rate $r = 0.03$, and a volatility $\sigma = 0.3$, and we display in Figure (1.2) the call price and its linear approximation versus different strike levels. Using the same settings, we display in Figure (1.3) the put price and its linear approximation versus different strike levels. The level of convexity of market prices depends on interest rate, repo rate, time to maturity, and volatility. As we move away from the ATM-forward strike, the convexity of market prices increases.

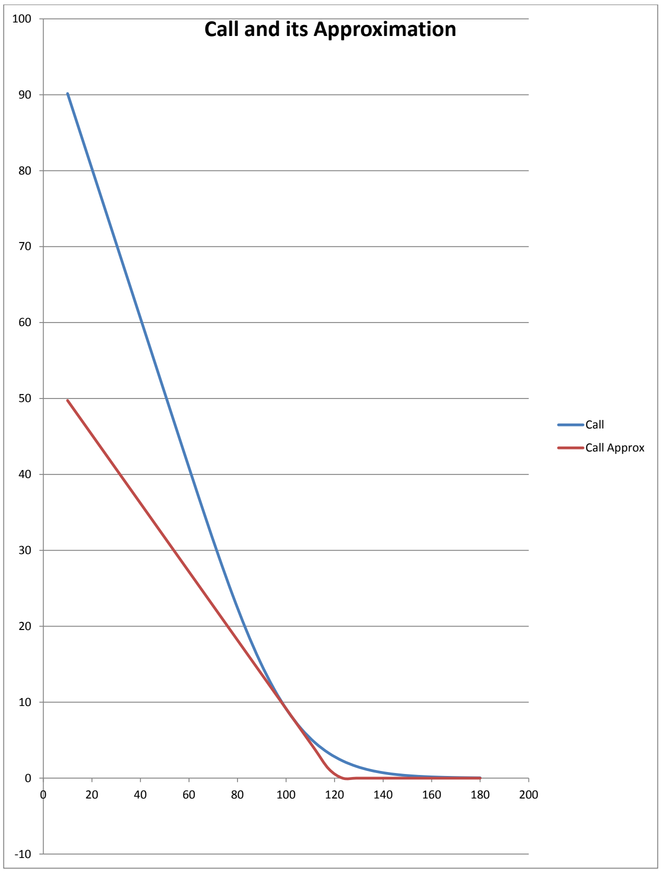


Figure 1.2: The call price and its linear approximation vs strikes.

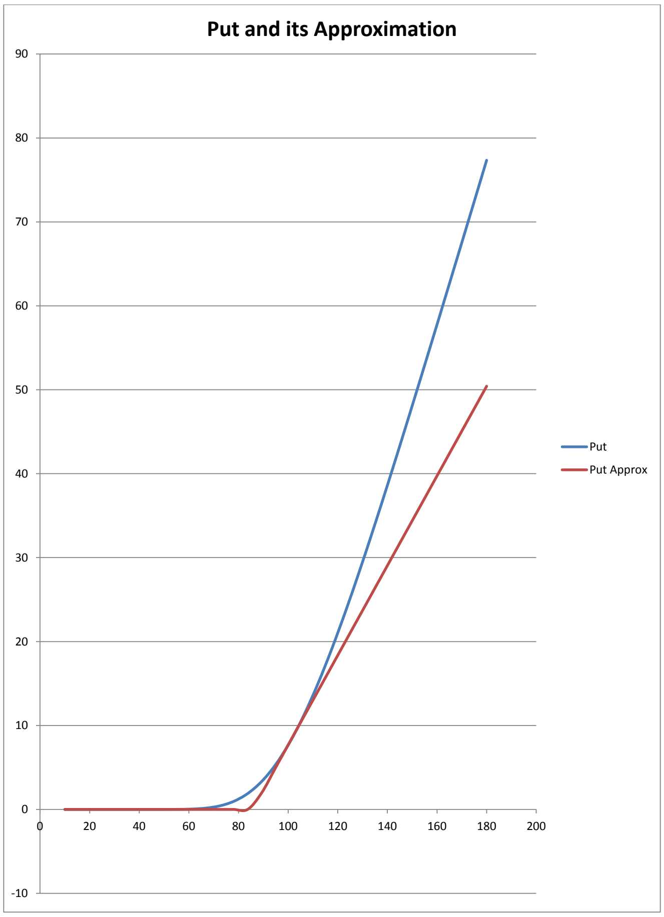


Figure 1.3: The put price and its linear approximation vs strikes.

1.5.2.2 The digital option

Given the stock price $(S_t)_{t \geq 0}$, a Digital option $D(K,T) = P(t,T)\Delta_K$ for strike K and maturity T pays \$1 when the stock price S_T is greater than the strike K, and zero otherwise. From a static replication argument, the price of the Digital option is given by

$$D(K,T) = \lim_{\Delta K \to 0} \frac{C(K,T) - C(K + \Delta K, T)}{\Delta K} = -\frac{\partial}{\partial K} C(K,T)$$

where $C(K,T)$ is a market call option. Hence, the digital options are uniquely priced from the volatility surface. Given $C(K,T) = C_{BS}(K,T;\Sigma_{BS}(K,T))$ where $\Sigma_{BS}(K,T)$ is the BS implied volatility for strike K and maturity T, and using the chain rule, the Digital option becomes

$$D(K,T) = -\frac{\partial}{\partial K} C_{BS}(K,T;\Sigma_{BS}(K,T)) = -\frac{\partial}{\partial K} C_{BS}(K,T;\Sigma_{BS}) - \frac{\partial}{\partial \Sigma} C_{BS}(K,T;\Sigma(K,T)) \frac{\partial}{\partial K} \Sigma(K,T)$$

We can express the Digital option in terms of the Vega and the Skew as

$$D(K,T) = -\frac{\partial}{\partial K} C_{BS}(K,T;\Sigma_{BS}) - Vega(K,T) Skew(K,T) \quad (1.5.28)$$

where $Vega(K,T;\Sigma_{BS}(K,T))$ is the Black-Scholes vega for the strike K and maturity T, and $\partial_K C_{BS}(K,T;\Sigma_{BS})$ is the BS digital price for the volatility $\Sigma_{BS}(K,T)$. In the special case where $r = q = 0$, $T = 1$ and for $S_0 = 100$ and $K = 100$, then $\eta = 1$ and we get $d_2 = -\frac{1}{2}\Sigma_{BS}\sqrt{T}$. Given a skew of 2.5% per 10% change in the strike and an ATM volatility $\Sigma_{ATM} = 25\%$, the Digital option simplifies to

$$\begin{aligned} D(100,1) &= N[-\frac{\Sigma_{ATM}}{2}] - S_0 n[\frac{\Sigma_{ATM}}{2}] \frac{-0.025}{0.1 S_0} \\ &\approx 0.45 + 0.25 \times 0.4 = 0.55 \end{aligned}$$

Ignoring the skew, the price is 45% of notional which is significantly lower than 55% of notional when the skew is included.

1.5.2.3 The butterfly option

Assuming that the volatility surface has been constructed from European option prices, we consider a butterfly strategy centered at K where we are long a call option with strike $K - \Delta K$, long a call option with strike $K + \Delta K$, and short two call options with strike K. The value of the butterfly for strike K and maturity T is

$$B(t_0,K,T) = C(K-\Delta K,T) - 2C(K,T) + C(K+\Delta K,T) \approx P(t_0,T)\phi(t_0;K,T)(\Delta K)^2$$

where $\phi(t_0;K,T)$ is the probability density function (PDF) of S_T evaluated at strike K. As a result, we have

$$\phi(t_0;K,T) \approx \frac{1}{P(t_0,T)} \frac{C(K-\Delta K,T) - 2C(K,T) + C(K+\Delta K,T)}{(\Delta K)^2}$$

where $A_{t_0}(K,T) = P(t_0,T)\phi(t_0;K,T)$ can be seen as the price at time t_0 of a security paying \$1 at state K and future time T, called the Arrow-Debreu security. Letting $\Delta K \to 0$ to be in continuous time, the density becomes

$$\phi(t_0;T,K) = \frac{1}{P(t_0,T)} \frac{\partial^2}{\partial K^2} C(K,T) \quad (1.5.29)$$

Hence, for any time T one can recover the marginal risk-neutral distribution of the stock price from the volatility surface. However, it tells us nothing about the joint distribution of the stock price at multiple times $T_1,..,T_n$. This is because the volatility surface is constructed from European options prices which only depend on the marginal distribution of S_T.

1.5.3 Some volatility products

1.5.3.1 The straddle

A straddle is an option spread where one buy (sell) both put and call options of the same strike price and maturity on the same underlying. A long straddle involves going long both a call option and a put option on the same stock with strike K, and satisfies

$$S_{BS}(t,x,K,T) = C_{BS}(t,x,K,T) + P_{BS}(t,x,K,T)$$

Given the definition of the BS-formula in Section (1.2.3), we can write the price of a straddle as

$$S_{BS}(t,x,K,T) = xe^{-q(T-t)}\big(2N(d_1(T-t,F(t,T),K)) - 1\big) - Ke^{-r(T-t)}\big(2N(d_2(T-t,F(t,T),K)) - 1\big)$$

Further, we can compute the limit case of a long straddle when the volatility tends to zero as

$$\lim_{\sigma \to 0} S_{BS}(t,x,K,T) = \big(xe^{-q(T-t)} - Ke^{-r(T-t)}\big)^+ + \big(Ke^{-r(T-t)} - xe^{-q(T-t)}\big)^+$$

which gives

$$\lim_{\sigma \to 0} S_{BS}(t,x,K,T) = \begin{cases} xe^{-q(T-t)} - Ke^{-r(T-t)} \text{ if } xe^{-q(T-t)} > Ke^{-r(T-t)} \\ Ke^{-r(T-t)} - xe^{-q(T-t)} \text{ if } xe^{-q(T-t)} < Ke^{-r(T-t)} \\ 0 \text{ if } xe^{-q(T-t)} = Ke^{-r(T-t)} \end{cases}$$

Hence, we can rewrite the limit of a long straddle when the volatility tends to zero as

$$\lim_{\sigma \to 0} S_{BS}(t,x,K,T) = |xe^{-q(T-t)} - Ke^{-r(T-t)}|$$

which is the absolute value of the payoff. When the volatility tends to infinity, we get

$$\lim_{\sigma \to \infty} S_{BS}(t,x,K,T) = xe^{-q(T-t)} + Ke^{-r(T-t)}$$

Assuming that the options in the straddle are ATM forward, we use Equation (1.2.12) and get

$$S_{BS}(t,x,K,T)\big|_{K=F(t,T)} \approx 0.8xe^{-q(T-t)}\sigma\sqrt{T-t}$$

From the above equation we can easily compute the delta of an ATM forward straddle option as

$$\Delta_S(t,x,K,T)\big|_{K=F(t,T)} \approx 0.8e^{-q(T-t)}\sigma\sqrt{T-t}$$

Straddles are a good strategy to pursue if an investor believes that a stock's price will move significantly, but is unsure as to which direction it will go. The stock price must move significantly if the investor is to make a profit. This position has a limited risk, since the purchaser of the strategy may lose at most the cost of both options, and there is unlimited profit potential. However, when shorting a straddle the profit is limited to the premium received from the sale of put and call. The risk is virtually unlimited as large moves of the underlying security's price either up or down will cause losses proportional to the magnitude of the price move. A maximum profit upon expiration is achieved if the underlying security trades exactly at the strike price of the straddle. In that case both puts and calls comprising the straddle expire worthless. Note, on stocks that are expected to jump, the market tends to price options at a higher premium, which ultimately reduces the expected payoff should the stock move significantly.

1.5.3.2 The strangle

A strangle option spread is similar to a straddle as one still buy (sell) both put and call options with the same maturity and on the same underlying, but different strikes are used. The investor use put options having a strike below the call options, and vice versa. Assuming a long strangle with strikes $K_1 < K_2$, we get

$$\widehat{S}_{BS}(t,x,K_1,K_2,T) = P_{BS}(t,x,K_1,T) + C_{BS}(t,x,K_2,T)$$

Given the definition of the BS-formula in Section (1.2.3), we can compute the limit case of a long strangle when the volatility tends to zero as

$$\lim_{\sigma \to 0} \widehat{S}_{BS}(t,x,K_1,K_2,T) = \big(K_1 e^{-r(T-t)} - xe^{-q(T-t)}\big)^+ + \big(xe^{-q(T-t)} - K_2 e^{-r(T-t)}\big)^+$$

which gives

$$\lim_{\sigma \to 0} \widehat{S}_{BS}(t,x,K_1,K_2,T) = \begin{cases} xe^{-q(T-t)} - K_2 e^{-r(T-t)} \text{ if } xe^{-q(T-t)} > K_2 e^{-r(T-t)} \\ K_1 e^{-r(T-t)} - xe^{-q(T-t)} \text{ if } xe^{-q(T-t)} < K_1 e^{-r(T-t)} \\ 0 \text{ if } K_1 e^{-r(T-t)} \leq xe^{-q(T-t)} \leq K_2 e^{-r(T-t)} \end{cases}$$

and when the volatility tends to infinity

$$\lim_{\sigma \to \infty} \widehat{S}_{BS}(t,x,K_1,K_2,T) = K_1 e^{-r(T-t)} + xe^{-q(T-t)}$$

Assuming that the options in the strangle are close to ATM forward, with $K_1 \approx F(t,T)$ and $K_2 \approx F(t,T)$, we use Equation (1.2.11) and get

$$\widehat{S}_{BS}(t,x,K_1,K_2,T) \approx K_1 e^{-r(T-t)}\big(\frac{1}{2} + \frac{1}{5}\sigma\sqrt{T-t}\big) + xe^{-q(T-t)}\frac{2}{5}\sigma\sqrt{T-t} - K_2 e^{-r(T-t)}\big(\frac{1}{2} - \frac{1}{5}\sigma\sqrt{T-t}\big)$$

Setting $K_1 = K_2 = F(t,T)$, we recover the formula, $S_{BS}(t,x,K,T)\big|_{K=F(t,T)}$, for the ATM forward straddle. In general, the strategy involves buying an out-of-the-money call and an out-of-the-money put option so that its intrinsic value is zero and stays there if the underlying asset hardly moves. A strangle is generally less expensive than a straddle as the contracts are purchased out of the money. The owner of a long strangle makes a profit if the underlying price moves far enough away from the current price, either above or below.

1.5.4 The call-spread option

1.5.4.1 Description

A spread position is entered by buying and selling equal number of options of the same class on the same underlying security, but with different strike prices or expiration dates. A bull call spread option is a vertical spread made of two calls with the same expiration but different strikes where the strike price of the short call is higher than the strike of the long call. The main purpose of the short call is to help pay for the upfront cost of the long call. The upside potential is capped. Hence, the call spread is decreasing with increasing volatility. The net volatility of an option spread is the volatility level such that the theoretical value of the spread trade is equal to the spread's market price. Assuming a symmetric smile, the effects of volatility shifts on the two contracts offset each other. However, in presence of skew the change in stock price can affect one leg more than the other. In any case, we can highlight a theoretical maximum value for all call spreads. When the spread market price is close to this theoretical maximum value, we sell it.

1.5.4.2 Some limits

We are going to study the properities of the call spread by analysing its limit behaviour. We consider a bull call spread with strikes $K_1 < K_2$ given by

$$CS_{BS}(t,x,K_1,K_2,T) = C_{BS}(t,x,K_1,T) - C_{BS}(t,x,K_2,T)$$

Given the definition of the BS-formula in Appendix (1.2.3), we can compute the limit cases of the call spread as follow

$$\lim_{\sigma \to 0} CS_{BS}(t,x,K_1,K_2,T) = \big(xe^{-q(T-t)} - K_1 e^{-r(T-t)}\big)^+ - \big(xe^{-q(T-t)} - K_2 e^{-r(T-t)}\big)^+$$

which gives

$$\lim_{\sigma \to 0} CS_{BS}(t,x,K_1,K_2,T) = \begin{cases} (K_2 - K_1)e^{-r(T-t)} \text{ if } xe^{-q(T-t)} \geq K_1 e^{-r(T-t)} \text{ and } xe^{-q(T-t)} \geq K_2 e^{-r(T-t)} \\ 0 \text{ if } xe^{-q(T-t)} < K_1 e^{-r(T-t)} \text{ and } xe^{-q(T-t)} < K_2 e^{-r(T-t)} \end{cases}$$

and

$$\lim_{\sigma \to \infty} CS_{BS}(t,x,K_1,K_2,T) = xe^{-q(T-t)} - xe^{-q(T-t)} = 0$$

Similarly, the put spread satisfies

$$PS_{BS}(t,x,K_1,K_2,T) = P_{BS}(t,x,K_2,T) - P_{BS}(t,x,K_1,T)$$

and the limit cases of the put spread are as follow

$$\lim_{\sigma \to 0} PS_{BS}(t,x,K_1,K_2,T) = \left(K_2 e^{-r(T-t)} - xe^{-q(T-t)}\right)^+ - \left(K_1 e^{-r(T-t)} - xe^{-q(T-t)}\right)^+$$

which gives

$$\lim_{\sigma \to 0} PS_{BS}(t,x,K_1,K_2,T) = \begin{cases} (K_2 - K_1)e^{-r(T-t)} \text{ if } xe^{-q(T-t)} \leq K_1 e^{-r(T-t)} \text{ and } xe^{-q(T-t)} \leq K_2 e^{-r(T-t)} \\ 0 \text{ if } xe^{-q(T-t)} > K_1 e^{-r(T-t)} \text{ and } xe^{-q(T-t)} > K_2 e^{-r(T-t)} \end{cases}$$

and

$$\lim_{\sigma \to \infty} PS_{BS}(t,x,K_1,K_2,T) = K_2 e^{-r(T-t)} - K_1 e^{-r(T-t)} = (K_2 - K_1)e^{-r(T-t)}$$

Note, in the special case where $xe^{-q(T-t)} < K_1 e^{-r(T-t)}$ and $xe^{-q(T-t)} < K_2 e^{-r(T-t)}$ then the limit of the call-spread when $\sigma \to 0$ and $\sigma \to \infty$ is zero. Similarly, the limit of the put-spread when $\sigma \to 0$ and $\sigma \to \infty$ is the difference in strike discounted, $(K_2 - K_1)e^{-r(T-t)}$, when $xe^{-q(T-t)} \leq K_1 e^{-r(T-t)}$ and $xe^{-q(T-t)} \leq K_2 e^{-r(T-t)}$. That is, stressing volatility values for out-of-the-money (OTM) call-spreads in the BS-formula, we observe that

- call option premium converges to 0 when volatility is very low.
- call option premium converges to spot price when volatility is very high. It is infinity when the volatility is infinity.

while stressing volatility values for in-the-money (ITM) put-spreads in the BS-formula, we get

- put option premium converges to discounted strike when volatility is very low.
- put option premium converges to the strike price when volatility is very high (infinity).

Since the call-spread and put-spread are continuous and differentiable with respect to volatility, we can apply the Theorem of de Rolle (B.5.1) and deduce an optimum when the vega equals zero. We plot in Figure (1.4) an OTM call-spread versus different level of volatility with $S_0 = 100$, $T = 0.5$, $r = 0.03$ together with strikes $K_1 = 130$ and $K_2 = 140$. We plot in Figure (1.4) its associated vega. We plot in Figure (1.5) an ITM put-spread versus different level of volatility with $S_0 = 100$, $T = 0.5$, $r = 0.03$ together with strikes $K_1 = 130$ and $K_2 = 140$. We also plot in Figure (1.5) its associated vega.

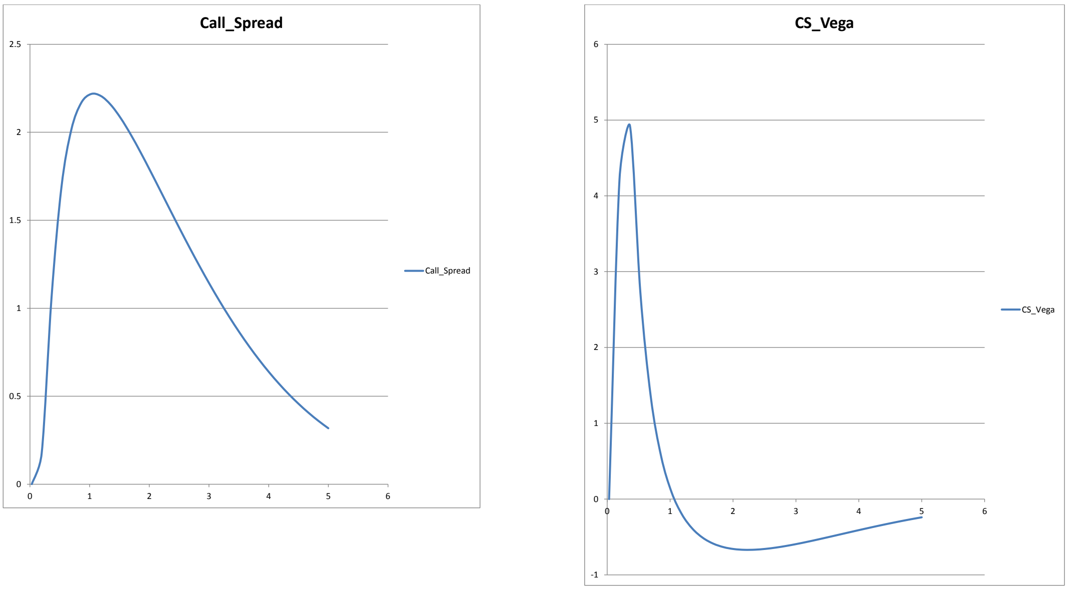
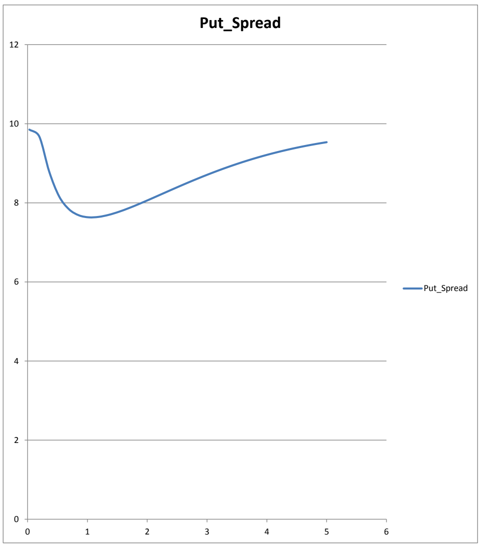


Figure 1.4: Call-spread and its vega with zero skew.



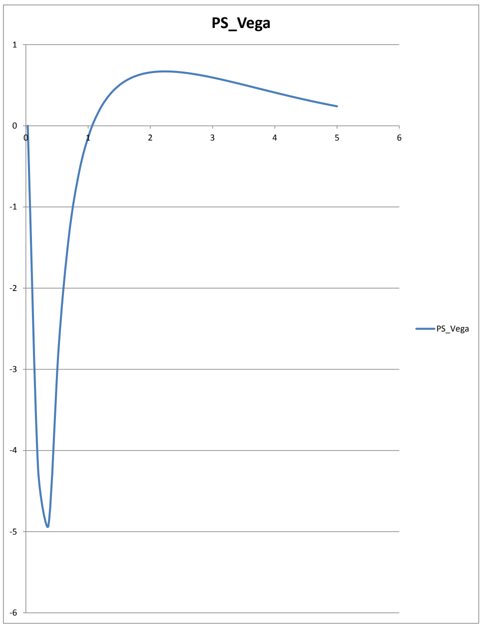


Figure 1.5: Put-spread and its vega with zero skew.

1.5.4.3 Some approximation

We are now going to approximate the call spread by letting the strikes $K_1 < K_2$ be close to the ATM forward. This way we will express the call spread as linear function of the volatility. We set $K_i = \alpha_i x e^{(r-q)(T-t)}$ for $i = 1, 2$ with $\alpha_i \in [1-\epsilon, 1+\epsilon]$ and $\epsilon > 0$. Using Equation (1.2.11), we approximate the call spread as follow

$$\begin{aligned} CS_{BS}(t,x,K_1,K_2,T) &= C_{BS}(t,x,K_1,T) - C_{BS}(t,x,K_2,T) \approx (K_2 - K_1)e^{-r(T-t)}\big(\frac{1}{2} - \frac{1}{5}\sigma\sqrt{T-t}\big) \\ &= \beta\epsilon x e^{-q(T-t)}\big(\frac{1}{2} - \frac{1}{5}\sigma\sqrt{T-t}\big) \end{aligned}$$

where $\beta \approx 2$. Since in a no-arbitrage world we have $C_{BS}(t,x,K_1,T) \geq C_{BS}(t,x,K_2,T)$, we must have

$$\frac{5}{2} \geq \sigma\sqrt{T-t}$$

Assuming now that we have some skew, such that $\sigma_1 > \sigma_2$ then we approximate the call spread as follow

$$\begin{aligned} & CS_{BS}(t,x,K_1,K_2,T) \approx x e^{-q(T-t)}\big(\frac{1}{2} + \frac{1}{5}\sigma_1\sqrt{T-t}\big) - K_1 e^{-r(T-t)}\big(\frac{1}{2} - \frac{1}{5}\sigma_1\sqrt{T-t}\big) \\ & - x e^{-q(T-t)}\big(\frac{1}{2} + \frac{1}{5}\sigma_2\sqrt{T-t}\big) - K_2 e^{-r(T-t)}\big(\frac{1}{2} + \frac{1}{5}\sigma_2\sqrt{T-t}\big) \\ & = x e^{-q(T-t)}\frac{1}{5}\big(\sigma_1\sqrt{T-t} - \sigma_2\sqrt{T-t}\big) + K_2 e^{-r(T-t)}\big(\frac{1}{2} - \frac{1}{5}\sigma_2\sqrt{T-t}\big) - K_1 e^{-r(T-t)}\big(\frac{1}{2} - \frac{1}{5}\sigma_1\sqrt{T-t}\big) \end{aligned}$$

Replacing K_i for $i = 1, 2$ as above, we get

$$\begin{aligned} & CS_{BS}(t,x,K_1,K_2,T) \approx x e^{-q(T-t)}\frac{1}{5}\big(\sigma_1\sqrt{T-t} - \sigma_2\sqrt{T-t}\big) + \alpha_2 x e^{-q(T-t)}\big(\frac{1}{2} - \frac{1}{5}\sigma_2\sqrt{T-t}\big) \\ & - \alpha_1 x e^{-q(T-t)}\big(\frac{1}{2} - \frac{1}{5}\sigma_1\sqrt{T-t}\big) \end{aligned}$$

which simplifies to

$$CS_{BS}(t,x,K_1,K_2,T) \approx (\alpha_2 - \alpha_1)x e^{-q(T-t)}\frac{1}{2} + x e^{-q(T-t)}\frac{1}{5}\Big[\big(\sigma_1\sqrt{T-t} - \sigma_2\sqrt{T-t}\big) + \big(\alpha_1\sigma_1\sqrt{T-t} - \alpha_2\sigma_2\sqrt{T-t}\big)\Big]$$

1.5.4.4 Computing the vega

In the special case where the volatility surface is flat, the vega of the call spread is given by

$$\frac{\partial}{\partial\sigma}CS_{BS}(t,x,K_1,K_2,T;\sigma) = \frac{\partial}{\partial\sigma}C_{BS}(t,x,K_1,T;\sigma) - \frac{\partial}{\partial\sigma}C_{BS}(t,x,K_2,T;\sigma)$$

Asumming now two disctinct volatilities, the BS-volatility σ_1 at the strike K_1 and the BS-volatility σ_2 at the strike K_2, the call spread becomes

$$CS_{BS}(t,x,K_1,K_2,T;\sigma_1,\sigma_2) = C_{BS}(t,x,K_1,T;\sigma_1) - C_{BS}(t,x,K_2,T;\sigma_2)$$

The volatility surface is no-longer flat, and we need to assume something about the way σ_1 and σ_2 are related. For simplicity, we assume

$$\sigma_2 = R_\alpha \sigma_1$$

where $R_\alpha = (1 + \alpha)$ and α is a percentage. In that case, differentiating the call spread with respect to σ_1, we get

$$\frac{\partial}{\partial \sigma_1} CS_{BS}(t, x, K_1, K_2, T; \sigma_1) = \frac{\partial}{\partial \sigma_1} C_{BS}(t, x, K_1, T; \sigma_1) - \frac{\partial}{\partial \sigma_2} C_{BS}(t, x, K_2, T; \sigma_2) \frac{d\sigma_2}{d\sigma_1}$$

with $\frac{d\sigma_2}{d\sigma_1} = R_\alpha$. Adding a skew of -0.1 to the eamples above, we plot in Figure (1.6) the call-spread and its associated vega in Figure (1.6). We also plot in Figure (1.7) the put-spread and its associated vega in Figure (1.7).

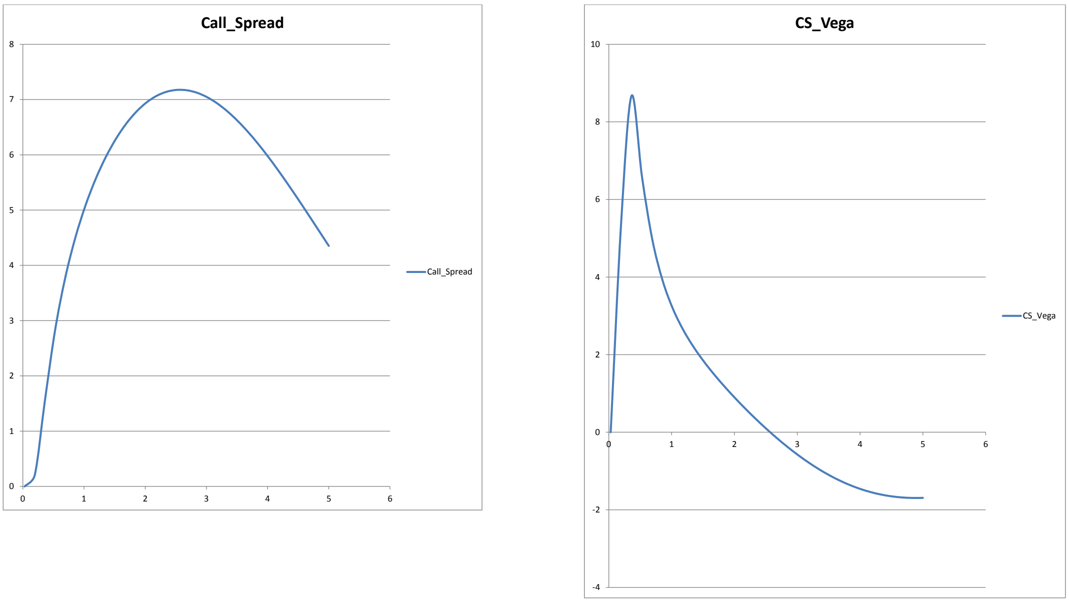
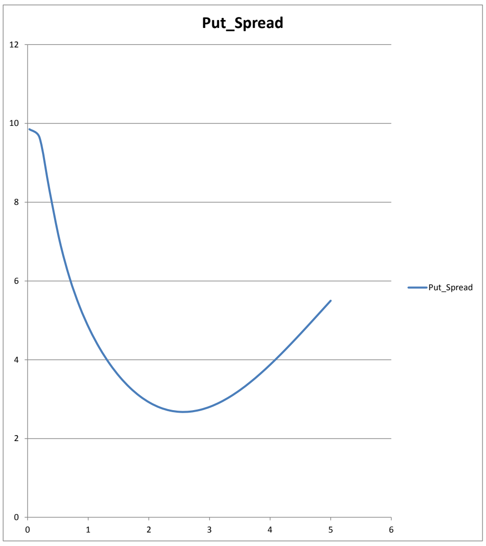


Figure 1.6: Call-spread and its vega with with -0.1 skew.



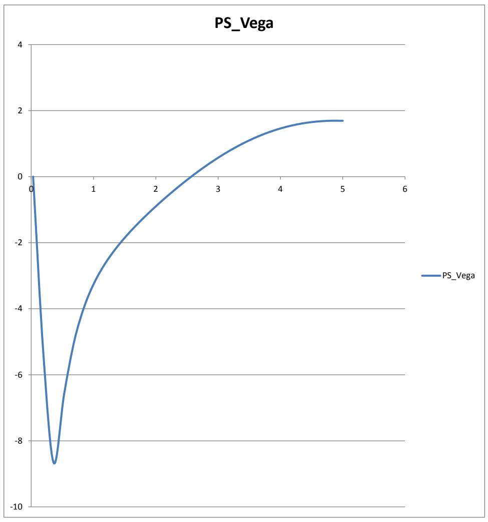


Figure 1.7: Put-spread and its vega with with -0.1 skew.

1.6 Introducing risk management

Hedging an option is the practice of taking offsetting positions in one market to balance against adverse movements in the value of our positions in another market. Assuming market returns to be lognormally distributed, we showed in Section (1.2.2) how Black and Scholes used a replicating argument to derive a theoretical valuation formula. The factors affecting option prices are the spot price S, the strike price K, the time to maturity T, the volatility of the underlying σ, and the risk-free interest rate r. Thus, when risk managing a book of options one should consider the sensitivity of option prices to these factors. In that framework, volatility is the key parameter for traders, allowing them to compute the necessary sensitivities, or Greeks, of their portfolio. We are going to introduce a few financial risks and present some tools used by traders to risk-mannage their book.

1.6.1 Defining financial risks

1.6.1.1 Identifying financial risks

For accounting purposes, the mark-to-market of a portfolio of contingent claims consists in determining the price of this portfolio using the prevailing market price of its components. Since the mark-to-market of a portfolio can be positive or negative, traders must define risks and quantify the conditions of a loss. Liquidity risk is the risk of not being able to liquidate the assets at the prevailing market price due to an insufficient depth of the market, while funding liquidity risk is the risk of not being able to raise capital for current operations. Operational risk refers to losses resulting from inadequate or failed internal process, people and systems, or from external events. One way of reducing that risk is to apply strict and frequent internal controls, up to date computer security and expert judgement. Model risk can be caused by an inadequate choice of pricing models, parameter errors (due to statistical estimation errors or nonstationary parameters), and inadequate implementation (unstable algorithm amplifying data errors). Even though parameter estimation errors may be quantified by varying the parameters within confidence bounds, inadequate models is more difficult to analyse and quantify. In order to control the risk of a portfolio we need to identify relevant risk factors and impose some limits on the sensitivities to these risk factors. The sensitivity is defined as the variation of the price of a financial product corresponding to a small change in the value of a given risk factor, when all other risk factors are kept constant. Examples of risk factors include FX rates, equity and equity index, yield curves, implied volatility, commodity term structures, etc. Hence, choosing the proper model accounting for the relevant risk factors associated to a given product class is of great importance.

1.6.1.2 Accounting for risk

Facing exposure to market direction in addition to volatility, option prices are not pure play trades on volatility. As a result, option traders must spend considerable time and financial resources delta hedging their position to get rid of directional risk. In addition, vanilla options are subject to theta decay, their value decreases simply with the passage of time, such that options can not be included as an asset in a passively managed portfolio. All these risks are incorporated into the market price of risk in such a way that the riskier the asset the higher the return. Assuming a risk-neutral world, market practice consists in using the BS-formula as a tool for representing market risk via the implied volatility surface. Once the market price of risk has been implicitly entered into the pricing equation, the traders are left to define their hedging strategies. For example, when delta hedging, the traders make the derivative of the portfolio with respect to the stock price zero at a point, but the portfolio still change in value if the asset moves a short distance from that point. However, if the distance is small, the change in value will be proportional to its square, whereas for a non delta hedged portfolio the change will be linear in the distance (see Fengler [2005]). Similarly, if we use options to hedge the portfolio, the second derivative can be matched and the portfolio's change in value for small changes in price of the underlying will be proportional to the cube of the change, which is a tiny quantity. Fortunately, one simplification to the trader's position arises from the fact that he will have bought and sold many different options on the same underlying. Each of these options has a delta and since in a complete market the model is linear, the traders can hedge them all by simply adding their deltas together. As the deltas of long and short positions have opposite signs, and if the portfolio is a mixture of such positions, the deltas will at least partially cancel each other.

1.6.1.3 The volatility risk

Even though the Black-Scholes formula assumes constant volatility, market practice consists in using one BS-model for every pair (K,T), leading to the generation of a non-flat IV surface. Since this approach is a proxy for accouting for stochastic variations of the volatility surface, traders use the wrong volatility parameter into the wrong formula. Traders must therefore consider the sensitivity of option prices to changes in the IV surface. When the volatility σ_t changes with a certain volatility of its own, the sensitivity of the option price with respect to the volatility is called Vega. Hence, the vega risk is associated with unexpected changes in volatility and should be accounted for (see Page et al. [1995]). Theoretically, the BS-formula does not apply to stochastic volatility, although the explicit BS-vega calculation is a useful approximation to the actual vega over a small time horizon. Market practice consists in bucketing the smile, that is, computing sensitivities to perturbations of specific sections of the smile. Note, this approach assumes that one small section of the smile moves while the other one remain constant, violating the no-arbitrage constraints (see Section (1.4.2.3)). One solution could be to use principal component analysis (PCA) to identify possible orthogonal deformation patterns, such as level changes, twists (skew), and convexity changes, and compute sensitivities to such deformations. All else constant, ATM options are more sensitive to changes in volatility than the OTM options, and this sensitivity is increasing with the initial level of the volatility. Note, the sensitivity in market value to changes in volatility is greater for far OTM options, and for lower initial volatility. The distinction between the absolute and relative sensitivities of option prices to volatility results from the fact that the price of an option declines more quickly than does its vega, as the option becomes more and more OTM and as the volatility parameter is lowered.

1.6.2 About the Greeks and their interpretation

We saw in Section (1.2.3) that the Black-Scholes formula was a function

$$v(t,x,r,q;T,K;\sigma)$$

where the underlying asset $(S_t^x)_{t\geq 0}$ and the time t are state variables, the risk-free rate r, the repo rate q and the volatility σ are fixed parameters of the model, the maturity T and the strike K are fixed parameters of the contingent claim. The sensitivities of a call price with respect to these parameters play a fundamental role in the risk management of derivative products. For instance, a call option with maturity T and strike K can be hedged with the quantity

$$\Delta(t,S_t)=\partial_x C_{BS}(t,S_t,K,T)=e^{-q(T-t)}N(d_1(T-t,F(t,T),K))$$

of the risky asset. For simplicity we let $t=0$ with initial value $S_0=x$ and we define $n(x)=\partial_x N(x)$. In the following we briefly present a few Greeks, for more details see Appendix (B.1.2).

$$\begin{aligned}
\partial_{xx}C &= \Gamma=\frac{e^{-q(T-t)}}{x\sigma\sqrt{T-t}}n(d_1)>0\\
\partial_{\sigma}C &= Vega=xe^{-q(T-t)}\sqrt{T-t}n(d_1)\\
\partial_K C &= -e^{-r(T-t)}N(d_2)\\
\partial_r C &= \rho=(T-t)Ke^{-r(T-t)}N(d_2)>0\\
\partial_t C &= \Theta=-\frac{x\sigma}{2\sqrt{T-t}}e^{-q(T-t)}n(d_1)+qxe^{-q(T-t)}N(d_1)-rKe^{-r(T-t)}N(d_2)
\end{aligned}$$

and the Black-Scholes partial differential equation (PDE) can be written as

$$\frac{1}{2}\sigma^2x^2\Gamma+(r-q)x\Delta-rC+\Theta=0 \tag{1.6.30}$$

Since we can express the Gamma in terms of the Vega (see Equation (B.1.6)), we can rewrite the PDE as

$$\frac{\sigma}{2T}Vega + (r-q)x\Delta - rC + \Theta = 0$$

While $\Delta(t, S_t)$ is a quantity of the risky asset to buy or sell when building the replicating portfolio, the other Greeks are not easily interpreted as financial quantities and must therefore be modified for risk management purposes. Assuming constant volatility, the dynamics of the stock price in discrete time is given by

$$\Delta S_t = \mu S_t \Delta t + \sigma S_t y \sqrt{\Delta t}$$

where ΔS_t is the variation of the stock price in the time interval Δt, and $y \sim N(0,1)$. As a result, $\frac{\Delta S_t}{S_t}$ follows a normal distribution with mean $\mu \Delta t$ and standard deviation $\sigma\sqrt{\Delta t}$. That is, $\frac{\Delta S_t}{S_t} \sim N(\mu \Delta t, \sigma\sqrt{\Delta t})$. We let $S^+ = S + \alpha$ where $\alpha = \epsilon S$ with $\epsilon << 1$, and we define a change in spot as $\Delta S = S^+ - S = \alpha$. In general, the market assumes $\epsilon = 1\%$. Using finite difference, we can approximate the Delta with a one-sided finite difference as

$$\Delta \approx \frac{C(S^+) - C(S)}{\alpha}$$

In order to include some convexity in the delta, by considering the difference in value between an option after an upward bump and one after a downward bump, we can use a two-sided finite difference as

$$\Delta \approx \frac{C(S^+) - C(S^-)}{2\alpha}$$

In the case of the one-sided delta we can deduce that

$$C(S^+) \approx \alpha\Delta + C(S)$$

We see that we can obtain a modified Delta as $\widehat{\Delta} = \alpha\Delta$ such that

$$C(S^+) \approx \widehat{\Delta} + C(S)$$

where we get the new price simply by adding $\widehat{\Delta}$ to the old price. Similarly, we can rewrite the Gamma in terms of the Delta as

$$\Gamma \approx \frac{\Delta(S^+) - \Delta(S)}{\alpha}$$

and deduce that

$$\Delta(S^+) \approx \alpha\Gamma + \Delta(S)$$

We see that we can obtain a modified Gamma as $\widehat{\Gamma} = \alpha\Gamma$ such that

$$\Delta(S^+) \approx \widehat{\Gamma} + \Delta(S)$$

where we get the new Delta simply by adding $\widehat{\Gamma}$ to the old Delta. In other words, if the spot price jumps from S to S^+, the trader must immediately modify his position in the risky asset by buying the quantity $\widehat{\Gamma}$ (or selling that quantity if $\widehat{\Gamma} < 0$) of the risky asset. Hence, the modified Gamma corresponds to a change in the hedging strategy induced by a sudden jump in the underlying. We now let $\sigma^+ = \sigma + \alpha$ where $\alpha = \epsilon\sigma$ with $\epsilon << 1$, and we define a change in volatility as $\Delta\sigma = \sigma^+ - \sigma = \alpha$. Using one-sided finite difference, we can approximate the Vega as

$$Vega \approx \frac{C(\sigma^+) - C(\sigma)}{\alpha}$$

Since volatility is already expressed in percentage, market practice is to directly set $\alpha = 1\%$ and obtain the modified Vega as $\widehat{V}ega = \alpha Vega$ such that

$$C(\sigma^+) \approx \widehat{V}ega + C(\sigma)$$

where we get the new price simply by adding $\widehat{V}ega$ to the old price. While theoretically the vega does not exist in the BS-formula, it provides an exposition of the call price to volatility risk. This is related to the IV surface and the fact that traders acknowledge a stochastic variation of the volatility. At last, the Theta above is calculated with yearly parameters and represents time decay over a yearly period. To express time decay over a one day period, we simply multiply the Theta with a fraction of time, getting

$$\widehat{\Theta} = \Theta \times \Delta t$$

where $\Delta t = \frac{1}{365}$.

When risk managing options, traders need to estimate their risk over a short time period. Relaxing the constraint of constant volatility and setting $V = V(t, S, \sigma)$, where σ represents a stochastic volatility, we apply a Taylor expansion on ΔV with respect to the time, the stock and the volatility, getting the variation

$$\Delta V_t = \frac{\partial V}{\partial t}\Delta t + \frac{\partial V}{\partial S}\Delta S_t + \frac{\partial V}{\partial \sigma}\Delta\sigma + \frac{1}{2}\big[\frac{\partial^2 V}{\partial S^2}(\Delta S_t)^2 + \frac{\partial^2 V}{\partial \sigma^2}(\Delta\sigma)^2 + 2\frac{\partial^2 V}{\partial S \partial \sigma}(\Delta S_t)(\Delta\sigma)\big] + ... \quad (1.6.31)$$

where $\Delta V_t = V(t + \Delta t, S + \Delta S, \sigma + \Delta\sigma) - V(t, S, \sigma)$. It is a quadratic function of ΔS_t where the linear terms behaves like the stock price and the quadratic term increases no matter how S_t moves. Using our previous notation, we can rewrite the portfolio variation as

$$\Delta V_t = \widehat{\Theta} + \widehat{\Delta} + \widehat{V}ega + \frac{1}{2}\widehat{\Gamma}\Delta S_t + \frac{1}{2}Volga(\Delta\sigma)^2 + Vanna(\Delta S_t)(\Delta\sigma) + ... \quad (1.6.32)$$

where $Volga = \frac{\partial^2 V}{\partial \sigma^2}$ and $Vanna = \frac{\partial^2 V}{\partial S \partial \sigma}$. Note, if $\epsilon = \frac{\Delta S}{S}$ we get $\widehat{\Delta} = (\frac{\Delta S}{S})S\Delta(t, S_t)$ where $S\Delta(t, S_t)$ is the equivalent stock position (ESP), or dollar delta. We see that the knowledge of the modified Greeks allows the traders to estimate the variation of the claim over a small period of time in the case where the stock price follows a diffusion process with stochastic volatility. More precisely, this portfolio variation assumes a one day period, with a spot price change of 1% and a volatility change of 1%, which is not very realistic. Further, ΔV_t does not correspond to the delta-hedged profit and loss, which we will discuss in Section (4.2.3). There are some limitations to computing the sensitivities of prices (see Tankov et al. [2010]), some of which being:

- they only relate changes in the values of derivative products to the changes of basic risk factors.
- they can not be aggregated across different underlyings and across different types of sensitivities (delta plus gamma). They are local measures and do not provide global information about the entire portfolio.
- they are only meaningful for small changes of risk factors, and do not provide accurate information about the change of portfolio to larger moves (jumps).
- the sensitivity to a given risk factor is associated to a very specific scenario of market dynamics applying only to the change in that risk factor while others are kept constant.

1.6.3 Some simple hedging strategies

Since we saw in Section (1.6.2) how to compute the most important sensitivities of European option prices, portfolios built with options can be hedged by taking the appropriate offsetting positions in the markets for the underlying, or, other options. For instance, when attempting to isolate skew or curvature in the volatility surface, the most important sensitivities to hedge against are the Delta and Vega, in order to remain immune to shifts in the underlying, and shifts in the overall level of volatility.

1.6.3.1 Delta hedging

Through delta-hedging, we hope to remain immune to movements in the underlying asset price that may negatively affect our position. Since the underlying asset has a Delta of one, we need only buy/sell the opposite number of units of the underlying (relative to our position) in order to achieve Delta-neutrality of our portfolio. For example, consider a trading strategy designed to isolate the skew in the volatility surface. When the implied risk-neutral density is more negatively skewed than the "real" density, OTM put options are potentially overpriced relative to ATM call or put options. Since skewness is known to be a mean-reverting process, a common approach is to "trade against the skew", by selling OTM put options and buying ATM or NTM call or put options. Such a strategy is commonly known as a Risk-Reversal. Consider a position short one put option at strike K_1, and long one call option at strike K_2 where $K_1 < K_2$. Then, Delta-neutrality can be achieved by forming the following portfolio:

$$\Pi_{RR} = C(S_t, K_2) - P(S_t, K_1) - \big(\Delta_C - \Delta_P\big) S_t \tag{1.6.33}$$

where Δ_C is the delta of a call option and Δ_P is the delta of a put option. Now consider a strategy designed to isolate the curvature of the surface. The strategy we will employ will mimic the concept of a butterfly trade, in that we buy/sell options at three different strike prices depending on which options we believe to be over-/under-valued. Specifically, if the risk-neutral density has a much higher kurtosis than the "real" density, the trader may wish to sell OTM options, and buy options ATM or NTM in order to capitalise on the mean-reversion quality of surface curvature. Specifically, consider a position short one put option at strike K_1, long one call and one put option at strike K_2, and short a further call option at strike K_3 where $K_1 < K_2 < K_3$. A Delta-neutral position can be achieved by forming the portfolio

$$\Pi_B = C(S_t, K_2) + P(S_t, K_2) - P(S_t, K_1) - C(S_t, K_3) - \big(\Delta_{C_2} + \Delta_{P_2} - \Delta_{P_1} - \Delta_{C_3}\big) S_t \tag{1.6.34}$$

1.6.3.2 Delta-vega hedging

Ideally, we may wish to hedge against movements in both the underlying asset price and the level of volatility that may negatively affect our position value. Through constructing strategies in this setting, we are able to isolate the "rotations" of the surface from its parallel movements, and hence profit as skew and curvature depart from, or, revert to their long-term mean. Since the underlying itself has a Vega of zero, we cannot rely on the underlying alone to establish a Vega-neutral position. Rather, by holding the appropriate ratio of options and the underlying, we can build a portfolio with both zero Delta and Vega. Since the Vega of a call and put option is typically not the same, we require Delta and Vega for both the call and put. Consider again the Risk-Reversal strategy discussed above, then the following portfolio achieves Delta-Vega-neutrality:

$$\Pi_{RR} = C(S_t, K_2) - \frac{\nu_c}{\nu_p} P(S_t, K_1) - \big(\Delta_C - \frac{\nu_c}{\nu_p} \Delta_P\big) S_t \tag{1.6.35}$$

where $\nu_c = \frac{\partial C}{\partial \sigma}$ is the vega of a call option and $\nu_p = \frac{\partial P}{\partial \sigma}$ is the vega of a put option. Likewise, in the case of the butterfly type strategy discussed above, a Delta-Vega-neutral portfolio is given by

$$\begin{aligned}
\Pi_B &= C(S_t, K_2) + \frac{\nu_{c_2}}{\nu_{p_2}} P(S_t, K_2) - \frac{\nu_{c_3}}{\nu_{p_1}} P(S_t, K_1) - C(S_t, K_3) \\
&- \big(\Delta_{C_2} + \frac{\nu_{c_2}}{\nu_{p_2}} \Delta_{P_2} - \frac{\nu_{c_3}}{\nu_{p_1}} \Delta_{P_1} - \Delta_{C_3}\big) S_t
\end{aligned}$$

1.6.3.3 Introducing dynamic hedging

It is worth expressing a note of caution with regards to the above. As discussed earlier, an arbitrage in the hedge fund sense is by no means risk-free. Hedging is an attempt to mitigate the risk of any downside you may face, but the Greeks are only locally correct during the hedging process, in the sense that they themselves are subject to sensitivities and so any attempt to hedge can only be a true hedge for an arbitrarily small time interval. As such, ones position must be dynamically adjusted to account for movements in the value of the Greeks. Whilst high-frequency trading systems may be capable of maintaining neutral positions to a reasonable accuracy, the impact of transactions costs and bid-ask spreads makes continuous adjustment infeasible in practice. It is instead standard for traders to adjust their positions daily and discretely approximate the continuous hedge. Others maintain a Delta within a certain interval, adjusting their hedge when the Delta exists this region, while some will hedge dependent upon the moving of the underlying. In order to mitigate the risks of inaccurate hedging, second moments are often considered, the most common including Gamma, Vanna, Vomma and Charm. Gamma hedging for instance, can reduce the impact that movements in the underlying will have on a positions delta, and hence reduce the inaccuracy of the Delta-hedge. In the case of the skew trade above, Δ_P increases more steadily than Δ_C when the underlying decreases (and vice versa as it increases), meaning a bullish market favours the position, whilst a bearish market is unfavourable. Hence, if the Delta-hedged position is left too long without re-balancing, the skew trade remains exposed to movements in the underlying asset.

1.6.4 Computing the Value at Risk

We saw in Section () that the Value at Risk (VaR) belongs to the class of Monetary risk measures quantifying the risk in dollar amount. It is a simple portfolio risk measure summarising all risk-factors affecting a large portfolio and accounting for correlations and other dependencies. It consists in identifying the risk factors, identifying the dependency of each product in the portfolio to these factors, and devising scenarios or probability laws for the dynamics of the factors. Then, all products in the portfolio are repriced with perturbed values of the risk factors. While VaR is relatively simple when applied to returns on cash securities, risks on non-linear contingent claims become much harder to measure since detailed historical returns on these claims are not available. Since we can obtain data on the prices and volatilities of the underlyings on which the claims are written we should proceed as follow:

1. simulate or devise secenarios for the risk factors.
2. calculate the return, or the change in value, of the contingent claims.
3. compute the relevant percentile of the resulting distribution of portfolio returns.

Going around, we can locally approximate the change in value of an option with the Taylor expansion given in Equation (1.6.32), and the VaR becomes concerned with the calculation of the Greeks (see Kwiatkowski [1997]). However, as discussed in Section (4.1.3), the Greeks are model dependent and we need to make some assumptions about the dynamics of the implied volatility surface. In general, assuming Gaussian risk factors, one can compute analytically the VaR for some asset class. Alternatively, we can use historical changes in risk factors getting historical VaR method, or we can simulate the variations of the risk factors from a fully calibrated model, getting Monte Carlo VaR method. It is very time consuming as the derivative prices must be evaluated for each simulation path. Even though complex derivatives need to be priced numerically, for vanilla options we can rely on the characteristic function method or some expansions.

1.6.4.1 The factor method

The VaR being concerned with the estimation of the cumulative distribution function (CDF) of the returns of a portfolio, we must be able to compute the change in value of the assets in a portfolio, $\Delta V(d)$, from time t to $t+d$. One solution for estimating the VaR of the entire portfolio is to shock a number of risk factors affecting the market value of its components (see Duffie et al. [1997]). We let $X_1, .., X_n$ be the risk factors and denote $C_{ij} = \sigma_i \sigma_j \rho_{ij}$ as the covariance between the risk factor X_i and X_j, where σ_i is the standard deviation of X_i and σ_j is that of X_j. We can then define the unexpected change in market value of the portfolio as a linear factor model

$$\Delta V = \beta_1 X_1 + ... + \beta_n X_n \tag{1.6.36}$$

where β_i is the direct exposure of the ith risk factor. It represents the dollar change in the market value of the portfolio in response to a unit change in the ith risk factor. Thus, the variance of the change in value of the portfolio is given by

$$\sigma_P^2 = Var(\Delta V) = \sum_{i,j=1}^{n} \beta_i \beta_j C_{ij}$$

In the special case where the factors $(X_i)_{i=1,..,n}$ are jointly normally distributed, then the VaR at 99% confidence level is the 99-percentile change for a normally distributed random variable $VaR(\Delta V) = z_p \sigma_P$ with $z_p = 2.33$. Irrespective of the probability distribution function (pdf), by simulating the factors $(X_i)_{i=1,..,n}$ we can simulate the total unexpected change in market value ΔV and estimate the VaR as the level of loss exceeded by a fraction of simulated outcomes of ΔV.

There exists several approaches for simulating the risk factors, one of which being to assume a stationary statistical environment and simulate underlying prices in an historically realistic manner by bootstrapping from historical data. Considering historical returns as the source of simulated returns we can capture the correlations, volatilities, tail fatness and skewness present in the data without assuming a model. However, market price changes and historical returns are significantly non-stationary in terms of volatilities and correlations. It was suggested to update the historical return distribution by for instance drawing from the returns $(\hat{R}_i)_{i=1,..,n}$ defined by

$$\hat{R}_i = R_i \frac{\hat{V}}{V}$$

where V is the historical volatility and $\hat{V}$ is a recent volatility estimate. Similar approaches were proposed for correlations, even though correlation estimates are relatively unstable. In general, the change in market value of an option is non-linear in the factor X_i such that the parameter β_i is not a constant as time passes. Nonetheless, when dealing with options it is assumed that for small changes in the underlying the delta approach is sufficiently accurate, such that we get the approximation

$$\Delta V \approx \Delta_1 X_1 + ... + \Delta_n X_n$$

where Δ_i is the delta of the total portfolio with respect to the ith factor. That is, assuming k options on the same underlying S with prices $P_1, .., P_k$, the delta of the portfolio satisfies

$$\frac{d}{dS}[P_1(S) + ... + P_k(S)] = \frac{\partial}{\partial S} P_1(S) + ... + \frac{\partial}{\partial S} P_k(S)$$

which is the sum of the individual deltas. Improving on the accuracy, the same method can be applied with the gamma of an option. In the case of a portfolio exposed to several underlying assets, the delta-gamma approximation of the market value of the portfolio can be computed in terms of the deltas and gammas of the book with respect to each underlying asset and each pair of underlying assets. The (i,j)-gamma of the portfolio for the pair (i,j) is merly the sum of the (i,j)-gammas of all individual positions. Combining all deltas and gammas the total change in value of the portfolio is approximated as

$$\Delta V_t = \sum_{j=1}^{n} \Delta_j X_j + \frac{1}{2} \sum_{j=1}^{n} \sum_{k=1}^{n} \Gamma_{jk} X_j X_k$$

We can then compute the variance of the portfolio to second-order accuracy as

$$Var(\Delta V_t) = \sum_{j,k}^{n} \Delta_j \Delta_k Cov(X_j, X_k) + \sum_{i,j,k} \Delta_i \Gamma_{jk} Cov(X_i, X_j X_k) + \frac{1}{4} \sum_{i,j,k,l} \Gamma_{jk} \Gamma_{kl} Cov(X_i X_j, X_j X_k)$$

In the case of normal returns $(X_i)_{i=1,..,n}$ the covariance terms involving the products $X_i X_j$ for $i,j = 1,..,n$ can be computed explicitly.

Using the BS-formula, we can calculate the value of the portfolio as a function of the underlying asset price.

1.6.4.2 The delta-gamma method

When the portfolio consists of plain vanilla options, a simple approach to computing the VaR is to consider the Taylor expansion of the option price derived in Equation (1.6.31). We are now going to introduce some popular methods for computing the VaR, all based on the first or second order local approximation of option prices. In that setting, the risk factors affecting option prices are reduced to the increment risk ΔS_t, such that the portfolio returns are either linear or quadratic in underlying risks. However, the time decay and the sensitivities with respect to the volatility surface present in Equation (1.6.32) are ignored. In the delta-gamma method, it is assumed that the option price is only a function of the spot price S and that the volatility is constant, that is, $V_t = V(S_t)$. As a result, the option variation simplifies to

$$\Delta V_t \approx \frac{\partial V}{\partial S} \Delta S_t = \Delta_S \Delta S_t$$

which correspond to the factor model in Equation (1.6.36) with exposure $\beta = \Delta_S$ and factor $X = \Delta S_t$. Setting $R_S = \frac{\Delta S}{S}$, such that $\Delta S_t = S_t R_S$, the return on the derivative is approximated by

$$\frac{\Delta V_t}{V_t} \approx \frac{S_t}{V_t} \Delta_S R_S$$

with variance given by

$$Var(\frac{\Delta V_t}{V_t}) \approx (\frac{S_t}{V_t} \Delta_S)^2 Var(R_S)$$

Thus, for small changes in the spot price, we approximate the change in market value of an option as that of a fixed position in the underlying whose size is the delta of the option. Since the delta Δ_S is a constant, the potential profit and loss ΔV_t is a linear function of the changes in S, and the worst loss for V can be obtained from the worst loss for S. Thus, $VaR(\Delta V) = |\Delta_S| \times VaR(\Delta S)$. Even though the delta-based approximation is fairly accurate for short periods of time, VaR is typically concerned with large changes. Fom the convexity of the BS-formula, we see that this approach over-estimates the loss on a long option position and under-estimates the loss for a short option position. Assuming that $R_S = \frac{\Delta S}{S}$ is a Gaussian variable with mean zero and volatility $\sigma_S \sqrt{\Delta t}$, then the volatility of the portfolio becomes $\sigma_P = \Delta_S \sigma_S \sqrt{\Delta t} S_t$. Note, we recover the volatility of the BS-price dynamics obtained by applying Ito's lemma to the option price and given in Equation (1.2.5). Hence, the VaR follows as

$$VaR(\Delta V_t) = |\Delta_S| \times z_p \sigma S_t$$

where $\sigma = \sigma_S \sqrt{\Delta t}$ is the standard deviation of R_S over the time period Δt. Given the spot price S_t, the worst loss around its mean is $z_p \sigma S_t$, and the new spot price becomes $S_t - z_p \sigma S_t$. Thus, the new price of the option becomes $V(S_t - z_p \sigma S_t)$ with the change in value given by $\Delta V_t = V(S_t - z_p \sigma S_t) - V(S_t)$. It corresponds to the modified delta $\widehat{\Delta} = \alpha \Delta$ with $\alpha = -z_p \sigma S_t$ and $\epsilon = z_p \sigma$. In that setting, we get

$$VaR(\Delta V_t) = V(S_t - z_p \sigma S_t) - V(S_t)$$

Given a 99% confidence level, a one-day horizon $\sqrt{\Delta t} = \sqrt{\frac{1}{250}} = 6.32\%$ and a volatility of 25% we get $\epsilon = 3.68\%$, while a 1% delta would correspond in that setting to a volatility of 6.78%. One can improve the accuracy of the delta method by adding a second-order approximation term to the Taylor expansion of the option price, getting

$$\Delta V_t = \frac{\partial V}{\partial S} \Delta S_t + \frac{1}{2} \frac{\partial^2 V}{\partial S^2} (\Delta S_t)^2 = \Delta_S \Delta S_t + \frac{1}{2} \Gamma (\Delta S_t)^2$$

Since the delta Δ_S and gamma Γ are constants, we can obtain the variance of the portfolio return as

$$Var(\Delta V_t) = \Delta_S^2 Var(\Delta S_t) + \frac{1}{2} \Gamma^2 Var((\Delta S_t)^2) + \Delta_S \Gamma Cov(\Delta S_t, (\Delta S_t)^2)$$

In the case of Gaussian returns, we can compute explicitly the moments of ΔS_t and obtain a VaR with a 99% confidence level given by $VaR(\Delta V_t) = z_p \sqrt{Var(\Delta V_t)}$ with $z_p = 2.33$. Alternatively, the worst loss of the spot price is $\Delta S_t = -z_p \sigma S_t$, so that the change in value of the portfolio becomes

$$\Delta V_t = -\Delta_S (z_p \sigma S_t) + \frac{1}{2} \Gamma (z_p \sigma S_t)^2$$

and we get the VaR

$$VaR(\Delta V_t) = |\Delta_S| \times z_p \sigma S_t - \frac{1}{2} \Gamma (z_p \sigma S_t)^2$$

Note, the approximation in the delta-gamma method corresponds to a parabolic function over-estimating long option positions resulting in an under-estimated VaR. Further, it assumes that the underlying risk and its square are joint normally distributed, which is not the case in practice. The underlying risks having non-normal distributions, some authors proposed closer approximations to the true distribution (see Britten-Jones et al. [1997]). The Greeks being model dependent and very sensitive to the dynamics of the volatility surface (see Section (4.1.3)), we should include the additional terms defined in Equation (1.6.31) when calculating the VaR. However, it vould implies a model for the dynamics of the IV surface.

1.6.4.3 Inconsistency in the delta-gamma method

The VaR is a risk measure quantifying the risk in dollar amount of each asset composing a portfolio by shocking the risk factors, repricing each assets and computing the relevant percentile of the resulting distribution of returns. We saw in Section (4.2.3) that when a trader sells an option $P(t)$ he will risk manage it by being long a replicating portfolio π_Δ made of a quantity of stock and some cash. However, when computing the Value at Risk, we evaluate the variation of the option and that of the stock independently. If we consider the replicating portfolio π_Δ with dynamics given in Equation (4.2.9) then the Tracking Error associated with (P, Δ) is

$$e(t) = \pi_\Delta(t) - P(t)$$

Even though the trader is delta-hedging the option with a misspecified volatility, the computation of the PnL account for the true volatility which is not the case in the delta-gamma VaR estimation. For instance, assuming a delta-method to approximate the variation of the option, the total variation of the portfolio becomes

$$\Delta V(t) = \Delta_\gamma \Delta S_t - \Delta_\gamma \Delta S_t = 0$$

Further, assuming a delta-gamma-method to approximate the variation of the option, the total variation of the portfolio becomes

$$\Delta V(t) = \Delta_\gamma \Delta S_t - \Delta_\gamma \Delta S_t - \frac{1}{2} \Gamma_\gamma (\Delta S_t)^2 = -\frac{1}{2} \Gamma_\gamma (\Delta S_t)^2$$

Accounting for the Theta of the option in the estimation of the variation of the option, the total variation of the portfolio becomes

$$\Delta V(t) = \Delta_\gamma \Delta S_t - \Theta_\gamma \Delta t - \Delta_\gamma \Delta S_t - \frac{1}{2} \Gamma_\gamma (\Delta S_t)^2 = -\Theta_\gamma \Delta t - \frac{1}{2} \Gamma_\gamma (\Delta S_t)^2$$

and rewriting the Theta in terms of the Gamma, the total variation of the portfolio becomes

$$\Delta V(t) = \Delta_\gamma \Delta S_t + \frac{1}{2} S_t^2 \gamma^2 \Gamma_\gamma \Delta t - \Delta_\gamma \Delta S_t - \frac{1}{2} S_t \Gamma_\gamma (\frac{\Delta S_t}{S_t})^2 = \frac{1}{2} S_t^2 \Gamma_\gamma \big(\gamma^2 \Delta t - (\frac{\Delta S_t}{S_t})^2 \big)$$

and the worst case happen when the trader is under-hedged

$$\gamma^2(t, S_t) < \sigma^2(t)$$

That is, the strategy $(P_\gamma, \Delta_\gamma)$ is a substrategy. We can obtain the variance of the portfolio return as

$$Var(\Delta V(t)) = \frac{1}{2} S_t^2 \Gamma_\gamma \Big[\gamma^2 \Delta t - \frac{1}{2} S_t^2 \Gamma_\gamma Var((\frac{\Delta S_t}{S_t})^2) \Big]$$

1.6.5 VaR in the entire portfolio

1.6.5.1 The delta method

In that setting, we assume a portfolio of N options written on d stocks, $S_t^1, ..., S_t^d$, and we let P_t^i be the price of the ith option at time t. It is further assumed that the option price not only depends on the spot price on which it is written, but also depends on the other underlyings constituting the portfolio. This is the case when a contingent claim is based on more than one underlying such as a cross-rate option exposed to two currencies simultaneously, or a spread option which depends on two or more underlying indices. As such, the option price satisfies $P_i(t, S_t^1, ..., S_t^d)$. In general, the delta approximation of an option exposed to two factors is to treat the position as a portfolio of two positions with Δ_i units of the first factor and Δ_j units of the second one where

$$\Delta_i(S^i, S^j) = \frac{\partial}{\partial S^i} P(S^i, S^j)$$

and likewise for $\Delta_j(S^i, S^j)$. Letting ω_i be the ith weight of the portfolio, the portfolio value becomes $V_t = \sum_{i=1}^N \omega_i P_t^i$. We then distinguish

The delta-normal approach assumes that the daily increments are normaly distributed $\Delta S_t \sim N(0, \Omega_t)$ and that the covariance matrix Ω_t is estimated from historical time series. Equivalently, $R_S = \frac{\Delta S_t}{S_t} \sim N(0, C_t)$ with $C_{t,ij} = \sigma_i \sigma_j \rho_{ij}$, such that given $\Delta S_t = S_t R_S$ we get $\Omega_{t,ij} = S_t^i S_t^j \sigma_i \sigma_j \rho_{ij}$. As a result, the variation of the portfolio is approximated by

$$\Delta V \approx \sum_{i=1}^N \sum_{j=1}^d \omega_i \frac{\partial P^i}{\partial S^i} \Delta S^j$$

and the variance of ΔV is given by

$$Var(\Delta V) \approx \sum_{ijkl} \omega_i \frac{\partial P^i}{\partial S^j} \omega_k \frac{\partial P^k}{\partial S^l} \Omega_{jl}$$

The daily VaR for a given confidence level α satisfies

$$VaR(\alpha) = N(\alpha) \sqrt{Var(\Delta V)}$$

where $N(\bullet)$ is the standard normal distribution function (non, z_p).

1.6.5.2 The historical VaR

The historical approach relies on historical data for obtaining risk factor scenarios. For example, considering one year of daily increments $(\Delta S_k^j)_{k=1,..,250}^{j=1,..,d}$, they are used as possible values for computing the portfolio price on the next day

$$\Delta V_k = \sum_{i=1}^{N} \omega_i \Big[P_i(t+\Delta t, S_t^1 + \Delta S_k^1, .., S_t^d + \Delta S_k^d) - P_i(t, S_t^1, .., S_t^d) \Big]$$

That is, the kth simulation trial assumes that the percetage changes in all market factors are as on the kth day, such that for the ith factor we get $S_{t+\Delta t}^i = S_t^i + \Delta S_k^i$. The VaR is then estimated as the corresponding empirical quantile of the change in value of the portfolio ΔV.

1.6.5.3 The Monte Carlo VaR

The Monte Carlo VaR is a full valuation method where the increments ΔS^i are simulated with a statistical model which may include fat tails, intertemporal dependencies such as stochastic volatility, correlations or copula-based cross-sectional dependencies, etc. For example, we sample once from the multivariate distributions of the ΔS^i and we use them to estimate the risk facors at the end of one day. We then revalue the portfolio at the end of the day and compute the change in portfolio ΔV. We repeat this process many times to generate a probability distribution for ΔV, and we estimate the VaR as a fractile of the distribution multiplied by the square root of the number of simulations. It is a flexible approach, but it is also very time consuming. The MC VaR allows for the diffusion of all risk factor types based on statistics and GBM-models, or even more complex models accounting for the non-normality of market returns. For instance, in the case of the implied volatility (IV) diffusion

- the IV surface is evolved based on the statistical characteristics of the at-the-money anchor.
- smile is diffused in order to simulate shift, twist and butterfly effects.
- smile movements must be constrained to avoid unrealistic shapes. The no-arbitrage relationship bwtween option prices must be preseved.

Simulation of a large number of scenarios is performed and a mark-to-market of the portfolio is computed for all scenarios. Hence, as a result of a stock bump, options are repriced under each scenario using a new moneyness on the smile. The VaR is then calculated as the worst distribution value for a given confidence interval.

1.7 The local volatility model

To the extent that the world deviates from the BS assumptions of constant volatility and a lognormal distribution to price changes, the BS model will be biased in certain, often predictable ways (see Hull and White [1987]). In reality, since the volatility is not constant it has a major impact on the values of certain options, especially those options that are away from the money, because the dynamics of the volatility process rapidly change the probability that a given out-of-the-money (OTM) option can reach the exercise price. Hence, the BS model consistently underestimates the value of an option to the extent that volatility is stochastic rather than constant as assumed. Further, since the implied volatilities inferred by the market varies with strike and maturity, using different BS-models for different vanilla prices makes risk management difficult (if not impossible) with the BS-model. The delta and vega risks calculated at one strike are not consistent with the same risks calculated at other strikes. When risk managing a book of options, we generally aggregate the delta and vega risks of all options on a given underlying before hedging, so that only the net exposure of the book is hedged. Hence, we need a single model to price market options for all strikes and all maturities as well as to hedge them. One approach is to consider the local volatility model (LVM), where vanilla option prices with different strikes and different maturities can be obtained by solving a single forward partial differential equation (PDE). However, the LVM is more of a framework than a model, as it is a non-parametric tweaking of Black-Scholes pricing (see Ayache et al. [2004]).

1.7.1 The setup

We consider the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ where $\mathcal{F}_t$ is a right continuous filtration including all $\mathbb{P}$ negligible sets in $\mathcal{F}$. For simplicity, we let the market be complete and assume that there exists an equivalent martingale measure $\mathbb{Q}$ as defined in a mixed diffusion model by Bellamy and Jeanblanc [2000]. In the presence of continuous dividend yield, that unique probability measure equivalent to $\mathbb{P}$ is such that the discounted stock price plus the cumulated dividends are martingale when the riskless asset is the numeraire. There are various ways of modelling discrete dividends, all resulting in different prices for path dependent options (see Appendix (B.2)). In a general setting, we let the underlying process $(S_t^x)_{t\geq 0}$ be a one-dimensional Ito process valued in the open subset D with dynamics under the risk-neutral measure $\mathbb{Q}$ being

$$\frac{dS_t}{S_t} = \mu_t dt + \sigma(t,\omega)dW_S(t) \tag{1.7.37}$$

where $\omega \in \Omega$ denotes dependence on some arbitrary variables and the drift $\mu : D \to \mathbb{R}$ as well as the diffusion $\sigma : D \to \mathbb{R}$ are regular enough to have a unique strong solution valued in D. We let the drift be $\mu_t = r_t^d - r_t^f - q(t)$ where $(r_t)_{t\geq 0}$ is the spot rate (domestic or foreign) process which can be either stochastic or deterministic, and $q(t)$ is a deterministic repo rate. Given the absence of arbitrage opportunity (AAO), the price of a European call option $C(t,x)$ on $]0,T]\times]0,+\infty[$ under the risk-neutral measure $\mathbb{Q}$ is

$$\begin{aligned} C(t,S_t) &= E^Q[e^{-\int_t^T r_s ds} h(S_T)|\mathcal{F}_t] \\ &= P(t,T)\int_{-\infty}^{\infty} h(x)\phi_S(t,x)dx \end{aligned} \tag{1.7.38}$$

where h is a sufficiently smooth payoff function and $\phi_S(t,\cdot)$ is a continuous density function. Furthermore, we assume that these price functions are of class $\mathcal{C}^2$, such that for $T > t$ the law of S_T knowing $S_t = x$ has a density $\phi(T,K)$ given by

$$\phi(T,K) = E_t^Q[\delta(S_T - K)] = \frac{1}{E_t^Q[e^{-\int_t^T r_s ds}|S_T = K]}\partial_{KK}C(t,S_t;K,T) \tag{1.7.39}$$

where $\delta(\cdot)$ is the Dirac function.

1.7.2 The local volatility of the underlying asset

We choose to introduce the local volatility model (LVM) now, because similarly to the BS-formula which corresponds both to a model and a representation of market risk via the implied volatility surface, the LVM also correponds to a model and a representation of market risk via the deterministic local volatility surface (DLV). LVM has a dynamycs of its own (which can be questionable) and provides a numerical methods for finding the DLV that fits option market prices. It is the projection of a multi-dimensional process into a one-factor Markov process with local variance being the local expected variance of the multi-dimensional process.

1.7.2.1 Reducing the dimension of the process

We consider the n-dimensional Ito process $(X_t)_{t\geq 0}$ valued in the open subset D with dynamics

$$\frac{dX_t}{X_t} = \alpha(t,\omega)dt + \beta(t,\omega)dW(t)$$

where $\omega \in \Omega$ denotes dependence on some arbitrary variables and the drift $\alpha : D \to \mathbb{R}^n$ as well as the diffusion $\beta : D \to \mathbb{R}^{n\times n}$ are regular enough to have a unique strong solution valued in D. Gyongy [1986] showed that there exists a Markov process $\tilde{X}_t$ with the same marginal distributions as those of our original process. That process follows the dynamics

$$\frac{d\tilde{X}_t}{\tilde{X}_t} = b(t, \tilde{X}_t)dt + v(t, \tilde{X}_t)d\tilde{W}(t)$$

where at time $t_0 < t$

$$\begin{aligned} b(t,x) &= E_{t_0}^Q[\alpha(t,\omega)|X_t = x] \\ v^2(t,x) &= E_{t_0}^Q[\beta(t,\omega)\beta(t,\omega)^\top|X_t = x] \end{aligned}$$

are respectively the local drift and local variance of the one-dimensional process. Note, they are still deterministic functions of time t and spot value x even though the instantaneous volatility and instantaneous drift are stochastic. Since the marginal distributions of our model are matched to those of a simple Markov process, we can use this approach to calibrate the parameters of a high-dimensional dynamics to European market prices. Therefore, given the one-dimensional underlying process S_t defined in Equation (1.7.37), the local drift and variance are given by

$$\begin{aligned} b(t,S) &= E_{t_0}^Q[\mu(t,\omega)|S_t = S] \\ v^2(t,S) &= E_{t_0}^Q[\sigma(t,\omega)\sigma(t,\omega)^\top|S_t = S] \end{aligned}$$

where S is an arbitrary realisation of S_t for $t \geq 0$. Using Bayes relationship, the local variance becomes

$$v^2(t,S) = \frac{E_{t_0}^Q[\sigma(t,\omega)\sigma(t,\omega)^\top\delta(S_t - S)]}{E_{t_0}^Q[\delta(S_t - S)]}$$

where $\delta(.)$ is the Dirac function. In general, in low dimensions, those expectations are solved numerically by diffusing forward the joint distributions of the underlying asset and the arbitrary variables (see Jex et al. [1999]).

1.7.2.2 The local volatility of the underlying with deterministic rates

Assuming deterministic rates, Dupire [1996] defined the local variance (LV) as the expectation of the future spot variance conditional on a given asset price level. More specifically, given the dynamics of the underlying process in Section (1.7.1) with deterministic interest rates, and applying Gyongy's results, we get

$$\frac{d\tilde{S}_t}{\tilde{S}_t} = \mu_t dt + \sigma_{LV}(t, \tilde{S}_t)d\tilde{W}_S(t)$$

with

$$\sigma_{LV}^2(t,S) = E_{t_0}^Q[\sigma(t,\omega)\sigma(t,\omega)^\top|S_t = S] \tag{1.7.40}$$

which is sometime called the stochastic local variance (SLV) (see Alexander et al. [2004]). Hence, the process $\tilde{S}_t$ has the same marginal distributions as those of the process S_t for all $t \geq 0$. However, S_t and $\tilde{S}_t$ do not follow the same dynamics, resulting in different exotic options prices as well as hedge ratios. In the special case where the instantaneous volatility is a deterministic function of time t and the spot price S_t, then the local volatility is equal to the instantaneous volatility, that is,

$$\sigma_{DLV}^2(t,S) = \sigma(t,S)\sigma(t,S)^\top$$

which we call the deterministic local volatility (DLV). It is the only case where the processes S_t and $\tilde{S}_t$ follow the same dynamics. Dupire [1994] showed that when $n = 1$, the deterministic local volatility can be estimated from market call and put prices of all strikes and maturities. He considered call prices as function of strike K and maturity T and solved the forward PDE

$$\partial_T C(T,K) = \frac{1}{2}\sigma^2_{DLV}(T,K)K^2\partial_{KK}C(T,K) - \mu_T K\partial_K C(T,K) \tag{1.7.41}$$

with initial condition $C(t_0, S_{t_0}, K, t_0) = (S_{t_0} - K)^+$, obtaining a unique solution to the function of volatility

$$\sigma^2_{DLV}(T,K) = \frac{\partial_T C(t,S_t,K,T) + K(r_T - r^f_T - q_T)\partial_K C(t,S_t,K,T) + (r^f_T + q_T)C(t,S_t,K,T)}{\frac{1}{2}C_{KK}(t,S_t,T,K)K^2} \tag{1.7.42}$$

This result is crucial because it allows perfect fit to the market prices leading to a complete modification of the equity market.

Remark 1.7.1 *The Black-Scholes implied volatility is no longer the only form of observable volatility, and one can directly use the local volatility for calibration.*

So, in principle one can directly diffuse the process S_t with stochastic instantaneous volatility provided that its local variance matches that of the DLV

$$\sigma^2_{LV}(t,S) = \sigma^2_{DLV}(t,S) \tag{1.7.43}$$

That is, we may impose any dynamics such that the above equality holds and the local variance stays non-negative. This is very appealing because we get perfect fit to the market prices, but it is difficult to implement in practice.

1.7.2.3 The local fixed point algorithm

We are going to describe the local fixed point algorithm used to solve the stochastic local variance (SLV) given in Equation (1.7.40). For simplicity of exposition we assume that the instantaneous volatility of the process in Equation (1.7.37) can be written as

$$\sigma(t,\omega) = \sigma(t,S_t)\sqrt{V_t}$$

where V_t is an Ito process. For the Markov process $\tilde{S}_t$ to match the marginal distributions of the original process S_t, we must equate their local volatilities as in Equation (1.7.43). Thus, we get

$$E^Q[\sigma^2(t,S_t)V_t|S_t = K] = E^Q[\sigma^2_{DLV}(t,\tilde{S}_t)|S_t = K]$$

which gives

$$\sigma^2_{SLV}(t,K) = \frac{\sigma^2_{DLV}(t,K)}{E^Q[V_t|S_t = K]}$$

In practice, we need to discretise the time domain into n time steps $t_0, t_1, ..., t_n$ and rewrite the above equation in discrte time as follow

$$\sigma^2_{SLV}(t_{i+1},K) = \frac{\sigma^2_{DLV}(t_{i+1},K)}{E^Q[V_{t_{i+1}}|S_{t_{i+1}} = K]}$$

If we assume that the process $S_{t_{i+1}}$ is a function of $\sigma(t_i,\cdot)$ and $\sigma(t_{i+1},\cdot)$, and if we further assume that over a small time increment Δt we set $\sigma(t_i,\cdot) \approx \sigma_{LV}(t_i,\cdot)$, then we can rewrite the local variance as

$$\sigma^2_{SLV}(t_{i+1},K) = \frac{\sigma^2_{DLV}(t_{i+1},K)}{E^Q[V_{t_{i+1}}|S_{t_{i+1}}(\sigma_{LV}(t_i,\cdot),\sigma_{LV}(t_{i+1},\cdot)) = K]}$$

where the local volatility $\sigma_{LV}(t_{i+1}, \cdot)$ now appears on both side of the equation. It is usually solved with a fixed point algorithm. Starting at time t_0 a first guess is chosen $\sigma_{LV}(t_0, \cdot)$ and then simulating the underlying process between the times t_i and t_{i+1} the local volatility $\sigma_{LV}(t_1, \cdot)$ is solved. Then, we iterate the process for all time t_i, $i = 1, ..., n$. In the special case where we assume that the process $S_{t_{i+1}}$ is only a function of $\sigma(t_i, \cdot)$, then simplifications can be made. For example if we discretise $S_{t_{i+1}}$ with the Euler scheme

$$S_{t_{i+1}} = S_{t_{i+1}}\big(1 + \sigma_{LV}(t_i, S_{t_i})\sqrt{V_{t_i}}(W_{t_{i+1}} - W_{t_i})\big)$$

the local variance simplies to

$$\sigma^2_{SLV}(t_{i+1}, K) = \frac{\sigma^2_{DLV}(t_{i+1}, K)}{E^Q[V_{t_{i+1}}|S_{t_{i+1}}\big(\sigma_{LV}(t_i, \cdot)\big) = K]}$$

and we no-longer need to iterate to solve for the local variance. This simplification is called the practical method. Note, a solution may not exist, leading the algorithm to diverge.

1.7.2.4 The risk-neutral density

From Remark (1.7.1), we can infer the market risk-neutral density from the local volatility.
First, we can express the local volatility directly in terms of implied volatility $\Sigma(T, K)$ as follow

$$\sigma^2(T, K) = \frac{\Sigma^2(T, K) + 2\Sigma(T, K)(T-t)\big(\partial_T\Sigma(T, K) + (r_T - q_T)K\partial_k\Sigma(T, K)\big)}{1 + \Sigma(T, K)(T-t)K^2\Big(\frac{2d_1}{K\Sigma(T,K)\sqrt{T-t}}\partial_K\Sigma(T, K) + \partial_{KK}\Sigma(T, K) + \frac{d_1 d_2}{\Sigma(T,K)}(\partial_K\Sigma(T, K))^2\Big)} \quad (1.7.44)$$

or we can factorise the numerator with $\Sigma(T, K)(T-t)$. The implied volatility being a function of the strike K and the maturity T, we use the chain rule to express the derivatives of the market prices in terms of the Black-Scholes Greeks. Concentrating on the density $C_{KK}(t, S_t, T, K)$ (see Equation (1.7.39)), we get

$$\begin{aligned} K^2\frac{\Sigma(T, K)(T-t)}{Vega(K, T)}C_{KK}(t, S_t, T, K) &= \big(1 + Kd_1\sqrt{T-t}\partial_K\Sigma(K, T)\big)^2 \\ + \quad K^2\Sigma(K, T)(T-t)\big(\partial_{KK}\Sigma(K, T) - d_1(\partial_K\Sigma(K, T))^2\sqrt{T-t}\big) \end{aligned}$$

with $d_1 = \frac{\ln(S_t/S_T) + (r + \frac{1}{2}\Sigma^2)(T-t)}{\Sigma\sqrt{T-t}}$. Rearranging, we get the convexity of the smile expressed in prices as

$$\begin{aligned} &\partial_{KK}\Sigma(K, T) = \quad (1.7.45) \\ &\frac{\partial_{KK}C(t, S_t, T, K)}{Vega(K, T)} - \frac{1}{K^2\Sigma(T, K)(T-t)}\Big[1 + 2Kd_1\sqrt{T-t}\partial_K\Sigma(K, T) + K^2 d_1 d_2(T-t)(\partial_K\Sigma(K, T))^2\Big] \end{aligned}$$

Quantitatively speaking, the implied risk-neutral probability density function (pdf) in the local volatility model is given by

$$\phi_{LV}(S_t, t; S_T, T) = \phi_{SN}(d_1)p(d_1) \quad (1.7.46)$$

where $\phi_{SN}(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$ is the standard normal density, and

$$p(d_1) = \frac{1}{S_T\Sigma\sqrt{T-t}}\left(\left(1 + S_T d_1\sqrt{T-t}\frac{\partial\Sigma}{\partial K}\right)^2 + S_T^2(T-t)\Sigma\left(\frac{\partial^2\Sigma}{\partial K^2} - d_1\left(\frac{\partial\Sigma}{\partial K}\right)^2\sqrt{T-t}\right)\right)$$

is a quadratic function in d_1. Note, the risk-neutral density is a function of the skew $\frac{\partial \Sigma(T,K)}{\partial K}$ and the curvature $\frac{\partial^2 \Sigma(T,K)}{\partial K^2}$. However, for the density to be valid we must impose some constraints on the shape of the skew and the curvature. We can easily observe that if the quantity

$$\frac{\partial^2 \Sigma}{\partial K^2} - d_1 \left(\frac{\partial \Sigma}{\partial K} \right)^2 \sqrt{T-t} \tag{1.7.47}$$

is too negative, which happens when the slope of the smile $\frac{\partial \Sigma}{\partial K}$ is too high, the pdf can become negative. This means that the European call prices are not the result of the expectancy under the risk-neutral measure of future pay-off and therefore allow for arbitrage opportunities [4]. Note, we can rewrite the risk-neutral density in terms of the skew $\frac{\partial \Sigma}{\partial K}$ and the curvature $\frac{\partial^2 \Sigma}{\partial K^2}$ as follow

$$p(d_1) = \frac{1}{S_T \Sigma \sqrt{T-t}} + \frac{2d_1}{\Sigma} \frac{\partial \Sigma}{\partial K} - S_T \sqrt{T-t} \big[\sqrt{T-t} d_1 - \frac{1}{\Sigma} d_1^2 \big] (\frac{\partial \Sigma}{\partial K})^2 + S_T \sqrt{T-t} \frac{\partial^2 \Sigma}{\partial K^2} \tag{1.7.48}$$

In the case where we model the logarithm of the stock price, we denote the implied volatility as $\Sigma(e^k, T)$ and get the local volatility as

$$E_t^Q[|\sigma(T,\omega)|^2 | S_T = e^k] = \frac{\partial_T C(t,S_t,e^k,T) + (r_T - q_T)\partial_k C(t,S_t,e^k,T) + q_T C(t,S_t,e^k,T)}{\frac{1}{2}\big(\partial_{kk} C(t,S_t,e^k,T) - \partial_k C(t,S_t,e^k,T)\big)}$$

Then, the risk-neutral density satisfies

$$\begin{aligned}
& \frac{\Sigma(T,e^k)(T-t)}{Vega(K,T)} \big(\partial_{kk} C(t,S_t,e^k,T) - \partial_k C(t,S_t,e^k,T)\big) \\
= \quad & 1 + \Sigma(T,e^k)(T-t)\Big(\big(\frac{2d_1}{\Sigma(T,e^k)\sqrt{T-t}} - 1 \big) \partial_k \Sigma(T,e^k) + \partial_{kk}\Sigma(T,e^k) + \frac{d_1 d_2}{\Sigma(T,e^k)} (\partial_k \Sigma(T,e^k))^2 \Big)
\end{aligned}$$

where $d_1 = \frac{x-k+(r-\frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}}$. Since $d_1 = d_2 + \sigma\sqrt{t}$ we can rewrite the density as

$$\begin{aligned}
& \frac{\Sigma(T,e^k)(T-t)}{Vega(K,T)} \big(\partial_{kk} C(t,S_t,e^k,T) - \partial_k C(t,S_t,e^k,T)\big) \\
= \quad & 1 + \Sigma(T,e^k)(T-t)\Big(\big(\frac{2d_1}{\Sigma(T,e^k)\sqrt{T-t}} - 1 \big) \partial_k \Sigma(T,e^k) + \partial_{kk}\Sigma(T,e^k) + \frac{d_1^2 - d_1 \Sigma(T,e^k)\sqrt{T-t}}{\Sigma(T,e^k)} (\partial_k \Sigma(T,e^k))^2 \Big)
\end{aligned}$$

Expanding and factorising, we get

$$\begin{aligned}
& \frac{\Sigma(T,e^k)(T-t)}{Vega(K,T)} \big(\partial_{kk} C(t,S_t,e^k,T) - \partial_k C(t,S_t,e^k,T)\big) \\
= \quad & \big(1 + d_1\sqrt{T-t}\partial_k \Sigma(T,e^k)\big)^2 + \Sigma(T,e^k)(T-t)\Big(\partial_{kk}\Sigma(T,e^k) - \partial_k \Sigma(T,e^k) - d_1\sqrt{T-t}(\partial_k \Sigma(T,e^k))^2 \Big)
\end{aligned}$$

The Vega for the logarithm of the stock is given by $Vega = e^x Re(t,T)\sqrt{T-t}\phi_{SN}(d_1)$, so that the density becomes

[4] In reality the slope of the smile can be higher than the one calculated using the risk-neutral measure as the market is incomplete. However we restrict ourselves to the complete market assumption.

$$p(d_1) = \frac{S_T}{\Sigma\sqrt{T-t}}\Big(\big(1+d_1\sqrt{T-t}\partial_k\Sigma(T,e^k)\big)^2+\Sigma(T,e^k)(T-t)\Big(\partial_{kk}\Sigma(T,e^k)-\partial_k\Sigma(T,e^k)-d_1\sqrt{T-t}(\partial_k\Sigma(T,e^k))^2\Big)\Big)$$

We can rewrite the risk-neutral density for the logarithm of the spot price in terms of skew $\partial_k\Sigma(T,e^k)$ and curvature $\partial_{kk}\Sigma(T,e^k)$ as follow

$$\begin{aligned}
p(d_1) &= \frac{S_T}{\Sigma\sqrt{T-t}}+S_T\big[2\frac{d_1}{\Sigma(T,e^k)}-\sqrt{T-t}\big]\partial_k\Sigma(T,e^k) \\
&+ S_T\sqrt{T-t}\big[\frac{d_1^2}{\Sigma(T,e^k)}-d_1\sqrt{T-t}\big](\partial_k\Sigma(T,e^k))^2+S_T\sqrt{T-t}\partial_{kk}\Sigma(T,e^k)
\end{aligned}$$

1.7.2.5 Point on Kolmogorov-Compatibility

Dealing with deterministic smile surface evolution always rises the question of whether the smile surface allows for arbitrage opportunities. It is natural to impose that future smile surface are compatible with today's prices of calls and puts. The problem is that defining a deterministic smile surface $\sigma(t,T,K,S_t)$ means defining a future deterministic density ϕ for stock process, and then a natural condition would be to impose that the future pdf is actually a conditional density. This is what we will formalise as the Kolmogorov-Compatibility. First of all, let's remind the relation between prices and pdf. Without doing any assumptions apart from that of no arbitrage opportunity, we have

$$\begin{aligned}
C(t,K,t_1-t_0) &= e^{-r(t_1-t_0)}\mathbb{E}\left((S_{t_1}-K)^+|\mathcal{F}_{t_1}\right)=e^{-r(t_1-t_0)}\int_K^{+\infty}(x_1-K)\phi(x_1,t_1,S_{t_0},t_0)d\lambda(x_1) \\
\frac{\partial C}{\partial K}(t,K,t_1-t_0) &= -e^{-r(t_1-t_0)}\int_K^{+\infty}\phi(x_1,t_1,S_{t_0},t_0)d\lambda(x_1) \\
\frac{\partial^2 C}{\partial K^2}(t,K,t_1-t_0) &= e^{-r(t_1-t_0)}\phi(K,t_1,S_{t_0},t_0)
\end{aligned}$$

By definition $C(t,K,t_1-t_0)=C_{BS}(S_{t_0};K,t_1-t_0;r,\Sigma(S_{t_0};K,t_1-t_0))$, such that applying the chain rule we have

$$\frac{\partial^2 C}{\partial K^2}=\frac{\partial^2 C_{BS}}{\partial K^2}+\frac{\partial^2 C_{BS}}{\partial K\partial\Sigma}\frac{\partial\Sigma}{\partial K}+\left[\frac{\partial^2 C_{BS}}{\partial K\partial\Sigma}+\frac{\partial^2 C_{BS}}{\partial K^2}\frac{\partial\Sigma}{\partial K}\right]\frac{\partial\Sigma}{\partial K}+\frac{\partial C_{BS}}{\partial\Sigma}\frac{\partial^2\Sigma}{\partial K^2}$$

For the sake of clarity we denote by Θ the right-hand-side operator.

Definition 1.7.1 (Kolmogorov-Compatibility) *Any future deterministic conditional density, or smile surface, such that*

$$\begin{aligned}
&\Theta(S_0;K,T-t_0;r,\sigma(t_0,T,K,S_{t_0}))= \\
&\int\Theta(S_0;K,t_1-t_0;r,\sigma(t_0,t_1,K,S_{t_0}))\Theta(S_{t_1};K,T-t_1;r,\sigma(t_1,T,K,S_{t_1}))d\lambda(S_{t_1})
\end{aligned}$$

is satisfied, defines a Kolmogorov-compatible density.

Property 1.7.1 *Given a current admissible*[5] *smile surface, if the future smile surface is Kolmogorov-compatible no model-independent strategy can generate arbitrage profits.*

[5] *i.e.* such that the associated call prices sarisfy $\frac{\partial Call(t,K,T-t)}{\partial K}<0$, $\frac{\partial^2 Call(t,K,T-t)}{\partial K^2}>0$, $\frac{\partial Call(t,K,T-t)}{\partial T}>0$, $\frac{\partial Put(t,K,T-t)}{\partial K}>0$, $Call(t,K,T-t)|_{K=0}=S_t$ and $lim_{K\to\infty}Call(t,K,T-t)=0$

There is in general an infinity of solutions for the forward density

$$\Theta(S_{t_1}; K, T - t_1; r, \sigma(t_1, T, K, S_{t_1}))$$

such that Definition (1.7.1) is satisfied. Therefore, even if we require the smile surface to be deterministic, there still exists an infinity of future smile surfaces compatible with today's prices of calls and puts. Additional information is needed to determine these conditional distributions. We need to know all forward smiles, that is, the future prices of all vanilla options as seen from all possible states of the world.

1.7.3 The relation between local and implied volatility

There is a link between the local volatility and the implied volatility. Berestycki et al. [2002] showed that when maturities are very short, the implied volatility is equal to the harmonic average of the local volatilities. As a result, given the implied volatility $\Sigma(T, K)$ and the forward log-moneyness $\overline{\eta} = \log \frac{SRe(0,T)}{KP(0,T)}$, the implied volatility in the log-moneyness metric becomes $\Sigma(T, K) = \tilde{\Sigma}(T, \overline{\eta})$ [6] . When the local volatility $\sigma_L(0, \overline{\eta})$ is differentiable at the strike $\overline{\eta} = 0$, we get

$$\frac{\partial}{\partial \overline{\eta}} \tilde{\Sigma}(0, 0) = \frac{1}{2} \frac{\partial}{\partial \overline{\eta}} \sigma_L(0, 0)$$

As a result, for short maturities, the at-the-money skew of the local volatility is twice that of the implied volatility. For more properties on the local volatility and for a detailed explaination of the interaction between the local volatility and the implied volatility see Derman [2008].
For example, the rule of 2 states that local volatilities grow approximately twice as fast with price as implied volatilities grow with strike. To gain intuition, Derman let the implied volatility with strike K and undrerlying asset S be roughly the average of local volatility in the region between S and K, given by

$$\Sigma(S, K) \approx \frac{1}{K - S} \int_S^K \sigma_L(u) du$$

Assuming the simple local volatility function

$$\sigma_L(S) = \sigma_0 + \beta S \, , \forall t$$

and integrating, the implied volatility becomes

$$\Sigma(S, K) \approx \sigma_0 + \frac{\beta}{2}(S + K)$$

which can be rewritten in terms of the local volatility as

$$\Sigma(S, K) \approx \sigma_L(S) + \frac{\beta}{2}(K - S)$$

More formally, we can use the local volatility in Equation (1.7.44) to understand its relation with the implied volatility. First, factorising the numerator with $\Sigma(T, K)(T - t)$ and re-arranging the denominator, the local volatility becomes

$$\sigma_L^2(T, K) = \frac{\frac{\Sigma(T,K)}{T-t} + 2\big(\partial_T \Sigma(T, K) + (r_T - q_T) K \partial_k \Sigma(T, K)\big)}{K^2 \Big(\frac{1}{\Sigma(T,K)(T-t)K^2} + \frac{2d_1}{K\Sigma(T,K)\sqrt{T-t}} \partial_K \Sigma(T, K) + \partial_{KK} \Sigma(T, K) + \frac{d_1 d_2}{\Sigma(T,K)} (\partial_K \Sigma(T, K))^2\Big)}$$

Noting that

[6] Here it is understood as $(T, K) \rightarrow \tilde{\Sigma}(T, \overline{\eta})$.

$$\frac{1}{\Sigma(T,K)}\Big(\frac{1}{(T-t)K^2}+\frac{2d_1}{K\sqrt{T-t}}\partial_K\Sigma(T,K)+d_1d_2(\partial_K\Sigma(T,K))^2\Big) = \frac{1}{\Sigma(T,K)}\big(\frac{1}{K\sqrt{T-t}}+d_1\partial_K\Sigma(T,K)\big)^2 - d_1\sqrt{T-t}(\partial_K\Sigma(T,K))^2$$

we recover

$$\sigma_L^2(T,K)=\frac{\frac{\Sigma(T,K)}{T-t}+2\big(\partial_T\Sigma(T,K)+(r_T-q_T)K\partial_K\Sigma(T,K)\big)}{K^2\Big(\frac{1}{\Sigma(T,K)}\big(\frac{1}{K\sqrt{T-t}}+d_1\partial_K\Sigma(T,K)\big)^2-d_1\sqrt{T-t}(\partial_K\Sigma(T,K))^2+\partial_{KK}\Sigma(T,K)\Big)}$$

Assuming no interest rates and no repo rates, we let the implied volatility be indenendent from the strike K. The local volatility simplifies to

$$\sigma_L^2(T,K)=\Sigma^2(T,K)+2\Sigma(T,K)(T-t)\partial_T\Sigma(T,K)$$

Setting $\tau=T-t$, it can be written as $\sigma_L^2(\tau)=\partial_\tau\big(\Sigma^2(\tau)\tau\big)$, so that

$$\tau\Sigma^2(\tau)=\int_0^\tau\sigma_L^2(u)du$$

which expresses the total variance as an average of forward variances.
Next, we let the implied volatility be independent from maturity, and assume a weak linear dependence of the skew on K, such that $\partial_{KK}\Sigma(T,K)$ and $\partial_K\Sigma(T,K)$ are negligible. The local volatility is approximated by

$$\sigma_L^2(T,K)\approx\frac{\Sigma^2(K)}{\big(1+d_1K\sqrt{T-t}\partial_K\Sigma(K)\big)^2} \tag{1.7.49}$$

such that

$$\sigma_L(K)\approx\frac{\Sigma(K)}{1+d_1K\sqrt{T-t}\partial_K\Sigma(K)}$$

If we now assume that the strike is close to the money, $K=S+\Delta K$, then

$$d_1\approx\frac{1}{\Sigma\sqrt{T-t}}\ln\frac{S}{K}\approx-\frac{\Delta K}{S(\Sigma\sqrt{T-t})}\approx-\frac{\Delta K}{K(\Sigma\sqrt{T-t})}$$

and the local volatility becomes

$$\sigma_L(K)\approx\frac{\Sigma(K)}{1-\frac{\Delta K}{\Sigma(K)}\partial_K\Sigma(K)}\approx\Sigma(K)\big(1+\frac{\Delta K}{\Sigma(K)}\partial_K\Sigma(K)\big)\approx\Sigma(K)+(\Delta K)\partial_K\Sigma(K)$$

As a result, we can write the local volatility as a function of $S+\Delta K$ as follow

$$\sigma_L(S+\Delta K)\approx\Sigma(S+\Delta K)+(\Delta K)\partial_K\Sigma(K)$$

and peforming a Taylor expansion around S, we get

$$\partial_S\sigma_L(S)=2\partial_K\Sigma(K)\big|_{K=S}$$

which says that the local volatility grows twice as fast with the stock price S as the implied volatility $\Sigma(K)$ grows with the strike. It is interesting to note that if we set $\tau=(T-t)$ in the local volatility in Equation (1.7.44), in the limit $\tau\to0$, we recover Equation (1.7.49). Further, $\sqrt{\tau}Kd_1\to\frac{K}{\Sigma}\ln\frac{S}{K}$ as $\tau\to0$, such that

$$\sigma_L(T,K) = \frac{\Sigma(K)}{1 + \frac{K}{\Sigma(K)} \ln \frac{S}{K} \partial_K \Sigma(K)}$$

which we can rewrite in the log-moneyness metric $\overline{\eta}$, with implied volatility $\tilde{\Sigma}(\overline{\eta})$, as

$$\sigma_L(\overline{\eta}) = \frac{\tilde{\Sigma}(\overline{\eta})}{1 - \frac{\overline{\eta}}{\tilde{\Sigma}(\overline{\eta})} \partial_{\overline{\eta}} \tilde{\Sigma}(\overline{\eta})}$$

since $\frac{dK}{d\overline{\eta}} = -K$. The solution satisfies

$$\frac{1}{\tilde{\Sigma}(\overline{\eta})} = \frac{1}{\overline{\eta}} \int_0^{\overline{\eta}} \frac{1}{\sigma_L(u)} du$$

and for very short time to maturity, the implied volatility in the log-moneyness metric is the harmonic mean of the local volatility.

1.7.4 Some simple local volatility models

1.7.4.1 The parametric local volatility

Another commonly used way of getting asymmetric risk-neutral distribution to recover the market implied skew is to use a parametric local volatility, such as the constant elasticity of variance (CEV) (see Cox et al. [1976]) with local volatility function being

$$\sigma(t,x) = \nu(t)(\frac{x}{L_t})^{\beta(t)-1}$$

and with simplified pricing formula given by Schroder [1989]. Alternatively, one can shift the underlying price with the function of time $\nu(t)$ getting the displaced diffusion model (see Rubinstein [1983]) such that the underlying dynamics become

$$dS_t = \sigma_D(t) Y_t dW_S(t) \tag{1.7.50}$$

where $\sigma_D(t) > 0$ and $Y_t = \nu(t)S_t + (1-\nu(t))S_0$. Applying Ito's lemma, we get the solution $Y_t = S_0 M_t$ where

$$M_t = e^{-\frac{1}{2}\int_0^t \nu^2(s)\sigma_D^2(s)ds + \int_0^t \nu(s)\sigma_D(s)dW(s)}$$

Note, $\nu(t)$ is a blending parameter taking value in the range $[0,1]$ and controlling the slope of the local volatility function $\psi(t,x) = \nu(t)x + (1-\nu(t))x_0$. In the special case where rates are deterministic, setting $\nu = 0$ we recover the Gaussian distribution, while when $\nu = 1$ we recover the lognormal distribution. Also, since the local volatility $\psi(t,x)$ is linear in x the model can be transformed into an Affine model and classical pricing formula of call options can be applied, provided that we modify the characteristic function of the logarithm of the forward price and that we shift the strike of the option. That is, the underlying price seen at time t is given by

$$S_t = \frac{S_0}{\nu(t)}\big(M(t) - (1-\nu(t))\big)$$

where the process $M(t)$ follows the dynamics

$$\frac{dM(t)}{M(t)} = \sigma_M(t) dW_S(t) \text{ with } M(0) = 1$$

with $\sigma_M(t) = \nu(t)\sigma_D(t)$. Note, the underlying asset can become negative even if $\nu(t) > 0$ unless $\nu(t) = 1$. Therefore, to generate some skew the displaced diffusion model must allow for the underlying price to become negative. One can calibrate the model parameters $\nu(t)$ and $\sigma(t)$ to the European call price C_t seen at time t with a strike K and maturity T satisfying

$$C(t,S_t,T,K) = P(t,T)E^Q\big[\big(\frac{S_t}{\nu(T)}M(T) - \overline{K}\big)^+|\mathcal{F}_t\big]$$

where $\overline{K} = \frac{1-\nu(T)}{\nu(T)}S_t + K$. The displaced diffusion model became popular among market practitioners since it is the simplest way to introduce an implied volatility skew and vanilla option are priced with the standard Black-Scholes formula with appropriate input parameters. In the presence of a non-zero drift, the dynamics of the underlying price become

$$\frac{dS_t}{S_t} = \mu_t dt + \gamma(t,S_t)dW(t)$$

where the local volatility function is

$$\gamma(t,x) = \sigma_D(t)\big(\nu(t) + (1-\nu(t))\frac{F(0,t)}{x}\big)$$

with $F(0,t)$ being the initial forward price with delivery date t. In that setting, $S_t + \frac{1-\nu(t)}{\nu(t)}F(0,t)$ is not a geometric Brownian motion due to the drift term and the Black-Scholes formula can no-longer be used to price vanilla options. One possibility is to set the interest rates to zero, to multiply the strike with the discount factor and to diffuse the process $\widehat{S}_t = \frac{S_t}{Re(0,t)}$ which is obtained by reinvesting the dividend in the stock price. When rates are stochastic, one should diffuse the forward price under the forward measure. However, contrary to deterministic local variance (DLV), the instantaneous variance of the forward price is not a function of the forward price but it is a function of the underlying asset $S_t = F(t,T)\frac{P^d(t,T)}{P^f(t,T)}$ with stochastic zero-coupon bond prices. It is therefore difficult to solve analytically, but approximations can be made. In the special case where the instantaneous volatility of the underlying asset is deterministic (CEV) and the rates are Gaussian, Piterbarg [2005] introduced the autonomous representation of the underlying process $\tilde{F}(t,T)$ and approximated its local volatility function.

1.7.4.2 Similarity between displaced diffusion and CEV model

Given the process $(S_t)_{t\geq 0}$, the dynamics of the underlying price are

$$dS_t = \sigma_\alpha\big(S_t + \alpha\big)dW_t$$

where α is the displacement parameter. The displaced diffusion process admits negative values for $\alpha > 0$. A general representation of the dynamics of the underlying price are

$$dS_t = \sigma\phi(S_t)dW_t$$

where $\phi(x)$ is a possibly non-linear function returning positive value. As an example, another commonly used model that exhibit an implied volatility skew is the CEV model ($\phi(x) = x^\beta$) introduced by Cox et al. [1976], where the dynamics of the underlying price are

$$dS_t = \sigma_\beta S_t^\beta dW_t$$

where the elasticity parameter β is a positive constant such that $0 \leq \beta \leq 1$ and where ϕ is not defined for negative arguments. For $\beta < 1$, the origin (zero) is an attainable and absorbing boundary for the process which is not desirable for numerical implementation especially in the pricing of path-dependent options. Andersen et al. [1998] described some of the properties of the CEV model. Marris [1999] showed that by setting

$$\sigma_\alpha = \sigma_\beta\beta S_0^{\beta-1} \text{ and } \alpha = \frac{1-\beta}{\beta}S_0$$

the displaced diffusion becomes

$$dS_t = \sigma_\beta S_0^{\beta-1}\big(\beta S_t + (1-\beta)S_0\big)dW_t$$

and produces vanilla option prices similar to those produced by the CEV model. Later, Svoboda-Greenwood [2007] proved that the parameters transformation were the result of a simple linearisation of the CEV dynamics around the initial value S_0. Therefore, considering the simple displaced diffusion model in Section (1.7.4.1), in order to recover the vanilla option prices produced by the CEV model, we must impose

$$\sigma_D = \sigma_\beta S_0^{\beta-1} \text{ and } \nu = \beta$$

That is, given the elasticity parameter β and the elasticity volatility σ_β, the displaced diffusion volatility depends in a non-linear way on the initial value S_0. Muck [2005] tested the approximated displaced diffusion against the CEV model on swaption prices and barrier swaption prices with $\beta = 0.5$, $\sigma_\beta = 0.12$ and initial forward rate $S_0 = 0.06$ which gives a displaced diffusion volatility of $\sigma_D = 0.4898$. However, it is not directly applicable to the Equity market where the index can be as high as $S_0 = 20000$ on the Japanese market, leading to an almost zero displaced diffusion volatility.

1.7.4.3 Mixture models

In a market with a limited number of option prices, a model of interpolation and extrapolation of the volatility surface should have few parameters with the ability of mapping a large family of surfaces. However, a deterministic local volatility model is not sufficiently flexible to calibrate a large number of points on the volatility surface when specified with a small number of parameters. Mixture models provide an alternative characterisation of the deterministic local volatility by finding a diffusion process with marginal law following a given evolving mixture probability densities. Melick et al. [1997] assumed the price of the underlying asset to follow a mixture of lognormals where the more lognormal distributions are included the better the fit to market option prices. Since then, mixture models have been well studied in the literature (see for example Brigo et al. [2000b] and Alexander [2003]). In particular Brigo et al. introduced a class of models for the dynamics of the asset price under the assumption that the asset-price density is given by a mixture of known basic densities. Assuming the dynamics of the asset price to be given by Equation (1.7.37), we let $\mu = r - q$ and set $\sigma(t, S_t) = \sigma(t)S_t$, to create a local volatility model, and fulfill the Lipschitz and linear growth conditions necessary to ensure a unique solution to the SDE. Letting $f_t^i(\bullet)$ be the density function of the ith stock S_t^i, Brigo et al. showed that

$$f_t := \frac{d}{dy}\mathbb{Q}\{S_t \le y\} = \sum_{i=1}^N \lambda_i \frac{d}{dy}\mathbb{Q}\{S_t^i \le y\} = \sum_{i=1}^N \lambda_i f_t^i(y)$$

is a $\mathbb{Q}$-density, where the λ_i are weights summing to one. An immediate consequence of their derivation is that under the $\mathbb{Q}$ measure, the price of a European contingent claim will be given by the weighted sum of prices on claims for each of the assets S_t^i. Setting $\omega = 1$ for a call option and $\omega = -1$ for a put option, then the price of the European option seen at time t for the maturity T and strike K becomes

$$\begin{aligned}
V(S_t, T, K, \sigma_t) &= P(t,T)\mathbb{E}_t^{\mathbb{Q}}\left[\omega(S_T - K)^+\right] \\
&= P(t,T)\int_0^\infty \left[\omega(y-K)^+\right]\sum_{i=1}^N \lambda_i f_T^i(y)dy \\
&= \sum_{i=1}^N P(t,T)\lambda_i \int_0^\infty \left[\omega(y-K)^+\right] f_T^i(y)dy \\
&= \sum_{i=1}^N \lambda_i V_i(S_t, T, K, \sigma_t^i)
\end{aligned}$$

where $P(t,T) = e^{-r(T-t)}$. Brigo et al. noted that the use of densities leading to explicit options prices makes extremely tractable models, whilst the potentially unlimited number of parameters in the asset price dynamics can help achieving satisfactory fits to market data. For example, setting $\sigma(t,S_t) = \sigma(t)S_t$ to generate a log-normal density model for each of the asset prices processes $(S_t)_{i=1}^N$, each stock S_t^i has a marginal density conditional on S_0 given by

$$f_t^i(y) = \frac{1}{y\Sigma_i(t)\sqrt{2\pi}} \exp\left(-\frac{1}{2\Sigma_i^2(t)}\left[\ln\frac{y}{S_0} - \mu t + \frac{1}{2}\Sigma_i^2(t)\right]^2\right)$$

where $\Sigma_i(t) = \sqrt{\int_0^t \sigma_i^2(s)ds}$ is the time-dependent volatility. Clearly, the resulting option prices, representing a weighted sum of expectations of log-normal densities, will reduce to a weighted sum of Black-Scholes prices

$$\begin{aligned}\frac{V(S_t,T,K,\sigma_t)}{P(t,T)} &= \omega\sum_{i=1}^N \lambda_i\left[S_t e^{\mu(T-t)}\Phi\left(\omega\frac{\ln\frac{S_t}{K} + (\mu+\frac{1}{2}\eta_i^2)(T-t))}{\eta_i\sqrt{T-t}}\right) - K\Phi\left(\omega\frac{\ln\frac{S_t}{K} + (\mu-\frac{1}{2}\eta_i^2)(T-t)}{\eta_i\sqrt{T-t}}\right)\right] \\ &= \sum_{i=1}^N \lambda_i V_{BS}(S_t^i,K,T,r,\sigma_t^i) \qquad (1.7.51)\end{aligned}$$

where $\eta_i := \frac{\Sigma_i(T-t)}{\sqrt{T-t}}$.

Chapter 2

The importance of asset returns

2.1 Statistical tools: A review

We saw in Section (1.1.2) that the decisions for the appropriate pricing of options are made contingent on the behaviour of the underlying securities. The uncertainty affecting the underlying is modelled by considering future trajectories of the risky asset seen as possible scenarios. Consequently, it is important to understand the properties of the market returns in order to devise their dynamics. We are first going to recall some statistical tools necessary to model asset returns.

2.1.1 Towards stable distributions

2.1.1.1 The summation of independent random variables

When formally studying random variables, the summation of independent random variables $X_1, X_2, ..., X_n$ has attracted a lot of attention. For instance we can consider a smooth function

$$Y = f(X_1, X_2, ..., X_n)$$

where the random variables $X_1, X_2, ..., X_n$ represent small and independent actions on the system under consideration. Similar situations take place when analysing observational errors arising in experiments. For instance, Laplace [1774] and Gauss [1809] associated the error distribution with the scheme of summation of random variables (see details in Uchaikin et al. [1999]). The theory they developed assumes that the random variables in the sum are small and independent. Originally, J. Bernoulli [1713] considered the sequence of normalised sums $\frac{1}{n}\sum_{i=1}^{n} X_i$ where independent random variables X_i take the value 1 with probability p and the value 0 with probability $1-p$. Bernoulli's therorm states that for any arbitrary small but fixed $\epsilon > 0$ we have

$$P(\big|\frac{1}{n}S_n - p\big| > \epsilon) \to 0 \, , \, n \to \infty$$

where $S_n = \sum_{i=1}^{n} X_i$. It is the Bernoulli's form of the law of large numbers. The law of large numbers and its various generalisations connect together the theory and practice. This is a special case of the Central Limit Theory (CLT) of probabilty theory. We now consider a sequence of independent random variables $X_1, X_2, ...$ possessing one and the same distribution function. Assuming that the mathematical expectation $a = E[X_i]$ and the variance $\sigma^2 = Var(X_i)$ of these variables are finite, we construct the corresponding sequence of the normalised sums $Z_1, Z_2, ...$

$$Z_n = \frac{S_n - na}{\sigma\sqrt{n}}$$

Then for any $x_1 < x_2$

$$P(x_1 < Z_n < x_2) \Rightarrow \int_{x_1}^{x_2} p^G(x)dx \, , n \to \infty$$

where

$$p^G(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

is the density of the standard normal law. The aim of the generalised CLT was the use of not only the normal law as limiting approximation, but also some other distributions of a certain analytical structure. Considering the i.i.d, sequence of random variables $X_1, X_2, ...$ witout any preliminary assumptions about their distribution, and using the real-valued constants $a_1, a_2, ...$ and positive constants $b_1, b_2, ...$, we introduce

$$Z_n = \frac{S_n - a_n}{b_n}$$

We want to find the constants a_n and b_n in such a way that the distribution functions of Z_n weakly converge to some limit distribution function $G(x)$, that is,

$$P(Z_n < x) \Rightarrow G(x) \, , n \to \infty$$

for any x which is a continuity point of the function $G(x)$. We will see that the stable law is a class of distributions playing the role of the limiting law.

2.1.1.2 Power-law distributions and self-similarity

The class of skew distribution functions (or Yule distribution) appears in a wide range of empirical data, particularly data describing sociological, biological and economic phenomena (see Simon [1955]). In the statistical extreme value theory (EVT), the extremes and the tail regions of a sample of independent and identically distributed (i.i.d.) random variables converge in distribution to one of only three types of limiting laws (see Reiss et al. [1997], Embrechts et al. [1999])

1. Exponential decay : it decreases at a rate proportional to its current value. That is, $X_t = X_0 e^{-\lambda t}$ where λ is the exponential decay constant and $\tau = \frac{1}{\lambda}$ is the mean lifetime.
2. Power-law decay : one quantity varies as a power of another. Power laws are scale-invariant, that is, for $f(x) = ax^{-\alpha}$ we get $f(cx) = c^{-\alpha} f(x) \propto f(x)$. A power law $x^{-\alpha}$ has a well-defined mean over $x \in [1, \infty]$ only if $\alpha > 2$ and it has a finite variance only if $\alpha > 3$.
3. The behaviour of distributions with finite endpoint of their support.

More generally, a power-law probability distribution is a distribution whose density function has the form

$$p(x) \propto L(x) x^{-\alpha}$$

where $\alpha > 1$ and $L(x)$ is a slowly varying function at infinity [1]. Thus, $p(x)$ is asymptotically scale invariant.
Heavy-tailed distributions are probability distributions whose tails are not exponentially bounded. That is, they have heavier tails than the exponential distribution. The distribution of a random variable X with distribution function F is said to have a heavy right tail if

$$\lim_{x \to \infty} e^{\lambda x} P(X > x) = \infty \text{ for all } \lambda > 0$$

[1] $\lim_{x \to \infty} \frac{L(ux)}{L(x)} = 1$, for any $u > 0$.

A fat-tailed distribution is a probability distribution exhibiting large skewness or kurtosis relative to the normal distribution. Some fat-tailed distributions have power law decay in the tail of the distribution, but do not necessarily follow a power law everywhere. The distribution of a random variable X is said to have a fat tail if

$$P(X > x) \sim x^{-\alpha} \text{ as } x \to \infty \, , \alpha > 0$$

That is, if X has a probability density function $f_X(x)$, then

$$f_X(x) \sim x^{-(1+\alpha)} \text{ as } x \to \infty \, , \alpha > 0$$

Since such a power is always bounded below by the probability density function of an exponential distribution, fat-tailed distributions are always heavy-tailed. Hence, the universe of fat-tailed distributions can be indexed by their tail index α with $\alpha \in (0, \infty)$. A standard measure of fat tails is the kurtosis, which is the expected fourth power of a process given by $E[X_t^4]$. Events deviating from the mean by five or more standard deviations in the normal distribution have lower probability, thus meaning that in the normal distribution rare events can happen but are likely to be milder in comparison to fat-tailed distributions. On the other hand, fat-tailed distributions such as the Cauchy distribution (and all other stable distributions with the exception of the normal distribution) are examples of fat-tailed distributions having undefined sigma (the variance is not bounded).

A stochastic process $Y = (Y(t), t \geq 0)$ is called self-similar if there is H such that for all $c > 0$ we have

$$(Y(ct), t \geq 0) = (c^H Y(t), t \geq 0) \tag{2.1.1}$$

where H is the exponent of self-similarity, or scaling exponent. Note, a self-similar process can not be stationary, but its increments can if they are stationary. If $X_i = Y(i) - Y(i-1)$, $i = 1, 2, ...$ is the increment process of Y, then the partial sum process $S_n = X_1 + ... + X_n$ clearly satisfies for $n \geq 1$,

$$S_n = Y(n) - Y(0) = n^H (Y(1) - Y(0)) = n^H S_1$$

If the process Y also has stationary increments, then the process $X = (X_1, X_2, ...)$ is stationary and the above equation shows that the scaling exponent H determines the distributional order of magnitude of the partial sum process of the stationary model X. Self-similar objects with parameters N and s are described by a power law such as

$$N = s^D$$

where $D = \frac{\ln(N)}{\ln(s)}$ is the dimension of the scaling law, known as the Hausdorff dimension.

2.1.2 The stable distributions

2.1.2.1 Definitions and properties

Stable distributions (or Levy alpha-stable distribution) are determined by characteristic function with the properties of being stable under addition, self-similar, and capable of producing high peaks at the mean and fat tails (see Nolan [2014]). L-stable distributions are characterised by tail indices $\alpha < 2$ (2 characterising the case of the normal distribution). All other distributions with a tail index smaller than 2 converge under summation to the Levy stable law with the same index, while all distributions with an asymptotic tail behaviour with $\alpha > 2$ converge under aggregation to the Gaussian law. This is a generalisation of the central limit theorem to random variables without second (and possibly first) order moments. Further, L-stable distributions are fractal because of their self-similar properties and are represented by $S(\alpha, \beta, c, \delta; k)$ where $\alpha \in (0, 2]$ is the characteristic exponent (or stability parameter), $\beta \in [-1, 1]$ is the skewness parameter, $\delta \in \mathbb{R}$ is the location parameter and $c > 0$ is the scale parameter. Note, k is an integer defining the different parametrisation. Some special case are the Gaussian distribution ($\alpha = 2$), the Levy distribution ($\alpha = \frac{1}{2}$, $\beta = 1$) and the Cauchy distribution ($\alpha = 1$, $\beta = 0$).

The stability parameter α define the existence, or not, of the variance. That is, the normal random variable is unbounded, and can take arbitrarily large (absolute) values, but can not take infinite values in the sense that

$$\lim_{x \to \infty} \int_x^{\infty} p(x^{'}) dx^{'} = \lim_{x \to -\infty} \int_{-\infty}^{x} p(x^{'}) dx^{'} = 0$$

Thus, a random variable distributed by the normal law (or any other stable) law takes finite values with probability one. They are called proper. Any non-degenerate stable distribution has a smooth (infinitely differentiable) density function, but for $\alpha < 2$ the density has an asymptotic behaviour of a heavy-tail distribution, leading to infinite variance. The meaning of infinite variance can be expressed as follow

$$\int_x^{\infty} p(x^{'}) dx^{'} + \int_{-\infty}^{-x} p(x^{'}) dx^{'} \propto x^{-\alpha} \, , x \to \infty \, , 0 < \alpha < 2$$

Hence, the existence of finite variance of the normal law is connected with just a faster decrease of tails as compared with the others. So, all the members of the stable law compete on equal terms, and only the common habit of using the variance to characterise distributions makes the normal law attractive.
To understand the relationship between power laws and moments we provide a brief discussion of the properties of the tails of non-Gaussian ($\alpha < 2$) stable laws:

- For all $\alpha < 2$ and $-1 < \beta < 1$, both tail probabilities and densities are asymptotically powers laws.
- When $\beta = -1$, the right tail of the distribution is not asymptotically a powers law.
- When $\beta = 1$, the left tail of the distribution is not asymptotically a powers law.

The smaller α is, the heavier the tails of the density. One consequence of heavy tails is that not all moments exits. The first two moments are not generally useful for heavy tailed distributions since the integral expressions for these expectations may diverge. Instead, we can use fractional absolute moments $E[|X|^p] = \int_{-\infty}^{\infty} |x|^p f(x) dx$, where p is any real number. One can show that for $\alpha < 2$,

- $E[|X|^p]$ is finite for $0 < p < \alpha$, and
- $E[|X|^p] = +\infty$ for $p \geq \alpha$.

Thus,

- when $\alpha \leq 1$, then $E[|X|] = +\infty$ and the mean of X is undefined.
- when $1 < \alpha \leq 2$, then $E[|X|] < \infty$ and the mean of X is defined.
- when $0 < \alpha < 2$, then $E[|X|^2] = E[X^2] = +\infty$ and stable distributions do not have finite second moments (variances).

We can define stable distributions as follow:

Definition 2.1.1 *The random variable X with values in $\mathbb{R}$ has a stable distribution if the following condition holds: If $n \in \mathbb{N}_+$ and $X_1, ..., X_n$ is a sequence of independent variables, each with the same distribution as X, then $S_n = \sum_{i=1}^n X_i$ has the same distribution as $a_n + b_n X$ for some $a_n \in \mathbb{R}$ and $b_n \in (0, \infty)$. X is strictly stable if and anly if $a_n = 0$ for all n.*

The problem of summing stable random variables, can be reduced to finding a_n and b_n. For a strictly stable random variable, $a_n = 0$, it simplifies to

$$S_n = \sum_{i=1}^{n} X_i = b_n X$$

which is easlily solved for the normal distribution (only case with finite variance). In the general case of strictly stable random variables, we get

$$S_n = \sum_{i=1}^{n} X_i = n^{\frac{1}{\alpha}} X$$

One can show that the only possible choice for the scaling constants is $b_n = n^{\frac{1}{\alpha}}$ for some $\alpha \in (0,2]$. All stable distributions remain stable under linear transformations, and one can choose some standard values of the shift and scale parameters.
Two important properties of the stable distributions are

1. the stability property, and
2. the generalisised central limit theorem.

A basic property of stable laws is that sums of α-stable random variables are α-stable. Using characteristic functions, we can show

Proposition 5 *If X_1 and X_2 are independent, and $X_i \sim S(\alpha, \beta_i, c_i, \delta_i)$, then $X_1 + X_2 \sim S(\alpha, \beta, c, \delta)$ with*

$$c = (c_1^\alpha + c_1^\alpha)^{\frac{1}{\alpha}} \ , \ \beta = \frac{\beta_1 c_1^\alpha + \beta_2 c_2^\alpha}{c_1^\alpha + c_1^\alpha} \ , \ \delta = \delta_1 + \delta_2$$

Definition 2.1.2 *The stability property states that the random variables $X_1, ..., X_n$ are independent and symmetrically stable with the same characteristic exponent, α, if and only if for any constants $a_1, ..., a_n$, the linear combination $\sum_{i=1}^n a_i X_i$ is symmetric α-stable.*

***Remark* 2.1.1** *When the summands in Definition (2.1.2) are dependent, the sum is stable, but the precise statement is more difficult and depends on the exact dependence structure.*

***Remark* 2.1.2** *In the case where the summands in Definition (2.1.2) do not all have the same α, the sum will not be stable.*

One can show that the sum of two independent stable random variables with different αs is not stable.

The central limit theorem (CLT) states that the sum of a number of independent and identically distributed (i.i.d.) random variables with finite variances will tend to a normal distribution as the number of variables grows. To be more precise, let $X_1, X_2, ...$ be i.i.d. random variables with mean μ and variance σ^2. CLT states that the sample mean $\overline{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ will have

$$a_n \sum_{i=1}^{n} X_i - b_n \to Z \sim N(0,1) \text{ as } n \to \infty$$

where $a_n = \frac{1}{\sigma \sqrt{n}}$ and $b_n = \frac{\sqrt{n}\mu}{\sigma}$. The generalisied CLT shows that if the finite variance assumption is dropped, the only possible resulting limits are stable. That is, the sum of a number of random variables with a power-law tail distribution decreasing as $|x|^{-(\alpha+1)}$ where $0 < \alpha < 2$ (having infinite variance) will tend to a stable distribution as the number of summands grows. If $\alpha > 2$, then the sum converges to a stable distribution with $\alpha = 2$, that is, a Gaussian distribution (see Gnedenko [1943] and Gnedenko et al. [1954]). It is mathematically formalised as follow:

Theorem 2.1.1 *Generalised Central Limit Theorem*
A nondegenerate random variable Z is α-stable for some $0 < \alpha \leq 2$ if and only if there is an i.i.d. sequence of random variables $X_1, X_2, ...$ constants $a_n > 0$, $b_n \in \mathbb{R}$ with

$$a_n \sum_{i=1}^{n} X_i - b_n \to Z \tag{2.1.2}$$

We can give an alternative definition of the generalised CLT as follow

Theorem 2.1.2 *Generalised Central Limit Theorem*
Let $X_1, .., X_n$ be i.i.d. random variables with the distribution function $F_X(x)$ satisfying the conditions

$$\begin{cases} 1 - F_X(x) \sim cs^{-\mu} \,,\, x \to \infty \\ F_X(x) \sim d|x|^{-\mu} \,,\, x \to \infty \end{cases}$$

with $\mu > 0$. Then there exist sequences a_n and $b_n > 0$ such that the distribution of the centered and normalised sum

$$Z_n = \frac{1}{b_n} (\sum_{i=1}^{n} X_i - a_n)$$

weakly converges to the stable distribution with parameters

$$\alpha = \begin{cases} \mu \,,\, \mu \leq 2 \\ 2 \,,\, \mu > 2 \end{cases}$$

and

$$\beta = \frac{c - d}{c + d}$$

as $n \to \infty$:

$$F_{Z_n}(x) \Rightarrow G^A(x : \alpha, \beta)$$

When discussing convergence of normalised sums, we can use the following definition:

Definition 2.1.3 *A random variable X is in the domain of attraction of Z if there exists constants $a_n > 0$, $b_n \in \mathbb{R}$ with*

$$a_n \sum_{i=1}^{n} X_i - b_n \to Z$$

where $X_1, X_2, ...$ are i.i.d. copies of X. $DA(Z)$ is the set of all random variables that are in the domain of attraction of Z.

Thus, the only possible nondegenerate distributions with a domain of attraction are stable. When $a_n = n^{-\frac{1}{\alpha}}$, X is said to belong to the normal domain of attraction of Z. Generally, $a_n = n^{-\frac{1}{\alpha}} h^{-1}(n)$ where $h(x)$ is a slowly varying function at infinity [2].

[2] $\lim_{x \to \infty} \frac{h(ux)}{h(x)} = 1$, for any $u > 0$.

2.1.2.2 Problems when estimating parameters

The lack of closed form formulas for most stable densities and distribution functions is very problematic when estimating parameters, or, simply trying to simulate sequences of stable variables. In the former, we need to rely on numerical approximations such as the fast Fourier transform (FFT) on the characteristic function, or apply direct numerical integration. The FFT can only compute the pdf, and the density must be numerically integrated to compute the cumulative distribution function (cdf). Further, it suffers from serious drawback when dealing with small values of α (see Mittnik et al. [1999]). When using direct integration, both the density and the distribution function are numerically integrated (see Nolan [2001]). In the latter, there is no analytical expressions for the inverse F^{-1} of the cdf. Borak et al. [2005] discussed various methods for estimating parameters and argued that a visual inspection test could help estimate the tail index. This is achieved by plotting the right tail of the empirical cdf on a double logarithmic graph. The slope of the linear regression for large values of x yields the estimate of the tail index, via $\alpha = -slope$. However, this method is very sensitive to the size of the sample and the choice of the number of observations used in the regression. Thus, it is subjective and can yield large estimation errors. Alternatively, ignoring a parametric form for the entire distribution function and focusing only on the tail behaviour, we can consider the Hill estimator to estimate the tail index (see Hill [1975]), but it tends to be overestimated when α is close to two and the sample size is not very large. Quantile estimation, characteristic function approaches (see Koutrouvelis [1980]), and maximum likelihood estimator (MLE) method are more accurate but much slower. Note, we use the MLE to estimate the distributional parameters only if we assume that the data are distributed according to some known heavy tail distribution.

2.1.2.3 Sums of dependent infinite variance random variables

Definition (2.1.1) established the limit relation for the sum, $S_n = \sum_{i=1}^{n} X_i$, of infinite variance stationary sequences. However, the convergence exists if and only if the random variable $X = X_1$ has a distribution with regularly varying tails with index $-\alpha \in (-2, 0)$. That is, there exist constants $p, q \geq 0$ with $p + q = 1$ and a slowly varying function h such that

$$P(X > x) \sim p\frac{h(x)}{x^\alpha} \text{ and } P(X \leq -x) \sim q\frac{h(x)}{x^\alpha} \text{ , } x \to \infty$$

This relation is called tail balance condition. Even though the limit in Equation (2.1.2) is a benchmark for weakly dependent stationary sequences with regularly varying marginal distribution, in the presence of dependence, conditions for the convergence of the partial sums towards a stable limit require special structure.
Going further, Davis et al. [1995] assumed the stronger condition that the strictly stationary sequence X_t is regularly varying with index $\alpha \in (0, 2)$. That is, the finite-dimensional distributions of X_t have a jointly regularly varying distribution in the following sense: for every $d \geq 1$, there exists a non-null Radon measure μ_d on the Borel σ-field of $\overline{\mathbb{R}}^d \setminus \{0\}$ (μ_d is finite on sets bounded away from zero), $\overline{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$, such that

$$nP(\frac{1}{a_n}(X_1, ..., X_d) \in \cdot) \xrightarrow{v} \mu_d(\cdot) \quad (2.1.3)$$

where $\xrightarrow{v}$ denotes vague convergence and a_n satisfies

$$nP(|X| > a_n) \sim 1 \quad (2.1.4)$$

The limiting measure has the property $\mu_d(xA) = x^{-\alpha}\mu_d(A)$, $t > 0$, for Borel sets A. The parameter α is the index of regular variation of X_t and its finite dimensional distributions.
Rather than directly assuming regular variation of X, Jakubowski [1993] [1997] required the conditions $T_+(d)$ and $T_-(d)$, $d \geq 1$, that is, the existenceof the limits

$$\lim_{n\to\infty} nP(S_d > a_n) = b_+(d) \text{ and } \lim_{n\to\infty} nP(S_d \leq -a_n) = b_-(d) \text{ , } d \geq 1 \quad (2.1.5)$$

Hence, if $b_+(d) + b_-(d) > 0$, the regular variation of a_n with index $\frac{1}{\alpha}$ is equivalent to regular variation of S_d with index α.

Regular variation (RV) of a stationary sequence X_t is well accepted in applied probability theory. Davis et al. [1995] proposed several conditions that we are going to briefly discuss.

Condition 1 *Regular variation*
The strictly stationary sequence X_t is regularly varying with index $\alpha \in (0,2)$ in the sense of Equation (2.1.3) with non-null Radon measures μ_d, $d \geq 1$, and a_n chosen in Equation (2.1.4).

In addition to Condition (1), Davis et al. [1995] required the mixing condition (MC) $\mathcal{A}(a_n)$ defined as follow:

Condition 2 *Mixing condition*
Given the point process $N_n = \sum_{t=1}^n \epsilon_{\frac{X_t}{a_n}}$ and assuming that there exists a sequence $m = m_n \to \infty$ such that $k_n = [\frac{n}{m_n}] \to \infty$ [3]*, the condition $\mathcal{A}(a_n)$ requires that*

$$E[e^{-\int f dN_n}] - \big(E[e^{-\int f dN_m}]\big)^{k_n} \to 0$$

where f belongs to a sufficiently rich class of non-negative measurable functions on $\mathbb{R}$ such that the convergence of the Laplace functional $E[e^{-\int f dN_n}]$ for all f from this class ensures weak convergence of N_n.

MC ensures that N_n can be approximated in law by a sum of k_n i.i.d. copies of N_m, hence the weak limits of N_n must be infinitely divisible point process. Further, Davis et al. required the anti-clustering condition (AC)

Condition 3 *Anti-clustering condition*

$$\lim_{d \to \infty} \lim_{n \to \infty} \sup P\big(\max_{d \leq |i| \leq m_n} |X_i| > x a_n \big| |X_0| > x a_n \big) = 0 \, , \, x > 0$$

where $m = m_n \to \infty$ is the block size used in the definition of the mixing condition $\mathcal{A}(a_n)$.

Davis et al. proved convergence of the normalised partial sums by showing that the limiting distribution is infinitely divisible with a Levy triplet corresponding to an α-stable distribution. They needed conditions to ensure that the sum of the small values in the sum $a_n^{-1} S_n$ does not contribute to the limit. The vanishing small values conditions (VSM) is as follow

Condition 4 *Vanishing small values conditions*

$$\lim_{\epsilon \to \infty} \lim_{n \to \infty} \sup P\Big(\big| \sum_{t=1}^n X_t I_{\{|X_t| \leq \epsilon a_n\}} - nE[X I_{\{|X| \leq \epsilon a_n\}}] \big| > x a_n \Big) = 0 \, , \, x > 0$$

for $\alpha \in (0,2)$.

Assuming that $(X_t)_{t>0}$ is a strictly stationary process satisfying the regular variation condition, mixing condition, the anti-clustering condition above as well as

- the TB condition: the limits

$$\lim_{d \to \infty} \big(b_+(d) - b_+(d-1) \big) = c_+ \text{ and } \lim_{d \to \infty} \big(b_-(d) - b_-(d-1) \big) = c_-$$

  exists, where $b_+(d)$ and $b_-(d)$ are the tail balance parameters given in Equation (2.1.5).

[3] $[x]$ denotes the integer part of x.

- the CT condition: for $\alpha > 1$ assume $E[X_1] = 0$ and for $\alpha = 1$

$$\lim_{d \to \infty} \lim_{n \to \infty} \sup n \big| E[\sin(a_n^{-1} S_d)] \big| = 0$$

Bartkiewicz et al. [2011] showed that c_+ and c_- are non-negative and $\frac{1}{a^n} S_n$ converges in distribution to an α-stable random variable and provided the characteristic function. They applied their results to some standard time series models, such as

- the stochatic volatility model

$$X_t = \sigma_t Z_t$$

  where the volatility sequence σ_t is strictly stationary independent of the i.i.d. noise sequence Z_t.

- the stochastic recurrence equation

$$X_t = A_t X_{t-1} + B_t \,,\, t \in \mathbb{Z}$$

  where (A_t, B_t) constitutes an i.i.d. sequence of non-negative random variables A_t and B_t. Examples of such process are the $ARCH(1)$ process and the volatility sequence of a $GARCH(1,1)$ process.

- the stochatic volatility model

$$X_t = \sigma_t Z_t$$

  where Z_t is an i.i.d. sequence with $E[Z] = 0$ and $var(Z) = 1$ and

$$\sigma_t^2 = a_0 + (a_1 Z_{t-1}^2 + b_1) \sigma_{t-1}^2$$

  where $a_0 > 0$ and the non-negative parameters a_1 and b_1 are chosen such that a strictly stationary solution to the above equation exists. Then, the process is strictly stationary. When $a_1 b_1 > 0$ we recover the $GARCH(1,1)$ and if $b_1 = 0$ and $a_1 > 0$ we get the $ARCH(1)$.

2.1.3 Correlations and dependencies

2.1.3.1 Understanding the concepts

Risk analyisis on sequences of random variables require two types of input:

1. the marginal distributions for the different variables, and
2. the dependencies between these variables.

The mathematical notion of independence of random variables is a fundamental part of probability theory since it models situations where the events do not cause one another. However, even though the concepts of linearly independent, orthogonal, and uncorrelated variables indicate lack of relationship between variables, the mathematical distinctions between them are confusing. Rodgers et al. [1984] discussed these concepts on fixed variabes, and proposed the following algebriac definition:

Definition 2.1.4 *Let $\underline{X}$ and $\underline{Y}$ be vector observations of the variables X and Y. Then*

1. *$\underline{X}$ and $\underline{Y}$ are linearly independent iff there exists no constant a such that $a\underline{X} - \underline{Y} = 0$.*

2. *$\underline{X}$ and $\underline{Y}$ are orthogonal iff $(\underline{X})^\top \underline{Y} = 0$.*

3. *$\underline{X}$ and $\underline{Y}$ are uncorrelated iff $(\underline{X} - \overline{X}I)^\top (\underline{Y} - \overline{Y}I) = 0$, where $\overline{X}$ and $\overline{Y}$ are the means of $\underline{X}$ and $\underline{Y}$, respectively, and I is a vector of ones.*

Note, linear independence and orthogonality are properties of the raw variables, while zero correlation is a property of the centered variables.
In a geometric framework, uncorrelated implies that once each variable is centered, then the vectors are perpendicular. That is, centering variables will change the angle between the two vectors. Thus, orthogonal denotes that raw variables are perpendicular, while uncorrelated denotes that the centered variables are perpendicular. For linearly independent variables we get the following situations

1. two variables that are perpendicular can become oblique once centered: orthogonal but not uncorrelated.

2. two variables not perpendicular can become perpendicular once centered: uncorrelated but not orthogonal.

3. two variables may be both orthogonal and uncorrelated.

In probability theory, the property of mutual independence of a set of random variables is some special property of its joint distributions. It is important to note that correlation is only a measure of linear dependence, but it does not necessarily imply something about other kinds of dependence. In the special case where X or Y has an expected value of 0, then the covariance is the expectation of the product, and X and Y are uncorrelated if and only if $E[XY] = 0$. For example, we let X be normally distributed with $E[X] = 0$ and we let $Y = X^2$ be a quadratic function of X. Thus, X and Y are dependent (quadratic dependence), but $Cov(X, Y) = E[XY] = E[X^3] = 0$. Therefore, X and Y are uncorrelated, but not independent (situation (2)). Note, X does not have to be normally distributed, and any density function symmetric about 0 and for which $\int |x|^3 dP$ exists, will do.

Many variables in complex natural and engineered systems are actually correlated, or, have some nonlinear interdependence. Even though assuming independent variables is mathematically convenient in presence of a lack of evidence about dependence, it is not correct. Ferson et al. [2005] reviewed the dangers of assuming all variables in an assessment are independent, and he showed how different dependencies can lead to quantitatively different results. Some important facts leading to confusion are

- independence between X and Y and between Y and Z implies nothing at all about the dependence between X and Z.
- independence implies that the correlation will be zero, but not vice versa.
- uncorrelatedness does not imply there is no relationship between the variables.
- a weak correlation does not guarantee a weak relationship.
- even negligible correlations can greatly influence the results of risk calculations. The estimated dispersion, and especially the tail probabilities, can be highly inaccurate.

2.1.3.2 Measuring dependence

Since we argued that dependencies between random variables should be modelled when it is present in time series, we need to measure that dependence. While we can use the correlation coefficient in the case of finite variance, it is not defined in the case of infinite variance. Thus, obtaining a sensible measure of dependence has been the object of considerable attention with several measures proposed (see Samorodnitsky et al. [1994]).

Finite variance In the case of finite variance, there are many different measures of correlation among which are the Pearson's product-moment correlation and Spearman's rank correlation coefficient. See Hutchinson et al. [1990] for a review. One can use the Pearson's product coefficient for computing the correlation between normally distributed variables since it only measures the degree of linear dependence. Otherwise, a more general measure of dependency between variables, such as the Spearman's rank correlation coefficient, can be used. In fact, it takes an entire dependence function, or copula, to fully specify the dependence between two random variables (see Clemen et al. [1999], Li [2000]). We introduce copulas in Section (2.1.3.3). There are many ways for dependence between variables, such as nonlinear relationships, clusters, subgroupings, impossible regions, and other complexities. As a result, copula function can not be completely characterised by a single dimension such as correlation coefficient. Nonetheless, there exists a few special cases where correlation completely determine the dependence, such as the Bernoulli marginals or when the correlation is extreme, ± 1. The above result naturally leads to the following fact

- varying correlations is insufficient to explore the possible nonlinear dependencies between variables.

Thus, a sensitivity analysis based on varying correlation gives an incomplete picture of uncertainty that is far too tight, even if correlation is varied in the range $[-1, 1]$.

Infinite variance In the case of infinite variance, these measures can be classified as

- using moments with orders less than two, like the coefficient of covariation, or
- not using any moment, like the rank-based correlation coefficients.

Note, the former is not bounded and not symmetric. To remedy this problem, Garel et al. [2005] proposed the symmetric coefficient of covariation which gives good results. Further, when estimating the multivariate extreme dependence of two series we can use the bivariate EVT method, but we still need to check if the two series are asymptotically independent or not (see Bekiros et al. [2008]). In a limiting sense, two stituations are possible, when quantifying dependence of multivariate extreme values:

- one where the extremes are dependent,
- the other where the extremes are independent.

However, asymptotic independence is not equivalent to pure statistical independence, as it allows a certain level of dependence between two series of finite samples. Non-parametric measures exist to describe the levels of asymptotic dependence or independence, but the data must be sampled independently (see Coles et al. [1999], Poon et al. [2003] [2004]). Thus, one can use a vector autoregressive model to filter for any serial correlation, and multivariate Garch models can be used to filter for any heteroskedasticity.

2.1.3.3 Introducing the copula

The copula between X and Y is the joint distribution function between the uniformly distributed variables $F_X(x)$ and $F_Y(y)$.

Definition 2.1.5 *A copula is a function* $C : [0,1] \times [0,1] \to [0,1]$ *such that*

- $C(a, 0) = C(0, a) = 0$*, and*
- $C(a, 1) = C(1, a) = a$ *for all* $a \in [0, 1]$*, and*
- $0 \leq C(a_2, b_2) - C(a_1, b_2) - C(a_2, b_1) + C(a_1, b_1)$ *for all* a, a_1, a_2, b_1, b_2 *in* $[0, 1]$ *when* $a_1 \leq a_2$ *and* $b_1 \leq b_2$*.*

While copulas are relatively simple structures compared to joint distributions, Sklar's theorem (see Sklar [1959]) explains how to compute the joint distribution function $H(x,y)$ from specified marginal distributions (CDFs) and a dependence function represented by a copula.

Theorem 2.1.3 *Sklar's theorem*
For any univariate distribution functions F_X and F_Y and any copula C, the function

$$H(x,y) = C(F_X(x), F_Y(y)) \tag{2.1.6}$$

is a two-dimensional distribution function having marginals F_X and F_Y.

The copula $C(a,b)$ is a joint density function mapping points on the unit square $(a, b \in [0,1] \times [0,1])$ to values between 0 and 1. Since it relates the quantiles (a and b) of the two distributions rather than the original variables (x and y), it is not affected by scaling. Note, Sklar's theorem was generalised to higher dimensions. One of the somplest copula is the product copula $C(a,b) = a * b$ such that $H(x,y) = F_X(x)F_Y(y)$, and the random variables are independent. Another example is the Gaussian copula. If $N(x,y)$ is the joint standard cumulative normal density function , and $N^{-1}(a)$ and $N^{-1}(b)$ are the standard normal quantile functions, then the Gaussian copula is

$$N(N^{-1}(a), N^{-1}(b))$$

which is computed with the method of inversion. Numerous parametric families of copula exist (see Nelsen [2010]), such as

- Archimedean (Clayton, Gumbel, Frank)
- Morgenstern
- Galambos
- Plackett

Taking the second cross-partial derivative of Equation (2.1.6) with respect to x and y, we get

$$h(x,y) = f_X(x)f_Y(y)c(F_X(x), F_Y(y)) \tag{2.1.7}$$

where $f_X(x)$ is the probability density function (PDF) of X, $h(x,y)$ is a two-dimensional density function of X and Y, and

$$\partial_{ab}C(a,b) = c(a,b)$$

is the copula density. Thus, the joint density function is the product of the marginal densities and the derivative of the copula evaluated at the marginal CDFs. Since the copula is a joint CDF, the copula density is a joint PDF and it is a scale-free measure of dependence. Taking the cross-partial derivatives of the copula, the product copula density is 1, and the normal density is

$$(N^{-1}(a))(N^{-1}(b))n(N^{-1}(a), N^{-1}(b))$$

Defining the dependence function as the copula density evaluated at the marginal CDFs ($c(F_X(x), F_Y(y))$), we can rewrite Equation (2.1.7) as

$$c(F_X(x), F_Y(y)) = \frac{h(x,y)}{f_X(x)f_Y(y)} \tag{2.1.8}$$

Thus, the dependence function is the ratio of marginal and joint PDFs, but it is not scale-free. The dependence function under the product copula is 1, and under the normal copula density it is

$$(N^{-1}(F_X(x)))(N^{-1}(F_Y(y)))n(N^{-1}(F_X(x)),N^{-1}(F_Y(y)))$$

A joint normal density has a Gaussian copula density with $F_X(x) = N(x)$ and $F_Y(y) = N(y)$, such that the dependence function simplifies to

$$c(N(x),N(y)) = \frac{n(x,y)}{n(x)n(y)}$$

which is a particular example of Equation (2.1.8).

2.1.4 Long range dependence

Long range dependence (LRD) (or long memory) is related to memory in a stochastic process, and it addresses a great variety of issues, and as such, has been defined in various different ways. We are going to discuss LRD and present various ways of defining it.

2.1.4.1 A historical review

Working on the Nile River Dam Project, Hurst [1951] observed non-randomness in the data, as larger than average overflows were more likely to be followed by more large overflows. He devised a statistical test with a power law asymptotic behaviour quantifying the relative tendency of a time series either to regress strongly to the mean or to cluster in a direction. The characteristic exponent, called the Hurst exponent, H, relates to the autocorrelations of the time series, and the rate at which these decrease as the lag between pairs of values increase. As a result, it is used as a measure of long-term memory of time series. In a series of articles, Mandelbrot [1965] and Mandelbrot et al. [1968] [1968b] analysed the importance of long memory in stochastic processes. Afterwards, Mandelbrot [1975] introduced fractals as geometric shapes being self-similar and having fractional dimensions. Since then, LRD has become closely associated with self-similar processes, which are stochastic models with the property that a scaling in time is equivalent to an appropriate scaling in space.

Hurst measured how a reservoir level (Nile River Dam project) fluctuated around its average level over time, and found that the range of the fluctuation would change, depending on the length of time used for measurement. If the series were random, the range would increase with the square root of time

$$R = T^{\frac{1}{2}}$$

but it was not the case. To standardise the measure over time, he created a dimensionless ratio by dividing the range by the standard deviation, S, of the observation, obtaining the Rescaled Range

$$\frac{R}{S} = k \times T^H$$

where k is a constant depending on the time series. Thus, by rescaling the data, the so called rescaled range analysis (R/S analysis) can compare diverse phenomena and time periods. He found that most natural phenomena follow a biased random walk, that is, a trend with noise which could be measured by how the rescaled range scales with time, or, how high the exponent H is above $\frac{1}{2}$. This phenomenon became known as the Hurst phenomenon (in the sense of $H > \frac{1}{2}$). The value of the Hurst exponent varies between 0 and 1, with

- $H = \frac{1}{2}$ implying a random walk, or an independent process.
- For $0 \leq H < \frac{1}{2}$ we have anti-persistence (or ergodicity) where the process covers less distance than a random walk (mean reverting).

- For $\frac{1}{2} < H \leq 1$ we have persistence (or trend-reinforcing) where the process covers more distance than a random walk (long memory effects).

The fractal dimension D of a time series measures how jagged the time series is. In order to calculate the fractal dimension of a time series we count the number of circles of a given, fixed diameter needed to cover the entire time series. We then increase the diameter and count again the number of circles required to cover the time series. Repeating that process, the number of circles has an exponential relationship with the radius of the circle satisfying

$$N \times d^D = 1$$

where N is the number of circles, and d is the diameter. Transforming the above equation [4], the fractal dimension is given by

$$D = \frac{\log(N)}{\log(\frac{1}{d})}$$

Thus, we get $s = \frac{1}{d}$, and we see that the fractal dimension of a time series is a function of scaling in time. As a result, the connection between the two types of scaling (time and space) is determined by the Hurst exponent. Mandelbrot showed that the Hurst Exponent, H, is related to the fractal dimension D for a self-similar surface in n-dimensional space by the relation

$$D = n + 1 - H$$

where $0 \leq H \leq 1$. Further, noises can be characterised by their power spectra, which follow simple inverse power laws, $P(f) \sim f^{-\beta_s}$, where $\beta_s \in [0, 4]$ is the spectral exponent (see Fourier analysis). For white noise $\beta_s = 0$, meaning that it is not related to frequency. When white noise is integrated. we get $\beta_s = 2$, the power spectrum for brown noise. If $0 < \beta_s < 2$, we have pink noise. Beyond brown noise, there is black noise where $\beta_s > 2$. There is a relationship between fractional noises and the Hurst exponent given by

$$\beta_s = \begin{cases} 2H + 1 \text{ if } H \neq \frac{1}{2} \\ 2H - 1 \text{ if } H = \frac{1}{2} \end{cases}$$

such that

- black noise is related to long memory effects with $H > \frac{1}{2}$, $2 < \beta_s \leq 4$.
- pink noise is related to antipersistence with $H < \frac{1}{2}$, $1 \leq \beta_s < 2$.

2.1.4.2 Modelling the Hurst phenomenon

Ever since, academics became concerned with finding a stochastic model that would explain the Hurst phenomenon (in the sense of $H > \frac{1}{2}$). Moran [1964] proposed to drop the assumption of finite variance of the data by assuming that the observations were in the domain of attraction of an infinite variance α-stable distribution with $\alpha < 2$. However, Mandelbrot et al. [1979] pointed out that the self-normalising feature of the R/S statistic prevented infinite variance alone from explaining the Hurst phenomenon. A proof is given by. Alternatively, Mandelbrot [1965] and Mandelbrot et al. [1968b] explained the Hurst phenomenon by assuming stationary process with finite variance, but with correlations decaying so slowly as to invalidate the Functional Central Limit Theorem. The simplest model of that sort is the fractional Brownian Gaussian noise (FBGN). Thus, the Hurst phenomenon can be explained in two different ways, by introducing in a model

[4]
$$N = (\frac{1}{d})^D$$

- unusually heavy tails via infinite variance. The Noah effect: extreme incidents of precipitation.
- unusually long memory. The Joseph effect: long stretches of time higher than average and lower than average precipitation.

By mixing these two effects (see Anh et al. [2002]), we get

- fractional stable processes: self-similar processes with LRD and heavy tails.

However, the Joseph effect (long memory) can be taken to indicate non-stationarity. Hence, the difficulty to distinguish between stationary long memory models and other non-stationary models. Bhattacharya et al. [1983] showed that the Hurst phenomenon could as well be explained by a simple non-stationary model. The model proposed is as follow

$$X_i = Y_i + (a+i)^{-\beta} \, , i = 1, 2, ...$$

where Y_i are i.i.d. random variables with a finite variance σ^2, $a \geq 0$ and $0 < \beta < \frac{1}{2}$. It has the same R/S statistic than the fBm with $H = 1 - \beta$. It can be generalised to the class of Regime Switching models, which is a very important class of non-stationary models resembling empirically long memory stationary models. They have break points with location changing with the sample size, in either a random or non-random manner. Such a change can affect the mean and the variance of the process after break points, so that to many sample statistics it will look like long memory. In fact, the sationary long memory processes form a layer among the stationary process that is near the boundary with non-stationary processes. Therefore, since most stationary processes do not have long memory, non-stationary processes can provide an alternative explanation to the notion of LRD.

2.1.4.3 How the Hurst exponent relates to other theories

The value of the Hurst exponent determines whether or not the increments of a self-similar process with stationary increments possess LRD. Peters [1994] showed that there is a relation between the characteristic exponent of α-stable distributions and the Hurst exponent given by

$$\alpha = \frac{1}{H}$$

which implyies that the fractal measure of a process is related to the statistical self-similarity of that process. That is, the fractal dimension of the probability space is related to the fractal dimension of the time series. Thus,

1. the Hurst exponent of the time series is related to the characteristic exponent.
2. the characteristic exponent ($\alpha = \frac{1}{H}$) is the fractal dimension of the probability space.
3. $D = 2 - H$ is the fractal dimension of the time series.

And we get the following representation:

- For $0 < \alpha \leq 1$ the distribution has infinite mean and infinite variance.
- For $H = 1$, we get $\alpha = 1$.
- Values of $\frac{1}{2} < H < 1$ implies $1 < \alpha < 2$, indicating undefined or infinite variance leading to the non-existence of a martingale process. We get the Pareto-Levy distributions.
- Values of $H = \frac{1}{2}$ implies $\alpha = 2$ and we get the Gaussian distribution.
- Values $0 < H < \frac{1}{2}$ implies $2 < \alpha < \infty$ and the distribution is not stable, implying that the process is not independent with identically distributed innovations. However, the distributions still have finite variance.

2.1.4.4 Correlation and frequency

The most popular definitions of LRD in the literature are based on the second-order properties of a stochastic process, since correlations are easy to understand and to estimate. Restricting this approach to second-order stationary processes, a common way to try detecting LRD is by looking for a slow decay of covariances and correlations. A time series possesses long-range dependence (LRD) if it has correlation persisting over all time scales. Formally we get the following definition in the time domain

Definition 2.1.6 *Time domain*
The weakly stationary time series $X(t)$ is said to be long-range dependent if $\sum_{k=-\infty}^{\infty} \rho_k$ diverges, where ρ_k is the autocorrelation function (ACF) of $X(t)$.

and in the frequency domain

Definition 2.1.7 *Frequency domain*
The weakly stationary time series $X(t)$ is said to be long-range dependent if its spectral density obeys $f(\lambda) \sim C_f |\lambda|^{-\beta}$ as $\lambda \to 0$, for some $C_f > 0$ and some real $\beta \in [0,1]$.

Note, the parameter β is related to the parameter α via $\alpha = 1 - \beta$, and the Hurst exponent is given by $H = \frac{1+\beta}{2}$. Hence,

1. in the time domain, LRD occurs at a high degree of correlation between distantly separated data points,
2. in the frequency domain, LRD occurs at a significant level of power at frequencies near zero,

making the estimation of the parameters α and β very difficult.

2.1.4.5 Monofractal scaling analysis

Since self-similarity can have very different origins, it is difficult to test for a particular parametric model, such as α-stable Levy processes or fractional Brownian motions. Rather than devising tests for parametric models, we need a general test for self-similarity. As a consequence of self-similarity we can set $c = \frac{1}{t}$ in Equation (2.1.1) and get

$$X_t = t^H X_1 \, , \forall t > 0$$

so that the distribution of X_t is completely determined by the distribution of X_1

$$F_t(x) = P(t^H X_1 \leq x) = F_1(\frac{x}{t^H})$$

Differentiating the above equation, we get the density ρ_t of F_t as

$$\rho_t(x) = \frac{1}{t^H} \rho_1(\frac{x}{t^H})$$

and the kth moment is

$$E[|X_t|^k] = t^{kH} E[|X_1|^k]$$

Assuming that the log-price process $X_t = \ln S_t$ has stationary increments, and since $X_{t+\tau}$ has the same law as X_τ, we can estimate the density and moments of X_τ from a sample of increments. We can compare the aggregation properties of empirical densities with $\rho_t(x)$ above. Using asset prices sampled at time inteval Δt, we compute returns at time horizons $n\Delta t$ for $n = 1, ..., M$ and estimate their marginal density via a histogram. The scaling property of the density implies that $\widehat{\rho}_{n\Delta t}(x)$ and $\frac{1}{n^H} \widehat{\rho}_{\Delta t}(\frac{x}{n^H})$ should coincide.

2.1.5 Introducing multifractality

2.1.5.1 A description

We saw in Section (2.1.4) that the long-range trial dependency of response time is numerically defined by a scaling exponent obtained by monofractal analyses. It is achieved by assuming that the response time series are Gaussian distributed, such that their variations are described by the second-order statistical moment alone. However, many time series do not exhibit a simple monofractal scaling behaviour described with a single scaling exponent. In complex system, such different scaling behaviour can be observed for many interwoven fractal subsets of the time series, in which case a multitude of scaling exponents is required for a full description of the scaling behaviour, and a multifractal analysis must be applied (see Kantelhardt et al. [2002]). In the 80s a new class of stochastic fractals developed termed multifractals, with processes originating in high-dimensional systems. In general, two different types of multifractality in time series can be distinguished

1. Multifractality due to a broad probability density function (pdf) for the values of the time series. One can not remove the multifractality by shuffling the series.
2. Multifractality due to different long-range (time-) correlations of the small and large fluctuations. In this case, the pdf of the values can be a regular distribution with finite moments. The corresponding shuffled series will exhibit nonmultifractal scaling, since all long-range correlations are destroyed by the shuffling procedure.

While multifractality, or anomalous scaling, allows for a richer variation of the behaviour of a process across different scales, there are various ways of defining multifractality. For instance,

- scaling analysis focus on global properties, such as moments and autocovariance.
- multifractal analysis adopt a more local viewpoint and examine the regularity of realised paths around a given instant.

2.1.5.2 Scaling analysis

Scaling analysis investigate the multiscaling properties of the self-affine function $f(X)$ by calculating the qth order height-height correlation function. It exhibits a nontrivial multiscaling behaviour if

$$C_q(\Delta t) \sim (\Delta t)^{qH(q)}$$

with $H(q)$ changing continuously with q at least for some region of the q values (see Barabasi et al. [1991]). As a result, multiscaling in empirical data is typically identified by differences in the scaling behaviour of different (absolute) moments

$$E[|X(t,\Delta t)|^q] = c(q)(\Delta t)^{qH(q)} = c(q)(\Delta t)^{\tau(q)}$$

where $c(q)$ and $\tau(q)$ are deterministic functions of the order of the moment q. While the unique coefficient H quantifies a global scaling property of the underlying process, the multiplicity of such coefficients in multifractal processes, called Holder exponents, can be viewed as local scaling rates governing various patches of a time series, leading to a characteristically heterogeneous (or intermittent) appearance of such series.

2.1.5.3 Multifractal analysis

Contrary to scaling analysis which concentrate on the scaling exponent $\tau(q)$, the multifractal analysis (MFA) studies how the (pointwise) local regularity of X fluctuates in time (or space). Local Holder regularity describes the regularity of sample paths of stochastic processes by means of a local comparison against a power law function and is therefore closely related to scaling in the limit of small scales. The exponent of this power law, $h(t)$, is called the (local)

Holder exponent and depends on both time and the sample path of X. Researchers focused on a global description of the regularity of the function of f in form of multifractal spectrum (also called the singularity spectrum) $D(h)$ reflecting the size of the set of points for which the Holder exponent takes a certain value h. The main idea being that the relative frequency of the local exponents can be represented by a renormalised density called the multifractal spectrum. Thus, multifractal analyses estimate a multifractal spectrum of scaling exponents. Under some assumptions on the homogeneity and isotropy of the statistics of local singularities, it is possible to derive a relation between self-similarity exponents $\tau(q)$ and the singularity spectrum $D(h)$. Various tools exist to define the scaling exponents of higher order statistical moments and the temporal modulation of a local scaling exponent.

2.1.5.4 Identifying multifractality

Multifractality is a property verified in the infinitesimal limit only, while empirical data have an inherent discrete nature. As a result, any tools designed to validate the multifractal character of a given signal faces several difficulties linked to the finite size and the discretisation of the data. As long as the statistical tools identifying multifractal behaviour were assuming stationary series, they have been the subject to dispute as apparent scaling could be found in the absence of true scaling. Several methods for the multifractal characterisation of non-stationary financial time series developed, among which are the multifractal fluctuation analysis (MFA) where the process X is detrended with a polynomial, or an averaging function. For instance, Peng et al. [1994] introduced the detrended fluctuation analysis (DFA), which filters the series not only from a constant trend, but also from higher order polynomials. Later, Kantelhardt et al. [2002]) proposed the multifractal detrended fluctuation analysis (MF-DFA) as a generalisation of the detrended fluctuation analysis (DFA). By comparing the MFDFA results for original series with those for shuffled series, one can distinguish multifractality due to long-range correlations from multifractality due to a broad probability density function. In parallel, wavelet multifractal methods developed for analysing multifractality from experimental data. These methods include the moment (M) method, the wavelet transform modulus maxima (WTMM) method, the gradient modulus wavelet projection (GMWP) method, and the gradient histogram (GH) method (see Turiel et al. [2006] for a review).

2.1.5.5 The need to localise outliers

There are cases when local information about scaling provides more relevant information than the global spectrum. It generally happens for time series where the scaling properties are non-stationary, due to intrinsic changes in the signal scaling characteristics, or even boundary effects. Using the wavelet transform multiscale decomposition (WTM), Struzik [1999] proposed stable procedures for both the local exponent and its global spectrum. We know now that financial time series have non-stationary scaling properties (stochastic Hurst exponent) with characteristics of abrupt changes in the fractal structure that can be related to the theory of outliers (see Czarnecki et al. [2008], Xu et al. [2009]). In general, outlier difference itself from noise through its inherently isolated and local character, leading to non-stationarity and highly erratic behaviour. Outliers in time series have an inherently isolated and local character, with erratic behaviour (spikes), that can be detected and localised in time with the help of the effective local Holder exponents (ELHE) (see Struzik et al. [2002]). Even though the width of the multifractal spectra is capable of indicating the presence of large shocks, the oscillating nature of the local Holder exponent characterise the continuously changing dynamics of the response time distribution. Thus, the local Hurst exponent gives valuable indications on extreme values (upward and downward) in data series. Bloch [2014] tested for the validity of the local Hurst exponent both on artificial data and on financial data, and demonstrated its ability at detecting large changes of fractal structure. Further, he showed that the obtained denoised local Holder exponents oscillate around a level of Hurst exponent with a succession of small and large amplitudes (similar to spikes), and is capable of sudden, or abrupt, change to a different level of Hurst exponent in presence of lage extreme price fluctuations related to market crashes. Thus, similarly to outliers, financial time series requires a methodology capable of determining the statistical nature of the non-stationary process both globally and locally, such as the effective local Holder exponent.

2.2 Modelling returns in the option pricing theory

It is widely accepted that asset returns follow a fat tailed non-normal distribution and that absolute returns possess long-range dependence (LRD). We saw in Section (2.1.2.1) that according to the generalised central limit theorem (GCLT), stable laws are the only possible limit distributions for properly normalised and centered sums of independent, identically distributed random variables. Further, we saw in Section (2.1.2.3) that the sums of dependent infinite variance random variables converges in distribution to an α-stable random variable for strictly stationary process satisfying the regular variation condition and other restrictive conditions. It is therefore natural to assume that asset returns are at least approximately governed by a stable distribution. Using standard statistical tools, a large number of empirical studies on asset prices have investegated self-similarity and LRD properties of asset returns. However, we also know that many variables in complex natural and engineered systems are actually correlated, or, have some nonlinear interdependence. Further, we know now that financial time series experience nonlinear interdependence with fractal structure capable of sudden, or abrupt, changes. We are first going to review some of the literature providing statistical tests on the estimation of scaling exponents and explain how these methods failed to capture the behaviour of financial returns. We will then discuss how researchers moved away from using a single exponent to considering a multitude of scaling exponents to account for extreme values in the data.

2.2.1 Assessing the properties of asset returns

2.2.1.1 Some empirical results

With the availability of new sources of financial data, the 1990s saw a surge of interest in whether or not there is long-range dependence (LRD) in asset returns. The concepts of self-similarity, scaling, fractional processes were used to assess market returns. For instance, various authors used statistical methods of inference to estimate the tail index without assuming a particular shape of the entire distribution (Jansen et al. [1991], Loretan et al. [1994], Lux [1996], Dacorogna et al. [2001]). The tail index, α, was found to be in the range of 3 to 4, giving weight to the stability of the tail behaviour under time aggregation. As a result, it was then assumed that the unconditional distribution of returns converged toward the Gaussian distribution, but was distinctly different from it at the daily (and higher) frequencies. Hence, the non-normal shape of the distribution motivated the quest for the best non-stable characterisation at intermediate levels of aggregation. While the dependency of long lasting autocorrelation was subject to debate for raw (signed) returns, it was plainly visible in absolute returns, squared returns, or any other measure of the extent of fluctuations (volatility). Still using statistical tests, other authors studied long term dependence with more or less success on finding deviations from the null hypothesis of short memory for raw asset returns, but strongly significant evidence of long memory in squared or absolute returns (see Lo [1991], Ding et al. [1993], Longin [1996]).

Eveven though it has been assumed little evidence of fractional integration in stock returns, long memory has been identified in the first differences of many economic series a long time ago. For instance, Mandelbrot [1974] argued against martingale models and Maheswaran et al. [1993] suggested potential applications in finance for processes lying outside the class of semimartingales. Econophysics developed to study the herd behaviour of financial markets via return fluctuations, leading to a better understanding of the scaling properties based on methods and approaches in scientific fields. To measure the multifractals of dynamical dissipative systems, the generalised dimension and the spectrum have effectively been used to calculate the trajectory of chaotic attractors that may be classified by the type and number of the unstable periodic orbits. Even though a time series can be tested for correlation in many different ways (see Taqqu et al. [1995]), some attempts at computing these statistical quantities emerged from the box-counting method, while others extended the R/S analysis (see Mandelbrot et al. [1979]). The moment-scaling properties of financial returns have been the object of a growing physics literature confirming that multiscaling was exhibited by many financial time series (see Vandewalle et al. [1998], Schmitt et al. [1999], Pasquini et al. [2000]). Since scaling analysis and multifractal analysis developed, various authors performed empirical analysis to identify anomalous scaling in financial data (see Calvet et al. [2002]).

Since the absence of long range dependence (LRD) in returns is still compatible with its presence in absolute returns (or volatility), several authors suggested models such as the Fractionally Integrated GARCH models, where returns have no autocorrelation but their amplitudes have LRD (see Baillie [1996]). Many continuous multifractal models, such as the multifractal model of asset returns (MMAR), have been proposed to capture the thick tails and long-memory volatility persistence exhibited in the financial time series (see Mandelbrot et al. [1997]). Such models are consistent with economic equilibrium, implying uncorrelated returns and semimartingale prices, thus precluding arbitrage in a standard two-asset economy. Returns have a finite variance, and their highest finite moment can take any value greater than 2. However, the distribution does not need to converge to a Gaussian distribution at low frequencies and never converges to a Gaussian at high frequencies, thus capturing the distributional nonlinearities observed in financial series. These multifractal models have long memory in the absolute value of returns, but the returns themselves have a white spectrum. That is, there is long memory in volatility, but absence of correlation in returns. Subsequent literature moved from the more combinatorial style of the multifractal model of asset returns (MMAR) to iterative causal models of similar design principles, such as the Markov-switching multifractal (MSM) model proposed by Calvet et al. [2004] and the multifractal random walk (MRW) introduced by Bacry et al. [2001], constituting the second generation of multifractal models.

Examining the stability of the Hurst exponent, H, on financial time series on the basis of characteristic values, such as rescaled ranges or fluctuations analysis, some authors (see Vandewalle et al. [1998c], Costa et al. [2003], Cajueiro et al. [2004], Grech [2005]) observed that the values of H could be significantly higher or lower for a specific scale. Using detrended fluctuation analysis (DFA), or detrended moving average (DMA), to analyse asset returns on different markets, various authors observed that the Hurst exponent would change over time indicating multifractal process. They found that the exponent values and the range over which the power law holds varied drastically from one underlying asset to another one, obtaining three categories, the persistent behaviour, the antipersistent behaviour, and the strictly random one. In addition, they showed in some stocks and some exchange rate that the Hurst exponent was changing with time with successive persistent and antiperisistent sequences (see Costa et al. [2003], Kim et al. [2004]). Several authors proposed to use methods of long-range analysis such as DFA or DMA to determine the local correlation degree of the series by calculating the local scaling exponent over partially overlapping subsets of the analysed series (see Vandewalle et al. [1998c], Costa et al. [2003], Cajueiro et al. [2004], Carbone et al. [2004], Matos et al. [2008]).

Later, measuring multifractality with either DFA, DMA, or wavelet analysis, and computing the local Hurst exponent on sliding windows, a large number of studies confirmed multifractality in stock market indices, commodities and FX markets, such as Matia et al. [2003], Kim et al. [2004], Matos et al. [2004], Norouzzadeh et al. [2005]. Further studies confirmed multifractality in stock market indices such as Zunino et al. [2007] [2008], Yuan et al. [2009], Wang et al. [2009], Barunik et al. [2012], Lye et al. [2012], Kristoufek et al. [2013], Niere [2013], to name but a few. Other studies confirmed multifractality on exchange rates such as Norouzzadeh et al. [2006], Wang et al. [2011b], Barunik et al. [2012], Oh et al. [2012], while some confirmed multifractality on interest rates such as Cajueiro et al. [2007], Lye et al. [2012], as well as on commodity such as Matia et al. [2003], Wang et al. [2011].

Then methods for the multifractal characterisation of nonstationary time series were developed based on the generalisation of DFA, such as the MFDFA by Kantelhardt et al. [2002]. Consequently, the multifractal properties as a measure of efficiency (or inefficiency) of financial markets were extensively studied in stock market indices, foreign exchange, commodities, traded volume and interest rates (see Matia et al. [2003], Ho et al. [2004], Moyano et al. [2006], Zunino et al. [2008], Stosic et al. [2014]). For instance, Zunino et al. [2008] used MFDFA to analyse the multifractality degree of a collection of developed and emerging stock market indices. Gu et al. [2010] analysed the return time series of the Shanghai Stock Exchange Composite (SSEC) Index with the one-dimensional MFDMA model within the time period from January 2003 to April 2008, and confirmed that the series exhibits multifractal nature, not caused by fat-tailedness of the return distribution. Lye et al. [2012] used MFDFA coupled with the rolling window approach to scrutinise the dynamics of weak form efficiency of Malaysian sectoral stock market, and showed that it was adversly affected by both Asian and global financial crises.

It was also shown that observable in the dynamics of financial markets have a richer multifractality for emerging markets than mature one. As a rule, the presence of multifractality signalises time series exhibiting a complex behaviour with long-range time correlations manifested on different intrinsic time scales. Considering an artificial multifractal process and daily records of the $S\&P$ 500 index gathered over a period of 50 years, and using multifractal detrended fluctuation analysis (MFDFA) and multifractal diffusion entropy analysis (MFDEA), Jizba et al. [2012] showed that the latter posses highly nonlinear, and long-ranged, interactions which is the manifestation of a number of interlocked driving dynamics operating at different time scales each with its own scaling function. Such a behaviour typically points to the presence of recurrent economic cycles, crises, large fluctuations (spikes or sudden jumps), and other non-linear phenomena that are out of reach of more conventional multivariate methods (see Mantegna et al. [2000]).

2.2.1.2 The difficulties of measuring empirically LRD

We saw in Section (2.2.1.1) that financial systems were assumed to display scaling properties similar to those of systems in statistical physics, such that the existence of long-term correlation could be empirically assessed using

1. fractal analysis
2. statistical self-similarity analysis (using various statistical methods)

However, these statistical approaches are based on the moment properties of stochastic processes and must be restricted to second-order stationary processes. In addition, fractal properties are only verified in the infinitesimal limit. Nonetheless, a large number of statistical studies on market prices checked for the existence of long-range and short-range power law correlation in financial data. Most of these studies did not find temporal correlations present in the system for price changes, but they did for absolute price changes, the average of absolute price changes, the square root of the variance, and the interquartile range of the distribution of price changes. We have seen in Section (2.1.4) that it was difficult to test for a particular parametric model, such as α-stable Levy processes or fractional Brownian motions, since self-similarity can have very different origins. For instance, we have seen in Section (2.1.2.2) that studies estimating the tail exponent directly from the tail observations tends to overestimate the characteristic exponent α. Also, the true tail behaviour of L-stable laws is only visible for extremely large data sets, leading to strongly misleading results when the sample is not large enough, such as the rejection of the L-stable regime (see McCulloch [1997]). Thus, we should use high-frequency data and consider the most outlying observations when estimating the tail index. In addition, if the time series of asset returns possess the two features of heavy tails and long range dependence, then many standard statistical tests will fail to work (see Resnick et al. [1999]). Also, Resnick et al. [2000] give examples of such processes where sample autocorrelations converge to random values as the sample size grows. Further, we have seen that slow decay of sample autocorrelation functions could indicate non-stationarity, making it hard to distinguish between stationary long memory models and other non-stationary models. Mikosch et al. [2003] observed that the non-stationarity of returns may generate spurious effects easily mistaken for LRD in the volatility. Granger et al. [2004] considered the interaction of LRD with non-stationarity by combining an underlying long memory process with occasional structural breaks.
More generally:

- statistical analysis applies to processes with stationary increments
- estimators are vulnerable to trends in data, periodicity, large fluctuations, etc.
- estimators of Hurst exponent works for very long, or infinite, time series
- in the time domain, at high lags only a few samples are available
- in the frequency domain, at frequencies near zero, measurement errors are largest

Yet, stock returns and FX rates suffer from systematic effects mainly due to the periodicity of human activities, and can not be considered as processes with stationary increments. Clegg [2006] showed that LRD was a very difficult property to measure in real life because the data must be measured at high lags/low frequencies where fewer readings are available, and all estimators are vulnerable to trends in the data, periodicity and other sources of corruption.
When eliminating these problems, academics found correlations present in systems and highlighted the multifractal nature of financial time series. It was shown that self-similar models were too restrictive to explain financial time series, as they were unable to capture more fully the complex dynamics of the series. While all these models assume that the response times are trial-independent random variables, financial series exhibit multifractal scaling behaviour since a multitude of scaling exponents are necessary to fully describe the series. Simply put, financial time series exhibit

- non-Gaussian distribution,
- the presence of extreme values (outliers),
- long-range dependent dynamics.

These characteristics of financial time series indicate:

1. that the variations can not be exclusively described by the scaling of the variance alone, but that the scaling of higher order statistical moments must be considered.
2. intermittent changes in the magnitude of response time variation which might be due to feedback effects, or changes in investor's behaviour. These changes provide temporal modulation of both the width and shape of the response time distribution, and consequently, temporal modulation of the scaling exponent.

2.2.2 Building a theory of option pricing

Bachelier's assumption of normally distributed returns (see Bachelier [1900]) led to the formulation of the weak form efficient market where the price changes are independent and may be a random walk (see Fama [1965a] (Fa65) [1970]). Relaxing the notion of independent returns, Samuelson [1965] (Sa65) assumed that the market is efficient if it follows a martingale process. Further, Delbaen et al. [2004] argued that discounted prices must be martingales under the risk-neutral measure $\mathbb{Q}$. Thus, the Efficient Market Hypothesis (EMH) states that in a free market all available information about an asset is already included in its price, so that there is no good buy. As a result, market prices must satisfy well known constraints of convexity. A direct consequence of the EMH is that the most important concepts in theoretical and empirical finance developed around the assumption that asset returns follow a normal distribution. This includes the classical portfolio theory, the Black-Scholes-Merton option pricing model and the RiskMetrics variance-covariance approach to Value at Risk (VaR). However, we saw in Section (2.2.1.1) that a number of academic studies have shown that the underlying price movements are neither normally nor lognormally distributed. Many financial assets exhibit more skewness and kurtosis than it is consistent with the Geometric Brownian Motion model of Black-Scholes (see Mandelbrot [1963]). Mandelbrot described the financial market as a system with fat tails, stable distributions and persistence. The argument of large data sets exhibiting heavy tails and skewness combined with the generalised CLT theorem was put together to justify the use of stables models. Thus, a pricing theory developed where returns belong to stable laws which are the only possible limit distributions for properly normalised and centered sums of i.i.d. random variables. Assuming stock prices to be the sum of many small terms, stable models have been considered to describe financial systems. Combining strong empirical evidence for fat tails and skewness with GCLT was enough to justify the use of stable models.

The Black-Scholes model [1973] for pricing European options assumes a continuous-time economy where trading can take place continuously with no differences between lending and borrowing rates, no taxes and short-sale constraints. Investors require no compensation for taking risk, and can construct a self-financing riskless hedge which must be continuously adjusted as the asset price changes over time. A major assumption of the BS model is that the underlying returns are normally distributed with a variance proportional to the length of time over which the asset trades. Even though the BS-model of option pricing is based on a normal distribution, if the distribution is actually a fat-tailed one, then the model will under-price options that are far out of the money, since a 5 or 7-sigma event is much more likely than the normal distribution would predict. The problem with the distributional assumption of the BS model means that it generally underestimates equity, or FX, option values because the likelihood of having an extreme price movement is greater than the model expects. Non-normal skewness and kurtosis in option-implied distributions have been found to contribute significantly to the phenomenon of volatility smile. For instance, Sterge [1989] found that very large (three or more standard deviations from the norm) price changes could be expected to occur two or three times as often as predicted by normality. Hence, given sufficiently long record of stock market, the observed largest historical price changes would be so unlikely under the normal law that one would have to wait for horizons beyond at least the history of stock markets to observe them occur with non-negligible probability. In reality, since the volatility is not constant it has a major impact on the values of certain options, especially those options that are away from the money, because the dynamics of the volatility process rapidly change the probability that a given out-of-the-money (OTM) option can reach the exercise price.

In order to justify the use of the Central Limit Theorem, it has been argued that financial asset returns are the cumulative outcome of a vast number of pieces of information and individual decisions arriving almost continuously in time (see McCulloch [1996]). Thus, returns have been modelled by the Gaussian distribution. However, financial markets are populated with interacting agents making only imperfect forecast and directly influencing each other. The uneven assimilation of information causes a biased random walk, with bias abruptly changing in direction or magnitude. These intermittent changes in the magnitude of response time variation provide temporal modulation of both the width and shape of the response time distribution, leading to temporal changes in the Hurst exponent. The presence of multifractality signalises time series exhibiting a complex behaviour with long-range time correlations manifested on different intrinsic time scales. Such a behaviour typically points to the presence of recurrent economic cycles, crises, large fluctuations, and other nonlinear phenomena that are out of reach of more conventional methods. As a result, the multifractal nature of financial markets contradicts the EMH, and thus the Black-Scholes pricing model and its associated assumptions. For instance, Bergman et al. [1996] showed that dependence of the volatility on a second Brownian motion, or jumps in the stock price, could lead to non-increasing, non-convex European call prices. In a complete market, El Karoui et al. [1998] showed that when the volatility of the underlying stock was allowed to be random in a path-dependent way, the value of a European call could decrease with increasing volatility, and even decrease with increasing stock price.

Nonetheless, it seems that the academics who developed the option pricing theory (OPT) focused on the empirical results that demonstrated thick tails and long-memory volatility persistence exhibited in the financial time series, but did not account for the evidence of long memory in raw returns. Such models of uncorrelated returns (white spectrum) and semimartingale prices are consistent with economic equilibrium and the EMH of Samuelson [1965]. It led to the well known stylised facts on volatility showing volatility clustering, asymmetry and mean reversion, comovements of volatilities across assets and financial markets, stronger correlation of volatility compared to that of raw returns, (semi-) heavy-tails of the distribution of returns, anomalous scaling behaviour, changes in shape of the return distribution over time horizons. For example, Engle [1982] proposed to model volatility as conditional upon its previous level, that is, high volatility levels are followed by more high volatility, while low volatility is followed by more low volatility. Engle [1982] and LeBaron [1992] found supportive evidence of the autoregressive conditional heteroskedastic (ARCH) model family, such that standard deviation is not a standard measure. This led jump-diffusion models to develop in continuous time, such as the Merton model (see Merton [1976]), the Heston model (see Heston [1993]), or a combination of both.

2.2.3 Towards jump-diffusion models

2.2.3.1 A simple model to generate the smile

As discussed by Clark [1973], distributions with fat tails can be explained with the notion of mixture of normal random variables with different variances. For instance, drawing at random the variance used to generate normal returns produces fat tails. We can build a jump-diffusion model differing from a BS-model only in the distribution of shocks. This is achieved by adding

- jumps to the process, where jumps means significant unexpected discontinuous changes in prices.
- stochastic volatility, meaning volatility changing at random over time.

For example, assuming an expected frequency of jumps per year, the process jumps by adding independent normal random variables. Usually, the jump arrival process follows a Poisson process. Fat tails can also be produced with a stochastic volatility σ_t changing at random, with persistence. Persistence means that relatively high recent volatility implies a relatively high forecast of volatility in the near future, while recent low volatility is associated with a prediction of lower volatility in the near future. Hence, a jump-diffusion model can also be approximated in a discrete time setting by an extreme version of a stochastic volatility model having non-persistent random volatility. That is, every day the volatility is drawn at random independently of the previous day.
We are now going to illustrate this approach with a very simple model. At a meeting of the Financial Research Initiative at Stanford University in March 1996, Robert Litterman proposed to simulate a random variable X of zero mean and unit variance but with a given degree of tail fatness (see Duffie et al. [1997]). To do so, we first draw a random variable Y whose outcomes are 1 and 0 with respective probabilities p and $1-p$, and then independently we draw a standard normal random variable Z, $Z \sim N(0,1)$. We let α and β be the standard deviations of the two normals to be mixed. The random variable X is defined as follow

$$X = \begin{cases} \alpha Z \text{ if the outcome of } Y \text{ is } 1 \\ \beta Z \text{ if the outcome of } Y \text{ is } 0 \end{cases}$$

that is, $X = \alpha Z I_{\{Y=1\}} + \beta Z I_{\{Y=0\}}$. From independence we get $E[Z I_{\{f(Y)\}}] = E[Z]E[I_{\{f(Y)\}}]$, such that $E[X] = 0$. The standard deviation α and β must be chosen such that the variance of X is 1, that is,

$$Var(X) = E[X^2] = p\alpha^2 + (1-p)\beta^2 = 1$$

We can then deduce β as

$$\beta = \sqrt{\frac{1-p\alpha^2}{1-p}}$$

We are left with choosing p and α to satisfy either a given kurtosis or a 0.99 critical value. The kurtosis of X is given by

$$E[X^4] = 3\big(p\alpha^4 + (1-p)\beta^4\big)$$

Note, this method is consistent with a jump-diffusion model.

2.2.3.2 Explaining the implied volatility

The financial literature comprises a lot of factors possibly responsible for smile and term structure patterns ranging from market microstructure frictions, such as liquidity constraints and transaction costs, to stochastic volatility and Levy processes for the underlying asset price process. Among the different explanations for that shift in implicit distributions, the most studied one is a change in investor's assessment of the underlying stochastic process. This is because the IV is a global measure of volatility and can not be used directly as an input factor since pricing requires a local measure of volatility. On the other hand, as discussed in Section (2.2.3.1), we can formulate the IV as an additional stochastic process entering the pricing equation of derivatives. It is well known that both the crash fears and the volatility evolution are explanatory factors for the negatively skewed implicit distribution, and that each of them implies a different relationship between the option maturity and the implicit skewness. Stochastic diffusion models imply a direct relationship between the option maturity and the magnitude of the implicit skew while that relationship is inversed in a finite variation jump model. This is because jump components address moneyness biases, while having stochastic latent variables allows distributions to evolve stochastically over time. Hence, the BS model consistently underestimates the value of an option to the extent that volatility is stochastic rather than constant as assumed. However, it appears that stochastic volatility smiles are too shallow, while jump diffusions imply the smile only for short maturity options. Moreover, when fitted to the substantial negative skewness and leptokurtosis implicit to the short maturity options prices, these models assign a huge probability of getting large weekly movements in the underlying returns which are not observed in practice. Nonetheless, according to Bakshi et al. [1997] only a combination of jump and stochastic volatility models is capable of capturing the IVS. Similarly, Bates [2000] studied empirically the impact of each explanatory factors on the shifted distribution and concluded that one need a combination of both to recover a good fit to the market distribution.

Chapter 3

Models generating the smile

As discussed in Section (1.5.2), a European call price can be decomposed into a Digital Stock, Δ_S, and a Digital Bond, Δ_K. In the Black-Scholes formula these terms simplifies to functions of the cumulative distribution function (CDF). The values of the CDF, $N(x)$, may be approximated by a variety of methods such as numerical integration, Taylor series, asymptotic series and continued fractions. As a result, numerous models have been proposed to generate the smile, and a considerable body of literature exist proposing methods for extracting the risk-neutral distribution from option prices (see Bahra [1997]). However, any successful generalisation of the Gaussian assumption must achieve flexibility and analytical tractability. We are now going to describe a few tractable models generating the smile. Assuming a time-homogeneous Markov process X_t valued in an open subset $D \in \mathbb{R}^n$, we let $V(X_t)$ be the instantaneous variance process of the stock price $\{S_t; t \in [0,T]\}$ over the life of the option price, and define the mean variance as

$$I^2(t,T) = \frac{1}{T-t} \int_t^T V(X_s) ds$$

In that setting, we are going to show that most models generating the smile can be decomposed as

$$C(t, S_t; K, T; I(t,T)) = C_{BS}(t, S_t, K, T; I_0(t,T)) + \alpha(K,T) \tag{3.0.1}$$

where $C_{BS}(t, x, K, T; \sigma)$ is the Black-Scholes call price with volatility σ, and $I_0^2(t,T) = E_t[I^2(t,T)]$ is the expected mean variance.

3.1 A family of local volatility models

3.1.1 Black-Scholes formula generating the smile

We are going to describe a few models capable of generating asymmetric distribution to recover the market implied skew by using the standard Black-Scholes formula with appropriate input parameters.

3.1.1.1 Displaced diffusion

We let the underlying process $K(t,T)$ be the forward price with fixed maturity T. We assume a fixed period of time $[0,T]$ where $t = 0$ is the inception of the product and T is its maturity. A commonly used way of getting asymmetric risk-neutral distribution to recover the market implied skew is to shift the forward price with a constant ν (see Section (1.7.4.1)). So, in our framework the dynamics of the shifted-forward price become

$$\frac{dK(t,T)}{\big(K(t,T) + a(T)\big)} = \nu\gamma\sigma(t) dZ_K(t)$$

where $a(T)$ is a constant in $[0, T]$ sometime given by $a(T) = \frac{1-\nu}{\nu} K(0, T)$. Note, ν is a blending parameter taking value in the range $[0, 1]$ and controlling the slope of the local volatility function $\psi(x) = \nu x + (1-\nu)x_0$. In the special cases where $\nu = 0$ the forward price is normally distributed, while when $\nu = 1$ it is log-normally distributed and both of them belong to the class of the Affine models. Moreover, since the local volatility $\psi(x)$ is linear in x the model can be transformed into an Affine model. We now assume that calculation time is $t_0 = 0$ and denote $K_L(t, T) = \ln\big(K(t, T) + a(T)\big)$ and then apply Ito's lemma to get its dynamics

$$dK_L(t, T) = -\frac{1}{2}\nu^2\gamma^2\sigma^2(t)dt + \nu\gamma\sigma(t)dZ_K(t)$$

We then integrate this SDE between $[0, t]$ and take the exponential, getting

$$K(t, T) = (K(0, T) + a(T))U(t, T) - a(T)$$

which is affine with respect to the process $U(t, T)$ and which we call the modified price. Note, in the special case where $a(T) = \frac{1-\nu}{\nu} K(0, T)$ we get $K(0, T) + a(T) = \frac{K(0,T)}{\nu}$. The solution to the process $U(t, T)$ at time t is

$$U(t, T) = e^{-\frac{1}{2}\nu^2\int_0^t \gamma^2\sigma^2(s)ds + \nu\int_0^t \gamma\sigma(s)dZ(s)} \tag{3.1.2}$$

and applying Ito's lemma, it satisfies the dynamics

$$\frac{dU(t, T)}{U(t, T)} = \nu\gamma\sigma(t)dZ(t) \; , U(0, T) = 1$$

We let $F(t, T) = (K(0, T) + a(T))U(t, T)$ be the modified forward price seen at time t with the fixed maturity T and dynamics

$$\frac{dF(t, T)}{F(t, T)} = \nu\gamma\sigma(t)dZ(t) \; , F(0, T) = K(0, T) + a(T)$$

and we re-express the European call price $C(t; K, T)$ seen at time t with a strike K and maturity T in terms of the logarithm of the forward price $Y(t, T)$. It satisfies the equation

$$C(t; K, T) = P(t, T)E^Q[((K(0, T) + a(T))U(T, T) - \overline{K})^+|\mathcal{F}_t] = P(t, T)E^Q[(F(T, T) - \overline{K})^+|\mathcal{F}_t]$$

where $\overline{K} = a(T) + K$. Expanding the solution of the forward price, we get

$$C(t; K, T) = P(t, T)F(t, T)N(\overline{d}_1) - \overline{K}P(t, T)N(\overline{d}_2)$$

where

$$\overline{d}_1 = d_1 - \frac{1}{\sigma_F\sqrt{(T-t)}}\log\big(1 + \frac{a(T)}{K}\big) \text{ and } \overline{d}_2 = \overline{d}_1 - \sigma_F\sqrt{(T-t)}$$

3.1.1.2 A modifed Black-Scholes formula

The price of the modified Black-Scholes call option seen at time t_0 with strike K, maturity T and shift parameter μ_T, under the risk-neutral probability measure $\mathbb{Q}$, is given by

$$C(t_0; K, T) = E^Q[e^{-\int_{t_0}^T r_s ds}(S_T - \overline{K})^+|\mathcal{F}_{t_0}] \tag{3.1.3}$$

where $\overline{K} = K(1 + \mu_T)$. Assuming a deterministic repo rate, we let $X_t = S_t Re(t, T)$ be the present value of the stock price S_T seen at time t, and set the exercise event to be $\xi = \{X_T > \overline{K}\}$. The call option price becomes

$$C(t_0;K,T)=S_{t_0}Re(t_0,T)N(\overline{d}_1)-\overline{K}P(t_0,T)N(\overline{d}_2)$$

where

$$\overline{d}_1=\frac{1}{\sigma\sqrt{(T-t_0)}}\log\frac{F(t_0,T)}{\overline{K}}+\frac{1}{2}\sigma\sqrt{(T-t_0)}\text{ and }\overline{d}_2=\overline{d}_1-\sigma\sqrt{(T-t_0)}$$

which we can write as $\overline{d}_1=d_1-\frac{1}{\sigma\sqrt{(T-t_0)}}\log(1+\mu_T)$. Setting $\mu_T=\frac{a(T)}{K}$, this model is similar in spirit to the displaced diffusion described in Section (3.1.1.1). Bloch [2012a] showed that the call price could be approximated as

$$C(t_0;K,T)\approx(1+\mu_T)C_{BS}(t_0,S_{t_0},K,T;\sigma)+\alpha(K,T)\tag{3.1.4}$$

with

$$\begin{aligned}\alpha(K,T)&=-\frac{1}{\sqrt{T-t_0}}Vega_{BS}(t_0,T)\Big(\mu_T\frac{N(d_1)}{n(d_1)}+\frac{1}{2}\frac{1}{\sigma^2(T-t_0)}\big(\log(1+\mu_T)\big)^2\big(d_1-(1+\mu_T)d_2\big)\\&-\mu_T\frac{1}{\sigma\sqrt{(T-t_0)}}\log(1+\mu_T)\Big)\end{aligned}$$

In the special case where the strike is at-the-money forward, we can use Equation (1.2.12), so that the call price simplifies to

$$\begin{aligned}&C(t_0;K,T)\big|_{K=F(t_0,T)}\approx(1+\mu_T)0.4S_{t_0}Re(t_0,T)\sqrt{\omega(T-t_0)}\\&-\frac{1}{\sqrt{T-t_0}}Vega_{BS}(t_0,T)\Big(\mu_T\frac{N(d_1)}{n(d_1)}+\frac{1}{2}\frac{1}{\sigma^2(T-t_0)}\big(\log(1+\mu_T)\big)^2\big(\sigma\sqrt{(T-t_0)}-\mu_T\big)\\&-\mu_T\frac{1}{\sigma\sqrt{(T-t_0)}}\log(1+\mu_T)\Big)\end{aligned}\tag{3.1.5}$$

with $d_1=\frac{1}{2}\sqrt{\omega(T-t_0)}$ and $d_2=-d_1$ for $\omega(T-t_0)=\sigma\sqrt{(T-t_0)}$.

***Remark* 3.1.1** *Contrary to the BS-formula, the ATM-forward call option price is not a linear function of the spot price and the square root of the total variance, as it depends in a complex way on the shift parameter μ_T.*

3.1.2 A stochastic volatility approximation

3.1.2.1 The collapse process model

Following market practice, we assume that the stochastic volatility process $(V_t)_{t\geq 0}$ takes values in $\mathbb{R}$ with the dynamics

$$dV_t=\mu(V_t)dt+\sigma(V_t)dW_V(t)$$

Further, we assume that the drift $\mu(V_t)$ is mean-reverting, and that the Brownian motions W_S and W_V are independent. Rather than imposing a dynamics to the volatility process as was done by Hull et al. [1987] and Stein et al. [1991], we are going to focus on the distribution of the stock price and consider a statistical model. It is well known that the asymptotic behaviour of an Ornstein-Uhlenbeck (OU) process in large mean reversion rate is to converge towards a collection of distinct independent random variables all sharing the same distribution. Hence, to make calculations feasible we assume that the stochastic volatility becomes the collapse process, $V_t=\widehat{V}$, with the same random variable for all time t. It becomes a non-persistent random volatility model consistent with a jump-diffusion model. In that setting the random variable is independent from the stock price and we let the instantaneous volatility be given by

$$\sigma(t,\omega) = (1 + \lambda \widehat{V})\psi(t, S_t)$$

where $\psi(t, S_t)$ is a deterministic function of time t and spot price S_t, and $\lambda \in [0, 1]$ is a constant controlling the intensity of the collapse process. Note, the modified instantaneous volatility is now stochastic and capable of rising or falling without a movement in spot prices as observed in the market. For $\lambda = 0$ we recover the local volatility $\psi(t, S_t)$ and for $\lambda = 1$ we get maximum variations from $\widehat{V}$.

3.1.2.2 Fast pricing

For simplicity of exposition, we let $\psi(t, S_t)$ be a function of time and define the square-root of the average variance as $\Sigma(t) = \sqrt{\frac{1}{t}\int_0^t \sigma^2(s)ds}$. Further, we let $I^2(t, T)$ be the mean variance over the life of the option price defined as

$$I^2(t,T) = \big[(1 + \lambda \widehat{V})\Sigma(T - t)\big]^2 \quad (3.1.6)$$

We then let $C(t, S_t, K, T; \widehat{V})$ be the price at time t of a contingent claim on climate with maturity T and strike K. Given the call price $C(T, S_T, K, T; \widehat{V})$ at maturity, the discounted call price must be a $\mathbb{Q}$-martingale so that its value at time t satisfies

$$C(t, T; x, y) = E_t^Q[e^{-\int_t^T r_s ds} C(T, S_T, K, T; \widehat{V}) | S_t = x, \widehat{V} = y] \quad (3.1.7)$$

which is solved numerically. However, the volatility $\widehat{V}$ being a random variable, we do not know with certainty its value at time t. Conditional on the mean variance $I^2(t, T)$, the process S_t is lognormally distributed. Therefore, from the properties of conditional expectation [1], the value of a call option in the stochastic local volatility model becomes

$$C(t, T; S_t) = E_t^Q[C(t, T; S_t, \widehat{V})]$$

where $C(t, T; S_t, \widehat{V}) = C_{BS}(t, S_t; T, K; I(t, T))$ is the BS-formula with volatility $\sigma_{BS} = I(t, T)$. Hence, if $\widehat{V}$ is a discrete random variable such that v_i is in the range of $\widehat{V}$, then $\widehat{V} = v_i$ is an event with nonzero probability and can be used as a partition of the sample space denoted B. Using the partition theorem, the price of the claim can be approximated as

$$C(t, T; S_t) = \sum_{i=1} C(t, T; S_t, v_i) P(\widehat{V} = v_i)$$

More generally, if B is a sufficiently nice subset of $\mathbb{R}$ then, for the variable X we get

$$P(X \in B) = \int_B f_X(x)dx$$

where $f_X(x)$ is the probability density function of X. We can then think of the pdf $f_X(x)dx$ as the element of the probability $P(X \in dx)$ since

$$P(X \in dx) \approx f_X(x)dx$$

For example, to get plausible values for the mean variance in Equation (3.1.6) that guarantee its positivity, we must draw $\widehat{V}$ in the range $[-1, 1]$. Beta and Kumaraswamy distributions are the most popular models to fit continuous bounded data. For ease of computation, we consider the Kumaraswamy's double bounded distribution defined on the interval $[z_{low}, z_{high}]$ with the shape parameters a and b (see details in Appendix ()). For $\lambda = 0$ the BS-volatility corresponds to the term structure $\sigma_{BS} = \Sigma(t)$, while for $\lambda =]0, 1]$ the variable $\widehat{V}$ is drawn in the range $[-\lambda, \lambda]$ automatically resizing the bounds z_{low} and z_{high}. By varying the parameters a and b we obtain a range of distributions

[1] $E[E[X|Y]] = E[X]$

for the variable $\widehat{V}$ which are either centered around zero (symmetric), or having positive or negative skew, generating skew and curvature to the smile. Then, the value of the call option is approximated as

$$C(t,T;S_t) \approx \sum_{i=1}^{n} C(t,T;S_t,v_i)\phi_{\widehat{V}}(v_i)dv_i$$

where n is a positve integer, and $\phi_{\widehat{V}}(\bullet)$ is the Kumaraswamy's density. Note, the value of n is numerically chosen such that

$$\sum_{i=1}^{n} \phi_{\widehat{V}}(v_i)dv_i = 1$$

to recover a valid density.

3.1.3 Approximating non-normal density

3.1.3.1 Pricing with the Gram-Charlier density

While we can model the smile in continuous time with jump-diffusion models, it can also be modelled in discrete time with Gram-Charlier expansions. In the latter, the conditional distribution of log-changes being specified directly with some flexibility over its form, the Gram-Charlier expansion is a normal density augmented with higher moments. While Jarrow et al. [1982] approximated the conditional distribution of the price, Corrado et al. [1996] and Backus et al. [1997] approximated that of the logarithm of the price and derived an implied volatility as a quadratic function of moneyness with coefficients related to the skewness and kurtosis of the price.
Assuming n periods of time, the stock price satisfies $S_{t+n} = S_t e^{x^n_{t+1}}$ where $x^n_{t+1} = \sum_{j=1}^n x_{t+j}$. Further, assuming that the n-period log-price change x^n_{t+1} is normal with conditional mean, μ_n, and standard deviation, σ_n, the standard normal variable becomes

$$w = \frac{x^n_{t+1} - \mu_n}{\sigma_n}$$

and the Gram-Charlier expansion around w is given by

$$f(w) = \phi(w) - \gamma_{1n}\frac{1}{3!}D^3\phi(w) + \gamma_{2n}\frac{1}{4!}D^4\phi(w)$$

where $\phi(\bullet)$ is the standard normal density and D^j is the jth derivative operator. Note, k_j is the jth cumulant of a random variable x, and m^c_j is its jth central moment. Then, the first cumulant is the mean $k_1 = E[x]$, the second is the variance $k_2 = \sigma^2 = E[(x-k_1)^2]$, the third is the central moment $k_3 = m^c_3 = E[(x-k_1)^3]$, and the fourth is $k_4 = E[(x-k_1)^4] - 3k_2^2$. Then, the skew is a scaled version of the third cumulant

$$\gamma_1 = \frac{k_3}{k_2^{\frac{3}{2}}} = \frac{k_3}{\sigma^3}$$

and the excess-kurtosis $K-3$ with $K = \frac{m^c_4}{(m^c_2)^2}$ is a scaled version of the fourth cumulant

$$\gamma_2 = \frac{k_4}{k_2^2} = \frac{k_4}{\sigma^4}$$

Even though all moments need to be defined to obtain the density, the authors assumed that for moderate values of γ_{1n} and γ_{2n}, then $f(w)$ is a proper density. However, conditions must be imposed on the expansion coefficients to obtain a valid density with $f(x) \geq 0, \forall x$. Jondeau et al. [2001] explained how to constrain skewness, γ_1, and excess kurtosis, γ_2, to make sure that the truncated series was a valid density, and obtained an optimisation problem under constraint.

In any case, the probability distribution must be sufficiently close to a normal distribution to be approximated by a Gram-Charlier expansion. Given the cumulant generating function

$$\psi(s;w) = logE[e^{sw}] = \frac{1}{2!}s^2 + \frac{1}{3!}\gamma_{1n}s^3 + \frac{1}{4!}\gamma_{2n}s^4$$

we can differentiate it to recover the coefficients

$$\frac{\partial \psi(s;w)}{\partial s^j}\big|_{s=0} = k_{jn}$$

Using that relation between the density and its higher moments, and applying the no-arbitrage condition

$$\mu_n = (r_{nt} - q_{nt})n - \frac{\sigma_n^2}{2} - \delta_{nt} \text{ with } \delta_{nt} = \frac{\sigma_n^3 \gamma_{1n}}{3!} + \frac{\sigma_n^4 \gamma_{2n}}{4!}$$

we get the price of a call option. Further, by discarding terms involving powers three and higher of σ_n, which are negligible, we get $\delta_{nt} = 0$ and the variable of integration in the density becomes

$$w = \frac{x_{t+1}^n - (r_{nt} - q_{nt})n + \frac{\sigma_n^2}{2}}{\sigma_n} = -d_2$$

with $S_{t+n} = K$. Solving the integral, we can approximate call price as

$$\begin{aligned} C(t;K,nt) &\approx S_t e^{-q_{nt}n} N(d) - K e^{-r_{nt}n} N(d - \sigma_n) \\ &+ S_t e^{-q_{nt}n} \phi(d) \sigma_n \big[\frac{1}{3!}\gamma_{1n}(2\sigma_n - d) - \frac{1}{4!}\gamma_{2n}\big(1 - d^2 + 3d\sigma_n - 3\sigma_n^2\big)\big] \end{aligned}$$

where

$$d = \frac{\log \frac{S_t}{K} + r_{nt}n + \frac{1}{2}\sigma_n^2}{\sigma_n}$$

which corresponds to the term d_1 in the BS-formula. From the definition of the BS-vega, and since σ_n is the standard deviation of x_{t+1}^n (incrporating square root of time), we can rewrite the call option price as

$$C(t;K,nt) \approx C_{BS}(t;K,nt,\sigma_n) + \alpha(K,nt) \tag{3.1.8}$$

where

$$\alpha(K,nt) = Vega(n)\sigma_n \big[\frac{1}{3!}\gamma_{1n}(2\sigma_n - d) - \frac{1}{4!}\gamma_{2n}\big(1 - d^2 + 3d\sigma_n - 3\sigma_n^2\big)\big]$$

A modified delta is given by

$$\frac{\partial}{\partial S_t e^{-q_{nt}n}} C(t;K,nt) \approx N(d) - \beta(K,nt)$$

where $\beta(K,nt)$ is the shadow delta

$$\beta(K,nt) = -\phi(d)\big[\frac{\gamma_{1n}}{3!}\big(1 - d^2 + 3d\sigma_n - 2\sigma_n^2\big) - \frac{\gamma_{2n}}{4!}\big(3d(1 + 2\sigma_n^2) + 4d^2\sigma_n - d^3 - 4\sigma_n + 3\sigma_n^3\big)\big]$$

Letting $\sigma_{BS}(d,n)$ be the implied volatility for the modified moneyness d, it is approximated by

$$\sigma_{BS}(d,n) \approx \sigma_n \big[1 - \frac{1}{3!}\gamma_{1n} d - \frac{1}{4!}\gamma_{2n}(1 - d^2) \big] \tag{3.1.9}$$

This approximation is obtained by discarding terms involving powers two and higher order of σ_n in the function $\alpha(K, nt)$. It is approximately a quadratic function of moneyness with coefficients related to skewness and excess-kurtosis. For $|d| < \frac{3}{2}$ the differences with the true implied volatility are small, but for larger values the true IV is smaller than the approximated one. In all cases, the call price, its associated delta, and the implied volatility correspond to the conventional BS-results plus additional terms involving skewness and excess-kurtosis.

Remark 3.1.2 *The additional term $\alpha(K, nt)$ is a deterministic function of the spot price S_t, so that the Gram-Charlier option price is associated to a local volatility model. Further, that dependency is expressed via the moneyness, d, in its associated implied volatility, corresponding to the sticky delta regime of Derman (see Section ((4.1.1.2))).*

As discussed by Corrado et al. [1997], the $S\&P$ 500 index have historical rates of return with mean in the range $[0.009, 0.13]$, standard deviation in the range $[0.11, 0.35]$, skewness in the range $[-1.12, 0.15]$, and kurtosis in the range $[0.22, 8.92]$. However, the Gram-Charlier density yields negative values for certain skewness and kurtosis parameters because it is a polynomial approximation. Rubinstein [1998] provided approximate kurtosis values for which the Gram-Charlier and Edgeworth expansions do not provide negative values. Jondeau et al. [2001] characterised the boundary delimiting the domain in the skewness-kurtosis space over which the Gram-Charlier expansion is positive. They identified the region $\mathcal{D}$ in the (γ_1, γ_2)-plane for which $f(w)$ is positive definite by requiring that the polynomial $p_4(w)$ be positive

$$p_4(w) = 1 - \gamma_{1n}\frac{1}{3!}D^3\phi(w) + \gamma_{2n}\frac{1}{4!}D^4\phi(w) \geq 0 \text{ , } \forall w$$

and they used notions of analytical geometry. They found that the excess-kurtosis must be inside the iterval $[0, 4]$, and that for each acceptable excess-kurtosis value there is a symmetrical interval for the skewness which must be solved numerically. The acceptable values for skewness were in the range $[-1.05, 1.05]$. To examine the range of applicability of the Gram-Charlier approximation, Backus et al. [1997] considered distributions generated by a Merton model (see Section (3.2.1)). In that model, the density of the n-period log-price change is a countable mixtutre of normals given by

$$f(x) = \sum_{j=0}^{\infty} p_j \phi(x; \mu_0 n + j\mu_j n, \sigma_0^2 n + j\sigma_j^2 n)$$

where $p_j = e^{-\lambda n}\frac{(\lambda n)^j}{j!}$ is the probability of j jumps, and (μ_0, σ_0^2) are the mean and standard deviation of the diffusion such that

$$r_{nt} - q_{nt} = \mu_0 + \frac{\sigma_0^2}{2} + \lambda\big(e^{\mu_j + \frac{\sigma_j^2}{2}} - 1\big)$$

The jump-diffusion density exhibits greater kurtosis than the normal one, and non-zero skewness if $\mu_j \neq 0$. The first four cumulants are

$$\begin{aligned}
k_1 &= (\mu_0 + \lambda\mu_j)n \\
k_2 &= \sigma_0^2 n + \lambda(\sigma_j^2 + \mu_j^2)n \\
k_3 &= \lambda(\mu_j^3 + 3\mu_j\sigma_j^2)n \\
k_4 &= \lambda(3\sigma_j^4 + 6\mu_j^2\sigma_j^2 + \mu_j^4)n
\end{aligned}$$

In their example, the moneyness vector, d, corresponds to $N(d) = 0.1, 0.2, ..., 0.9$, annualised jump intensities range in $[1, 50]$. In the benchmark case, $r_t = q_t = 0$, $\sigma = 0.1$ (annualised), $\lambda = 10$, $\gamma_1 = 0$ and $\gamma_2 = 1$, and maturity is 1 month. In Panel A, the approximated volatility smile produced $\gamma_1 = -0.006$ and $\gamma_2 = 0.898$, while using the more accurate price equation they obtained $\gamma_1 = -0.013$ and $\gamma_2 = 0.917$. Modifying the jump intensity in Panel B, they got $\gamma_2 = 1.046$ with $\lambda = 15$, and $\gamma = 0.748$ with $\lambda = 5$. Note, for $\lambda < 5$, impying larger less frequent jumps, the estimated γ_2 were well below those of the Merton model. For six-month options the Gram-Charlier's model was not capable of reproducing $\gamma_2 = 1$ with the benchmark value of λ, which had to be lowered to $\gamma_2 = 0.5$ to get a match. Adding skewness was even more difficult as the model would underestimate the absolute amount of skewness, and for larger values underestimates the kurtosis as well. The great difficulties of the Gram-Charlier approximation to reproduce the jump-diffusion model highlited the differences between the two models.

Corrado [2007] imposed a martingale restriction in the pricing formula by using a simple parameter space reduction for the Gram-Charlier expansion coefficients. The pricing equation is expressed in terms of Hermite polynomials defined as follow:

Definition 3.1.1 *The Hermite polynomials $He_i(\bullet)$ are defined by the identity*

$$(-D)^i \phi(x) = He_i(x)\phi(x)$$

where $D = \frac{d}{dx}$ is the differential operator and $\phi(x)$ is the standard normal density function.

Hermite polynomials begin with $He_0 = 1$, $He_1(x) = x$ and $He_2(x) = x^2 - 1$. Then, for $r > 2$, they are generated recursively as

$$He_r(x) = xHe_{r-1}(x) - (r-1)He_{r-2}(x)$$

The density function corresponding to an mth-order Gram-Charlier expansion is given by

$$f^m(x) = \sum_{j=0}^m c_j He_j(x)\phi(x)$$

where the coefficients are derived from the Hermite polynomials as follow

$$\begin{aligned} c_r &= \frac{1}{r!}\int_{-\infty}^{\infty} f(x)He_r(x)dx = E_f[He_r(x)] \\ &= \frac{1}{r!}\big(\mu_r^{'} - \frac{r^{[2]}}{2 \times 1!}\mu_{r-2}^{'} + \frac{r^{[4]}}{2^2 \times 2!}\mu_{r-4}^{'} - ...\big) \end{aligned}$$

where $r^{[k]} = \frac{r!}{(r-k)!}$ is a partial factorial, and $\mu_r^{'}$ represents the rth raw moment of a target density function. In general, the function $f(x)$ is standardised to get $\mu_2^{'} = 1$ by setting $c_0 = \mu_0^{'} = 1$, $c_1 = \mu_1^{'}$ and $c_2 = 0$. Then the coefficients become

$$c_3 = \frac{(\mu_3^{'} - 3\mu_1^{'})}{6}, c_4 = \frac{(\mu_4^{'} - 3)}{24}, ...$$

and so on. In that setting, we get the skew $\gamma_1 = \frac{\mu_3^{'}}{3!}$ and excess-kurtosis $\gamma_2 = \frac{\mu_4^{'}}{4!}$, which correspond to γ_3 and γ_4 in Corrado. Further, the martingale restriction imposes that $P(0,t)E^m[S_t] = S_0$ under the risk-neutral measure. This is not the case when deriving the call option price formula with the Gram-Charlier expansion. To recover the call-put parity, we must impose the constraint

$$\sum_{j=1}^m c_j \sigma^j = 0$$

Corrado proposed to fix one degree of freedom to recover the constraint. The local coefficient c_1 is modified as follow

$$\bar{c}_1 = -\sum_{j=3}^{m} c_j \sigma^{j-1}$$

leaving the coeffiecients $c_j, j \geq 3$ free to determine the shape of the density function. The polynomial accounting for the hidden martingale restriction is

$$\bar{p}_4 = 1 + \gamma_3(He_3(x) - x\sigma^2 t) + \gamma_4(He_4(x) - x\sigma^3 t^{\frac{3}{2}})$$

Note, Backus et al. [1997] used δ_{nt} as an explicit martingale restriction, while Corrado [2007] considered an implicit martingale restriction, the latter being a particular case of the former. Schlogl [2010] proposed an option pricing formula in terms of the full series and presented a fitting algorithm ensuring that a series truncated at an arbitrary moment represents a valid probability density. The following assumption describe a sufficient condition to obtain no-arbitrage at a fixed maturity T.

Assumption 1 *The standardised risk-neutral distribution, seen at time t for the maturity T, of the logarithm of the asset price $S_T = X(T)$ is given by the Gram-Charlier expansion*

$$f(x) = \sum_{j=0}^{\infty} c_j He_j(x)\phi(x)$$

where

$$c_r = \frac{1}{r!} \int_{-\infty}^{\infty} f(x) He_r(x) dx$$

and $\phi(x)$ is the standard normal density function.

Proposition 6 *Under the previous assumption, the risk-neutral expected value μ of the logarithm of the asset price $S(T)$ must satisfy*

$$\mu = \ln \frac{S(t)}{P(t,T)} - \ln \sum_{j=0}^{\infty} c_j \sigma^j - \frac{1}{2}\sigma^2$$

where $S(t)$ is the current asset value and σ is the standard deviation of $\ln S(T)$.

Given the previous Assumption, the price of a call option at time t on the asset $S(T)$ with expiry T is given by

$$\begin{aligned} C(t,K,T) &= S(t)N(d^*) - KP(t,T)N(d^* - \sigma) \\ &+ S(t)\Big(\sum_{j=0}^{\infty} c_j \sigma^j\Big)^{-1} \phi(d^*) \sum_{j=2}^{\infty} \sum_{i=1}^{j-1} c_j \sigma^i He_{j-1-i}(-d^* + \sigma) \end{aligned}$$

where

$$d^* = \frac{\mu - \ln K + \sigma^2}{\sigma}$$

To make sure that the calibrated expansion coefficients yield a valid density, the author adapted an unconstrained non-linear optimisation algorithm. He let $c_0 = 1$, standardised $f(x)$ by setting $c_1 = c_2 = 0$, and truncated the series by setting $c_j = 0 \ \forall j > k$ for some choice of $k \geq 4$. The set C_k for $k = 3, .., k$ defines a convex set. The boundaries of C_k along any given direction $d \in \mathbb{R}^{k-2}$ is obtained by solving two one-dimensional searches. Then a multiple dimensions minimisation algorithm is used to calibrate the model to market prices.

3.1.3.2 A link with the local volatility density

We derived in Equation (1.7.46) the risk-neutral density function of a local volatility model, and we showed in Equation (1.7.48) that it was a function of the skew $\partial_K \Sigma(K,T)$ and the curvature $\partial_{KK}\Sigma(K,T)$ of the implied volatility surface. One way forward, is to assume that the IV surface is a quadratic function of the moneyness $m = \frac{K}{S_t}$, and to compute its associated skew and curvature. However, to conform with the Gram-Charlier expantion in Section (3.1.3.1), we will assume that the IVS is a quadratic function of the Black-Scholes term d_1, given by

$$\Sigma(d_1) = a + bd_1 + cd_1^2$$

Since $\frac{dd_1}{dK} = -\frac{1}{S_T \Sigma \sqrt{T-t}}$ and $\frac{d}{dd_1}\Sigma(d_1) = b + 2cd_1$, we get the skew in terms of b and c as

$$\partial_K \Sigma(K,T) = -\frac{1}{S_T \Sigma \sqrt{T-t}}(b + 2cd_1)$$

and since $\frac{d^2 d_1}{dK^2} = \frac{1}{S_T^2 \Sigma \sqrt{T-t}}$, the curvature satisfies

$$\partial_{KK}\Sigma(K,T) = (\frac{1}{S_T \Sigma \sqrt{T-t}})^2 2c + \frac{1}{S_T^2 \Sigma \sqrt{T-t}}(b + 2cd_1)$$

Putting terms together, the risk-neutral density in Equation (1.7.46) becomes

$$\begin{aligned} p(d_1) &= \frac{1}{S_T \Sigma \sqrt{T-t}} + \frac{2d_1}{S_T \Sigma^2 \sqrt{T-t}}(b+2cd_1) - \left[\sqrt{T-t}d_1 - \frac{1}{\Sigma}d_1^2\right]\frac{1}{S_T\Sigma\sqrt{T-t}}(b+2cd_1)^2 \\ &+ \frac{1}{S_T \Sigma^2 \sqrt{T-t}}2c + \frac{1}{S_T \Sigma}(b + 2cd_1) \end{aligned}$$

which is a polynomial in d_1 of the fourth order. In the special case where we consider the Gram-Charlier approximation $\Sigma(d_1)\sqrt{T-t}$ in Equation (3.1.9), we get

$$a = \Sigma(K,T)(1 - \frac{\gamma_{2n}}{4!})\ ,\ b = -\frac{\Sigma(K,T)\gamma_{1n}}{3!}\ ,\ c = -\frac{\Sigma(K,T)\gamma_{2n}}{4!}$$

which we replace in $p(d_1)$ to recover the density expressed in terms of γ_{1n} and γ_{2n}.

3.2 Jump-diffusion in Affine models

3.2.1 The Merton model

We are now going to add jumps to the dynamic of the stock price in the Black-Scholes model. We will follow the approach of Merton [1976] which consist in conditioning the expectation of the call price with respect to the number of jumps, obtaining a weighted sum of BS-models. We assume that the dynamics of the jump-diffusion process in the Merton model are

$$dS_t = S_{t_-}\left(rdt + \sigma(t)dW(t) + (J-1)dN_t - \lambda E[J-1]dt\right)$$

where the volatility $\sigma(t)$ is a deterministic function of time, and N_t denotes a Poisson process with positive jump size J and compensator $\lambda E[J-1]$. In the Merton model, the size of the jump J is a random variable such that $\log(J)$ is normally distributed with constant mean μ and constant variance γ^2, that is,

$$J = e^{\mu + \gamma Z_t}$$

where Z_t follows a standard normal law. We apply Ito's lemma to the logarithm of the stock price and integrate in the range $[0,T]$ to get

$$S_T = S_0 e^{\int_0^T \left(r - \lambda E[J-1] - \frac{1}{2}\sigma^2(s)\right) ds + \int_0^T \sigma(s) dW(s) + \sum_{i=1}^{N_T} \log(J)}$$

If we allow for n fixed jumps between $[0,T]$, so that the number of jumps is no longer random, we get

$$S_T = S_0 e^{\int_0^T \left(r - \lambda E[J-1] - \frac{1}{2}\sigma^2(s)\right) ds + \int_0^T \sigma(s) dW(s) + n\mu + \sum_{i=1}^{n} \gamma \epsilon_i}$$

where $\frac{W_t}{\sqrt{t}}$ and ϵ_i are independent standard normal variables. The variance of the stock price is given by

$$\int_0^T \sigma_n^2(T) ds = \int_0^T \sigma^2(s) ds + n\gamma^2 = \int_0^T \big(\frac{n}{T}\gamma^2 + \sigma^2(s)\big) ds$$

so that the stock price becomes

$$S_T = S_0 e^{\int_0^T \left(r_n(t) - \frac{1}{2}(\frac{n}{T}\gamma^2 + \sigma^2(s))\right) ds + \int_0^T \sqrt{(\frac{n}{T}\gamma^2 + \sigma^2(s))} dW(s)}$$

where $r_n(t) = r - \lambda E[J-1] + \frac{n}{T}\log(E[J])$. Using the dynamics of the underlying stock price previously defined, we define the call price of a vanilla option in the Merton framework as

$$\begin{aligned} C_M(t, S_t, T, K, r; \Theta_M) &= E[e^{-\int_t^T r_s ds}(S_T - K)^+ | \mathcal{F}_t] \\ &= e^{-\int_t^T r_s ds} E[E[(S_T - K)^+ | \mathcal{F}_t, N_T - N_t]] \end{aligned}$$

where Θ_M are the model parameters of the Merton model. Using the property of independent increments of a Poisson process, the number of jumps between dates t and T, that is $N_T - N_t$, has the same law as N_{T-t}, namely a Poisson law with intensity, or frequency, $\lambda(T-t)$. The call price can be re-written as

$$C_M(t, S_t, T, K, r; \Theta_M) = e^{-\int_t^T r_s ds} E[E[(S_T - K)^+ | \mathcal{F}_t, N_{T-t}]]$$

Solving the conditional expectation with the density of the Poisson process, we get

$$\begin{aligned} C_M(t, S_t, T, K, r; \Theta_M) &= e^{-\int_t^T r_s ds} \sum_{n=0}^{\infty} e^{-\lambda(T-t)} \frac{(\lambda(T-t))^n}{n!} E[(S_T - K)^+ | \mathcal{F}_t, N_{T-t} = n] \\ &= e^{-\int_t^T r_s ds} \sum_{n=0}^{\infty} e^{-\lambda(T-t)} \frac{(\lambda(T-t))^n}{n!} C_{BS}(t, S_t, T, K, r_n, \sigma_n) e^{\int_t^T r_n(s) ds} \end{aligned}$$

since $C_{BS}(t, S_t, T, K, r_n, \sigma_n) = e^{-\int_t^T r_n(s) ds} E[(S_T - K)^+ | \mathcal{F}_t, N_{T-t} = n]$. Recall that $\lambda' = \lambda E[J]$, so that $E[J] = \frac{\lambda'}{\lambda}$. Replacing in the equation and simplifying the exponentional terms, we get

$$e^{-\int_t^T r_s ds - \int_t^T \lambda ds + \int_t^T r_n(s) ds} = e^{\int_t^T \left(-\lambda' + \frac{n}{T-t}\log(E[J])\right) ds} = (E[J])^n e^{\int_t^T -\lambda' ds}$$

so that the call price becomes

$$C_M(t, S_t, T, K, r; \Theta_M) = \sum_{n=0}^{\infty} e^{-\int_t^T \lambda' ds} \frac{(\lambda'(T-t))^n}{n!} C_{BS}(t, S_t, T, K, r_n, \sigma_n)$$

This is a weighted sum of Black-Scholes option prices each of which assuming n jumps. The weights correspond to the probability that there is n jumps before maturity T. In this framework the implied volatility surface obtained by inverting the Black and Scholes formula must be computed numerically. It is well known that a diffusion model is convexity preserving, that is, when the payoff of option prices is convex with respect to the underlying then the option price is convex. It implies that for two call prices P_A and P_B with the same expiration date and identical convex payoff function such that $\sigma_B(s,t) \geq \sigma_A(s,t)$ then $P_A(s,t) \geq P_B(s,t)$. However, Bergman et al. [1996] showed that when the underlying price follows a discontinuous process then call prices can have properties very different from those of a diffusion model. Ekstrom and Tysk [2006] gave necessary condition for convexity to be preserved in jump-diffusion models in arbitrary dimensions. Using that condition they showed that the only higher-dimensional convexity preserving models are the ones with linear coefficients. The Affine models are therefore convexity preserving and the Merton model belongs to the class of Affine models. We can then assume in our setting that if the option prices are convex at all time then they are increasing in the volatility. Bellamy and Jeanblanc [2000] showed that in a jump-diffusion model the option prices model were bounded by the underlying and were dominating the diffusion prices. We can therefore assume the implied volatility surface inferred from a jump-diffusion model dominate the one inferred from a diffusion model.

3.2.2 The yield-factor approach

3.2.2.1 Some definitions

We consider the probability space $(\Omega, \mathcal{F}, P)$ where $\{\mathcal{F}_t : t \in [0,\infty)\}$ is a right continuous filtration including all P negligible sets in $\mathcal{F}$. Each claim (Z,T) is assigned a price process given a semimartingale $S^{Z,T}$. There is no arbitrage if and only if there exist a probability measure $\mathbb{Q}$ equivalent to $\mathbb{P}$ and such that the price process is

$$S_t^{Z,T} = E^Q[e^{-\int_t^T r_u du} Z|\mathcal{F}_t]$$

Assumption 2 *We assume there exists a time-homogeneous Markov process X_t valued in an open subset $D \in \mathbb{R}^n$ such that the market value at time t of an asset maturing at $t+\tau$ is of the form*

$$f(X_t,\tau)$$

where $f \in C^{2,1}(D \times [0,\infty[)$.

Proposition 7 *Let the Markov process X_t defined on $D \in \mathbb{R}^n$ with dynamics*

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t + dZ_t \tag{3.2.10}$$

where W_t is an $\mathcal{F}_t$ standard Brownian motion on $\mathbb{R}^n$ under the risk-neutral probability measure $\mathbb{Q}$. The drift $\mu : D \to \mathbb{R}^n$ and diffusion $\sigma : D \to \mathbb{R}^{n \times n}$ are regular enough to have a unique strong solution valued in D.

In the Yield factor model, the asset $f(X_t,\tau)$ in assumption (2) is the zero coupon bond of maturity $t+\tau$ (see Duffie et al. [1996]).

Assumption 3 *We assume that there exists a measurable function $R : D \to \mathbb{R}$ such that the short rate at time t is $r_t = R(X_t)$ with*

$$R(x) = \lim_{\tau \to 0^+} -\frac{\log f(x,\tau)}{\tau} \text{ for } x \in D$$

Given the definition of the risk-free zero coupon bond, its price is given by

$$f(X_t, T-t) = E^Q[e^{-\int_t^T R(X_s)ds}|\mathcal{F}_t] \tag{3.2.11}$$

We consider Affine models and assume

Assumption 4 *For an affine process X_t, we assume there exists $A : [0,\infty[\to \mathbb{R}$ and $B : [0,\infty[\to \mathbb{R}^n$ such that $\tau \geq 0$ and $x \in D$ such that*

$$f(x,\tau) = e^{A(\tau)+B(\tau)x}$$

Proposition 8 *Under mild non-degeneracy conditions the exponential-affine assumption (4) implies that $\mu(\cdot)$ and $\sigma\sigma^\top(\cdot)$ are affine functions on D.*

We can re-think the variance-covariance matrix $\sigma\sigma^\top(x) = (\alpha_{i,j} + \beta_{i,j}x)_{1 \leq i,j \leq n}$ in terms of Bessel processes and get the following proposition

Proposition 9 *For $\sigma\sigma^\top(x)$ affine in x then under non-degeneracy conditions we can take $\sigma(x)$ as*

$$\sigma(x) = \begin{bmatrix} \sqrt{v_1(x)} & ... & 0 \\ . & ... & . \\ 0 & ... & \sqrt{v_n(x)} \end{bmatrix} \Sigma$$

where $\Sigma \in \mathbb{R}^{n \times n}$ is a positive semi-definite matrix and $v_i(x) = \alpha_i + \beta_i x$ with $\alpha_i \in \mathbb{R}$ and $\beta_i \in \mathbb{R}^n$.

The coefficient vectors β_i for $i = 1, ..., n$ generate stochastic volatility. If they are set to zero, the process in Equation (3.2.10) is a Gauss-Markov process. In that case, provided that $D = \mathbb{R}^n$, there exists a unique solution to the process X_t. However, to ensure stong solution to the process (3.2.10) when the volatility follows Proposition (9), it must be nonnegative for all i and t. Therefore, the domain D must satisfy

$$D = \{x \in \mathbb{R}^n, v_i(x) > 0 \text{ for all } i\}$$

The Bessel process can take negative trajectories, such that the volatility process $v_i(x)$ leaves the domain of definition D. However, it is well known that the drift can be constraint such that the volatility process remains positive on the ith boundary segment $\partial D_i = \{x \in \overline{D} : v_i(x) = 0\}$. We therefore must impose a restriction on the model parameters a, b, β and $\Sigma\Sigma^\top$ such that for all $i, 1 \leq i \leq n$ we must have

- for all x such that $v_i(x) = 0$, $\beta_i^\top(ax + b) > \frac{1}{2}\beta_i^\top \Sigma\Sigma^\top \beta_i$.
- for all j, if $(\beta_i^\top \Sigma)_j \neq 0$ then $v_i = v_i$.

If the constraints are satisfied, there exists a unique strong solution $(X_t)_{t \geq 0}$ to the stochastic differential equation (3.2.10) with $v_i(x) = \alpha_i + \beta_i x$ and $v_i(X_t) > 0$. Therefore, it is possible to define a non-negative variance process $V(X_t)$ such that

$$V(X_t) = \sum_i \gamma_i^2 v_i(X_t) \text{ for } \gamma_i \geq 0 \tag{3.2.12}$$

Since the variance is a linear combination of affine processes, we can re-write it as

$$V(X_t) = \overline{\alpha} + \overline{\beta} X_t$$

We denote $V_{BS}(t)$ as the square of the Black-Scholes local volatility satisfying

$$\Sigma(t; K_{ATM}, T) = \sqrt{\frac{1}{T-t} \int_t^T V_{BS}(s) ds}$$

with $\Sigma(\cdot)$ being the market implied volatility. If we differentiate the last expression with respect to time t, we can re-express the local variance in terms of the market implied volatility, that is $V_{BS}(t) = \Sigma(\cdot) - (T-t)\frac{\partial}{\partial t}\Sigma(\cdot)$. For example, one can use a spline to fit the Black-Scholes at-the-money implied volatility $\Sigma(\cdot)$ and then deduce the variance $V_{BS}(t)$. Therefore, by construction, we can impose that the Black-Scholes variance term structure be fitted exactly

$$V(X_t) = V_{BS}(t) + \overline{\beta} X_t$$

and such that the model variance is a pertubation arournd it.

In some cases we can express the characteristic function of a stock price as the price of a bond. In order to price exotic equity products within the Yield Factor approach, we need to calculate the characteristic function $\phi_Y(u,T)$ of the process $s_\tau = \ln \frac{S_T}{S_t}$ with $\tau = T - t$, and where the stock price S_t under the risk-neutral measure is given by

$$\frac{dS_t}{S_t} = r_t dt + \sum_{j=1}^n \gamma_j \sqrt{v_j(X(s))} dZ_j(t)$$

for $j = 1, ..., n$. In that setting, the characteristic function of the process s_τ becomes

$$\phi_Y(u,T) = E[e^{ius_\tau}|\mathcal{F}_t] = E[e^{iu\int_t^T r_s - \frac{1}{2}\sum_{j=1}^n \gamma_j^2 v_j(X(s))ds + iu\sum_{j=1}^n \gamma_j \int_t^T \sqrt{v_j(X(s))}dZ_j(s)ds}]$$

We first assume that the process X and the Brownian motions Z_j are independent, and then we condition the expectation in the characteristic function with respect to the variance $V(X(t))$, to get

$$\begin{aligned}\phi_Y(u,T) &= E\Big[E[e^{iu\int_t^T \big(r_s - \sum_{j=1}^n \frac{1}{2}\gamma_j^2 v_j(X(s))\big)ds + iu\sum_{j=1}^n \gamma_j \int_t^T \sqrt{v_j(X(s))}dZ_j(s)ds}]|V(X(t))\Big] \\ &= E[e^{iu\int_t^T r_s ds - \lambda\sum_{j=1}^n \int_t^T \gamma_j^2 v_j(X(s))ds}]\end{aligned}$$

with $\lambda = \frac{1}{2}(iu + u^2)$. By analogy to the Yield factor model in interest rates, and given the price of the zero coupon bond in Equation (3.2.11), we obtain the measurable function R as

$$R(X_t) = -iur_t + \lambda\sum_{j=1}^n \gamma_j^2 v_j(X(t))$$

The solution to the characteristic function $\phi_Y(u,T)$ satisfies the assumption (4), where $A(\tau)$ and $B(\tau)$ are obtained by solving a system of Ricatti equations (see Duffie et al. [2000]).

3.2.2.2 Introducing stochastic volatility models

Several authors proposed another way of getting asymmetric risk-neutral distribution to recover the market implied skew by considering a stochastic volatility model with dynamics given by

$$\begin{aligned}\frac{dS_t}{S_t} &= \sqrt{v(X(t))}dZ(t) \\ dX(t) &= \mu(X(t))dt + \sigma(X(t))dW_X(t)\end{aligned}$$

with $< dZ, dW_X >_t = \rho dt$. For instance, Hull and white [1987] considered the instantaneous variance to be given by $v(X(t)) = X(t)$, and they assumed a log-normal volatility process with dynamics

$$dX(t) = bX(t)dt + \xi X(t)dW_X(t)$$

where $\xi \geq 0$ is the volatility of the instantaneous variance. They obtained semi closed-form solution in the special case where $< dZ, dW_X >_t = 0$. Heston [1993] assumed the variance to follow a square-root process

$$dX(t) = \big(a - bX(t)\big)dt + \xi\sqrt{X(t)}dW_X(t)$$

where $V_\infty = \frac{a}{b}$ is the long term variance level and b is the speed of mean reversion of the variance. However, due to the nature of the squared Bessel process, the variance can reach zero value unless we apply the constraint

$$bV_\infty > \frac{1}{2}\xi^2$$

Stein et al. [1991] let the volatility process $Y_t = \sqrt{X_t}$ be an arithmetic Ornstein-Uhlenbeck process

$$dY(t) = \big(a - bY(t)\big)dt + \sigma dW_Y(t)$$

and derived explicit closed-form solution to the price of a call option. Even though the volatility Y_t can become negative, it is used in a squared fashion so that this formulation is equivalent to putting a reflecting barrier at $Y_t = 0$ in the volatility process. Further, for a wide range of relevant parameter values, the probability of actually reaching the point $Y_t = 0$ is very small.
In a qudratic volatility model, we let the process X_t be the OU-process above and define the variance process as $Z_t = X_t^2$ with dynamics given by

$$dZ(t) = \big(\sigma^2 + 2aX_t - 2bZ(t)\big)dt + 2\sigma X_t dW_X(t)$$

In any case, we then integrate the stock price SDE in the range $[t,T]$ and take the exponential, getting the solution

$$S_T = e^{-\frac{1}{2}\int_t^T v(X(s))ds + \int_t^T \sqrt{v(X(s))}dZ_X(s)}$$

The European call price $C(t;K,T)$ seen at time t with a strike K and maturity T satisfy $C(t;K,T) = P(t,T)E^Q[(S_T - K)^+|\mathcal{F}_t]$.
Following Hull et al. [1987], we let $I^2(t,T)$ be the mean variance over the life of the option price defined as

$$I^2(t,T) = \frac{1}{T-t}\int_t^T V(X_s)ds$$

where $V(X_t)$ is variance process, and such that the distribution function of S_T can be written as

$$p(S_T|V(X_t)) = \int g(S_T|I^2(t,T))h(I^2(t,T)|V(X_t))dI^2(t,T)$$

We can then condition the expectation with respect to the mean variance $I^2(t,T)$ and get the option price

$$C(t;K,T) = P(t,T)E^Q\big[E[(S(T)-K)^+]|I^2(t,T)\big]$$

Hence, conditional on the mean variance $I^2(t,T)$ the process S_t is lognormally distributed and we can re-write the option price as

$$\begin{aligned} C(t;K,T) &= P(t,T)E^Q\big[E[\big(S_T - K\big)^+]|I^2(t,T)\big] \\ &= \int C_{BS}\big(S_T, K, I^2(t,T)\big)h(I^2(t,T)|V(X_t))dI^2(t,T) \end{aligned}$$

where $C_{BS}(S_T, K, I^2(t,T))$ with $\sigma_{BS}^2 = I^2(t,T)$ is the Black-Scholes formula, and $h(\bullet)$ is the conditional density of the mean variance. This equation is always true when the stock price and the volatility are instantaneously uncorrelated. In that setting, the option price is the BS-price integrated over the distribution of the mean variance. In the spirit of Hull et al., we could calculate the moments of $I^2(t,T)$ and then expand the price $C(t;K,T;I^2(t,T))$ in a Taylor series about its mean $I_0^2(t,T) = E[I^2(t,T)]$, getting

$$\begin{aligned} f(S_t, X_t) &= C_{BS}(S_T, K, I_0^2(t,T)) + \alpha(K,T) \qquad (3.2.13) \\ \alpha(K,T) &= \frac{1}{2}\frac{\partial^2 C_{BS}}{\partial(I^2(t,T))^2}|_{I_0^2(t,T)} Var(I^2(t,T)) + \frac{1}{6}\frac{\partial^3 C_{BS}}{\partial(I^2(t,T))^3}|_{I_0^2(t,T)} Skew(I^2(t,T)) + ... \end{aligned}$$

where $Var(I^2(t,T))$ and $Skew(I^2(t,T))$ are the second and third moments of the mean variance $I^2(t,T)$. The series converges very quickly for small values of $k=\xi^2(T-t)$. That is, if the variance of X_t is small the distribution of the process $\log\frac{S_T}{S_0}$ is close to a normal distribution and few extra moments are necessary.
Alternatively, to price exotic equity products within the Yield Factor approach we need to calculate the characteristic function $\phi_Y(u,T)$. The process of interest to us is $s(t,T,\nu)=\ln\frac{S_T}{S_t}$

$$s(t,T,\nu)=-\frac{1}{2}\nu^2\int_t^T X(s)ds+\nu\int_t^T\sqrt{X(s)}dZ(s)$$

In that case the measurable function R becomes

$$R(X_t)=\overline{\lambda}X(t)$$

where $\overline{\lambda}=\lambda\nu^2$.

3.2.2.3 Adding Jumps to stochastic volatility

So far we have considered stochastic volatility models with a linear combination of independent Brownian motions. We know that Brownian motions have symmetric increments, while the risk-neutral distribution is asymmetric. Therefore, in such models the implied volatility skew is completely determined by the correlation of the Brownian increments. As an example, in a bivariate diffusion model when the correlation is set to zero that model can only produce a symetric implied volatility smile. One way to go around and to recover the market implied skew is to introduce jumps in these models. For simplicity we add the Merton dynamics to the present diffusion, getting

$$\frac{dS_t}{S_t}=(r_t-\lambda m)dt+\sqrt{v(X(t))}dZ(t)+(J-1)dN_t$$

where $J=e^{\mu+\gamma N(0,1)}$, $m=E(J)-1=e^{\mu+\frac{1}{2}\gamma^2}$ and so we have

$$dY_t=d\log(S_t)=(r_t-\lambda m-\frac{1}{2}v(X(t)))dt+\sqrt{v(X(t))}dZ(t)+\log(J)dN_t$$

We can now compute the characteristic function of the log of the forward price as

$$\phi_T(u)=E[e^{ius_\tau}]=E[e^{iu\int_t^T(r_s-\lambda m-\frac{1}{2}v(X(t)))ds+iu\int_t^T\sqrt{v(X(s))}dZ(s)+iu\sum_{N_t}^{N_T}\log(J)}]$$

We first condition the expectation in the characteristic function with respect to the variance $V(X(t))$ and the Jump, getting

$$\begin{aligned}\phi_T(u) &= E\big[E[e^{iu\int_t^T(r_s-\lambda m-\frac{1}{2}v(X(t)))ds+iu\int_t^T\sqrt{v(X(s))}dZ(s)+iu\sum_{N_t}^{N_T}\log(J)}]|V(X(t)),N_{T-t}\big]\\ &= E[e^{iu\int_t^T(r_s-\lambda m-\frac{1}{2}v(X(t)))ds-\frac{1}{2}u^2\int_t^T v(X(s))ds+iu\sum_{N_t}^{N_T}\log(J)}]\end{aligned}$$

Then we condition the expectation in the characteristic function only with respect to the variance $V(X(t))$ and get

$$\begin{aligned}\phi_T(u) &= E\big[e^{iu\int_t^T(r_s-\lambda m-\frac{1}{2}v(X(t)))ds-\frac{1}{2}u^2\int_t^T v(X(s))ds+\lambda\int_t^T[e^{\mu u+\frac{1}{2}u^2\gamma^2}-1]ds}\big]\\ &= E[e^{iu\int_t^T(r_s-\frac{1}{2}v(X(t)))ds-\frac{1}{2}u^2\int_t^T v(X(s))ds}]E[e^{-iu\int_t^T\lambda mds+\lambda\int_t^T\left(e^{\mu u+\frac{1}{2}u^2\gamma^2}-1\right)ds}]\end{aligned}$$

We recognise that $E[e^{iu\int_t^T(r_s-\frac{1}{2}v(X(t)))ds-\frac{1}{2}u^2\int_t^T v(X(s))ds}]$ can be written as $e^{A(\tau)+B(\tau)x}$ where A and B are solutions of the Ricatti equations from above. Let $A_0(\tau)=A(\tau)-iu\int_t^T\lambda mds+\lambda\int_t^T[e^{\mu u+\frac{1}{2}u^2\gamma^2}]$, and then we have

$$\phi_T(u) = e^{A_0(\tau)+B(\tau)x}$$

with the Ricatti equation

$$\dot{A}_0(\tau) = \dot{A}(\tau) - iu\lambda m + \lambda(e^{\mu u + \frac{1}{2}u^2\gamma^2} - 1)$$

and the same boundary condition than before $A_0(0) = 0$.

3.3 Option pricing approximation

3.3.1 Introduction

Under general Markov processes for the model parameters, the conditional probabilities of the Digital Bond and Digital Share (see Section (1.5.1)) are difficult to solve analytically under any probability measure, and numerical tools must be used. In the special case of constant spot rate, repo rate and uncorrelated stochastic volatility Hull and white [1987] obtained the pricing formula for a call option

$$E^Q\left[C_{BS}(t,x;I(t,T))|\mathcal{F}_t\right] , I(t,T) = \sqrt{\frac{1}{T-t}\int_t^T \sigma_s^2 ds}$$

and derived, in the case of correlated volatility, an approximation from a second-order Taylor series expansion around a constant volatility specification (see Equation (3.2.13)). Closed-form solutions with correlated stochastic volatility were found (see Heston [1993]) but not expressed in terms of the Black-Scholes formula. More recently, the derivative price with general volatility process was decomposed into combination of Black-Scholes terms by means of the Ito's lemma. For example, Alos [2004] and Alos et al. [2007] introduced Malliavin calculus in the derivation of option prices, obtaining approximations to the call prices in both Heston [1993] and Hull et al. [1987] models. They derived four terms identifying the expected future volatility, the correlation between the volatility and the stock price, the market price of volatility risk and the quadratic variation of the expected volatility process. Assuming that both the market price of volatility risk and the quadratic variation of the expected volatility process are small enough, they obtained a first order approximation to option prices.

We saw in Section (1.2.3) that when the stock price is log-normally distributed with constant model parameters, the European call prices satisfy $C(t,x) = C_{BS}(t,x,K,T)$ for $t \in [0,T]$. We also know that when the model parameters are general Markov processes we can no-longer use the Black-Scholes Equation (1.2.6). However, we still want to express the solution to the European call options in terms of the Black-Scholes formula, which can then be seen as a function of the stochastic spot rate, repo rate and volatility. Based on the work of Alos [2004] and Alos et al. [2007], Bloch [2010a] have described a general pricing approximation technique for call options in a jump-diffusion model with stochastic interest rates and stochastic repo rates. At maturity, the European option prices must satisfy

$$C(T,x) = C_{BS}\Big(t,x,K,T;\sqrt{\frac{1}{T-t}\int_t^T \sigma_s^2 ds},\sqrt{\frac{1}{T-t}\int_t^T r_s ds},\sqrt{\frac{1}{T-t}\int_t^T q_s^2 ds}\Big)\Big|_{t=T}$$

and since $e^{-\int_0^t r_s ds}C(t,x)$ is a $\mathbb{Q}$-martingale, we get

$$\begin{aligned} & e^{-\int_0^t r_s ds}C(t,x) = E^Q[e^{-\int_0^T r_s ds}C(T,x)|\mathcal{F}_t] \\ = \; & E^Q[e^{-\int_0^T r_s ds}C_{BS}\Big(t,x,K,T;\sqrt{\frac{1}{T-t}\int_t^T \sigma_s^2 ds},\sqrt{\frac{1}{T-t}\int_t^T r_s ds},\sqrt{\frac{1}{T-t}\int_t^T q_s^2 ds}\Big)\Big|_{t=T}|\mathcal{F}_t] \end{aligned}$$

Since the discounted price $\overline{C}(t,x)=\frac{C(t,x)}{P(t,T)}$ is a $\mathbb{Q}^T$ martingale, we can take the zero-coupon bond price as numeraire and use the change of measure formula to neutralise the interest rate movements getting

$$\overline{C}(t,S_t)=E^{Q^T}[h(S_T)|\mathcal{F}_t]$$

where the dynamics of the processes involved are now expressed under the forward measure. That is, the spot rate is now implicit and the underlying is the forward price $F(t,T)$ with its associated volatility σ_F and the modified repo factor q_F expressed under the forward measure. When the model parameters are constants, this is equivalent to using the Black formula where we know that we only need to calculate the forward price and its associated volatility. Therefore, pursuing our previous approach, we are going to express the price of European call options with stochastic model parameters as a combination of Black formula. In that case, we let

$$I(t,T)=\sqrt{\frac{1}{T-t}\int_t^T\sigma_F^2(s)ds}$$

be the future average volatility (and $I^2(t,T)$ is the mean variance) and

$$D(t,T)=\frac{1}{T-t}\int_t^T q_F(s)ds$$

be the future average dividend, and we obtain the discounted call price as

$$\overline{C}(t,S_T)=E^{Q^T}[\overline{C}(T,S_T)|\mathcal{F}_t]=E^{Q^T}[C_B\Big(T,F(T,T),K,T;I(T,T),D(T,T)\Big)|\mathcal{F}_t]$$

Using the linearity of both the Black formula and the expected value, we get

$$\begin{aligned}\overline{C}(t,S_T) &= E^{Q^T}[F(T,T)N(d_1(T-t,F(T,T),K;I(T,T),D(T,T)))|\mathcal{F}_t]\\ &- KE^{Q^T}[N(d_2(T-t,F(T,T),K;I(T,T),D(T,T)))|\mathcal{F}_t]\end{aligned}$$

which we can express in terms of the Black Digital Bond $D_B(S,t,T;\xi)$ and Black Digital Share $D_S(S,t,T;\xi)$ as

$$\overline{C}(t,S_T)=E^{Q^T}[S(F,T,T;\xi;I(T,T),D(T,T))|\mathcal{F}_t]-KE^{Q^T}[D(F,T,T;\xi;I(T,T),D(T,T))|\mathcal{F}_t]$$

We can therefore approximate the Digital Bond and the Digital Share and combine them to price other path-independent contingent claims.

3.3.2 Approximating the call price

We can summaries the approximation method proposed by Alos [2004], and improved over time, by considering as an example a jump-diffusion process X_t valued in $\mathbb{R}^2$ where $Y_T(t)$ is the logarithm of the forward price seen at time t with fixed maturity T, and V_t is its associated instantaneous variance. We then augment the size of our process X_t with the future average volatility $I(t,T)$, getting the augmented process $\mathcal{X}(t,T)=[Y_T(t),V_t,I(t,T)]$. In this framework the approximated call price under the forward probability measure $\mathbb{Q}^T$ is

$$\begin{aligned}\frac{C(t,K)}{P(t,T)} &= E^{Q^T}[C_B(t,Y_T(t),I(t,T))|\mathcal{F}_t]+\frac{1}{2}E^{Q^T}[\int_t^T H(s,Y_T(s),I(s,T))\Lambda_s ds|\mathcal{F}_t]\\ &- \lambda(\theta_Y(1)-1)E^{Q^T}[\int_t^T\partial_2 C_B(s,Y_T(s),I(s,T))ds|\mathcal{F}_t]\\ &+ E^{Q^T}[\int_t^T\lambda\int_{\mathbb{R}^2}\big(C_B(s,Y(s_-,T)+z,I(s,T))-C_B(s,Y(s_-,T),I(s,T))\big)F_Y(\omega,dz)ds|\mathcal{F}_t]\end{aligned}$$

where

- $C_B(t,x,\sigma)$ is the Black formula,
- $H(t,x,\sigma)=\big(\frac{\partial^3}{\partial x^3}-\frac{\partial^2}{\partial x^2}\big)C_B(t,x;\sigma)$, and
- $\Lambda_s=\big(\int_s^T D_s^W V_r dr\big)\sqrt{V_s}$.

Following Hull and White [1987], we can approximate this price by approximating the future average volatility with its expectation $I_0(t,T)=E^{Q^T}[I(t,T)|\mathcal{F}_t]$. So the approximated call price becomes

$$\begin{aligned}\frac{C(t,K)}{P(t,T)} &= C_B(t,Y_T(t),I_0(t,T))+\frac{1}{2}H(t,Y_T(t),I_0(t,T))E^{Q^T}[\int_t^T\Lambda_s ds|\mathcal{F}_t]\\ &- \lambda(\theta_Y(1)-1)\int_t^T\partial_2 C_B(s,Y_T(s),I_0(s,T))ds\\ &+ \int_t^T\lambda\int_{\mathbb{R}^2}\big(C_B(s,Y(s_-,T)+z,I_0(s,T))-C_B(s,Y(s_-,T),I_0(s,T))\big)F_Y(\omega,dz)ds\end{aligned}$$

The computation of the call option price is now as simple as the Black formula.
Assuming deterministic rates, Alos et al. [2007] tested, within the Heston model, the quality of their approximation against the original implementation given by Heston. They chose a very short maturity of $T=0.1$ where skew is most pronounced and found fairly accurate results. They also found that the larger part of the error was produced by replacing $E[C_{BS}(t,T;I(t,T))|\mathcal{F}_t]$ with the approximation $C_{BS}(t,T;I_0(t,T))$. It is well known that this approximation corresponds to the first term of the Taylor expansion of $E[C_{BS}(t,T;I(t,T))|\mathcal{F}_t]$ around $I_0(t,T)$ (see Equation (3.2.13)). So, we can improve the accuracy of the approximation of the call option price by improving the computation of $E[C_{BS}(t,T;I(t,T))|\mathcal{F}_t]$ by adding extra terms from the Taylor expansion. Doing so, we get

$$E[C_{BS}(t,T;I(t,T))|\mathcal{F}_t]=C_{BS}(t,T;I_0(t,T))+\frac{1}{2}var(I^2(t,T))\frac{\partial^2}{\partial(I^2(t,T))^2}C_{BS}(t,T;I(t,T))\Big|_{I_0(t,T)} \tag{3.3.14}$$

It is not difficult to calculate explicitly the derivative expression of the second term of the expansion as

$$\frac{\partial^2}{\partial(I^2(t,T))^2}C_{BS}(t,T;(t,T))=\frac{vega}{4(I^2(t,T))^{3/2}}\big(\frac{(\log(m))^2}{I^2(t,T)(T-t)}-\frac{I^2(t,T)(T-t)}{4}-1\big)$$

where $m=\frac{S_t}{Ke^{-r(T-t)}}=\eta$ is the forward moneyness and $vega=\frac{\partial}{\partial\sigma}C_{BS}(t,T;\sigma)=S_tN^{'}(d_1)\sqrt{T-t}$ is the BS-vega. To get an explicit solution to the price expansion we are left with calculating explicitly the term $var(I(t,T))$. In most cases, we will use the results of Dufresne [2001] where he found an explicit formula via a recursive method for the calculation of any moment of an integrated square root process which is extremely fast, especially for higher moments. However, in the Quadratic model the variance does not follow a square root process as the drift is modified with an Ornstein-Uhlenbeck process. So, in order to compute the variance of that integrated process we solve a series of ordinary differential equations (see Bloch [2007]). We can now write a better approximation to the price of a call option in a jump-diffusion model, which we can solve analytically. The pricing equation is given as follow:

$$\frac{C(t,K)}{P(t,T)}=C_B(t,Y_T(t),I_0(t,T))+\alpha(K,T) \tag{3.3.15}$$

where

$$\begin{aligned}\alpha(K,T) &= \frac{1}{2}var(I^2(t,T))\frac{\partial^2}{\partial(I^2(t,T))^2}C_B(t,T;I(t,T))\Big|_{I(t,T)=I_0(t,T)} \\ &+ \frac{1}{2}H(t,Y_T(t),I_0(t,T))E^{Q^T}[\int_t^T \Lambda_s ds|\mathcal{F}_t] - \lambda(\theta_Y(1)-1)\int_t^T \partial_2 C_B(s,Y_T(s),I_0(s,T))ds \\ &+ \int_t^T \lambda\int_{\mathbb{R}^2}\big(C_B(s,Y(s_-,T)+z,I_0(s,T)) - C_B(s,Y(s_-,T),I_0(s,T))\big)F_Y(\omega,dz)ds\end{aligned}$$

3.3.2.1 The Heston model revisited

The dynamics of the stock price $(S_t)_{t\geq 0}$ in the Heston model (see Heston [1993]) is given by

$$\begin{aligned}dS_t &= S_t\big(rdt + \sqrt{V_t}dW(t)\big) \\ dV_t &= \kappa(\theta - V_t)dt + \nu\sqrt{V_t}dZ(t)\end{aligned}$$

with $< dW, dZ >_t = \rho dt$, and $\kappa > 0$, $\theta > 0$ and $\nu > 0$. Alos et al. [2007] give a proof of the Malliavin differentiability of the Heston volatility as well as its square root process and gives explicit expression for their derivatives. The Malliavin derivative of the instantaneous variance is

$$D_r V_t = \nu e^{\int_r^t -\frac{\kappa}{2} - (\frac{\kappa\theta}{2} - \frac{\nu^2}{8})\frac{1}{V_t}ds}\sqrt{V_t} \tag{3.3.16}$$

Then, the pricing of a call price is considered. Letting $I^2(t,T) = \frac{1}{T-t}\int_t^T V_s ds$ be the square of average future volatility starting at time t and maturing at time T, the call price at maturity T is given by $C_{BS}(T,T;I(T,T)) = C(T,T)$. Since the discounted call price must be a $\mathbb{Q}$-martingale, we get

$$e^{-rt}C(t,T) = E[e^{-rT}C(T,T)|\mathcal{F}_t] = E[e^{-rT}C_{BS}(T,T;I(T,T))|\mathcal{F}_t]$$

We then apply the modified Ito's formula to the process $e^{-rt}C_{BS}(t,T;I(t,T))$, getting

$$C(t,T) = E[C_{BS}(t,T;I(t,T))|\mathcal{F}_t] + \frac{1}{2}E\big[\int_t^T e^{-r(s-t)}H(s,X_s,I(s,T))dU_s|\mathcal{F}_t\big]$$

where

$$U_t = \int_t^T \Big(\int_s^T D_s^W V_r dr\Big)\sqrt{V_s}ds$$

and

$$H(t,X_t,I(t,T)) = \Big(\frac{\partial^3}{\partial x^3} - \frac{\partial^2}{\partial x^2}\Big)C_{BS}(t,T;I(t,T))$$

Note that Alos et al. expanded the derivative with respect to the process U_t in the integral and approximated the option formula by considering only a first order approximation to the option price.

$$E[C_{BS}(t,T;I(t,T))|\mathcal{F}_t] + \frac{1}{2}E[H(t,X_t,I(t,T))U_t|\mathcal{F}_t]$$

Obviously, the first term of this approximation corresponds to the non-correlated case which was approximated by Hull et al. [1987] by replacing $I(t,T)$ with $I_0(t,T) = E[I(t,T)|\mathcal{F}_t]$. In the same spirit they used this approximation in the calculation of the function H, so that the approximated pricing formula becomes

$$C(t,T)=C_{BS}(t,T;I_0(t,T))+\frac{1}{2}H(t,X_t,I_0(t,T))E[U_t|\mathcal{F}_t]$$

Now, denoting $\Lambda_s=\big(\int_s^T D_s^W V_r dr\big)\sqrt{V_s}$, the call price becomes

$$C(t,T)=C_{BS}(t,T;I_0(t,T))+\frac{\rho}{2}H(t,X_t,I_0(t,T))E[\int_t^T \Lambda_s ds|\mathcal{F}_t]$$

So, in order to evaluate the option price we need to compute two quantities, that is, $I_0(t,T)$ and $E[\int_t^T \Lambda_s ds|\mathcal{F}_t]$. In the Heston model, the mean variance $I_0^2(0,T)$ seen at time $t_0=0$ is given by

$$I_0^2(0,T)=\frac{1}{T}\int_0^T E[V_s]ds=\theta+(V_0-\theta)\frac{1-e^{-\kappa T}}{\kappa T}$$

and the total variance is $\omega(T)=\theta T+(V_0-\theta)\frac{1-e^{-\kappa T}}{\kappa}$ with derivative with respect to maturity being

$$\partial_T\omega(T)=V_0+\theta(1-e^{-\kappa T})$$

which is always positive. The second term $E[\int_t^T \Lambda_s ds|\mathcal{F}_t]$ is by definition

$$E[\int_t^T \Lambda_s ds|\mathcal{F}_t]=E\Big[\int_t^T\big(\int_s^T E[D_s^W V_r|\mathcal{F}_s]dr\big)\sqrt{V_s}ds\Big]$$

We know that when $2\kappa\theta\geq\nu^2$, then

$$E[D_s^W V_r|\mathcal{F}_s]=\nu e^{-\kappa(r-s)}\sqrt{V_s}$$

Replacing in the equation, the integrand becomes $\nu\int_s^T e^{-\kappa(r-s)}drV_s$, and taking the expectation and integrating, we get

$$E[\int_0^T \Lambda_s ds]=\frac{\nu}{\kappa^2}\big[\theta(\kappa T-2)+V_0+e^{-\kappa T}(\kappa T(\theta-V_0)+2\theta-V_0)\big]$$

In the simple call option price we can derive an explicit solution for the function H, getting

$$\frac{1}{2}H(0,x,\sigma)=\frac{e^x}{2\sigma\sqrt{T}}N^{'}(d_1)(1-\frac{d_1}{\sigma\sqrt{T}})$$

where $d_1=\frac{x-\log K+rT}{\sigma\sqrt{T}}+\frac{1}{2}\sigma\sqrt{T}$.
We saw in Section (3.3.2) that we could improve the accuracy of the approximation of the call option price by improving the computation of $E[C_{BS}(t,T;I(t,T))|\mathcal{F}_t]$ by adding extra terms from the Taylor expansion (see Equation (3.3.14)). To get an explicit solution to the expansion we are left with calculating explicitly the term $var(I(t,T)$. We will use the results given by Dufresne in [2001] where he found an explicit formula via a recursive method for the calculation of any moment of an integrated square root process which is extremely fast especially for higher moments. We can now write a better approximation to the price of a call option in the Heston model, which we solve analytically. We get

$$\begin{aligned} C(t,T) &= C_{BS}(t,T;I_0(t,T))+\frac{1}{2}var(I^2(t,T))\frac{\partial^2}{\partial(I^2(t,T))^2}C_{BS}(t,T;I(t,T))\Big|_{I(t,T)=I_0(t,T)} \\ &+ \frac{\rho}{2}H(t,X_t,I_0(t,T))E[\int_t^T \Lambda_s ds|\mathcal{F}_t] \end{aligned} \tag{3.3.17}$$

Knowing the price of the call option in a fast and robust way we can then invert the Black and Scholes formula to recover the equivalent implied volatility surface. Using the implied volatility surface applied to the Heston model, we get

$$
\begin{aligned}
C_{BS}^{-1}(C(t,T)) &= I_0(t,T) + \frac{1}{2} var(I^2(t,T)) \frac{1}{4(I_0^2(t,T))^{3/2}} \big(\frac{(\log(m))^2}{I_0^2(t,T)(T-t)} - \frac{I_0^2(t,T)(T-t)}{4} - 1 \big) \\
&+ \frac{\rho}{2} \frac{1}{I_0(t,T)\sqrt{T-t}} (1 - \frac{d_1}{I_0(t,T)\sqrt{T-t}}) E[\int_t^T \Lambda_s ds | \mathcal{F}_t]
\end{aligned}
$$

Chapter 4

The importance of smile dynamics

4.1 The need to understand the smile dynamics

We saw in Section (2.2) that financial time series exhibit non-Gaussian response time distribution leading to more skewness and kurtosis than it is consistent with the Geometric Brownian Motion model of Black-Scholes. To the extent that the world deviates from the BS assumptions of constant volatility and a lognormal distribution to price changes, the BS model will be biased in certain, often predictable ways (see Hull and White [1987]). Various models have been proposed to describe the dynamics of returns and their deviation from lognormality, among which some of them were discussed in Section (3). However, option pricing is not just about matching vanilla prices, but it is also about getting a model capable of generating all the types of volatility regime observed in the market. Understanding the dynamics of the volatility surface is vital for two reasons

1. to get consistent pricing of exotic options,
2. to choose accordingly the delta that we should be using when hedging vanilla options.

Focusing on the latter, we will show that smile dynamics and hedging rationale are inter-related. The problem being how to decide on a particular smile dynamics given that this information is not contained in vanilla prices. We can either infer the smile dynamics from the prices of exotic options, such as forward start options, or we can estimate them statistically. Alternatively, we can simply take a view. In the case of systematic trading of vanilla options we are less concerned with valuing exotic options and more inclined in devising statistical dynamics to the volatility surface.

4.1.1 Some standard evolution of the implied volatility surface

In the financial literature, some authors analysed the properties of a pricing model to deduce the dynamics of the smile inferred by that model, while others analysed the market and devised regimes of volatility. We briefly present both approaches.

4.1.1.1 The local volatility and its associated smile dynamics

In order to understand the smile dynamics generated by the deterministic local volatility (DLV) model, Hagan et al. [1999] considered the special case where the local volatility is only a function of the forward price F. In that setting, the forward price satisfies the dynamics

$$dF = \sigma_L(F)FdW \, , F(0) = f$$

Using singular perturbation methods, they showed that European call and put prices were given by the Black-Scholes formula with implied volatility

$$\sigma_B(K,f) = \sigma_L(f_{av})\Big\{1 + \frac{1}{24}\frac{\sigma_L^{''}(f_{av})}{\sigma_L(f_{av})}(f-K)^2 + ...\Big\}$$

with $f_{av} = \frac{1}{2}(f+K)$, so that $K = 2F - f$. The first term dominates the solution and the second term provides a much smaller correction. Analysing the dynamics of the local volatility, Hagan et al. [2002] calibrated the model to market prices, obtaining

$$\sigma_L(F) = \sigma_B^0(2F - f_0)\{1 + ...\}$$

where $\sigma_B^0(K)$ is the implied volatility seen at time t_0 with forward price f_0. Assuming that the forward price changes from f_0 to f, the new implied volatility becomes

$$\sigma_B(K,f) = \sigma_B^0(K + f - f_0)\{1 + ...\}$$

Thus, if the forward price increases from f_0 to f, the IV curve moves to the left, and if it decreases, the IV curve moves to the right. In presence of negative skew, local volatility models predict that the skew moves in the opposit direction as the price of the underlying asset. They illustrated their findings by considering the quadratic smile

$$\sigma_B^0(K) = a + b(K - f_0)^2$$

obtaining the local volatility

$$\sigma_L(F) = a + 3b(F - f_0)^2 + ...$$

and they showed that when the forward moves from f_0 to f, then the implied volatility becomes

$$\sigma_B(K,f) = a + b\big[K - (\frac{3}{2}f_0 - \frac{1}{2}f)\big]^2 + \frac{3}{4}b(f - f_0)^2 + ...$$

where $\partial_f \sigma_B(K,f) = b[K + (2f - 3f_0)]$. The implied volatility shows that the curve also shift upward regardless of whether f increases or decreases. This is only true because the smile is quadratic, which is rarely the case in equity.

4.1.1.2 The regimes of volatility

We present some recently proposed deterministic implied volatility models, defined by assuming that either the per-delta or the per-strike implied volatility surface has a deterministic evolution. We then give some rules of thumb used by practitioners to compute their plain-vanilla prices. To do so, we look at some standard evolutions of the implied volatility surface denoted by $\Sigma(t,S_t;K,T)$ and first described by Derman [1999] in "Regimes of Volatility" and extended by various authors such as Daglish, Hull and Suo [2006]. It provides an indication of some possible behaviour for the smile that one might expect, even though they are not all **arbitrage-free**.

1. sticky delta : the strike of the option is rescaled according to how the current spot evolved with respect to the spot at inception
   $\Sigma(t,S_t;K,T) = \Sigma_0^{obs}(K\frac{S_0}{S_t}, T - t)$

2. absolute floating : the future implied volatility is obtained from the original smile surface by simply reducing time to maturity and linearly offsetting the strike by how much the spot has moved
   $\Sigma(t,S_t;K,T) = \Sigma_0^{obs}(K + (S_0 - S_t), T - t)$

3. absolute sticky (sticky strike) : any dependence on the current spot level or calendar time is ignored
   $\Sigma(t,S_t,K,T) = \Sigma_0^{obs}(K, T - t)$

where Σ^{obs} is the current smile. These regimes of volatility were slightly extended by Derman [2008], who for simplicity of exposition, considered the linear smile (only true when the strike is near the money)

$$\Sigma(t_0, S_0; K, T) = a - b(K - S_0) , b > 0$$

where a can be seen as the ATM-volatility, and studied its variation when the underying moves away from S_0.

1. sticky moneyness : The dynamics of option prices derive entirely from the moneyness

$$\Sigma(t, S_t; K, T) = a - b(\frac{K}{S_t} - 1)S_0 , b > 0$$

It shifts the skew as the stock price moves by adjusting for moneyness, assuming that the market mean reverts to the value a independently from market level. When the skew is nearly linear we get $\Sigma \approx \Sigma(S_t - K)$, so that the implied volatility rises when S rises. In the BS-delta (see Equation (B.1.3)), the dependence on K and S occurs only through the moneyness η, such that for ATM option we get $|\Delta| = \frac{1}{2}$. However, the delta depends on the scaled moneyness $\frac{1}{\sqrt{\omega(t)}} \log \eta$, which should be the fixed point. In presence of negative skew, the IV increases with increasing spot, leading to a greater delta than BS-delta. Stochastic volatility models and jump-diffusion models have this property (assuming the other variables do not change).

2. sticky implied tree : In the implied tree, local volatility increases as the underlying decreases. Assuming that the IV is approximately the linear average of local volatilities between the initial spot price and the strike price, we can extract the future local and implied volatilities from current implied volatilities. We get the linear approximation to the skew

$$\Sigma(t, S_t; K, T) = a - b(K + S_t - 2S_0) , b > 0$$

so that the implied volatilities decrease as K or S increases, leading to a smaller delta than BS-delta. Hence, when linearly approximating the local volatility we get $\Sigma \approx f(K + S_t)$ and volatilities are inversly correlated with the underlying spot price. The sticky implied tree regime corresponds to a fear of higher market volatility in the case of a fall of the spot price. Local volatility models have this property.

However, these simplified regimes of volatility have a couple of major drawbacks. First of all, they are not consistent with market observable data, and secondly the way they are used to price forward call options is not derived from the theory of No Arbitrage.

4.1.2 Characterising the smile dynamics

4.1.2.1 Space homogeneity versus space inhomogeneity

Even though the dynamics of implied volatility surface (IVS) are neither stationary nor Markovian, we saw in Section (4.1.1) that they could be classified in terms of regimes which might change over time. These regimes explain the evolution of the vanilla smiles in terms of the strike price and the underlying spot price. At the two extremes, the smile can either be expressed in terms of moneyness, or in terms of absolute value for the strike and spot prices. The former is generated by space homogeneous models, like the Heston model, and follow the sticky delta rule. The latter is generated by space inhomogeneous models, such as certain types of local volatility models, and follow the sticky implied tree rule. In between, practitioners developed a range of universal models, where the diffusion term is a mix of stochastic volatility and local volatility (see Lipton [2002], Blacher [2001]). The sticky delta rule leads to a delta greater than the BS-delta, meaning that the dynamics of the IV follow that of the underlying asset. In the sticky implied tree rule the delta is smaller than the BS-delta, meaning that the dynamics of the IVS is inversly related to the underlying asset. It is interesting to note that some authors justified their models by the type of dynamics they produce. For Blacher [2001], homogenous smile do not predict a systematic move of the relative smile when spot changes. Observing the equity index, the smile does not move coincidentally with the underlying asset, so that to modify the dynamics of the smile he had to re-introduce inhomogeneity in the stochastic model. On the contrary, for Hagan et al. [2002] local volatility models (LVMs) predict smile dynamics opposite of observed market behaviour. LVMs predict that the smile shifts to higer prices when the underlying asset decreases, and that the smile shifts to lower price when spot price increases. Observing interest rate smiles, where asset prices and market smiles move in the same direction, they devised the SABR model, which is space inomogeneous for $\beta \neq 1$ and allows to capture the correct dynamics of the smile (in the rates market). Thus, we see that the specific model chosen characterise a particular type of dynamics for the IVS, which in turn impose a particular delta hedging ratio.

4.1.2.2 A measure of smile dynamics

In space homogeneous models, Euler's theorem implies

$$C(K,T) = S\frac{\partial C}{\partial S} + K\frac{\partial C}{\partial K}$$

such that for a fixed smile surface, $\Delta = \frac{\partial C}{\partial S}$ is fixed. As a result, two homogeneous models will agree on the option delta, when calibrated to the same smile, even though they have different stochastic structures. However, space inhomogeneous models, with an explicit relation between the diffusion coefficient and the underlying, can yield a different delta. While Δ is the measure of smile dynamics in LVMs, it is not the case in models including other state variables. For example, in a stochastic volatility model where the volatility σ_t is correlated with the underlying, the volatility will move as well, such that the Vega, $\frac{\partial C}{\partial \sigma}$, and the Vanna, $\frac{\partial C}{\partial S \partial \sigma}$, will also capture part of the smile dynamics. That is, some of the vega risks are due to changes in the price of the underlying asset and should be hedged as part of the delta risks. A trader is interested in knowing the real time dynamics of option prices as he constanly need to re-evaluate his book as the spot moves and can not aford systematic recalibration. As explained by Ayache et al. [2004], using the chain rule in LVMs leads to real time delta, but it can not be done when the diffusion term is a stochactic function of the underlying (correlation in SVMs). This is related to the notion of market incompleteness, leading to optimal hedging in some sense (see Section (5.3)). A trader is interested in smile dynamics because he needs to know what the expected value of the vanilla prices will be, conditionally on the underlying trading at some level S for some future date t. That does not mean that he will hedge his positions accordingly.

4.1.3 The Greeks are model dependent

As explained in Section (4.1.1) different option pricing models imply different smile dynamics leading to different Greeks. Hence, getting the right delta for a given option is a complex problem as it is model dependent (see Benhamou [2001]). We are now going to discuss how the smile dynamics impact the computation of the Greeks.

4.1.3.1 The implied delta

We saw in Section (1.6.2) that when the implied volatility does not depend on the underlying asset S_t, but only on the strike and on time, the delta of the option satisfies

$$\Delta(t,S_t) = \partial_x C(t,S_t,K,T) = \partial_x C_{BS}(t,S_t,K,T)$$

which is the Black-Scholes delta. Derman [1999] called that regime of volatility the sticky strike regime. In the markets, this assumption is rarely satisfied, as one observe a change in the implied volatility surface when the spot price moves. Taking this change into consideration, and assuming that the volatility $\Sigma = \Sigma(S_t)$ depends on the spot price, the delta becomes

$$\Delta(t,S_t) = \partial_x C(t,S_t,K,T) = \partial_x C_{BS}(t,S_t,K,T;\Sigma) + \partial_\Sigma C_{BS}(t,S_t,K,T;\Sigma)\frac{\partial \Sigma(t,S_t,K,T)}{\partial x} \quad (4.1.1)$$

where $\frac{\partial \Sigma(t,S_t,K,T)}{\partial x}$ is the sensitivity of the implied volatility with respect to the spot price $S_t = x$, $\partial_\Sigma C_{BS}(t,S_t,K,T;\Sigma)$ is the Vega given in Equation (B.1.4), and $\partial_\Sigma C_{BS}(t,S_t,K,T;\Sigma)\frac{\partial \Sigma(t,S_t,K,T)}{\partial x}$ is called the shadow delta.

Remark 4.1.1 *The new delta will be either bigger or smaller than the Black-Scholes delta, depending on the shape of the implied volatility with repect to the spot price. If $\frac{\partial \Sigma}{\partial x} > 0$, the new delta will be bigger, but If $\frac{\partial \Sigma}{\partial x} < 0$, the new delta will be smaller.*

The difference between the two deltas is treated as a new source of risk, called the vega risk which can be hedged by making the portfolio vega-neutral. Hence, given a portfolio of European options, the sum of the derivatives of each option with respect to its implied volatility should be null, that is, $\sum_{k=1}^N \partial_\sigma C_{BS}(t,S_t)|_{\sigma=\Sigma_k} = 0$. To be fully hedged, we must also assume that the implied volatility $\Sigma(t,S_t;K,T)$ for different strikes and different maturities changes in the same way with respect to the underlying, that is, $\frac{\partial \Sigma}{\partial x}$ does not depend on K. Again, this assumption is rarely satisfied.
Another approach taking into consideration the shadow delta is to make an assumption on the dynamics of the implied volatility with respect to the underlying.

- For instance, Derman [1999] proposed the sticky delta regime also called the sticky moneyness regime. In that regime, it is assumed that the dynamics of the implied volatility depends on the ratio $m = \frac{K}{x}$ (moneyness) for $x = S_t$, but not on the two variables separately. Letting $\Sigma = \Sigma(\frac{K}{x})$ and using the chain rule, we get

  $$\frac{\partial}{\partial x}\Sigma(m) = \Sigma^{'}(m)\frac{dm}{dx}$$

  where $\Sigma^{'}(m) = \frac{\partial}{\partial m}\Sigma(m)$. Thus, the implied delta becomes

  $$\Delta(t,S_t) = \partial_x C(t,S_t,K,T) = \partial_x C_{BS}(t,S_t,K,T) - \partial_\Sigma C_{BS}(t,S_t,K,T;\Sigma)\frac{K}{S^2}\Sigma^{'}(m)$$

  Since in general the volatility decreases with increasing strikes and the Vega is positive, then $\Sigma^{'}(m) < 0$ and the sticky delta regime implies

  $$\partial_x C(t,S_t,K,T) > \partial_x C_{BS}(t,S_t,K,T)$$

  That is, the delta of the call option is greater than the Black-Scholes delta. For example, assuming a smile given by

  $$\Sigma(m) = a - bm \ , b > 0$$

  then $\frac{d}{dm}\Sigma(m) = -b$ so that $\frac{\partial}{\partial x}\Sigma(m) = b\frac{K}{x^2}$ (which is positive).

- Alternatively, considering the sticky implied tree regime, the implied volatility satisfies $\Sigma \approx f(K+x)$, such that $\frac{d}{dx}f(K+x) = f^{'}(K+x)$ and the delta becomes

  $$\Delta(t,S_t) = \partial_x C(t,S_t,K,T) = \partial_x C_{BS}(t,S_t,K,T) + \partial_\Sigma C_{BS}(t,S_t,K,T;\Sigma)f^{'}(K+x)$$

  Assuming the linear smile

  $$\Sigma(t,S_t;K,T) = a - b(K+S_t-2S_0) \ , b > 0$$

  we get $\Sigma^{'}(K+x) = -b$, such that the sticky implied tree regime implies

  $$\partial_x C(t,S_t,K,T) < \partial_x C_{BS}(t,S_t,K,T)$$

When delta-hedging a portfolio, the best hedge ratio is the one minimising the variance of the PnL. Comparing delta hedging between the Black-Scholes model and the local volatility model, Crepey [2004] showed that BS hedging strategy performed worst in fast selloffs or slow rise, while local volatility hedging strategy performed worst in slow selloffs and fast rises. The equity index markets being negatively skewed, Crepey argued that the BS hedging strategy is worse than the local volatility strategy.

4.1.3.2 The implied gamma

Differentiating the delta one more time with respect to the spot price, we get the gamma expressed in terms of the shadow gamma as follow

$$\begin{aligned}\Gamma(t,S_t) &= \partial_{xx}C(t,S_t,K,T) = \partial_{xx}C_{BS}(t,S_t,K,T) + 2\partial_{x\Sigma}C_{BS}(t,S_t,K,T;\Sigma)\frac{\partial\Sigma}{\partial x} \\ &+ \partial_{\Sigma\Sigma}C_{BS}(t,S_t,K,T;\Sigma)\big(\frac{\partial\Sigma}{\partial x}\big)^2 + \partial_{\Sigma}C_{BS}(t,S_t,K,T;\Sigma)\frac{\partial^2\Sigma}{\partial x^2}\end{aligned} \tag{4.1.2}$$

where $\partial_{x\Sigma}C_{BS}(t,S_t,K,T;\Sigma)$ is the Vanna given in Equation (B.1.7) and $\partial_{\Sigma\Sigma}C_{BS}(t,S_t,K,T;\Sigma)$ is the Volga given in Equation (B.1.8). Note, the gamma also depends on the second derivative of the implied volatility with respect to the spot price S_t. In the sticky delta regime, using the chain rule, we get

$$\frac{\partial^2}{\partial x^2}\Sigma(m) = \frac{\partial^2}{\partial m^2}\Sigma(m)\big(\frac{dm}{dx}\big)^2 + \frac{\partial}{\partial m}\Sigma(m)\frac{d^2m}{dx^2}$$

Applying that result to the example above, the second derivatives becomes $\frac{\partial^2}{\partial x^2}\Sigma(m) = -2b\frac{K}{x^3}$. However, in the simple example of the sticky tree, the second derivative is null.

4.1.3.3 Accounting for other risks

Note, we have only considered local volatility models, but in principle we could also account for jumps in the underlying price, transaction costs, etc. Market practice is to take a simple model and add some reserve for additional risk not accounted for by the model. This is part of model risk. If the trader has a view on the direction of the market (mean-reverting), he can formulate an optimisation problem and solve a specific Hamilton Jacobi Bellman equation. In case of a trending market, a shadow delta would suffice.

4.1.3.4 Sticky delta regime in the Black-Scholes model

Given the option price $C(t,K,T)$ at time t for a strike K and a maturity T, the Black-Scholes implied volatility $\Sigma(K,T)$ is obtained by inverting the Black-Scholes formula $C_{BS}^{-1}(C(t,K,T);K,T)$. Further, expressing the strike in terms of the forward price $K=\eta F(t,T)$, the call price is given by Equation (1.2.9). Hence, by varying the strike for a fixed level of spot price, we can express the implied volatility in terms of the forward moneyness. If we bump the initial spot price by $\alpha=\epsilon x$ where $\epsilon << 1$, that is, $x+\alpha$, the new forward moneyness becomes $\eta^{\epsilon} = \frac{1}{(1+\epsilon)}\frac{K}{x}e^{-(r-q)(T-t)} = \frac{1}{(1+\epsilon)}\eta$, implying that the strike is simply rescaled to $K\frac{1}{(1+\epsilon)}$ [1]. As a result, the new option price becomes

$$C_{TV}(t,x,\eta,T;\epsilon) = x(1+\epsilon)e^{-q(T-t)}\big(N(d_1(T-t,\eta^{\epsilon})) - \eta^{\epsilon}N(d_2(T-t,\eta^{\epsilon}))\big) = (1+\epsilon)C_{TV}(t,x,\eta^{\epsilon},T)$$

[1]
$$K\frac{x}{x+\alpha} = K\frac{x}{x(1+\epsilon)} = K\frac{1}{1+\epsilon}$$

and the new Black-Scholes implied volatility corresponds to a translation in the forward moneyness of the initial implied volatility. In fact, its dynamics follow the sticky delta regime of volatility described in Section (4.1.1.2) with strike $K\frac{S_t}{S_t+\alpha} = K\frac{1}{(1+\epsilon)}$.
Note, from the property of homogeneity of the BS-prices given in Equation (1.2.13) we can set $K^\epsilon = K(1+\epsilon)$, such that η^ϵ recovers the initial forward moneyness η, and the call price becomes

$$C_{TV}(t,x,\eta(1+\epsilon),T;\epsilon) = x(1+\epsilon)e^{-q(T-t)}\big(N(d_1(T-t,\eta)) - \eta N(d_2(T-t,\eta))\big) = (1+\epsilon)C_{TV}(t,x,\eta,T)$$

and the Black-Scholes volatility expressed in forward moneyness is unchanged.

4.1.4 Characterising the implied Greeks

In order to characterise the implied delta in Equation (4.1.1) and the implied gamma in Equation (4.1.2) we need define the variations of the volatility surface $\Sigma(t,S_t;K,T)$ with respect to the spot price S_t. That is, we need a model to describe how the IVS is modified when the underlying asset moves. One way forward is to consider the regimes of volatility given in Section (4.1.1), and relate these variations to the Skew and Curvature of the surface. Using the chain rule, we can relate the first derivative of the implied volatility with respect to the spot price S_t to that with respect to the Skew as follow

$$\frac{\partial\Sigma(t,S_t;K,T)}{\partial S_t} = k_S\frac{\partial\Sigma(t,S_t;K,T)}{\partial K} \tag{4.1.3}$$

where

$$k_S = \begin{cases} 0 \text{ if sticky strike} \\ 1 \text{ if sticky implied tree} \\ -\frac{K}{S_t} \text{ if sticky delta} \\ \frac{K}{S_t} \text{ if minimum variance} \end{cases}$$

where the minimum variance (MV) was introduced by Lee [2001]. Further, using the chain rule and making some approximations [2], we can relate the second derivative of the implied volatility with respect to the spot price S_t to that with respect to the Curvature as follow

$$\frac{\partial^2\Sigma(t,S_t;K,T)}{\partial S_t^2} \approx k_C\frac{\partial^2\Sigma(t,S_t;K,T)}{\partial K^2} \tag{4.1.4}$$

where

$$k_C = \begin{cases} 0 \text{ if sticky strike} \\ 1 \text{ if sticky implied tree} \\ (\frac{K}{S_t})^2 \text{ if sticky delta and minimum variance} \end{cases}$$

We can also introduce the forward moneyness $\eta_t = \frac{K}{F(t,T)}$, where $F(t,T) = S_te^{(r-q)(T-t)}$, and express the volatility in terms of the forward moneyness metric

$$\Theta(t,S_t;\eta_t,T) = \Sigma(t,S_t;\eta_tF(t,T),T)$$

In that setting, the price of a call option seen at time t with strike K and maturity T becomes

$$C_{BS}(t,S_t,K,T) = S_te^{-q(T-t)}\big(N(d_1(T-t,\eta_t)) - \eta_tN(d_2(T-t,\eta_t))\big)$$

[2] $\frac{d^2m}{dx^2} = 2\frac{K}{x^3} << 1$.

where

$$d_2(t,\eta_t) = \frac{1}{\Theta\sqrt{t}} \log \frac{1}{\eta_t} - \frac{1}{2}\Theta\sqrt{t} \text{ and } d_1(t,\eta_t) = d_2(t,\eta_t) + \Theta\sqrt{t}$$

Note, given the total variance $\omega(t) = \Theta^2 t$, we recover the scaled price in Equation (1.2.9). Thus, the delta in Equation (4.1.1) becomes

$$\Delta(t,S_t) = \partial_x C(t,S_t,\eta_t,T) = \partial_x C_{BS}(t,S_t,\eta_t,T) + \partial_\Theta C_{BS}(t,S_t,\eta_t,T;\Theta)\frac{\partial\Theta(t,S_t,\eta_t,T;\Theta)}{\partial x}$$

Using the chain rule, we have

$$\frac{\partial\Sigma(t,S_t;K,T)}{\partial K} = \frac{\partial\Theta(t,S_t;\eta_t,T)}{\partial\eta}\frac{d\eta_t}{dK} = \frac{1}{S_t}\frac{P(t,T)}{Re(t,T)}\frac{\partial\Theta(t,S_t;\eta_t,T)}{\partial\eta}$$

and

$$\frac{\partial\Sigma(t,S_t;K,T)}{\partial S_t} = \frac{\partial\Theta(t,S_t;\eta_t,T)}{\partial\eta}\frac{d\eta_t}{dS_t} = -\frac{K}{S_t^2}\frac{P(t,T)}{Re(t,T)}\frac{\partial\Theta(t,S_t;\eta_t,T)}{\partial\eta}$$

Since we have the relation $\partial_x\Sigma(t,S_t;K,T) = \partial_x\Theta(t,S_t;\eta_t,T)$, we get

$$\frac{\partial\Theta(t,S_t;\eta_t,T)}{\partial S_t} = k_S(\eta_t)\frac{\partial\Theta(t,S_t;\eta_t,T)}{\partial\eta}$$

where

$$k_S(\eta_t) = \begin{cases} 0 \text{ if sticky strike} \\ \frac{1}{S_t}\frac{P(t,T)}{Re(t,T)} = \frac{\eta_t}{K} \text{ if sticky implied tree} \\ -\frac{K}{S_t^2}\frac{P(t,T)}{Re(t,T)} = -\frac{\eta_t}{S_t} \text{ if sticky delta} \\ \frac{\eta_t}{S_t} \text{ if minimum variance} \end{cases}$$

That is, we move from k_S to $k_S(\eta_t)$ by a factor of $\frac{1}{S_t}\frac{P(t,T)}{Re(t,T)} = \frac{\eta_t}{K}$. Note, the minimum variance (MV) delta, given in the moneyness metric by $k_S(\eta_t) = \frac{\eta_t}{S_t}$, is the exact opposit of the term in the sticky delta regime (see Lee [2001]). In any case, to compute the adjusted delta we need to estimate the market skew by computing either $\frac{\partial\Sigma(t,S_t;K,T)}{\partial K}$ or $\frac{\partial\Theta(t,S_t;\eta_t,T)}{\partial\eta}$.
For example, Alexander et al. [2010] chose to fit the smile with the cubic polynomial

$$\Theta(t,S_t;\eta_t,T) = a_0(t) + a_1(t)\eta_t + a_2(t)\eta_t^2 + a_3(t)\eta_t^3$$

getting the skew

$$\frac{\partial\Theta(t,S_t;\eta_t,T)}{\partial\eta} = \widehat{a}_1(t) + 2\widehat{a}_2(t)\eta_t^2 + 3\widehat{a}_3(t)\eta_t^3$$

To account for the fact that the implied volatility is non-Markovian, Alexander et al. modified the IV sensitivity to the spot price, $\partial_x\Theta(t,S_t,\eta_t,T;\Theta)$, by letting the correction term $k_S(\eta_t)$ be dependent on the regime of volatility observed over a recent period of historical data. At each time t, they fixed $k_S(t,S_t;\eta_t,T)$, computed ex-post the delta hedging error on a sample of option prices, and chose the value for $k_S(t,S_t;\eta_t,T)$ minimising the resulting standard deviation. The delta obtained has a regime dependent (RD) smile adjustment. To improve the smile dynamics further, they added an extra term, independent from the moneyness, to the sensitivity of the IV with respect to the spot price, getting

$$\frac{\partial \Theta(t, S_t; \eta_t, T)}{\partial S_t} = \alpha(t, S_t; T) + k_S(\eta_t) \frac{\partial \Theta(t, S_t; \eta_t, T)}{\partial \eta}$$

Performing a linear regression of daily changes in ATM volatility on the daily log return of the underlying asset, they set $\widehat{\alpha}(t, S_t; T)$ equal to the estimated slope coefficient, divided by the spot price. The delta obtained has a regime-dependent shift (RS).

4.1.5 Using exotic options to devise the smile dynamics

4.1.5.1 The importance of forward starting options

The problem with pricing forward start options is that today's market prices do not provide us with the future anticipations of stock process evolution. Dupire [2003] reached the same conclusion when he said that a model must be calibrated to the initial volatility surface, but also to its evolution in time. This is because the implied volatility surface is not stationary and not Markovian. Basically, it means that if we denote by t_0 today's date, and t_1 and t_2 two future dates with $t_2 > t_1 > t_0$, then market prices do not directly give us the probability density function (*pdf*)

$$\phi(S_{t_2} \in dx | \mathcal{F}_{t_1})$$

where $\mathcal{F}_{t_1}$ can be considered all available information at time t_1. Yet, the knowledge of this function is of fundamental importance if we want to price forward start call options. This is the problem with Dupire's local volatility model, that is, while assuming the evolution of the volatility surface to be Markovian and considering an infinity of fixing dates [3], one can not control its evolution between the start date (previously denoted t_1) and the maturity of the forward start option, that is, its conditional density. However, when the number of fixing dates is finite, there is an infinity of conditional densities, which is not the case in Dupire's model. In fact, Rebonato [2002] explained that in a discret-time model if the no-arbitrage violation is to be allowed there is an infinity of solutions for the future conditional deterministic density, which means that the future is unknown and can not be derived from today's information. One solution is to add some structure by choosing a particular model with extra state variables other than the underlying spot price. These extra processes need to be calibrated to forward starting options, which are directly linked to the forward smiles (see Ayache et al. [2004]). This way, we let the market decide on the structure of the conditional distributions. Put another way, we let the market decide on the dynamics of the volatility surface. However, the market of forward start options is an OTC business which is model dependent and reflects the views of the parties but not the market as a whole. As a result, volatility trading is about taking a view on the smile dynamics that will prevail, where traders take a bet with or against the future smile. As explained by Derman [1999], depending on the regime you think the market is, you adjust accordingly your BS-hedge. An alternative approach is to satisfy this infinity of solutions (conditionals) by giving the forward smile a shape consistent with its historical evolution.

4.1.5.2 Pricing forward start option

Even though, in the local volatility model, the future implied volatility surface is assumed static (a function of today's one), we can provide the implied volatility surface with general dynamics. One way forward is to link the future implied volatility to a stochastic process. For instance, we can have the surface depends on the factor process $(X_t)_{t \geq 0}$, which is assumed Markov. Strong of this result, Bloch [2008] derived the consequences of such an hypothesis on the pricing of *European forward start call option*. Let's denote $C^F(t_0, t_1, k, T)$ the value at time t_0 of a forward start call option starting at time t_1 with risky strike kS_{t_1} and maturity T. The No Arbitrage conditions lead to

$$C^F(t_0, t_1, k, T) = e^{-r(T-t_0)} \mathbb{E}\left((S_T - kS_{t_1})^+ | \mathcal{F}_{t_0}\right)$$

[3] It is a continuous time model

if we look at forward start call option with *absolute* pay-off [4] $(S_T - kS_{t_1})^+$. By conditioning the forward start option with information available at time t_1 *i.e.* by the filtration $\mathcal{F}_{t_1}$ we get

$$\begin{aligned} C^F(t_0, t_1, k, T) &= e^{-r(T-t_0)}\mathbb{E}\left((S_T - kS_{t_1})^+ | \mathcal{F}_{t_0}\right) \\ &= e^{-r(T-t_0)}\mathbb{E}\left(\mathbb{E}\left((S_T - kS_{t_1})^+ | \mathcal{F}_{t_1}\right) | \mathcal{F}_{t_0}\right) \end{aligned}$$

Recall that the implied volatility for a given strike and maturity is the number that plugged in Black-Scholes formula fit the European call market prices. So, from now on we consider a general form of implied volatility surface (IVS) and assume that it is driven by a factor process, X, such as the one in Equation (3.2.10). Then, we get the following hypothesis

$$(\mathcal{H}) \quad \forall\, T \quad \forall t < T, \quad \forall\, K \quad \Sigma_t(K,T) = \Sigma(X_t; K, T) \tag{4.1.5}$$

Therefore, we can rewrite our assumption in terms of the Jump-Diffusion formula as

***Assumption* 4.1.1**

$$(\mathcal{H}) \quad \forall\, T \quad \forall t < T, \quad \forall\, K \quad C(t, S_t, K, T-t) = C_{JD}(t, X_t, K, T)^{5} \tag{4.1.6}$$

As there exists a one-to-one relation between the implied volatility surface Σ and the pdf ϕ with respect to the Lebesgue measure λ, defined as the function such that for any set $\mathcal{S}$ the probability that $X_{t_1} \in \mathcal{S}$ at t_1 given X_{t_0} at t_0 is

$$P(X_{t_1} \in \mathcal{S} | X_{t_0}) = \int_{\mathcal{S}} \phi(x, t_1, X_{t_0}, t_0)\lambda(dx)$$

The current factor evolution directly influences its future increment pdf, which means that economic agent anticipations are *dynamic*. Note that these dynamics neither appear in the Black-Scholes model nor in Dupire model, where agent anticipations are assumed to be static. Linking the factors and smile together is to say that the smile effect, which characterises agent beliefs for future evolution of the stock, is dynamically modified according to factors realisations. This is a way of modelising dynamic, *i.e.* rational anticipations.

A consequence of such an hypothesis (4.1.5) is that call prices are all determined by factor levels and consequently that the factor process X holds the whole market risk. Basically, the σ-algebra $\mathcal{F}_t$, which is *a priori* the σ-algebra generated by $\{\eta_u, u \leq t\}$ where $\eta_u = \{X_u \cup \bigcup_{K,T} C(u, S_u, K, T-u)\}$ coincides with $\sigma(X_u,\, u \leq t)$ for η_u is fully determined by X_u:

$$\mathcal{F}_t = \sigma\{\eta_u, u \leq t\}$$

Moreover the dependency of Σ_t on X_t conserves the *Markov property* for X. Therefore, considering the payoff of a call option in Section (1.7.1), the forward starting call option becomes

$$C^F(t_0, t_1, k, T) = e^{-r(T-t_0)}\mathbb{E}\left[e^{d.X_{t_1}}\mathbb{E}\left((\frac{e^{d.X_T}}{e^{d.X_{t_1}}} - k)^+ | X_{t_1}\right) | X_{t_0}\right]$$

Since the stock price at time t_1 is $S_{t_1} = e^{d.X_{t_1}}$, we will keep this notation for clarity when appropriate. Contrary to the lognormal hypothesis where the increments of the process $\log S$ are stationary and independent (see Section (1.2)) this is no longer the case in the jump-diffusion model and one must rely on taking the stock price at time t_1 as numeraire where $\mathbb{Q}^S$ is a martingale measure equivalent to $\mathbb{Q}$ such that

[4]There exists forward start call options which pay their owner the difference, if positive, between the performance and the relative strike k *i.e.* $(\frac{S_T}{S_{t_1}} - k)^+$

[5]In the Black & Sholes formula, we get $C_{BS}(t, x, \sigma) = e^x N[d_1] - Ke^{-r(T-t)}N[d_2]$ with $d_2 = \frac{x - \log K + (r - \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}}$.

$$\frac{dQ^S}{dQ} = e^{-\frac{1}{2}\int_t^T V_s ds + \int_t^T \sqrt{V_s} dW_Y(s)}$$

and performing the change of measure

$$dW_Y^S(t) = dW_Y(t) - \sqrt{V_t} dt$$

which is a Brownian motion under the $\mathbb{Q}^S$ probability measure. When the Brownian motions of the forward price and that of its instantaneous variance V_t are independent under the $\mathbb{Q}$ probability measure then the process V_t is unchanged when changing probability measures. However, when the Brownian motions are correlated then the process V_t is modified but only the order of magnitude of its parameters is changed. Therefore, for clarity of exposition we will not change our notation when describing the instantaneous variance. Examples can be found in Zhu [2007] in the case of the libor process and in Kruse [2003] in the case of the Heston model. The forward starting call option under the $\mathbb{Q}^S$ probability measure becomes

$$C^F(t_0, t_1, k, T) = \frac{B(t_1)}{B(T)} S_{t_0} \mathbb{E}^S \left[\frac{1}{S_{t_1}} \mathbb{E}\left((S_T - kS_{t_1})^+ | X_{t_1} \right) | X_{t_0} \right]$$

where $B(\cdot)$ is the money market. Introducing the pdf ϕ^S we get

$$C^F(t_0, t_1, k, T) = \frac{B(t_1)}{B(T)} S_{t_0} \int e^{-d.x_{t_1}} \mathbb{E}\left((S_T - ke^{d.x_{t_1}})^+ | X_{t_1} = x_{t_1} \right) \phi^S(x_{t_1}, t_1, X_{t_0}, t_0) \lambda(dx_{t_1})$$

for some $d \in \mathbb{R}^n$. Now, the integrand is

$$e^{-d.x_{t_1}} \mathbb{E}\left((S_T - ke^{d.x_{t_1}})^+ | X_{t_1} = x_{t_1} \right) = e^{-d.x_{t_1}} \int (e^{d.x_T} - ke^{d.x_{t_1}})^+ \phi(x_T, T, x_{t_1}, t_1) \lambda(dx_T) \quad (4.1.7)$$

which is independent from the stock price S_{t_1}. As we linked the implied volatility surface Σ to the factor level X_t and the time to maturity $T - t$, and as there exists a one-to-one relation between the implied volatility surface Σ and the pdf ϕ, we can consider ϕ a function $\phi(., T, X_t, t)$ of factor level and time to maturity. It actually means that, given X_{t_1}, the density $\phi(., T, X_{t_1}, t_1)$ is *known*.

We can now compute the quantity $\mathbb{E}\left((S_T - ke^{d.x_{t_1}})^+ | X_{t_1} = x_{t_1} \right)$ for any moneyness k. In our Jump-Diffusion model, we have :

$$\frac{1}{S_{t_1}} \mathbb{E}\left((S_T - ke^{d.x_{t_1}})^+ | X_{t_1} = x_{t_1} \right) = \frac{1}{P(t_1, T)} \frac{C_{JD}(t_1, X_{t_1}; kS_{t_1}, T)}{S_{t_1}}$$

Therefore we can rewrite Equation (4.1.5.2) as

$$\begin{aligned} C^F(t_0, t_1, k, T) &= P(t_1, T) S_{t_0} \int \frac{e^{-d.x_{t_1}}}{P(t_1, T)} C_{JD}(t_1, x_{t_1}; ke^{d.x_{t_1}}, T) \phi^S(x_{t_1}, t_1, X_{t_0}, t_0) d\lambda(x_{t_1}) \quad (4.1.8) \\ &= S_{t_0} \mathbb{E}^S \left[\frac{1}{S_{t_1}} C_{JD}(t_1, X_{t_1}; kS_{t_1}, T) | X_{t_0} \right] \end{aligned}$$

which implies that the forward start call option price is given by the discounted expectancy under the pricing measure $\mathbb{Q}^S$ of future European calls prices starting at t_1 and fixing at T. Note that these products do not exist and cannot be priced today unless we make some assumptions on future anticipations. This is actually what we are doing here. There are different methods for computing such an expectancy. For example, we can consider pricing the forward start option by using the characteristic function of the call price, we can also consider using the density of the underlying process at the determination time of the strike, or we can consider an approach consisting in applying Ito's lemma to the approximated call price at time t_1.

4.2 Dynamic risk management

4.2.1 Delta-hedging with a misspecified volatility

We saw in Section (1.2.2) that the BS-price of a short call option depends on the creation of a hedging portfolio consisting of being long a quantity of the stock price and the money account. Further, we saw in Section (4.1) that knowing the market implied volatility (IV) is key to risk management, as one can define hedging strategies by computing the appropriate Greeks. Thus, assuming some dynamics for the underlying stock price, we can dynamically manage the evolution of the replicating portfolio. However, in practice we do not know the true volatility of the process, but only a misspecified one. Hence, when risk managing an option we are concerned with the profit or loss (PnL) to be made when hedging options that are mispriced by the market. While volatility of the underlying can be statistically estimated on historical data, it can also be implied from the market since vanilla options are actively traded. The implied volatility differs from the historical one as it represents the future expected volatility from the market, and includes the market price of risk. Hence, we can either estimate statistically our own future volatility (actual volatility), or, use that of the market as measured by the implied volatility. In any case, using implied, or actual, volatility have different effects in the classical Black-Scholes delta formula, resulting in different profit or loss. As a result, in a world of constant volatility, there exists different delta-hedging strategies leading to different replicating portfolios. However, the IV is not stationary in time and one need to expand on the problem of hedging using the implied volatility by considering implied volatility varying stochastically (see El Karoui et al. [1998], Carr and Verma [2005]). Whichever hedging volatility we use, we will get different risk/return profiles.

In order to understand the performance of a hedging portfolio derived from misspecified volatility, we consider an option on a stock whose volatility is unknown and stochastic. That is, observing the true stock price, the hedger incorrectly computes the price of the contingent claim and the corresponding hedging portfolio. Given deterministic interest rates, and assuming a volatility dependent only on time and the current stock price, El Karoui et al. [1998] showed that the price of a European and American contingent claim with convex payoffs is a convex function of the price of the stock. Examining the performance of a hedging portfolio derived from misspecified volatility, they found that if the misspecified volatility dominates (or is dominated by) the true volatility, then the self-financing value of the misspecified hedging portfolio exceeds (or is dominated by) the payoff of the contingent claim at expiration. Alternatively, assuming that they know the true forecasted constant volatility in the Black-Scholes world, Ahmad et al. [2005] built a replicating portfolio by buying an underpriced option and delta hedging it to expiry with the actual volatility.

We are going to illustrate these two approaches by considering a trader expecting the underlying stock price S to go down, and choosing to sell a call option with payoff $h(S_T)$ at maturity T. Assuming the Black-Scholes world, to hedge himself the trader can use the option premium π_0 to build a replicating portfolio $V(t) = \pi_\Delta(t)$ for $0 \leq t \leq T$. He is then going to dynamically buy a quantity of the underlying stock to hedge the risk of paying the payoff at maturity. That is, he is going to monitor the option price $P(t)$ for $t \in [0, T]$ and risk-manage the self-financing portfolio $V(t)$ with initial value being the option premium $x = \pi_0$, making sure that his tracking error $\epsilon(T) = V(T) - h(S_T)$ is null at maturity. While the market prices $v_\gamma(t, S_t)$ with misspecified volatility γ correspond to the expectected value of the future financial flux to be paid, the replicating portfolio is directly observed on the market. Hence, the remaining question is that of the choice of the volatility when computing the delta-hedging portfolio.

4.2.2 Replication and measure of errors

4.2.2.1 European option

Following El Karoui et al. [1998], we consider a continuous-time economy with a positive finite horizon T and two traded assets in a frictionless market.

Assumption 4.2.1 *We define the money market $M(t)$ and the stock price $S(t)$ on a probability space (Ω, F, P) and adapted to a filtration $\{\mathcal{F}_t; 0 \leq t \leq T\}$. In addition their dynamics are given by*

$$\begin{aligned} M(t) &= e^{\int_0^t r(s)ds} \\ \frac{dS_t}{S_t} &= r(t)dt + \sigma(t)dW(t) \end{aligned}$$

where $\{W(t); 0 \leq t \leq T\}$ is a one-dimensional Brownian motion adapted to $\mathcal{F}_t$ for $0 \leq t \leq T$, σ is a non-negative volatility process satisfying $E[\int_0^T \sigma^2(t)dt] < \infty$, and r is a deterministic interest rate process.

Letting $\frac{S}{M}$ be a martringale and assuming $E[\frac{S^2(t)}{M^2(t)}] < \infty$ for $0 \leq t \leq T$, we consider a bounded adapted portfolio process $\Delta(t)$ for $0 \leq t \leq T$ and define the portfolio value at time t as

$$\pi_\Delta(t) = \Delta(t)S(t) + (\pi_\Delta(t) - \Delta(t)S(t))$$

where $(\pi_\Delta(t) - \Delta(t)S(t))$ is a quantity of cash.

Definition 4.2.1 *Given a non-random initial portfolio value $\pi_\Delta(0)$, the self-financing value of a portfolio process Δ is the solution of the linear SDE*

$$d\pi_\Delta(t) = r(t)\big[\pi_\Delta(t) - \Delta(t)S(t)\big]dt + \Delta(t)dS(t) \tag{4.2.9}$$

with solution

$$\pi_\Delta(t) = M(t)\big[\pi_\Delta(0) + \int_0^t \Delta(u)d(\frac{S(u)}{M(u)})\big]$$

Since $\frac{S(t)}{M(t)}$ is assumed to be a square-integrable martingale, and Δ is bounded, the discounted portfolio value $\frac{\pi_\Delta(t)}{M(t)}$ is a square integrable martingale. We now define a payoff convex function h on $(0, \infty)$, having bounded one-sided derivatives as

$$\Big|h^{'}(x\pm)\Big| \leq C \,, \forall x > 0$$

for a positive constant C. In that setting, a price process for a European option is any adapted process $\{P(t); 0 \leq t \leq T\}$ satisfying the terminal condition $P(T) = h(S(T))$ a.s.. In a complete market the arbitrage price of a European option is

$$P_E(t) = M(t)E\big[\frac{h(S(T))}{M(T)}|F(t)\big] \,, 0 \leq t \leq T \tag{4.2.10}$$

We assume that we short the option P and hedge it by being long a quantity Δ of the stock and some cash. We now define our measure of error, the Tracking Error associated with (P, Δ), which is the process

$$e(t) = \pi_\Delta(t) - P(t)$$

with initial condition $\pi_\Delta(0) = P(0)$. If the discounted tracking error $\frac{e(t)}{M(t)}$ is

1. identically equal to zero, then (P, Δ) is a replicating strategy

2. non-decreasing, then (P, Δ) is a superstrategy

3. non-increasing, then (P, Δ) is a substrategy

***Remark* 4.2.1** *A hedger who incorrectly estimates the volatility of the underlying will incorrectly compute the European option price and hedging portfolio.*

Let (P, Δ) be the result of such a computation. If the hedger starts with a portfolio with initial value $\pi_\Delta(0) = P(0)$ and uses the portfolio process Δ, at maturity, the portfolio will be $\pi_\Delta(T)$. If (P, Δ) is a superstrategy, then $\frac{e}{M}$ is non-decreasing, and since $e(0) = 0$ we have

$$\pi_\Delta(T) = P(T) + e(T) \geq h(S(T))$$

and

- the hedger has successfully hedged a short position in the contingent claim.
- A substrategy (the negative of the portfolio process of a substrategy) hedges a long position, and the above inequality is reversed.
- If the market is complete, there exists a unique replicating strategy for the European contingent claim with price process given in Equation (4.2.10).

4.2.3 Delta hedging in a complete market

4.2.3.1 Using implied volatility

We are now going to quantify the distribution of the tracking error. In addition to the true volatility process, σ, we assume that we price an option with a misspecified volatility, γ, which may be stochastic only through dependence on the current stock price.

***Remark* 4.2.2** *Allowing for the misspecified volatility γ to be only dependent on the current stock price, comparisons between σ and γ lead to comparisons of contingent claim prices and performance of hedging portfolios.*

***Assumption* 4.2.2** *The functions $r : [0, T] \to \mathbb{R}$ and $\gamma : [0, T] \times (0, \infty) \to [0, \infty)$ are Holder continuous.*

We write the dynamics of the misspecified underlying asset price process as

$$dS_\gamma(u) = S_\gamma(u)\big(r(u)du + \gamma(u, S_\gamma(u))dW(u)\big) \ , t \leq u \leq T$$

with initial condition $S_\gamma(t) = x$ for $x > 0$. The misspecified option value at time t is

$$v_\gamma(t, x) = E\big[e^{-\int_t^T r(u)du} h(S_\gamma(T))\big] \ , 0 \leq t \leq T \ , x > 0 \tag{4.2.11}$$

***Remark* 4.2.3** *If the stock prices $S_\gamma(t)$ were governed by the above SDE, the market would be complete and $v_\gamma(t, x)$ would be the arbitrage price of the contingent claim.*

The option prices satisfy the PDE in Equation (1.6.30) which we rewrite as

$$\mathcal{L}_\gamma v_\gamma(t, x) - r v_\gamma(t, x) = 0 \ , 0 \leq t \leq T \ , x > 0$$

where $\mathcal{L}$ is the diffusion operator. That PDE can be expressed in terms of the true stock process $S(t)$ as

$$\frac{\partial v_\gamma}{\partial t}(t, S(t)) = r(t) v_\gamma(t, S(t)) - r(t) S(t) \Delta_\gamma(t) - \frac{1}{2} \gamma(t, S(t))^2 S(t)^2 \frac{\partial^2 v_\gamma}{\partial x^2}(t, S(t))$$

where $\Delta_\gamma(t) = \frac{\partial v_\gamma}{\partial x}$ is bounded. Observing the true stock price $S(t)$, the hedger (incorrectly) computes the contingent claim price $P_\gamma(t) = v_\gamma(t, S(t))$ and uses the hedging portfolio $\Delta_\gamma(t) = \partial_x v_\gamma(t, S(t))$. Starting with the initial value $v_\gamma(0, S(0))$, the dynamics of the self-financing replicating portfolio is given in Equation (4.2.9). Alternatively, we can express the option price process with respect to the true stock process as $P_\gamma(t) = v_\gamma(t, S(t))$ and find its SDE using Ito's lemma

$$dP_\gamma(t) = \frac{\partial v_\gamma}{\partial t}(t, S(t))dt + \frac{\partial v_\gamma}{\partial x}(t, S(t))dS(t) + \frac{1}{2}\sigma^2(t)S^2(t)\frac{\partial^2 v_\gamma}{\partial x^2}(t, S(t))dt$$

Now, replacing $\frac{\partial v_\gamma}{\partial t}(t, S(t))$ in the above equation with the value previously calculated with the misspecified volatility, we get

$$dP_\gamma(t) = r(t)P_\gamma(t)dt + \Delta_\gamma[dS(t) - r(t)S(t)dt] + \frac{1}{2}[\sigma^2(t) - \gamma^2(t, S(t))^2]S^2(t)\frac{\partial^2 v_\gamma}{\partial x^2}(t, S(t))dt$$

Given the tracking error $e_\gamma(t)$, with SDE

$$de_\gamma(t) = d\pi_{\Delta_\gamma}(t) - dP_\gamma(t)$$

and after rearranging, we get the dynamics

$$de_\gamma(t) = r(t)e_\gamma(t)dt + \frac{1}{2}[\gamma^2(t, S(t)) - \sigma^2(t)]S^2(t)\frac{\partial^2 v_\gamma}{\partial x^2}(t, S(t))dt$$

Using the relation between the gamma and the vega in Equation (B.1.6), we can rewrite the dynamics of the tracking error as

$$de_\gamma(t) = r(t)e_\gamma(t)dt + \frac{1}{2}\frac{\gamma(t, S(t))}{t}[1 - \frac{\sigma^2(t)}{\gamma^2(t, S(t))}]Vega(t, S(t))dt$$

Using $e^{rt}d(e^{-rt}e_\gamma(t))$ [6] and discounting, we rewrite the SDE as

$$d(e^{-rt}e_\gamma(t)) = e^{-rt}\frac{1}{2}[\gamma^2(t, S(t)) - \sigma^2(t)]S^2(t)\frac{\partial^2 v_\gamma}{\partial x^2}(t, S(t))dt$$

Integrating in the range $[0, t]$, the tracking error $e_\gamma(t)$ is given by

$$e_\gamma(t) = \frac{1}{2}M(t)\int_0^t \frac{1}{M(u)}[\gamma^2(u, S(u)) - \sigma^2(u)]S^2(u)\frac{\partial^2 v_\gamma}{\partial x^2}(u, S(u))du \tag{4.2.12}$$

such that to be over-hedged we require

$$\gamma^2(t, S(t)) \geq \sigma^2(t)$$

Theorem 4.2.1 *Given the Assumptions (4.2.1), (4.2.2), if*

$$\sigma(t) \leq \gamma(t, S(t))$$

for Lebesgue-almost all $t \in [0, T]$, almost surely, then $(P_\gamma, \Delta_\gamma)$ is a superstrategy and

$$\pi_{\Delta_\gamma}(T) \geq h(S(T))$$

Inversly, if $\sigma(t) \geq \gamma(t, S(t))$, then $(P_\gamma, \Delta_\gamma)$ is a substrategy.

[6] $e^{rt}d(e^{-rt}e_\gamma(t)) = de_\gamma(t) - r(t)e_\gamma(t)dt$

This theorem shows that successful hedging is possible even under significant model error. It depends on the relationship between the misspecified volatility $\gamma(t, S(t))$ and the true volatility $\sigma(t)$. In presence of a superstrategy, the trader makes a profit with probability one even though the true price model is different from the assumed one as long as $\Gamma_t = \frac{\partial^2 v_\gamma}{\partial x^2}(t, S(t)) \geq 0$. Further, since the hedging error depends on the option convexity gamma, for small gamma the hedging error is small, irrespective of the model. The expected PnL is maximum when the stock price is close to the money at maturity, where Gamma is largest, and when the instantaneous volatility is large at maturity. That is, the PnL depends on the path taken by the underlying asset. Two trajectories with the same average volatility ending at-the-money can lead to two very different PnL.

4.2.3.2 Using forecasted volatility

Rather than using the implied volatility to construct the replicating portfolio, Ahmad et al. [2005] assumed that they know the true forecasted volatility. Following the argument given in Remark (4.2.2), we can extend their demonstration by allowing for the misspecified volatility γ to be only dependent on the current stock price. In our setting, we are short the option and long the replicating portfolio. Hence, the replicating portfolio value at time t is

$$\pi_{\Delta_a}(t) = \Delta_a(t) S_t + (\pi_{\Delta_a(t)} - \Delta_a(t) S_t)$$

which is solution of the SDE

$$d\pi_{\Delta_a}(t) = r(t)(\pi_{\Delta_a(t)} - \Delta_a(t) S_t) dt + \Delta_a(t) dS_t$$

where Δ_a is the quantity of stock computed with the actual volatility σ. In order to mark-to-market prices, the dynamics of the price process with misspecified volatility follow the SDE

$$dC_\gamma(t) = r(t) C_\gamma(t) dt + \Delta_\gamma [dS(t) - r(t) S(t) dt] + \frac{1}{2}[\sigma^2(t) - \gamma^2(t, S(t))] S^2(t) \frac{\partial^2 v_\gamma}{\partial x^2}(t, S(t)) dt$$

where $\gamma(t, S(t))$ is the misspecified volatility. The Tracking Error associated with (C_γ, Δ_a) is the process

$$\epsilon(t) = \pi_{\Delta_a}(t) - C_\gamma(t)$$

with dynamics

$$d\epsilon(t) = r\epsilon(t) dt + (\Delta_a(t) - \Delta_\gamma(t))[dS(t) - r(t) S(t) dt] + \frac{1}{2}[\gamma^2(t, S(t)) - \sigma^2(t)] S^2(t) \frac{\partial^2 v_\gamma}{\partial x^2}(t, S(t)) dt$$

which is equal to

$$e^{rt} d(e^{-rt}\epsilon(t)) = (\Delta_a(t) - \Delta_\gamma(t))[dS(t) - r(t) S(t) dt] + \frac{1}{2}[\gamma^2(t, S(t)) - \sigma^2(t)] S^2(t) \frac{\partial^2 v_\gamma}{\partial x^2}(t, S(t)) dt$$

and correspond to the Tracking Error from time t to $t + dt$. Discounting with $P(t_0, t) = e^{-r(t-t_0)}$ to get the Tracking Error at time t_0, we get

$$e^{rt_0} d(e^{-rt}\epsilon(t)) = P(t_0, t)(\Delta_a(t) - \Delta_\gamma(t))[dS(t) - r(t) S(t) dt] + \frac{1}{2} P(t_0, t)[\gamma^2(t, S(t)) - \sigma^2(t)] S^2(t) \frac{\partial^2 v_\gamma}{\partial x^2}(t, S(t)) dt$$

So, the total Tracking Error from time t_0 to maturity T is

$$
\begin{aligned}
e^{rt_0}\int_{t_0}^{T} d(e^{-rt}\epsilon(t)) = P(t_0,T)\epsilon(T) = \int_{t_0}^{T} P(t_0,u)(\Delta_a(u)-\Delta_\gamma(u))[dS(u)-r(u)S(u)du] \\
+ \quad \frac{1}{2}\int_{t_0}^{T} P(t_0,u)[\gamma^2(u,S(u))-\sigma^2(u)]S^2(u)\frac{\partial^2 v_\gamma}{\partial x^2}(u,S(u))du
\end{aligned}
$$

since $\epsilon(t_0)=0$.

In the case of time-dependent volatility, both the dynamics of the call price $C_{BS}(t,S_t;\sigma(t))$ with true forecasted volatility $\sigma(t)$ and the dynamics of the call price $C_{BS}(t,S_t;\gamma(t))$ with misspecified volatility $\gamma(t)$ satify the Equation (1.2.5). Further, the two BS prices are solutions to the parabolic equation (1.6.30) with flat volatility $\sigma(t)$ and $\gamma(t)$ respectively. As we assumed that our forecasted volatility was correct, the dynamics of the former price is

$$
dC_{BS}(t,S_t;\sigma(t)) = rC_{BS}(t,S_t) + \Delta_a(t)(dS_t - rS_t dt)
$$

with Tracking Error $d\epsilon_a(t) = d\pi_{\Delta_a}(t) - dC_a(t) = r(t)\epsilon_a(t)dt$, while the dynamics of the latter price is

$$
dC_{BS}(t,S_t;\gamma(t)) = \frac{1}{2}(\sigma_t^2-\gamma^2(t))S_t^2\partial_{xx}C_{BS}(t,S_t;\gamma(t)) + rC_{BS}(t,S_t;\gamma(t)) + \Delta_\gamma(t)(dS_t - rS_t dt)
$$

Now, taking the difference between these two theoretical option values with different volatilities, we get

$$
\begin{aligned}
dC_{BS}(t,S;\sigma(t)) - dC_{BS}(t,S;\gamma(t)) \quad &= \quad r(C_{BS}(t,S;\sigma(t)) - C_{BS}(t,S;\gamma(t))dt + (\Delta_a(t)-\Delta_\gamma(t))(dS_t - rS_t dt) \\
&+ \quad \frac{1}{2}(\gamma^2(t)-\sigma_t^2)S_t^2\partial_{xx}C_{BS}(t,S_t;\gamma(t))
\end{aligned}
$$

which is equal to

$$
\begin{aligned}
e^{rt}d\big[e^{-rt}(C_{BS}(t,S;\sigma(t)) - C_{BS}(t,S;\gamma(t)))\big] = (\Delta_a(t)-\Delta_\gamma(t))(dS_t - rS_t dt) \\
+ \quad \frac{1}{2}(\gamma^2(t)-\sigma_t^2)S_t^2\partial_{xx}C_{BS}(t,S_t;\gamma(t))
\end{aligned}
$$

and correspond to the Tracking Error from time t to $t+dt$. Discounting with the zero-coupon bond $P(t_0,t) = e^{-r(t-t_0)}$, to get the Tracking Error at time t_0, we get

$$
\begin{aligned}
e^{rt_0}d\big[e^{-rt}(C_{BS}(t,S;\sigma(t)) - C_{BS}(t,S;\gamma(t)))\big] = P(t_0,t)(\Delta_a(t)-\Delta_\gamma(t))(dS_t - rS_t dt) \\
+ \quad \frac{1}{2}P(t_0,t)(\gamma^2(t)-\sigma_t^2)S_t^2\partial_{xx}C_{BS}(t,S_t;\gamma(t))
\end{aligned}
$$

So, the total Tracking Error from time t_0 to maturity T is

$$
\begin{aligned}
e^{rt_0}\int_{t_0}^{T} d\big[e^{-rt}(C_{BS}(t,S;\sigma(t)) - C_{BS}(t,S;\gamma(t)))\big] = C_{BS}(t_0,S;\gamma(t)) - C_{BS}(t_0,S;\sigma(t)) \\
= \quad \int_{t_0}^{T} P(t_0,u)(\Delta_a(u)-\Delta_\gamma(u))(dS_u - rS_u du) + \frac{1}{2}\int_{t_0}^{T} P(t_0,u)(\gamma^2(u)-\sigma^2(u))S_u^2\partial_{xx}C_{BS}(u,S_u;\gamma(u))
\end{aligned}
$$

which corresponds to $P(t_0,T)\epsilon(T)$ above. Hence, we can conclude that

$$
P(t_0,T)\epsilon(T) = C_{BS}(t_0,S;\gamma(t)) - C_{BS}(t_0,S;\sigma(t))
$$

which is the difference between the option with the misspecified volatility and that with the true forecasted one.

4.2.4 Comments on delta hedging

When delta-hedging with forecasted volatility, even though the final profit is guaranteed and known for $\gamma(t) > \sigma(t)$, $t \in [0,T]$ (when shorting options), it is a random variable due to the extra term $[dS(t) - r(t)S(t)dt]$ in the tracking error. As a result, on a mark-to-market basis we could lose money before making a gain. This leaves us with some noise in our mark-to-market PnL which is not ideal from a risk management point of view. It is similar to owning a bond, as for a bond there is a guaranteed outcome, but we may lose on a mark-to-market basis in the meantime. On the other hand, when hedging with implied volatility we are balancing the random fluctuations in the mark-to-market option value with the fluctuations in the stock price so that the evolution of the portfolio value is deterministic. From a risk management perspective this is much better behaved. Further, we do not need to know the actual volatility. To make a profit, all we need to know is that the actual volatility is always going to be greater than the implied volatility (if we buying), or always less than the implied volatility (if we are selling). This is because in the case of a convex price, the gamma is positive and the trader will make a profit as long as the chosen volatility dominates the true one.

In the case of forecasted volatility, when integrating the present value of all of the profits over the life of the option to get a total profit, the result is always positive, but highly path dependent. Hence, maximising the total profit will depend on the trajectory of the stock price, and more precisely its historical drift term. In that setting, the PnL is always increasing in value but the end result is random, and we are left with maximising $S^2(t)\frac{\partial^2 v_\gamma}{\partial x^2}(t,S(t))$. That is, we want maximum fluctuations in the trajectory of the stock price. However, when hedging with a delta based on implied volatility, the profit made each day is deterministic (with constant volatility), but the present value of total profit up to expiration is path dependent, and given by the tracking error $\epsilon(T)$. Ahmad et al. [2005] considered $P(t,S,\epsilon)$ to be the real expected value of the PnL with known terminal condition $P(T,S,\epsilon) = \epsilon(T)$, and derived its partial differential equation, assuming a solution of this equation of the form

$$P(t,S,\epsilon) = \epsilon + F(t,S)$$

Obtaining analytical solutions, they analysed the expected profit versus the growth rate μ of the stock price. They found that to maximise the expected profit, the growth rate must ensure that the stock ends up close to, or, at the money at expiration, where gamma is largest. Further, for most realistic parameter regimes the maximum expected profit hedging with implied volatility is similar to the guaranteed profit hedging with actual volatility. Practically, considering the expected profit versus the strike K and the drift μ, the higher the growth rate the larger the strike price at the maximum. Introducing the skew effect, when the actual volatility is ATM there is no maximum, and profitability increases with distance away from the money. Similarly, they calculated the variance in the final profit $Var(\epsilon(T))$ using implied volatility by solving the PDE of $v(t,S,\epsilon)$.

4.2.5 Delta hedging American option

Working under the Hypothesis (4.2.1), we let the Snell envelope of $\{\frac{h(S(t))}{M(t)}; 0 \leq t \leq T\}$ be the smallest supermartingale dominating this process, and given by

$$\text{ess} \sup_{\tau \in \mathcal{T}_t} E[\frac{h(S(\tau))}{M(\tau)}|\mathcal{F}(t)] \, , 0 \leq t \leq T$$

where $\mathcal{T}_t$ is the set of stopping times τ satisfying $t \leq \tau \leq T$ almost surely. Bensoussan [1984] and Karatzas [1988] showed that in a complete market, the arbitrage price process P_A of an American option with payoff h satisfies

$$P_A(t) = M(t) \, \text{ess} \sup_{\tau \in \mathcal{T}_t} E[\frac{h(S(\tau))}{M(\tau)}|\mathcal{F}(t)] \, , 0 \leq t \leq T$$

An optimal exercise time is

$$D = \inf\{t \in [0,T]; P_A(t) = h(S(t))\}$$

and the process $\{\frac{P_A(t\wedge D)}{M(t\wedge D)}, 0 \leq t \leq T\}$ is a martingale. Under certain conditions, El Karoui et al. [1998] showed that

$$P_A(t) = M(t)E[\frac{h(S(T))}{M(T)}|\mathcal{F}(t)] \, , 0 \leq t \leq T$$

and the price processes for the European and American options agree.

Theorem 4.2.2 *Working under the Hypothesis (4.2.1), in a complete market, if $r(t) \geq 0$ for all $t \in [0,T]$ and $h(0) = 0$, then the European and American contingent claim price processes coincide.*

In the case of American option, the misspecified option value at time t is

$$v_\gamma(t,x) = \sup_{\tau \in \mathcal{T}_t} E\big[e^{-\int_t^\tau r(u)du} h(S_\gamma(\tau))\big] \tag{4.2.13}$$

When shorting the American option, the hedger must be prepared to hedge it all the way to maturity. However, if he is long the option the optimal exercise time is

$$D_\gamma = \inf\{t \in [0,T]; v_\gamma(t,S(t)) = h(S(t))\}$$

and he only need to hedge until that time. El Karoui et al. [1998] obtained the following theorem

Theorem 4.2.3 *Given the Assumptions (4.2.1), (4.2.2) and the optimal exercise time D_γ, if*

$$\sigma(t) \leq \gamma(t,S(t)) \, , 0 \leq t \leq T$$

then $(P_\gamma, \Delta_\gamma)$ is a superstrategy for American option. Inversly, given

$$D = T \wedge \inf\{t \in [0,T]; P(t) = h(S(t))\}$$

if $\sigma(t) \geq \gamma(t,S(t))$ for $0 \leq t \leq D$, then $(P_\gamma, \Delta_\gamma)$ is a substrategy for American option.

4.2.6 Delta hedging with stochastic volatility

We are now going to discuss delta headging in incomplete markets. For simplicity, we consider a stochastic volatility model as a typical model used in incomplete markets. We let the stock price in continuous time follow under the $\mathbb{Q}$-measure the dynamics

$$dS_t = rS_t dt + \sigma_t S_t dW_S(t)$$

where we assume a general type of dynamics for the instantaneous volatility σ_t given by

$$d\sigma_t = \mu_\sigma dt + \xi dW_\sigma(t)$$

where ξ is the volatility of volatility, and $< dW_S, dW_\sigma >_t = \rho_{S,\sigma} dt$. Given an option V_t written on the asset S_t, the hedged portfolio consists in being long the option and short the replicating portfolio $\pi_\delta(t)$ made of a quantity δ of the stock price and a certain amount of cash. Letting $\epsilon(t) = (V_t - \pi_\delta(t))$ be the tracking error, the PnL at time t is

$$\epsilon(t) = V_t - \delta S_t + (\delta S_t - \pi_\delta(t))$$

where $\pi_\delta(t) = \delta S_t + (\pi_\delta(t) - \delta S_t)$ is the replicating portfolio. The variation of the tracking error over the period $[t, t+dt]$ is given by

$$d\epsilon(t) = dV_t - \delta dS_t + (\delta S_t - \pi_\delta(t))rdt$$

where the first part corresponds to the price variation of the option, the second one to the stock price move, and the last one is the risk-free return of the amount of cash. Setting $V = V(t, S, \sigma)$ and applying Ito's lemma on dV_t, we get

$$dV_t = \Theta dt + \frac{\partial V}{\partial S} dS_t + \frac{\partial V}{\partial \sigma} d\sigma_t + \frac{1}{2}\big[\frac{\partial^2 V}{\partial S^2} < dS, dS >_t + \frac{\partial^2 V}{\partial \sigma^2} < d\sigma, d\sigma >_t + 2\frac{\partial^2 V}{\partial S \partial \sigma} < dS, d\sigma >_t\big] + ...$$

where σ_t represents the stochastic volatility. In the PnL formula, we can replace the term Θdt with its value given in the Black-Scholes PDE, calculated with the implied volatility that had to be input to determine the amount of cash to lock the position. Hence, we obtain

$$\begin{aligned} d\epsilon(t) &= -(\frac{1}{2}\sigma_{imp}^2 S_t^2 \frac{\partial^2 V}{\partial S^2} + rS_t \frac{\partial V}{\partial S} - rV_t)dt + \frac{\partial V}{\partial S} dS_t + \frac{\partial V}{\partial \sigma} d\sigma_t \\ &+ \frac{1}{2}\big[\frac{\partial^2 V}{\partial S^2} < dS, dS >_t + \frac{\partial^2 V}{\partial \sigma^2} < d\sigma, d\sigma >_t + 2\frac{\partial^2 V}{\partial S \partial \sigma} < dS, d\sigma >_t\big] - \delta dS_t + (\delta S_t - \pi_\delta(t))rdt \end{aligned}$$

Assuming the trader delta-hedge himself with the quantity $\delta = \frac{\partial V}{\partial S}$, the PnL simplifies to

$$\pi_V(t, t+dt) = \frac{1}{2}\Gamma S_t^2 \big[(\frac{dS_t}{S_t})^2 - \sigma_{imp}^2 dt\big] + \frac{\partial V}{\partial \sigma} d\sigma_t + \frac{1}{2}\frac{\partial^2 V}{\partial \sigma^2} < d\sigma, d\sigma >_t + \frac{\partial^2 V}{\partial S \partial \sigma} \sigma_t S_t < dW, d\sigma >_t$$

where $\pi_V(t, t+dt) = e^{rt} d(e^{-rt}\epsilon(t))$. In trading terms, it can be expressed as

$$\pi_V(t, t+dt) = \frac{1}{2}\Gamma S_t^2 \big[(\frac{dS_t}{S_t})^2 - \sigma_{imp}^2 dt\big] + \text{ Vega } d\sigma_t + \frac{1}{2} \text{ Volga } < d\sigma, d\sigma >_t + \text{ Vanna } \sigma_t S_t \xi \rho_{S,\sigma} dt \quad (4.2.14)$$

where $\rho_{S,\sigma}$ is the correlation between the stock price and the volatility, and ξ is the volatility of volatility.

4.2.7 Delta-hedged portfolio revised

Rather than using a single PnL equation to describe either a long option position or a short one, we assume that a profit is positive and a loss is negative, and we get one PnL equation for the former and another one for the latter. From the definition of the BS-theta in Equation (B.1.10), we can get the following approximation

$$\Theta = \frac{\partial}{\partial t} C_{BS}(t, S_t, K, T) \approx -\frac{1}{2} S_t^2 \sigma^2 \Gamma(t, S_t)$$

so that we can approximate the PnL for being long the option and short the replicating portfolio as

$$\pi_V(t, t+dt) \approx -\Theta\big[(\frac{dS_t}{S_t \sigma \sqrt{dt}})^2 - 1\big]dt = \Theta\big[1 - (\frac{dS_t}{S_t \sigma \sqrt{dt}})^2\big]dt \quad (4.2.15)$$

where $n_S = \frac{dS_t}{S_t \sigma \sqrt{dt}}$ is the standardise move of the stock price on the considered period. Similarly, should we sell the option and be long the stock price, the PnL for the portfolio on $[t, t+dt]$ would be

$$\pi_V(t, t+dt) = \frac{1}{2}\Gamma S_t^2 \big[\sigma^2 dt - (\frac{dS_t}{S_t})^2\big]$$

and we get the approximate PnL as

$$\pi_V(t, t+dt) \approx -\Theta\big[1 - (\frac{dS_t}{S_t \sigma \sqrt{dt}})^2\big]dt = \Theta\big[(\frac{dS_t}{S_t \sigma \sqrt{dt}})^2 - 1\big]dt \quad (4.2.16)$$

Note, in the literature some authors use only Equation (4.2.16) to represent the PnL of an option, where $\Theta > 0$ when we short the option and $\Theta < 0$ when it is bought. When using a stochastic volatility model, we can approximate the PnL for being long the option and short the replicating portfolio as

$$\pi_V(t, t+dt) \approx \Theta\big[1 - (\frac{dS_t}{S_t \sigma \sqrt{dt}})^2\big] dt + \text{ Vega } d\sigma + \frac{1}{2} \text{ Volga } < d\sigma, d\sigma >_t + \text{ Vanna } \sigma_t S_t < dW, d\sigma >_t \quad (4.2.17)$$

Part II

Quantitative trading in inefficient markets

Chapter 5

A multifractal market

5.1 Market anomalies

5.1.1 Introducing discrete dividends

We discussed in Section (1.2.2) the pricing of vanilla option in the Black-Scholes world where it is assumed ideal conditions in the market for the underlying asset and its associated option prices. However, in presence of repo rates and discrete dividends we get into complicated problems. For instance, even though we could convert discrete dividends into dividend yields, as an ex-dividend date approaches the dividend yield could grow arbitrarily high. We would also need a different dividend yield for each option maturity. Further, the Greeks can become distorted when discrete dividends are replaced with a continuous dividend yield.

5.1.1.1 The pure stock process

We can always rewrite the spot price with discrete dividends in terms of a process Y with no discrete dividends and no rates by using the transformation

$$S_t = a(t)Y_t + b(t) \tag{5.1.1}$$

where $a(t)$ and $b(t)$ are deterministic functions of time (see Overhaus et al. [2002]). In that setting, the price of a call option on S is

$$C_S(t;K,T) = P(t,T)a(t,T)E_t[(Y_T - k)^+] = P(t,T)a(t,T)C_Y(t;k,T) = C(t,Y_t;k,T)$$

where $a(t,T)$ and $b(t,T)$ are deterministic functions to be specified, and where $k = \frac{K-b(t,T)}{a(t,T)}$ and $C_Y(t;k,T) = E_t[(Y_T - k)^+|Y_y = y]$. Hence, a discrete set of option prices on S determines a discrete set of option prices on Y. If we reverse the process we get $K = a(t,T)x + b(t,T)$, for some $x > 0$, and the call price becomes

$$C_Y(t;x,T) = \frac{1}{P(t,T)a(t,T)}C_S(t;a(t,T)x + b(t,T),T)$$

However, since the market only quotes strike K, we should use the price

$$C_Y(t;k,T) = \frac{1}{P(t,T)a(t,T)}C_S(t;K,T) \tag{5.1.2}$$

expressed in terms of (k,T) when stock price has discrete dividends. As explained by Buehler [2008], in presence of discrete dividends the no-arbitrage conditions no-longer hold for S, but it does for the pure process Y. As a result, one must satisfy the no-arbitrage conditions in the Y-space and have positive butterflies $\partial_{kk}C_Y(t;k,T) \geq 0$ and positive calendar spreads $\partial_T C_Y(t;k,T) \geq 0$ for all maturities $T \geq 0$ and all strikes $k \geq 0$. Hence, the Black-Scholes implied volatility defined in Equation (1.4.14) is not the appropriate number for observing pure volatility risk in order to compare option prices. When using this mapping device, traders have no tools for identifying the inherent violation of arbitrage bounds implied by the presence of discrete dividends and can no-longer properly manage their books. Since the no-arbitrage constraints are only defined in terms of the pure process Y, one should use a mapping device on Y by directly considering the implied volatility of Y. Hence, given the option price $C_Y(t;k,T)$ at time t for a strike k and a maturity T, the pure implied volatility $\Sigma(t,Y_t;K,T)$ satisfies

$$C_Y(t;k,T) = C_{BS}(t,Y_t,k,T;\Sigma_Y(k,T)) \tag{5.1.3}$$

where $C_{BS}(t,Y_t,k,T;\sigma)$ is the Black-Scholes formula for a call option with volatility σ.

5.1.1.2 The dynamics of the stock price

We are going to introduce discrete dividends to the stock price $\{S_t, t \in [0,T]\}$ and describe our framework. We let

$$C(t,T) = \frac{Re(t,T)}{P(t,T)} = e^{\int_t^T \mu_s ds}$$

be the capitalisation factor from time t until time T. Further, we let D_t be the dividends paid till time t given by

$$D_t = \sum_{i=0}^{\infty} \mathcal{H}(t - t_{d_i}) d_i C(t_{d_i}, t)$$

with dynamics given by

$$dD_t = \mu_t D_t dt + D_t^{'} dt$$

where $D_t^{'} = \sum_{i=0}^{\infty} \delta(t - t_{d_i}) d_i C(t_{d_i}, t)$. Similarly, the dividends paid from time t till time T and capitalised at maturity T are given by

$$D(t,T) = D_T - C(t,T) D_t$$

Discounting these dividends to time t with the rate μ, we get the present value $D_{PV}(t;t,T)$ of the dividends paid between t and T

$$D_{PV}(t;t,T) = \frac{D(t,T)}{C(t,T)} = \frac{D_T}{C(t,T)} - D_t$$

Having made an assumption on discrete dividends, for simplicity of exposition, we are going to consider the dynamics of the stock price in the Spot model (see details in Appendix (B.2.3)). The stock price dynamics in the spot model and under the risk-neutral measure $\mathbb{Q}$ become

$$\begin{aligned} S_t &= C(0,t) Y_t - D_t \\ dY_t &= Y_t \sigma_{Y,t} dW_t \\ Y_0 &= S_0 \end{aligned}$$

with $a(t) = C(0,t)$, $b(t) = -D_t$ and where $\sigma_{Y,t}$ is the instantaneous volatility which is independent from the discrete dividend D_t. Using Ito's calculus we can derive the process for the stock price dynamics as

$$dS_t = \mu_t S_t + (S_t + D_t)\sigma_{Y,t} dW_t - D_t^{'} dt \tag{5.1.4}$$

In that setting the price of a European call option with strike K and maturity T is

$$C(t,Y_t,K^{''},T)=E_t^Q[e^{-\int_t^T q_s ds}(Y_T-K^{''})^+]$$

where $K^{''}=\frac{K+D(t,T)}{C(t,T)}$.
For notational purpose we let $Z_t=C(0,t)Y_t$ be the asset price S_t plus the forward value of all dividends paid from inception up to time t. Thus, the spot price becomes

$$S_t=Z_t-D_t$$

and Z_t is a continuous process with dynamics given by

$$dZ_t=\mu_t Z_t+\sigma_{Z,t}Z_t dW_t \tag{5.1.5}$$

This model is popular mainly due to the fact that adding the already paid dividends to the strike of the option one can get a closed-form solution for European option prices

$$C_S(K,T)=P(t,T)E_t^Q[(S_T-K)^+]=P(t,T)E_t^Q[(Z_T-K^{'})^+] \tag{5.1.6}$$

where $K^{'}=K+D(t,T)$ such that $a(t,T)=1$ and $b(t,T)=-D(t,T)$.

5.1.1.3 The multi-asset dynamics

The stock price dynamics in the spot model for the N-assets involved satisfy

$$\begin{aligned} S_t^i &= C^i(0,t)Y_t^i-D_t^i \\ dY_t^i &= Y_t^i\sigma_t^i dW_t^i \\ Y_0^i &= S_0^i \end{aligned}$$

with a correlation structure between the n-dimensional Brownian motion $(W_t^i)_{1\leq i\leq N,t>0}$ given by

$$\begin{aligned} <dW_t^i,dW_t^j> &= \rho_{ij}dt \\ \rho_{ii} &= 1 \text{ and } -1\leq\rho_{ij}\leq 1 \end{aligned}$$

We consider a basket B_t of N stocks at time t given by

$$B_t=\sum_{i=1}^N w_i\frac{S_i(t)}{Ref_i(0)}$$

and let the payoff of a call option on the basket at maturity T and strike K satisfies

$$\big(B_T-K\big)^+=\big(\sum_{i=1}^N w_i\frac{S_i(T)}{Ref_i(0)}-K\big)^+$$

with ATM strike $K=\sum_{i=1}^N w_i\frac{S_i(0)}{Ref_i(0)}$. In the spot model the payoff of the basket option simplifies to

$$\Big(\sum_{i=1}^N w_iC^i(0,T)\frac{Y_T^i}{Ref_i(0)}-(K+\sum_{i=1}^N w_i\frac{D_T^i}{Ref_i(0)})\Big)^+$$

5.1.2 The impacts of discrete dividends

5.1.2.1 The limits of the no-arbitrage conditions

Derivatives prices depend on the fundamental law of the market which states that in a very liquid market with no transaction cost, no limitation on buying or selling the underlying asset, there is no arbitrage opportunities. As a result, we can deduce the price of a few financial products independently from a model. For example, we let F(t;S,T) be the forward contract initiated at time t to get the spot price S delivered at maturity T. To guarantee S at time T, we can either buy the spot S at time t and keep it till maturity, or buy the forward contract at time t. Discounting the forward price and equating the two strategies, we recover the price of the forward contract today. But, the price of the underlying must be bounded and one must be able to buy or sell the underlying at all time t. However, in the case of a dividend paying stock, the stock holder gets the dividend but not the holder of the forward contract, and similarly in case of shortage, the stock holder has a clear advantage over the holder of the forward contract. In general, in a market where the holder of the underlying has a benefit, or a cost, over the holder of the forward contract, the no-arbitrage condition no-longer holds. This is the case when a commodity is non-storable, such as electricity, or when the stock pays a dividend, and it can be extended to illiquid markets.

5.1.2.2 The forward price

We introduced in Section (1.1.1.2) the put-call parity in the ideal world of no repo rates and no discrete dividends. However, in a more realistic world with repo rates and discrete dividends, the classical forward price and call-put parity must be modified. We let the underlying stock pays disctrete dividends d_i, $i = 1, ..., N$, at time $t_1, ..., t_N$ in the period $[t, T]$, as seen in Figure (5.1). We want to compute the fair strike of the forward contract using replication arguments. We can reinvest all proceeds from holding the stock (repo and proportional dividends) to buy more shares, and forward-sell the cash dividends we would receive from holding these shares (see Bermudez et al. [2006]).

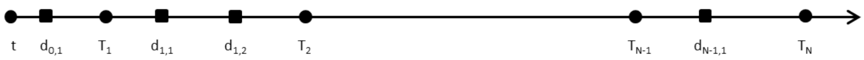


Figure 5.1: Example timeline of forward contracts

As explained in Appendix (B.2), in the spot model the forward price is given by

$$F(t,T) = S_t C(t,T) - D(t,T) = S_t \frac{Re(t,T)}{P(t,T)} - \sum_{t_i \in [t,T]} d_i \frac{Re(t_i,T)}{P(t_i,T)} \tag{5.1.7}$$

where $D(t,T) = \sum_{t_i \in [t,T]} d_i \frac{Re(t_i,T)}{P(t_i,T)}$. For indices, we usually use an approximation by projecting the discrete dividends to a continuous dividend yield, that is, S_t pays continuously $qS_t dt$. Note, since the share price S_t can not become negative, it implies that the forward price must be positive. Merton [1973] first noted that reducing a call option by the amount of any dividend paid on the stock could reduce the price of the stock and the value of the option to zero no matter what adjustment was made in the exercise price of the option. He pointed out that there maybe was no adjustment in terms of the option capable of providing adequate protection against a large dividend. From the definition of the forward price above, we can get the condition

$$S_t \geq \frac{D(t,T)}{C(t,T)} = D_{PV}(t;t,T)$$

meaning that the price process $(S_t)_{[t,T]}$ is floored by the deterministic process $D_{PV}(t;t,T)$. Hence, the stock price must always exceed the discounted expected value of all forthcoming dividends. As a result, the stock price can not be modelled with a random process which can be arbitrarily close to zero. When modelling the stock price, we must separate the volatility risk of the stock from the other charactersistic of the equity, which is what we did in the spot model. Alternatively, Buehler [2008] proposed to modify the spot price as $S_0^* = S_0 - D_{PV}(0;0,T)$ and obtained the formula

$$S_t^* = S_0^* C(0,t) Y_t + D_{PV}(t;t,T)$$

where $S_0^* C(0,t) Y_t$ is a pure equity process.

5.1.2.3 The put-call parity

We now consider the put-call parity where we are long a European call option and short a European put option. In the presence of repo rates and discrete dividends, we can always buy one stock and borrow at time t the quantity $\sum_{i=1}^N d_i \frac{P(t,t_i)}{Re(t,t_i)} + KP(t,T)$ to recover the flux $(S_T - K)$ at maturity T. Further, the put-call parity must satisfies

$$C(t,S_t,K,T) - P(t,S_t,K,T) = P(t,T)E_t[S_T - K]$$

and using the definition of S_T in our model, it simplifies to

$$\begin{aligned} C(t,S_t,K,T) - P(t,S_t,K,T) &= P(t,T)\big(S_t C(t,T) - D(t,T)\big) - KP(t,T) \\ &= S_t Re(t,T) - P(t,T)D(t,T) - KP(t,T) \end{aligned} \quad (5.1.8)$$

Dividing Equation (5.1.8) by the zero-coupon bond $P(t,T)$, we get the put-call parity expressed as

$$\overline{C}(t,S_t,K,T) - \overline{P}(t,S_t,K,T) = S_t \frac{Re(t,T)}{P(t,T)} - D(t,T) - K \quad (5.1.9)$$

where $\overline{C}(t,S_t,K,T) = \frac{C(t,S_t,K,T)}{P(t,T)}$ and $\overline{P}(t,S_t,K,T) = \frac{P(t,S_t,K,T)}{P(t,T)}$ are the discounted call price and discounted put price, respectively.
Note, the put-call parity only holds for European options and not for American options because the latter can be exercised at any time prior to the expiration date. The right to exercise options early implies that the value of an American option can never drop below its intrinsic value. The time at which the holder decide to exercise the option depends on the value of the underlying stock price. For American options on non-dividend paying stocks, one can show that the following inequality holds

$$S_t \frac{Re(t,T)}{P(t,T)} - K \frac{1}{P(t,T)} \leq \overline{C}_A(t,S_t,K,T) - \overline{P}_A(t,S_t,K,T) \leq S_t \frac{Re(t,T)}{P(t,T)} - K \quad (5.1.10)$$

In presence of discrete dividends the inequality becomes

$$S_t \frac{Re(t,T)}{P(t,T)} - D(t,T) - K \frac{1}{P(t,T)} \leq \overline{C}_A(t,S_t,K,T) - \overline{P}_A(t,S_t,K,T) \leq S_t \frac{Re(t,T)}{P(t,T)} - K$$

5.1.2.4 Inferring the implied yield from market prices

When using a model to discribe the dynamics of the spot price, we first need to estimate the continuous dividend yield, or repo rate, that we place in our model in view of matching the prices and volatilities observed in the market. Even though it is not uncommon in the equity market to observe a yield, or repo rate of zero across all times to maturity, it may result in discrepancies between the quoted prices and the implied forward price required to satisfy the put-call parity relationship. Therefore, we "mark-to-market" the repo curve by comparing the synthetic forward price implied by the put-call parity relationship with the theoretical value of a forward for an equity paying discrete dividends. Specifically, for an equity stock paying discrete dividends d_i, $i = 1,...,N$ between inception time t, and maturity

T, the theoretical value of a forward contract between times t and T is given by Equation (5.1.7), where $t_i = t_{d_i}$ represents the ex dividend date of the i^{th} dividend. We assume a series of forward contracts, increasing in maturity, such as that seen in Figure (5.1). We assume that there exist N forward contracts, each with expiry T_i, $i = 1, ..., N$ between times t and T_N. Further, between each forward maturity T_i, the equity in question may also pay dividends. However, the theoretical price of each forward contract does not allow for clear identification of the repo rate in each period. We therefore seek an alternative representation to that given in Equation (5.1.7) by allowing for separation of the repo factors. Note initially that the value of the first forward, expiring at date T_1 remain unchanged. Then, due to the multiplicative nature of the repo and discount factors when rates are deterministic, and by splitting the dividend payments, we can rewrite the value of the second forward contract as

$$\begin{aligned} F(t,T_2) &= S_t \frac{Re(t,T_1)Re(T_1,T_2)}{P(t,T_1)P(T_1,T_2)} - \sum_{t_{d_i} \in [t,T_1]} d_i \frac{Re(t_{d_i},T1)Re(T_1,T_2)}{P(t_{d_i},T_2)} - \sum_{t_{d_i} \in [T_1,T_2]} d_i \frac{Re(T_1,T_2)}{P(t,T_2)} \\ &= F(t,T_1) \frac{Re(T_1,T_2)}{P(T_1,T_2)} - \sum_{t_{d_i} \in [T_1,T_2]} d_i \frac{Re(t_{d_i},T_2)}{P(t_{d_i},T_2)} \end{aligned}$$

Hence we are able to express the T_2 maturity forward in terms of the T_1 maturity forward, and in doing so separate the repo factors in the two time periods. Similarly, we can iterate this process through to time T_N, resulting in the recursive relationship

$$F(t,T_N) = F(t,T_{N-1}) \frac{Re(T_{N-1},T_N)}{P(T_{N-1},T_N)} - \sum_{t_{d_i} \in [T_{N-1},T_N]} d_i \frac{Re(t_{d_i},T_N)}{P(t_{d_i},T_N)} \tag{5.1.11}$$

This completes one side of the argument.
In the absence of liquid market prices for forward contracts, we must construct synthetic forwards from the quotes on call and put options, by making use of the put-call parity relationship in Equation (5.1.9). Taking the strike K over to the left hand side of the equation, we are left with the synthetic forward price

$$F^s(t,T) = \overline{C}(t,S_t,K,T) - \overline{P}(t,S_t,K,T) + K$$

All values on the right hand side are observable in the market, yielding a synthetic forward with which we can compare our theoretical forward. The final step is to match the synthetic forward values with the theoretical forward values parameterised by the repo rate, q. Assuming piecwise constant repo rate

$$q(T_i) = \frac{1}{T_i} \sum_{j=1}^{i} (T_j - T_{j-1}) q_{T_{j-1},T_j}$$

the repo factor satisfies $Re(0,t) = e^{-q(t)t}$. We set $F^s(t,T) = F(t,T,q)$, and find $q_i = q(T_i)$ such that

$$G(t,T_i,q_i) = F^s(t,T_i) - F(t,T_i,q_i) = 0 \quad \text{for} \quad i = 1, ..., N$$

This is a relatively simple numerical task given the recursive relationship for the forward in Equation (5.1.11), and it can be solved using any root finding method. Newton's method works adequately well, producing convergence in fractions of a second with the right initial guess. To do so, we Taylor expand the repo factor to the first order, getting $e^{-q(t)t} \approx 1 - q(t)$ and replace it in the theoretical forward equation. Letting $\widehat{F}(t,T)$ be a theoretical forward price with no repo rate

$$\widehat{F}(t,T) = \frac{S_t}{P(t,T)} - \sum_{t_i \in [t,T]} \frac{d_i}{P(t_i,T)}$$

and setting

$$F(t,T_i) = F(t,T_{i-1})\frac{e^{-q_{T_{i-1},T_i}(T_i-T_{i-1})}}{P(T_{i-1},T_i)} - \sum_{t_{d_j}\in[T_{i-1},T_i]} d_j \frac{e^{-q_{T_{i-1},T_i}(T_i-t_j)}}{P(t_j,T_i)}$$

we can approximate the repo rate q_{T_{i-1},T_i} for the period $[T_{i-1},T_i]$ as

$$q_{T_{i-1},T_i} \approx -\frac{F^s(t,T_i) - \widehat{F}(t,T_i)}{\widehat{F}'(t,T_i)}$$

where $\widehat{F}'(t,T) = \frac{S_t}{P(t,T)}(T-t) - \sum_{t_i\in[t,T]} \frac{d_i}{P(t_i,T)}(T-t_i)$.

5.2 The difficulties associated with American options

5.2.1 Accounting for American options

5.2.1.1 The markets trading American options

In practice, most exchange traded options (stock and futures options) are American rather than European style, while financial index options can be issued as either American- or European-style options. Among the broad-based indices, only limited indices such as the $S\&P$ 100 have American-style options. Major broad-based indices, such as the $S\&P$ 500, have very actively traded European-style options. Option contracts traded on futures exchanges are mainly American-style, whereas those traded over-the-counter are mainly European. Commodity options can be either style. All optionable stocks and exchange traded funds (ETFs) have American-style options. Currently, all equity options traded on US option exchanges, including LEAPS, are American-style, as are certain index options. For example, stock options issued for companies like Microsoft and Sony, are American-style options. Trading volume provides important insight into the strength of the current market direction for the option's underlying stock. The volume, or market breadth, is measured in shares and tells us how meaningful the price movement in the market is. We present statistics depicting the number of contracts for single stock options traded at selected stock exchanges worldwide in 2009. International Securities Exchange (ISE) (672,429,815), Chicago Board Options Exchange (CBOE) (634,710,477), Eurex (146,286,452), NYSE Liffe (European Markets) (141,604,421), Montreal Exchange (14,507,261), Tokyo Stock Exchange Group (660,875), MexDer (345,718), Tel Aviv SE (321,735).

5.2.1.2 The implied American volatility

Given $C_E(t,S_t)$ the price of a European call option defined in Equation (1.7.38) and $C_A(t,S_t)$ the price of an American call option defined in Equation (4.2.13), then we can write the American price as

$$C_A(t,S_t) = C_E(t,S_t) + \mathcal{A}_\epsilon \tag{5.2.12}$$

where $\mathcal{A}_\epsilon$ is the value of the early exercise premium (EEP). Since market prices are convex functions with respect to the underlying (see Section (1.4.2)), so must be the early exercise premium A_ϵ. We can then define the American implied volatility as follow

Definition 5.2.1 *We define the American implied volatility as the level of volatility we need to plug into a basic pricer (bp) to recover the market prices of the option.*

Hence, given the American option price $C_A(t,S_t,K,T)$ at time t for a strike K and a maturity T, the market implied volatility $\Sigma_A(t,S_t;K,T)$ satisfies

$$C_A(t,S_t,K,T) = C_{bp}(t,S_t,K,T;\Sigma_A(K,T)) \tag{5.2.13}$$

where $C_{bp}(t, S_t, K, T; \sigma)$ is the basic pricer for an American option under the univariate lognormal model (UL).
As explained in Section (5.1.2.3), the put-call parity exists for European vanilla option prices, where the call option price can be calculated from the put option price with the same parameters, but it does not hold for American options. Thus, one can not infer the market forward price from American options. Further, the American smile for call options does not equal the one for put options. We give an idea of how to prove that result. Given the decomposition of the American call option in Equation (5.2.12), if we assume that the early exercise premium $\mathcal{A}_\epsilon$ is small with respect to $C_A(t, S_t)$, then we can perform a Taylor expansion on prices (see Appendix (B.3.1)) to approximate the American implied volatility. Given the volatility decomposition in Equation (B.3.14), we can approximate the American volatility as follow

$$\Sigma_A(K,T) \approx \Sigma_E(K,T) + \frac{1}{Vega(\Sigma_E(K,T))}\mathcal{A}_\epsilon + ...+ \quad (5.2.14)$$

where $Vega(\sigma) = \partial_\sigma C_A(t, S_t)$. Since it may be optimal to exercise immediately American put options, which is not the case for American call options, the early exercise premium for call options differ from that for put options. Thus, the American smile for call options is not the same as that for put options.

5.2.1.3 Approximating American options

Barone-Adesi et al. [1987] provided the first standard analytical approximation to American options. Bjerksund et al. [1993] proposed an algorithm that compares favourably in accuracy and speed to the Barone-Adesi et al. quadratic approximation. Ingersoll [1998] proposed an approximation method based on barrier options using constant and exponential barriers. Considering a gradient search method over a few variables to get the maximum of a convex function, he obtained tight lower bounds to the true option values. Later, Ju [1998] and Ju et al. [1999] approximated American option by computing the sum of a European option and an early exercise premium estimated by solving a quadratic approximation to the partial differential equation (PDE). Infering the American implied volatility, we remove the estimated early exercise premium from the American option price and get an estimate of the European option price. Corrado et al. [1996c] considered a univariate Edgeworth expansion to fit the probability density function of the option by employing higher-order moments. Flamouris et al. [2002] proposed to apply this method to infer the implied density of American options.

5.2.1.4 European versus American volatility

The payoff of Euopean options is entirely driven by the ending asset price distribution, so that the total variance of the asset over the life of the option determines the value of a European option. However, for American options, the option value is determined by its payoffs on the early boundary as well as at its maturity (see Section (4.2.5)). Hence, the total variance of the asset over the life of the option does not properly reflect the fraction of option value derived from the early exercise boundary. For instance, due to early exercise, the value of an American option is equal, or close, to its intrinsic value if it is sufficiently in-the-money (ITM) when early exercise is more likely to occur. As a result, the value of the American option is no-longer sensitive to changes in the volatility of the underlying asset. On the other hand, for out-of-the-money (OTM) options, the probability of early exercise is much lower, leading to a smaller difference in implied volatility between American and European options. Therefore, we can use OTM options to minimise the impact of these problems. Further, OTM options is advantageous because the implied volatility calculation is more reliable due to higher sensitivity of option price to changes in volatility.

5.2.2 Estimating the implied risk-neutral density

5.2.2.1 Describing the problem

Since the only unobservable quantity in the BS-formula (see Equation (1.2.6)) for a European call price is the volatility, satisfying the put-call parity in Equation (5.1.8) leads

- to no-arbitrage opportunity, and
- an implied volatility (IV) for a European call option that is the same as that for a European put option.

That is, the put-call parity in the BS-formula holds if and only if $d_1^{call} = d_1^{put}$ which leads to $\sigma_{BS}^{call} = \sigma_{BS}^{put}$. Even though different implementations of the BS-formula, due to different methods of handling dividends, result in different implied volatility surfaces, we expect the volatility surfaces to be very similar in shape (see Haugh [2009]). The seminal work of Breeden et al. [1978] on option-implied risk-neutral distribution paved the way to implied trees, model-free implied volatility and implied skewness and kurtosis. However, this option-implied approach requires market prices of European options across a wide range of strike prices. This is because the absence of butterfly arbitrage, which corresponds to non-negative density (see Section (1.7.2.4)), only applies to market prices of European options across a wide range of strike prices. In addition, as explained in Section (5.1.2.3), the put-call parity exists for European vanilla option prices, where the call option price can be calculated from the put option price with the same parameters, but it does not hold for American options. As a result, when considering American options there is no guarantee that the implied volatility (IV) call equals the IV put, leading to two different smiles of volatility for a given maturity with the largest discrepancy being around-the-money forward. As an example of this discrepancy, we plot in Figure (5.2) a European smile versus an American smile. It is therefore pointless to fit an American implied smile with a polynomial model. However, when systematicall trading options we rely on an option-implied model to fit the implied volatility surface. We must therefore modify the American prices to recover a single, smooth, smile maturity per maturity. A similar problem arise when one has to infer the option-implied risk-neutral distribution and requires the European option prices in the estimation of these moments. Hence, the inability to handle American options may prevent the application of the option-implied methodology to a broader range of option classes.

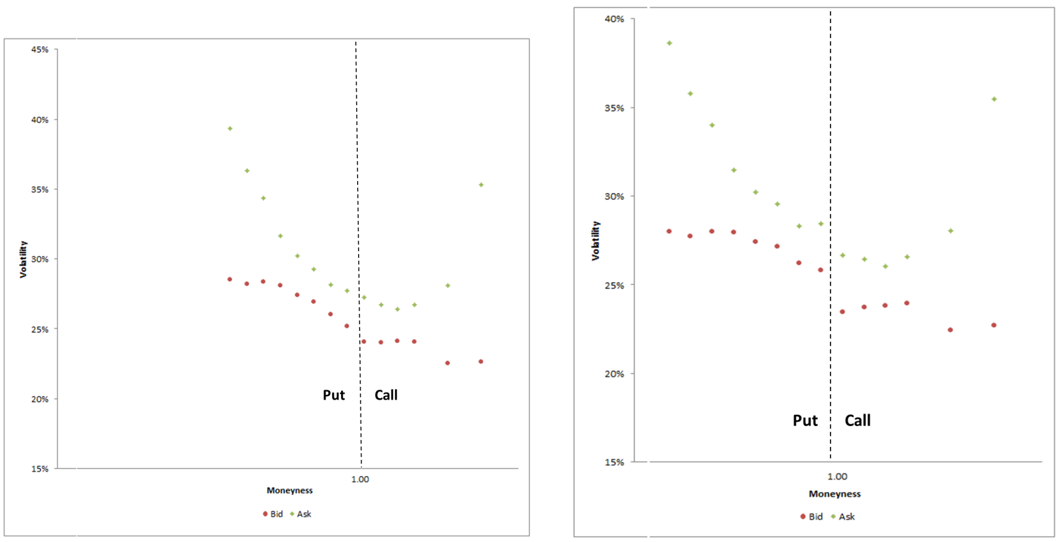


Figure 5.2: Market Implied Volatilities: European vs. American options

5.2.2.2 Market practice

Since American call and put prices are at least as valuable as their European counterparts, one can first extract synthetic European option prices from American option prices, and then apply the usual implied method to the extracted European option prices. A common approach to extracting European option prices $C_E(t, S_t)$ from American prices is to use the CRR binomial tree (see Section (1.3)) as a basic pricer. We first build a CRR tree to price American option for a given maturity and strike, and back out the corresponding American implied volatility using Equation (5.2.13). We then use this American implied volatility in the same CRR tree, but without the American feature, to determine the price of the corresponding European option. The only assumption made in this approach is that the American implied volatility is identical to the European implied volatility. An alternative approach is to use an analytical approximation to extract the early exercise premium. Even though these methods are simple to implement, they do not work when the options are deep in-the-money, and they treat each option independently. In the former approach we use the CRR binomial tree as a tool to map out the implied volatility smile across American options. However, in the latter approach we rely on the lognormality assumption, which is not consistent with the market. Tian [2010] proposed an accurate method for extracting risk-neutral density and its moments from American option prices by using an iterative Implied Binomial (iIB) tree. Calculating the early exercise premiums on the American options, they are then removed from the market prices of the American options until the extracted early exercise premiums converge. We are now going to present that method.

5.2.2.3 The iterative implied binomial tree

Using a single binomial tree to match simultaneously the market prices of all American options for a given maturity, Tian [2010] proposed an iterative approach (iIB) which is simpler to implement and quite accurate. Given an initial guess of the European option prices, we fit a set of ending nodal probabilities, obtaining the fitted ending nodal probabilities necessary to build an IB tree. Both American and European options are computed for a given maturity, and their difference at each strike is calculated, obtaining an estimate of the early exercise premium (EEP). Subtracting the EEP from each observed American option price, we get a refined estimate of the European option price, which are used in turn to obtain a new set of ending nodal probabilities. Repeating the process, we estimate a refined EEP leading to new estimated European option prices. We continue this iterative procedure until the estimated EEP converge. The measure of convergence chosen is the sum of squared incremental changes in EEP between two iterations. Note, the American option prices are chosen as initial guess, and turn out to be quite robust as they act as implicit control variate. Considering OTM options to recover European options from American ones, Tian [2010] showed that the iIB and CRR models had good results on prices. Translating the pricing errors into differences in BS implied volatilities, the volatilitiy smiles produced were virtually indistinguishable, which were very close to the true volatility smile. He also investigated the numerical performance across option maturities and different model parameters and found that both models were consistently accurate as long as OTM call and put options were used. Similar results were obtained on traded options from the XEO for European options, and from the OEX for American options. The difference between the iIB and CRR models essentially occured for ITM options were the American implied volatility is not identical to the European implied volatility.

5.2.2.4 Approximating the implied risk-neutral density

We can assume that there exists a process $X_t = \{X_t \in \mathbb{R}^+, t \geq 0\}$ with continuous density $f_X(\cdot)$, such that the expected value of the payoff function $h(X_t)$ under the risk-neutral measure $\mathbb{Q}$ will be equal to the American option. That is,

$$\begin{aligned} C_A(t, S_t) &= C_E(t, X_t) \\ &= P(t,T) E^Q[h(X_T)|\mathcal{F}_t] \\ &= P(t,T) \int_{-\infty}^{\infty} h(x) \phi_X(t,x) dx \end{aligned}$$

where $C_E(t, X_t)$ is a synthetic European option with S_t is replaced by X_t. We are therefore interesed in finding the density $\phi_X(t, \cdot)$, called the implied risk-neutral density (IRND) of the American option, such that

$$C_E(t, X_t) = C_E(t, S_t) + \mathcal{A}_\epsilon \tag{5.2.15}$$

Note, given the market anomalies, many different densities $\phi_X(t,\cdot)$ can satisfy Equation (5.2.15). That is, the implied risk-neutral density is model dependent.

***Remark* 5.2.1** *In the Black-Scholes framework, the only degrees of freedom we have to find a process X_t satisfying Equation (5.2.15) is to either play with the repo rates or with the volatility.*

Assuming that the European options $C_E(t,S_t)$ are known, several authors proposed to use the European density $\phi_S(t,\cdot)$ to approximate the IRND density $\phi_X(t,\cdot)$. Using the multivariate extension of the Edgeworth expansion (MGEE) (see Jarrow et al. [1982]), Arismendi [2013] proposed to approximate the density as follow

$$\phi_X \approx \phi_S + \sum_{j=1}^{2} M_{[l_1,[l_2]]} \frac{(-1)^j}{j!} \frac{\partial^j}{\partial_{s_{[l_1]}} \partial_{s_{[l_2]}}} \phi_S \tag{5.2.16}$$

where the tensor notation is

$$M_{[l_1,[l_2]]} \frac{(-1)^j}{j!} \frac{\partial^j}{\partial_{s_{[l_1]}} \partial_{s_{[l_2]}}} \phi_S = \sum_{l_1=1}^{n} \Big(M_{l_1}(-1) \partial_{s_{l_1}} \phi_S + \sum_{l_2=1}^{n} \big(M_{l_1,l_2}(\frac{1}{2}) \partial_{s_{l_1} s_{l_2}} \phi_S \big) \Big)$$

where M_{l_1} and M_{l_1}, l_2 are the differences between the first and second-order moments of the risk-neutral densities $\phi_X(t,\cdot)$ and $\phi_S(t,\cdot)$. There are two ways of considering Equation (5.2.16):

- D1 The density $\phi_S(t,\cdot)$ is known, we can then estimate the difference of the cumulants $M[l_1,[l_2]]$ and compute the implied density $\phi_X(t,\cdot)$.
- D2 The density $\phi_S(t,\cdot)$ is not known, but the density $\phi_X(t,\cdot)$ and the difference of the cumulants $M[l_1,[l_2]]$ are known. We can then approximate the density $\phi_S(t,\cdot)$.

Tian [2010] considered the problem $D2$ in the univariate case. Arismendi et al. [2014] considered the problem $D1$ in the multivariate case. As discussed in Section (5.2.2.1), single stock options are American, while European options are OTC. Thus, in practice one do not know the density $\phi_S(t,\cdot)$ and can not solve the problem $D1$.

5.2.3 Inferring the implied dividends from American options

As explained in Section (5.1.2.3), the put-call parity does not hold for American options. Thus, for a fixed maturity T, one can not infer the market forward price $F(t,T)$ from American option prices. By construction, a univariate lognormal model pricing American options, such as a CRR tree or an approximation, assume a single forward price per maturity. It is based around the knowledge of the forward price. In these models, since the only degree of freedom to match American option prices is the volatility, then the American implied volatility for call options obtained by solving Equation (5.2.13) can not be identical to that for put options. We must therefore find ways around to obtain a smooth volatility smile from the market call and put prices. The solutions discussed depend on the model chosen for discrete dividends and the way the put-call parity is enforced.

5.2.3.1 ATM American options

The solution presented in Section (5.1.2.4) to infer the implied dividend yield is often sufficient to remove any discrepancy in the put-call parity relationship for European options, since that relationship holds for any given strike on a fixed expiry. However, as explained in Section (5.1.2.3), since the put-call parity does not hold for American options, the implied volatility for an American call option is not necessarily the same as that for an American put option. At best, a relationship exists in the form of a bounded region for the difference between the prices of call and put options, for a given strike and time to maturity (see Equation (5.1.10)). Hence, if we use a CRR binomial tree to infer the American implied volatility strike by strike for a fixed maturity, we still obtain a discontinuity in the smile around the money forward. That is, the inequality in Equation (5.1.10) is not strong enough to guarantee the equality of implied volatility smile. Nonetheless, practitioners assume that the put-call parity holds for ATM American options within reasonable approximation. Ignoring discrete dividends, the implied dividend d_{imp} is computed as follow

$$d_{imp}(t,T) = -\frac{1}{T-t} \ln \Big[\frac{P(t,T)}{S_t} \big(\overline{C}_A(t,S_t,K_{ATM},T) - \overline{P}_A(t,S_t,K_{ATM},T) + K_{ATM} \big) \Big]$$

Difficulty arises when we can not find American options in the market with strike close enough to K_{ATM}. In the single stock market, there does not exist a continuum of quoted option prices for all possible strike prices, meaning that it is often unlikely to find the relevant option prices. One way forward is to linearly interpolate the option prices to achieve a synthetic price at-the-money. Although linear interpolation is a rudimentary technique, the option prices should behave almost linearly at-the-money. Doing so, we should expect some appreciable improvement in the smile when implemented, but this is not the case because the non-existence of the put-call parity applies to all strikes and not just the ATM one. Thus, the European options that we would recover from this smile would not satisfy the put-call parity and we would not be able to apply the option implied methodology to infer the risk-neutral distribution, or equivalently, the volatility surface.

5.2.3.2 The implied forwards and dividends

To compute an equity implied volatility surface for American options, Cui et al. [2011] proposed to transorm these options to their equivalent European prices, estimate the market forward price, and then modify the dividends to match that forward. Given a list of discrete dividents (dates and cash amounts) and a yield curve, the model forward is computed. Then, for a fixed maturity T, using Equation (5.2.13), the American implied volatilities $\Sigma_A(T,K)$ are estimated from American call and put prices. At this stage, the American call smile is not identical to the American put smile. Using the model Forward, and assuming that the American implied volatility surface is equal to the European one, the European call and put prices are calculated. Since in that setting $\sigma_{BS}^{call} \neq \sigma_{BS}^{put}$, the put-call parity for these European options is not satisfied, and one obtain the synthetic forward price at strike K_i, given by

$$F^s(t,T,K_i) = \overline{C}_E(t,S_t,K_i,T) - \overline{P}_E(t,S_t,K_i,T) + K_i \;, \forall i$$

For a fixed maturity T, the median of the forward prices is calculated, and the market forward price is estimated by averaging n forward prices around that median

$$\widehat{F}(t,T) = \frac{1}{n} \sum_{i=1}^n F^s(t,T,K_i) \;, F_{med} = F^s(t,T,K_{int(\frac{n}{2})})$$

where $int(x)$ is the integer part of x [1]. If the error between the estimated forward and the model forward, $\epsilon = \widehat{F}(t,T) - F(t,T)$, is greater than a threshold, the listed dividends are modified such that the model forward matches the estimated one. Having a single forward price per maturity for European options, we get the modified implied volatility $\widehat{\Sigma}_E(T^*,K)$. This approach does not guarantee that the put-call parity is matched at all strikes. Further, if we need to interpolate or extrapolate the European volatility surface for a pair (T,K), it becomes very difficult to recover the associated American price. For example, given listed options with expiries $T_1 < T_2 < ... < T_N$, we can use the model of our choice to interpolate the modified European implied volatility $\widehat{\Sigma}(T^*,K)$ for the pair (T^*,K) with $T_i < T^* < T_{i+1}$. However, we can not perform reverse engineering to recover the price of the equivalent American option, since we can interpolate the forward price $\widehat{F}(t,T^*)$ but we do not know the synthetic forward prices $F^s(t,T^*,K_i)$ composing that price. Thus, we can not convert the modified implied volatility $\widehat{\Sigma}_E(T^*,K)$ back to the American implied volatility $\Sigma_A(T^*,K)$, hence, we can not recover the associated American price.

[1]
$$int(x) = \begin{cases} \lfloor x \rfloor \text{ for } x \geq 0 \\ \lceil x \rceil \text{ for } x < 0 \end{cases}$$

where $\lfloor x \rfloor$ and $\lceil x \rceil$ are the floor and ceiling functions.

5.2.3.3 The implied repo rates

Rather than estimating the American volatilities, computing European prices with different market forward prices for a fixed maturity and averaging them to estimate the forward price, we can directly modify the dividends in the American pricer to recover a single smile. This is done by forcing the put-call parity strike by strike. We can compute the synthetic forward price at strike K_i, given by

$$F^s(t,T,K_i) = \overline{C}_A(t,S_t,K_i,T) - \overline{P}_A(t,S_t,K_i,T) + K_i \; , \forall i \tag{5.2.17}$$

where $\overline{C}_A(t,S_t,K_i,T)$ and $\overline{P}_A(t,S_t,K_i,T)$ are, respectively, the discounted American call price and the discounted American put price at maturity T.
We saw in Section (1.6.1.3) that practitioners accounted for stochastic volatility by using one BS-model for every pair (T,K), leading to the generation of a non-flat implied volatility surface (IVS). Similarly, to account for the fact that American options generate different forward prices for a fixed maturity, practitioners introduced the implied repo rates $q(T,K)$ as a function of maturity and strike. This way, they forced the put-call parity in time and space by adjusting the repo rates strike by strike for a fixed maturity, obtaining the implied repo surface

$$(T,K) \to q(T,K)$$

If we want to implement path-dependent options on single stocks we need to devise a stochastic process for the repo rates. One solution would be to assume that the repo rates follow the local function $q(t,S_t)$ (in the sense of the local volatility), and derive that function to match the synthetic forwards prices. However, since the respective call and put smiles in American options result from the inclusion of discrete dividends, with the discrepancy appearing only after the first ex-dividend date, adjusting the continuous repo rates is not an ideal solution due to the impact it also has on the theoretical forward through the multiplying factor on the stock price.

5.2.3.4 The implied cumulative dividend

A better solution consists in considering the Spot model described in Section (5.1.1.2) with stochastic discrete dividends, and dynamics of the pure diffusion process given in Equation (5.1.5). The strength of the model is that we do not diffuse directly the underlying process S_t, but the pure diffusion process Z_t, and that we shift the strike with a step function including the past discrete dividends capitalised to time t. In our setting, the cumulative dividend is now a function of time and an external source of risk ω, given by

$$D(t,T) = \sum_{t_{d_i} \in [t,T]} d_i(t,\omega) \frac{Re(t_{d_i},T)}{P(t_{d_i},T)}$$

where $d_i(t,\omega)$ is a stochastic discrete dividend. Thus, we now need to infer the implied cumulative dividends $D(t,T,K_i)$ for all strikes K_i such that the put-call parity is satisfied. Specifically, given the definition of the theoretical forward price in Equation (5.1.7), we get

$$D(t,T,K_i) = S_t \frac{Re(t,T)}{P(t,T)} - F(t,T,K_i) \; , \forall i \tag{5.2.18}$$

where $D(t,T,K_i)$ is the implied cumulative dividends between time t and T. The model forward price $F(t,T,K_i)$ is calculated by using the synthetic forward price computed at strike K_i using Equation (5.2.17). Thus, we infer the implied cumulative dividend for each strike present in the market, and on each expiration date, getting the surface

$$(T,K) \to D(t,T,K)$$

Since we do not have to diffuse the process S_t, we do not need to compute the stochastic discrete dividends $d_i(t,\omega)$. Thus, given the diffusion process Z_t and the implied cumulative dividends $D(t,T,K_i)$, we can use Equation (5.1.6) for each strike, K_i, shifted by the amount $D(t,T,K_i)$ as follow

$$K_i^{'} = K_i + D(t, T, K_i) \;, \forall i \tag{5.2.19}$$

to estimate the modified American implied volatility $\widetilde{\Sigma}_A(T, K_i)$. The modified implied volatility is obtained by solving

$$C_A(t, S_t, K_i, T) = C_{bp}(t, Z_t, K_i, T; D(t, T, K_i); \widetilde{\Sigma}_A(T, K_i)) \;, \forall i$$

where $C_{bp}(t, Z_t, K, T; D(t, T, K); \sigma)$ is the basic pricer for the Spot model with strike shifted by $D(t, T, K)$. Following this method, we are able to ensure that our model matches the synthetic forward prices, enforcing the put-call parity relationship at each strike K_i and at each maturity T, and yielding a smooth market volatility smile. We are left with decomposing our model for American prices into European options and early exercise premiums. To do so, we need to define our forward model. We choose to use the initial list of discrete dividents and to modify the dividend yield such that the synthetic forward $F^s(t, T, K_i)$ around the forward moneyness $K_i \approx \eta = \frac{S_t}{P(t,T)}$ is matched. Since the modified implied volatility is now identical for American call and put options, and assuming that $\widetilde{\Sigma}_A(T, K_i)$ is the same for American and European options (in the case of OTM), we can then decompose the American price as

$$C_A(t, Z_t, T, K_i; D(t, T, K_i); \widetilde{\Sigma}_A(T, K_i)) = C_E(t, Z_t, T, K_i; P(t, T), Re(t, T), D(t, T); \widetilde{\Sigma}_A(T, K_i)) + \widetilde{\mathcal{A}_\epsilon} \;, \forall i$$

where $D(t, T)$ is the cumulative dividend computed from the list of discrete dividents, $Re(t, T)$ is the repo rates matching the forward price, and $\widetilde{\mathcal{A}_\epsilon}$ is a modified early exercise premium. Plugging $\widetilde{\Sigma}_A(T, K_i)$ in a CRR tree, without the American feature, we can then calculate the synthetic European options $C_E(t, Z_t)$. An estimate for the early exercise premium (EEP) is obtained by taking the difference between the American price and the synthetic European price. Then, we can implement the option-implied approach on the synthetic European options. For example, given listed options with expiries $T_1 < T_2 < ... < T_N$, we can use the model of our choice to interpolate the European implied volatility for the pair (T^*, K) with $T_i < T^* < T_{i+1}$. We perform reverse engineering to recover the price of the equivalent American option. We interpolate the surface of calibrated implied cumulative dividend at (T^*, K), set the repo rates and the discrete dividends to zero, and compute the price as follow

$$C_A(t, S_t, K, T^*) = C_{bp}(t, Z_t, K, T^*; D(t, T^*, K); \widetilde{\Sigma}_A(T^*, K))$$

As an example, we consider options on BASF with evaluation date $18/01/16$, and we graph in Figure (5.3) the European smile at maturity $15/12/17$ and its associated American prices.

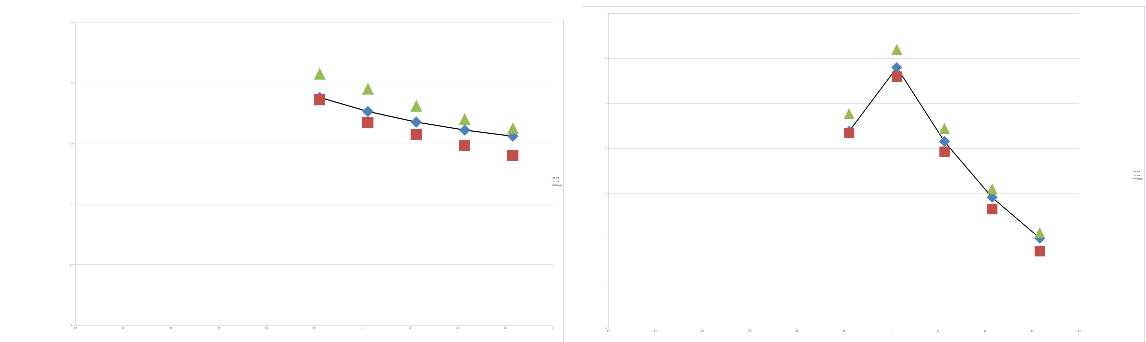


Figure 5.3: European Implied Volatilities and associated American prices

5.3 Pricing and hedging in incomplete markets

5.3.1 Accounting for unhedgeable risks

We saw in Section (1.2) that option prices evaluated using the Black-Scholes formula were subject to strict explicit assumptions. However, we saw in Section (5.1) that such a market does not exist in presence of discrete dividends, and that it is not possible to buy and sell continuously any amount of the underlying. Consequently, in financial markets, perfect hedges do not exist and option prices induce market risks called gamma risk and vega risk whose order of magnitude is much larger than market corrections such as transaction costs and other imperfections. For instance, when the stock prices exhibit large downwards jumps the traders have no time to delta hedge their book against these market movements, and experience losses. In general, these risks can not be hedged away even in continuous time trading, and hedging becomes approximating a target payoff with a trading strategy (see Section (1.1.3)). The value of the option price (see Equation (1.1.2)) is thus the cost of the hedging strategy plus a risk premium required by the seller to cover his residual risk which is unhedgeable. The no-arbitrage pricing theory tells us about the first component of the option value, while the second component depends on the preferences of investors (subjective). Thus, the unhedgeable portion is a risky asset and one must decide how much he is willing to pay for taking the risk. The no-arbitrage argument implies a unique price for that extra risk called the market price of risk. However, in incomplete markets, this risk is no-longer unique. Hence, when pricing in incomplete market, the market price of risk enter explicitly the pricing equation leading to a distribution of prices rather than a single price, such that one must consider bounds. Therefore, one can either simply ignore the risk premium associated to a discontinuity in the underlying, or one can choose any equivalent martingale measure as a self-consistent pricing rule, but in that case the option price does not correspond to the cost of a specific hedging strategy. Hence, one should first discuss a hedging strategy, and then derive a valuation formula for the options in terms of the cost of hedging plus an intrinsic risk premium.

By definition, a market is incomplete when some payoffs can not be replicated by trading in marketed securities (see Staum [2008]). One cause of incompleteness is the insufficient span of marketed assets, when a payoff does not depend uniquely on market prices. One of the most used pricing methods under market incompleteness is the calibration of a risk-neutral probability measure $\mathbb{Q}$ (see Section (1.1.3)). However, Staum [2008] stated that this approach wrongly prices the claim because it is using a fictitious replicating portfolio that can not be constructed in a real market. Moreover, in illiquid markets with large bid-ask prices, it is difficult to use market price of derivatives to calibrate the risk-neutral probability. As a result, we will assume an OTC [2] market, where the contingent claims are traded between two distinct entities. On one hand there is a hedger, or speculator, who wishes to trade the product, and on the other hand there is an OTC market-maker that will provide bid/ask prices to proceed the trade (see Staum [2008]). To define these prices we will then apply indifference pricing as a method combining wealth investors and investment bank methodologies (see Carmona et al. [2009]).

5.3.2 Pricing under the intrinsic price of risk

In view of explaining market anomalies, such as volatility smile and convexity adjustment on the forward price, Le [2014] postulated the existence of an intrinsic risk in every contingent claim. It lead to a unified framework to price and risk manage these claims under a risk-subjective measures. We consider the continuous true price process X with dynamics

$$dX(t) = \mu(t)X(t)dt + \sigma(t)X(t)dB(t)$$

where $B(t)$ is a Brownian motion, $\mu(t)$ and $\sigma(t)$ are Lipschitz continuous functions so that a solution exists. The risk-free money account satisfies the dynamics

$$dM(t) = \nu(t)M(t)dt$$

[2] Over-the-counter

where the interest rate, $\nu(t)$, is a Lipschitz continuous function. Expanding the portfolio, we get

$$dV(t) = \nu(t)V(t)dt + \alpha(t)\sigma(t)X(t)dW(t)$$

where $W(t)$ is a $\mathbb{Q}$-Brownian motion given by

$$dW(t) = \lambda(t)dt + dB(t)$$

with

$$\lambda(t) = \frac{\mu(t) - \nu(t)}{\sigma(t)}$$

The dynamics of the price process X under $\mathbb{Q}$ become

$$dX(t) = \nu(t)V(t)dt + \sigma(t)X(t)dW(t)$$

and the risk-neutral valuation is given by

$$V(t) = M(t)E^Q[\frac{1}{M(T)}H|\mathcal{F}_t] = M(t)E^Q[\frac{1}{M(T)}h(X(T))|\mathcal{F}_t]$$

In practice, market completeness and uniqueness of derivative prices are no-longer valid. Le [2014] defined the measure of intrinsic risk as follow

Definition 5.3.1 *The measure of intrinsic risk in a time interval dt is defined by $dG(t,T) = \xi(t,T)X(t)dt$, where $\xi(t,T)$ is a continuous adapted process representing a rate of intrinsic risk.*

That measure implies adding, or removing, capital in a time interval dt, leading the trading strategy to be adaptable, that is,

$$dX(t) = \nu(t)V(t)dt + \alpha(t)\sigma(t)X(t)dZ(t)$$

where $Z(t)$ is an S-Brownian motion given by

$$dZ(t) = \lambda^* dt + dB(t) = \frac{\xi(t,T)}{\sigma(t)}dt + dW(t)$$

with

$$\lambda^* = \frac{\mu(t) + \xi(t,T) - \nu(t)}{\sigma(t)}$$

and S is a measure equivalent to P, so that $S \in \mathcal{Q}$. Equivalently, $\frac{\xi}{\sigma}$ is the intrinsic price of risk. Thus, the formula for the price of a contingent claim becomes

$$V(t) = M(t)E^S[\frac{1}{M(T)}H|\mathcal{F}_t] = M(t)E^S[\frac{1}{M(T)}h(X(T))|\mathcal{F}_t]$$

One must therefore add extra sources of uncertainty that are epistemic (subjective) rather than purely random. In fact, traders and market makers do not fully observe the true price process X, but only part of it. We let $Y(t,T)$ be the price process of a contingent claim in the horizon $[0,T]$ with dynamics

$$dY(t,T) = \nu(t)Y(t,T)dt + \overline{\sigma}(t,T)Y(t,T)dZ(t)$$

where the misspecified volatility $\overline{\sigma}$ is a Lipschitz continuous function. It is an implied price process with misspecification in the underlying asset as in El Karoui et al. [1998]. Le showed that Z was an S-Brownian motion and that Y could be a replacement of the true price process X with dynmics

$$dX(t) = \big(\nu(t) - \xi(t,T)\big)X(t)dt + \sigma(t)X(t)dZ(t)$$

where

$$\xi(t,T) = \frac{\nu(t)}{\overline{\sigma}^2(t,T)}\big(\overline{\sigma}^2(t,T) - \sigma^2(t)\big)$$

We see that the hedging error introduced by El Karoui et al. [1998] in the complete market becomes an intrinsic price of risk, presented as traded asset, in the incomplete market theory. The measure S is subjective since the valuation of the claim depends on the exogenous measure of risk ξ. That is, it becomes explicit in the pricing equation

$$\partial_t V(t,x) + \frac{1}{2}\sigma^2(t)x^2\partial_{xx}V(t,x) + (\nu(t) - \xi(t,T))x\partial_x V(t,x) - \nu(t)V(t,x) = 0$$

with $X(t) = x$ and $V(T,x) = h(x)$. That measure is called the risk-subjective measure, and it implies that possible arbitrage exists in the market. Nonetheless, the growth of the portfolio value is still at the risk-free rate ν. While forward contracts are associated with the underlying asset, their prices should be obtained by assuming complete market and using Equation (). However ... such that they can be priced in incomplete market where the measure of intrinsic risk ξ corresponds to the convexity adjustment. Similarly, in presence of dividends, the put-call parity is not necessarily satisfied, suggesting that we can price vanila options in incomplete market by letting the measure of intrinsic risk be the dividend yield.

5.3.3 Indifference pricing

5.3.3.1 Introduction

Barrieu et al. [2009] developed indifference pricing as a method that depends directly on the counterparts and not on some market probabilities. It consists in finding a price from which the buyer is willing to buy the contingent claim and another price up to which the seller is willing to sell it. These prices are computed relative to a corresponding view of the risk from the entities, represented by their utility function. As described by Barrieu et al. [2009], the buyer wants his investment net present value (NPV) to be greater or equal to his initial wealth, given his view on risk represented by its utility function. His indifference price is the maximum amount that he is ready to pay for the claim so that he does not loose money by carrying the transaction. Similarly, the seller's indifference price is the minimum amount that he is willing to receive to process the trade.

We identify the increasing concave functions $U_b : \mathbb{R} \to \mathbb{R} \cup \{-\infty\}$ and $U_s : \mathbb{R} \to \mathbb{R} \cup \{-\infty\}$ as the buyer and the seller utility functions, respectively. We let $w_b \geq 0$, $w_s \geq 0$ be their initial wealth and $P(X)$ be the claim's discounted payoff on the underlying X. We define $\Pi_b(P) \geq 0$ and $\Pi_s(P) \geq 0$ as the indifference prices of the buyer and seller, given by the equations:

$$\mathbb{E}_{\mathbb{P}}[U_b(w_b)] = \mathbb{E}_{\mathbb{P}}[U_b(w_b + P(X) - \Pi_b(X))] \quad (5.3.20)$$

$$\mathbb{E}_{\mathbb{P}}[U_s(w_s)] = \mathbb{E}_{\mathbb{P}}[U_s(w_s - P(X) + \Pi_s(X))] \quad (5.3.21)$$

We can then find analytically, or numerically, the two prices $\Pi_b(P)$ and $\Pi_s(P)$ providing the range of prices that can be settled between the two counterpart. In the case where $\Pi_b(P) < \Pi_s(P)$ there is no price which they can agree on, otherwise the price range is $[\Pi_s(P), \Pi_b(P)]$.

5.3.3.2 Exponential utility function

We let the processes X^S and X^B correspond to the seller's and buyer's view, respectively, and compute the indifference prices in Equation (5.3.20) and (5.3.21). We consider the exponential utility function

$$U(x) := 1 - e^{-\lambda x}, \forall x \in \mathbb{R}$$

where $\lambda > 0$ is a known parameter corresponding to the coefficient of risk aversion representing the counterpart view on financial risk. This utility function is widely used in the literature and has no constraint on the sign of the cash-flow. Moreover, the indifference prices over this utility are independent of the initial wealth, which simplifies the number of parameters to include in the computation of the prices (see Barrieu et al. [2009]). Using the exponential utility function with the coefficient of risk aversion λ_b from (5.3.20) we have:

$$\Pi_b(X^B) = -\frac{1}{\lambda_b} \log \mathbb{E}_{\mathbb{P}}[e^{-\lambda_b P(X^B)}] \tag{5.3.22}$$

Similarly for the seller with an exponential utility function with a coefficient of risk aversion λ_s from (5.3.21) we have:

$$\Pi_s(X^S) = \frac{1}{\lambda_s} \log \mathbb{E}_{\mathbb{P}}[e^{\lambda_s P(X^S)}] \tag{5.3.23}$$

5.3.3.3 Pricing a call option

For simplicity of exposition we are going to price a call option with payoff defined as follow:

$$P := B_T U (X_T - K)^+ \tag{5.3.24}$$

where $U > 0$, T is the maturity, K is the strike, $B_T := B(0,T)$ is the discount factor, and X_t is the underlying asset at time t. We let X_t be a Gaussian process such that $X_t \sim N(m_X, \sigma_X^2)$ where m_X is the mean and σ_X is the standard deviation. Given the indifference price of a buyer having an exponential utility function with risk aversion parameter λ_b in Equation (5.3.22), we get after some algebra

$$\mathbb{E}[e^{-\lambda_b P(X)}] = \Phi\left(\frac{K - m_X}{\sigma_X}\right) + \exp\left(\frac{\sigma_X^2(\lambda_b B_T U)^2}{2} + \lambda_b(K - m_X)\right)\Phi\left(\frac{m_X - \sigma_X^2\lambda_b - K}{\sigma_X}\right)$$

We can then deduce the buyer indifference price as

$$\begin{aligned} &\Pi_b(X) = \\ &-\frac{1}{\lambda_b}\log\left[\Phi\left(\frac{K - m_X}{\sigma_X}\right) + \exp\left(\frac{\sigma_X^2(\lambda_b B_T U)^2}{2} + \lambda_b B_T U(K - m_X)\right)\Phi\left(\frac{m_{X_T} - \sigma_X^2\lambda_b B_T U - K}{\sigma_X}\right)\right] \end{aligned} \tag{5.3.25}$$

Similarly, we consider the indifference price of the seller having an exponential utility function with risk aversion parameter λ_s in Equation (5.3.23). Since the computation of $\mathbb{E}[e^{-\lambda' P(X)}]$ in Equation (5.3.25) does not depend on the sign of λ_b, we can have $\lambda_b = -\lambda_s \ \forall \lambda_s$ and $X = X^S$. Then, we get

$$\mathbb{E}[e^{\lambda_s P(X)}] = \Phi\left(\frac{K - m_X}{\sigma_X}\right) + \exp\left(\frac{\sigma_X^2(\lambda_s B_T U)^2}{2} - \lambda_s B_T U(K - m_X)\right)\Phi\left(\frac{m_X + \sigma_X^2\lambda_s B_T U - K}{\sigma_X}\right)$$

We can then deduce the seller's indifference price $\Pi_s(X)$ as

$$\begin{aligned} &\Pi_s(X) = \\ &\frac{1}{\lambda}\log\left[\Phi\left(\frac{K - m_X}{\sigma_X}\right) + \exp\left(\frac{\sigma_X^2(\lambda_s B_T U)^2}{2} - \lambda_s B_T U(K - m_X)\right)\Phi\left(\frac{m_X + \sigma_X^2\lambda_s B_T U - K}{\sigma_X}\right)\right] \end{aligned} \tag{5.3.26}$$

5.4 Basic option strategies

5.4.1 Understanding the existence of option strategies

5.4.1.1 Pricing principles

Options provide leverage and give the ability to take a view on volatility as well as equity direction. The nature of options makes it possible to create a considerable number of speculative trading strategies. Those can be based on different approaches encompassing a variety of principles and valuation techniques. Option strategies are often done to gain exposure to a specific type of opportunity or risk, while eliminating other risks as part of a trading strategy. However, parsimonious information on option prices is available in time and space, and can only be accounted for with the No-Dominance law stating that valuation obeys to a trivial monotonicity rule. A dominant trading strategy is a portfolio costing the same as another one, but which is always guaranted to outperform it. Equivalently, a dominant trading strategy exists if it is possible to start with no money, and make a guaranteed trading profit. It can be formalised in the following proposition:

***Proposition* 10** *No dominance principle*
Let X be the gain from a portfolio strategy with initial cost x. If $X \geq 0$ in every state of the world, then $x \geq 0$.

The dominance principle is based on the following fundamental rules:

(D1) The law of one price: the same asset must trade at the same price on all markets.

(D2) Two assets with identical cash flows in the future must have the same price to start with.

(D3) An asset with a known future price must trade today at the price discounted at the risk free rate.

The law of one price is a less demanding requirement than the non-existence of dominant trading strategies or no-arbitrage. It can be proved that

- if there is no arbitrage there is no dominant trading strategy, but there may be arbitrage opportunities even if there are no dominant trading strategies.
- if there is no dominant trading strategy then the law of one price holds, but the law of one price may hold even when trading strategies exist.
- if there are no dominant trading strategies then it can be shown that there must exist a linear pricing measure.

As an example of financial instruments following these rules we have:

1. Futures: they must satisfy rule number $(D3)$, any deviation from this equality leads to arbitrage.
2. Derivaties: buy and sell the same asset on two different markets, namely the spot market and the derivative market. Hence, they must satisfy rule number $(D1)$.

The No-Dominance law is most easly expressed in terms of relative value which is the attractiveness of one instrument relative to another instrument measured in terms of risk, liquidity, and return. Given that options are a derivative instrument, meaning they derive their value from an underlying security, options themselves have value relative to other options. Hence, when comparing two options, one option's value can be deduced from or defined relative to another option's value. In practice, even though options are quoted on the basis of price, option traders assess relative value on the basis of volatility. That is, knowing the price of the option, one can solve for the IV of the underlying stock by inverting the Black-Scholes formula (see Section (1.4.4)). More generally, all of financial modelling is based upon comparison between assets. No one would buy the more expensive portfolio if the cheaper one would produce the same cash flows. If these rules are not satisfied arbitrage opportunities will arise and traders will take advantage of market's imbalance.

5.4.1.2 Accounting for inefficient markets

Using proper tools to systematically identify when option prices become undervalued, one can generate a profit by buying these options and risk-managing them with a non-Gassian model. Further, the properties of market prices in a complete market are such that it is optimal to hold a delta-hedged position on options until maturity as long as the misspecified volatility is higher/lower than the true one over the life of the option (see Section (4.2.3)). However, in an incomplete market, prices are no-longer always increasing with volatility and the replicating portfolio is no-longer the price of the option (see Section (5.3)). As a result, opportunities (or arbitrage) occur during the life of the option, and it may become optimal to unwind a position on options before maturity. In general, the realised profit is due to subsequent non-Gaussian market moves ahead of the option maturities. Hence, algorithms can be put in place to systematically detect such arbitrage and take advantage of them. They rely on a combination of statistical and quantitative market timing tools, pattern recognition, volatility and direction indicators. Systematic option trading is about monetising these price discrepancies by buying / selling undervalued / overvalued options identified by a model.

5.4.1.3 The difficulty of choosing options

While options have the ability to limit a trader's total investment, options can also expose traders to volatility risk, and increase opportunity costs. The Greeks - delta, gamma, vega, theta and rho - measure different levels of risk in an option, each one adding a different level of complexity to the decision-making process. The Greeks are designed to assess the various levels of volatility, time decay and the underlying asset in relation to the option. As a result, choosing the right option is a difficult task because there is the constant fear for the trader to pay too much for the option or that it may lose value before you have a chance to gain profits. Further, it is also difficult to discern what makes one option strike price better than any other option strike price. A trader must assume that the underlying asset will not only achieve the strike price level, but will exceed it in order for a profit to be made. In the case where the wrong strike price is chosen, the entire option investment is lost. Most of these problems can be minimised or eliminated when a trader decides to use either a synthetic option, or relative trading, instead of simply purchasing an option. A relative value hedge fund seeks to exploit differences in the price or rate of the same or similar securities. The relative value fund trades on gaps, rather than the price of a specific security alone. The relative value fund may take positions if the gap between prices or rates is considered to have reached its peak and is thus expected to shrink, or may take a position in a security if similar securities are experiencing price changes. Relative value trades can be expressed with many spread types such as a simple spread $(A - B)$, a ratio $\frac{A}{B}$, a percentage change $100 \times (A\Delta\% - B\Delta\%)$, merger arbitrage $(xB + \$ - A)$, etc.

5.4.2 Some statistical arbitrages

Statistical arbitrages are market neutral attempts at identifying and exploiting (usually short live) anomalies in pricing, but they are not real arbitrage since they do not deliver a guaranteed profit. The strategies are market neutral because long positions are matched against similar short positions. They rely on the existence of violations of weak form market efficiency, which are likely to be small and transient. We now describe a few well known statistical arbitrage strategies on options.

5.4.2.1 The put-call parity

Following the fundamental rule number $(D2)$ there is an equivalence relationship between call/put prices and a forward price given by

$$C(t, S_t, K, T) - P(t, S_t, K, T) = S_t Re(t, T) - \sum_{i=1}^{N} d_i \frac{P(t, t_i)}{Re(t, t_i)} - KP(t, T)$$

where $P(t, T)$ is a zero-coupon bond with maturity T. Hence, the put-call parity posits a risk neutral equivalence relationship between a call, a put and some amount of the underlier. If there is a difference between the two assets there is an arbitrage which must be corrected. That is, arbitrageurs can make a sure profit by shorting the securities in the expensive portfolio and buying the securities in the cheaper one. For instance, assuming no discrete dividends, in the case where $C(t,S_t,K,T)-P(t,S_t,K,T) < S_tRe(t,T)-KP(t,T)$, one should short both the put option and the stock and use the proceeds to buy the call option and invest the difference $P(t,S_t,K,T)+S_tRe(t,T)-C(t,S_t,K,T)$ at the risk-free rate. At expiration, one of the two options is ITM and the other one is OTM. In either case, the arbitrageur ends up buying one share at the price K, leaving him with a profit of $P(t,T)\big(P(t,S_t,K,T)+S_tRe(t,T)-C(t,S_t,K,T)\big)-K$. In the case of American options the put-call parity does not exit because it can be exercised at any time prior to the expiration date. Only an inequality holds given in Equation (5.1.10). Still, by adjusting for dividends and rates, in the case of at-the-money call and put options, some authors assume the existance of such a relationship.

5.4.2.2 Synthetic positions

Following the fundamental rule number $(D1)$ and $(D2)$ we can duplicate the risk profile of options with complex strategies having the same strike K and the same maturity T. In that case, the risk and rewards are identical and the synthetic positions must have the same price. There are two types of synthetic options, the synthetic calls and the synthetic puts. Both types of synthetics require a cash or futures position combined with an option, where the former is the primary position and the latter is the protective position. Being long in the cash or futures position and purchasing a put option is known as a synthetic call, while a short cash or futures position combined with the purchase of a call option is known as a synthetic put. Note, since we are holding on to a cash position or futures contract, if the market begins to move against the cash or futures position it is losing money in real time. However, it is assumed that the option will move up in value at the same speed to cover the losses. This is best accomplished by buying an at-the-money option, but they are more expensive than out-of-the-money options. In any case, one must have a money management strategy to help you determine when to get out of the cash or futures position. At last, if the traded market has little to no activity, the at-the-money protective option can begin to lose value due to time decay, forcing a trader to abandon a trade early.

As an example of such synthetic position is the call/put parity described in Section (5.4.2.1). However, there is a dividend difference between the two type of assets. In that setting, we distinguish two cases:

- conversion: buy a put and sell a call, while buying the stock.
- reverse conversion (a synthetic long position): sell a put and buy a call, while shorting the stock.

which leads to

- gains from price difference in a time frame of seconds/minutes.
- profit only if prices move out of alignment (violate put-call parity), otherwise make a loss.

As long as the put and call have the same underlying, strike price and expiration date, a synthetic long position will have the same risk/return profile as owning an equivalent amount of the underlying stock. However, it can be affected by the borrowing cost of shorting the stock and the put option. Unlike the conversion, where you have to pay for the position with a net cash outlay, a reversal involves selling short the stock and the put, which brings in a credit to the account.

5.4.2.3 Volatility arbitrage

Volatility arbitrage is a type of statistical arbitrage that is implemented by trading a delta neutral portfolio of an option and its underlier. The objective being to take advantage of the differences between the implied volatility of the option, and a forecast of future realised volatility of the option's underlier. So long as the trading is done delta-neutral, buying an option is a bet that the underlier's future realised volatility will be high, while selling an option is a bet that future realised volatility will be low. Due to the put-call parity, being long a delta-hedged call results in the same returns as being long a delta-hedged put. We let the price of a call option $C = f(S, \sigma, T-t, K, ...)$ be a function of the underlying S, the volatility σ, the time to maturity $T-t$, the strike K. We consider a Δ-neutral portfolio (options plus a quantity of the underlying), and look at the difference between the implied volatility (IV) and the future realised volatility (FRV). The trade is based on the assumption that if IV is far away from FRV it will mean revert to FRV at a certain speed called the speed of mean reversion. Hence,

1. if $IV >> FRV$: sell option
2. if $IV << FRV$: buy option

However, it is not a true arbitrage since

- we need to predict the future direction of the IV, and we can face Black Swan (abrupt changes in IV).
- we need to forecast the FRV.

5.4.3 From directional to neutral trading

In portfolio theory, directional trading corresponds to the basic strategy of going long if the market, or security, is perceived as heading higher, or taking short positions if the direction is assumed downward. In the trading of stocks, commodities, currencies, and other linear assets, all profits and losses are directly proportional to their prices. Hence, most trading strategies dealing with plain assets (not options) are based on the forecast of the direction of their price movement (we will call them directional strategies). Options can also be used in such strategies by only using a fraction of the underlying. That is, because of their leverage, directional trading with options can be attempted even if the anticipated movement in the underlying stock is not expected to be large. Further, a number of strategies can be used to capitalise on a move higher or lower in the broad market or a particular stock. For example, spread options (see Section (1.5.4)) (combination of call and put option with same strike and same maturity) benefit from a moderate increase in the underlying price (bull spreads), or from its decline (bear spread). Some options strategies allow to profit from movements in the underlying that are bullish, bearish, or neutral:

1. Bullish options strategies are employed when the options trader expects the underlying stock price to move upwards. It is necessary to assess how high the stock price can go and the time frame in which the rally will occur in order to select the optimum trading strategy. The simplest of unlimited profit bullish options strategies is the Long Call strategy where you buy a longer term call option and hold on to it to maturity. Being moderately bullish means that you expect the price of the underlying stock to increase to a certain pre-determined price instead of being bullish for an unknown extended period of time to an unknown high price. Bullish options strategies with limited risk and limited profit are typically debit spread strategies that profit only if the price of the underlying stock rises beyond a certain breakeven point.

2. Neutral strategies in options trading are employed when the options trader does not know whether the underlying stock price will rise or fall. Also known as non-directional strategies, they are so named because the potential to profit does not depend on whether the underlying stock price will go upwards. Rather, the correct neutral strategy to employ depends on the expected volatility of the underlying stock price. Neutral trading strategies that are bullish on volatility profit when the underlying stock price experiences big moves upwards or downwards. They include the long straddle, long strangle, short condor and short butterfly (see Section (1.5.2.3) and (1.5.3)). Neutral trading strategies that profit when the underlying stock price experience little or no movement include the short straddle, short strangle, ratio spreads, long condors and long butterflies.

3. A volatility strategy means profiting from the market no matter if the price of the underlying stock breaks out to upside or downside. Volatile options strategies are options strategies designed to profit under such conditions of uncertainty and are commonly used ahead of major news or earnings releases in order to profit from either direction depending on how the release turn out.

Despite the fact that trading strategies based on such option combinations possess many features distinguishing them from plain assets strategies, the main determinant of their performance is the accuracy of price forecasts. In the case of options, however, position profitability depends not only on the direction of the price movement, but on many other factors such as volatility and time left until the expiration. While directional trading requires the trader to have a strong conviction about the market or security's near-term direction, the trader also needs to have a risk mitigation strategy in place to protect investment capital in the event of a move to the opposite direction. For instance, market-neutrality is a trading strategy where small changes in the underlying price do not lead to a significant change in the position value, and given larger price movements, the position value changes by approximately the same amount regardless of the direction of the underlying price movement. The main analytical instrument used to create market-neutral positions is the delta of an option. The position of a portfolio is market-neutral if the sum of the deltas of all its components is equal to or close to zero. If we consider strategies containing certain directional elements, the reduction to zero of the delta is not an obligatory requirement. Forecasts of the directions of future spot price movements represent an integral part of such a strategy. These forecasts can be incorporated into the strategy structure in the form of biased probability distributions or asymmetrical option combinations. Directional strategies are often used to engineer a particular risk profile to movements in the underlying security. Buying a butterfly spread (long one K_1 call, short two K_2 calls, and long one K_3 call) allows a trader to profit if the stock price on the expiration date is near the middle exercise price K_2, and does not expose the trader to a large loss (see Section (1.5.2.3)). One well-known strategy is the covered call, in which a trader buys a stock (or holds a previously-purchased long stock position) and sells a call option. If the stock price rises above the strike price, the call will be exercised and the trader will get a fixed profit. If the stock price falls, the call will not be exercised, and any loss incurred to the trader will be partially offset by the premium received from selling the call. Overall, the payoffs match the payoffs from selling a put due to the put-call parity. A bull Call Spread is constructed by buying a call option with a low exercise price K_L, and selling another call option with a higher exercise price K_H with the same maturity. Often the call with the lower exercise price will be at-the-money while the call with the higher exercise price is out-of-the-money. It is designed to profit from a moderate rise in the price of the underlying security. A short straddle is a non-directional options trading strategy that involves simultaneously selling a put and a call of the same underlying security, strike price and expiration date. The profit is limited to the premiums of the put and call, but it is risky because if the underlying security's price goes very high up or very low down, the potential losses are virtually unlimited. One of the advantage of selling an ATM straddle as opposed to a call option is that its sensitivity with respect to a change in spot price is nearly null at inception. It is a volatility product as long as the spot price does not deviate from the strike price, otherwise it becomes very risky and one should buy it back. That is, in the case where the spot price S_t deviates too much from the strike K, one of the two legs of the straddle brings non-necessary risk to the seller of the contract who should buy it back. Selling a straddle (selling both a put and a call at the same strike price K) would give a trader a greater profit than a butterfly if the final stock price is near the strike price, but might result in a large loss. Similar to the straddle is the strangle which is also constructed by a call and a put, but whose strikes are different, reducing the net debit of the trade, but also reducing the risk of loss in the trade.

5.4.4 From pair trading to skew trading

5.4.4.1 Pair trading

One of the simplest relative-value arbitrage called pair trading is an investment strategy seeking to take advantage of price differences between related financial instruments by simultaneously buying and selling the different instruments, thereby allowing investors to potentially profit from the relative value of the two products. The simultaneous purchase and sale of two similar products whose prices are not in synchrony with what the trader believes to be "true value" is called an arbitrage in the hedge funds world. For example, acting on the assumption that option prices will revert to their true value over time, traders will sell short the overpriced security and buy the underpriced one. Once prices revert to their true value, the trade can be liquidated at a profit. It is clear from this example that what hedge funds call an arbitrage is simply the fact that their views about the future performance of the underlying asset differ from the market's view. It is a purely directional bet on the expected future dynamics of the underlying stock price. However, in the classical option pricing theory, arbitrages are quite different (see Shreve [2004]). One can reconcile the two approaches by providing the traders with tools capable of properly quantifying market arbitrages to option prices, together with tools modeling dynamically the agents rational anticipations.

5.4.4.2 Skew trading

Not only are volatile options strategies used for speculating in an uncertain price breakout, but they can also be used to speculate on implied volatility (IV). The IV being a biased predictor of the future volatility, it bears valuable information on the asset price process and its dynamics. That is, the options market provides a remarkable outlook on future expectations of the value or performance of an underlying asset, allowing traders to compare various strike prices over different maturities. Since supply and demand ultimately drive prices, traders can learn which options are cheap or expensive relative to others, as measured by the implied volatility of each option. This relative value is defined as options skewness, or skew, and can be used to identify trading opportunities. Trading "with" the skew is defined as buying higher valuation options and selling lower valuation options, while trading "against" the skew is defined as buying lower valuation options and selling higher valuation options. Traders can then use quantitative tools and decide to either trade "with" or "against" the skew or smile. Even though there are different explanations for the skew, one of the most straightforward is leverage. Skew is priced to reflect the market's assessment of future risk, which takes into account an asset's current price, pricing trends, and the potential for a sudden price jump in either direction. The basis of skew is that even though options are founded on a risk-neutral concept, market participants have risk profiles that affect the supply-demand relationship of the options market. For example, many equity option traders tend to sell upside calls and purchase downside puts (termed options collar) to reduce their overall risk exposure. Given that many participants have predetermined preferences, options with lower strikes tend to have higher implied volatilities relative to options with higher strikes. Consequently, if you believe that the skew is wrongly priced, then the market is either underestimating or overestimating the probability of a large upside or downside move in the underlying instrument. Hence, one can look at the volatility skews of an index or a single stock over different maturities and compare their relative predicting power for the movement of the underlying index or stock. For example, in the case where the short term months have very low implied volatilities relative to intermediate term options, it is usually the case that the options market is expecting news likely to move the index or stock in a dramatic way in a few months' time, but not in the shorter term. The trading strategies should be based on the trader's prediction for market movements, or lack thereof, relative to market expectations (in the time period selected). However, one must recognise the conditions where a strategy might fail.

5.4.4.3 Taking a view on the Skew

When trading "with" the skew, the market is willing to overpay for a certain strike price or time frame. A trader can purchase the higher implied volatility and sell a different strike price, or month, at a lower implied volatility. He chooses to purchase higher-valued options and sell lower-valued options because his market forecast could mirror the options market (more demand relative to supply creates the skew) where he gives away theoretical edge. Alternatively, he can trade "against" the skew, assuming he forecasted a symmetric event and wanted to gain theoretical edge by selling higher-valued options and buying lower-valued options. Trading against the skew is a "reversion to the mean" strategy, meaning that we are implementing an options strategy that benefits from a more normalised trading scenario (think of symmetrical rather than asymmetric underlying moves). Recall, "cheap" or "expensive" attributions are based on a symmetric mathematical options model, such as the Black-Scholes formula. For example, if the trader believes the market will remain calm in the near-term but that, in the coming months, it's going to become more volatile, he can implement a long calendar position. This is considered trading "with" the skew if the shorter-dated options are priced much cheaper than the longer-dated options. There are many variations of trading "with" the skew. The only requirement is that the option you sold has a lower implied volatility than the option you purchased. To conclude, since the IV is an indicator of potential outcomes for an underlying asset, with insight into market expectations, traders can choose to trade "with" the skew (smile) or "against" it.

5.4.5 Strategies on different maturities

5.4.5.1 Defining the term structure of volatility

Choosing an expiry date for an option is a difficult task as one must determine the likely dates a stock will converge to a forecast target price. Most well-established option trading strategies in the literature operate by trading multiple contracts in a single chain (maturity). However, we can not treat each testing chain independently as there is a correlation between chains with different expiration dates for the same underlying. One must consider a time dimension to incorporate information between the chains of different expiration dates for the same underlying. Hence, when considering a prior distribution, it should be over paths of prices over time, such that when contracts are traded at expirations the prior affects the probability distribution over those paths. Furthermore, different expiration dates are linked through the underlying, that is, a trading agent may make plans to purchase an underlying now to sell at a specific date in the future, or to sell conditional on how prices move over time. The term structure, represented as a curve, shows how the maturity date of an option will change the implied volatility over time. Describing the pattern of options with the same strike price but different maturities, the term structures of implied volatility helps investors get a better expectation of whether an option with a short-term expiration date will rise or fall in the future. If the implied volatility of long-term options are higher than that of short-term options, then the short-term implied volatility is expected to rise. Similarly, if the implied volatility of short-term options is lower than those for long-term options, then short-term volatility is expected to fall. The slope of the volatility term structure is the difference between implied volatilities (IV) of long-dated and short-dated ATM options. In general, the slope of the implied volatility term structure is positively related with future option returns. It has been shown that option portfolios with high slopes of the volatility term structure outperform option portfolios with low slopes by an economically and statistically significant amount (see Jones et al. [2002]).

5.4.5.2 Modelling the term-structure of volatility

Options at different dates or tenors T_i for $i = 1, .., n$ reveal a term structure of implied volatilities that is a useful tool in relative value analysis. Long dated options are priced relative to the value of short dated options. The effect is similar to that observed in fixed income yield curves. In general, long dated implied volatilities do not fully respond to transitory movements in short-dated implied volatility because of the mean reverting nature of the series. That is, a shock to short-dated volatility is likely to dissipate over a longer time interval as the market reverts to its normal or average level. Finally, greater volatility of short-dated volatility reduces the value of short dated volatility as a benchmark for valuing long dated volatility. As a result, it tends to lower long-dated implied volatilities. Using this intuition, practitioners (see Simpson [2003]) have devised statistical models where the term-structure of the IV is a function of two factors, the short-dated volatility and the volatility of short-dated volatility

$$\sigma_l(t) = \alpha + \beta_1 \sigma_s(t) + \beta_2 \sigma_s^2(t) + \epsilon$$

where $\sigma_l(t)$ is the volatility of the long-dated options and $\sigma_s(t)$ is that of the short-dated option. The parameters of the model are α and β_i for $i = 1, 2$ and ϵ is a white noise. Deviations from the expected values from this econometric model suggest a useful approach to determining whether implied volatilities in a specific portion of the yield curve are expensive or cheap.

5.4.5.3 The calendar spread

When buying an option, we are betting that the market will move in a certain direction within a certain amount of time. The theta of an option is an estimate of how much an option would decrease per day from time decay when there is no outside movement or volatility in the underlying futures contract. At-the-money options tend to have higher thetas and are more susceptible to time decay as expiration approaches. The thetas are progressively lower for options as you get away from the strike price at the money. Option decay is not linear, it is exponential and increases as the option gets closer to expiration. The calendar spread (time spread) refers to a family of spreads involving options of the same underlying stock, same strike prices, but different expiration months. They can be created with either all calls or all puts. Using calls, the calendar spread strategy can be setup by buying long term calls and simultaneously writing an equal number of near-month at-the-money or slightly out-of-the-money calls of the same underlying security with the same strike price. The investor hopes to reduce the cost of purchasing a longer-term call option. He hopes that the price of the underlying remains unchanged at expiration of the near month options so that they expire worthless. As the time decay of near month options is at a faster rate than longer term options, his long term options still retain much of their value. The options trader can then either own the longer term calls for less or write some more calls and repeat the process. If the options trader is bullish for the long term and is selling the near month calls with the intention to ride the long call for free, he is implementing the bull calendar spread strategy. The maximum loss would occur should the two options reach parity, which happen if the underlying stock declined enough that both options became worthless, or if the stock rose enough that both options went deep in-the-money and traded at their intrinsic value. In either case, the loss would be the premium paid to put on the position. At the expiration of the near-term option, the maximum gain would occur should the underlying stock be at the strike price of the expiring option. If the stock were any higher, the expiring option would have intrinsic value, and if the stock were any lower, the longer-term option would have less value. An increase in implied volatility, all other things equal, would have an extremely positive impact on this strategy since longer-term options have a greater sensitivity to changes in market volatility. The difference in time to expiration of these two call options results in their having a different theta, delta and gamma. The near-term call suffers more from time decay (greater theta), but it has a lower delta and a higher gamma (if the strike is at-the-money). This means that if the stock moves sharply higher, the near-term call becomes much more sensitive to the stock price and its value approaches that of the more expensive longer-term call.

5.4.6 Volatility trading

5.4.6.1 Some examples

Volatility is a tradable asset enabling consistent returns via the construction of portfolio of options. There exists several strategies based on volatility trading. The goal being to buy or sell a contingent claim in order to be exposed to volatility, while hedging the other exposures (to the movement of underlying or interest rate) as much as possible. For instance, a volatility dispersion trade is a hedged strategy taking advantage of relative value differences in implied volatilities between an index and a basket of component stocks. One frequently used derivative to execute dispersion trade is the variance swap. Another strategy consists in trading volatilities of the least volatile stocks (such as utilities) against the most volatile ones (such as financials). The trade is usually based on the historic difference between volatilities and is executed whenever a substantial (or large enough to overcome transaction costs) difference between the historic one and current one occur. The direction of the trade is such that an investor is betting that the current difference will converge to historic one. As explained by Hutchinson [2011], there exists a large number of option value strategies among which the most famous one are

1. Volatility Surface Relative Value: It optimises a portfolio of long and short option positions, on the same underlying security, to harvest gains from changes in the shape of the volatility surface.

2. Implied Correlation / Dispersion: It trades the relative value relationship between the implied volatility of an Index option and the implied volatilities of the Component Stocks that comprise the Index.

3. Capital Structure Arbitrage: It trades the relative value relationship between options on various segments of the capital structure of an individual company, or the options embedded in the same company.

4. Generic Volatility Long/Short: It trades the relative value relationship between options on a full array of securities, and is not limited to one underlying security or issuer of securities.

5. Directional Volatility: It creates an option portfolio that exhibits a continuous short or long exposure to changes in implied volatility.

5.4.6.2 Volatility surface relative value

As the IV surface is a three-dimensional array obtained by plotting the inverted option prices along two axes, the time to maturity and the strike prices (see Section (1.4.4)), one must rely on tools quantifying the relative value of option prices not only on space but also on time. Since the IV surface changes shape as investors change their expectation of risk to come, and thus the price they are willing to pay for options on the surface, one can evaluates the IV surface and optimises a portfolio of long and short option positions. For example, point pairs on the surface may be initiated when the pair relationship is at statistical extreme, or it reflects misplaced expectations. In that setting the investment strategy relies on some fundamental analysis consisting of

- examination of the IVS
- analysis of option market values relative to model values
- statistical analysis of relationships between points on the IVS
- objective consideration of factors affecting expectations of risk (IVS dynamics)

Then the trades are implemented and executed based on some targeted opportunities. Note, trades can also be identified for risk management purposes by offsetting aggregated portfolio sensitivities (Greeks). In general the market volatility surface is analysed relative to a model volatility surface and an electronic eye is used to identify and display trade opportunities. Using quantitative tools, traders should be able to assess opportunities, identify hedges and execute orders. Therefore, one can identify potential opportunities by profiling a smile of IV generated by proprietary modeling against the IV of options actually observed in the market place. For instance, one can consider cheap IV versus model as well as the expensive one against model. For that relation to exist the model must be arbitrage-free, not just in space but also in time. This is the case with the parametric MixVol model described in Section (6.2). In addition, the market surface can also be described statistically by looking at the relationship between each point. For example, one can look at the ratios and spreads of 10% OTM puts for various maturities, placed in one-year historical context via a percentile calculation. A pair with a Vol Ratio Percentile of less than 15% would be coloured blue or green, while high ratios would be coloured red.

5.4.6.3 Dispersion relative value

Volatility dispersion trading is a popular hedged strategy designed to take advantage of relative value differences in implied volatilities between an index and a basket of component stocks, looking for a high degree of dispersion. This strategy typically involves short option positions on an index, against which long option positions are taken on a set of components of the index (see Section (7.6)). Assuming that we know the IV of a portfolio (or index) together with the weights and IV of the individual stocks in the portfolio, we can solve for the expected correlation, also called the implied correlation (IC). The IC is traded via positions in single stock options offset by index option positions. As the relationship between index volatilities and their component stock volatilities change quite a lot over time, so does the implied correlation. The IC is usually viewed by practitioners as one of the option market's expectation of systematic risk. In principle it is increasing when investors sell stocks, and decrease as more benign expectations filter in.

The growth of the variance swap (VS) market and the success of the VIX raised the profile of volatility trading as an asset being negatively correlated to the underlying equity market. Rather than solving the difficult questions of knowing which volatility to own, and how to mitigate the expense of decay, the VS is a liquid, non-strike dependent hedge (see Section (7.1)). One strategy consists in owning a basket of single stock volatility, assuming that volatilities would increase when the underlying shares sell off, and then take advantage of the elevated implied correlation. Rather than selling the basket of single volatility, the relatively more elevated index volatility is sold against the basket of single stock volatilities. Based on empirical studies, it is assumed that a basket of higher beta single stock volatilities may have a larger absolute move up than index volatility during a sell-off. However, single stock volatility having wider bid/offer spreads, exhibit a greater expected cost over the course of a year, and potentially unmanageable bid/offer spreads. In addition, the relationship between single stock volatilities and index volatility may not perform as historically expected, leading to basis risk. Nonetheless, some opportunities each year to sell elevated index volatility present the possibility of significantly mitigating the expense.

Chapter 6

A parametric volatility model

6.1 The choice of a volatility model

In complete markets, pricing models such as Dupire's formula (see Equation (1.7.42)), require the knowledge of Europran call and put prices for all strikes and all maturities. However, in practice we can only observe a few market prices from standard strikes and maturities. For an index, one can easily obtain more than 100 points (prices or implied volatility) where some of them are listed while the others are produced by brokers. On the other hand, there are only a few points for single stock (from 0 to 30) produced only by brokers. These points do not necessary satisfy the no-arbitrage conditions, can have wide or narrow spreads depending on the liquidity on the market and the volume traded. As a result, the market is incomplete and there are more than one acceptable price (volatility) surface satisfying the no-arbitrage conditions. Hence, multiple risk-neutral distributions can fit the option prices, so that one needs some additional criteria to generate a unique probability distribution function (pdf). To do so, one can either impose a functional form to the probability distribution and estimate its parameters using option data, or, one can choose non-parametric methods obtaining perfect fit to market data (see Jackwerth [1999], Bondarenko [2003] for a review of these methods). However, non-parametric methods are less adapted to the extrapolation problem than the parametric ones, and they tolerate less control over the generated volatility surface. We are therefore going to use a parametric representation of the market call prices in order to smooth the data and get nice probability distribution functions (pdf).

6.1.1 A nonparametric model

We briefly describe the nonparametric model proposed by Ait-Sahalia et al. [1998] for the estimation of the risk-neutral marginal densities. While the general method requires a kernel regression of call prices on five variables, they discussed a dimensionality reduction method where the call price function is given by the BS-formula with the volatility replaced ny a nonparametrically estimated function. Given a collection of option prices ($i = 1, ..., n$) and a spot price S_t, we let $\kappa(\cdot)$ be the Nadaraya-Watson kernel estimator with a bandwidth of h. Then, the model price of a call option is given by

$$C_M(t; K, T) = C_{BS}(t, S_t, K, T; \widehat{\sigma}(F(t, T), K, T))$$

where

$$\widehat{\sigma}(F(t, T), K, T) = \frac{\sum_{i=1}^n \kappa_F\big(\frac{(F(t,T)-F(t_i,T))}{h_F}\big)\kappa_K\big(\frac{(K-K_i)}{h_K}\big)\kappa_T\big(\frac{(T-T_i)}{h_T}\big)\sigma_i}{\sum_{i=1}^n \kappa_F\big(\frac{(F(t,T)-F(t_i,T))}{h_F}\big)\kappa_K\big(\frac{(K-K_i)}{h_K}\big)\kappa_T\big(\frac{(T-T_i)}{h_T}\big)}$$

The risk-neutral density function is obtained by taking the second derivative of the call option pricing formula with respect to the strike price. Considering a centered finite difference approximation, the density is computed as follow

$$P(0,T)f^Q_{S_T}(K) = \partial_{KK}C_M(t) \approx \frac{C_M(t;\widehat{\sigma}(K+\epsilon)),K+\epsilon) - 2C_M(t;\widehat{\sigma}(K)),K) + C_M(t;\widehat{\sigma}(K-\epsilon)),K-\epsilon)}{\epsilon}$$

6.1.2 The parametric models

Using parametric representations of the market prices, proper option portfolio management requires suitably chosen interpolation and extrapolation techniques of the IV surface, forcing practitioners to use a fully specified model. The standard approach is to interpolate and extrapolate market prices, or volatilities, to complete the market. Practitioners use the Black-Scholes implied volatility and smooth the prices in a parabolic way to generate the missing prices. Due to the difficulty of defining a single parametric function for the entire volatility surface, a typical solution is to estimate each smile independently with some nonlinear function. Then, the IV surface is reconstructed by interpolating total variances along the forward moneyness.

6.1.2.1 The polynomial models

One way forward is to interpolate the smile within the region of observed prices with a polynomial, as was done by Malz [1997], and to cut the volatility outside that region. Similarly, Daglish et al. [2006] performed a Taylor expansion up to the second order of the implied volatility surface around the money forward level. In general, the implied volatility is modelled with a functional form of the smile around the money forward as follow:

$$\Sigma(t,S_t;K,T-t) = f\left(t,T-t,\, X_t^{T-t}, Y_t^{T-t}, Z_t^{T-t};K\right)$$

where the processes X_t^{T-t}, Y_t^{T-t} and Z_t^{T-t} represent respectively the at-the-money volatility, the Skew and the Curvature of the smile at time t for call options with maturity T *i.e.* time to maturity $T-t$. Using historical data, some authors studied the dynamics of these processes. For instance, assuming an Ornstein-Uhlenbeck dynamic for the processes, Cont et al. [2002] explained the deformation of the volatility surface. Alternatively, Bloch et al. [2002] assumed the parameters to be led by the spot process holding the whole market risk. When it comes to generating a volatility surface, the evaluation time t is fixed and practitioners estimate one set of parameters X, Y and Z per trading maturities. Then, they rely on some interpolation and extrapolation of parameters in time with no guarantee of satisfying the calendar spread. Further, in the presence of discrete dividends it is not an easy task to satisfy the time constraint given by the calendar spread.
As an example, setting $\overline{\eta}(t,T) = \ln\left(\frac{KP(t,T)}{S_t Re(t,T)}\right)$, we can consider the functional form

$$f(t,T-t,X,Y,Z;K) = X - Y\overline{\eta}(t,T) + Z\left(\overline{\eta}(t,T)\right)^2 \tag{6.1.1}$$

with $P(t,T) = e^{-r(T-t)}$ and $X = \Sigma(t,KP(t,T);KP(t,T),T-t)$, $Y = \frac{\partial\Sigma(t,S_t;K,T-t)}{\partial K}$ and $Z = \frac{\partial^2\Sigma(t,S_t;K,T-t)}{\partial K^2}$. Note, whatever the shape we take, the smile needs to be caped and floored in the lowest and highest strikes in order to avoid any arbitrage opportunity. If the slope of the smile $\frac{\partial\Sigma}{\partial K}$ is too high, the implicit *pdf* can be negative (see Equation (1.7.47)).
Among the polynomial models describing the whole volatility surface with one equation are the Cubic model and the Spline model. In the former the IV is a cubic function of moneyness $\overline{\eta} = \ln\frac{K}{S}$ (or the time adjusted moneyness $\widehat{\eta} = \frac{1}{\sqrt{\tau}}\ln\frac{K}{S}$) and a quadratic function of time to expiry τ

$$\Sigma(t,S_t;K,T) = a_0 + a_1\overline{\eta} + a_2(\overline{\eta})^2 + a_3(\overline{\eta})^3 + a_4\tau + a_5\tau^2$$

where a_i for $i=0,..,5$ are parameters to be estimated. Similarly, one can build a Spline model of the volatility surface as

$$\Sigma(t,S_t;K,T) = a_0 + a_1\overline{\eta} + a_2(\overline{\eta})^2 + a_3\tau + a_4\tau^2 + D\big(a_5 + a_6\overline{\eta} + a_7(\overline{\eta})^2 + a_8\tau + a_9\tau^2\big)$$

with

$$D = \begin{cases} 0 \text{ if } \overline{\eta} < 0 \\ 1 \text{ if } \overline{\eta} \geq 0 \end{cases}$$

together with the constraints

$$a_5 + a_6.0 + a_7.0^2 + a_8\tau + a_9\tau^2 = 0 \, , \, Da_6 = 0$$

for the volatility function to be continuous and differentiable.

6.1.2.2 The stochastic volatility models

Alternatively, one can fit with little control a parametric form for the implied volatility derived from a model which is usually the result of an asymptotic expansion of a stochastic volatility model. For example, the SABR model of Hagan et al. [2002], or the Stochastic Volatility Inspired (SVI) model introduced by Gatheral [2004], assume some behaviour of the underlying asset and connections with the values of the implied volatility. These models fit a wide range of smile patterns, both from empirical observations and those arising from several jump-diffusion models. However, being time slice parametric models, they present a number of difficulties when interpolating through time. For instance, direct interpolation and extrapolation of implied volatility surfaces does not guarantee a resulting smooth risk-neutral density, hence a proper local volatility surface.

In the case of the SABR model, it can either be used as a model for a whole volatility surface, or for the skew. Under the first approach, the parameters α, β, ρ and ν are calibrated for all given times to expiration τ_i for $i = 1, 2, ...$ In the second approach, one fit a SABR skew for each observed time to expiration, and then interpolate the values of implied volatility for any arbitrary τ. The piecewise SABR (PSABR) parameters α_i, β_i, ρ_i and ν_i are calculated separately for each time to expiration τ_i, and the IV surface is built as a linear approximation of separate skews.

The SVI is smooth in the strike direction, and it has five parameters at each maturity with intuitive interpretations in terms of implied volatility changes. It is a slice parametrisation of the implied variance, which should be relatively easy to fit liquid listed option prices. However, as noticed by Zeliade Systems [2009], the least-square fit of the SVI to the market is not trivial, as there are often local minima. The same smile can be very well calibrated with sets of parameters that are totally different one from the other. To overcome this problem, they decomposed the minimisation problem into an analytical one and a numerical one, which improved the situation by reducing the number of dimensions of the numerical optimisation. That is, they downsised the minimisation problem from dimension five to dimension two, while the optimisation over the remaining three is performed explicitly. Another problem with the SVI model is that parameter set cannot be directly comparable between smile with different maturities. Intuitively, we would like to rescale the parameters related to the skew so that we can compare the skewness of smile between different maturities. In other words, we would like to normalise the three skew related parameters by certain time factor. Gurrieri [2011] derived a set of sufficient conditions on the parametric forms of the SVI in order to satisfy the no-calendar spread arbitrage constraint, while preserving the condition of no-strike arbitrage. Adapting the strategy of Bloch [2010b] to the case of SVI, he allowed the model parameters to be functions of time and derived a set of sufficient conditions on these functional forms such that no calendar spread arbitrage can exist. He further proved the existence of solutions, and proposed a strategy to find explicit examples.

6.1.3 Dynamics of the parametric models

The main advantage of both the SABR model and the SVI model is the use of a set of parameters at each maturity with intuitive interpretations in terms of implied volatility changes. This intuitive parametric representation of the smile permits traders to manually modify its shape when the underlying price moves. They must do so, because traders using the Black-Scholes model to hedge their book keep having to change the volatility assumption when the spot moves in order to match market prices. This is due to the fact that today's market prices do not provide us with the right future anticipations of the stock price process, because the implied volatility surface is neither stationary nor Markovian but stochastic. As a result, it stresses the necessity to take into consideration the dynamics of the volatility surface when pricing and risk-managing a portfolio of options. Practitioners use rules of thumb to compute the Greeks by considering some standard evolutions of the implied volatility surface providing an indication of some possible behaviour for the smile that one might expect (see Section (4.1.1.2)).

While a model with a large number of parameters may calibrate well the volatility surface on a given day, the same model parameters may give poor results on the next day. On the other hand, any risk management system tries to estimate the future (short term forecast) behaviour of the volatility surface. As a result, one need a model that accurately fit market prices with stable and robust parameters. Ladokhin [2009] focused on two different approaches to model the dynamics of the IV surfaces, one applicable to the Cubic and the Spline model and the other used for SABR models. In the first approach, the dynamics of the surface is treated as dynamics of these parameters. To reduce the dimensionality of the problem, he applied Principal Component Analysis (PCA) to the values of the parameters a_i for $i = 0, .., n$. This is to switch to another space of the no-correlated factors, that fully describe the dynamic of the implied volatility surface. Since the first few principal components explain most of the variance of the calibrated parameters, the dynamics of the volatility surface over time is explained by the dynamics of the first two principal components modelled with an Autoregressive moving average model (ARMA). In the second approach, the SABR model already assumes certain dynamics of the volatility and the underlying asset expressed by a system of stochastic differential equations. Consequently, given the calibrated parameters, the spot price can be simulated with the Monte Carlo method, and for each path, the IV surface generated. The forecast of the implied volatility surface is an average surface over the simulated paths.

Strong of these dynamics, Ladokhin [2009] used the rolling horizon technique to build a 1 and 5 day forecasts of the volatility skews. Setting $N = 100$, he used observations from days $t - N$ till $t - 1$ to calibrate the dynamic models in order to build a forecast of the skew for day t. Then, at time $t + 1$ the horizon is rolled so that days $t - N + 1$ till t are used for calibration. An equivalent technique is used for the five days ahead forecast. He then tested how the models can hold the volatility skew pattern by performing a static test. Calibrating the models to observed IV surface at date $t - 1$ or $t - 5$ he calculated the weighted mean square error (WMSE) between the model results and the IV surface on day t. Cubic and Spline models approximate the implied volatility surface with the function of a certain form. The Spline model is perhaps, the most effective to minimise the fitting error. Good performance of the dynamic version of the Spline model is an empirical evidence of the dependence of the dynamics of the surface of two principal components. Both of these models use much less parameters, than the PSABR model. Even though the SABR and the PSABR assume a certain model for the joint dynamics of the volatility and the underlying asset, their fitting results have a higher error than the polynomial models. Nonetheless, the PSABR tends to model the skew rather effectively in case of insufficient or bad data. Because these models assume some shape of the IV surface, they are predisposed to give a more theoretical shape of the skew resulting in lower relative forecasting error. As a result, models relating prices in time and space perform better on an incomplete market or with missing data. To conclude, no single method exhibits superior accuracy in the analysis of every data set. Some methods perform better for certain underlying assets, while other methods are more suitable for the other.

6.1.4 The difficulty of generating arbitrage-free smiles

We saw in Section (1.4.2) that various necessary conditions to obtain an implied volatility surface (IVS) free from static arbitrage were introduced by Durrleman [2003]. For instance, there are strict constraints that one must observe for the convergence of the implied volatility to low and high strikes in order to guaranty no-arbitrage (see Lee [2004]), as well as for infinitely large expiries (see Tehranchi [2009]). We saw in Section (1.4.3) that Rogers et al. [2010] gave a necessary condition for no-strike arbitrage. Further, Roper [2010] established necessary and sufficient conditions for an IVS to be free from static arbitrage, which are summarised in Theorem (1.4.2). These results have provided us with a set of tools and methods to check whether a given volatility parametrisation is free from arbitrage or not. Even though the SVI model is consistent with Lee's moment formula for extreme strikes (see Lee [2004]), it is well-known that the SVI smiles may be arbitrageable. Roper investigated well known models such as SVI, Avellaneda's SABR, Quadratic parametrisation, and showed that they still admit arbitrage under certain parameter classifications. Setting $x = \ln\left(\frac{K}{F(t,T)}\right)$ and given the time scaled implied volatility Ξ defined in Equation (1.4.18), the parametrisation of the SVI follows

$$\Xi^2_{SVI}(x,\tau) = a + b\Big(\rho(x-m) + \sqrt{(x-m)^2+\sigma^2}\Big)$$

and the parameters chosen are $a = 0.04$, $b = 0.8$, $\sigma = 0.1$, $\rho = -0.4$ and $m = 0$. The parametrisation of the SABR model proposed by Avellaneda [2005] is

$$\Xi_{SABR}(x,1) = \frac{k|x|}{\ln\left(k|f(x)| + \sqrt{1+k^2 f^2(x)}\right)}$$

where

$$f(x) = \frac{1-e^{-\beta x}}{\sigma_0 \beta}$$

with parameters $\sigma_0 = 0.2$, $\beta = -4.0$ and $k = 0.5$. At last, the Quadratic parametrisation is given by

$$\Xi_Q(x,1) = 0.16 - 0.34x + 4.45x^2$$

Specifically, through deriving necessary and sufficient conditions for the IVS to be free from static arbitrage, Roper demonstrated that the Durrleman Condition (see Equation (1.4.19)) was violated and that the resulting parametrisation was not arbitrage free.

To remedy the problem of arbitrageable smiles, Gatheral et al. [2012] considered a class of SVI volatility surfaces with a simple closed-form representation, for which absence of static arbitrage is guaranteed. They found a class of SVI smiles for which the absence of butterfly arbitrage is guaranteed which is an extension of the natural parametrisation on total variance given by $\omega(\overline{\eta},t) = \theta_t SVI_\rho(\overline{\eta}\phi(\theta_t))$ where $\overline{\eta}$ is the logarithm of the forward moneyness and SVI_ρ is the classical SVI parametrisation. Note that this representation amounts to considering the volatility surface in terms of ATM variance time θ_t, instead of standard calendar time. In their new class of SVI volatility surfaces, they defined the absence of calendar spread arbitrage and butterfly arbitrage by the non-negativity of the numerator and denominator of Dupire's equation in implied volatility. However, in light of the Fokker-Plank equation, these properties are necessary, but not sufficient for no-arbitrage. As discussed by Guo et al. [2012], the right statement is that there exists an inhomogeneous diffusion process associated to the implied volatility through Dupire's equation. To this end, the local volatility defined through Dupire's equation should have adequate continuity and growth properties. Alternatively, one can consider the additional conditions defined by Roper [2010]. Guo et al. [2012] proposed a generalisation of the work by Gatheral et al. [2012] to volatility surfaces parametrised as $\omega(\overline{\eta},t) = \theta_t \Psi(\overline{\eta}\phi(\theta_t))$ for some general function Ψ, obtaining necessary and sufficient conditions such that the corresponding implied volatility surface is free of arbitrage. To facilitate the calibration of the entire volatility surface, they preconised to parametrise the set of functions ϕ to get easy-to-implement calibration algorithms among the whole admissible class. They provided an example of non-SVI parametric family, but the relation between the model parameters and the implied volatility is no-longer intuitive.

6.1.5 Towards a globally arbitrage-free parametric model

In Section (6.1.4), we discussed some of the difficulties associated with the generation of arbitrage-free smiles, and we saw that the solution was to define a globally arbitrage-free parametric model in time and space. Howver, the modelling of the whole volatility surface requires complex calibration algorithms and the use of sophisticated optimisation engine to avoid local minima. To this end, the model must be capable of reproducing a large range of implied volatility surfaces, it must be flexible and fast to compute. Among the different techniques proposed for obtaining from market prices a smooth volatility surface, Rebonato et al. [2004] argued that modelling directly the density was the most desirable approach. They extended the mixture of normals approach proposed by Alexander [2001], obtaining a density with non-zero skew and satisfying the risk-neutral forward condition while retaining an unconstrained numerical search. However, in markets with long maturity products and discrete dividends, it is important for model pricing to obtain a reliable volatility surface satisfying the no-arbitrage constraint not only in space but also in time. So, we intend to generate a surface without arbitrage in time and in space as closely as possible to market data. Interpolation techniques to recover a globally arbitrage-free call price function have been proposed by Kahalé [2004] where he considered a piecewise convex polynomials, and by Wang et al. [2004] who suggested the use of a cubic B-spline interpolation. Later, Fengler [2009] considered smoothing call prices with a natural cubic splines by choosing to minimise a penalised sum of squares resulting in an iterative quadratic minimisation problem under constraints. Similarly, Bloch [2010b] proposed to impose smoothness and value constraints directly on the market prices and their resulting implied volatility surface. He imposed the market prices to satisfy the no-arbitrage conditions defined by Durrleman [2003] and Roper [2010], and smoothed the implied volatility surface by fitting a special functional form to the observed market prices. A parametric representation of the market call prices under constraints was considered in order to smooth the data and get nice probability distribution functions (pdf). The Differential Evolution algorithm described by Bloch et al. [2011] was used to calibrate the model's parameters to a finite set of option prices.

6.2 The parametric MixVol model

Tankov [2000] and Baude [2001] developed a parametric interpolation and extrapolation of the implied volatility (IV) surface by decomposing the market option prices into a weighted sum of strike shifted Black-Scholes counterparts. That model was detailed in an internal document by Baude et al. [2002]. To infer plausible deformation of the IV surface both in space and in time, Bloch [2010b] [2012a] introduced a term structure of volatility and used a Differential Evolution algorithm to solve a non-linear optimisation problem under constraints. We are now going to describe that model and emphasise the necessary improvements to fit a large family of surfaces with very long maturities.

6.2.1 Description of the model

As explained in Section (1.7.4.3), since we can always convert a density into call prices, we can then convert a mixture of normal densities into a linear combination of Black-Scholes formula (see Brigo et al. [2000]). Therefore, to obtain a pronounced skew we consider a sum of shifted log-normal distributions, that is, using the Black-Scholes formula with shifted strike (modified by the parameters $\mu_i(t)$) as an interpolation function. In our parametric model, the market option price $C_M(K,t)$, for a strike K and maturity t, is estimated at time $t_0 = 0$ by the weighted sum

$$C_M(t_0, S_0, P_t, R_t, D_t; K, t) = \sum_{i=1}^{n} a_i(t) Call_{BS}(t_0, S_0, R_t, P_t, \overline{K}(K,t), t, \Sigma_i(t)) \tag{6.2.2}$$

where $a_i(t)$ for $i = 1, .., n$ are the weights, $\overline{K}(K,t) = K^{'}(K,t)(1 + \mu_i(t))$ with $K^{'}(K,t) = K + D_t$. In that setting $R_t = Re(0,t)$ is the repo factor in the range $[0,t]$, $P_t = P(0,t) = e^{-rt}$ is the zero-coupon bond price, $C_t = C(0,t) = \frac{R_t}{P_t}$ is the cost of carry and $D_t = D(0,t)$ is the compounded sum of discrete dividends between $[0,t]$. We let the time function $t \to \Sigma_i(t)$ be regular enough, and choose to model directly the square-root of the average variance defined as

$$\Sigma_i^2(t) = \frac{1}{t} \int_0^t \sigma^2(s) ds$$

where $\sigma(t)$ is the instantaneous volatility of the underlying process. To guarantee the positivity of the local volatility given in Equation (1.4.20), the average variance must verify

$$\Sigma_i^2(t) + 2\Sigma_i(t)t\partial_t\Sigma_i(t) \geq 0 \,, \forall i$$

and since $\Sigma_i(t) > 0$, the constraint simplifies to

$$\Sigma_i(t) \geq -2t\partial_t\Sigma_i(t) \,, \forall i \tag{6.2.3}$$

Further, the no-arbitrage theory imposes time and space constraints on market prices. Introducing the time dependent parameters $a_i(t)$ and $\mu_i(t)$, the simplest way of ensuring these constraints is to take the same time dependency for each μ, that is, $\mu_i(t) = \mu_i f(t, \beta_i)$. As the pdf of the equity price should tend toward a single Dirac when $t \to 0$, to get control on $\mu_i(t)$, we choose to let the function $f(t,x)$ tend to 1 when $t \to +\infty$, getting

$$f(t,x) = 1 - \frac{2}{1 + (1 + \frac{t}{x})^2}$$

with $f^{'}(t,x) = [1 - f(t,x)]^2 \frac{1}{x}(1 + \frac{t}{x})$. Moreover, to keep manageable the no-free lunch constraints, we make the weight $a_i(t)$ proportional to $\frac{a_i^0}{f(t,\beta_i)}$ for some constant $a_i^0 > 0$, getting the representation

$$\mu_i(t) = \mu_i^0 f(t,\beta_i) \text{ and } a_i(t) = \frac{a_i^0}{f(t,\beta_i) \times norm}$$

where $norm = \sum_{i=1}^n \frac{a_i^0}{f(t,\beta_i)}$. As a result, with seperable functions of time, we can prove that the no-free lunch constraints simplify.

Theorem 6.2.1 *No-Arbitrage Constraints in the MixVol Model*
We let the price of a European call option be given by Equation (6.2.10) and assume seperable functions of time $a_i(t)$, $\mu_i(t)$. Then the resulting implied volatility surface is free from static arbitrage (according to Definition (1.4.6)) provided the following parameter restrictions hold:

$$\begin{aligned} a_i^0 &\geq 0 \\ \sum_{i=1}^n a_i(t) &= 1 \\ \sum_{i=1}^n a_i^0 \mu_i^0 &= 0 \\ \mu_i^0 &\geq -1 \end{aligned} \tag{6.2.4}$$

6.2.2 Defining arbitrage constraints

We are going to show that the model in Equation (6.2.10) is capable of generating a wide range of volatility surfaces that are free from arbitrage both in space and in time.

6.2.2.1 Checking the no-arbitrage conditions

In order to check the no-arbitrage conditions defined by Durrleman [2003] and Roper [2010], we are going to prove that this argument is true given the restrictions in the model parameters detailed in Theorem (6.2.1). We must identify restrictions on the model parameters such that each conditions in Definition (1.4.6) remain valid. Proceeding sequentially, we get

(A1) Monotonicity of total variance is guaranted provided that Equation (6.2.3) is satisfied. This will be dependent upon the choice of the volatility functions $\Sigma_i(t)$, and will be discussed in Section (6.2.2.3).

(A2) Taking the partial derivative of the call price with respect to the strike K, we get

$$\frac{\partial}{\partial K}C_M(t_0,S_0,P_t,R_t,D_t;K,t)=\frac{1}{norm}\sum_{i=1}^{n}\overline{a}_i(t)\frac{\partial}{\partial\tilde{K}}Call_{BS}(t_0,S_0,R_t,1,\tilde{K}(K,t),t,\Sigma_i(t))\frac{\partial\tilde{K}}{\partial K}$$

where $\frac{\partial\tilde{K}}{\partial K}=P(t_0,t)(1+\mu_i(t))$. Substituting for the partial derivative of the BS-formula with respect to the strike yields

$$\frac{\partial}{\partial K}C_M(t_0,S_0,P_t,R_t,D_t;K,t)=-\frac{1}{norm}\sum_{i=1}^{n}\overline{a}_i(t)P(t_0,t)(1+\mu_i(t))N(d_2^i(t-t_0,S_t,\tilde{K}(K,t)))$$

Differentiating one more time the call price with respect to the strike, and using Equation (B.1.9), we get

$$\begin{aligned}\frac{\partial^2}{\partial K^2}C_M(t_0,S_0,P_t,R_t,D_t;K,t)&=\frac{1}{norm}\sum_{i=1}^{n}\overline{a}_i(t)\frac{\partial^2}{\partial(\tilde{K})^2}Call_{BS}(t_0,S_0,R_t,1,\tilde{K}(K,t),t,\Sigma_i(t))(\frac{\partial\tilde{K}}{\partial K})^2\\ &=\frac{1}{norm}\sum_{i=1}^{n}\overline{a}_i(t)[P(t_0,t)(1+\mu_i(t))]^2\frac{1}{\tilde{K}^2(K,t)\Sigma_i(t)(t-t_0)}Vega(\tilde{K}(K,t),t;\Sigma_i(t))\end{aligned}$$

which is greater or equal to zero if and only if $a_i^0\geq 0$ and $\mu_i^0\geq -1$, $\forall i$, given $\Sigma_i(t)>0$.

(A3) The ith forward moneyness is defined in the model as

$$\eta_{M,i}=\frac{\overline{K}(K,t)P(t_0,t)}{xRe(t_0,t)}=\eta_{BS}(1+\mu_i(t))+\frac{D_t(1+\mu_i(t))P(t_0,t)}{xRe(t_0,t)}\ ,\forall i$$

where $\eta_{BS}=\frac{KP(t_0,t)}{xRe(t_0,t)}$. Then, writting $d_1(x,y,t)$ in terms of the forward moneyness (see Equation (1.2.10)), we have

$$\lim_{\eta_{M,i}\to\infty}d_1(\eta_{M,i},\omega(\eta_{M,i},t))=-\frac{1}{\sqrt{\omega_i(t)}}\lim_{\eta_{M,i}\to\infty}\left\{\ln(\eta_{M,i})-\frac{1}{2}\omega_i(t)\right\}=-\infty$$

Since $d_2(\eta_{M,i},\omega(\eta_{M,i},t))=d_1(\eta_{M,i},\omega(\eta_{M,i},t))-\sqrt{\omega_i(t)}$, the second result follows immediately.

(A4) We are going to use the equivalent condition of Small-Moneyness-Behaviour on the call price surface given in Equation (1.4.23). Given the partial derivative of the model price with respect to the strike in (A2), taking the limit $K\to 0$, we get

$$\begin{aligned}
\lim_{K\to 0}\frac{\partial}{\partial K}C_M(t_0,S_0,P_t,R_t,D_t;K,t) &= -\frac{1}{norm}\sum_{i=1}^n \overline{a}_i(t)P(t_0,t)(1+\mu_i(t))\lim_{K\to 0}N(d_2^i(t-t_0,S_t,\tilde{K}(K,t))) \\
&= -P(t_0,t)\big(\sum_{i=1}^n a_i(t)+\sum_{i=1}^n a_i(t)\mu_i(t)\big) \\
&= -P(t_0,t)\Big(\sum_{i=1}^n a_i(t)+\sum_{i=1}^n \frac{a_i^0}{f(t,\beta_i)\sum_{j=1}^n \frac{a_j^0}{f(t,\beta_j)}}\mu_i^0 f(t,\beta_i)\Big) \\
&= -P(t_0,t)\big(1+\frac{1}{\sum_{j=1}^n \frac{a_j^0}{f(t,\beta_j)}}\sum_{i=1}^n a_i^0\mu_i^0\big) \\
&= -P(t_0,t)
\end{aligned}$$

if and only if $\sum_{i=1}^n a_i(t)=1$, $\sum_{i=1}^n a_i^0\mu_i^0=0$ and $\mu_i^0\geq -1$, $\forall i$.

By construction we set $\sum_{i=1}^n a_i(t)=1$ to ensure a normalised risk-neutral probability density function (pdf), whilst $\sum_{i=1}^n a_i^0\mu_i^0=0$ safeguards the martingale property of the induced pdf. Further, $a_i^0\geq 0$ is used to guarantee convexity of the price function, and $\mu_i^0\geq -1$ is included to obtain non-degenerate Black-Scholes functions.

6.2.2.2 The put-call parity

While the constraints in Theorem (6.2.1) must be satisfied by the model to prevent static arbitrage of the implied volatility surface, they are also necessary to ensure the put-call parity. Given the parametric model in Equation (6.2.10), we are going to derive the constraint it must meet for the put-call parity to be satisfied. Taking the difference between a call and a put, we have

$$C_M(t_0,S_0;K,t)-P_M(t_0,S_0;K,t)=\frac{1}{norm}\sum_{i=1}^n \overline{a}_i(t)\big(C_{BS}(t_0,S_0;\overline{K}_i(K,t),t)-P_{BS}(t_0,S_0;\overline{K}_i(K,t),t)\big)$$

where $\overline{K}_i(K,t)=K^{'}(K,t)(1+\mu_i(t))$. From the classical call-put parity we must have

$$\forall i\ ,\ \overline{a}_i(t)\big(C_{BS}(t_0,S_0;\overline{K}_i(K,t),t)-P_{BS}(t_0,S_0;\overline{K}_i(K,t),t)\big)=\overline{a}_i(t)(S_0Re(t_0,t)-\overline{K}_i(K,t)P(t_0,t))$$

Further, since from Equation (5.1.8) we must have

$$C_M(t_0,S_0;K,t)-P_M(t_0,S_0;K,t)=S_0Re(t_0,t)-D_{PV}(t_0;t_0,t)-KP(t_0,t)$$

we can plug in that formula the put-call parity equation for a single i, getting

$$\frac{1}{norm}\sum_{i=1}^n \overline{a}_i(t)(S_0Re(t_0,t)-\overline{K}_i(K,t)P(t_0,t))=S_0Re(t_0,t)-D_{PV}(t_0;t_0,t)-KP(t_0,t)$$

Since $\frac{1}{norm}\sum_{i=1}^n \overline{a}_i(t)=\sum_{i=1}^n a_i(t)=1$ we get the constraint

$$\sum_{i=1}^n a_i(t)\overline{K}_i(K,t)P(t_0,t)=KP(t_0,t)+D_{PV}(t_0;t_0,t)$$

which simplifies to

$$K^{'}(K,t)P(t_0,t)+K^{'}(K,t)P(t_0,t)\sum_{i=1}^{n}a_i(t)\mu_i(t)=KP(t_0,t)+D_{PV}(t_0;t_0,t)$$

Given $K^{'}(K,t)=K+D(t_0,t)$ and since $D(t_0,t)P(t_0,t)=D_{PV}(t_0;t_0,t)$, the constraint becomes

$$\sum_{i=1}^{n}a_i(t)\mu_i(t)=0$$

Now, given $\mu_i(t)=\mu_i^0 f(t,\beta_i)$ and $a_i(t)=\frac{a_i^0}{f(t,\beta_i)\times norm}$, the constraint simplifies to

$$\sum_{i=1}^{n}a_i^0\mu_i^0=0$$

6.2.2.3 The term structure

At last, the no-arbitrage condition with respect to time holds if and only if the total variance $\omega(K,t)=\Sigma^2_{imp}(K,t)t$ is an increasing function of time t, that is,

$$\partial_t\omega_i(t)\geq 0\ ,\ \forall i \tag{6.2.5}$$

In order to satisfy the constraint in Equation (6.2.5) we can directly model the derivative of the total variance with respect to maturity, $F_i(t)=\partial_t\omega_i(t)$, as a positive function. Following this route, Gurrieri [2011] considered the function

$$F(t)=s^2_\infty+\big(Bt+s^2_0-s^2_\infty\big)e^{-\frac{t}{\tau}}$$

where s_∞, s_0, B, and τ are parameters satisfying the constraints $\tau>0$ and

$$B\geq 0$$

or

$$\begin{aligned} B &< 0 \\ s^2_\infty+B\tau e^{-1+\frac{s^2_0-s^2_\infty}{B\tau}} &\geq 0 \end{aligned}$$

for the function $F(\bullet)$ to be positive. We get the limit behaviour $\lim_{t\to 0}F(t)=s^2_0$ and $\lim_{t\to\infty}F(t)=s^2_\infty$. Integrating that function in the range $[0,t]$, we get the total variance term structure

$$\omega_i(t)=\big(-B_i\tau_i+s^2_{\infty,i}\big)t-\tau_i\big(B_it+B_i\tau_i+s^2_{0,i}-s^2_{\infty,i}\big)(e^{-\frac{t}{\tau_i}}-1)$$

Note, in the special case where $B_i=0$, the total variance simplifies to

$$\omega_i(t)=s^2_{\infty,i}t+\tau_i\big(s^2_{0,i}-s^2_{\infty,i}\big)(1-e^{-\frac{t}{\tau_i}})$$

which corresponds to the total variance of the Heston model described in Section (3.3.2.1).
Since we chose to model directly the square-root of the average variance, to guarantee the positivity of the local volatility we must verify

$$\Sigma^2_i(t)+2\Sigma_i(t)t\partial_t\Sigma_i(t)\geq 0$$

In the model above, the volatility becomes

$$\Sigma_i(t) = \frac{1}{\sqrt{t}} \sqrt{\big(- B_i \tau_i + s^2_{\infty,i} \big) t - \tau_i \big(B_i t + B_i \tau_i + s^2_{0,i} - s^2_{\infty,i} \big) (e^{-\frac{t}{\tau_i}} - 1)}$$

which simplifies to

$$\Sigma_i(t) = \sqrt{\big(- B_i \tau_i + s^2_{\infty,i} \big) - \frac{\tau_i}{t} \big(B_i t + B_i \tau_i + s^2_{0,i} - s^2_{\infty,i} \big) (e^{-\frac{t}{\tau_i}} - 1)}$$

and derivative with respect to time t as

$$\begin{aligned} \Sigma_i^{'}(t) &= \frac{1}{2} \Big(\big(- B_i \tau_i + s^2_{\infty,i} \big) - \frac{\tau_i}{t} \big(B_i t + B_i \tau_i + s^2_{0,i} - s^2_{\infty,i} \big) (e^{-\frac{t}{\tau_i}} - 1) \Big)^{-\frac{1}{2}} \\ &\times \Big(\frac{\tau_i}{t^2} \big(B_i \tau_i + s^2_{0,i} - s^2_{\infty,i} \big) (e^{-\frac{t}{\tau_i}} - 1) + \frac{1}{t} \big(B_i t + B_i \tau_i + s^2_{0,i} - s^2_{\infty,i} \big) e^{-\frac{t}{\tau_i}} \Big) \end{aligned}$$

which we can rewrite as

$$\Sigma_i^{'}(t) = \frac{1}{2\Sigma_i(t)} \Big(\frac{\tau_i}{t^2} \big(B_i \tau_i + s^2_{0,i} - s^2_{\infty,i} \big) (e^{-\frac{t}{\tau_i}} - 1) + \frac{1}{t} \big(B_i t + B_i \tau_i + s^2_{0,i} - s^2_{\infty,i} \big) e^{-\frac{t}{\tau_i}} \Big)$$

Another approach is to specify the time-dependent volatility $\Sigma_i(t)$ that capture the term structure of the implied volatility surface, and derive the constraints in Equation (6.2.3) that it must satisfy. In the special case where the user has no information on the term-structure of the implied volatility surface, we set $\Sigma_i(t) = d_i$ where $d_i > 0$. On all the other cases, the user can choose among different term-structures based on his information of the implied volatility surface. To make sure that the IV surface flattens for infinitely large expiries, we impose the limit behaviour

$$\lim_{t \to \infty} \Sigma_i(t) = d_i \tag{6.2.6}$$

For example, Bloch [2010b] discussed the volatility function

$$\Sigma_i(t) = \big(a_i + b_i t \big) e^{-c_i t} + d_i$$

where one must impose $b_i \geq 0$ for $\Sigma_i(.)$ to remain positive, generating only upward humps. Its derivative with respect to time is $\Sigma_i^{'}(t) = -c_i \big(a_i + b_i t \big) e^{-c_i t} + b_i e^{-c_i t}$ and its second derivative with respect to time is

$$\Sigma_i^{''}(t) = c_i^2 \big(a_i + b_i t \big) e^{-c_i t} - 2 c_i b_i e^{-c_i t}$$

However, if there are humps in the volatility surfaces they tend to be downwards humps. One possibility would be to let b_i be negative and to introduce a penalty in the target function by for instance letting the Black-Scholes price be arbitrarly high when $\Sigma_i(t) < 0$. Alternatively, we can solve an optimisation problem under constraints. The no-arbitrage constraint in Equation (6.2.3) must satisfy

$$\big(a_i + b_i t \big) e^{-c_i t} + d_i \geq -2t \big(- a_i c_i + b_i - b_i c_i t \big) e^{-c_i t}$$

which is true at $t = 0$ and $t = \infty$. For $t > \frac{1}{c_i}$ and c_i sufficiently large, the term d_i should dominate $(a_i + b_i t)$. Hence, at time $t = \frac{1}{c_i}$ we get

$$d_i \geq \big(a_i - \frac{b_i}{c_i} \big) e^{-1}$$

To get a general volatility function capable of generating both an upward hump or a downward one, Bloch [2012a] proposed the volatility function

$$\Sigma_i(t) = \big(a_i + b_i \ln{(1 + e_i t)} \big) e^{-c_i t} + d_i$$

where $c_i > 0$, $d_i > 0$ and $a_i \in \mathbb{R}$, $b_i \in \mathbb{R}$ and $e_i \in]-\frac{1}{t}, \infty[$. The derivative of the function with respect to time t is

$$\Sigma_i^{'}(t) = \big(- a_i c_i + b_i \frac{e_i}{1 + e_i t} - b_i c_i \ln{(1 + e_i t)} \big) e^{-c_i t}$$

and the no-arbitrage constraint must satisfy

$$e^{-c_i t} \big[(1 - 2tc_i) \big(a_i + b_i \ln{(1 + e_i t)} \big) + 2tb_i \frac{e_i}{1 + e_i t} \big] + d_i \geq 0$$

Since $a_i \in \mathbb{R}$, at time $t = 0$ the left hand side of the inequality can becomes negative. Consequently, we must impose the constraint

$$a_i + d_i > 0$$

to get the constraint satisfied at $t = 0$. Further, when $t > \frac{1}{c_i}$ then for c_i sufficiently large the constant d_i will dominates $\big(a_i + b_i \ln{(1 + e_i t)} \big)$ ensuring positivity of the left hand side. Hence, at time $t = \frac{1}{c_i}$ we must impose

$$d_i \geq \big(a_i - 2 \frac{b_i}{c_i} \frac{e_i}{1 + \frac{e_i}{c_i}} + b_i \ln{(1 + \frac{e_i}{c_i})} \big) e^{-1}$$

6.2.3 Some results

6.2.3.1 A simple test on the term structure

We are first going to test the type of term structures generated by directly modelling the derivative of the total variance with respect to maturity, $F(t)$. We want to achieve the steepest possible down hump for the term structure with $s_0 = 0.3$ and $s_\infty = 0.35$. To do so we set $B = -0.55$ and $\tau = 0.543$ obtaining the no-arbitrage constraint 1.9×10^{-6}. We dispay the volatility in Figure (6.1), and the total variance in Figure (6.2) and note that the minimum is reached at $t = 0.9$ which is not realistic in financial markets.

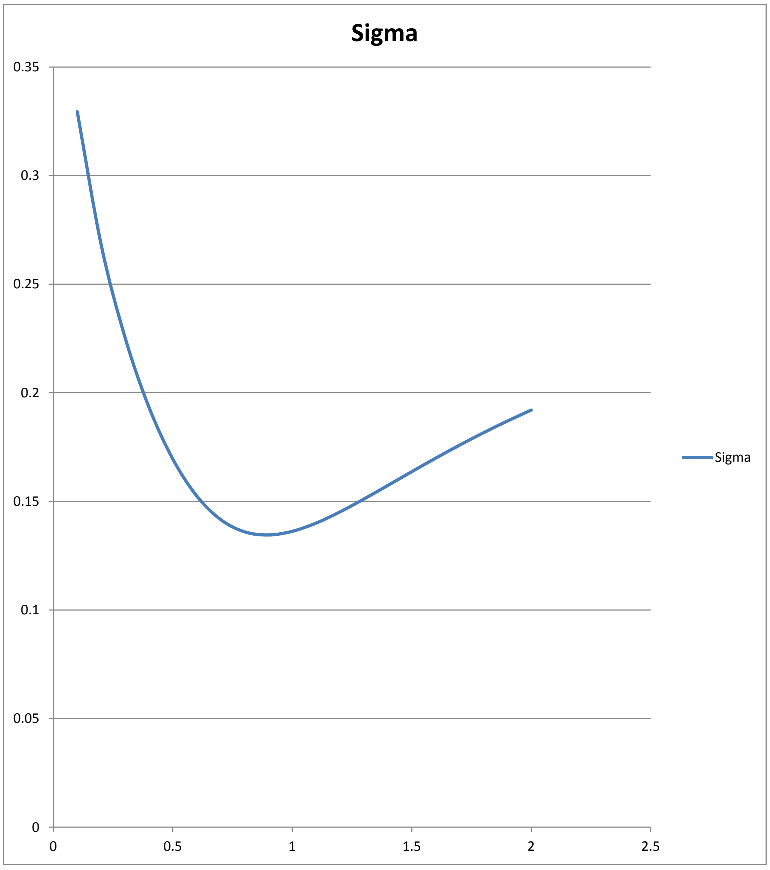


Figure 6.1: Term structure with $B = -0.55$ and $\tau = 0.543$.

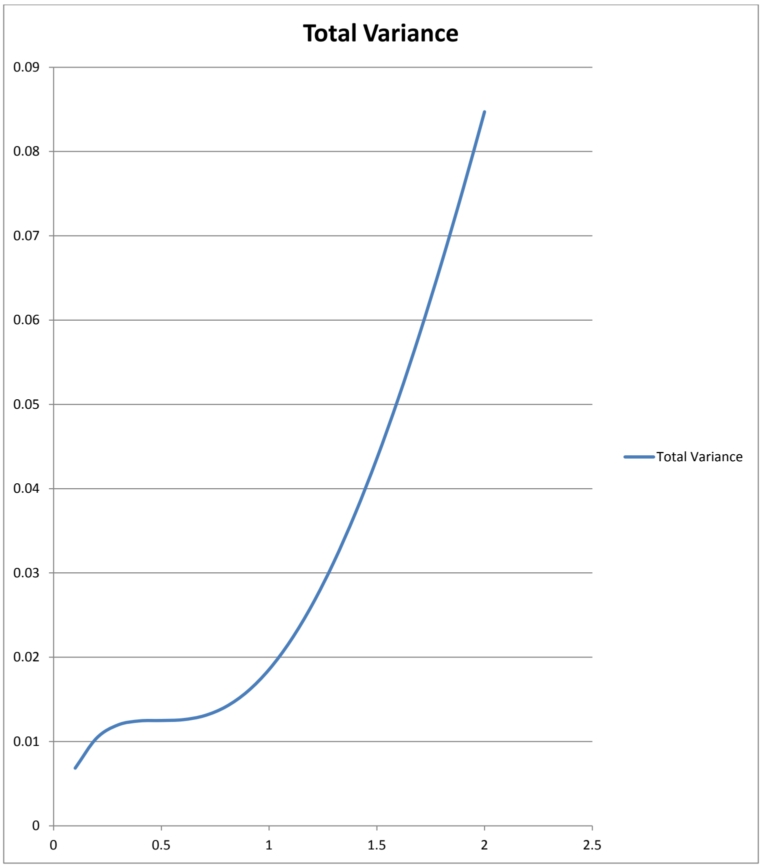


Figure 6.2: Total variance with $B = -0.55$ and $\tau = 0.543$.

6.2.3.2 Example of term structure

We are now going to test the capacity of different models to generate term structutre of volatility reaching the steepest possible down hump in a no-arbitrage setup. We constrain the initial volatility to be $s_0 = 0.35$ and the infinite volatility to be $s_\infty = 0.15$, and we calibrate the models to reach a minimum at the earliest possible maturity. We test the Heston model, the Gurrieri model, and Expo2 and Expo3 models with settings given in Table (6.1). For the Heston model the infinite variance is $\theta = 0.0225$ and the speed of mean reversion is $\kappa = 72$, while for the Gurrieri model we get $B = -3.5$ and the speed of mean reversion is $\kappa = 30$, leading to $\tau = \frac{1}{\kappa} = 0.033$, with a no-arbitrage constraint of 0.0042. We show the results in Figure (6.3) and display a zoom in Figure (6.4) for maturities in the range $[0, 0.1]$. Discretising time in the range $[0, 0.1]$ with time step $\Delta t = \frac{1}{150}$, we also display the values of the volatility as well as the no-arbitrage constraints in Table (6.2).

Fct	a	b	λ	σ_D	σ_A	σ_C	σ_B	σ_E
Expo2	0	0	0	0.2	-1.7	8.919	0.15	0
Expo3	0	0	0	0.2	-0.375	8.088	0.15	7.4
KumaExpo2	1.3	3	0.45	0.22	-2.144	9.015	0.179	0
KumaExpo3	1.3	3	0.45	0.23	-0.43	8.9	0.179	8.8

Table 6.1: BMW: Expo3 caibration

T	**Expo2**	**Const**	**Expo3**	**Const**	**Heston**	**Gurrieri**	**KE2**	**Const**	**KE3**	**Const**
0	0.35	0.35	0.35	0.35	0.35	0.35	0.336	0.399	0.344	0.409
0.006667	0.327	0.285	0.322	0.270	0.319	0.320	0.314	0.322	0.313	0.304
0.0133	0.307	0.229	0.297	0.205	0.294	0.294	0.293	0.257	0.287	0.222
0.02	0.288	0.182	0.276	0.153	0.274	0.272	0.275	0.201	0.263	0.157
0.02667	0.271	0.142	0.256	0.111	0.258	0.251	0.258	0.155	0.243	0.107
0.033	0.256	0.108	0.239	0.077	0.245	0.234	0.243	0.117	0.225	0.070
0.04	0.242	0.081	0.224	0.052	0.235	0.218	0.229	0.085	0.210	0.042
0.046667	0.229	0.058	0.210	0.032	0.226	0.205	0.217	0.060	0.196	0.022
0.0533	0.217	0.040	0.198	0.018	0.219	0.194	0.205	0.040	0.184	0.010
0.06	0.207	0.026	0.188	0.009	0.212	0.184	0.195	0.024	0.174	0.002
0.066667	0.197	0.015	0.178	0.003	0.207	0.176	0.186	0.013	0.165	0.00007
0.0733	0.189	0.008	0.170	0.0003	0.203	0.169	0.178	0.006	0.157	0.0008
0.08	0.181	0.003	0.163	0.00002	0.199	0.163	0.170	0.001	0.151	0.004
0.086667	0.174	0.0006	0.157	0.001	0.196	0.158	0.164	0.00006	0.145	0.010
0.0933	0.167	0.00001	0.151	0.005	0.193	0.154	0.158	0.0006	0.140	0.018
0.1	0.162	0.0010	0.146	0.010	0.190	0.151	0.152	0.003	0.136	0.026

Table 6.2: BMW: Expo3 caibration

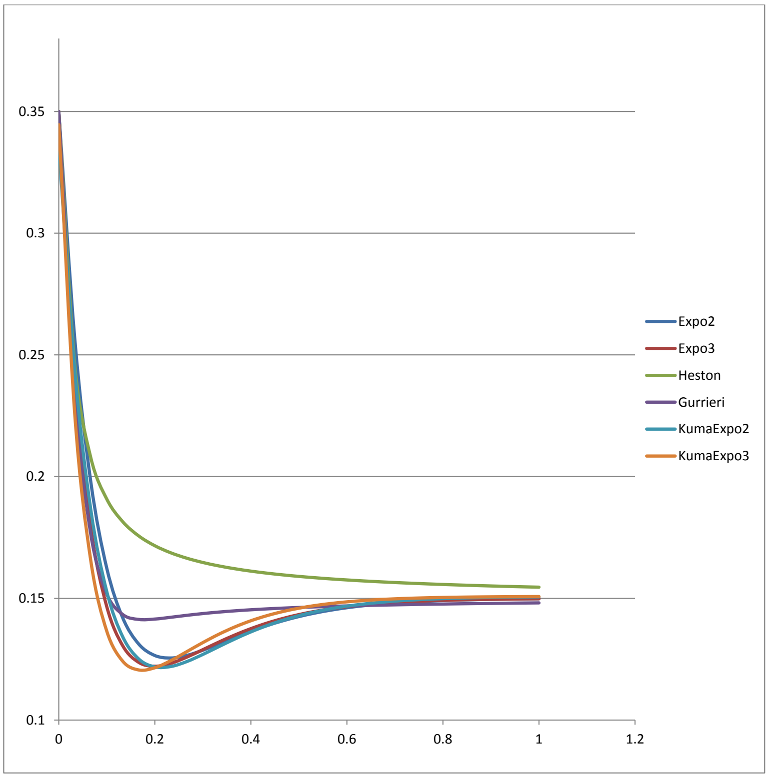


Figure 6.3: Term structure with $B = -0.55$ and $\tau = 0.543$.

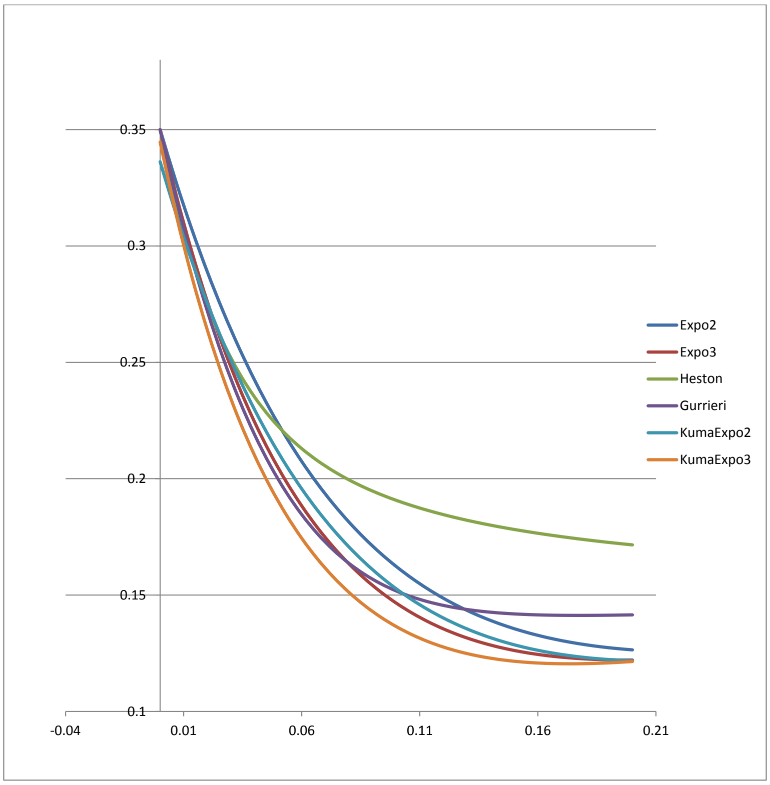


Figure 6.4: Term structure with $B = -0.55$ and $\tau = 0.543$.

6.2.3.3 Example of smile

As an example, we consider a set of American options on a single stock which we transform into European prices and calibrate with the MixVol model. The calculation time is $25/03/2013$, the spot price is 4.8. We display the smile in Figure (6.5) and the prices in Figure (6.6) for the maturity $27/06/2013$.

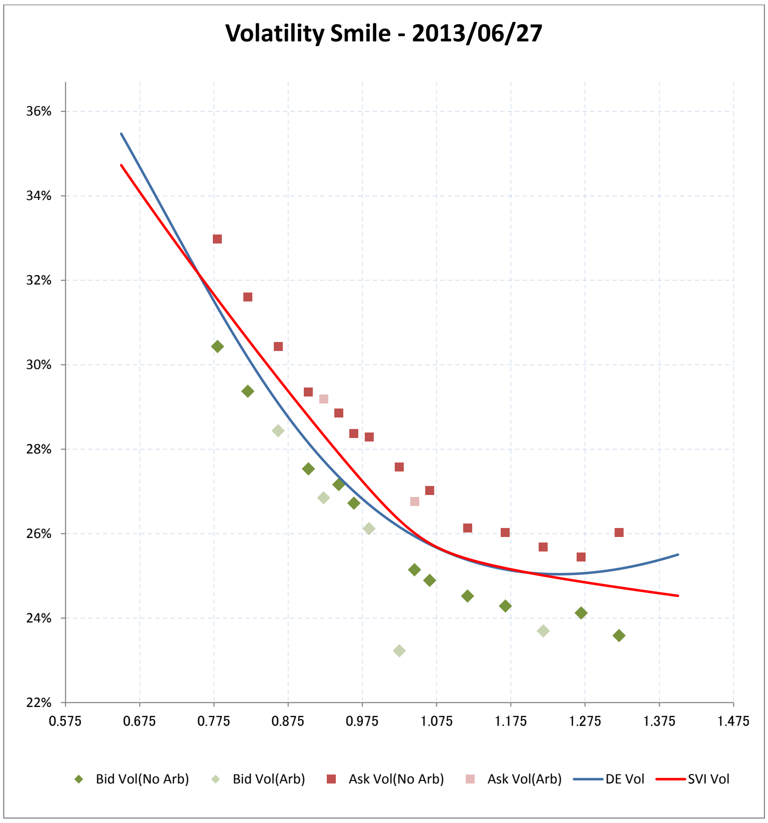


Figure 6.5: Smile for the maturity $27/06/2013$.

6.2.3.4 Example of volatility surface

As an example, we consider American options on the single stock BMW with evaluation date $15/04/2015$, spot level $S = 116.33$, and discrete dividends given in Table (6.3). The options' maturities are $17/04/2015$, $15/05/2015$, $19/06/2015$, $18/09/2015$, $18/12/2015$, $18/03/2016$, $17/06/2016$, and $16/12/2016$. The results of the calibration with the volatility model Expo3 are given in Table (6.4) and that with the volatility model ExpoTV are given in Table (6.5). The volatility surfaces generated by Expo3 and ExpoTV are given in Figure (6.7) and in Figure (6.8),

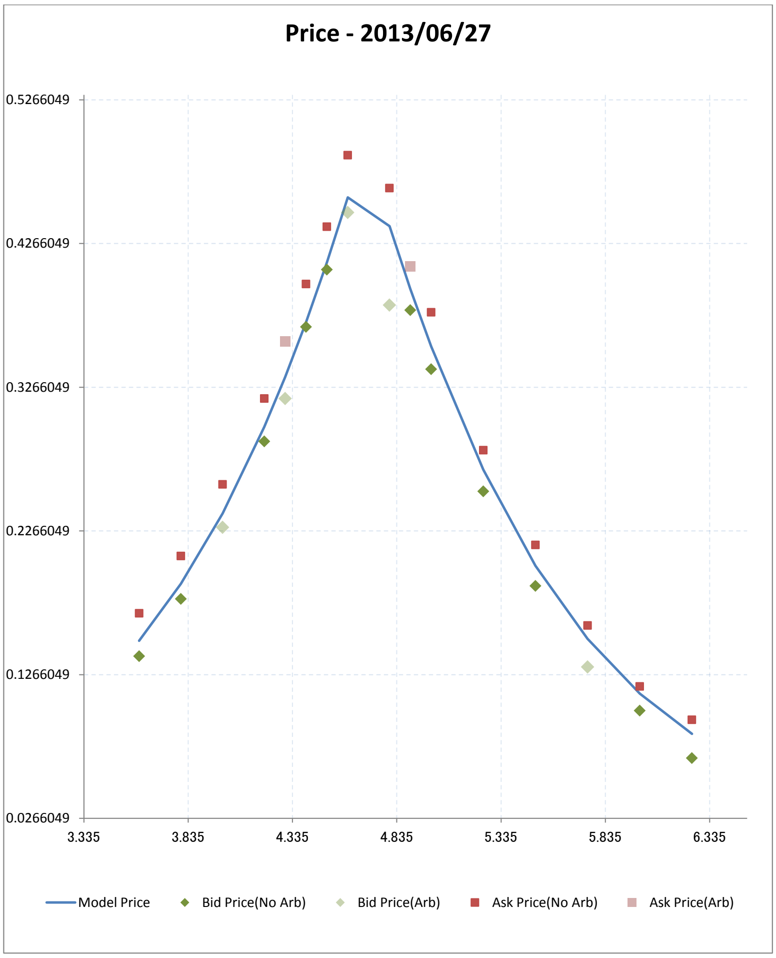


Figure 6.6: Prices for the maturity $27/06/2013$.

respectively, for maturities ranging from 1 week till 10 years. The weighted sum of term structures generated by Expo3 and ExpoTV are given in Figure (6.9) and in Figure (6.10), respectively, for maturities ranging from 0 to 2 years. We clearly see that the ExpoTV model is not capable of generating a steep downward hump as exhibited in the market.

Date	Cash
14/05/2015	2.73
16/05/2016	2.73
15/05/2017	2.73
15/05/2018	2.73
15/05/2019	2.73

Table 6.3: BMW: discrete dividends

Fct	μ_0	a_0	β_0	σ_D	σ_A	σ_C	σ_B	σ_E
1	0.2376	0.1449	0.7612	0.3929	-0.3000	54.3117	1.5994	45.1605
2	-0.0404	0.8550	0.4959	0.2094	0.1653	56.2481	-0.1742	49.5995

Table 6.4: BMW: Expo3 caibration

Fct	μ_0	a_0	β_0	σ_D	σ_A	σ_C	σ_B	σ_E
1	-0.1663	0.7488	3.5245	0.2093	0.2370	0.2915	-0.0158	0
2	0.4957	0.2511	0.4111	0.2567	0.9674	19.9976	-6.9933	0

Table 6.5: BMW: ExpoTV caibration

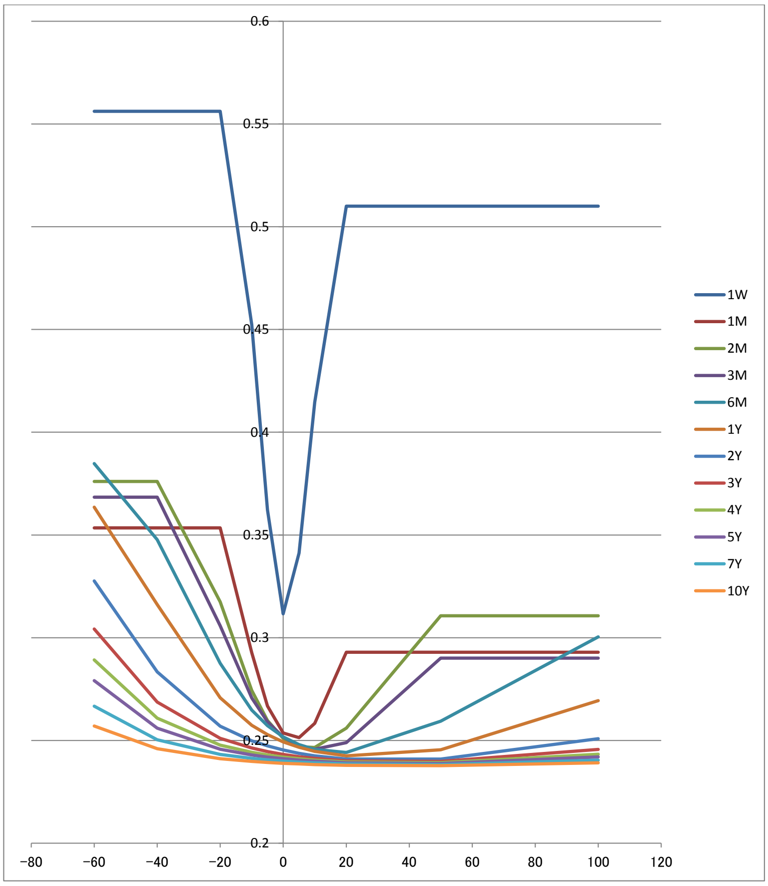


Figure 6.7: BMW: Expo3 volatility surface.

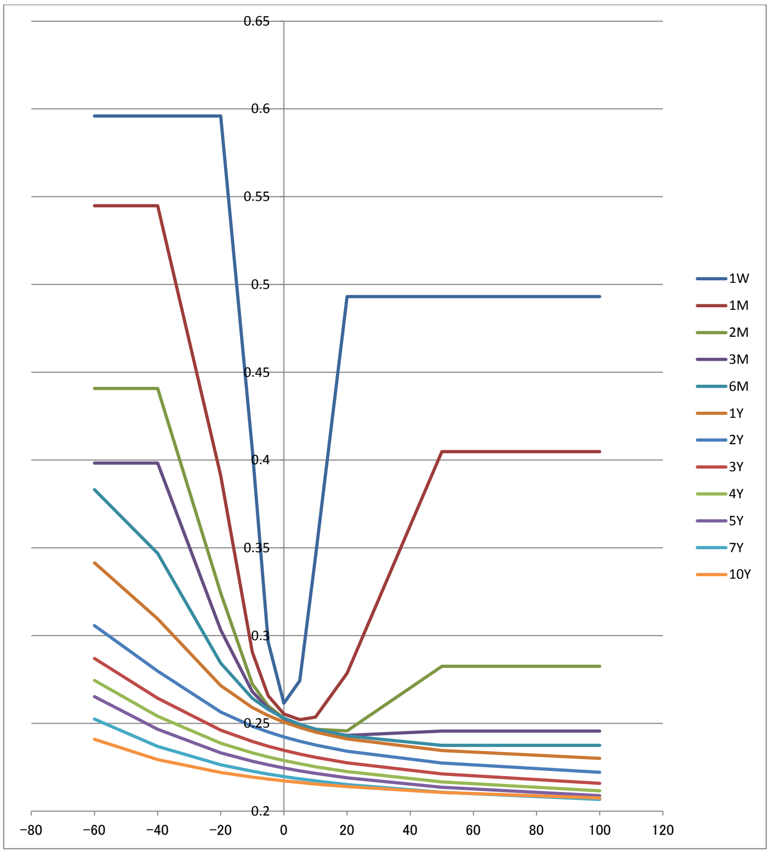


Figure 6.8: BMW: ExpoTV volatility surface.

6.2.4 An analysis of the implied volatility surface

We are going to expand the parametric MixVol model described in Equation (6.2.10) to infer the deformations of the implied volatility surface it generates. The main idea is to express each function $F_i(t_0, S_0, P_T, R_T, D_T; K, t)$ in terms of the standard Black-Scholes price plus some terms accounting for the skew and the curvature of the smile as follow

$$Call_{BS}(t_0, S_0, R_t, P_t, \overline{K}(K,t), t, \sigma) \approx C_{BS}(t_0, S_{t_0}, K^{'}(K,t), t; \sigma) + \alpha(K^{'}(K,t), t; \sigma)$$

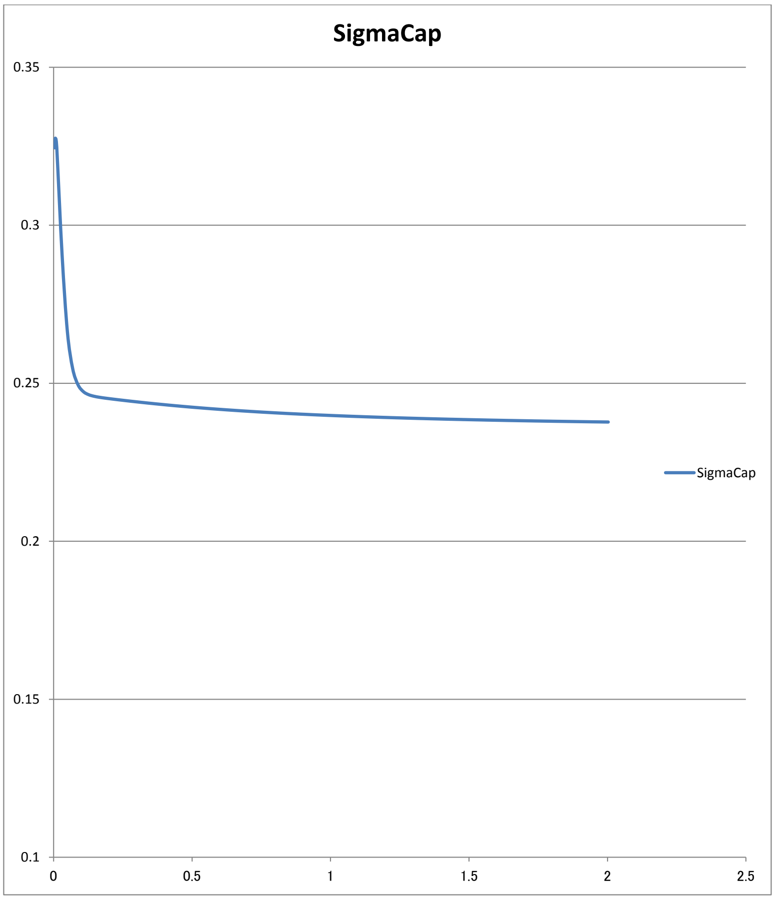


Figure 6.9: BMW: Expo3 weighted term structure.

As an example, considering the modifed Black-Scholes price in Appendix (3.1.1.2), we get

$$\alpha(K^{'}(K,t),t;\sigma) = S_{t_0}D(t_0,t)\big(N(\overline{d}_1) - N(d_1)\big) - K^{'}(K,t)P(t_0,t)\big(N(\overline{d}_2) - N(d_2)\big) - K^{'}(K,t)\mu_t P(t_0,t)N(\overline{d}_2)$$

with

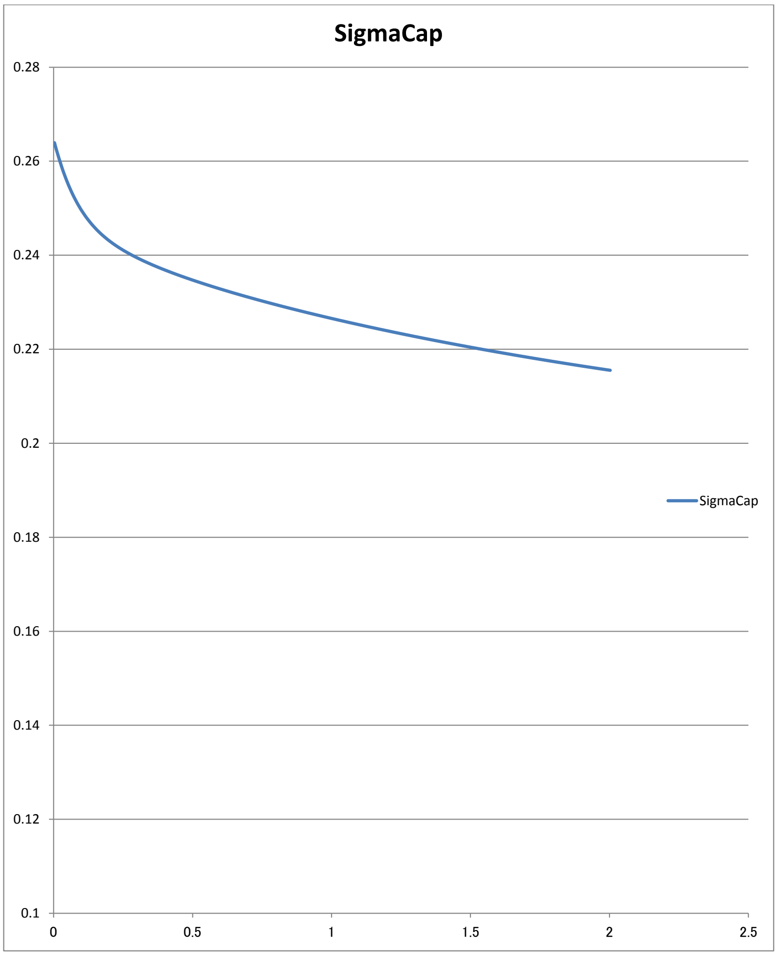


Figure 6.10: BMW: ExpoTV weighted term structure.

$$\overline{d}_1(t-t_0, x\frac{Re(t_0,t)}{P(t_0,t)}, K^{'}(K,t)) = d_1(t-t_0, x\frac{Re(t_0,t)}{P(t_0,t)}, K^{'}(K,t)) - \frac{1}{\sigma\sqrt{(t-t_0)}}\log{(1+\mu_t)}$$

In this model, $\lim_{\mu_t \to 0} \alpha(\sigma) = 0$, so that the parameter μ_t control the skew and the curvature of the volatility surface. Combining terms together, the parametric price model is approximated by

$$C_M(t_0,S_0,P_t,R_t,D_t;K,t) \approx \frac{1}{norm}\sum_{i=1}^n \overline{a}_i(t)C_{BS}(t_0,S_{t_0},K^{'}(K,t),t;\Sigma_i(t)) + \frac{1}{norm}\sum_{i=1}^n \overline{a}_i(t)\alpha_i(K^{'}(K,t),t;\Sigma_i(t))$$

where $\overline{a}_i(t) = \frac{a_i^0}{f(t,\beta_i)}$ for $i=1,..,n$. Note, the first term of the parametric model

$$y(0) = \frac{1}{norm}\sum_{i=1}^n \overline{a}_i(t)C_{BS}(t_0,S_{t_0},K^{'}(K,t),t;\Sigma_i(t))$$

is a weighted sum of lognormal prices generating a symetric smile. We define the pure implied volatility for the price $y(0)$ with strike K and maturity t as

$$\widehat{\sigma} : C_{BS}(K^{'}(K,t),t;\widehat{\sigma}(K^{'}(K,t),t)) - y(0) = 0 \tag{6.2.7}$$

There are various ways of estimating $\widehat{\sigma}(K^{'}(K,t),t))$, and for simplicity of exposition we use the property of lineartity of the Black-Scholes formula with respect to the volatility when the spot price S_0 is close to the at-the-money forward strike (see Equation (1.2.11)), by assuming

$$C_{BS}(t_0,t;\widehat{\sigma}) \approx \sum_{i=1}^n a_i C_{BS}(t_0,t;\sigma_i)$$

where $\widehat{\sigma} = \sum_{i=1}^n a_i\sigma_i$. That is, the implied volatility $\widehat{\sigma}$ is no-longer a function of the strike but only a function of time. We will call it the MixVol term structure. As a result, the approximated parametric model simplifies to

$$C_M(t_0,S_{t_0},P_t,R_t,D_t;K,t) \approx C_{BS}(t_0,S_{t_0},K^{'}(K,t),t;\widehat{\Sigma}(t)) + \alpha(t_0,S_{t_0},K^{'}(K,t),t)$$

where

$$\alpha(t_0,S_{t_0},K^{'}(K,t),t) = \frac{1}{norm}\sum_{i=1}^n \overline{a}_i(t)\alpha_i(K^{'}(K,t),t;\Sigma_i(t))$$

and

$$\widehat{\sigma} = \widehat{\Sigma}(t) = \frac{1}{norm}\sum_{i=1}^n \overline{a}_i(t)\Sigma_i(t) \tag{6.2.8}$$

By performing a Taylor series expansion around the weighted Black-Scholes formula $y(0)$, we deduce the approximated implied volatility for strike K and maturity t as follow

$$\begin{aligned} \Sigma(K^{'}(K,t),t) &\approx \widehat{\sigma} + \frac{1}{vega(\widehat{\sigma})}\alpha(t_0,S_{t_0},K^{'}(K,t),t) - \frac{1}{2}\frac{d_1(\widehat{\sigma})d_2(\widehat{\sigma})}{\widehat{\sigma}}\frac{1}{[vega(\widehat{\sigma})]^2}\alpha^2(t_0,S_{t_0},K^{'}(K,t),t) \\ &+ \frac{1}{6}\frac{d_1(\widehat{\sigma})d_2(\widehat{\sigma}) + d_2^2(\widehat{\sigma}) + d_1^2(\widehat{\sigma}) + 2d_1^2(\widehat{\sigma})d_2^2(\widehat{\sigma})}{\widehat{\sigma}^2}\frac{1}{[vega(\widehat{\sigma})]^3}\alpha^3(t_0,S_{t_0},K^{'}(K,t),t) + ...+ \end{aligned} \tag{6.2.9}$$

From this representation of the IVS we see that it is a highly non-linear function of the term structures $\Sigma_i(t)$ as well as the shift parameters $\mu_i(t)$. Further, given the properties of the Vega in Section (B.1.2), this IVS is not defined for short maturities $(T-t) << 0$ and for the strike range where the Vega is too small. One possibility, if one wants to directly use $\Sigma(K^{'}(K,t),t)$, is to cap the inverse Vega

$$\min\big(\frac{1}{vega(\widehat{\sigma})}, cap\big)$$

with a positive constant. Since the Black-Scholes Vega reaches its highest values at-the-money where the BS-formula is given in Equation (1.2.12), we choose

$$cap = 0.4x\sqrt{T-t}$$

6.2.5 Digital contracts in the MixVol model

Given our discrete dividends assumption on the dynamics of the underlying stock, we should always consider the modified strike $K^{'}(K,t) = K + D_t$ with event $\xi = \{Z_t > K^{'}\}$ when pricing a Digital contracts. We are now going to obtain analytical solution to the Digital Bond and the Digital Share in the special case where the instantaneous volatility of the stock price is a deterministic function of time and the stock price. Setting the interest rate to zero and multiplying the strike with the discount factor, the parametric model for a call option price of maturity t becomes

$$C_M(t_0,S_0,P_t,R_t,D_t;K,t) = \frac{1}{norm}\sum_{i=1}^n \overline{a}_i(t)Call_{BS}(t_0,S_0,R_t,1,\tilde{K}(K,t),t,\Sigma_i(t)) \quad (6.2.10)$$

where $\tilde{K}(K,t) = P(t_0,t)\overline{K}(K,t)$ and $\overline{K}(K,t) = K^{'}(K,t)(1+\mu_i(t))$ and $\overline{a}_i(t) = \frac{a_i^0}{f(t,\beta_i)}$ for $i = 1,..,n$. As a result, to each modified Black-Scholes formula corresponds the event $\xi_i = \{Z_t > \overline{K}\}$. Given the definition of the Black-Scholes formula, we can always rewrite the model call price as

$$\begin{aligned}
C_M(t_0,S_0,P_t,R_t,D_t;K,t) &= xRe(t_0,t)\frac{1}{norm}\sum_{i=1}^n \overline{a}_i(t)N(d_1^i(t-t_0,x,\tilde{K}(K,t)) \\
&- K^{'}(K,t)\frac{1}{norm}\sum_{i=1}^n \overline{a}_i P(t_0,t)(1+\mu_i(t))N(d_2^i(t-t_0,x,\tilde{K}(K,t)))
\end{aligned}$$

Given the derivative of the parametric model with respect to the strike estimated in Equation (6.3.13), and combining terms together, the parametric model for a call option becomes

$$C_M(t_0,S_0,P_t,R_t,D_t;K,t) = xRe(t_0,t)\frac{1}{norm}\sum_{i=1}^n \overline{a}_i(t)N(d_1^i(t-t_0,x,\tilde{K}(K,t)) + K^{'}(K,t)\frac{\partial}{\partial K}C_M(t_0,S_0,P_t,R_t,D_t;K,t)$$

Since the Digital Bond can be expressed in terms of a digital option as $D(S,t_0,t;\xi) = -\partial_K C(t_0,S_{t_0},t,K)$, its value in the parametric model is

$$D_M(S,t_0,t;\xi) = \frac{1}{norm}\sum_{i=1}^n \overline{a}_i(t)P(t_0,t)(1+\mu_i(t))N(d_2^i(t-t_0,x,\tilde{K}(K,t))) \quad (6.2.11)$$

which we can write as a shifted weighted sum of digital options on a shifted strike

$$D_M(S,t_0,t;\xi) = \frac{1}{norm}\sum_{i=1}^n \overline{a}_i(t)(1+\mu_i(t))D(S,t_0,t;\xi_i)$$

We can easily show that the Small-Moneyness Behaviour (SMB) defined by Roper [2010] is satisfied by construction in the MixVol model. Letting the strike tends to zero in the model, and using Equation (1.4.23) on each Black-Scholes formula, we get the limit

$$\lim_{K\to 0} D_M(S,t_0,t;\xi) = -\frac{P(t_0,t)}{norm}\sum_{i=1}^n \overline{a}_i(t)(1+\mu_i(t)) = -P(t_0,t)(\sum_{i=1}^n a_i(t) + \sum_{i=1}^n a_i^0\mu_i^0) = -P(t_0,t)$$

from the no-free lunch constraints in Equation (6.2.4). Similarly, by analogy to the call price in Equation (1.5.27), the Digital Share is

$$S_M(S,t_0,t;\xi)=xRe(t_0,t)\frac{1}{norm}\sum_{i=1}^{n}\overline{a}_i(t)N(d_1^i(t-t_0,x,\tilde{K}(K,t))$$

so that the call option becomes

$$C_M(t_0,S_0,P_t,R_t,D_t;K,t)=S_M(S,t_0,t;Z_t>K^{'})-K^{'}(K,t)D_M(S,t_0,t;Z_t>K^{'})$$

Given $\frac{\partial}{\partial d_2}N(d_2)=\frac{xRe(t_0,t)}{KP(t_0,t)}\frac{\partial}{\partial d_1}N(d_1)$ and $\frac{dd_2}{dx}=\frac{1}{x\sigma\sqrt{t-t_0}}$ the delta of a Digital option is

$$\frac{\partial}{\partial KS}C_M(t_0,S_0,P_t,R_t,D_t;K,t)=-\frac{1}{norm}\sum_{i=1}^{n}\overline{a}_i(t)P(t_0,t)(1+\mu_i(t))n(d_2^i(t-t_0,x,\tilde{K}(K,t)))\frac{1}{x\Sigma_i(t)\sqrt{t-t_0}}$$

6.2.6 Computing the Skew and Curvature analytically

For every model that one can define, we need to estimate the vector Ψ of model parameters from either the market prices or their implied volatility quotes. When the parametric model is calibrated to the market quotes and the optimum vector Ψ^* is obtained, the model call and put prices must equate the market ones. Since our model can retrieve analytically European prices as well as Digital price, we show how to infer analytically the skew and curvature of the IV surface for all strikes and maturities.

6.2.6.1 Computing the Skew

When the parametric model is calibrated to the market quotes and the optimum vector Ψ^* is obtained, the model Digital Bond in Equation (6.2.11) must equate the market digital price in Equation (1.5.28). As a result, we can infer analytically the skew of the parametric model for the strike K and the maturity t

$$Skew(K,t)=\frac{1}{Vega(K,t;\Sigma_{BS}(K,t))}\big[-\frac{\partial}{\partial K}C_{BS}(K,t;\Sigma_{BS})-D_M(S,t_0,t;\xi)\big]$$

Given the definition of the Black-Scholes digital option, we get

$$Skew(K,t)=\frac{1}{Vega(K,t;\Sigma_{BS}(K,t))}\big[D_{BS}(S,t_0,t;\xi)-D_M(S,t_0,t;\xi)\big]$$

which gives

$$Skew(K,t)=\frac{1}{Vega(K,t;\Sigma_{BS}(K,t))}\big[D_{BS}(S,t_0,t;\xi)-\frac{1}{norm}\sum_{i=1}^{n}\overline{a}_i(t)(1+\mu_i(t))D(S,t_0,t;\xi_i;\Sigma_i))\big]$$

When the shift terms are set to zero, $\mu_i(t)=0$ for $i=1,..,n$, the Digital Bond simplifies to $\frac{1}{norm}\sum_{i=1}^{n}\overline{a}_i(t)D(S,t_0,t;\xi;\Sigma_i)$ which is a weighted sum of digital option on a GBM. Note, it still has a skew but much less prononced.

6.2.6.2 Computing the Curvature

When the parametric model is calibrated to the market quotes, the parametric density in Equation (6.3.14) must equate the market density in Equation (1.5.29). From the formula of the convexity of the smile expressed in prices in Equation (1.7.45), we can infer analytically the curvature of the parametric model for the strike K and the maturity t

$$\partial_{KK}\Sigma(K,t) =$$
$$\frac{\partial_{KK}C_M(t_0,S_0;K,t)}{Vega(K,t;\Sigma_{BS}(K,t))} - \frac{1}{K^2\Sigma(K,t)(t-t_0)}\big[1+2Kd_1\sqrt{t-t_0}Skew(K,t)+K^2d_1d_2(t-t_0)(Skew(K,t))^2\big]$$

Given the optimum vector Ψ^* of model parameters, we can compute analytically the European call and put prices for all maturity t and strike K. Inverting the Black-Scholes formula, we recover the implied volatility surface $\Sigma(K,t)$. We can then use that surface to compute exactly the skew and curvature of the parametric IV surface. Alternatively, one can approximate the skew and curvature of the IV surface around the money by considering the ATM volatility $\widehat{\sigma} = \widehat{\Sigma}(t) = \frac{1}{norm}\sum_{i=1}^n \overline{a}_i(t)\Sigma_i(t)$. As a result, given a volatility function $\Sigma(K,t)$ obtained with a parametric function such as a SVI or SABR, one can directly fit our parametric model to volatilities without inverting the Black-Scoles formula.

6.2.7 A modified MixVol model

We described in Section (6.2.1) the MixVol model as a weighted sum of modified Black-Scholes formula with a term structure of volatility. Rather than using modified Black-Scholes formula in the sum, we are going to consider Bla Black-Scholes formula. In that setting, the market option price $C_M(K,t)$ of strike K and maturity t is estimated at time $t_0 = 0$ by the weighted sum

$$C_M(t_0,S_0,P_t,R_t,D_t;K,t) = \sum_{i=1}^n a_i(t)\sum_{j=1}^m Call_{BS}(t_0,S_0,R_t,P_t,\overline{K}(K,t),t,\Sigma_i(t)+v_j)\phi_{\widehat{V}}(v_j)dv_j \quad (6.2.12)$$

where $a_i(t)$ for $i=1,..,n$ are the weights, $\overline{K}(K,t) = K^{'}(K,t)(1+\mu_i(t))$ with $K^{'}(K,t) = K + D_t$, and $\phi_{\widehat{V}}(\bullet)$ is the lognormal density. Letting $\Sigma_{ij}(t) = \Sigma_i(t) + v_j$ be the new term structure of volatility, the total variance must be an increasing function of time. From the constraint in Equation (6.2.3), the volatility $\Sigma_{ij}(t)$ must satisfies

$$\Sigma_i(t) + v_j \geq -2t\partial_t\Sigma_i(t)$$

For positive v_j, the term structure $\Sigma_i(t)$ can now have a steeper decreasing slope while preseving the no-arbitrage condition. Following the same approach as in Section (6.2.2.2), we can prove that the new model satisfies the call-put parity. Taking the difference between a call and a put, from linearity we have

$$C_M(t_0,S_0;K,t) - P_M(t_0,S_0;K,t) =$$
$$\frac{1}{norm}\sum_{i=1}^n \overline{a}_i(t)\sum_{j=1}^m\big(C_{BS}(t_0,S_0;\overline{K}_i(K,t),t,\Sigma_{ij}(t)) - P_{BS}(t_0,S_0;\overline{K}_i(K,t),t,\Sigma_{ij}(t))\big)\phi_{\widehat{V}}(v_j)dv_j$$

To recover the call-put parity, the model must satisfy

$$\frac{1}{norm}\sum_{i=1}^n \overline{a}_i(t)\sum_{j=1}^m\big(S_0Re(t_0,t) - \overline{K}_i(K,t)P(t_0,t)\big)\phi_{\widehat{V}}(v_j)dv_j = S_0Re(t_0,t) - D_{PV}(t_0;t_0,t) - KP(t_0,t)$$

Since $S_0Re(t_0,t) - \overline{K}_i(K,t)P(t_0,t)$ does not depend on v_j, we can rewrite the equation as

$$\frac{1}{norm}\sum_{i=1}^n \overline{a}_i(t)\big(S_0Re(t_0,t) - \overline{K}_i(K,t)P(t_0,t)\big)\sum_{j=1}^m\phi_{\widehat{V}}(v_j)dv_j = S_0Re(t_0,t) - D_{PV}(t_0;t_0,t) - KP(t_0,t)$$

Further, $\sum_{j=1}^{m} \phi_{\widehat{V}}(v_j)dv_j = 1$, such that the equation simplifies to

$$\frac{1}{norm}\sum_{i=1}^{n}\overline{a}_i(t)\big(S_0Re(t_0,t)-\overline{K}_i(K,t)P(t_0,t)\big)=S_0Re(t_0,t)-D_{PV}(t_0;t_0,t)-KP(t_0,t)$$

and we recover the constraint for the MixVol model in Section (6.2.2.2).

6.3 Risk management with the MixVol model

Since the implied volatility surface (IVS) is stochastic, we saw in Section (6.1) that traders require an intuitive parametric representation of the smile allowing them to manually modify its shape when the underlying price moves. They must do so, because traders using the Black-Scholes model to hedge their book keep having to change the volatility assumption when the spot moves in order to match market prices. This stresses the necessity to take into consideration the dynamics of the volatility surface when pricing and risk-managing a portfolio of options. Thus, practitioners use the IVS as the market of primary hedging instruments against market risks, such as delta risk (first order sensitivity with respect to the spot), gamma risk (second order sensitivity with respect to the spot), and volatility risk (see Section (1.6)). In addition, model risk can lead to the mispricing of derivative products. The Bayesian literature considered model averaging approaches where model uncertainty is not distinguishable from market risk (see Hoeting et al. [1999]). However, as explained by Routledge et al. [2001], practitioners use probabilistic models to value the former, while the latter is valued through a worst case approach such as stress testing of portfolios. Since the IV surface depends on the strike and maturity of an option, one must consider the risks associated with its dynamics. For instance, risk managers use stress scenarios defined on the IV surface to visualise and quantify the risk inherent to option portfolios. Therefore, one distinguishes between risk uncertainty on outcomes with known probabilities and ambiguity or model uncertainty when several specifications are possible for such probabilities. We are now going to explain how the Greeks and stress scenarios are calculated analytically in the parametric MixVol model described in Equation (6.2.10), without recalibration of the model parameters.

6.3.1 Computing the Greeks

6.3.1.1 Analytical solutions for European options

Assuming that the dynamics of the spot price with discrete dividends follow the Spot model, our parametric model in Equation (6.2.10) corresponds to a weighted sum of Black-Scholes formulas in the Z-space defined in Section (5.1.1.2) with $a(t,T)=1$ and $b(t,T)=-D(t,T)$. In that setting, the derivative of a call option price with respect to the spot is

$$\frac{\partial}{\partial S_0}C_M(t_0,S_0,P_t,R_t,D_t;K,t)=\frac{1}{norm}\sum_{i=1}^{n}\overline{a}_i(t)\frac{\partial}{\partial S_0}Call_{BS}(t_0,S_0,R_t,1,\tilde{K}(K,t),t,\Sigma_i(t))$$

where $\tilde{K}(K,t)=P(t_0,t)\overline{K}(K,t)$ and $\overline{K}(K,t)=K^{'}(K,t)(1+\mu_i(t))$ and $\overline{a}_i(t)=\frac{a_i^0}{f(t,\beta_i)}$. Differentiating one more time the call price with respect to the spot, we get

$$\frac{\partial^2}{\partial S_0^2}C_M(t_0,S_0,P_t,R_t,D_t;K,t)=\frac{1}{norm}\sum_{i=1}^{n}\overline{a}_i(t)\frac{\partial^2}{\partial S_0^2}Call_{BS}(t_0,S_0,R_t,1,\tilde{K}(K,t),t,\Sigma_i(t))$$

Differentiating the call price with respect to the strike, we get

$$\frac{\partial}{\partial K}C_M(t_0,S_0,P_t,R_t,D_t;K,t)=\frac{1}{norm}\sum_{i=1}^{n}\overline{a}_i(t)\frac{\partial}{\partial\tilde{K}}Call_{BS}(t_0,S_0,R_t,1,\tilde{K}(K,t),t,\Sigma_i(t))\frac{\partial\tilde{K}}{\partial K}\quad(6.3.13)$$

where $\frac{\partial \tilde{K}}{\partial K} = P(t_0,t)(1+\mu_i(t))$. Differentiating one more time the call price with respect to the strike, we get

$$\frac{\partial^2}{\partial K^2} C_M(t_0,S_0,P_t,R_t,D_t;K,t) = \frac{1}{norm} \sum_{i=1}^n \overline{a}_i(t) \frac{\partial^2}{\partial (\tilde{K})^2} Call_{BS}(t_0,S_0,R_t,1,\tilde{K}(K,t),t,\Sigma_i(t)) (\frac{\partial \tilde{K}}{\partial K})^2 \tag{6.3.14}$$

Differentiating the call price with respect to time t_0, we get

$$\begin{aligned}
& \frac{\partial}{\partial t_0} C_M(t_0,S_0,P_t,R_t,D_t;K,t) = \frac{d}{dt_0}(\frac{1}{norm}) C_M(t_0,S_0,P_t,R_t,D_t;K,t) norm \\
+ \quad & \frac{1}{norm}\Big(\sum_{i=1}^n \frac{d}{dt_0}\overline{a}_i(t) Call_{BS}(t_0,S_0,R_t,1,\tilde{K}(K,t),t,\Sigma_i(t)) \\
+ \quad & \sum_{i=1}^n \overline{a}_i(t) \frac{d}{dt_0}\tilde{K}(K,t) \frac{\partial Call_{BS}}{\partial \tilde{K}}(t_0,S_0,R_t,1,\tilde{K}(K,t),t,\Sigma_i(t)) \\
+ \quad & \sum_{i=1}^n \overline{a}_i(t) \frac{d}{dt_0}\Sigma_i(t) \frac{\partial Call_{BS}}{\partial \Sigma_i}(t_0,S_0,R_t,1,\tilde{K}(K,t),t,\Sigma_i(t)) \\
+ \quad & \sum_{i=1}^n \overline{a}_i(t) \frac{\partial Call_{BS}}{\partial t_0}(t_0,S_0,R_t,1,\tilde{K}(K,t),t,\Sigma_i(t))\Big)
\end{aligned}$$

where

$$\frac{d}{dt_0}\tilde{K}(K,t) = \frac{d}{dt_0}D(t_0,t)P(t_0,t)(1+\mu_i(t)) + r_{t_0}\tilde{K}(K,t) + P(t_0,t)(K+D(t_0,t))\frac{d}{dt_0}(1+\mu_i(t))$$

with $\frac{d}{dt_0}\mu_i(t) = \mu_i^0 \frac{d}{dt_0} f(t,\beta_i)$. Note, when there is no dividends between t_0 and $t_0+\epsilon$ then the cumulative dividends term $D(t_0,t) = D_t$ does not depend on t_0 and the derivative becomes $\frac{d}{dt_0}\overline{K}(K,t) = r_{t_0}\tilde{K}(K,t) + P(t_0,t)(K + D(t_0,t))\frac{d}{dt_0}(1+\mu_i(t))$.

6.3.1.2 Computing the Greeks for exotic options

In the case of exotic options, the Greeks are obtained by bumping the relevant variable, while keeping the other ones constant, and computing the difference in the resulting prices. For example, in the case of the Delta and the Gamma, we bump the initial spot price for a fixed level of interest rate and volatility. In order to compute the Greeks for exotic options in a fast and robust way, we are going to judiciously bump the parameters of the MixVol model described in Section (6.2) in such a way that it corresponds to a particular shift of the implied volatility surface.
Since the MixVol model is a weighted sum of Black-Scholes functions, the results described in Section (1.2.3) on the properties of the Black-Scholes prices apply to the MixVol model. Further, we saw in Section (4.1.3.4) that the dynamics of the BS-volatility correspond to the sticky delta regime, and that multiplying the strike with $(1+\epsilon)$ leaves the smile unchanged. Thus, we distinguish two types of Delta:

1. The fixed delta: The spot price is bumped but the IVS is unchanged, which correspond to the sticky strike regime of volatility described in Section (4.1.1.2).

2. The floating delta: The spot price is bumped and the strike of the option rescaled to preserve moneyness, which correspond to the sticky delta regime of volatility.

Remark 6.3.1 *Since in the MixVol model the strike is shifted by the cumulative dividends $K^{'}(K,t) = K + D_t$, the regimes of volatility must apply to the strike $K^{'}$.*

Hence, for a Fixed Delta with an upwind finite difference and a bumped spot $S_t = S_0 + \alpha$, where $\alpha = \epsilon S_0$, multiplying the strike with $(1+\epsilon)$, the shifted option price becomes

$$\begin{aligned} C_M(t_0, S_0, P_t, R_t, D_t; K, t; \epsilon) &= \frac{1}{norm} \sum_{i=1}^{n} \overline{a}_i(t) Call_{BS}(t_0, S_0(1+\epsilon), R_t, P_t, \overline{K}(K,t)(1+\epsilon), t, \Sigma_i(t)) \\ &= (1+\epsilon) C_M(t_0, S_0, P_t, R_t, D_t; K, t) \end{aligned}$$

Similarly, for a Floating Delta, the shifted option price in our parametric model becomes

$$\begin{aligned} C_M(t_0, S_0, P_t, R_t, D_t; K, t; \epsilon) &= \frac{1}{norm} \sum_{i=1}^{n} \overline{a}_i(t) Call_{BS}(t_0, S_0(1+\epsilon), R_t, P_t, \overline{K}(K,t), t, \Sigma_i(t)) \\ &= (1+\epsilon) C_M(t_0, S_0, P_t, R_t, D_t; K \frac{1}{(1+\epsilon)}, t) \end{aligned}$$

In that setting, given the strike K at time t, the old strike is $K \frac{S_0}{S_0+\epsilon}$, or vice versa, given K at inception $t_0 = 0$, the new strike is $K \frac{S_t}{S_0} = K \frac{S_0+\alpha}{S_0}$. It is interesting to note that the Floating Delta obtained with finite difference is an approximation to the weighted sum of Black-Scholes deltas

$$C_M^{\Delta}(t_0, S_0, P_t, R_t, D_t; K, t) = \frac{1}{norm} \sum_{i=1}^{n} \overline{a}_i(t) Call_{BS}^{\Delta}(t_0, S_0, R_t, P_t, \overline{K}(K,t), t, \Sigma_i(t))$$

Further, assuming that the discrete dividend impact is negligable between t_0 and $t_0 + \epsilon$, we bump the initial time with a constant ϵ to compute the shifted price. It is an approximation to differentiating the parametric model with respect to time t_0 in order to get the annualised analytic Theta. One can show that

$$\frac{\partial C_M^{\Theta}(t_0, S_0, P_t, R_t, D_t; K, t)}{\partial t_0} = \frac{\partial C_M^{\Theta}(t_0, S_0, P_t, R_t, D_t; K, t)}{\partial \tau} \frac{d\tau}{dt_0}$$

where $\tau = t - t_0$. Hence, since our parametric model is homogeneous in τ, when $t_0 = 0$, we can use the derivative of the price with respect to the maturity t with formula given in Section (6.3.1.1).

6.3.1.3 Controlling the Greeks

Since practitioners can not continuously modify the shape of the implied volatility surface (IVS) when the underlying price moves, they use rules of thumb to compute the Greeks by considering some standard evolutions of the IVS, which provides an indication of some possible behaviour for the smile that one might expect (see Section (4.1.1.2)). Assuming that the evolution of the IVS only depends on time and the spot price, $\Sigma = \Sigma(t, S_t, K, T)$, we obtain the implied delta in Equation (4.1.1) and the implied gamma in Equation (4.1.2), which are functions of the first derivative of the IV with respect to the spot price $\partial_x \Sigma$, and the second derivative of the IV with respect to the spot price $\partial_{xx} \Sigma$. Further, we saw in Section (4.1.4) that these variations could be related to the Skew and Curvature of the implied volatility surface, respectively, via Equation (4.1.3) and Equation (4.1.4). Fortunately, in Section (6.2.6), we computed the Skew and Curvature analytically in the MixVol model. Thus, after calibration, we infer the Black-Scholes volatility $\Sigma_{BS}(K,T)$ for the pair (K,T) from the model price $C_M(K,T)$, and we compute the implied delta as

$$\partial_x C_M(t, S_t, K, T) = \partial_x C_{BS}(t, S_t, K, T; \Sigma) + \partial_\Sigma C_{BS}(t, S_t, K, T; \Sigma) k_S \frac{\partial \Sigma}{\partial K} \tag{6.3.15}$$

and the implied gamma as

$$\partial_{xx}C_M(t,S_t,K,T) = \partial_{xx}C_{BS}(t,S_t,K,T;\Sigma)+2\partial_{x\Sigma}C_{BS}(t,S_t,K,T;\Sigma)k_S\frac{\partial\Sigma}{\partial K} \quad (6.3.16)$$
$$+ \partial_{\Sigma\Sigma}C_{BS}(t,S_t,K,T;\Sigma)\big(k_S\frac{\partial\Sigma}{\partial K}\big)^2+\partial_\Sigma C_{BS}(t,S_t,K,T;\Sigma)k_C\frac{\partial^2\Sigma}{\partial K^2}$$

where k_S and k_C satisfy

$$k_S = \begin{cases} 0 \text{ if sticky strike} \\ 1 \text{ if sticky implied tree} \\ -\frac{K}{S_t} \text{ if sticky delta} \\ \frac{K}{S_t} \text{ if minimum variance} \end{cases} \quad k_C = \begin{cases} 0 \text{ if sticky strike} \\ 1 \text{ if sticky implied tree} \\ (\frac{K}{S_t})^2 \text{ if sticky delta and minimum variance} \end{cases}$$

6.3.2 Computing the Vegas

When it comes to Vega hedging a book of options, one must create a realistic range of scenarios of volatility surface which are not arbitragable. As explained in Section (1.4.2.3), one can not blindly impose dynamics on the IV surface, as a Vega obtained by bumping the IV surface with parallel shifts is inconsistent. Given the need for a volatility model, we described in Section (6.2) a tractable parametric model satisfying no-arbitrage conditions both in space and in time. Even though the model generates a robust and smooth implied volatility surface and allows for analytic local volatility, it is difficult for the traders to apprehend the deformation of the implied volatility surface it produces when bumping one or more model parameters. To gain insight in the types of volatility surfaces the model can generate, we performed in Section (6.2.4) a simple expansion of the parametric model obtaining a representation of its implied volatility surface. We showed that the first order approximation of the IVS in the MixVol model was the parametric term structure $\widehat{\Sigma}(t)$ given in Equation (6.2.8), which is independent from the strike. Using the analytical properties of the MixVol model we can therefore obtain a simple representation of the Vega given by

$$\frac{\partial}{\partial\widehat{\Sigma}}C_M(t_0,S_0,P_t,R_t,D_t;K,t) = \frac{1}{norm}\sum_{i=1}^n\overline{a}_i(t)\frac{\partial}{\partial\Sigma_i}Call_{BS}(t_0,S_0,R_t,1,\tilde{K}(K,t),t,\Sigma_i(t))\frac{\partial\Sigma_i}{\partial\widehat{\Sigma}}$$

Since $\frac{\partial\Sigma_i}{\partial\widehat{\Sigma}}=\frac{norm}{\overline{a}_i(t)}$, the Vega simplifies to

$$\frac{\partial}{\partial\widehat{\Sigma}}C_M(t_0,S_0,P_t,R_t,D_t;K,t) = \sum_{i=1}^n\frac{\partial}{\partial\Sigma_i}Call_{BS}(t_0,S_0,R_t,1,\tilde{K}(K,t),t,\Sigma_i(t))$$

Further, from the expansion of the IVS, we were capable of understanding how to modify our parametric model price in order to generate a parallel shift, a skew shift, or a curvature shift of the implied volatility surface. Hence, in view of adding control to the generated volatility surface, we propose to modify the model in Equation (6.2.10) with an extra term as follow

$$C_M(t_0,S_0,P_t,R_t,D_t;K,t) = \frac{1}{norm}\sum_{i=1}^n\overline{a}_i(t)Call_{BS}(t_0,S_0,R_t,P_t,\overline{K}(K,t),t,\Sigma_i(t)) \quad (6.3.17)$$
$$+ vega_{BS}(K^{'}(K,t),t;\widehat{\sigma}(K^{'}(K,t),t))h(K,t;\epsilon_{ATM},\epsilon_S,\epsilon_C)$$

where $\widehat{\sigma}(K^{'}(K,t),t))$ satisfy Equation (6.2.7) and where

$$h(K,t;\epsilon_{ATM},\epsilon_S,\epsilon_C) = \big(\epsilon_{ATM}-\epsilon_S\ln\frac{K^{'}(K,t)}{x}+\epsilon_C(\ln\frac{K^{'}(K,t)}{x})^2\big)$$

with $\epsilon_{ATM} << 1$, $\epsilon_S << 1$, and $\epsilon_C \in \mathbb{R}^+$. The parameter ϵ_{ATM} produces a parallel shift of the volatility surface and should be of the order of the percentage (for example 1%). The skew ϵ_S is given by $\frac{\partial}{\partial K}\Sigma(t,S_t;K,T-t)$, and using finite differences, it is expressed as a few percents per 10% change in the strike. For example, a skew of 2% per 10% change in the strike is 0.2. Similarly, using finite differences to discretise the curve $\frac{\partial^2}{\partial K^2}\Sigma(t,S_t;K,T-t)$ in the previous example, gives $\epsilon_C = \frac{4\%}{(10\%)^2} = 4$. One can calibrate the model parameters a_i^0, μ_i^0, and $\Sigma_i(t)$ to the market and then as time evolves modify manually ϵ_{ATM} to generate parallel shift to the implied volatility surface, ϵ_S to generate skew shift to the implied volatility surface, and ϵ_C to generate curve shift to the implied volatility surface. The only additional constraint is for the price in Equation (6.3.17) to remain positive. Note, the parameters ϵ_{ATM}, ϵ_S and ϵ_C can also be used to generate analytically the new local volatility when computing the Vegas of an option.

6.3.3 Scenarios analysis

As shocks across the IV surface are highly correlated, its dynamics can be decomposed into a small number of driving factors (see Fengler [2005]). Consequently, rather than bumping the volatility surface for some given strikes and maturities when computing the Vega, we choose to consider a range of plausible scenarios for the dynamics of the implied volatility surface. As explained in Section (1.4.2), consistent scenarios are difficult to identify, and one should avoid letting the IV surface moves up and down with parallel shifts. Using time series of option prices, and applying Principal Component Analysis to the implied volatility surface, Cont et al. [2002] identified three orthogonal random factors explaining most of its dynamics (the first three eigenmodes account for about 98% of the variance). Usually:

- The first factor represent a parallel shift of the IVS called the level factor.
- The second eigenmode changes sign at the money, such that a positive shock along that direction increases the volatility of out-of-the-money calls, while decreasing those of out-the-money puts. Hence, by biasing the IVS towards high strikes, positive movements in that direction increases the skewness of the risk neutral density.
- The third eigenmode is a butterfly mode corresponding to a change in the convexity of the surface. Movements along this direction contribute to the flattening of both tails of the risk neutral density, but they only contribute to about 0.8% of the overall variance.

This approach provides a decomposition of volatility risk as a sum of contributions from empirically identifiable factors. Studying the properties of the mixture model of prices described in Section (6.2), each Black-Scholes function is composed of two term structure of volatility $\Sigma_i(t)$ controlling the behaviour of the IVS along the money, but also along the out-of-the money strikes. In addition, each function $f_i(t)$ controls the weights and the departure from a symmetric smile. Hence, to be consistent with the factor decomposition of the IVS, we are going to bump the term structure $\Sigma_i(t)$, the function $f_i(t)$ as well as the shift function $\mu_i(t)$ for each Black-Scholes function of our model with a constant $\epsilon > 0$.

6.3.3.1 Arbitrageable scenarios

Note, when the implied volatility surface experience some skew and curvature, then some of the parameters μ_i^0 will be positive while others will be negative. Hence, bumping each parameter μ_i^0 with the same positive constant will increase the volatility of out-of-the-money calls while decreasing those of out-the-money puts, which is exactly the behaviour observed empirically by the second eigenmode. Consequently, under that scenario, one of the shifted option price becomes

$$C_M(t_0,S_0,P_t,R_t,D_t;K,t;\epsilon) = \frac{1}{norm}\sum_{i=1}^n \overline{a}_i(t)Call_{BS}(t_0,S_0,R_t,P_t,K^{'}(K,t)(1+(\mu_i^0+\epsilon)f_i(t)),t,\Sigma_i(t))$$

where ϵ is a constant of the order of magnitude of μ_i^0, and such that $\mu_i^0+\epsilon \geq -1$. In that scenario, the martingale property of the induced risk-neutral pdf given in Equation (6.2.4) becomes

$$\sum_{i=1}^{n} a_i^0 \mu_i^0 = \epsilon$$

that is, we are ϵ away from the no-arbitrage.

6.3.3.2 Non-arbitrageable scenarios

The scenarios defined in the previous section no-longer satisfy the no-arbitrage constraints given in Equation (6.2.4). We are now going to modify the model parameters such that the the no-arbitrage constraints are satisfied. In order to obtain an upward or downward movement of the skew and curvature on the IVS, we are going to bump each function's curve $f_i(t)$ inside the shift function $\mu_i(t)$ with a constant. As a result, one of the shifted parametric price becomes

$$C_M(t_0, S_0, P_t, R_t, D_t; K, t; \epsilon) = \frac{1}{norm} \sum_{i=1}^{n} \overline{a}_i(t) Call_{BS}(t_0, S_0, R_t, P_t, K^{'}(K,t)(1 + \mu_i^0(f_i(t) + \epsilon)), t, \Sigma_i(t))$$

Rather than modifying the function $f_i(t)$, we can directly modify the parameter β_i in the previous parametric price $C_M(t_0, S_0; K, t)$, getting $f(t, \beta_i(1 + \epsilon))$ with $\epsilon << 1$. In addition, multiplying each parameter μ_i^0 with the same constant $(1 + \epsilon)$, rather than adding a constant, will preserve the no-arbitrage constraints. Hence, under that scenario, the shifted option price becomes

$$C_M(t_0, S_0, P_t, R_t, D_t; K, t; \epsilon) = \frac{1}{norm} \sum_{i=1}^{n} \overline{a}_i(t) Call_{BS}(t_0, S_0, R_t, P_t, K^{'}(K,t)(1 + (1 + \epsilon)\mu_i^0 f_i(t)), t, \Sigma_i(t))$$

where $\epsilon << 1$. Alternatively, we can multiply each parameter μ_i^0 with the constant $(1 + \epsilon_i)$ if we force one degree of liberty as

$$\epsilon_1 = -\frac{1}{a_1^0 \mu_1^0} \sum_{i=2}^{n} a_i^0 \mu_i^0 (1 + \epsilon_i)$$

One can also bump each μ_i^0 with ϵ_2 for $i = 1, .., n$, but in order to satisfy the no-arbitrage constraint we also need to bump each weight a_i^0 with ϵ_1 such that

$$\epsilon_1 = -\frac{\epsilon_2}{\sum_{i=1}^{n} \mu_i^0 + n\epsilon_2}$$

As the representation of the IVS in Equation (6.2.9) is a highly non-linear function of the term structures $\Sigma_i(t)$ as well as the shift parameters $\mu_i(t)$, adding a positive constant ϵ to each term structure $\Sigma_i(t)$ will produce a parallel shift in the MixVol term structure $\widehat{\sigma} = \sum_{i=1}^{n} a_i(t)\Sigma_i(t)$ but not in the IVS itself. That is, for $t_1 > t_0$ the movement of the new implied volatility surface $\Sigma(t_1; K^{'}, t)$ produced by that scenario depends on the shape of the IVS at inception $\Sigma(t_0; K^{'}, t)$ which is in accordance with the result found by Rogers et al. [2010] given in Equation (1.4.17). The smaller the skew, the larger the impact of $\widehat{\sigma}$ on the IV surface. As a result, we can shift each BS function $\Sigma_i(t)$ with a positive constant to generate a parallel shift type of movements of the IVS around the money, getting the shifted parametric price

$$C_M(t_0, S_0, P_t, R_t, D_t; K, t; \epsilon) = \frac{1}{norm} \sum_{i=1}^{n} \overline{a}_i(t) Call_{BS}(t_0, S_0, R_t, P_t, \overline{K}(K,t), t, \Sigma_i(t) + \epsilon)$$

where $\epsilon = 1\%$ to be consistent with the Black-Scholes vega. We can also define another type of volatility scenario dependent on the maturity with

$$(1+\epsilon)\Sigma_i^2(t) = (1+\epsilon)\frac{1}{t}\int_0^t \sigma^2(s)ds$$

where $\epsilon << 1$, so that the shifted option price becomes

$$C_M(t_0, S_0, P_t, R_t, D_t; K, t; \epsilon) = \frac{1}{norm}\sum_{i=1}^n \overline{a}_i(t) Call_{BS}(t_0, S_0, R_t, P_t, \overline{K}(K,t), t, \Sigma_i(t)(1+\epsilon))$$

Combining all these bumps together, we obtain a no-arbitrage deformation of the IV surface generated by the shifted option price

$$\begin{aligned}
& C_M(t_0, S_0, P_t, R_t, D_t; K, t; \epsilon_\mu, \epsilon_\beta, \epsilon_\sigma) \\
= \quad & \frac{1}{norm}\sum_{i=1}^n \overline{a}_i(t) Call_{BS}(t_0, S_0, R_t, P_t, K^{'}(K,t)(1+(1+\epsilon_\mu)\mu_i^0 f(t, \beta_i(1+\epsilon_\beta))), t, \Sigma_i(t)+\epsilon_\sigma)
\end{aligned}$$

For risk management purposes, we can therefore compute the modified vega as

$$\widehat{V}ega = C_M(t_0, S_0, P_t, R_t, D_t; K, t; \epsilon_\mu, \epsilon_\beta, \epsilon_\sigma) - C_M(t_0, S_0, P_t, R_t, D_t; K, t)$$

Chapter 7

Complex option strategies

We are going to describe a few complex strategies on single stock options and we will then discuss some well known strategies on multi-asset options. Multi-asset options is an example of contingent claims necessitating the analysis of sequences of random variables requiring two types of input:

1. the marginal distributions for the different variables, and
2. the dependencies between these variables.

A direct consequence of the Efficient Market Hypothesis (EMH) is that the most important concepts in theoretical and empirical finance developed around the assumption that asset returns follow a normal distribution. This includes the classical portfolio theory, the Black-Scholes-Merton option pricing model and the RiskMetrics variance-covariance approach to Value at Risk (VaR). One consequence of the capital asset pricing model (CAPM), BS-formula and Var is that statistical approaches must be restricted to second-oder stationary processes. Thus, the portfolio risk becomes a weighted sum of covariation of all stocks in the portfolio and one is left with modelling the variance and correlation of Gaussian random variables. However, financial time series exhibit multifractal scaling behaviour indicating a complex behaviour with long-range time correlations manifested on different intrinsic time scales. Hence, variance and covariance are not the proper measures of risk when analysing financial derivatives. These phenomena can easily be observed in the market via the existence of an implied volatility surface for single stock options and an implied correlation matrix for multi-asset options. Moreover, since the variance and covariance of financial time series are not necessarily defined, their historical dynamics are stochastic with erratic jumps. Thus, in order to account for these dynamics when pricing path-dependent options, practitioners and academics focused on modelling the variance and covariace with jump-diffusion processes.

7.1 The variance swap

7.1.1 Description of the variance swap

The variance swap is a tool for trading and managing volatility risk. It gives investors exposure to variance without directly exposing themselves to movements in the underlying since it does not require delta hedging offering them a strong advantage over the more traditional buying and selling of vanilla options and delta hedging. The payoff at maturity of the variance swap is the difference between the realised variance of the underlying stock price and a fixed strike multiplied by a notional. It can be seen as an interest rate swap where the floating leg correspond to the realised variance of the stock price. So, in essence the price of the variance swap is a forward contract on the realised annualised variance of the stock price.

The variance swaps have a large number of applications such as hedging, volatility trading and relative value strategies, portfolio diversification and structured products. They can be used in association with a view, to speculate on future levels of volatility, or for relative value trades including volatility pairs and dispersion trades. They can also be used as a way to diversify a portfolio when considered as an asset class. To conclude they can be used to structure products as they can be embedded into certificates, notes and other structured products.

7.1.2 Defining the variance swap contract

A variance swap is a forward contract on the realised annualised variance. Given the period $[0, T]$ with business days $0 = t_0 < ... < T_n = T$, its payoff at time T is

$$U\Big(A \times \Big[\frac{1}{n}\sum_{i=1}^{n}(\log\frac{S_i}{S_{i-1}})^2 - (\frac{1}{n}\log\frac{S_n}{S_0})^2\Big] - K_{var}\Big)$$

where $(S_i)_{i=1,..,n}$ is the stock price at time t_i, $i = 1, .., n$, U is the notional amount of the swap, A the annualisation factor and K_{var} is the strike price.
We assume that the stock price process $(S_t)_{t\geq 0}$ is a strictly positive semimartingale with stochastic instantaneous volatility $\sigma_S(t)$. Further, we assume that the stock price S_t is build from a jump-diffusion process, so that the positions and the amplitudes of its jumps are described by a jump random measure. From the definition of the quadratic variation of a jump-diffusion process (see Cont et al. [2003] for reference) we can deduce that when the number of sampling dates n tends to infinity the price of the variance swap converge to the quadratic variation of the logarithm of the stock price denoted by L_t. We now apply the definition of the quadratic variation to the Lévy process L_t with the random measure J_L [1] and approximate the so-called realised variance

$$\overline{V}_n = \frac{A}{n}\sum_{i=1}^{n}\big(\ln\frac{S_{t_i}}{S_{t_{i-1}}}\big)^2$$

by letting n tends to infinity, getting

$$\overline{V}_t = [L, L]_t = \int_0^t \sigma^2(\omega, s)ds + \sum_{s\in[0,t]}|\Delta L_s|^2 = \int_0^t \sigma^2(\omega, s)ds + \int_{[0,t]}\int_{\mathbb{R}} z^2 J_L(ds \times dz)$$

Proposition 11 *If we take the limit and let n tends to infinity, then the realised annualised variance $\overline{V}_n$ is approximated by the quadratic variation of the stock price, that is,*

$$\overline{V}_T = \frac{1}{T}[L, L]_T = \frac{1}{T}\Big(\int_0^T \sigma(\omega, s)^2 ds + \int_0^T d\widehat{J}_L(s)\Big)$$

where $\widehat{J}_L(t) = \sum_{s\in[0,t]}^{\Delta L_s \neq 0}|\Delta L_s|^2 = \sum_j^{N_t}|Z_L(j)|^2$ is a pure jump process with $E[\widehat{J}_L(t)] = Var[J_L(t)] = \int_{[0,t]}\lambda\int_{\mathbb{R}} z^2 F_L(dz)ds$.

From the absence of arbitrage opportunities, variance swap prices are martingales under the risk-neutral measure. Therefore, the forward contract is given by $E^Q[e^{-\int_t^T r_s ds}(\overline{V}_T - K_{var})|\mathcal{F}_t]$ and setting it to zero we obtain its fair delivery price. From now on we will consider the annualisation factor $\frac{1}{T}$ just as a weight which multiply an expectation called the variance swap, or variance bond. So, it leads to the definition of the variance swap price

Definition 7.1.1 *We let the price of the variance swap contract $(V_A(t, T))_{t\geq 0}$, seen at time t with maturity T, satisfies under the risk-neutral measure $\mathbb{Q}$*

$$V_A(t, T) = E^Q[e^{-\int_t^T r_s ds}\big(\int_0^T \sigma^2(\omega, s)ds + \int_0^T d\widehat{J}_L(s)\big)|\mathcal{F}_t]$$

[1] J_L is a pure jump process with intensity λ_L and whose distribution F_L has the transform θ_L associated to the logarithm of the stock price.

where the payoff of the contingent claim is the quadratic variation of a Lévy process.

7.1.3 The logarithm contract

We let $FV_A(t,T) = \partial_T V_A(t,T)$ be the forward variance price, given as

$$FV_A(t,T) = -r_T V_A(t,T) + E^Q[e^{-\int_t^T r_s ds}|\sigma(\omega,T)|^2|\mathcal{F}_t]$$

In a pure diffusion case, Carr et al. [1998b] linked the price of the variance swap to the price of the logarithm contract. We can integrate the general diffusion equation for the stock price and express the logarithm contract as

$$\log \frac{S_T}{S_{t_0}} = \int_0^T \frac{dS_t}{S_t} - \frac{1}{2}\int_0^T |\sigma(t,\omega)|^2 dt$$

We denote $\overline{W}_T$ the total variance over the period $[0,T]$ and get

$$\overline{W}_T = \int_0^T |\sigma(t,\omega)|^2 dt = 2\big(\int_0^T \frac{dS_t}{S_t} - \log \frac{S_T}{S_{t_0}}\big)$$

The first term in the brackets is the net outcome of continuously rebalancing a stock position so that it is always instantaneously long $\frac{1}{S_t}$ shares of stock worth 1. So the strategy of holding a portfolio of shares constantly equal to one unit of currency defines the payoff, that is the cost or profit of the strategy. The second term in the brackets is a static short position in a log-contract which can be replicated with a weighted combination of European options. Therefore, following this strategy we can replicate the total variance $\overline{W}_T$ in a model-independent way as long as the stock price is a diffusion. Taking the expectation of the total variance and using the Breeden and Litzenberger [1978] results, we get

$$\begin{aligned} E_{t_0}[\int_0^T |\sigma(t,\omega)|^2 dt] &= 2E_{t_0}[\int_0^T \frac{dS_t}{S_t}] - 2E_{t_0}[\log \frac{S_T}{S_{t_0}}] \\ &= 2\big(\int_0^T r_t dt - (P(0,T)^{-1} - 1) + \int_0^{S_{t_0}} \frac{1}{K^2} P(T,K)dK + \int_{S_{t_0}}^\infty \frac{1}{K^2} C(T,K)dK\big) \end{aligned}$$

where $C(T,\cdot)$ and $P(T,\cdot)$ are call and put prices. In the case where we center the expansion on the forward price (with deterministic rates) we get

$$\log \frac{S_T}{F_T(0)} = \int_0^T \frac{dS_t}{S_t} - \frac{1}{2}\int_0^T |\sigma(t,\omega)|^2 dt + \log P(0,T)$$

and the total realised variance over the period $[0,T]$ becomes

$$\overline{W}_T = \int_0^T |\sigma(t,\omega)|^2 dt = 2\big(\int_0^T \frac{dS_t}{S_t} - \log \frac{S_T}{F_T(0)} + \log \frac{S_{t_0}}{F_T(0)}\big)$$

Taking the expectation of the total variance we get

$$E_{t_0}[\int_0^T |\sigma(t,\omega)|^2 dt] = -2E_{t_0}[\log \frac{S_T}{F_T(0)}] = 2\big(\int_0^{F_T(0)} \frac{1}{K^2} \overline{P}(T,K)dK + \int_{F_T(0)}^\infty \frac{1}{K^2} \overline{C}(T,K)dK\big) \quad (7.1.1)$$

where $\overline{C}(T,\cdot)$ and $\overline{P}(T,\cdot)$ are undiscounted call and put prices. From the properties of the law of iterated expectation for random variables [2] we can recover the variance swap from the local volatility of the underlying price. That is, we have the approximation in space

[2] $E[X] = E[E[X|Y]]$

$$\sum_{K=Kmin}^{Kmax} E_{t_0}[|\sigma(T,\omega)|^2|S_T = K]P(S_T = K) \approx E_{t_0}[|\sigma(T,\omega)|^2]$$

So that the weighted sum in time and in space of the local variance of a diffusion process with deterministic rates gives back the variance swap.

***Remark* 7.1.1** *Since the instantaneous volatility $\sigma(t,\omega)$ is not a function of time T, differentiating the variance swap (or the expected log-contract) with respect to maturity T and conditioning the resulting expected value with $S_T = K$ we can express the forward variance swap in terms of the local variance.*

7.1.4 The limits of the logarithm contract

We now ask ourselves if the logarithm contract is a good estimator of the quadratic variation of the logarithm of the stock price in presence of jumps. So far we have assumed that the stock price was a semimartingale build from a Lévy process. We are now going to explicit that Lévy process which is the logarithm of the stock price.
As an example, we let the logarithm of the stock price L_t satisfies

$$L_t = L_0 + \int_0^t b_s ds - \lambda \int_0^t (\theta(1) - 1)ds - \frac{1}{2}\int_0^t \sigma_s^2 ds + \int_0^t \sigma_s dW(s) + \sum_{i=1}^{N_t} \Delta L_i$$

where the Laplace transform of the jump size is $\theta(c) = E[e^{c.z}] = \int_{\mathbb{R}} e^{c.z} F_L(dz)$ and which is such that the discounted stock price is a martingale. We then compute the variation of the process $S_t = e^{L_t}$, corresponding to the exponential of the Lévy process defined above. Applying Ito's lemma, we get its variation as

$$S_t = S_0 + \int_0^t S_s b_s ds - \lambda \int_0^t S_s(\theta(1) - 1)ds + \int_0^t \sigma_s S_s dW(s) + \sum_{0 \leq s \leq t, \Delta L_s \neq 0} \big(e^{L_{s_-} + \Delta L_s} - e^{L_{s_-}}\big)$$

Using the theory on Lévy processes, we can express the relationship between the dynamic of the process S_t and the dynamic of its logarithm L_t, that is, transform the exponential of a Lévy process into a stochastic Doleans-Dade exponential. Integrating that equation between $[0,T]$, we get

$$2\Big(\int_{[0,T]} \frac{dS_t}{S_{t_-}} - L_T + L_0\Big) = \int_0^T \sigma_s^2 ds + 2\int_{[0,T]}\int_{\mathbb{R}} \big(e^z - 1 - z\big) J_L(ds \times dz)$$

Taking the expectation of the price of the variance swap expressed in terms of the price of the logarithm contract, we get

$$2\big(E[\int_{[0,T]} \frac{dS_t}{S_{t_-}}] - E[L_T] + L_0\big) = E[\int_0^T |\sigma_s|^2 ds] + 2\lambda \int_{[0,T]}\int_{\mathbb{R}} \big(e^z - 1 - z\big) F_L(dz) ds$$

We know from Proposition (11) that the fair price of the variance swap is

$$E\big[[L,L]_t\big] = E[\int_0^t |\sigma_s|^2 ds] + \lambda \int_{[0,t]}\int_{\mathbb{R}} z^2 F_L(dz) ds$$

We see that the logarithm contract is a bias estimator of the variance swap as the right order of magnitude of the jump term is $\mathcal{O}(z^2)$ when computing the quadratic variation of the logarithm of the stock price, while when using the logarithm contract it is of the order $\mathcal{O}(z)$. The error term at maturity T is calculated by taking the difference between the two approaches, getting

$$E\big[[L,L]_T\big] - E[< U >_T] = 2\lambda \int_{[0,T]} \int_{\mathbb{R}} \big(\frac{1}{2}z^2 - e^z + 1 + z\big) F_L(dz)ds$$

Thus, replicating the price of the variance swap with a continuum of call and put prices does not approximate correctly the quadratic variation of the logarithm of the stock price in presence of jumps. It incurs an error term which is model dependent (depending on the distribution of the jump size) and varies depending on the maturity considered. We therefore need a model (incorporating stochastic volatility and jumps) to correctly value the price of the variance swap for all time t.

7.2 Approximating Asian option

7.2.1 The payoff

We consider the payoff of an Asan option as

$$\big(\frac{1}{\theta}\int_{T-\theta}^{T} S(u)du - K\big)^+$$

where

$$dS(t) = S(t)\big[rdt + \sigma dW_t\big]$$

The price of an Asian call option is given by

$$C_A(t) = e^{-r(T-t)} E[\big(\frac{1}{\theta}\int_{T-\theta}^{T} S(u)du - K\big)^+] \tag{7.2.2}$$

7.2.2 A first approximation

We first provide a quasi-analytic approximation in the Black-Scholes world. It gives a tight bound to Monte-Carlo simulation. This is a good approximation if we assume that we are in the presence of a sum of Log-Normal distribution with constant volatilities. However, it is possible that some of the stocks might be Pareto distributed or Levy distributed and even incorporate stochastic volatilities (see Section (2.1.2)). In which case a price in the Black-Scholes world is always lower than the price in a stochastic world.
We let the stochastic differential equation (SDE) of the underlying asset at time t_i, under the risk-neutral measure $\mathbb{Q}$, be given by

$$dS(t_i) = S(t_i)\big[r_t dt + \sigma(t_i) dW_{t_i}\big]$$

Integrating in the range $[0, t_i]$ to get the stock price process, we get

$$S(t_i) = S(0) e^{\int_0^{t_i} r_s - d_s - \frac{1}{2}\sigma^2(s)ds + \int_0^{t_i} \sigma(s)dW_s}$$

In the special case where all parameters are constants, the stock price simplifies to

$$S(t_i) = S(0) e^{\left(r - d - \frac{1}{2}\sigma^2\right)t_i + \sigma(t_i)W_{t_i}}$$

We are interested in the law of a weighted sum of Geometric Brownian Motion

$$Y = \sum_{i=1}^{n} \alpha_i S(t_i)$$

which can be written as

$$Y = \sum_{i=1}^{n} \mu_i e^{\sigma(t_i)W(t_i)}$$

with $\mu_i = \alpha_i e^{\left(r-d-\frac{1}{2}\sigma^2\right)t_i}$. This is a Linear Combination of n state variables, which can be approximated by a weighted sum of only one state variable. We consider the process

$$Z = \sum_{i=1}^{n} \alpha_i W(t_i)$$

Assuming the call option payoff to be convex, the call option is convex and we can apply Jensen's inequality, getting

$$\begin{aligned} E[Y^+] &= E[E(Y^+|Z)] \\ &\geq E[E(Y|Z)^+] \end{aligned}$$

Therefore, the law of Y is approximated by the law of $Y|Z$, where we consider the process $U_i = W_{t_i}$ normalised, to give the state variable

$$U \simeq N(0,1)$$

This can be written as

$$Y|Z = \sum_{i=1}^{n} \tilde{\mu}_i e^{\gamma(t_i)U} \tag{7.2.3}$$

or as a weighted sum of one state variable with probabilities P_i

$$Y|Z = \sum_{i=1}^{n} P_i e^{q_i U - \frac{1}{2}q_i^2}$$

where

$$q_i = \rho\sigma(t_i)$$

There is no more process, and we are only left with the manipulation of the terminal distribution. We are not concerned by a possible evolution of a given process. We have mapped the weighted sum of n Geometric Brownian Motions into a one dimensional world where the volatilities are multiplied by their respective correlation. Note that as Z gets closer to Y, its distribution tends to the Dirac distribution. We are therefore trying to reduce $Y|Z$ to a deterministic function. We have actually obtained a variance reduction, since

$$var[Y] > var[Y|Z^*]$$

This is a good approximation as long as the law of $Y|Z^*$ is close to the law of Y, that is Log-Normal. However, there is a problem if the law of $Y|Z^*$ is far away from the true law of Y. One way to overtake the distribution problem would be to incorporate the smile. However, it is not clear that the Greeks will be correctly calculated.

7.2.3 Upper bound using Black-Scholes with stochastic volatility

Assuming stochastic volatility, we are now concerned in finding out, how the Black-Scholes model behave in the presence of stochastic volatilities when pricing Asian options. To do so, we need to readjust the Black-Scholes model to be sure to be on an upper bound, whatever the stochastic volatility we experience. Given the option price in Equation (7.2.2), we follow El Karoui et al. [1998] and consider a new variable replicating $\frac{1}{\theta}\int_{T-\theta}^{T} S(u)du$ at maturity. Let's look at the process

$$X(t) = e^{-r(T-t)} E[\frac{1}{\theta}\int_{T-\theta}^{T} S(u)du]$$

Applying Fubini's theorem and interchanging the expectation with the integral operator, we get

$$X(t) = \frac{e^{-r(T-t)}}{\theta}\int_{T-\theta}^{T} E[S(u)du]$$

Recall that for $t \geq u$, we get

$$E[S(u)|F_t] = S(u)$$

so that the process becomes

$$X(t) = \frac{e^{-r(T-t)}}{\theta}\int_{T-\theta}^{t\vee(T-\theta)} S(u)du$$

where $\vee$ stands for max. Now, for $t \leq u$

$$E[S(u)|F_t] = e^{r(u-t)}S(t)$$

so that the process becomes

$$X(t) = \rho(t)S(t)$$

where

$$\rho(t) = \frac{1 - e^{-rT + r(t\vee(T-\theta))}}{r\theta}$$

Adding the two terms together, we get

$$X(t) = \rho(t)S(t) + \frac{e^{-r(T-t)}}{\theta}\int_{T-\theta}^{t\vee(T-\theta)} S(u)du \tag{7.2.4}$$

The true forward is given by setting $t = 0$

$$X(0) = \rho(0)S(0)$$

where

$$\rho(0) = \frac{1 - e^{-rT}}{r\theta}$$

Now, we can write the process as

$$X(t) = \rho(t)S(t) + \frac{e^{-r(T-t)}}{\theta}\int_{T-\theta}^{t\vee(T-\theta)} S(u)du$$

which satisfies the SDE

$$dX(t) = rX(t)dt + \sigma\rho(t)S(t)dW_t$$

It is now clear that X(t) is another underlying asset whose price at time T is $\int_{T-\theta}^{T} S(u)du$. The SDE can be rewritten as

$$dX(t) = X(t)\big[rdt + \frac{\sigma\rho(t)S(t)}{X(t)}dW_t\big] \tag{7.2.5}$$

where the true volatility of $X(t)$ is given by

$$\frac{\sigma\rho(t)S(t)}{X(t)} \tag{7.2.6}$$

Since we have the relation

$$\frac{\rho(t)S(t)}{X(t)} < 1 \tag{7.2.7}$$

it can be seen that the true volatility is dominated by the misspecified volatility σ. We now consider Theorem (4.2.1) and write

- the true volatility as $\sigma(t) = \frac{\sigma\rho(t)S(t)}{X(t)}$
- the misspecified volatility as $\gamma(t, S(t)) = \sigma$

Then, the theorem implies that

$$\begin{aligned} C_A(0) &= E[e^{-rT}\big(X(T) - K\big)^+] \\ &\leq BS(T, X(0); r, \sigma) \end{aligned}$$

Therefore, the price of an asian option is dominated by a Black-Scholes price. However, this Black-Scholes price is not a tight upper bound. This is because among all possible choice of the misspecified volatility $\gamma(t, S(t))$, we chose the average volatility to be a constant. Remember that our underlying replicate $\int_{T-\theta}^{T} S(u)du$ at maturity, and that at maturity, it is a well known event with volatility equal to zero. Thus, our Black-Scholes price over estimate the true price as time tends to maturity. To overtake this problem, we need to take a well chosen time dependent misspecified volatility, with the property of being zero at maturity. In that case, we obtain a good approximation of the stochastic volatility by a deterministic one. In terms of hedging, the pair process

- $P(t) = BS(T - t, X(t); r, \sigma)$
- $\Delta(t) = \rho(t)\frac{\partial}{\partial x}BS(T - t, X(t); r, \sigma)$

is a superstrategy.

7.2.4 Model comparison

In the Black-Scholes world, the sum of N Geometric Brownian Motions is mapped into a single weighted Log-Normal distribution

$$Y|Z = \sum_{i=1}^{n} P_i e^{q_i U - \frac{1}{2}q_i^2}$$

In the case of the Black-Scholes with stochastic volatility, we take the Log-Normal distribution of the true forward with the dominated misspecified volatility

$$Y = X(0)e^{\sigma W_T - \frac{1}{2}\sigma^2}$$

We see that the problem is to choose the best possible misspecified time dependent volatility which always satisfies

$$\sigma(t) \leq \gamma(t, S(t))$$

Knowing the true diffusion and the true volatility, we can study them in some special cases, such as studing their limit behaviours, and therefore we are in a better position to define the optimum $\gamma(t, S(t))$.

7.3 Introducing the pricing of multi-asset options

In order to be exposed to more than one underlying and to diversify risk, financial markets considered the use of portfolio as weighted sum of assets. Then options on the performance of multiple assets developed, providing a hedge to positions on multiple assets, or serving as speculative tools on multiple assets. In general, multi-asset options are cheaper than single stock options, such that institutional, retail and private banking buy the correlation, while investment banks sell it. The correlation derivatives include spread options, worst-of options, index dispersion trades, basket dispersion tradres, correlation swaps etc.

7.3.1 Pricing European options

For simplicity of exposition we briefly describe the pricing of a European multi-asset option with two underlyings $S_{1,t}^{x_1}$ and $S_{2,t}^{x_2}$. Given the payoff $h(S_{1,T}, S_{2,T})$ at maturity T, the price seen at time t under the risk-neutral measure $\mathbb{Q}$ is

$$BC(t,T) = P(t,T)E^Q[h(S_{1,T}, S_{2,T})|\mathcal{F}_t]$$

The price can also be expressed as a double integral as follow

$$BC(t,T) = P(t,T)\int_0^\infty \int_0^\infty h(x_1, x_2)\phi_B(x_1, x_2)dx_1 dx_2 \quad (7.3.8)$$

where $\phi_B(\cdot, \cdot)$ is the joint density of $S_{1,t}$ and $S_{2,t}$ under the risk-neutral measure. However, in the case where we can directly buy and sell options on a traded index I_t^x, like stock index options (SIO), we can directly price these options as

$$C_I(t,T) = P(t,T)E^Q[h(I_T)|\mathcal{F}_t]$$

which gives

$$C_I(t,T) = P(t,T)\int_0^\infty h(x)\phi_I(x)dx$$

where $\phi_I(\cdot)$ is the density of the index I_t under the risk-neutral measure.

7.3.2 Introducing implied correlation

We saw in Section (1.4.1.2) that the implied volatility was observed after the market crash of 1987 and that it is computed by inverting the Black-Scholes formula $C_{BS}^{-1}(C(t, S_t, K, T); K, T)$. As a result, the market referred to the two-dimensional map

$$(K,T) \to \Sigma(K,T)$$

as the implied volatility surface (IVS). Rebonato once stated that implied volatility was the wrong number to put in the wrong formula to get the right price. Nonetheless, it contains additional information about expected future volatility not included in the historical volatility. After 1981, a market for stock index options (SIO) developed to manage and hedge portfolio exposure, where the Black-Scoles formula was used to price these complex options. Similarly to single stock options, after the market crash of 1987, we observed an implied volatility surface when pricing stock index options. Further, combining the constituent single stock options with a Gaussian copula to price an index option would not fully explain the volatility skew of that index option, leading to the notion of implied correlation (IC). Thus, the market attempted at inferring an implied correlation from option prices on multiple assets. It would require liquid markets for such products in the multi-asset case and the existence of a closed-form solution to the multivariate lognormal model, equivalent to the BS-formula in single asset. However, neither of them exist. It did not stop practitioners from defining different implied correlation indicators based on the different derivative products considered. For example, they defined an implied correlation for spread options, one for worst-of options, one for index dispersion, and one for correlation swaps. Obviously, these implied correlations are not consistent with one another. Moreover, they observed a non-flat implied correlation surface, indicating that returns are non-Gaussian and most likely that their increments are not independent. By analogy with the implied volatility, the implied correlation can be viewed as the wrong number to put in the wrong pricer given a wrong volatility model to get the right price (see Delanoe [2014]). Rather than modifying the Browninan motions, they chose to consider stochastic and local correlation models to explain the correlation skew.

7.3.3 The multi-underlying processes

We assume a no-arbitrage world where all asset prices follow Ito's processes. We consider the Markov process $(X_t)_{t\geq 0}$ taking values in $\mathbb{R}$ to be the logarithm of the underlying price of a contingent claim with dynamics under the risk-neutral measure $\mathbb{Q}$ being

$$dX_t = \mu_t dt + \sigma dW(t)$$

For any two processes $X_{1,t}$ and $X_{2,t}$, we get $< dW_1, dW_2 >_t = \rho_{1,2} dt$ where $\rho_{1,2}$ is the instantaneous correlation between the two assets. Assuming that the correlation matrix is symmetric and positive definite, we use Cholesky decomposition to project the correlated Brownian motions W into a basis of independent orthogonal Brownian motions B. In that setting, the dynamics of the one-dimensional Markov process become

$$dX_t = \mu_t dt + \sigma_t dB(t) = \mu_t dt + < \sigma_t, dB(t) > \tag{7.3.9}$$

where $\mu_t = r_t - \frac{1}{2}\sigma_t \sigma_t^\top - q_t$. The Brownian motion $B(t) = \big(B_1(t), ..., B_n(t)\big)^\top$ is a column vector of dimension $(n,1)$ and σ_t is a matrix of dimension $(1,n)$ with element $\sigma_i(t)$. We assume that σ_t, r_t and q_t are second-order stochastic processes adapted to the filtration generated by B. That is, we have assumed a structure of correlation modelled with a finite number of independent Brownian motions and a vector of volatility σ_t. When all the elements $\sigma_i(t)$ are constants, the model reduces to the classic multivariate lognormal model (generalisation of Samuelson lognormal model described in Section (1.2.2)). Hence, we get a family of random functions of volatility, such that $\sigma_t dB(t) = \sum_{j=1}^n \sigma_j(t) dB_j(t)$, and the dynamics of the process rewrite as

$$dX_t = \mu_t dt + \sum_{j=1}^n \sigma_j dB_j(t)$$

Using the properties of the stochastic integral and the Brownian motion, we can re-express the above equation in terms of the volatility $vol_S(t)$ and a unidimensional Brownian motion $Z(t)$ as

$$< \sigma_t, dB(t) > \; = vol_S(t) dZ(t)$$

so that the link between the vector of local volatilities and the underlying volatility is

$$V(t) = vol_S^2(t) = \sum_{j=1}^n \sigma_j^2(t) = ||\sigma_t||^2 = \sigma_t \sigma_t^\top$$

with average variance in the range $[t, T]$ given by

$$I^2(t,T) = \frac{1}{T-t} \int_t^T V(s) ds$$

and total variance given by

$$\tilde{I}(t,T) = (T-t) I^2(t,T)$$

The stochastic differential equation (SDE) of the process becomes

$$dX_t = \mu_t dt + \sqrt{V(t)} dZ(t)$$

with

$$dZ(t) = \frac{\sum_{j=1}^n \sigma_j dB_j(t)}{\sqrt{\sum_{j=1}^n \sigma_j^2(t)}}$$

Using vectorial notation, the instantaneous volatility is

$$vol_S(t) = ||\sigma_t||$$

and the instantaneous covariance between different underlyings $X_{1,t}$ and $X_{2,t}$ is

$$Cov(dX_{1,t}, dX_{2,t}) =< \sigma_t^1, \sigma_t^2 > dt = \sigma_t^1 (\sigma_t^2)^\top dt$$

since $X_{i,t} = \log S_{i,t}$ for $i = 1, .., n$. Hence, in the case where we have not yet decorrelated the Brownian motions, the covariance between two stocks $S_{1,t}$ and $S_{2,t}$ is

$$Cov(\frac{dS_{1,t}}{S_{1,t}}, \frac{dS_{2,t}}{S_{2,t}}) =< \sigma_t^1, \sigma_t^2 > dt = \rho_{12}(t) vol_{S_1}(t) vol_{S_2}(t) dt$$

and the instantaneous correlation is given by

$$\rho_{12}(t) = \frac{< \sigma_t^1, \sigma_t^2 >}{||\sigma_t^1|| ||\sigma_t^2||}$$

Recall, the volatility of the product of two assets is given by

$$vol_{S_1 S_2}(t) = ||\sigma_t^1 + \sigma_t^2|| = ||\sigma_t^1||^2 + ||\sigma_t^2||^2 + 2\rho_{12}(t) ||\sigma_t^1|| ||\sigma_t^2||$$

7.3.4 The basket in the BS-world

We derive the stochastic differential equation of a basket and compute its instantaneous variance. For simplicity of exposition, we assume a multivariate lognormal model and note that the analysis can apply to instantaneous volatility being deterministic function of time and the spot price.

7.3.4.1 Type 1

In a multidimensional Black-Scholes model with N stocks $S_1, .., S_N$, the risk-neutral dynamics are given by

$$\frac{dS_i(t)}{S_i(t)} = r(t)dt + \sum_{j=1}^{n} \sigma_{ij}(t)dB_j(t) \, , \, i = 1, ..., N$$

where B_j for $j = 1, .., n$ are independent standard Brownian motions. Further, the interest rate and the volatilities $\sigma_{ij}(t)$ can be deterministic functions of time. For simplicity, we denote $\sigma_{ij} = \int_0^t \sigma_{ij}(s)ds$ the time dependent volatility function. Correlations among different stocks are captured through the matrix (σ_{ij}). The solutions to the above SDE are

$$S_i(t) = S_i(0)e^{(r - \frac{1}{2}\sum_{j=1}^{n}(\sigma_{ij})^2)t + \sum_{j=1}^{n} \sigma_{ij} B_j(t)} \, , \, i = 1, ..., N$$

In a general framework, we let X be the random variable

$$X = \sum_{i=0}^{N} \epsilon_i x_i e^{G_i \sqrt{t} - \frac{1}{2} Var(G_i) t}$$

where $(G_i)_{i=0,...,N}$ is a mean zero Gaussian vector of size $(N+1)$ and covariance matrix Σ, $\epsilon_i = \pm 1$, and $x_i > 0$ for all $i = 0, ..., N$. Note, Σ is symmetric positive semi-definite but not necessarily definite.
The risk-neutral dynamics of the forward price $F_i(t,T)$ are given by

$$\frac{dF_i(t,T)}{F_i(t,T)} = \sum_{j=1}^{n} \sigma_{ij}(t)dB_j(t) \, , \, i = 1, ..., N$$

Then, given the forward basket price $B(t,T) = \sum_{i=1}^{N} \omega_i F_i(t,T)$, with weights $\omega_i \geq 0$, its dynamics satisfy the SDE

$$dB(t,T) = \sum_{i=1}^{N} \omega_i F_i(t,T) \sum_{j=1}^{n} \sigma_{ij}(t)dB_j(t)$$

which we can rewrite as

$$\frac{dB(t,T)}{B(t,T)} = \sum_{i=1}^{N} \widehat{\omega}_i(t) \sum_{j=1}^{n} \sigma_{ij}(t)dB_j(t) = \sum_{j=1}^{n} \Big(\sum_{i=1}^{N} \widehat{\omega}_i(t)\sigma_{ij}(t) \Big) dB_j(t)$$

where

$$\widehat{\omega}_i(t) = \frac{\omega_i F_i(t,T)}{\sum_{i=1}^{N} \omega_i F_i(t,T)} = \frac{\omega_i F_i(t,T)}{\omega^\top F(t,T)} \, , \, i = 1, ..., N \quad (7.3.10)$$

with $0 \leq \widehat{\omega}_i(t) \leq 1$ and $\sum_{i=1}^{N} \widehat{\omega}_i(t) = 1$.
We can define the matrix of volatility of the basket $\Gamma_B(t,T)$ of size $(1,n)$ as

$$\Gamma_B(t,T) = [\Gamma_{B,1}(t,T), ..., \Gamma_{B,n}(t,T)]$$

with jth element being a linear combination of stock's volatility

$$\Gamma_{B,j}(t,T) = \sum_{i=1}^{N} \hat{\omega}_i(t)\sigma_{ij}(t) \, , \, j = 1, ..., n$$

and the dynamics of the basket become

$$\frac{dB(t,T)}{B(t,T)} = \Gamma_B(t,T)dZ(t)$$

with solution

$$B(t,T) = B(0,T)e^{-\frac{1}{2}\int_0^t V_B(s)ds + \int_0^t \sum_{j=1}^n \Gamma_{B,j}(t,T)dZ_j(s)}$$

where the instantaneous variance of the basket is given by

$$V_B(t) = ||\Gamma_B(t,T)||^2 = \sum_{j=1}^n \Gamma_{B,j}^2(t,T) = \sum_{j=1}^n \Big(\sum_{i=1}^N \widehat{\omega}_i(t)\sigma_{ij}(t)\Big)^2$$

and the basket average future variance is given by

$$I_B^2(t,T) = \frac{1}{T-t}\int_t^T V_B(s)ds$$

7.3.4.2 Type 2

Alternatively, following the notation in Section (7.3.3), we can rewrite the dynamics of the underlying forward price as follow

$$\frac{dF_i(t,T)}{F_i(t,T)} = \sigma_t^i dZ(t)$$

where σ_t^i is a matrix of dimension $(1,n)$ such that $\sigma_t^i dZ(t) = \sum_{j=1}^n \sigma_{ij}(t)dZ_j(t)$ and the stock's instantaneous variance is given by $V_i(t) = ||\sigma_t^i||^2$. Then, the dynamics to the forward basket price $B(t,T)$ satisfy

$$dB(t,T) = \sum_{i=1}^N \omega_i F_i(t,T)\sigma_t^i dZ(t)$$

which we can rewrite as

$$\frac{dB(t,T)}{B(t,T)} = \sum_{i=1}^N \widehat{\omega}_i(t)\sigma_t^i dZ(t) = \big(\sum_{i=1}^N \widehat{\omega}_i(t)\sigma_t^i\big)dZ(t)$$

By analogy to the basket, $\Gamma_{B,j}(t,T) = \sum_{i=1}^N \widehat{\omega}_i\sigma_{ij}(t)$ is the jth element of $\big(\sum_{i=1}^N \widehat{\omega}_i(t)\sigma_t^i\big)$ since the sum of vectors is a vector. We can construct the variance-covariance matrix as follow

$$\sigma\sigma^\top = \big(\sum_{i=1}^N \widehat{\omega}_i(t)\sigma_t^i\big)\big(\sum_{i=1}^N \widehat{\omega}_i(t)\sigma_t^i\big)^\top$$

Setting $\Omega_t = \widehat{\omega}_t\widehat{\omega}_t^\top$, where the modified weight $\widehat{\omega}_t = \big(\widehat{\omega}_1(t), ..., \widehat{\omega}_N(t)\big)^\top$ is a column vector of dimension $(N,1)$, we can compute the trace of $\Omega_t X_u$ as follow

$$\begin{aligned} Tr(\Omega_t X_u) &= \sum_{i,j=1}^N \Omega_{i,j}(t)X_{i,j}(t) \\ &= \sum_{i,j=1}^N \widehat{\omega}_i(t)\widehat{\omega}_j(t)\sigma_u^i\sigma_u^j \end{aligned}$$

We can also set

$$\tilde{\sigma}_t^i = \sigma_t^i - \sigma_t^B \text{ with } \sigma_t^B = \sum_{i=1}^N \widehat{\omega}_i(t)\sigma_t^i$$

and note that σ_t^B is $\mathcal{F}_t$-measurable. Hence, we can rewrite the dynamics of the basket as

$$\frac{dB(t,T)}{B(t,T)} = \sigma_t^B dZ(t)$$

where $\sigma_t^B dZ(t) = \sum_{i=1}^N \widehat{\omega}_i(t) \sum_{j=1}^n \sigma_{ij}(t) dB_j(t)$ and with instantaneous variance given by

$$V_B(t) = \sigma_t^B(\sigma_t^B)^\top = ||\sigma_t^B||^2$$

From the properties of independent Brownian motions in a norm (see Section (7.3.3)), we get

$$V_B(t) = ||\sum_{i=1}^N \widehat{\omega}_i(t)\sigma_t^i||^2 = \sum_{i=1}^N ||\widehat{\omega}_i(t)\sigma_t^i||^2$$

7.3.4.3 Type 3

An alternative approach is to consider the dynamics of the basket as

$$\frac{dB(t,T)}{B(t,T)} = \sum_{i=1}^N \widehat{\omega}_i(t) \frac{dF_i(t,T)}{F_i(t,T)} \tag{7.3.11}$$

where the instantaneous return of the basket is a weighted sum of the instantaneous returns of single stocks with stochastic weights. We can compute the instantaneous variance of the basket as

$$V_B(t) = Var(\frac{dB(t,T)}{B(t,T)}) = Var(\sum_{i=1}^N \widehat{\omega}_i(t) \frac{dF_i(t,T)}{F_i(t,T)})$$

From independence of the Brownian motions, the instantaneous variance simplifies to

$$V_B(t) = \sum_{i=1}^N Var(\widehat{\omega}_i(t) \frac{dF_i(t,T)}{F_i(t,T)}) = \sum_{i=1}^N ||\widehat{\omega}_i(t)\sigma_t^i||^2$$

In the case where we have not projected the Brownian motions W in a basis of independent Brownian motions B, the instantaneous variance of the basket becomes

$$V_B(t) = \sum_{i,j=1}^N \widehat{\omega}_i(t)\widehat{\omega}_j(t) Cov(\frac{dS_i(t)}{S_i(t)}, \frac{dS_j(t)}{S_j(t)})$$

Since the Brownian motions are correlated, each stock has a single Brownian motion instead of a vector of Brownian motions, and the instantaneous variance of the basket becomes

$$V_B(t) = \sum_{i,j=1}^N \widehat{\omega}_i(t)\widehat{\omega}_j(t) < \sigma_i(t), \sigma_j(t) > dt = \sum_{i,j=1}^N \widehat{\omega}_i(t)\widehat{\omega}_j(t)\rho_{ij}(t)\sigma_i(t)\sigma_j(t) dt \tag{7.3.12}$$

From the definition of the stochastic weight $\widehat{\omega}_i(t)$ in Equation (7.3.10) we can rewrite the instantaneous variance of the basket as

$$V_B(t) = \frac{1}{B^2(t,T)} \sum_{i,j=1}^{N} \omega_i \omega_j \rho_{ij}(t) F_i(t,T) F_j(t,T) \sigma_i(t) \sigma_j(t) dt \tag{7.3.13}$$

Since the weights ω_i are positive, the maximal covariance at time t is achieved when $\rho_{ij}(t) = 1$ for all i and all j. Thus, we must have

$$V_B(t) \leq \sum_{i,j=1}^{N} \widehat{\omega}_i(t) \widehat{\omega}_j(t) \sigma_i(t) \sigma_j(t) dt \tag{7.3.14}$$

Remark 7.3.1 *The relation in Equation (7.3.14) between the instantaneous variance of a basket and that of its constituents derives from stochastic calculus and can not be used to relate basket options to individual options.*

7.4 The spread option

7.4.1 The payoff

Denoting $h(S_{1,T}, S_{2,T})$ the payoff at maturity of a multi-asset option, its price at time t is

$$C(t, x_1, x_2, T) = P(t,T) E^Q[h(S_{1,T}, S_{2,T}) | S_{1,t} = x_1, S_{2,t} = x_2]$$

The payoff of a spread option satisfies

$$\left(S_{1,T} - \alpha S_{2,T} - K \right)^+$$

where α is a ratio and K is the strike. The dynamics of the processes $S_{i,t}$, $i = 1, 2$, are given in Equation (7.3.9). In the special case where $K = 0$ and the model is a multivariate lognormal (ML) model, Margrabe [1978] obtained a closed-form solution. Otherwise, for $K \neq 0$ there is no closed-form solution and one needs to rely on approximations and numerical tools. For K close to zero, Kirk [1995] provided a good approximation to spread options. Bjerksund et al. [2006] extended Kirk's idea for larger strikes K, obtaining an accurate closed-form solutions. Carmona et al. [2003] proposed bounds to spread options very close to the true prices.

7.4.2 The implied correlation

While the multivariate lognormal model assume constant correlations between Brownian motions, the empirical correlations vary over time, often experiencing peaks during financial crises (see Moskowitz [2003], Ahdida et al. [2012]). Assuming a particular approximation method (Kirk's formula) for the computation of spread options, the market defined the implied correlation (IC) as the level of correlation to input in that formula in order to recover market prices. It is defined as a function of the ratio α and the strike K, providing a two-dimensional surface. When the surface is not flat, practitioners call it the implied correlation skew (ICS), indicating that tail distributions are not properly captured in the multivariate lognormal framework. We formally define the implied correlation as follow:

Definition 7.4.1 *We define the implied correlation of a spread option as the level of correlation we need to plug into a basic pricer to recover the market prices of the option.*

Hence, given the option price $C(t, S_{1,t}, S_{2,t}, K, T)$ at time t for a strike K and a maturity T, the market implied correlation $\rho_S(t, S_{1,t}, S_{2,t}; K, T)$ satisfies

$$C(t, S_{1,t}, S_{2,t}, K, T) = C_{bp}(t, S_{1,t}, S_{2,t}, K, T; \Sigma_1(K,T), \Sigma_2(K,T); \rho_S) \tag{7.4.15}$$

where $C_{bp}(t, S_{1,t}, S_{2,t}, K, T; \sigma_1, \sigma_2; \rho)$ is the basic pricer of a spread option under the multivariate lognormal (ML) model. When $K = 0$ we use Magrabe's formula, and the price C_{bp} is monotonically decreasing in the correlation ρ_S. Otherwise, for $K \neq 0$, the monotonicity depends on the choice of the approximation pricer. Kirk's formula, for K close to zero, preserves that property. The implied correlation is obtained by inverting the basic pricer formula $C_{bp}^{-1}(C(t, S_{1,t}, S_{2,t}, K, T); K, T)$. Consequently, for a fixed maturity T, we refer to the two-dimensional map

$$(\alpha, K) \to \rho_S(\alpha, K)$$

as the implied correlation surface. We can therefore define the 3d-volume of implied correlation as

$$(\alpha, K, T) \to \rho_S(\alpha, K, T)$$

The trading volume of spread options is much smaller than that of options on single stock making it very difficult to calibrate implied correlation on market spread options prices. It has been suggested to calibrate part of the IC coefficients to single stock options, fitting the individual implied volatility surfaces (IVS). Then, the correlation ρ is seen as an additional input that can be calibrated to recover the impled correlation curve. However, as discussed in Section (**??**), nearly all stock and equity options are American options. Nonetheless, mostly all the authors discussed the calibration of their model to single stock European options. To overcome this problem, we suggested to first extract synthetic European option prices from American option prices, and then apply the usual implied method to the extracted European option prices.

7.4.3 Modelling the implied correlation

Academics argued that new models were needed to capture the ICS, which is another way of saying that returns are not Gaussian and must be modelled accordingly. Alexander et al. [2004b] proposed a bivariate normal mixture model for the underlying assets at maturity. Some authors proposed to extract correlation information from historical data and market prices. Cont et al. [2008] suggested to statistically extract an implied correlation matrix from index option prices. However, when pricing path-dependent options we need to account for the dynamics of the underlyings.

7.4.3.1 The stochastic correlation

Alternatively, new models developed such as stochastic correlation models (see Ma [2009a] [2009b], Fengler et al. [2009]). For example, assuming the Bachelier model with stochastic correlation

$$dS_{1,t} = \sigma_1 dW_{1,t} \quad (7.4.16)$$
$$dS_{2,t} = \sigma_2 dW_{2,t} \quad (7.4.17)$$
$$d < W_1, W_2 >_t = \rho_t dt \quad (7.4.18)$$

Ma [2009b] used the Jacobi diffusion model (for details see Delbaen et al [2002]) to bound the stochastic correlation

$$d\rho_t = \kappa\big(\rho_\infty - \rho_t\big)dt + \xi\sqrt{(1-\rho_t)(1+\rho_t)}dW$$

where W is a Brownian motion independent from W_1 and W_2. Further, $\kappa > 0$ is the speed of mean reversion, $\xi > 0$ is the volatility of the instantaneous correlation, $-1 < \rho_0 < 1$ and $-1 < \rho_\infty < 1$. Under the constraint

$$\frac{\xi^2}{\kappa} - 1 < \rho_\infty < 1 - \frac{\xi^2}{\kappa}$$

the correlation ρ_t does not exit the interval $[-1, 1]$. Generalising the multiscale model of Fouque et al. [2003] to options on multi-assets, Carmona et al. [2011] developed a two factor multiscale stochastic volatility model, and used asymptotic methods to approximate the prices of spread options. In Fouque's model, two factors Z_t and V_t are used to capture fast and slow scale volatilities respectively. The former is fast mean-reverting with $\epsilon >$ corresponding to the fast time scale of the process. The latter is a slow-varying factor with small parameter δ. In that model, the implied volatility is approximated by

$$\Sigma(K,T) \approx \beta(T) + \alpha(T)\frac{1}{T}\ln\frac{K}{x}$$

which is linear with respect to the log-moneyness to maturity $\frac{1}{T}\ln\frac{K}{x}$. More precisely, the implied volatility satisfies

$$\Sigma(K,T-t) \approx \overline{\sigma} + b^{\epsilon} + \frac{a^{\epsilon}}{T-t}\ln\frac{K}{x} + a^{\delta}\ln\frac{K}{x} + b^{\epsilon}(T-t)$$

where $\overline{\sigma}$ is the average implied volatility, a^{δ} and b^{ϵ} are explicit function of the model. Allowing for more flexibility than the multivariate lognormal model, they can account for non-Gaussian returns in a multi-assets framework, but the market becomes incomplete.

7.4.3.2 The local correlation

Following Dupire's LV model (see Section (1.7.2.2)), Carmona et al. [2011] developed a Local Correlation (LC) theory to price spread options, where the instantaneous correlation is a deterministic function of time and the underlying prices. The correlation is determined by the conditional expectation

$$\rho(t,x_1,x_2) = E[\rho_t|S_{1,t}=x_1,S_{2,t}=x_2]$$

To compute the local correlation we first need to define the Radon transform.

Definition 7.4.2 *Given $f(x,y)$ an integrable continuous function defined on $\mathbb{R}^2$, the Radon transform of f is the function Rf defined by the line integral*

$$Rf(\alpha,K) = \int_{-\infty}^{\infty} f(K+\alpha y,y)dy$$

In the special case where the volatilities of the two processes are constant, $\sigma_1(t,x_1,x_2)=\sigma_1$ and $\sigma_2(t,x_1,x_2)=\sigma_2$, the local correlation becomes

$$\rho(t,x_1,x_2) = \frac{R^{-1}\Big(\partial_{KK}C\big(\frac{\sigma_1}{2\alpha\sigma_2}+\frac{\alpha\sigma_2}{2\sigma_1}\big) - \partial_t C\frac{1}{\alpha\sigma_1\sigma_2}\Big)}{R^{-1}(\partial_{KK}C)}$$

When the volatilities σ_1 and σ_2 are not constants the local correlation formula is more involved as it requires computing the density function

$$f_{S_{1,t},S_{2,t}} = R^{-1}(\partial_{KK}C(t,\cdot,\cdot))$$

and then the terms $R\big(\sigma_i^2(t,\cdot,\cdot)f_{S_{i,t},S_{j,t}}(\cdot,\cdot)\big)$ for $i=1,2$.

Problems We saw in Section (1.7.2.1) that Gyongy [1986] gave sufficient conditions for an n-dimentional Ito process to have the same one dimensional marginal distributions as a Markov diffusion process with drift and volatility being deterministic functions of time and the underlying process. Since the marginal distributions of the original model are matched to the simple Markov process, this approach is used to calibrate the parameters of a high-dimensional dynamics to European option prices. That is, the theoretical foundation of projecting an n-dimentional Ito process into a Markov process only applies to European options and not to American options because the joint distributions are not matched. However, in practice spread options are not liquid and practitioners calibrate their models to single stock options, which are American options.

7.5 The option on a basket

7.5.1 Presenting stock index options

We briefly introduce stock index options (SIO) as we will talk about them when presenting dispersion trading. Introduced in 1981, stock index options are options on stock market indices, both on the spot and on futures contracts. The most common ones are options on the SP500, SP100, on NYSE and on the major market index (MMI) as well as options on the institutional index (XII). Note, there is no trading going on in the underlying index itself, it is a calculated value and exists only on paper. The options only allow one to speculate on the price direction of the underlying index, or to hedge all or some part of a portfolio that might correlate closely to that particular index. Some indices are price weighted, where the value of the index is essentially the sum of the total value of an equal number of stocks that make up that index. Another approach to valuation is known as capitalisation weighing where the total capitalisation of each of the stocks in the index are added together and divided by some divisor amount that reduces that total to some reasonable number. Most of the major index options, such as options on the SP500 (SPX), Russell 2000 (RUT), Nasdaq 100 (NDX), Dow Jones 30 (DJX), XII, are European type, while some of them, such as options on the SP100, are America type. On the other hand, all optionable stocks and exchange traded funds (ETFs) have American-style options. Options on stock indices are similar to ETFs, the difference being that ETFs values change throughout the day whereas the value on stock index options change at the end of each trading day. Even though these options are complex to price, the market used the Black-Scoles formula, such that some academics investigated the model errors (see Figlewski [1985]). Brenner et al. [1987] identified three major differences between stock index options and single stock options

- IO1 the irregular pattern of discrete dividends on the index. It impacts the valuation of American options.
- IO2 the effect of interest rate volatility. It has an effect on the probability of early exercise of American options.
- IO3 the fact that the stock index is not directly traded, only the component stocks are. It creates a hedging problem for options written on large indices. Thus, there is no arbitrage relationship guaranteeing a particular link between the options and the index. The use of futures contracts, when they exist, can improve the hedge. However, it is problematic in the case of American options as futures price can stronly deviate from its theoretical price.

Brenner et al. found that in the case of narrow index the discrete dividend had an impact on the valuation formula. When futures price can not reproduce its theoretical level, basis risk is introduced which can not be hedged away. Further, stock index options are settled with cash, so that sellers are facing the risk of early cash settlment.

7.5.2 Pricing basket option

7.5.2.1 Definition

A basket option is a path-independent option whose terminal payoff is a function of several asset prices at the maturity date. While its value depends on the implied volatilities of its constituents, it also depends on the correlations between them. Thus, combining multiple assets can reduce the level of risk via diversification. However, empirical correlations can change dramatically through time, making its use unreliable and more importantly the risk management of basket options difficult.
We let a basket composed of N stock prices be given by $B(t) = \sum_{i=1}^N \omega_i S_i(t)$, where ω_i represent the weight of the basket, and define the price of the basket option at evaluation time $t = 0$ for strike K and maturity T, under the risk-neutral measure $\mathbb{Q}$, as

$$BC(N,T,K) = P(0,T)E^Q[(B(T) - K)^+] = P(0,T)E^Q[(\sum_{i=1}^N \omega_i S_i(T) - K)^+]$$

Note, the weights ω_i are known at the outset so that they stay the same over time.

***Proposition* 12** *The value of a portfolio of European/American call/put options, with common strike and maturity, always exceeds the value of the corresponding basket option.*

By sublinearity of the maximum, we get the following relation on the intrinsic value

$$\max(B(T)-K,0) \leq \sum_{i=1}^{N} \omega_i \max(S_i(T)-K,0)$$

If the portfolio of options is exercised on the optimal excercise date of the option on the portfolio, the payoff of the former is never less than that on the latter. Hence, by the dominance principle (see Proposition 10) we see that a basket option is cheaper than the corresponding portfolio of plain vanilla options. In addition, a basket option takes into account the imperfect correlation between the asset components, and minimise the transaction costs. However, if we assume that the i-th stock price follows the dynamics of an Affine model, such as the Heston one factor model, and apply Ito's Lemma to the forward basket price $B(t,T) = \sum_{i=1}^{N} \omega_i F_i(t,T)$ where $F_i(t,T) = \frac{S_i(t)}{P(t,T)}$, then its dynamics are no-longer affine because of the processes $F_i(t,T)$ in the diffusion terms. In general, there is no explicit analytical expression available for the distribution of the weighted sum of the assets, and using the Black-Sholes formula on a collection of underlying stocks will not produce a closed-form solution to the price of a basket option. As a result, determining the price of a basket option is not a trivial problem.
The most straightforward extension of the univariate BS-model is the Gaussian copula model, where the stocks composing the basket are lognormally distributed and a Gaussian copula connetcs the marginals (see Hull et al. [1993], Milevsky et al. [1998]). When asset prices evolve according to a multivariate binomial tree, the multivariate risk-neutral density converge to a multivariate lognormal in the limits as the time step approaches zero (see Boyle [1988], Boyle et al. [1989]). Rosenberg [1998] developed a parametrised density function for the valuation of multi-asset options which was extended to the case of nonparametric multivariate density (see Rosenberg [2003]). El Karoui et al. [1998] provided upper and lower bounds on a European call on the arithmetic average of two independent stock price processes. In order to obtain bounds to the price of a basket option several authors applied methods used for Asian options combined with the notion of comonotonicity (see Nielsen et al. [2003]). That is, Rogers et al. [1995] proposed a method of conditioning to price Asian options. Kaas et al. [2000] developed a general technique based on comonotonic risks to derive upper and lower bounds for stop-loss premiums of sums of dependent random variables. Since the problem of pricing arithmetic basket options is equivalent to calculating stop-loss premiums of a sum of dependent risks, Deelstra et al. [2002] combined these results with the conditioning technique of Rogers et al. to get upper and lower bounds to the price of basket options. Alternatively, using a moment-matching method, Brigo et al. [2004] found an equivalent log-normal random variable having the same mean and variance as the weighted basket of its constituents. Thus, the BS-formula could be applied.

7.5.2.2 The Gaussian copula

Even though Gaussian copula model has been used to approximate the price of multi-asset European payoffs, the normality assumption for the marginals is too restrictive. Other copulas have been proposed and a framework has been described by Cherubini et al. [2004], but the pricing of basket options is not straightforward. Since option pricing is done under the risk-neutral measure $\mathbb{Q}$, we generally assume that the copula under $\mathbb{Q}$ belongs to the same family as the one under the $\mathbb{P}$-measure. Some authors impose they have the same parameters under both measres, others allow for different parameters.
For simplicity of exposition we briefly describe the pricing of a multi-asset option with two underlyings $S_{1,t}^{x_1}$ and $S_{2,t}^{x_2}$. The price can be expressed as a double integral as follow

$$BC(t,T) = P(t,T) \int_0^\infty \int_0^\infty h(x_1,x_2)\phi_B(x_1,x_2)dx_1dx_2$$

where $\phi_B(\cdot,\cdot)$ is the joint density of $S_{1,t}$ and $S_{2,t}$ under the risk-neutral measure. From Skalar's theorem, we can express that joint density in terms of the marginal densities ϕ_1 and ϕ_2 of $S_{1,t}$ and $S_{2,t}$ as follow

$$\phi_B(x_1, x_2) = c_{12}\big(F_1(x_1), F_2(x_2)\big) f_1(x_1) f_2(x_2)$$

where $F_i(\cdot)$ is the cumulative distribution function (cdf) of $S_{i,t}$ (see details in Section (2.1.3.3)). Note, $c_{12} = \partial_{y_1,y_2} C(y_1, y_2)$ where $C(\cdot,\cdot)$ is the copula between $S_{1,t}$ and $S_{2,t}$ under the $\mathbb{Q}$-measure. Thus, all we need to compute the option price is the marginal distributions F_i, $i = 1, 2$, and the copula $C(\cdot,\cdot)$. For simplicity, we now describe the Gaussian copula. Assuming a continuum of prices on the basket option market and single stock option market, we use Equation (1.5.25) to compute the cumulative distribution $P(B(T) < K)$ of the basket and Equation (1.5.29) to compute its density $\phi_B(t_0; T, K)$. Similarly, we can compute the cumulative distribution $P(S_i(T) < K)$ of the ith single stock and its associated density $\phi_{S_i}(t_0; T, K)$. We let ψ be a n-dimensional standard normal correlated Gaussian variate with

$$E^Q[\psi_i \psi_j] = \rho_{ij} \; , \, i \neq j$$

and

$$E^Q[\psi_i] = 0 \; , \, E^Q[\psi_i^2] = 1$$

for a given correlation matrix ρ_{ij}. Note, the random variables ψ_i are driven by a systematic factor $W(\rho)$ and a stock specific factor $W_i(1-\rho)$, where W and W_i are standard Brownian motions. The ith stock price at maturity T satisfies

$$S_i(T) = S_i(t_0) e^{\mu_i (T-t_0) + \sigma_i \sqrt{T-t_0} \psi_i}$$

such that the log returns are modelled by Normal marginals and a Gaussian copula. We can construct the joint distribution of the basket from its constituents by inverting

$$n_i \equiv N(\psi_i) = P(S_i(T) < K) \; , \, i = 1, ..., N$$

The resulting multivariate distribution is denoted by

$$\phi_B(t_0; S_1(T), ..., S_N(T); T, K)$$

Thus, we can approximate the price of the basket option as follow

$$\widehat{BC}(N, T, K) = P(0, T) \int_0^\infty \big(\sum_{i=1}^N \omega_i S_i - K\big)^+ \phi_B(t_0; S_1, ..., S_N; T, K) dS_1 ... dS_N$$

7.5.2.3 The nonparametric density model

Using results on copulas, Rosenberg [2003] created a multivariate risk-neutral sensity by combining nonparametrically estimated marginal risk-neutral densities with nonparametric dependence function. In the bivariate case, the pricing of the option is performed with Equation (7.3.8) with densities under the risk-neutral measure. The joint probability density $f^Q(S_{1,T}, S_{2,T})$ can be expressed as the product of the marginals $f^Q(S_{i,T})$, $i = 1, 2$, and the risk-neutral dependence function (see Equation (2.1.7)). Then, the risk-neutral dependence function $c^Q(\cdot,\cdot)$ (see Equation (2.1.8)) can be expressed as the objective dependence function scaled by a dependence risk-adjustment. If we let $M(S_{i,T})$ be the pricing kernel projected onto $S_{i,T}$ and $M(S_{i,T}, S_{j,T})$ be the pricing kernel projected onto $S_{i,T}$ and $S_{j,T}$, then each risk-neutral density can be expressed as

$$\begin{aligned}
f^Q(S_{i,T}) &= P^{-1}(T) M(S_{i,T}) f(S_{i,T}) \; , \, i = 1, 2 \\
f^Q(S_{1,T}, S_{2,T}) &= P^{-1}(T) M(S_{1,T}, S_{2,T}) f(S_{1,T}, S_{2,T})
\end{aligned}$$

Thus, replacing in the risk-neutral dependence function, we get

$$c^Q(F^Q(S_{1,T}), F^Q(S_{2,T})) = \frac{f(S_{1,T}, S_{2,T})}{f(S_{1,T})f(S_{2,T})}[P(T)\frac{M(S_{1,T}, S_{2,T})}{M(S_{1,T})M(S_{2,T})}]$$

Rosenberg assumed the dependence risk-adjustment to be equal to one, with the pricing kernel

$$M(S_{1,T}, S_{2,T}) = e^{z_1(S_{1,T}) + z_2(S_{2,T})}$$

for arbitrary continuous functions $z_i(\cdot)$, $i = 1, 2$. Putting terms together the risk-neutral joint density becomes

$$f^Q(S_{1,T}, S_{2,T}) = f^Q(S_{1,T})f^Q(S_{2,T})[\frac{f(S_{1,T}, S_{2,T})}{f(S_{1,T})f(S_{2,T})}]$$

The dependence function is estimated by using historical returns and the marginal densities are fitted to option prices.

7.5.2.4 Simple bounds

We follow El Karoui et al. [1998] and describe simple bounds to the pricing of options on arithmetic basket. In the BS-world, assuming σ_{ij} is a diagonal matrix, then the stock price follow the dynamics

$$\frac{dS_i(t)}{S_i(t)} = r(t)dt + \sigma_i(t)dZ_i^Q(t) \, , \, i = 1, ..., N$$

with independent Brownian motions Z_i^Q for $i = 1, ..., N$ where the interest rate and the volatilities $\sigma_i(t)$ are deterministic functions of time. Then, the dynamics to the forward basket price $B(t,T) = \sum_{i=1}^N \omega_i F_i(t,T)$ satisfy

$$dB(t,T) = \sum_{i=1}^N \omega_i F_i(t,T)\sigma_i(t)dZ_i^Q(t)$$

which we can rewrite as

$$\frac{dB(t,T)}{B(t,T)} = \sum_{i=1}^N \widehat{\omega}_i(t)\sigma_i(t)dZ_i^Q(t)$$

where

$$\widehat{\omega}_i(t) = \frac{\omega_i F_i(t,T)}{\sum_{i=1}^N \omega_i F_i(t,T)} = \frac{\omega_i F_i(t,T)}{\omega^\top F(t,T)} \, , \, i = 1, ..., N$$

with $0 \leq \widehat{\omega}_i(t) \leq 1$ and $\sum_{i=1}^N \widehat{\omega}_i(t) = 1$. Assuming independend Brownian motions Z_i^Q for $i = 1, .., N$, we define a new Brownian motion

$$Z^Q(t) = \int_0^t \frac{\sum_{i=1}^N \omega_i F_i(u,T)\sigma_i(u)dZ_i^Q(u)}{\sqrt{\sum_{i=1}^N \omega_i^2 F_i^2(u,T)\sigma_i^2(u)}}$$

and set the volatility of the basket as

$$\sigma_{B,t} = \frac{\sqrt{\sum_{i=1}^N \omega_i^2 F_i^2(t,T)\sigma_i^2(t)}}{B(t,T)}$$

We can rewrite the dynamics of the forward basket as follow

$$\frac{dB(t,T)}{B(t,T)} = \sigma_{B,t}dZ^Q(t)$$

where $\sigma_{B,t}$ is a stochastic volatility. In the special case of an arithmetic mean, the weights become $\omega_i = \frac{1}{N}$, and we can rewrite the Brownian motion as

$$Z^Q(t) = \int_0^t \frac{\sum_{i=1}^N F_i(u,T)\sigma_i(u)dZ_i^Q(u)}{\sqrt{\sum_{i=1}^N F_i^2(u,T)\sigma_i^2(u)}}$$

and the stochastic volatility as

$$\sigma_{B,t} = \frac{\sqrt{\sum_{i=1}^N F_i^2(t,T)\sigma_i^2(t)}}{NB(t,T)}$$

In order to obtain bounds to the price of a basket option, we use the property of convexity of call/put option payoffs to find a lower bound α and an upper bound β to the basket volatility $\sigma_{B,t}$.
Setting $\beta = \sigma_1(t) \vee \sigma_2(t) \vee .... \vee \sigma_N(t)$ and replacing each volatility σ_i, $i = 1, ..., N$ in the basket volatility with β, we get

$$\sigma_{B,t} \leq \frac{\beta\sqrt{\sum_{i=1}^N F_i^2(t,T)}}{NB(t,T)} \leq \frac{\beta\sum_{i=1}^N F_i(t,T)}{NB(t,T)} = \beta$$

Then, setting $\alpha = \frac{\prod_{i=1}^N \sigma_i(t)}{\sqrt{\sum_{i=1}^N \sigma_i^2(t)}}$ we need to prove that $\alpha \leq \sigma_{B,t}$ which is more difficult that in the case of the upper bound. We can set the following bounds to the volatility

$$\begin{aligned} \alpha &= \frac{1}{\sqrt{\sum_{i=1}^N \frac{1}{\sigma_i^2(t)}}} \\ \beta &= \sigma_1(t) \vee \sigma_2(t) \vee .... \vee \sigma_N(t) \end{aligned}$$

such that we have $\alpha \leq \sigma_t \leq \beta$ for all $t \in [0,T]$, almost surely. Hence, we can bound the price of a basket option as

$$C_{BS}(t, B(t,T); \alpha) \leq BC(N, t.T, K) \leq C_{BS}(t, B(t,T); \beta)$$

where the bounds are computed explicitely.
In addition, a hedger selling the option for $C_{BS}(t, B(t,T); \beta)$ and using the hedging portfolio

$$\Delta_\beta(t) = \frac{\partial}{\partial x} C_{BS}(t, B(t,T); \beta)$$

is guaranted to have at least $(B(T,T) - K)^+$ at maturity T, and to be overhedged at each time $t \in [0,T)$.
Similarly, a hedger borrowing $C_{BS}(t, B(t,T); \alpha)$ to buy the option and using the hedging portfolio

$$\Delta_\alpha(t) = -\frac{\partial}{\partial x} C_{BS}(t, B(t,T); \alpha)$$

is guaranteed to have accumulated no more than $(B(T,T) - K)^+$ in debt at maturity T.

Alternatively, we can obtain a lower bound to the basket option by noting that the geometric mean $G(t) = \prod_{i=1}^N (S_i(t))^{\omega_i}$, with $\omega_i = \frac{1}{N}$, always lies below the arithmetic basket $B(t)$. Taking the logarithm, we get $LG(t) = \sum_{i=1}^N \omega_i \log S_i(t)$, so that the dynamics of $LG(t)$ becomes

$$dLG(t,T) = \sum_{i=1}^N \omega_i d\log F_i(t,T) = -\frac{1}{2N}\sum_{i=1}^N \sigma_i^2(t)dt + \frac{1}{N}\sum_{i=1}^N \sigma_i(t)dZ_i^Q(t)$$

since $\omega_i = \frac{1}{N}$. Again, we set the Brownian motion as

$$\tilde{Z}^Q(t) = \int_0^t \frac{\sum_{i=1}^N \sigma_i(u) dZ_i^Q(u)}{\sqrt{\sum_{i=1}^N \sigma_i^2(u)}}$$

and the stochastic volatility as

$$\tilde{\sigma}_t = \sqrt{\sum_{i=1}^N \sigma_i^2(t)}$$

Note, this volatility corresponds to the denominator of the lower bound α of an option on the arithmetic basket.

7.5.2.5 The comonotonic upper bound

We briefly describe the comonotonic upper bound proposed by Deelstra et al. [2002]. Rewriting the basket as

$$B(t) = \sum_{i=1}^N X_i(t) = \sum_{i=1}^N \alpha_i(t) e^{Y_i(t)}$$

where $\alpha_i(t) = \omega_i S_i(0) e^{(r - \frac{1}{2}\sigma_i^2)t}$ and $Y_i(t) = \sigma_i W_i(t) \sim N(0, \sigma_i^2 t)$. That is, $X_i(t) = \omega_i S_i(t)$. Hence, the process $X_i(t)$ is lognormally distributed

$$X_i(t) \sim LN\big(\omega_i S_i(0) + (r - \frac{1}{2}\sigma_i^2)t, \sigma_i^2 t\big)$$

and the stop-loss premium $E^Q[(X_i - d_i)^+]$ for some retention d_i can be computed. Given $\ln(X_i(t)) \sim N(\mu_i(t), \sigma^2_{Y_i(t)})$ with $\mu_i(t) = \ln(\alpha_i(t))$ and $\sigma_{Y_i(t)} = \sigma_i \sqrt{t}$, we get

$$E^Q[(X_i - d_i)^+] = e^{\mu_i(t) + \frac{1}{2}Y_i^2(t)} N(d_{i,1}(t)) - d_i N(d_{i,2}(t))$$

where

$$d_{i,1}(t) = \frac{\mu_i(t) + \sigma^2_{Y_i(t)} - \ln(d_i)}{\sigma_{Y_i(t)}} \text{ and } d_{i,2}(t) = d_{i,1}(t) - \sigma_{Y_i(t)}$$

and $N(\bullet)$ is the cdf of the $N(0,1)$ distribution. We can derive the inverse cdf $F_{X_i}^{-1}$ and get

$$B_c(t) = \sum_{i=1}^N \alpha_i(t) e^{\frac{1}{2}\sigma^2_{Y_i(t)} N^{-1}(U)}$$

for any uniformaly distributed random variable U distributed on the unit interval. The upper bound for any $K > 0$ is given by

$$BC(N,T,K) \leq \sum_{i=1}^N \omega_i S_i(0) N(\sigma_i \sqrt{T} - N^{-1}(F_{B_c}(K))) - e^{-rT} K (1 - F_{B_c}(K))$$

where $F_{B_c}(K)$ satisfies

$$\sum_{i=1}^N \omega_i S_i(0) e^{(r - \frac{1}{2}\sigma_i^2)T + \sigma_i \sqrt{T} N^{-1}(F_{B_c}(K))} = K$$

Following Simon et al. [2000] we can rewrite this upper bound as a combination of Black-Scholes prices. As a result of the lognormality of S_i, we get

$$F^{-1}_{\omega_i S_i}(p) = \omega_i F^{-1}_{S_i}(p) \text{ for all } p \in [0,1]$$

Remark 7.5.1 *The comonotonic upper bound is the smallest linear combination of European BS-prices dominating the basket option price.*

We can write it in terms of BS-price with strike $K_i = F^{-1}_{S_i}(F_{B_c}(K))$ as follow

$$\begin{aligned} BC(N,T,K) &\leq P(0,T) \sum_{i=1}^N \omega_i E^Q[(S_i(T) - F^{-1}_{S_i}(F_{B_c}(K)))^+] \\ &= \sum_{i=1}^N \omega_i \big(S_i(0) N(d_{i,1}) - P(0,T) K_i N(d_{2,1}) \big) \end{aligned}$$

where K_i satisfies

$$K_i = F^{-1}_{S_i}(F_{B_c}(K)) = S_i(0) e^{(r - \frac{1}{2}\sigma_i^2)T + \sigma_i \sqrt{T} N^{-1}(F_{B_c}(K))}$$

and

$$d_{i,1} = \frac{\ln \frac{S_i(0)}{K_i} + (r + \frac{1}{2}\sigma_i^2)T}{\sigma_i \sqrt{T}} \text{ and } d_{i,2} = d_{i,1} - \sigma_i \sqrt{T}$$

Remark 7.5.2 *This comonotonic upper bound does not dependent on the instantaneous correlations ρ_{ij}, $i = 1, ..., N$.*

Hence, we still have an upper bound to the price of the basket option even if the correlations are not known. Deelstra et al. [2002] improved this upper bound by taking into account the instantaneous correlations.

7.5.3 The correlation

We have introduced in Section (7.3.2) the notion of implied correlation (IC) and we presented in Section (7.4.3.2) a local correlation model (LCM) for spread options. However, since there is no closed-form solution to index options, one can not use Definition (7.4.1) to define the implied correlation of index options. Nonetheless, given that there exist a market for stock index options (SIO), practitioners tried to explain the observed index option prices in terms of their constituent distributions. They observed that the implied correlation would increase as the index decreases and called this effect the correlation skew. We are now going to discuss the implied correlation for index options and introduce the local correlation model.

7.5.3.1 The implied correlation

Assuming that the change in composition of the index during the life of the option is a rare event, the weights have been considered fixed and the multidimensional integration in Equation (7.3.8) has been used with a Gaussian copula to compute the basket option price $\widehat{BC}(t;K,T)$. It is argued that if the Gaussian copula model was a good model to price index options we should get

$$C_I(t;K,T) = \widehat{BC}(t;K,T) \; \forall K,T$$

However, as discussed in Section (7.5.1), most of the major index options are European type, and there is no trading going on in the underlying index itself, creating a hedging problem for options written on large indices. On the other hand, single stock options are American type and one can not enter the American volatility in the pricing equation of the basket option $\widehat{BC}(t;K,T)$. Thus, the stock index options and the basket options seems hardly comparable. Still, various authors analysed the volatility skew of the index options against that of its constituents.

For example, Langnau [2009] [2010] compared the index option prices of the Euro Stoxx 50 against theirs constituents prices. The distribution of the Euro Stoxx 50 index together with that of all the constituent of the basket are computed and an approximated option price with Gaussian copula is derived. It is observed that half the volatility skew of the index is explained by the skew of the index constituents. It is concluded that the correlation depends on the strike of the option impacting the distribution of the index on the downside. The market impled volatility smile is compared with the one generated with a Gaussian copula at maturity $T = 2$ and $T = 4$ years. Defining the skew as $skew = \frac{\Delta\Sigma}{\Delta\eta}$, in the former, the skew around the money ($\eta \in [0.9, 1]$) is about $skew \approx \frac{0.28-0.26}{0.1}$, and in the latter it is about $skew \approx \frac{0.285-0.265}{0.1}$. When anaysing the market skew of index options, the choice of these maturities is rather strange considering that the liquidity ranges from one week to one year, and that the skew is most prononced for short maturities and decreases over time. Further, the one year ATM spread of volatility on the Euro Stoxx 50 is about 0.004, the two year ATM spread is about 0.006 and that a four year ATM spread is above 0.01 (1 vol = 100 basis points). In the case of single stock options, these maturities are not quoted and the ATM spread could easily reach a few vol. In addition, the spread of volatility increases as we move away from the money, even if we account for the vega. However, in that example, the slope of the smile at maturity $T = 2$ and $T = 4$ is roughly constant over the moneyness. It would be more reasonable to conclude that the difference of volatility smile between the index option and the one obtained by combining its constituents with a Gaussian copula is absorbed by market liquidity.

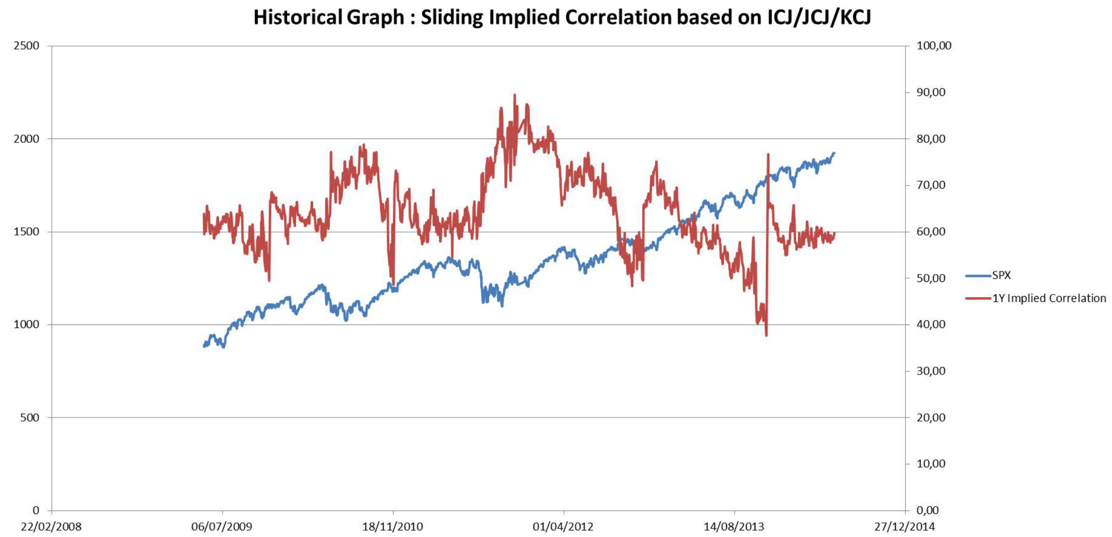


Figure 7.1: Historical graph: sliding implied correlation Delanoe [2014].

Similarly, Delanoe [2014] considered the evolution of the SP500 and its associated $1Y$ sliding implied correlation, and concluded that the implied correlation tended to increase for decreasing index. On the graph in Figure (7.1), the index is slowly increasing from below 1000 on the $6/07/09$ to below 2000 around the $06/14$, with a jump occuring towards the end of 2011. However, the $1Y$ IC is moving randomly around a level of 60 with increasing behaviours both when the index increases as well as when it decreases. The process is quite eratic, with jumps, and nothing can be concluded about a possible relationship between the implied correlation and the index. He also analysed the SP500 index smile for a one year maturity with a stock model having local volatilities and constant correlation calibrated to match ATM index option, and concluded that additional correlation needed to be added. For the correlation matrix to remain positive semi-definite he used convexity property of the matrix obtaining

$$\widehat{\rho}_{ij} = (1 - \varpi)\rho_{ij}^H + \varpi \tag{7.5.19}$$

where $\widehat{\rho}_{ij}$ is the bumped correlation matrix, ρ_{ij}^H is the historical correlation matrix and $\varpi \approx 15\%$.

7.5.3.2 The local correlation

In the spirit of Dupire's LV model (see Section (1.7.2.2)), Langnau [2009] [2010] and Reghai [2010] generalised the approach to multi-assets options obtaining a Local Correlation (LC) model. Langnau considered a local volatility model for each asset and combined them with a local correlation matrix. The local correlation is based on a family of perturbations of a given center correlation matrix preserving all the relevant properties of the system. A particular form of the correlation matrix that fit the index skew is derived. However, the form of the correlation matrix is derived by assuming that the correlation matrix satisfies $\rho_{ij}(t) \geq 0$ and that if the inequality in Equation (7.3.14) holds, there is no-dispersion arbitrage between individual option markets and the option market of the basket. Thus, the local correlation depends on the knowledge of the individual local volatility together with that of the basket option at all time t and for all level of the constiuent assets. Ignoring interest rates, Reghai [2010] assumed that the dynamics of the constituent assets follow

$$\frac{dS_{i,t}}{S_{i,t}} = \sigma(t, S_{i,t})\big(\sqrt{1 - \varpi(t, B(t))}dW_{i,t} + \sqrt{\varpi(t, B(t))}dW_{B,t}\big)$$

where $B(t) = \sum_{i=1}^N \omega_i S_{i,t}$ is the value of the basket at time t, and

$$< dW_i, dW_j >_t = \rho_{ij}^0 dt \; , \; < dW_i, dW_B >_t = 0 \; \forall i$$

such that the correlation satisfies Equation (7.5.19) with $\rho_{ij}^H = \rho_{ij}^0$. Using Cholesky decomposition on ρ_{ij}^0, we can only bump the correlation matrix in a positive way. That is, we must impose $\varpi \in [0, 1]$. To obtain an expression for the local correlation, Reghai [2010] considered a fixed-point algorithm on implied quantities. Alternatively, Delanoe [2014] considered a fixed-point algorithm on local quantities (conditional expectations) and derived the call price of the basket option with respect to the maturity. Letting $H_t = h(S_{1,t}, ..., S_{N,t})$ be the payoff of the multi-asset option, he obtained

$$\begin{aligned}
\partial_T C &= \frac{1}{2} K^2 \partial_{KK} C \sigma_{H,LV}^2(K,T) \\
\partial_T C &= \frac{1}{2} E^Q\Big[\Big(\big(\sum_{i,j}^N \partial_{x_i x_j} h I_{\{H_T > K\}} + \partial_{x_i} h \partial_{x_j} h I_{\{H_T = K\}}\big) S_{i,T} S_{j,T} \sigma_i \sigma_j \rho_{ij}\Big)\Big]
\end{aligned}$$

where $\sigma_{H,LV}^2(K,T)$ is the local variance of the multi-asset option. Equating the two models, we must have

$$K^2 \sigma_{H,LV}^2(K,T) = -\frac{\partial_K C}{\partial_{KK} C} E^Q[\sum_{i,j}^N \partial_{x_i x_j} h \sigma_i \sigma_j \rho_{ij} | H_T > K] + E^Q[\sum_{i,j}^N \partial_{x_i} h \partial_{x_j} h S_{i,T} S_{j,T} \sigma_i \sigma_j \rho_{ij} | H_T = K]$$

with

$$\rho_{ij}(S_{1,t}, ..., S_{N,t}) = \big(1 - \varpi(t, h(S_{1,t}, ..., S_{N,t}))\big)\rho_{ij}^0 + \varpi(t, h(S_{1,t}, ..., S_{N,t}))$$

In the case of the basket option, we get $H_t = B(t)$, and the above equation simplifies to

$$K^2 \sigma_{B,LV}^2(K,T) = E^Q[\sum_{i,j}^N \omega_i \omega_j S_{i,T} S_{j,T} \sigma_i \sigma_j \rho_{ij} | B(T) = K]$$

Thus, the local volatility $\sigma_{B,LV}(K,T)$ of a basket option is a condtitional expectation (see Avellaneda et al. [2002b]).

Remark 7.5.3 *This equation corresponds to expectation of the instantaneous variance of the basket $V_B(T)$ at maturity T, given in Equation (7.3.13), conditional on the basket being equal to the strike at maturity. That is,*

$$\sigma^2_{B,LV}(K,T) = E^Q[V_B(T)|B(T) = K]$$

Given the definition of the correlation matrix, we obtain

$$\varpi(K,T) = \frac{K^2\sigma^2_{B,LV}(K,T) - E^Q[\sum_{i,j}^N \omega_i\omega_j S_{i,T}S_{j,T}\sigma_i\sigma_j\rho^0_{ij}|B(T) = K]}{E^Q[\sum_{i,j}^N w_i w_j S_{i,T}S_{j,T}\sigma_i\sigma_j(1-\rho^0_{ij})|B(T) = K]}$$

For computation stability this equation can be rewritten as follow

$$\varpi(K,T) = \frac{\sigma^2_{B,LV}(K,T) - E^Q[\frac{\sum_{i,j}^N \omega_i\omega_j S_{i,T}S_{j,T}\sigma_i\sigma_j\rho^0_{ij}}{\sum_{i,j}^N \omega_i\omega_j S_{i,T}S_{j,T}}|B(T) = K]}{E^Q[\frac{\sum_{i,j}^N \omega_i\omega_j S_{i,T}S_{j,T}\sigma_i\sigma_j(1-\rho^0_{ij})}{\sum_{i,j}^N \omega_i\omega_j S_{i,T}S_{j,T}}|B(T) = K]}$$

7.5.3.3 Problems

Again, the theoretical foundation of projecting an n-dimentional Ito process into a Markov process only applies to European options and not to American options because the joint distributions are not matched. However, European options on the constituents of indices, such as the Euro Stoxx 50 index or the SP500, are not listed, and the liquid options are only American options. Hence, one should question the relevance of a local correlation model.

7.6 Dispersion trading

7.6.1 Introduction

7.6.1.1 From portfolio theory to option pricing theory

A direct consequence of the Efficient Market Hypothesis (EMH) is that the most important concepts in theoretical and empirical finance developed around the assumption that asset returns follow a normal distribution. This includes the classical portfolio theory, the Black-Scholes-Merton option pricing model and the RiskMetrics variance-covariance approach to Value at Risk (VaR). When measuring portfolio risk, VAR is the risk expressed in dollar terms showing what amount of money your portfolio could lose during a defined interval with a given probability. It is commonly understood that VAR is in essence the volatility of a portfolio expressed in dollar terms. This concept has been applied not only to estimate the risks of a portfolio of assets, but also for trading the market itself, since the market is considered as one large portfolio of various assets each having a weight in the global market place. However, given the size of such a portfolio, it is more efficient to explore the same concepts focusing on a few subsets of the global markets, such as indices or sectors. For example, an index measures the price performance of a portfolio of selected stocks, and its risk can be evaluated as the risk of the portfolio of stock constituents.

In the problem of portfolio selection, Markowitz [1952] introduced the mean-variance approach as a simple trade-off between return and uncertainty, where one is left with the choice of one free parameter, the amount of variance acceptable to the individual investor. The mean-variance efficient portfolios are obtained as the solution to a quadratic optimisation program, with theoretical justification requiring either a quadratic utility function or some fairly restrictive assumptions on the class of return distribution, such as the assumption of normally distributed returns. For instance, we assume zero transaction costs and a portfolio with price V_t taking values in $\mathbb{R}$ and following the geometric Brownian motion with dynamics under the historical probability measure $\mathbb{P}$ given by

$$\frac{dV_t}{V_t} = \mu dt + \sigma_V dW_t \tag{7.6.20}$$

where μ is the drift, σ_V is the volatility and W_t is a standard Brownian motion. Markowitz postulated that while diversification would reduce risk, it would not eliminate it, so that an investor should maximise the expected portfolio return μ, while minimising portfolio variance of return σ_V^2. The argument follows from the relation between the variance of the return of the portfolio σ_V^2 and the variance of return of its constituent securities σ_j^2 for $j = 1, 2, ..., N$ given by

$$\sigma_V^2 = \sum_j w_j^2 \sigma_j^2 + \sum_j \sum_{k \neq j} w_j w_k \rho_{jk} \sigma_j \sigma_k \tag{7.6.21}$$

where the w_j are the portfolio weights such that $\sum_j w_j = 1$, and ρ_{jk} is the correlation of the returns of securities j and k. Therefore, $\rho_{jk}\sigma_j\sigma_k$ is the covariance of their returns. So, holding various assets not perfectly correlated in a portfolio should offer a reduced risk exposure to a specific asset. Thus, the decision to hold any security would depend on what other securities the investor wants to hold. That is, securities can not be properly evaluated in isolation, but only as a group, resulting in the notion of efficient portfolios.

One consequence of the capital asset pricing model (CAPM) is that statistical approaches must be restricted to second-oder stationary processes. Thus, the portfolio risk becomes a weighted sum of covariation of all stocks in the portfolio. Using volatilities and correlations calculated on any desired historical time interval for price data and different forecast times, we can then compute the risk of the portfolio, or the index, for different terms. Rather than using volatilities and correlations statistically forecasted in Equation (7.6.21), it has been suggested to use implied volatility from option prices because they are biased estimators of the future volatility. Since implied volatility on options on an index as well as implied volatilities on options on single stocks are assumed easy to estimate, it has been proposed to use them directly in Equation (7.6.21). In addition, practitoners defined Correlation as the statistic $\rho_{imp} \approx 1$ and Dispersion as the statistic $\rho_{imp} \approx -1$, leading to a new type of correlation products called dispersion trading (DT). Yet, since several indices are actively trading in the option market, a market on basket options developed. In that market, the dynamics of the basket are assumed to follow a geometric Brownian motion and the BS-formula is used to price and risk manage basket options. However, we saw in Section (7.5) that there was no explicit analytical expression available for the distribution of the weighted sum of the assets, such that using the Black-Sholes formula on a collection of underlying stocks will not produce a closed-form solution to the price of a basket option. That is, due to the non-linearity of option prices with respect to volatility, no such relationship exists in the option pricing theory (OPT). As a result, Markowitz equation has been wrongly used to relate the implied volatility of an option on an index to the implied volatilities of its components. Moreover, practitoners wrongly used this pseudo relation to calculate a value expressing the correlation level between the implied volatilities of the stocks and the index implied volatility observed on the market. Nonetheless, from the no-dominance principle (see Proposition (10)), we must get consistancy between the two markets, and if it is not the case arbitrage opportunities will arise. As a results, traders need a computationally fast and robust model to relate the two markets together and to devise indicators measuring the occurrence of arbitrage oppotunities.

7.6.1.2 Definitions of implied correlation

Given an $\mathbb{R}$ random variable X, we can compute its characteristic function Φ_X, which completely characterise its law, and if it exists, its n-th moment m_n (see Appendix (A.3)). Defining a portfolio as a weighted sum of random variables, Markowitz's portfolio theory is derived from the variance of that sum given in Equation (A.3.2). Following the notation in Section (7.3.4), we consider an index I (or basket) made of N underlying X_i for $i = 1, .., N$ with volatility given by σ_i. In the portfolio theory, the correlation comes from the composition of the index, that is, the variance of the index satisfies

$$
\begin{aligned}
Var(I) &= \sigma_I^2 = \sum_{i,j=1}^{N} w_i w_j Cov(X_i, X_j) \\
&= \sum_{i=1}^{N} w_i^2 Var(X_i) + 2 \sum_{j=1}^{N} \sum_{i<j} w_i w_j Cov(X_i, X_j) \\
&= \sum_{i=1}^{N} w_i^2 \sigma_i^2 + 2 \sum_{j=1}^{N} \sum_{i<j} w_i w_j \sigma_i \sigma_j \rho_{ij} = \sum_{i=1}^{N} w_i^2 \sigma_i^2 + 2 \sum_{i=1}^{N} \sum_{j>i} w_i w_j \sigma_i \sigma_j \rho_{ij}
\end{aligned}
$$

where σ_I is the actual index volatility (also noted σ_P for portfolio volatility), w_i is the weight of the ith-stock in the basket, and ρ_{ij} is the instantaneous correlation between the ith-stock and the jth-stock in the basket.

***Remark* 7.6.1** *The portfolio variance σ_I^2 is not to be confused with the instantaneous basket variance $V_B(t)$ defined in Equation (7.3.12). The former is derived from a probability formula (see Equation (A.3.2)) and has constant weights, while the latter is derived from stochastic calculs within infinitesimal time and has stochastic weights.*

In order to relate portfolio theory (PT) to option pricing theory (OPT), several authors played with the volatilities of the ith component, σ_i, and that of the jth component, σ_j, in the above equation by allowing them to be either historically estimated or implied from option prices.

***Remark* 7.6.2** *From the moment we are plugging implied volatility to both the LHS and RHS of the portfolio variance, we are trying to relate an option on a basket with a sum of options on its component. We enter the option pricing theory (OPT), where the no-dominance principle apply, and this relationship is no-longer valid.*

Nonetheless, over time this approach constituded a market standard. For instance, assuming σ_i and σ_j for the ith and jth component of an index I are implied volatilities, and letting $\rho = \rho_{ij}$ for $i \neq j$, practitioners defined the so-called Implied Correlation (IC) of this portfolio, or average level of correlation, as follow

$$
\rho_{imp} = \frac{\sigma_I^2 - \sum_{i=1}^{N} w_i^2 \sigma_i^2}{2 \sum_{i=1}^{N} \sum_{j>i} w_i w_j \sigma_i \sigma_j} \tag{7.6.22}
$$

which we can also write as

$$
\rho_{imp} = \sum_{i=1}^{N} \sum_{j>i} \frac{w_i w_j \sigma_i \sigma_j}{\sum_{i=1}^{N} \sum_{j>i} \sigma_i \sigma_j} \rho_{ij}
$$

The Implied Index Correlation (IC) attempts at defining the correlation level between the actual implied volatility of the index and the implied volatilities of its stock components. This is in itself absurd. The greater the implied correlation ρ_{imp}, the stronger the correlation between the index implied volatility and that of its constituent stocks, and therefore the more suitable the market conditions for deploying a dispersion strategy. Under some reasonable conditions, Bossu [2006] assumed that the term $\sum_{i=1}^{N} w_i^2 \sigma_i^2$ was close to zero, and showed that a good proxy for the implied correlation was

$$
\rho_{imp}^* = \frac{\sigma_I^2}{\left(\sum_{i=1}^{N} w_i \sigma_i \right)^2} \tag{7.6.23}
$$

since $\left(\sum_{i=1}^{N} w_i \sigma_i \right)^2 = \sum_{i=1}^{N} w_i^2 \sigma_i^2 + \sum_{i \neq j} w_i w_j \sigma_i \sigma_j$. We can call

$$\sigma_{Avg} = \sum_{i=1}^{N} w_i \sigma_i$$

the average implied volatility of the components. Thus, we can approximate the index volatility as follow

$$\sigma_I \approx \sqrt{\rho^*_{imp}} \sigma_{Avg}$$

for $\rho^*_{imp} > 0.15$ and $N > 20$.
The Realised Correlation (RC) was proposed as an alternative to the implied correlation. Denoted $\widehat{\rho}$, it is defined as

$$\widehat{\rho} = \frac{\sum_{i \neq j}^{N} w_i w_j \rho_{ij}}{\sum_{i \neq j}^{N} w_i w_j} \tag{7.6.24}$$

A different definition, which is not market practice, is given by

$$\widehat{\rho} = \frac{\sum_{i \neq j}^{N} w_i w_j \widehat{\sigma}_i \widehat{\sigma}_j \rho_{ij}}{\sum_{i \neq j}^{N} w_i w_j \widehat{\sigma}_i \widehat{\sigma}_j}$$

where $(\widehat{\sigma}_I, \widehat{\sigma}_1, ..., \widehat{\sigma}_N)$ represent realised volatilities. It can be approximated as

$$\widehat{\rho} = \frac{\widehat{\sigma}_I^2}{\left(\sum_{i=1}^{N} w_i \widehat{\sigma}_i\right)^2} = \frac{\widehat{\sigma}_I^2}{(\widehat{\sigma}_{Avg})^2}$$

7.6.1.3 A surface of implied correlation

We have seen in Section (1.4) that the implied volatility (IV) of a single asset was far from being flat, and experienced skew and curvature at the pair (K, T). Similarly, for a fixed maturity, the prices of basket options vary with different strikes leading to skew and curvature. We have also seen that the Implied Correlation in Equation (7.6.22), for a fixed maturity T, was computed with ATM volatilities. To account for the observed smiles, some pratitioners assumed they could compute an Implied Correlation for the pair (K, T) made of the IV of the basket option and that of its constituents calculated at (K, T). Thus, they obtained a surface of IC, where for a fixed maturity T, the curve of the IC plotted against the strikes is called the Correlation Skew (see Bouzoubaa et al. [2010]). For example, given the price of the 90% strike basket option and that of the 100% strike, we can compute the $90 - 100$ correlation skew. This is a complete nonsense because the BS-formula can only be linearised at-the-money forward, such that Equation (7.6.21) can not be applied for another strike. For a fixed strike K, the curve obtained is a correlation term structure. As a consequence of the existence of a pseudo surface of implied correlation, practitioners developed models accounting for implied correlation skew (ICS).

Alternatively, we can define the implied correlation with the help of Markov Functional modelling. In that setting, the price of a call option is given by

$$C(t, S_t; T, K) = E[\big(F(t, T)e^{\phi(W_{T-t}, T-t)} - K\big)^+ | \mathcal{F}_t]$$

where $F(t, T)$ is the forward price seen at time t. The function ϕ is estimated in terms of the implied volatility $\Sigma(K, T)$ as follow

$$\phi^{-1}(x, \tau) = \sqrt{\tau} N^{-1}\Big(N(-d_2) + F(t, T)n(d_1)\sqrt{\tau}\partial_K \Sigma(F(t, T)e^x, T - t)\Big)$$

where $\tau = T - t$ and x is the log forward moneyness. It depends on the market skew $\partial_K \Sigma(K, T - t)$. In a multi-asset framework we define the Multidimensional Gaussian Markov Functional (MGMF) as the Markov Functionals together with Gaussian copula to capture the dependence of the marginal distributions. We can therefore define the implied correlation as follow

Definition 7.6.1 *The implied correlation is the (Gaussian) corelation number we plug in the MGMF to recover the market price.*

7.6.2 Some market practice

7.6.2.1 Defining dispersion trading

Dispersion trading (DT) consists in selling the index option and buying options on the index components, or, buying the index option and selling options on the index components. Since all trades are delta-neutral, dispersion trading

- sells index volatility and buys volatility on the index components
- buys index volatility and sells volatility on the index components

Volatility dispersion trading is essentially a hedged strategy designed at taking advantage of relative value differences in implied volatilities between an index and a basket of component stocks. The idea being to profit from price differences in volatility markets using index options and options on individual stocks due to temporary shifts in correlations between assets, idiosyncratic news on individual stocks, etc. It is achieved by looking for a high degree of dispersion. Typical strategies involve short option position on an index, against which long option positions are taken on a set of components of the index. There are multiple instruments available to implement dispersion trading. Since DT is a non-directional strategy exposed to the volatility of the underlying components, when using vanilla options it is necessary to delta-hedge each option. Straddles and strangles (see Section (1.5.2)) are two basic volatility spreads suitable for these requirements since they have volatility exposure with limited delta. Thus, it is market practice to devise straddles and strangles strategies when building dispersion exposure. It is common to consider a short position of a straddle, or near-ATM strangle, on the index and long positions of straddles or strangles on 30% to 40% of the stocks that make up the index. This is because at-the-money straddle, or out-of-the-money strangle positions, have delta exposures very close to zero. Hence, a dispersion strategy that buys index straddles/strangles and sells straddle/strangle positions on individual components is hedged against large market movement and has low volatility risk. If maximum dispersion is realised, the strategy will make money on the long options on the individual stocks and will lose very little on the short option position on the index, since the latter would have moved very little. The strategy is typically a very low-premium strategy, with very low initial Delta and typically a small net long vega. Rather than using vanilla options to build dispersion trading, Ganatra [2004] considered using variance swaps. However, this approach is difficult to implement in period of low liquidity for single stock variance swaps. Marshall [2008] studied the performance of dispersion trading on the $S\&P$ 500 index and underlying stocks by using implied volatility obtained from synthetic VIX-indexes calculated on instruments weighted to 30-days time period.

7.6.2.2 Empirical evidence

Since empirical evidences suggest that index options, especially index puts, are more expensive than their theoretical Black-Scholes prices (see Bakshi et al. [2003a] and [2003b]), different explanations regarding the expensiveness of index versus individual options flourished. Several hypotheses have been put forward in the literature to explain the source of the profitability of dispersion strategy amoung which are

1. the risk-based hypothesis
2. the market inefficiency

The former argues that the index options are more expensive relative to individual stock options because they bear some risk premium, such as volatility risk and correlation risk, that are absent from individual stock options. Driessen et al. [2005] argued that index options have a risk premium which is absent in stock options because the former hedge correlation risk, which is pronounced in index option puts. The latter argues that options market demand and supply drive option premiums to deviate from their theoretical values. Garleanu et al. [2009] developed a model showing that option premium increase with market demand. They showed that end investors are net short single stock options and net long index options, translating into a net demand for index options and a net supply of single stock options. Taking advantage of market structural changes during late 1999 and 2000, which reduced the costs of arbitraging any differential pricing of individual equity and index options via dispersion trading, Deng [2008] found that the dispersion strategy was quite profitable through the year 2000, after which it disappeared. These results provided support for the market inefficiency hypothesis against the risk-based explanation. Marshall [2008] developed empirical evidence for dispersion trading in the US by showing that $S\&P$ 500 index-option-implied volatility (both for calls and puts) tended to exceed Markowitz implied volatility (MIV), and assumed that the difference came from the correlation risk premium. This is because correlation tends to rise in stressful times so that a seller of the option requires extra premium to hedge this risk. Lozovaia et al. [2005] explained that the best timing to execute the direct strategy (short index, long constituents) was when implied index volatility exceeded the realised (or historical) one and when the implied index correlation was close to its maximum registered value, since the strategy works better if implied volatility of the index is highly correlated with implied volatilities of its components. Lisauskas [2010] found that the difference between index option implied volatility and the theoretical implied volatility was largely dependent on the time period used to calculate the correlation between stocks.

7.6.2.3 Market risks

When building a dispersion trade, practitioners must decide which stocks to pick and how to weight the index. While broad index ($S\&P$ 500, Nasdaq) have many constituents, so that one must buy many component options and manage them, narrow index (industry groups) have few component options, making them an easier trade to execute and risk manage. This is because we can not hedge the entire index, but only a selected group of stocks. In general, we consider

- the heavily weighted stocks - large market
- the ones with large liquidity
- the ones with large implied volatility numbers

We must then decide on the number of stocks selected. As it is not possible to place all the legs of this trade simultaneously, the trade is exposed to lag risk. Dispersion trading being a delta-neutral strategy, it is not affected by directional price moves in the underlying asset. Since both straddle and strangle options are almost delta-neutral when they are ATM, or perfectly centred around ATM for strangle, they are good candidates for DT. However, as the stocks and index move, these options may no longer be close to ATM and delta exposure will arise, leading to possible readjustment of the options position. It is also necessary to delta hedge the portfolio of options in order to realise the actual volatility for which straddles and/or strangles are bought cheaper and sold more expensive against. In general, dynamic delta hedging is necessary for dispersion trading since no other suitable instruments exist with cash flows equal to that generated from the delta exposure to be hedged. Since the dispersion trade is a bet that implied correlation will return to normal levels, if it continues to increase (due to all stocks moving in the same direction) the trade may lose money.
Further, with many stocks in the index, some options on its components may not exist, or be illiquid, such that we can not uniquely determine an implied correlation matrix. One approach consists in using the historical correlation and the implied volatility and compare the resulting volatility with the actual implied volatility of the index. This method implies the use of some tools to assess the relative implied volatility differential. Moreover, as the stocks move up and down, their relative weightings in the index change, impacting on the index implied volatility. If a stock moves down dramatically, then its weight in the index drops, but its implied volatility increases. However, these effects tend to partially cancel each other.
All these risks can be attenuated by carefully choosing the weights of the portfolio. We distinguish three main weighting strategies

- Vega-hedging weighting: the dispersion is built such that the vega of the index equals the sum of the vega of the constituents. This way the trader will be immune against short moves in volatility.

- Gamma-hedging weighting: the dispersion is built such that the gamma of the index is worth the sum of the gammas of the constituents. This way the trader will be immune against large move in the stocks.
- Theta-hedging weighting: this strategy results in both a short vega and a short gamma position.

7.6.3 Modelling dispersion trading

The only way of relating the market of basket options with that of single stock options is to use a theoretical model, consistent with the absence of arbitrage opportunities (AAO), to price the single stock options and the basket option. While dispersion trading is about replicating a basket option with options on its components, we saw in Section (7.5.2) that we can only bound the price of a basket option with a combination of Black-Scholes prices. Different methods exists, all depending on the way to select the strikes of the single options (see El Karoui et al. [1998], Deelstra et al. [2002]). As we can only get bounds to the price of a basket option, when focussing on the upper bound, its replication becomes a super-replication. Once the super-replicating portfolio has been constructed, we must delta-hedge the constituting options to relalise the volatility differential between the two markets and make a profit. The dynamics of these options being constrained by a relation of no-arbitrage both in time and space, we need a computationally fast and robust model to risk manage the trade over time. We are first going to define the problem and derive a simple upper bound in the special case where the options are at-the-money forward. We will then use the MixVol model to compute the super-replicating portfolio.

7.6.3.1 Defining the problem

By properly defining the relationship between an option on a basket and a weighted sum of options on its components, we can assess the difficulties of implementing dispersion trading and explain the assumptions and approximations made by practitioners for using Markowitz's portfolio theory.
In Section (7.5.2.1) we defined a basket as $B(t,T) = \sum_{i=1}^N \omega_i F_i(t,T)$, computed its associated instantaneous variance $V_B(t)$, and introduced the pricing of a basket option. In equity, the instantaneous variance of a basket is used as fair value relation for volatilities, assuming a given correlation matrix. We must therefore project this volatility relation into single option prices. That is, setting the strike on the index option as

$$K = \sum_{i=1}^N \omega_i K_i$$

and applying Jensen's inequality, we get the following relation on the intrinsic value

$$\max\left(B(T,T) - K, 0\right) \leq \sum_{i=1}^N \omega_i \max\left(F_i(T,T) - K_i, 0\right) \tag{7.6.25}$$

which states that the premium from an option on the index is less or equal to that from its components. Taking the expectation of the payoff under the risk-neutral measure and discounting with a zero-coupon bond, we can rewrite this relation in terms of prices as

$$C_B(t, B(t,T), K, T) \leq \sum_{i=1}^N \omega_i C_i(t, F_i(t,T), K_i, T) \tag{7.6.26}$$

Hence, we see that an option on a basket and a weighted sum of options on its components are two different products, the latter being an upper bound to the former. As a result, the Equation (7.6.26) defines a no-arbitrage condition beteween an option on a basket and a weighted sum of options on its components.

We know that a simple way of understanding the properities of linear combination of European option prices in the Black-Scholes world is to assume options are at-the-money (ATM) forward and to linearise them (see Equation

(1.2.12)). In that setting, given $K = B(t,T)$ and $K_i = F_i(t,T)$, the implied volatility of each European option no-longer depend on the strike, and the price relation above simplifies to

$$B(t,T)e^{-q(T-t)}\sigma_B\sqrt{T-t} \leq \sum_{i=1}^N \omega_i F_i(t,T)e^{-q_i(T-t)}\sigma_{F,i}\sqrt{T-t}$$

where σ_B is the ATM implied volatility of an option on a basket and $\sigma_{F,i}$ is the ATM volatility of its ith component. From the definition of the stochastic weight $\widehat{\omega}_i(t)$ in Equation (7.3.10), and assuming zero dividend yield, we can rewrite the relation as

$$\sigma_B \leq \sum_{i=1}^N \widehat{\omega}_i(t)\sigma_{F,i}$$

Note, $\sigma_{F,i}$ is the implied volatility of the ith component, that is, the constant which is plugged in the BS-formula to recover the market option price. Even though there are instantaneous correlations in the portfolio's theory, there is no explicit notion of implied correlation in the option pricing theory. At most, the instantaneous correlations of the Brownian motions in the dynamics of the basket have been implicitely embedded in the BS-volatility σ_B. Further, since an option on a basket depends on the stochastic weights $\widehat{\omega}_i(t)$, $i = 1, .., N$, the relation between an option on a basket and a weighted sum of options on its components expressed in terms of implied volatilities also depends on these stochastic weights.

Assuming that the stochastic weights $\widehat{\omega}_i(t)$ for $i = 1, .., N$ are stable, we can approximate them with their time zero values, getting $\overline{\omega}_i$, $i = 1, .., N$. We can then approximate the no-arbitrage relation for ATM implied volatility of an option on a basket as

$$\sigma_B \leq \sum_{i=1}^N \overline{\omega}_i \sigma_{F,i} = \sigma_{WB}$$

where $\sigma_{WB} = \sum_{i=1}^N \overline{\omega}_i \sigma_{F,i}$ is the approximated average implied volatility of the basket components. Since volatilities and weights are positive, we can square both sides, getting

$$\sigma_B^2 \leq \big(\sum_{i=1}^N \overline{\omega}_i \sigma_{F,i}\big)^2 = \sum_{i=1}^N (\overline{\omega}_i)^2 \sigma_{F,i}^2 + \sum_{i \neq j}^N \overline{\omega}_i \overline{\omega}_j \sigma_{F,i}\sigma_{F,j}$$

We can therefore define an ATM indicator of arbitrage between an option on a basket and a weighted sum of options on its components as follow

$$M_\rho = \frac{\sigma_B^2}{\sigma_{WB}^2} \leq 1 \tag{7.6.27}$$

If M_ρ is close to 1, then the implied volatility of a basket option, σ_B, is very expensive, and it is likely to move down. Further, if this indicator is greater than one, there is an arbitrage and we should sell the option on the basket. However, this indicator does not represent the market's expectation of the future realised correlation. Only a correlation swap can be seen as such. Note, the indicator M_ρ is similar to Bossu's indicator ρ^*_{imp} in Equation (7.6.23), but with different weights.

Remark 7.6.3 *One can only relate Markowitz's portfolio theory to option pricing theory when the options have been linearised and the stochastic weights have been frozen.*

7.6.3.2 Approximating the basket option implied volatility

We are going to show that due to the non-linearity of the BS-formula with respect to volatility, expressing the implied volatility $\Sigma_B(K,T)$ of an option on a basket in terms of implied volatilities $\Sigma_{F_i}(K_i,T)$ of its components is not an easy task.
For example, to generate the smile at maturity T, we decompose the market price as in Equation (3.0.1) with a BS-formula and a convex function $\alpha(K,T)$ producing the skew, or curvatutre. The model is expressed as

$$C(t,S_t;K,T;I(t,T)) = C_{BS}(t,S_t,K,T;I_0(t,T)) + \alpha(K,T)$$

where $C_{BS}(t,x,K,T;\sigma)$ is the Black-Scholes call price with volatility σ, and $I_0^2(t,T) = E_t[I^2(t,T)]$ is the expected mean variance. Assuming $\alpha(K,T)$ to be small, we Taylor expanded the implied volatility of the market price in Appendix (B.3.1.1), obtaining the approximated implied volatility in Equation (B.3.14). Then, assuming that the market price could be decomposed into the linear combination

$$C(t,x,K,T) = \sum_{i=1}^n a_i \big(C_{BS}(t,x,K,T;\Sigma_i(K,T)) + \alpha_i(K,T) \big)$$

where the weights $a_i > 0$ for $i = 1,...,n$ are positive and such that $\sum_{i=1}^n a_i = 1$. In that setting, we obtained the approximated volatility in Equation (B.3.16) with the first few terms given by

$$\Sigma(K,T) \approx \widehat{\sigma}(K,T) + \frac{1}{vega(\widehat{\sigma})} \big(\sum_{i=1}^n a_i \alpha_i \big) - ...$$

where $\widehat{\sigma}$ is the implied volatility of a weighted sum of lognormal prices generating a symmetric smile approximated by

$$\widehat{\sigma}(K,T) \approx \overline{\sigma} + \frac{1}{vega(\overline{\sigma})} \big(\sum_{i=1}^n a_i C_{BS}(t,x,K,T;\sigma_i) - C_{BS}(t_0,t;\overline{\sigma}) \big) - ...$$

and $\overline{\sigma} = \sum_{i=1}^n a_i \sigma_i$.

Remark 7.6.4 *For a fixed maturity, the weighted sum of BS-formula with constant volatility σ_i generates a symmetric smile centered around $\overline{\sigma}$.*

In the special case where the strike is ATM-forward, the approximate implied volatility simplifies to (see Equation (B.3.17))

$$\Sigma(K,T)\big|_{K=F(T)} \approx \overline{\sigma} + \frac{1}{vega(\overline{\sigma})} \big(\sum_{i=1}^n a_i \alpha_i(K,T) \big)$$

Going further and assuming that the IVS of each constituent of the basket option is flat, we set $\alpha_i(K,T) = 0, \forall i$, and we get

$$\Sigma(K,T)\big|_{K=F(T)} \approx \overline{\sigma}$$

which gives an idea of the over-simplification made by pratitioners when using Markowitz's portfolio theory with ATM volatilities.

Remark 7.6.5 *Linearising the BS-formula to express the implied volatility of a basket option in terms of that of its constituents, via Markowitz's equation, only applies around the ATM-forward strike. We see that the notion of Correlation Skew is meaningless.*

7.6.3.3 Replicating basket option with the MixVol model

When dealing with dispersion trading, we must be concerned by the ability of controlling the risks involved. As the basket components move, we must be able to price each individual option and compute its Greeks. Assuming that we could replicate the price of a basket option with a linear combination of single option prices in the sense of Equation (7.6.26), we could use the Black-Scholes formula to compute the option price of each component of the replicating portfolio. However, the BS-formula is based on the lognormal assumption, do not relate prices with different strikes, and do not accounts for the dynamics of the volatility surface. Thus, we need a pricing model capable of relating prices with different strikes and evolving them in time in an arbitrage-free way. We can use the MixVol model defined in Section (6.2) to relate single option prices with different strikes, and compute their Greeks. In that setting, the super-replicating portfolio $\widehat{C}_B(t_0, K, t)$ is given by

$$\widehat{C}_B(t_0, K, t) = \sum_{i=1}^N \omega_i \Big(\sum_{j=1}^n a_{ij}(t) Call_{BS}(t_0, F_i(0), R_i(t), P_i(t), \overline{K}(K, t), t, \Sigma_{ij}(t)) \Big)$$

where $a_{ij}(t)$ for $j = 1, .., n$ are the weights for the ith underlying, $\Sigma_{ij}(t)$ is the volatility for the ith underlying. Given the strike K_i, we get $\overline{K}(K, t) = K^{'}(K, t)(1 + \mu_{ij}(t))$ with $K^{'}(K, t) = K_i + D_t$. We calibrate independently the MixVol model's parameters $a_{ij}(t)$, μ_{ij}, $\Sigma_{ij}(t)$ to the liquid market prices of each underlying $F_i(0)$ for $i = 1, .., N$. From linearity of the model, we can interchange the summation operators and rewrite the super-replication portfolio as

$$\widehat{C}_B(t_0, K, t) = \sum_{j=1}^n \Big(\sum_{i=1}^N \omega_{ij}(t) Call_{BS}(t_0, F_i(0), R_i(t), P_i(t), \overline{K}(K, t), t, \Sigma_{ij}(t)) \Big)$$

where $\omega_{ij}(t) = \omega_i a_{ij}(t)$. Since that model is a weighted sum of modified BS-formula, we can easily compute the Greeks of the super-replicating portfolio. Differentiating with respect to the initial stock price, the delta becomes

$$\Delta_{\widehat{C}_B}(t_0, K, t) = \sum_{j=1}^n \Big(\sum_{i=1}^N \omega_{ij}(t) \frac{\partial}{\partial F_i(0)} Call_{BS}(t_0, F_i(0), R_i(t), P_i(t), \overline{K}(K, t), t, \Sigma_{ij}(t)) \Big)$$

Differentiating one more time with respect to the initial stock price, the gamma is given by

$$\Gamma_{\widehat{C}_B}(t_0, K, t) = \sum_{j=1}^n \Big(\sum_{i=1}^N \omega_{ij}(t) \frac{\partial^2}{\partial F_i^2(0)} Call_{BS}(t_0, F_i(0), R_i(t), P_i(t), \overline{K}(K, t), t, \Sigma_{ij}(t)) \Big)$$

and differentiating with respect to the BS-volatility, the vega is given by

$$Vega_{\widehat{C}_B}(t_0, K, t) = \sum_{j=1}^n \Big(\sum_{i=1}^N \omega_{ij}(t) \frac{\partial}{\partial \Sigma_{ij}} Call_{BS}(t_0, F_i(0), R_i(t), P_i(t), \overline{K}(K, t), t, \Sigma_{ij}(t)) \Big)$$

We can use the approximation to the implied volatility generated by the MiVol model given in Equation (6.2.9) to get a feel of what the implied volatility of the replicating portfolio looks like. Assuming first order expansion, we get

$$\Sigma_{\widehat{B}}(K, t) \approx \sum_{i=1}^N \omega_i \Sigma_i(K^{'}(K, t), t)$$

such that

$$\Sigma_{\hat{B}}(K,t) \approx \tag{7.6.28}$$

$$\sum_{i=1}^N \omega_i \widehat{\sigma}_i + \sum_{i=1}^N \omega_i \frac{1}{vega(\widehat{\sigma}_i)} \alpha_i(t_0, S_{t_0}, K^{'}(K,t),t) - \frac{1}{2}\sum_{i=1}^N \omega_i \frac{d_1(\widehat{\sigma}_i)d_2(\widehat{\sigma})}{\widehat{\sigma}_i} \frac{1}{[vega(\widehat{\sigma}_i)]^2} \alpha_i^2(t_0, S_{t_0}, K^{'}(K,t),t)$$

$$+ \quad \frac{1}{6}\sum_{i=1}^N \omega_i \frac{d_1(\widehat{\sigma}_i)d_2(\widehat{\sigma}_i) + d_2^2(\widehat{\sigma}_i) + d_1^2(\widehat{\sigma}_i) + 2d_1^2(\widehat{\sigma}_i)d_2^2(\widehat{\sigma}_i)}{\widehat{\sigma}_i^2} \frac{1}{[vega(\widehat{\sigma}_i)]^3} \alpha_i^3(t_0, S_{t_0}, K^{'}(K,t),t) + ... + \tag{7.6.29}$$

where

$$\begin{aligned}
\widehat{\sigma}_i &= \frac{1}{norm_i} \sum_{j=1}^n \overline{a}_{ij}(t) \Sigma_{ij}(t) \\
\alpha_i(t_0, S_{t_0}, K^{'}(K,t),t) &= \frac{1}{norm_i} \sum_{j=1}^n \overline{a}_{ij}(t) \alpha_{ij}(K^{'}(K,t),t;\Sigma_{ij}(t))
\end{aligned}$$

7.6.4 Incorrect correlation analysis

When the dispersion strategy consists in long positions on individual volatilities and short position on index volatility, then profits are made when the realised volatilities of individual stocks become high and the realised volatility of the index become low. It will happen when the realised average correlation turns out to be lower than the implied one. Hence, this strategy is exposed to the variation of correlation between individual component stocks. It is argued that the implied IC can be greater than 1, since the individual stock options and the index options markets are actually separate markets. The only fact we can state, is that if the indicator of implied correlation, M_ρ, is greater than 1 then there is an arbitrage. Lozovaia et al. [2005] claimed that an implied IC greater than one implies that actual implied volatility of the index, σ_I, substantially exceeds the theoretical volatility σ_P calculated from the formula of the portfolio risk (theoretically index options are too overpriced). That is , for $\rho_{imp} > 1$ we get

$$\sigma_I^2 > \sum_{i=1}^N w_i^2 \sigma_i^2 + 2\sum_{i=1}^N \sum_{j>i} w_i w_j \sigma_i \sigma_j > \sum_{i=1}^N w_i^2 \sigma_i^2 + 2\sum_{i=1}^N \sum_{j>i} w_i w_j \sigma_i \sigma_j \rho_{ij} = \sigma_P^2$$

However, as explained in Section (7.6.3) we can not compute the theoretical volatility σ_P with implied volatilities since that relation does not hold.
One method for calculating an index volatility is to consider it as a weighted sum of volatilities of its components $\sigma_{WC}^2 = \left(\sum_{i=1}^N w_i \sigma_i\right)^2$, which is called weighted components implied volatility, or the weighted volatility of index. That is, it expresses overall implied volatility of the index components, but ignores correlation between component stocks. The ratio of the components implied volatility to the actual implied index volatility, $\frac{\sigma_{WC}}{\sigma_I}$, is called a first volatility level coefficient. Given

$$\sigma_I^2 = \sum_{i=1}^N w_i^2 \sigma_i^2 + 2\rho_{imp} \sum_{i=1}^N \sum_{j>i} w_i w_j \sigma_i \sigma_j$$

and since $\sigma_{WC}^2 = \left(\sum_{i=1}^N w_i \sigma_i\right)^2 = \sum_{i=1}^N w_i^2 \sigma_i^2 + \sum_{i \neq j} w_i w_j \sigma_i \sigma_j$, we get

$$\sigma_I^2 = \sigma_{WC}^2 + 2(\rho_{imp} - 1) \sum_{i=1}^N \sum_{j>i} w_i w_j \sigma_i \sigma_j$$

so that for $\rho_{imp} = 1$ the actual implied volatility of index is exactly the weighted sum of the individual stock constituent implied volatilities. When $\rho_{imp} < 1$ then $\sigma_{WC} > \sigma_I$ and the first volatility level coefficient is greater than 1, and when $\rho_{imp} > 1$ then $\sigma_{WC} < \sigma_I$ and the first volatility level coefficient is less than 1. Hence, the main factor affecting the first volatility level coefficient is the changing correlation between stock and index volatility. Note, this interpretation is eroneous and leads to misleading conclusions.
When the index is defined as a portfolio of component stocks with the corresponding weights, the risk of the index is given by the risk of portfolio σ_P (volatility of the portfolio). Replacing the volatility of each component σ_i with the implied volatility index of a component stock $\sigma_{I,i}$, we obtain the theoretical correlated implied variance of an index calculated on a correlation adjusted basis

$$\sigma_{PI}^2 = \sum_{i=1}^N w_i^2 \sigma_{I,i}^2 + 2 \sum_{i=1}^N \sum_{j>i} w_i w_j \sigma_{I,i} \sigma_{I,j} \rho_{I,ij}$$

where $\rho_{I,ij}$ is the correlation between implied volatility indexes of two stocks. The ratio of theoretical correlated implied volatility of the index to the actual implied volatility $\frac{\sigma_{PI}}{\sigma_I}$ is calculated to estimate the difference between theoretical and real prices of index options. It is called a second volatility level coefficient. One can show that the theoretical correlated implied volatility of an index can not be greater than the weighted implied volatility, which ignores correlations. As a result, the second volatility level coefficient is always less or equal to first coefficient level. Further, if the second volatility level is less than one, and relatively low, then theoretically, index options are too overpriced and it is profitable to sell index options. Conversely, if the second volatility level coefficient is greater than one, and relatively high, then theoretically, index options are conservatively priced and it is profitable to buy them.
The theoretical historical volatility of index can be calculated from the portfolio volatility on the basis of historical volatilities of each component and correlations between stock prices. The ratio of theoretical historical volatility of index to actual volatility is called the third volatility level coefficient and it indicates how much theoretical historical volatility differs from actual volatility. If the third coefficient is less than 1 and low, then theoretical performance of the index is less then actual volatility and we can generate profit by selling index options.

7.6.5 Some strategies

7.6.5.1 A simple strategy

A simple strategy consists in selling a portfolio of near-ATM (moneyness varying from 0.85 to 1.15) straddles on an index for a given maturity T, and buying a portfolio of near-ATM straddles on its component stocks for the same maturity. This strategy is hold until the expiration date and rolled over on the first trading day following the expiration date of the options (say one month). Letting t be the investment date, the payoff $H_L(t,T)$ for the long side is given by

$$H_L(t,T) = \sum_{i=1}^n \nu_{i,t} |S_{i,T} - K_{i,t}|$$

where $S_{i,T}$ is the ith stock price at maturity T, $K_{i,t}$ is the ith strike price, and $\nu_{i,t}$ is the number of individual straddles traded at t. The payoff from the short side of the straddle is

$$H_S(t,T) = |S_{I,T} - K_{I,t}|$$

where $S_{I,T}$ is the index level at expiration and $K_{I,t}$ the index option strike price. In that setting the number of individual straddles traded at t is

$$\nu_{i,t} = \frac{N_{i,t} S_{I,t}}{\sum_{i=1}^n N_{i,t} S_{i,t}}$$

where $N_{i,t}$ is the number of shares outstanding of stock i. Since $S_{I,t} = \sum_{i=1}^n \nu_{i,t} S_{i,t}$, we choose $\nu_{I,t}$ such that the payoff of the index straddle is matched as closely as possible to the total payoff of the individual straddles. Hence, the strategy is protected by construction against large stock market movements. The return of the strategy over the risk-free rate is given by

$$R(t,T)\begin{cases}\frac{V_T-V_t}{V_t}-\frac{1}{P(t,T)} \text{ if } V_t\geq 0\\ -\frac{V_T-V_t}{V_t}+\frac{1}{P(t,T)} \text{ if } V_t<0\end{cases}$$

where $V_T = H_L(t,T) - H_S(t,T)$ is the payoff from the portfolio at expiration, and

$$V_t = \sum_{i=1}^{n}\nu_{i,t}\big(C_i(t,T,K_i)+P_i(t,T,K_i)\big)-\big(C_I(t,T,K_I)+P_I(t,T,K_I)\big)$$

is the initial price paid for the portfolio. Since the index options are European-style and the individual options are American-style, assuming the option portfolio is hold till expiration may underestimate the resulting returns since we are selling index options and buying individual options. The rate of return net of transaction costs is given by

$$R_{net}(t,T)=R(t,T)-\frac{\delta_t}{V_t}$$

where δ_t is the transaction costs at time t (half the bid-ask spread).

7.6.5.2 A delta hedged strategy

Even though the simple dispersion strategy initially has very low delta exposure, as the prices of the underlying stocks change, the deltas of the straddle positions will also change, leading to higher exposure to delta risk. One way forward is to delta hedge the individual stock options with the underlying stock. Following the simple strategy, the long leg of the dispersion trade has a delta exposure to stock i given by

$$\Delta_{L,i,t}=\Delta_{C,i,t}+\Delta_{P,i,t}$$

where $\Delta_{C,i,t}$ and $\Delta_{P,i,t}$ are the BS-delta for a call and a put option on the ith stock, respectively. The short leg has a delta exposure to stock i given by

$$\Delta_{S,i,t}=\nu_{i,t}\big(\Delta_{C,I,t}+\Delta_{P,I,t}\big)$$

where $\Delta_{C,I,t}$ and $\Delta_{P,I,t}$ are the BS-delta for a call and a put option on the index, respectively. The deltas are computed at the market close each day using closing stock prices and index level. The delta exposure of the dispersion strategy (DS) to stock i satisfies

$$\Delta_{DS,i,t}=\Delta_{L,i,t}-\Delta_{S,i,t}$$

which is rebalanced everyday such that the trade keeps delta-neutral until the expiration date. The return of the delta-hedged disperion trading over the life of the options is given by

$$R(t,T)=\frac{V_T-V_t-\sum_{s=t}^{T-1}\sum_{i=1}^{n}\nu_{i,s}\Delta_{DS,i,s}\big(S_{i,s+1}+D_{s,s}-S_{i,s}\big)P^{-1}(s,T)}{V_t-\sum_{i=1}^{n}\nu_{i,t}\Delta_{DS,i,t}\big(S_{i,t}-P^{-1}(t,T)\big)}$$

and the net return is defined as above. Note, while both these studies estimates the efficiency of dispersion trading as an every day rolled strategy, an alternative is to wait for profitable opportunities which are most likely caused by market inefficiency.

7.6.5.3 The intrinsic value hedge

To solve the problem defined in Section (7.6.3.1), Avellaneda [2002] considered the intrinsic value hedge (IVH) which uses index weights for option hedge. We now need to estimate the strike K_i of each option. In the Black-Scholes world, the value of the basket at time t is given by

$$B(t,T) = \sum_{i=1}^{N} \omega_i F_i(t,T) = \sum_{i=1}^{N} \omega_i F_i(0,T) e^{-\frac{1}{2}\sigma_i^2 t + \sigma_i W_i(t)}$$

where $< dW_i, dW_j >= \rho_{ij} dt$. Assuming perfect correlation, $\rho_{ij} = 1$ for all i,j then all the stocks have the same Brownian motion. In that setting we know how to compute analytically a bond option in a 1-factor model (see El Karoui et al. [1989]). Applying the same method, we can solve for X in

$$K = \sum_{i=1}^{N} \omega_i F_i(0,T) e^{-\frac{1}{2}\sigma_i^2 t + \sigma_i X}$$

where $X = Y\sqrt{t}$ with $Y \sim N(0,1)$. Once we know X we recover the strike K_i as

$$K_i = F_i(0,T) e^{-\frac{1}{2}\sigma_i^2 t + \sigma_i X}$$

Under the assumption of perfect correlation, Equation (7.6.25) simplifies, and we get equality of the intrinsic values

$$\max\left(B(T,T) - K, 0\right) = \sum_{i=1}^{N} \omega_i \max\left(F_i(T,T) - K_i, 0\right), \forall T$$

In that setting, we can solve for Y, getting

$$Y = \frac{1}{\sigma_i \sqrt{t}} \ln \frac{K_i}{F_i(t,T)} + \frac{1}{2}\sigma_i \sqrt{t} = -d_2$$

The assumption of perfect correlation is clearly not satisfactory. Further, the whole notion of no-dominance principle has been taken out of the original problem.

7.7 Model risks for dispersion trading

The following risk analysis assume that we can use Markowitz Equation (7.6.21) to relate an option on the index to options on single stocks.

7.7.1 Delta-hedged dispersion PnL

Avellaneda [2002] proposed a method for estimating the delta-hedged dispersion profit and loss (PnL), which we are going to describe. We derived in Section (4.2.6) the PnL of a delta-hedged portfolio, for being short a single option and long the replicating portfolio, in a world of stochastic volatility. Given the profit and loss $\pi_V(t,t+dt)$ in Equation (4.2.17), and ignoring some derivatives of the call option with respect to volatility and stock price in the calculation of the PnL, it can be further approximated as

$$\pi_F(t,t+dt) \approx \Theta(n^2 - 1) + V_N \frac{d\sigma}{\sigma}$$

where Θ is the time decay, $n = \frac{\Delta F(t,T)}{F(t,T)\sigma\sqrt{\Delta t}}$ is the standardised move of the forward price, and $V_N = \sigma \frac{\partial C}{\partial \sigma}$ is the normalised vega. Following the same argument, we can compute the PnL for being short the index option.
In addition, assuming $d\sigma = 0$, the PnL of the basket simplifies to

$$\pi_B(t,t+dt) \approx \Theta_B(n_B^2 - 1)$$

where Θ_B is the time decay of the basket and $n_B = \frac{\Delta B(t,T)}{B(t,T)\sigma_B\sqrt{\Delta t}}$ is the standardised move of the basket.
From the definition of the instantaneous return of the basket in Equation (7.3.11), we can rewrite the standardised move of the basket price, n_B, as

$$n_B = \sum_{i=1}^N \widehat{\omega}_i(t) \frac{\Delta F_i(t,T)}{F_i(t,T)} \frac{1}{\sigma_B\sqrt{\Delta t}} = \sum_{i=1}^N \widehat{\omega}_i(t) \frac{\sigma_i}{\sigma_B} n_i$$

so that the profit and loss for the index option becomes

$$\pi_B(t,t+dt) \approx \Theta_B\big(\big(\sum_{i=1}^N \widehat{\omega}_i(t) \frac{\sigma_i}{\sigma_B} n_i\big)^2 - 1\big)$$

Further, we can decompose the standardised move of the basket as

$$\begin{aligned} n_B^2 &= \big(\sum_{i=1}^N \widehat{\omega}_i(t) \frac{\sigma_i}{\sigma_B} n_i\big)^2 \\ &= \sum_{i=1}^N \widehat{\omega}_i^2(t) \frac{\sigma_i^2}{\sigma_B^2} n_i^2 + \sum_{i \neq j} \widehat{\omega}_i(t)\widehat{\omega}_j(t) \frac{\sigma_i\sigma_j}{\sigma_B^2} n_i n_j \end{aligned}$$

and rewrite the PnL for the index option as

$$\pi_B(t,t+dt) \approx \Theta_B\Big[\big(\sum_{i=1}^N \widehat{\omega}_i^2(t) \frac{\sigma_i^2}{\sigma_B^2} n_i^2 + \sum_{i \neq j} \widehat{\omega}_i(t)\widehat{\omega}_j(t) \frac{\sigma_i\sigma_j}{\sigma_B^2} n_i n_j - 1\Big]$$

Assuming that we can use Markowitz Equation (7.6.21) to relate an option on the index to options on single stocks, we can replace the implied volatilities in the PnL equation with historical volatilies. The instantaneous variance of the basket in Equation (7.3.12) is

$$\sigma_B^2 = \sum_{i,j=1}^N \widehat{\omega}_i(t)\widehat{\omega}_j(t)\sigma_i\sigma_j\rho_{ij}$$

which can be decomposed as

$$\sigma_B^2 = \sum_{i=1}^N \widehat{\omega}_i^2(t)\sigma_i^2 + \sum_{i \neq j} \widehat{\omega}_i(t)\widehat{\omega}_j(t)\sigma_i\sigma_j\rho_{ij}$$

since $\rho_{ii} = 1$. Since we have $\frac{\sigma_B^2}{\sigma_B^2} = 1$, we can rewrite the PnL for the index option as

$$\pi_B(t,t+dt) \approx \Theta_B \sum_{i=1}^N \widehat{\omega}_i^2(t) \frac{\sigma_i^2}{\sigma_B^2} n_i^2 + \Theta_B \sum_{i \neq j} \widehat{\omega}_i(t)\widehat{\omega}_j(t) \frac{\sigma_i\sigma_j}{\sigma_B^2} n_i n_j - \Theta_B \sum_{i=1}^N \widehat{\omega}_i^2(t) \frac{\sigma_i^2}{\sigma_B^2} - \Theta_B \sum_{i \neq j} \widehat{\omega}_i(t)\widehat{\omega}_j(t) \frac{\sigma_i\sigma_j}{\sigma_B^2} \rho_{ij}$$

Combining terms together, the PnL of the basket becomes

$$\pi_B(t,t+dt) \approx \Theta_B \sum_{i=1}^N \widehat{\omega}_i^2(t)\frac{\sigma_i^2}{\sigma_B^2}(n_i^2-1) + \Theta_B \sum_{i \neq j} \widehat{\omega}_i(t)\widehat{\omega}_j(t)\frac{\sigma_i\sigma_j}{\sigma_B^2}(n_i n_j - \rho_{ij}) \quad (7.7.30)$$

Recall, in the dispersion trade we short the index option and buy the options on the stocks. Combining the PnL for all the options, where the PnL for being long the ith option is $\pi_{F_i} \approx -\Theta_i(n_i^2-1)$, with that of the index option, we get the Gamma PnL for the dispersion trade as

$$\begin{aligned}
\pi_D(t,t+dt) &\approx \sum_{i=1}^N \pi_{F_i}(t,t+dt) + \pi_B(t,t+dt) \\
&\approx \sum_{i=1}^N \big(-\Theta_i + \widehat{\omega}_i^2(t)\frac{\sigma_i^2}{\sigma_B^2}\Theta_B\big)(n_i^2-1) + \Theta_B \sum_{i \neq j} \widehat{\omega}_i(t)\widehat{\omega}_j(t)\frac{\sigma_i\sigma_j}{\sigma_B^2}(n_i n_j - \rho_{ij})
\end{aligned}$$

where the two terms correspond to

- diagonal term: realised single stock movements vs. implied volatilities
- off-diagonal term: realised cross-market movements vs. implied correlation

7.7.2 Problems with the correlation

A correlation swap pays at maturity the notional multiplied by the difference between the realised correlation $\widehat{\rho}$, given in Equation (7.6.24), and a strike K. The payoff is defined as

$$\Pi = \frac{\sum_{1 \leq i < j \leq n} w_i w_j \widehat{\rho}_{ij}}{\sum_{1 \leq i < j \leq n} w_i w_j} - K$$

Further, the realised correlation can be seen as the ratio between two traded products, through variance swaps, or variance dispersion trades. Based on this proxy, Bossu [2006] proved that

- the correlation swap can be dynamically quasi-replicated with a variance dispersion trade.
- the PnL of a variance dispersion trading is worth $\sum_{i=1}^n w_i \sigma_i^2(1-\widehat{\rho})$, where $\widehat{\rho}$ is the realised correlation.

The problem with dispersion trading is that the correlations between the stocks vary with the market volatility. One way of looking at it is to study the derivative of the variance of the index with respect to the correlation, that is,

$$\frac{\partial Var(I)}{\partial \rho_{ij}} = 2\sum_j \sum_{i<j} w_i w_j \sigma_i \sigma_j$$

which depends on the volatility of the i-th stock and the j-th stock. It is not a pure correlation product. On the contrary, a correlation swap with the payoff at maturity being $(\rho - K)$ plays directly with the realised correlation. It is a pure correlation product since its derivative with respect to the correlation is 1, that is $\frac{\partial RS(t,T)}{\partial \rho} = 1$.

In the context of dispersion, we assume that the correlations ρ_{ij} are equal to an average one, denoted ρ. Since this correlation makes the implied variance of the index and the implied variance of the weighted sum of the components equal, it represents the implied correlation (IC). We observe empirically a spread of a few basis points between the strike of a correlation swap and the dispersion implied correlation. That spread was studied more formally studied by Jacquier et al. [2007]. Following their arguments, we can express the relation between a dispersion trade in terms of variance swaps and a correlation swap as follow

$$\widehat{\sigma}_B^2 - \sigma_B^2 = \sum_{i=1}^N \widehat{\omega}_i^2(t)(\widehat{\sigma}_i^2 - \sigma_i^2) + \sum_{i \neq j} \widehat{\omega}_i(t)\widehat{\omega}_j(t)\big(\widehat{\sigma}_i\widehat{\sigma}_i\widehat{\rho} - \sigma_i\sigma_i\rho\big)$$

where σ_B is the implied volatility of a basket and $\widehat{\sigma}_B$ is its realised volatility. We can rewrite the difference as

$$\begin{aligned}
\widehat{\sigma}_B^2 - \sigma_B^2 &= \sum_{i=1}^N \widehat{\omega}_i^2(t)(\widehat{\sigma}_i^2 - \sigma_i^2) + \sum_{i \neq j} \widehat{\omega}_i(t)\widehat{\omega}_j(t)\big(\sigma_i\sigma_i(\widehat{\rho} - \rho) + (\widehat{\sigma}_i\widehat{\sigma}_i - \sigma_i\sigma_i)\widehat{\rho}\big) \\
&= \sum_{i=1}^N \widehat{\omega}_i^2(t)(\widehat{\sigma}_i^2 - \sigma_i^2) + \sum_{i \neq j} \widehat{\omega}_i(t)\widehat{\omega}_j(t)\sigma_i\sigma_i(\widehat{\rho} - \rho) + \sum_{i \neq j} \widehat{\rho}(\widehat{\sigma}_i\widehat{\sigma}_i - \sigma_i\sigma_i)
\end{aligned}$$

which leads to

$$(\widehat{\sigma}_B^2 - \sigma_B^2) - \sum_{i=1}^N \widehat{\omega}_i^2(t)(\widehat{\sigma}_i^2 - \sigma_i^2) - \sum_{i \neq j} \widehat{\omega}_i(t)\widehat{\omega}_j(t)\sigma_i\sigma_i(\widehat{\rho} - \rho) = \sum_{i \neq j} \widehat{\rho}(\widehat{\sigma}_i\widehat{\sigma}_i - \sigma_i\sigma_i)$$

It is the PnL for being short a dispersion trade through variance swap and long a correlation swap.

7.7.3 Gamma PnL of the dispersion trade

Jacquier et al. [2007] considered dispersion trades via variance swaps, shwowing correlation exposure, with implied correlation (IC) higher than the strike of a correlation swap. We are going to briefly describe their approach. Given the PnL of a basket option in Equation (7.7.30), and using the definition of theta in Equation (B.1.10) applied to the basket option as follow

$$\Theta_B \approx -\frac{1}{2}B^2(t,T)\sigma_B^2\Gamma_B(t, B(t,T))$$

we can then rewrite the PnL of the basket option with $d\sigma = 0$ in terms of the gamma as follow

$$\begin{aligned}
\pi_B(t, t+dt) &\approx -\frac{1}{2}B^2(t,T)\Gamma_B \sum_{i=1}^N \widehat{\omega}_i^2(t)\big[(\frac{\Delta F_i(t,T)}{F_i(t,T)})^2 - \sigma_i^2\Delta t\big] \\
&- \frac{1}{2}B^2(t,T)\Gamma_B \sum_{i \neq j} \widehat{\omega}_i(t)\widehat{\omega}_j(t)\sigma_i\sigma_j\big[\frac{\Delta F_i(t,T)}{F_i(t,T)\sigma_i}\frac{\Delta F_j(t,T)}{F_j(t,T)\sigma_j} - \rho_{ij}\Delta t\big]
\end{aligned}$$

We assume that the correlation ρ_{ij} between two components are all equal to an average correlation ρ, which is in fact the implied correlation in the dispersion trade. Hence, $\frac{\Delta F_i(t,T)}{F_i(t,T)}\frac{\Delta F_j(t,T)}{F_j(t,T)}$ is the instantaneous realised covariance between two underlyings $F_i(t,T)$ and $F_j(t,T)$. Therefore, $\frac{\Delta F_i(t,T)}{F_i(t,T)\sigma_i}\frac{\Delta F_j(t,T)}{F_j(t,T)\sigma_j}$ is the instantaneous realised correlation between the two forward prices. Thus, assuming the same implied and realised correlation for all the pairs of forward prices, $\widehat{\rho}$, and expending the stochastic weights $\widehat{\omega}_i(t)$, then the PnL for the index option becomes

$$\begin{aligned}
\pi_B(t, t+dt) &\approx -\frac{1}{2}\Gamma_B \sum_{i=1}^N \omega_i^2 F_i^2(t,T)\big[(\frac{\Delta F_i(t,T)}{F_i(t,T)})^2 - \sigma_i^2\Delta t\big] \\
&- \frac{1}{2}\Gamma_B \sum_{i \neq j} \omega_i\omega_j F_i^2(t,T)F_j^2(t,T)\sigma_i\sigma_j(\widehat{\rho} - \rho)\Delta t
\end{aligned}$$

That PnL is made of a spread between the implied and the realised correlation over the period of time $[t,t+dt]$ together with a volatility exposure.
We can also consider a position in a dispersion trade with variance swaps (α_i represents the proportion of variance swaps for the ith stock), the Gamma PnL becomes

$$\begin{aligned}
\pi_D^\Gamma(t,t+dt) &= \sum_{i=1}^N \alpha_i \pi_{F_i}(t,t+dt) - \pi_B(t,t+dt) \\
&\approx \frac{1}{2}\sum_{i=1}^N F_i^2(t,T)\big[(\frac{\Delta F_i(t,T)}{F_i(t,T)})^2 - \sigma_i^2 \Delta t\big]\big(\alpha_i \Gamma_i - \omega_i^2 \Gamma_B\big) \\
&- \frac{1}{2}\Gamma_B \sum_{i \neq j} \omega_i \omega_j F_i(t,T) F_j(t,T) \sigma_i \sigma_j (\widehat{\rho} - \rho)\Delta t
\end{aligned}$$

The dispersion trade PnL is equal to the sum of a spread between the implied and the realised correlation over the period of time $[t,t+dt]$ (pure correlation exposure) together with a volatility exposure. Setting $\Gamma_i = \frac{2}{TF_i^2(t,T)}$ and $\Gamma_B = \frac{2}{TB^2(t,T)}$, the Gamma PnL becomes

$$\begin{aligned}
\pi_D^\Gamma(t,t+dt) &\approx \frac{1}{T}\sum_{i=1}^N \big[(\frac{\Delta F_i(t,T)}{F_i(t,T)})^2 - \sigma_i^2 \Delta t\big]\big(\alpha_i - \omega_i^2 \frac{F_i^2(t,T)}{B^2(t,T)}\big) \\
&- \frac{1}{TB^2(t,T)} \sum_{i \neq j} \omega_i \omega_j F_i(t,T) F_j(t,T) \sigma_i \sigma_j (\widehat{\rho} - \rho)\Delta t
\end{aligned}$$

which we can rewrite as

$$\begin{aligned}
\pi_D^\Gamma(t,t+dt) &\approx \frac{1}{T}\sum_{i=1}^N \big[(\frac{\Delta F_i(t,T)}{F_i(t,T)})^2 - \sigma_i^2 \Delta t\big]\big(\alpha_i - \widehat{\omega}_i^2(t)\big) \\
&- \frac{1}{T} \sum_{i \neq j} \widehat{\omega}_i(t) \widehat{\omega}_j(t) \sigma_i \sigma_j (\widehat{\rho} - \rho)\Delta t
\end{aligned}$$

Setting $\beta_V = \frac{1}{T}\sum_{i \neq j} \widehat{\omega}_i(t)\widehat{\omega}_j(t)\sigma_i\sigma_j$, which only depends on the components of the index, we rewrite the PnL as

$$\pi_D^\Gamma(t,t+dt) \approx \frac{1}{T}\sum_{i=1}^N \big[(\frac{\Delta F_i(t,T)}{F_i(t,T)})^2 - \sigma_i^2 \Delta t\big]\big(\alpha_i - \widehat{\omega}_i^2(t)\big) - \beta_V (\widehat{\rho} - \rho)\Delta t$$

Choosing $\alpha_i = \widehat{\omega}_i^2(t)$, the Gamma PnL of the dispersion trade corresponds to the spread between the implied and the realised correlation, multiplied by the factor β_V

$$\pi_D^\Gamma(t,t+dt) = \sum_{i=1}^N \alpha_i \pi_{F_i}(t,t+dt) - \pi_B(t,t+dt) \approx -\beta_V (\widehat{\rho} - \rho)\Delta t$$

Thus, the dispersion trade is close to a pure correlation trade as it is multiplied by the quantity β_V, which is a weighted average variance of the components of the index.

Considering the delta-hedged PnL with stochastic volatility in Equation (4.2.14), we can add the volatility terms when $d\sigma \neq 0$, obtaining

$$
\begin{aligned}
\pi_D^{Vol}(t,t+dt) &= \sum_{i=1}^N \alpha_i \Big(\text{Vega(i)}\ d\sigma_{i,t} + \frac{1}{2}\ \text{Volga(i)}\ < d\sigma_i, d\sigma_i >_t + \text{Vanna(i)}\ \sigma_{i,t} S_{i,t} \xi_i \rho_{S_i,\sigma_i} dt \Big) \\
&- \Big(\text{Vega(B)}\ d\sigma_{B,t} + \frac{1}{2}\ \text{Volga(B)}\ < d\sigma_B, d\sigma_B >_t + \text{Vanna(B)}\ \sigma_{B,t} S_{B,t} \xi_B \rho_{B,\sigma_B} dt \Big)
\end{aligned}
$$

Thus, the total PnL of the dispersion trade becomes

$$
\pi_D(t,t+dt) = \pi_D^{\Gamma}(t,t+dt) + \pi_D^{Vol}(t,t+dt)
$$

Hence, the difference between the implied correlation of a dispersion trade and the strike of the correlation swap with the same characteristics is due to the volatility terms (vega, volga, and vanna).

7.7.4 Dispersion statistic

According to Avellaneda [2002], in order to define a measure of dispersion, we can define the dispersion statistic as

$$
D^2 = \sum_{i=1}^N \widehat{\omega}_i(t) \big(R_{F_i} - R_B \big)^2
$$

where $R_{F_i} = \frac{\Delta F_i(t,T)}{F_i(t,T)}$ and $R_B = \frac{\Delta B(t,T)}{B(t,T)}$, which is a proxy for the mean return of the basket. Since the standardised move of the ith stock is $n_i = \frac{\Delta F_i(t,T)}{F_i(t,T)\sigma_i \sqrt{\Delta t}}$, we can deduce that the return of the ith forward is

$$
R_{F_i} = \frac{\Delta F_i(t,T)}{F_i(t,T)} = n_i \sigma_i \sqrt{\Delta t}
$$

Further, since $\Delta B(t,T) = \sum_{i=1}^N \omega_i \Delta F_i(t,T)$, we can deduce that the return of the basket is

$$
R_B = \frac{\Delta B(t,T)}{B(t,T)} = \sum_{i=1}^N \widehat{\omega}_i(t) n_i \sigma_i \sqrt{\Delta t} = \sum_{i=1}^N \widehat{\omega}_i(t) R_{F_i}
$$

Hence, $R_{F_i}^2 = n_i^2 \sigma_i^2 \Delta t$ and

$$
\begin{aligned}
R_B^2 &= \big(\sum_{i=1}^N \widehat{\omega}_i(t) n_i \sigma_i \sqrt{\Delta t} \big)^2 \\
&= \sum_{i=1}^N \widehat{\omega}_i^2(t) n_i^2 \sigma_i^2 \Delta t + \sum_{i \neq j} \widehat{\omega}_i(t) \widehat{\omega}_j(t) \sigma_i \sigma_j n_i n_j \Delta t
\end{aligned}
$$

Now, the cross-term is

$$
R_{F_i} R_B = n_i \sigma_i \sqrt{\Delta t} \sum_{j=1}^N \widehat{\omega}_j(t) n_j \sigma_j \sqrt{\Delta t}
$$

which leads to two cases, $i = j$ and $i \neq j$, so that

$$
R_{F_i} R_B = \widehat{\omega}_i(t) n_i^2 \sigma_i^2 \Delta t + n_i \sigma_i \sum_{i \neq j} \widehat{\omega}_j(t) n_j \sigma_j \Delta t
$$

Putting terms together

$$R_{F_i}^2 + R_B^2 - 2R_{F_i}R_B = n_i^2\sigma_i^2\Delta t + \sum_{i=1}^N \widehat{\omega}_i^2(t)n_i^2\sigma_i^2\Delta t + \sum_{i\neq j}\widehat{\omega}_i(t)\widehat{\omega}_j(t)\sigma_i\sigma_j n_i n_j\Delta t$$
$$- \quad 2\widehat{\omega}_i(t)n_i^2\sigma_i^2\Delta t - 2n_i\sigma_i\sum_{i\neq j}\widehat{\omega}_j(t)n_j\sigma_j\Delta t$$

From the definition of the standardised move of the basket n_B^2, we have

$$\sigma_B^2 n_B^2 = \sum_{i=1}^N \widehat{\omega}_i^2(t)\sigma_i^2 n_i^2 + \sum_{i\neq j}\widehat{\omega}_i(t)\widehat{\omega}_j(t)\sigma_i\sigma_j n_i n_j$$

Replacing in the above equation, we get

$$R_{F_i}^2 + R_B^2 - 2R_{F_i}R_B = n_i^2\sigma_i^2\Delta t + \sigma_B^2 n_B^2\Delta t - 2\widehat{\omega}_i(t)n_i^2\sigma_i^2\Delta t - 2n_i\sigma_i\sum_{i\neq j}\widehat{\omega}_j(t)n_j\sigma_j\Delta t$$

Hence, we can rewrite the dispersion statistic as

$$D^2 = \sum_{i=1}^N \widehat{\omega}_i(t)n_i^2\sigma_i^2\Delta t + \sigma_B^2 n_B^2\Delta t - 2\sum_{i=1}^N \widehat{\omega}_i^2(t)n_i^2\sigma_i^2\Delta t - 2\sum_{i=1}^N \widehat{\omega}_i(t)n_i\sigma_i\sum_{i\neq j}\widehat{\omega}_j(t)n_j\sigma_j\Delta t$$

since $\sum_{i=1}^N \widehat{\omega}_i(t) = 1$. The statistic simplifies to

$$D^2 = \sum_{i=1}^N \widehat{\omega}_i(t)n_i^2\sigma_i^2\Delta t - \sigma_B^2 n_B^2\Delta t$$

7.7.5 Expressing the PnL in terms of the dispersion statistic

Starting from the definition of the PnL of the dispersion

$$\pi_D \approx \sum_{i=1}^N \pi_{F_i} - \pi_B = \sum_{i=1}^N \Theta_i(n_i^2 - 1) + \Theta_B(n_B^2 - 1)$$

we can rewrite it as

$$\pi_D \approx \sum_{i=1}^N \Theta_i n_i^2 + \Theta_B n_B^2 - \Theta_s$$

with $\Theta_s = \sum_{i=1}^N \Theta_i + \Theta_B$. Adding and substracting the quantity $\Theta_B\sum_{i=1}^N \widehat{\omega}_i(t)\frac{\sigma_i^2}{\sigma_B^2}n_i^2$, we get

$$\pi_D \approx \sum_{i=1}^N \Theta_i n_i^2 + \Theta_B\sum_{i=1}^N \widehat{\omega}_i(t)\frac{\sigma_i^2}{\sigma_B^2}n_i^2 - \Theta_B\sum_{i=1}^N \widehat{\omega}_i(t)\frac{\sigma_i^2}{\sigma_B^2}n_i^2 + \Theta_B n_B^2 - \Theta_s$$

and putting terms together, it simplifies to

$$\pi_D \approx \sum_{i=1}^N\big(\Theta_i + \Theta_B\widehat{\omega}_i(t)\frac{\sigma_i^2}{\sigma_B^2}\big)n_i^2 - \frac{\Theta_B}{\sigma_B^2}\big(\sum_{i=1}^N \widehat{\omega}_i(t)\sigma_i^2 n_i^2 - \sigma_B^2 n_B^2\big) - \Theta_s$$

Replacing with the dispersion statistic, the PnL of the dispersion becomes

$$\pi_D \approx \sum_{i=1}^N \big(\Theta_i + \Theta_B \widehat{\omega}_i(t) \frac{\sigma_i^2}{\sigma_B^2}\big) n_i^2 - \frac{\Theta_B}{\sigma_B^2} D^2 - \Theta_s$$

where the first term is the Idiosyncratic gamma, the second term is the Dispersion gamma, and the last term is the time-decay.

7.7.6 Some Greeks

We want to bump the volatility by a constant according to

$$\sigma_i \to \sigma_i + \Delta\sigma \,, \forall i$$

and compute $\frac{\sigma_{B,\sigma}^2 - \sigma_B^2}{\sigma_B^2}$. Replacing the single option variance in the basket instantaneous variance, we get

$$\sigma_{B,\sigma}^2 \to \sum_{i,j=1}^N \widehat{\omega}_i(t)\widehat{\omega}_j(t)(\sigma_i + \Delta\sigma)(\sigma_j + \Delta\sigma)\rho_{ij}$$

which gives

$$\sigma_{B,\sigma}^2 \to \sum_{i,j=1}^N \widehat{\omega}_i(t)\widehat{\omega}_j(t)\sigma_i\sigma_j\rho_{ij} + \sum_{i,j=1}^N \widehat{\omega}_i(t)\widehat{\omega}_j(t)(\sigma_i + \sigma_j)(\Delta\sigma)\rho_{ij} + \sum_{i,j=1}^N \widehat{\omega}_i(t)\widehat{\omega}_j(t)(\Delta\sigma)^2\rho_{ij}$$

The correlation ρ_{ij} between two assets appears in the instantaneous variance of the basket σ_B^2. The variance can be decomposed as

$$\sigma_B^2 = \sum_{i=1}^N \widehat{\omega}_i^2(t)\sigma_i^2 + \sum_{i \neq j} \widehat{\omega}_i(t)\widehat{\omega}_j(t)\sigma_i\sigma_j\rho_{ij}$$

We want to bump the correlations by a constant according to

$$\rho_{ij} \to \rho_{ij} + \Delta\rho \,, i \neq j$$

and compute $\frac{\sigma_{B,\rho}^2 - \sigma_B^2}{\sigma_B^2}$. Replacing the correlation in the instantaneous variance, we get

$$\sigma_{B,\rho}^2 \to \sum_{i=1}^N \widehat{\omega}_i^2(t)\sigma_i^2 + \sum_{i \neq j} \widehat{\omega}_i(t)\widehat{\omega}_j(t)\sigma_i\sigma_j(\rho_{ij} + \Delta\rho)$$

which gives

$$\sigma_{B,\rho}^2 \to \sum_{i,j=1}^N \widehat{\omega}_i(t)\widehat{\omega}_j(t)\sigma_i\sigma_j\rho_{ij} + \sum_{i \neq j} \widehat{\omega}_i(t)\widehat{\omega}_j(t)\sigma_i\sigma_j\Delta\rho$$

Writting

$$\sigma_1 = \sum_{i=1}^N \widehat{\omega}_i(t)\sigma_i \text{ and } \sigma_0 = \sqrt{\sum_{i=1}^N \widehat{\omega}_i^2(t)\sigma_i^2}$$

we get

$$[\sigma_1^2 - \sigma_0^2]\Delta\rho = \sum_{i \neq j} \widehat{\omega}_i(t)\widehat{\omega}_j(t)\sigma_i\sigma_j\Delta\rho = \Delta_\rho\sigma_B^2$$

Appendices

Part III

Appendices

Appendix A

Some probabilities

For details see text books by Grimmett et al. [1992], Oksendal [1998] and Jacod et al. [2004].

A.1 Some definitions

Definition A.1.1 *The set of all possible outcomes of an experiment is called the sample space and is denoted* Ω.

Definition A.1.2 *An event is a property which can be observed either to hold or not to hold after the experiment is done. In mathematical terms, an event is a subset of* Ω.

We think of the collection of events as a subcollection $\mathcal{F}$ of the set of all subsets of Ω such that

1. if $A, B \in \mathcal{F}$ then $A \cup B \in \mathcal{F}$ and $A \cap B \in \mathcal{F}$
2. if $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$
3. the empty set $\oslash$ belongs to $\mathcal{F}$

Any collection $\mathcal{F}$ of subsets of Ω which satisfies these three conditions is called a field. It follows from the properties of a field $\mathcal{F}$ that

$$\text{if } A_1, A_2, .., A_n \in \mathcal{F} \text{ then } \bigcup_{i=1}^{n} A_i \in \mathcal{F}$$

so that $\mathcal{F}$ is closed under finite unions and hence under finite intersections also.

Definition A.1.3 *A collection* $\mathcal{F}$ *of subsets of* Ω *is called a* σ*-field if it satisfies the following conditions*

1. *the empty set* $\oslash$ *belongs to* $\mathcal{F}$
2. *if* $A_1, A_2, ... \in \mathcal{F}$ *then* $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$
3. *if* $A \in \mathcal{F}$ *then* $A^c \in \mathcal{F}$

We consider a set E in $\mathbb{R}^d$ and let the measure μ on E associate to some measurable subsets $A \subset E$ be a positive number $\mu(A) \in [0, \infty]$ called the measure of A. The domain of definition of a measure on E is a collection of subsets of E called a σ -algebra which contains the empty set, is stable under unions and contains the complementary of every element. We define the counting measure $\mu_X = \sum_i \delta_{x_i}$ on a countable set of points $X = \{x_i, i = 0, 1, ...\} \subset E$

where $\delta_x(A) = 1$ if $x \in A$ and $\delta_x(A) = 0$ if $x \neq A$ is a Dirac measure such that for any $A \subset E$, $\mu_X(A)$ counts the number of points x_i in A

$$\mu(A) = \#\{i \, , x_i \in A\} = \sum_{i \geq 1} I_{x_i \in A}$$

It is an integer valued measure. A finite measure with mass 1 is called a probability measure.

Definition A.1.4 *Let $E \subset \mathbb{R}^d$. A Radon measure on $(E, \mathcal{B})$ is a measure μ such that for every compact measurable set $B \in \mathcal{B}$, $\mu(B) < \infty$.*

A measure μ_0 which gives zero mass to any point is said to be diffusive or atomless, that is $\forall x \in E, \mu_0(\{x\}) = 0$. Any Radon measure can be decomposed into a diffusive part and a sum of Dirac measures

***Proposition* 13** *Any Radon measure μ can be decomposed into a diffusive part μ_0 and a linear combination of Dirac measures*

$$\mu = \mu_0 + \sum_{j \geq 1} b_j \delta_{x_j} \; x_j \in E, b_j > 0$$

We can now look at measurable functions

Definition A.1.5 *We consider two measurable spaces $(E, \mathcal{E})$ and $(F, \mathcal{F})$, then a function $f : E \to F$ is called measurable if for any measurable set $A \in \mathcal{F}$, the set*

$$f^{-1}(A) = \{x \in E, f(x) \in A\}$$

is a measurable subset of E.

If the measure μ can be decomposed as in Proposition (13) then the integral of μ with respect to f denoted by $\mu(f)$ is

$$\mu(f) = \int f(x) \mu_0(dx) + \sum_{j \geq 1} b_j f(x_j)$$

We let Ω be the set of scenarios equipped with a σ-algebra $\mathcal{F}$ and consider a probability measure on $(\Omega, \mathcal{F})$ which is a positive finite measure $\mathbb{P}$ with total mass 1. Therefore, $(\Omega, \mathcal{F}, \mathbb{P})$ is called a probability space and any measurable set $A \in \mathcal{F}$ called an event is a set of scenarios to which a probability can be assigned. The probability measure assigns value in $[0, 1]$ to each event such that

$$\begin{aligned} \mathbb{P} : \quad & \mathcal{F} && \to [0, 1] \\ & A && \to \mathbb{P}(A) \end{aligned}$$

An event A with probability $\mathbb{P}(A) = 1$ is said to occur almost surely and if $\mathbb{P}(A) = 0$ the event is impossible. We will say that a property holds $\mathbb{P}$-almost surely if the set of $\omega \in \Omega$ for which the property does not hold is a null set (subset of an impossible event). Two probability measures $\mathbb{P}$ and $\mathbb{Q}$ on $(\Omega, \mathcal{F})$ are equivalent if they define the same impossible events

$$\mathbb{P} \sim \mathbb{Q} \Longleftrightarrow [\forall A \in \mathcal{F}, \mathbb{P}(A) = 0 \Longleftrightarrow \mathbb{Q}(A) = 0]$$

A random variable X taking values in E is a measurable map $X : \Omega \to E$ where $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space. $X(\omega)$ represents the outcome of the random variable if the scenario ω happens and is called the realisation of X in the scenario ω. The law of X is the probability measure on E defined by $\mu_X(A) = \mathbb{P}(X \in A)$.

A.2 Random variables

Definition A.2.1 *A random variable is a function $X : \Omega \to \mathbb{R}$ with the property that $\{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F}$ for each $x \in \mathbb{R}$.*

Definition A.2.2 *The distribution function of a random variable X is the function $F : \mathbb{R} \to [0,1]$ given by*

$$F(x) = P(X \leq x)$$

A.2.1 Discrete random variables

Definition A.2.3 *The random variable X is called discrete if it takes values in some countable subset $\{x_1, x_1, ..\}$, only, of $\mathbb{R}$.*

Its distribution function $F(x) = P(X \leq x)$ is a jump function.

Definition A.2.4 *The probability mass function of a discrete random variable X is the function $f : \mathbb{R} \to [0,1]$ given by $f(x) = P(X = x)$.*

The distribution and mass functions are related by

$$F(x) = \sum_{i:x_i \leq x} f(x_i)$$

Lemma A.2.1 *The probability mass function $f : \mathbb{R} \to [0,1]$ satisfies*

1. *$f(x) \neq 0$ if and only if x belongs to some countable set $\{x_1, x_2, ..\}$*
2. *$\sum_i f(x_i) = 1$*

Let $x_1, x_2, .., x_N$ be the numerical outcomes of N repetitions of some experiment. The average of these outcomes is

$$m = \frac{1}{N} \sum_i x_i$$

In advance of performing these experiments we can represent their outcomes by a sequence $X_1, X_2, .., X_N$ of random variables, and assume that these variables are discrete with a common mass function f. Then, roughly speaking, for each possible value x, about $Nf(x)$ of the X_i will take that value x. So, the average m is about

$$m \approx \frac{1}{N} \sum_x xNf(x) = \sum_x xf(x)$$

where the summation is over all possible values of the X_i. This average is the expectation or mean value of the underlying distribution with mass function f.

Definition A.2.5 *The mean value or expectation of X with mass function f is defined to be*

$$E[X] = \sum_{x:f(x)>0} xf(x)$$

whenever this sum is absolutely convergent.

Definition A.2.6 *If k is a positive integer, then the kth moment m_k of X is*

$$m_k = E[X^k]$$

The kth central moment σ_k is

$$\sigma_k = E[(X - m_1)^k]$$

The two moments of most use are $m_1 = E[X]$ and $\sigma_2 = E[(X - E(X))^2]$ called the mean and variance of X.

A.2.2 Continuous random variables

Definition A.2.7 *The random variable X is called continuous if its distribution function can be expressed as*

$$F(x) = \int_{-\infty}^{x} f(u)du \, , \, x \in \mathbb{R}$$

for some integrable function $f : \mathbb{R} \to [0, \infty)$.

The expectation of a discrete variable X is $E[X] = \sum_x xP(X = x)$ which is an average of the possible values of X, each value being weighted by its probability. For continuous variables, expectations are defined as integrals.

Definition A.2.8 *The expectation of a continuous random variable X with density function f is*

$$E[X] = \int_{-\infty}^{\infty} xf(x)dx$$

whenever this integral exists.

We shall allow the existence of $\int g(x)dx$ only if $\int |g(x)|dx < \infty$. Note, the definition of the kth moment m_k in Appendix (A.2.1) applies to continuous random variables, but the moments of X may not exist since the integral

$$E[X^k] = \int x^k f(x)dx$$

may not converge.
We are now going to define independence of random variables.

Definition A.2.9 *Random variables X and Y are independent if their joint distribution function factors into the product of their marginal distribution functions*

$$F_{X,Y}(x, y) = F_X(x)F_Y(y)$$

Theorem A.2.1 *Suppose X and Y are jointly continuous random variables. X and Y are independent if and only if given any two densities for X and Y their product is the joint density for the pair (X, Y), that is,*

$$f_{X,Y}(x, y) = f_X(x)f_Y(y)$$

Independence requires that the set of points where the joint density is positive must be the Cartesian product of the set of points where the marginal densities are positive, that is, the set of points where $f_{X,Y}(x, y) > 0$ must be (possibly infinite) rectangles.

A.3 The characteristic function, moments and cumulants

A.3.1 Definitions

We start by recalling some definitions together with the properties of the characteristic functions. The characteristic function of a random variable is the Fourier transform of its distribution

Definition A.3.1 *The characteristic function of a $\mathbb{R}^d$ random variable X is the function $\Phi_X : \mathbb{R}^d \to \mathbb{C}$ defined by*

$$\Phi_X(z) = E[e^{iz.X}] = \int_{\mathbb{R}^d} e^{iz.x} d\mu_X(x) \, \text{for } z \in \mathbb{R}^d$$

where μ_X is the measure of X.

The characteristic function of a random variable completely characterises its law. Smoothness properties of Φ_X depend on the existence of moments of the random variable X which is related on how fast the distribution μ_X decays at infinity. If it exists, the n-th moment m_n of a random variable X on $\mathbb{R}$ is

$$m_n = E[X^n]$$

The first moment of X called the mean or expectation measures the central location of the distribution. Denoting the mean of X by μ_X, the nth central moment of X, if it exists, is defined as

$$m_n^c = E[(X - \mu_X)^n]$$

The second central moment σ_X^2 called the variance of X measures the variability of X. The third central moment measures the symmetry of X with respect to its mean, and the fourth central moment measures the tail behaviour of X. In statistics, skewness and kurtosis, respectively the normalised third and fourth central moments of X are used to summarise the extent of asymmetry and tail thickness. The are defined as

$$S = \frac{m_3^c}{(m_2^c)^{\frac{3}{2}}} = E[\frac{(X-\mu_X)^3}{\sigma_X^3}] \, , K = \frac{m_4^c}{(m_2^c)^2} = E[\frac{(X-\mu_X)^4}{\sigma_X^4}]$$

Since $K = 3$ for a normal distribution, the quantity $K - 3$ is called the excess kurtosis. The moments of a random variable are related to the derivatives at 0 of its characteristic function.

Proposition 14 *If $E[|X|^n] < \infty$ then Φ_X has n continuous derivatives at $z = 0$ and*

$$m_k = E[X^k] = \frac{1}{i^k} \frac{\partial^k \Phi_X}{\partial z^k}(0)$$

Proposition 15 *X possesses finite moments of all orders iff $z \to \Phi_X(z)$ is C^∞ at $z = 0$. Then the moments of X are related to the derivatives of Φ_X by*

$$m_n = E[X^n] = \frac{1}{i^n} \frac{\partial^n \Phi_X}{\partial z^n}(0)$$

If X_i with $i = 1, .., n$ are independent random variables, the characteristic function of $S_n = X_1 + ... + X_n$ is the product of characteristic functions of individual variables X_i

$$\Phi_{S_n}(z) = \prod_{i=1}^n \Phi_{X_i}(z) \tag{A.3.1}$$

We see that $\Phi_X(0) = 1$ and that the characteristic function Φ_X is continuous at $z = 0$ and $\Phi_X(z) \neq 0$ in the neighborhood of $z = 0$. It leads to the definition of the cumulant generating function or log characteristic function of X.

Definition A.3.2 *There exists a unique continuous function Ψ_X called the cumulant generating function defined around zero such that*

$$\Psi_X(0) = 0 \text{ and } \Phi_X(z) = e^{\Psi_X(z)}$$

The cumulants k_n of a probability distribution are a set of quantities providing an alternative to the moments of the distribution. It is defined via the cumulant-generating function $\Psi_X(z)$, which is the natural logarithm of the moment generating function

$$\Psi_X(z) = \ln \Phi_X(z)$$

The cumulants k_n are obtained from the power series expansion of the cumulant generating function

$$\Psi_X(z) = \sum_{n=1}^{\infty} k_n \frac{z^n}{n!}$$

so that the nth cumulant can be obtained by differentiating the above equation n-times and evaluating the result at zero

$$k_n = \Psi_X^{(n)}(z)\big|_{z=0}$$

A.3.2 The first two moments

We consider the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and let X and Y be two discrete random variables. We are going to present some properties of expectations involving joint distributions. For constants $a, b \in \mathbb{R}$, we have

$$E[aX + bY] = aE[X] + bE[Y]$$

Next, given the definition of the kth moment m_k and the kth central moment σ_k in Appendix (A.2.1) we get the following definition

Definition A.3.3

$$\begin{aligned}
Cov(X,Y) &= E[XY] - E[X]E[Y] = E\big[(X - E[X])(Y - E[Y])\big] \\
E[XY] &= \sum_{x,y} xy f_{XY}(x,y) \\
Var(X) &= Cov(X,X) = E[X^2] - (E[X])^2 = E\big[(X - E[X])^2\big] \\
\rho(X,Y) &= \frac{Cov(X,Y)}{\big(Var(X)Var(Y)\big)^{\frac{1}{2}}} \text{ whenever } Var(X), Var(Y) \neq 0 \text{ and all these quantities exist}
\end{aligned}$$

The covariance measures whether, or not, $(X - E[X])$ and $(Y - E[Y])$ have the same sign. Further, the correlation is scale invariant

$$\rho(aX + b, cY + d) = \rho(X,Y)$$

For random variables X, Y, Z and constants $a, b, c, d \in \mathbb{R}$ then

- $Cov(aX + b, cY + d) = acCov(X,Y)$
- $Cov(X + Y, Z) = Cov(X,Z) + Cov(Y,Z)$
- $Cov(X,Y) = Cov(Y,X)$

Theorem A.3.1 *For X and Y random variables, whenever the correlation $\rho(X,Y)$ exists, it must satisfy*

$$-1 \leq \rho(X,Y) \leq 1$$

The correlation $\rho(X,Y)$ is a measure of the strength and direction of the linear relationship between X and Y. Also,

- if X, Y have non-zero variance, then $\rho \in [-1,1]$.
- Y is a linearly increasing function of X if and only if $\rho(X,Y) = 1$.
- Y is a linearly decreasing function of X if and only if $\rho(X,Y) = -1$.

If X and Y are independent, then $\rho(X,Y) = 0$ and we must have $Cov(X,Y) = 0$ which leads to

$$E[XY] = E[X]E[Y]$$

Remark **A.3.1** *Independence of X and Y implies $Cov(X,Y) = 0$ but not vice versa. That is, some times $\rho(X,Y) = 0$ and X and Y are dependent.*

Moreover, if X and Y are independent, we get

$$Var(X+Y) = Var(X) + Var(Y)$$

Otherwise, if they are correlated, that is $\rho(X,Y) \neq 0$ we set $Z = X + Y$ and plug it back into the variance equation

$$\begin{aligned}
Var(Z) &= E[Z^2] - (E[Z])^2 = E[(X+Y)^2] - (E[X+Y])^2 \\
&= E[X^2 + Y^2 + 2XY] - \big(E[X] + E[Y]\big)^2 \\
&= E[X^2] + E[Y^2] + E[2XY] - (E[X])^2 - (E[Y])^2 - 2E[X]E[Y] \\
&= E[X^2] - (E[X])^2 + E[Y^2] - (E[Y])^2 + 2\big(E[XY] - E[X]E[Y]\big) \\
&= Var(X) + Var(Y) + 2Cov(X,Y)
\end{aligned}$$

More generally, for n random variables $X_1, ..., X_n$ the variance becomes

$$Var(\sum_{i=1}^n X_i) = \sum_{i=1}^n Var(X_i) + \sum_{i \neq j}^n Cov(X_i, X_j) \quad \text{(A.3.2)}$$

A.4 Introducing stochastic processes

We let $X_t(\omega)$ be a random process with discrete time t, then the choice of a point ω in Ω results in a infinite sequence

$$...X_{-1}(\omega), X_0(\omega), X_1(\omega), ...$$

or

$$X_1(\omega), ..., X_k(\omega), X_{k+1}(\omega), ...$$

which is a single realisation of the process $X_t(\omega)$. When time t is continuous, we get a real-valued function defined on the axis t. Let $X = (X_1, ..., X_k)$, $k \geq 2$ be some set of random variables, we denote the distribution function of the random variable X_i by $F_i(x_i)$, $i = 1, .., k$ and its joint distribution by $F_X(x_1, ..., x_k)$. We say that the random variables of this set are independent if

$$F_X(x_1,x_2,...,x_k)=F_1(x_1)F_2(x_2)...F_k(x_k)$$

for any $x_1,...,x_k$. Thus, the property of mutual independence of random variables of a set X is some special property of its joint distributions. The above equation is a special case of the more general property of probability distributions

$$P(X_1\in B_1,X_2\in B_2,...,X_k\in B_k)=P(X_1\in B_1)P(X_2\in B_2)...P(X_k\in B_k)$$

which holds for any Borel sets $B_1,...,B_k$. The mathematical notion of independence in probability theory is of particular importance since it allows to model the situation where the events do not cause one another. A process $X_t(\omega)$ is called narrow-sense stationary if for any finite set of times $t_1<t_2<...<t_n$ and any real τ, the joint distributionsof the vectors

$$\big(X_{t_1}(\omega),X_{t_2}(\omega),...,X_{t_k}(\omega)\big)$$

and

$$\big(X_{t_1+\tau}(\omega),X_{t_2+\tau}(\omega),...,X_{t_k+\tau}(\omega)\big)$$

coincide. Combining results from Birhhoff on the theory of dynamic systems and Khinchin on the relation between dynamic systems and narrow-sense stationary random processes we get the following theorem

Theorem A.4.1 *Birhhoff-Khinchin*
If for a stationary process $X_t(\omega)$ the mathematical expectation $E[|X_t(\omega)|]$ exists, then the limit

$$\lim_{(T-S)\to}\frac{1}{(T-S)}\sum_{t=S+1}^{T}X_t(\omega)=\widehat{X}$$

for a process with discrete time, or the limit

$$\lim_{(T-S)\to}\frac{1}{(T-S)}\int_S^T X_t(\omega)dt=\widehat{X}$$

for a process with continuous time, exists with probability one.

To recover the strong law of large numbers of Kolmogorov, the expectation $\widehat{X}=m=E[X_t(\omega)]$ is not sufficient, and we must consider an extra condition consisting in the existence of the variance $\sigma^2=Var(X_t(\omega))$ and the property

$$\lim_{T\to}\frac{1}{T}\sum_{\tau=0}^{T-1}B(\tau)=0$$

for a process with discrete time, or

$$\lim_{T\to}\frac{1}{T}\int_0^T B(\tau)d\tau=0$$

for a process with continuous time, where

$$B(\tau)=E[X_{t+\tau}(\omega)X_t(\omega)]-m^2$$

is the correlation function of the process $X_t(\omega)$.

Definition A.4.1 *We let $\{X_t;t=1,2,...\}$ be a sequence of random variables on a given probability space $(\Omega,\mathcal{F},\mathbb{P})$ with $E[X_t=0]$ and $\{\mathcal{F}_t;t=1,2,...\}$ a current of σ-algebras on the measurable space $(\Omega,\mathcal{F})$, where Ω is the complete universe of all possible events. Then $\{X_t\}$ is a sequence of independent random variables with respect to $\{\mathcal{F}_t\}$ if X_t is measurable with respect to $\mathcal{F}_t$ and is independent of $\mathcal{F}_{t-1}$ for all $t=1,2,...$*

We saw that if $X_1, X_2, ...$ is a stationary process, then the passage of time correspond to a shift process $X_{1+k}, X_{2+k}, ...$, for some time shift k. Thus, the notion of memory is related to the connections between the process and its shifts. Since the process is stationary, the shifts do not change the distribution of the process.

Definition A.4.2 *A shift transformation T is ergodic if for every measurable function $f \in L_1(\Omega, \mathcal{F}, P)$ we have*

$$\lim_{n \to \infty} \frac{1}{n} \sum_{j=0}^{n-1} f(T^j X) = E[f(X)] \; a.s.$$

where $T^j X$ is the jth shift of the process X:

$$T^j(..., X_{-1}, X_0, X_1, ...) = (..., X_{j-1}, X_j, X_{j+1}, ...)$$

A transformation T of a probability space is mixing if for every two measurable sets A and B we have $P(A \cap T^{-n}B) \to P(A)P(B)$ as $n \to \infty$, and a stationary process is called mixing if the corresponding shift transformation T is mixing. Thus, a mixing stationary process is ergodic and we can construct ergodic but non-mixing stationary processes.

Proposition 16 *A stationary stochastic process is not ergodic if and only if it can be represented as a non-trivial mixture of two different stationary stochastic processes*

$$X = \begin{cases} Y \text{ with probability } p \\ Z \text{ with probability } 1-p \end{cases}$$

where $0 < p < 1$ and Y and Z are stationary stochastic processes with different finite-dimensional distribution.

In general, we say that a non-ergodic stationary process X has infinite memory. However, the mixing property is not sufficiently strong to say that a stationary process with this property has short memory. A stronger requirement is needed, and it is given by the strong mixing conditions:

Definition A.4.3 *Let $X = (..., X_{-1}, X_0, X_1, ...)$ be a stationary process. Define for $n \geq 1$*

$$\alpha_X(n) = \sup\{|P(A \cap B) - P(A)P(B)| \, , \, A \in \sigma(X_k, k \leq 0) \, , \, B \in \sigma(X_k, k \geq n)\}$$

Then, the process X is called strongly mixing if $\alpha_X(n) \to 0$ as $n \to \infty$.

A martingale process is defined on the basis of semi-martingales.

Definition A.4.4 *A random process $(X_t)_{t>0}$ is called a submartingale if*

$$E[|X_t|] < \infty$$

and

$$E[X_t|\mathcal{F}_s] \geq X_s \, , \, s < t$$

a supermartingale if, instead

$$E[X_t|\mathcal{F}_s] \leq X_s \, , \, s < t$$

and it is a martingale if the process is both a submartingale and a supermartingale.

We briefly present the Kunita-Watanabe decomposition (see Kunita et al. [1967]) used in portfolio replication.

Definition A.4.5 *Kunita-Watanabe decomposition*
Let M be a continuous local martingale. Then every continuous local martingale N can be uniquely written as

$$N = H \bullet M + L$$

where $H \bullet M = \int HdM$, $H \in L^2_{loc}(M)$, where L is a continuous local martingale strongly orthogonal to M.

Note, in some cases we can drop the continuity assumption on N (see Ansel et al. [1993]). We now provide the definitions of a Markov process and that of an independent process.

Definition A.4.6 *A random process $(X_t)_{t>0}$ is called a a Markov process if, for each n and every $i_0, .., i_n$, then*

$$P(X_{n+1} = j | X_0 = i_0, ..., X_n = i_n) = P(X_{n+1} = j | X_n = i_n)$$

where $P(\bullet|\bullet)$ denotes conditional probability.

We can now define a random walk (RW) process and geometric Brownian motion (GBM).

Definition A.4.7 *A random walk is a Markov process with independent innovations*

$$X_t - X_{t-1} = \epsilon_t$$

where $\epsilon_t \approx IID$, standing for independent and identically distributed process.

Definition A.4.8 *A geometric Brownian motion is a random walk of natural logarithm of the original process X_t, where $L_t = \ln(X_t)$, so that*

$$\Delta L_t = L_t - L_{t-1} = \epsilon_t$$

where $\epsilon_t \approx IID$.

Note, martingale is more general than random walk since semi-martingales allow for dependence in the process. Thus, random walk implies martingale but martingale does not imply random walk in the process.

A.5 Standard Theorems and Proofs

We state and prove the Conditional Jensen Inequality as presented by Shreve [2003].

Theorem A.5.1 (The Conditional Jensen Inequality) *Let $\phi : \mathbb{R} \to \mathbb{R}$ be a convex function and $\mathbb{E}|\phi(X)| < \infty$ then*

$$\mathbb{E}[\phi(X)|\mathcal{F}] \geq \phi(\mathbb{E}[X|\mathcal{F}])$$

■ Since ϕ is convex, we can express it as

$$\phi(x) = \max\{g | \, g(x) \leq \phi(x) \, \forall x \in \mathbb{R}; \text{ g is linear}\}$$

Since g is linear, we may write $g(x) = ax + b$ for $a, b, \in \mathbb{R}$. Hence

$$\begin{aligned}
\mathbb{E}[\phi(X)|\mathcal{F}] &\geq \mathbb{E}[aX + b|\mathcal{F}] \\
&= a\mathbb{E}[X|\mathcal{F}] + b \\
&= g(\mathbb{E}[X|\mathcal{F}])
\end{aligned}$$

Therefore

$$
\begin{aligned}
\mathbb{E}[\phi(X)|\mathcal{F}] &\geq \max_{g\leq\phi} g(\mathbb{E}[X|\mathcal{F}]) \\
&= \phi(\mathbb{E}[X|\mathcal{F}])
\end{aligned}
$$

We give the special case of the Optional Sampling Theorem necessary for use in Section (1.1.4.2). The theorem and proof that follow are given by Williams [1991].

Theorem A.5.2 (Optional Sampling Theorem) *Let X be a supermartingale. Let τ be a stopping time such that for some $K \in \mathbb{N}$, $\tau(\omega) \leq K, \forall\omega$. Then $X_\tau \in \mathcal{L}^1(\Omega, \mathcal{F}_\tau, \mathbb{P})$ and*

$$
\mathbb{E}[X_K|\mathcal{F}_\tau] \leq X_\tau
$$

If X is a submartingale then

$$
\mathbb{E}[X_K|\mathcal{F}_\tau] \geq X_\tau
$$

■ Let $A \in \mathcal{F}_\tau$. Then

$$
\begin{aligned}
\mathbb{E}[X_K]\mathbf{I}_A &= \sum_{n\leq K} \mathbb{E}[X_K]\mathbf{I}_{A\cap\{\tau=n\}} \\
&\leq \sum_{n\leq K} \mathbb{E}[X_n]\mathbf{I}_{A\cap\{\tau=n\}} \\
&= \mathbb{E}[X_\tau]\mathbf{I}_A
\end{aligned}
$$

The fact that $|X_\tau| \leq \sum_{n\leq K}|X_n|$ guarantees the integrability of X_τ. If we consider instead the process $-X$, so that $-X$ is a submartingale, then analogously to above, we obtain the result in the second part of the theorem. We give the special one-dimensional case of the Girsanov Theorem with Brownian Motion, necessary for the probabalistic approach to the Black-Scholes model in Section (1.2.3). We also give the Levy Characterisation Theorem that is needed during its proof. We follow the construction given by Zheng [2015].

Theorem A.5.3 (Girsanov's Theorem) *Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}), \mathbb{P})$ be a filtered probability space, where $T \in (0,\infty)$ and assume for simplicity that $\mathcal{F} = \mathcal{F}_T$. Let W_t be an adapted one-dimensional Brownian Motion and $\theta(t)$ an adapted process such that $\int_0^T \theta^2(s)ds < \infty$ almost surely. Then define*

$$
Z_t = e^{\left(\int_0^t \theta(s)dW_s - \frac{1}{2}\int_0^t \theta^2(s)ds\right)}
$$

If $\mathbb{E}\left[\int_0^T \theta^2(s)Z_t^2 ds < \infty\right]$ then $\mathbb{E}[Z_T] = 1$. Furthermore, define a measure $\mathbb{Q}$ on $(\Omega, \mathcal{F})$ by $\frac{d\mathbb{Q}}{d\mathbb{P}} = Z_T$, then under the measure $\mathbb{Q}$

$$
\tilde{W}_t = W_t - \int_0^t \theta(s)ds
$$

is an $\mathcal{F}_t$ Brownian Motion.

■ Assuming that $\mathbb{E}[Z_T] = 1$ ensures that $\mathbb{Q}$ will indeed be a probability measure. Now applying the product rule

$$
\begin{aligned}
d(Z_t W_t) &= Z_t dW_t + W_t dZ_t + d < Z_t, W_t > \\
&= Z_t(W_t\theta(t) + 1)dW_t
\end{aligned}
$$

Hence, $(Z_t W_t)_{t\geq 0}$ is a local martingale. Then, since

$$\mathbb{E}^{\mathbb{Q}}[\tilde{W}_t|\mathcal{F}_s] = \frac{\mathbb{E}[\tilde{W}_t Z_t|\mathcal{F}_s]}{\mathbb{E}[Z_t|\mathcal{F}_s]} = \frac{\mathbb{E}[\tilde{W}_t Z_t|\mathcal{F}_s]}{Z_s}$$

then $(\tilde{W}_t Z_t)_{t\geq 0}$ is a martingale if and only if $\tilde{W}_t$ is a $\mathbb{Q}$-martingale. Hence, the latter must be true. Further, $\tilde{W}_t$ has continuous sample paths and $[\tilde{W}_t] = t$. So, by the Levy Characterisation Theorem, $\tilde{W}_t$ is a $\mathbb{Q}$-Brownian Motion.

Theorem A.5.4 (Levy Characterisation Theorem) *Let M_t be a continious martingale on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}), \mathbb{P})$ and the quadratic variation of M be $[M_t] = t$. Then M_t is an $\mathcal{F}_t$-Brownian Motion.*

■ We suppose that M_t is a continuous martingale with $[M_t] = t$ and consider the function $g(t,x) = e^{i\theta x + \theta^2 \frac{t}{2}}$. By applying Itô's formula we obtain

$$dg(t, M_t) = \frac{1}{2}\theta^2 g(t, M_t)dt + i\theta g(t, M_t)dM_t - \frac{1}{2}\theta^2 g(t, M_t)d[M]_t$$

so that $g(t, M_t)$ is a martingale if $[M]_t = t$. Hence for $t > s$

$$\mathbb{E}[e^{i\theta(M_t - M_s)}|\mathcal{F}_s] = e^{-\frac{1}{2}\theta^2(t-s)}$$

Let X be any $\mathcal{F}_s$-measurable random variable and ψ_X its characteristic function. Then by the Tower Property of conditional expectation, the joint characteristic function of X and $M_t - M_s$ is

$$\begin{aligned}
\psi_{X, M_t - M_s}(\theta, \phi) &= \mathbb{E}[e^{i(\phi X + \theta(M_t - M_s))}] \\
&= \mathbb{E}[e^{i\phi X} e^{i\theta(M_t - M_s)}|\mathcal{F}_s] \\
&= \mathbb{E}[e^{i\phi X}\mathbb{E}[e^{i\theta(M_t - M_s)}|\mathcal{F}_s]] \\
&= \mathbb{E}[e^{i\phi X}]e^{-\frac{1}{2}\theta^2(t-s)} \\
&= \psi_X(\phi)\psi_{M_t - M_s}(\theta)
\end{aligned}$$

So, X and $M_t - M_s$ are independent, meaning $(M_t - M_s)$ is independent of $\mathcal{F}_s$. Also from above, $(M_t - M_s) \sim N(0, t-s)$. Hence, (M_t) is an $(\mathcal{F}_t)$-Brownian Motion.

A.6 Some continuous variables and their distributions

For details see text book by Grimmett et al. [1992].

A.6.1 Some popular distributions

A.6.1.1 Exponential distribution

The exponential distribution is the probability distribution describing the time between events in a Poisson process, that is, a process where events occur continuously and independently at a constant average rate. It is the continuous analogue of the geometric distribution, and it has the key property of being memoryless. X is exponential with parameter $\lambda (> 0)$ if

$$F(x) = \begin{cases} 1 - e^{-\lambda x} \,, x \geq 0 \\ 0 \text{ otherwise} \end{cases}$$

which is the continuous limit of the waiting time distribution. The probability density function is given by

$$f(x) = \begin{cases} \lambda e^{-\lambda x} \text{ for } x \geq 0 \\ 0 \text{ otherwise} \end{cases}$$

The mean of X is given by

$$E[X] = \int_0^\infty [1 - F(x)] dx = \frac{1}{\lambda}$$

and the variance of X is given by $Var(X) = \frac{1}{\lambda^2}$ so that the standard deviation of X is equal to its mean. The moments of X for $n = 1, 2, ..$ are given by

$$E[X^n] = \frac{n!}{\lambda^n}$$

An exponentially distributed random variable T follows the relation

$$P(T > s + t | T > s) = P(T > t) \;, \forall s, t \geq 0$$

A.6.1.2 Gamma distribution

The gamma distribution is a two-parameter family of continuous probability distributions where the exponential distribution and the chi-squared distribution are special cases. It is the maximum entropy probability distribution for a random variable X for which $E[X] = k\theta = \frac{\alpha}{\beta}$ is fixed and greater than zero, and $E[\ln X] = \psi(k) + \ln\theta = \psi(\alpha) - \ln\beta$ is fixed and $\psi(\bullet)$ is the digamma function. In the following we denote the pair (k, θ) by $(t, \frac{1}{\lambda})$ corresponding to $\alpha = t$ and $\beta = \lambda$. The variable X has the gamma distribution with parameters λ, $t > 0$, denoted by $\Gamma(\lambda, t)$ if it has the probability density function

$$f(x) = \frac{1}{\Gamma(t)} \lambda^t x^{t-1} e^{-\lambda x} \;, x \geq 0$$

where $\Gamma(t)$ is the gamma function

$$\Gamma(t) = \int_0^\infty x^{t-1} e^{-x} dx$$

If $t = 1$ then X is Exponentially distributed with parameter λ. The cumulative distribution function is the regularised gamma function

$$F(x) = \int_0^x f(u) du = \frac{\gamma(t, \lambda x)}{\Gamma(t)}$$

where $\gamma(t, \lambda x)$ is the lower incomplete gamma function. If t is a positive integer, the CDF follows the series expansion

$$F(x) = e^{-\lambda x} \sum_{i=t}^\infty \frac{(\lambda x)^i}{i!}$$

The skewness $\frac{2}{\sqrt{t}}$ depends only on the shape parameter t and approaches a normal distribution when t is large (when $t > 10$).

A.6.1.3 Beta distribution

The beta distribution is a family of continuous probability distributions defined on the interval $[0,1]$ with two shape parameters α and β. It is applied to model the behaviour of random variables limited to interval of finite length. The probability density function of the beta distribution for $0 \leq x \leq 1$ is a power function of the variable X and of its reflection $(1-X)$

$$\begin{aligned}
f(x) &= cst \times x^{\alpha-1}(1-x)^{\beta-1} = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\int_0^1 u^{\alpha-1}(1-u)^{\beta-1}du} \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} \\
&= \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}
\end{aligned}$$

where $\Gamma(\bullet)$ is the gamma function, and $B(\bullet,\bullet)$ is a normalisation constant ensuring that the total probability integrates to 1. Note, this definition includes both ends $x=0$ and $x=1$, but some authors choose to exclude them and consider $0 < x < 1$ instead. Beta densities are symmetric (for $\alpha = \beta = 1$), unimodal ($\alpha, \beta > 1$), uniantimodal, increasing, decreasing or constant depending on the values α and β relative to 1, and have many more attractive properties. The beta density is U-shaped when $\alpha, \beta < 1$, it has positive skew when $\alpha < \beta$ and negative skew when $\alpha > q$. Since the beta distribution approaches the Bernoulli distribution in the limit when both shape parameters α and β approach zero, some authors denote the pair (α, β) by (p,q). The slope of the pdf is given by

$$\begin{aligned}
f^{'}(x) &= f(x) \frac{(\alpha+\beta-2)x - (\alpha-1)}{(x-1)x} \\
&= -\frac{x^{\alpha-2}(1-x)^{\beta-2}}{B(\alpha,\beta)} \big((\alpha+\beta-2)x - (\alpha-1) \big)
\end{aligned}$$

and at $x = \frac{1}{2}$, for $\alpha = \beta$, the slope of the pdf is zero. Further, we get the differential equation

$$(x-1)xf^{'}(x) + \big(\alpha - 1 - (\alpha+\beta-2)x\big) f(x) = 0$$

The cumulative distribution function is given by

$$F(x) = \frac{B(x;\alpha,\beta)}{B(\alpha,\beta)} = I_x(\alpha,\beta)$$

where $B(\bullet;\alpha,\beta)$ is the incomplete beta function and $I_\bullet(\alpha,\beta)$ is the regularised incomplete beta function. The mode of a beta distributed random variable X with $\alpha, \beta > 1$ (corresponding to the peak in the PDF) is given by

$$\frac{\alpha-1}{\alpha+\beta-2}$$

The mean of X is given by

$$\begin{aligned}
\mu = E[X] = \int_0^1 xf(x)dx &= \int_0^1 x \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)} dx \\
&= \frac{\alpha}{\alpha+\beta} = \frac{1}{1+\frac{\beta}{\alpha}}
\end{aligned}$$

which only depends on the ratio $\frac{\beta}{\alpha}$. For $\alpha = \beta$, we get the mean $\mu = \frac{1}{2}$, which is at the center of the (symmetric) distribution. We also get the following limits

$$\lim_{\frac{\beta}{\alpha} \to 0} \mu = 1 \text{ and } \lim_{\frac{\beta}{\alpha} \to \infty} \mu = 0$$

For the former limit ratio, the beta distribution becomes a one-point degenerate distribution with a Dirac delta function spike at $x = 1$ with probability 1 and zero probability elsewhere else. Similarly, for the latter limit ratio the spike is at $x = 0$. Next, we consider the limit cases where one parameter is finite (non-zero) and the other one approaches the limits

$$\lim_{\beta \to 0} \mu = \lim_{\alpha \to \infty} \mu = 1 \text{ and } \lim_{\beta \to \infty} \mu = \lim_{\alpha \to 0} \mu = 0$$

The variance of a beta distributed random variable X is given by

$$Var(X) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$$

and for $\alpha = \beta$ the variance simplifies to

$$Var(X) = \frac{1}{4(2\beta + 1)}$$

Setting $\alpha = \beta = 0$ in the above equation, we obtain the maximum variance $Var(X) = \frac{1}{4}$. The skewness of the beta distribution is

$$\gamma_1 = \frac{E[(X - \mu)^3]}{(Var(X))^{\frac{3}{2}}} = \frac{2(\beta - \alpha)\sqrt{\alpha + \beta + 1}}{(\alpha + \beta + 2)\sqrt{\alpha\beta}}$$

For $\alpha = \beta$ in the above equation, we obtain $\gamma_1 = 0$ showing that in that setting the distribution is symmetric. We get positive skew (right-tailed) for $\alpha < \beta$ and negative skew (left-tailed) for $\alpha > \beta$.

A.6.1.4 Kumaraswamy distribution

The Kumaraswamy's double bounded distribution (also known as minmax distribution) is a family of continuous probability distributions defined on the interval $[0, 1]$ with two shape parameters a and b (see Kumaraswamy [1980]). While similar to the beta distribution, the Kumaraswamy distribution is simpler to use due to the simple closed form of both its probability density function and cumulative distribution function. The pdf is given by

$$f(x) = abx^{a-1}(1 - x^a)^{b-1} \text{ , } 0 < x < 1 \text{ , } a, b > 0$$

and the cdf is given by

$$F(x) = 1 - (1 - x^a)^b$$

In a more general form, using linear transformation, the normalised variable x is replaced with the unshifted and unscaled variable z where

$$x = \frac{z - z_{min}}{z_{max - z_{min}}} \text{ , } z_{min} \leq z \leq z_{max}$$

In a discrete setting where the density satisfies $P(X \in dx) \approx f_X(x)dx$, then when considering the unshifted variable z we get $dx = \frac{1}{z_{max - z_{min}}}$. We can invert the distribution function to obtain the quantile function

$$Q(y) = F^{-1}(y) = \left[1 - (1 - y)^{\frac{1}{b}}\right]^{\frac{1}{a}} \text{ , } 0 < y < 1$$

We can then trivially generate random variable as for $U \sim U(0,1)$, then $X \sim f$ if

$$X = (1 - U^{\frac{1}{b}})^{\frac{1}{a}}$$

The Kumaraswamy distribution has the same basic shape properties as the beta distribution (see Jones [2009])

- unimodal: $a > 1, b > 1$.
- uniantimodal: $a < 1, b < 1$.
- increasing: $a > 1, b \leq 1$.
- decreasing: $a \leq 1, b > 1$.
- constant: $a = b = 1$.

and the Kumaraswamy density also matches that of the beta density at the boundaries of their support

- $f(x) \sim x^{a-1}$ as $x \to 0$.
- $f(x) \sim (1-x)^{b-1}$ as $x \to 1$

The raw moments of the Kumaraswamy distribution are given by

$$m_n = \frac{b\Gamma(1+\frac{n}{a})\Gamma(b)}{\Gamma(1+b+\frac{n}{a})} = bB(1+\frac{n}{a}, b)$$

where $B(\bullet, \bullet)$ is the Beta function. Similarly to the beta distribution, they exist for all $n > -a$. The variance, skewness, and excess kurtosis can be computed from these raw moments. For instance, the variance is given by $\sigma^2 = m_2 - m_1^2$

$$Var(X) = bB(1+\frac{2}{a}, b) - \big(bB(1+\frac{1}{a}, b)\big)^2$$

Assuming that $X_{a,b}$ is a Kumaraswamy distributed random variable, then it is the a-th root of a suitably defined $Beta$ distributed random variable (see Jones [2009]). Let $Y_{1,b}$ denote a Beta r.v. with parameters $\alpha = 1$ and $\beta = b$, then $X_{a,b} = Y_{1,b}^{\frac{1}{a}}$ with equality in distribution

$$P(X_{a,b} \leq x) = \int_0^x abu^{a-1}(1-u^a)^{b-1}du = \int_0^{x^a} b(1-u)^{b-1}du = P(Y_{1,b} \leq x^a) = P(Y_{1,b}^{\frac{1}{a}} \leq x)$$

Considering a generalised distribution with Beta distributed r.v. $Y_{\alpha,\beta}^{\frac{1}{\gamma}}$ with $\gamma > 0$, the raw moments are given by

$$m_n = \frac{\Gamma(\alpha+\beta)\Gamma(\alpha+\frac{n}{\gamma})}{\Gamma(\alpha)\Gamma(\alpha+\beta+\frac{n}{\gamma})}$$

where the original moments are recovered by setting $\alpha = 1$,$\beta = b$ and $\gamma = a$. However, in general the CDF does not have a closed form solution. Note,

- if $X \sim K(1,1)$ then $X \sim U(0,1)$.
- if $X \sim B(1,b)$ then $X \sim K(1,b)$.
- if $X \sim B(a,1)$ then $X \sim K(a,1)$.
- if $X \sim K(a,b)$ then $X \sim GB1(a,1,1,b)$.

where $GB1(a,1,1,b)$ is the generalised beta distribution of the first kind.

A.6.1.5 Generalised beta distribution

McDonald [1984] introduced the generalised beta distribution of the first kind characterised by its density function

$$f_{GB1}(x) = B^{-1}(p,q)\big[ax^{ap-1}(1-x^a)^{q-1}\big] \text{ , } 0 < x < 1$$

where we recover the classical beta distribution for $a = 1$ $(1,p,q)$, and the Kumaraswamy distribution for $p = 1$ $(a,1,q)$. It is the distribution of the $\frac{1}{a}$ power of a $B(p,q)$ random variable or of the p-th order statistic of a sample of size $p+q-1$ from the power function distribution $B(a,1)$. Jones [2004] introduced the general class of beta-generated distibutions characterised by their density function

$$f_{BG}(x) = B^{-1}(\alpha,\beta)f(x)[F(x)]^{\alpha-1}[1-F(x)]^{\beta-1} \text{ , } x \in \mathcal{I}$$

where $F(x)$ is the parent distribution function and $f(x)$ is its density. Jones concentrated on the cases where F is symmetric about zero with no free parameter other than location and scale, and where $\mathcal{I}$ is the whole real line. Since then, Beta-generated distributions with more general parents have been studied extensively. Even though the shapes of the BG-distributions are more flexible than the beta-normal, they depend on two parameters adding only a limited structure to the generated distribution. Alexander et al. [2010] proposed to use a more flexible generator distribution such as a generalised beta distribution of the first kind $GB1$(a,p,q). They recover theclassical BG and Kumaraswamy generated distributions as special cases. Given a parent distribution $F(x)$ with density $f(x)$, the generalised beta-generated (GBG) density is given by

$$f(GBG)(x) = B^{-1}(p,q)f(x)\big[aF^{ap-1}(x)(1-F^a(x))^{q-1}\big] \text{ , } x \in \mathcal{I}$$

where we recover the beta-generated distribution for $a = 1$ and the Kumaraswamy distribution for $p = 1$. Further, the GBG with parent $F(x)$ is a standard beta-generated distribution with parent $F^a(x)$.

A.6.2 Approximating the probability distribution

The Gram-Charlier A series and the Edgeworth series are series approximating a probability distribution in terms of its cumulants. The series are the same, but the arrangement of terms differ.

A.6.2.1 The Gram-Charlier A series

The idea of the Gram-Charlier A series is to approximate the characteristic function of the distribution, whose probability density function is f, in terms of the characteristic function of a distribution with known and suitable properties. We can then recover f with the inverse Fourier transform. Given a continuous random variable X, we let Ψ be the CF of its distribution with pdf f, and k_r its cumulants (see details in Appendix (A.3)). We expand in terms of a known distribution with pdf ϕ, CF Φ, and cumulants γ_r. The density function ϕ is generally chosen to be that of the normal distribution. By definition of the cumulants, we have

$$\Psi(t) = e^{\sum_{r=1}^{\infty}(k_r-\gamma_r)\frac{(it)^r}{r!}}\Phi(t)$$

By the properties of the Fourier transform, $(it)^r\Phi(t)$ is the Fourier transform of $(-1)^r[D^r\phi](-x)$, where D is the differential operator with respect to x. Thus, setting $x = -x$ in the above equation, we ger the formal expansion

$$f(x) = e^{\sum_{r=1}^{\infty}(k_r-\gamma_r)\frac{(-D)^r}{r!}}\phi(x)$$

If ϕ is the normal density with mean $\mu = k_1$ and variance $\sigma^2 = k_2$, then the expansion of the density becomes

$$f(x) = e^{\sum_{r=3}^{\infty}k_r\frac{(-D)^r}{r!}}\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Expanding the exponential and collecting terms according to the order of the derivatives we obtain the Gram-Charlier A series. Focussing on the first two correction terms to the normal distribution, we get

$$f(x) \approx \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \big[1 + \frac{k_3}{3!\sigma^3} He_3(\frac{x-\mu}{\sigma}) + \frac{k_4}{4!\sigma^4} He_4(\frac{x-\mu}{\sigma})\big]$$

where $He_3(x) = x^3 - 3x$ and $He_4(x) = x^4 - 6x^2 + 3$ are Hermite polynomials. Note, this expression is not guaranteed to be positive, so that it is not a valid probability distribution. There are many cases of interest where the Gram-Charlier A series diverges. It converges only if the density $f(x)$ falls at a faster rate than $e^{-\frac{x^2}{4}}$ at infinity (see Cramer [1957]). Hence, when it does not converge the series is not a true asymptotic expansion because it is not possible to estimate the error of the expansion.

A.6.2.2 The Edgeworth series

Edgeworth developed a similar expansion as an improvement of the central limit theorem. The main advantage of Edgeworth series is that the error is controlled, leading to a true asymptotic expansion. Given $\{X_i\}$ a sequence of i.i.d. random variables with mean μ and variance σ^2, we let Y_n be their standardised sum

$$Y_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{X_i - \mu}{\sigma}$$

We let F_n be the cumulative distrbution functions of the variables Y_n. Then, by the central limit theorem, we get

$$\lim_{n \to \infty} F_n(x) = N(x) = \int_{-\infty}^x \phi(u) du$$

with $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ for every x, as long as the mean and variance are finite. If we now assume that the random variables X_i with $i = 1, .., n$ have higher cumulants $k_r = \sigma^r \lambda_r$, and if we expand in terms of the standardised normal distribution $\phi(x)$, then the cumulant differences in the formal expansion of the CF $\Psi_n(t)$ of F_n are

$$\begin{aligned}
&k_1^{F_n} - \gamma_1 = 0 \\
&k_2^{F_n} - \gamma_2 = 0 \\
&k_r^{F_n} - \gamma_r = \frac{k_r}{\sigma^r n^{\frac{r}{2}-1}} = \frac{\lambda_r}{n^{\frac{r}{2}-1}} \, , r \geq 3
\end{aligned}$$

Following the same approach as the Gram-Charlier A series except that terms are collected according to powers of n, we get

$$\Psi_n(t) = \big[1 + \sum_{j=1}^\infty \frac{P_j(it)}{n^{\frac{j}{2}}}\big] e^{-\frac{t^2}{2}}$$

where $P_j(x)$ is a polynomial of degree $3j$. After applying the inverse Fourier transform, the distribution function is given by

$$F_n(x) = N(x) + \sum_{j=1}^\infty \frac{P_j(-D)}{n^{\frac{j}{2}}} N(x)$$

Letting $N^{(j)}(x)$ be the jth derivative of $N(\bullet)$ at point x, we can recover the first few terms of the expansion. Further, since the derivatives of the density of the normal distribution are related to the normal density by $\phi^{(n)}(x)$ is $(-1)^n He_n(x)\phi(x)$, then we obtain an alternative representation in terms of the density function. However, Edgeworth expansions are not guaranteed to be a proper probability distribution since the integral of the density needs not to integrate to one, and the probabilities can be negative. Further, they can be inaccurate, especially in the tails, because they are obtained under a Taylor series around the mean, and they guarantee (asymptotically) an absolute error but not a relative one.

A.6.3 The stable distributions

The class of stable distributions was characterised by Paul Levy in his study of sums of independent identically distributed terms in the 1920's. It is a rich class of probability distributions allowing for skewness and heavy tails with many intriguing properties. We are first going to define stable distributions and explain the reason why they are called stable. We follow Ouyang [2006] and Nolan [2014]. Note, we omit the letter d above the equal sign or an arrow to indicate equality in distribution.

Definition A.6.1 *X is called stable, if for any positive number a and b, there exists a positive number c and real number d, such that*

$$aX_1 + bX_2 = cX + d \tag{A.6.3}$$

where X_1 and X_2 are independent copies of X. If the equation holds for $d = 0$, then it is called strictly stable. If it $X = -X$, then it is called symmetric.

Proposition 17 *For any stable random variable X, there exists $\alpha \in (0, 2]$, such that numbers in Equation (A.6.3) satisfies*

$$a^\alpha + b^\alpha = c^\alpha$$

In the Gaussian case, $X \sim N(\mu, \sigma^2)$ and we get $c = \sqrt{a^2 + b^2}$ and $d = (a + b - c)\mu$ with $\alpha = 2$.

Definition A.6.2 *X is called stable, if for any integer $n \geq 2$, there exists positive numbers c_n, and real numbers d_n, such that*

$$\sum_{i=1}^n X_i = c_n X + d_n \tag{A.6.4}$$

where X_i are independent copies of X.

If Equation (A.6.4) holds, then $c_n = n^{\frac{1}{\alpha}}$.

A sequence of random variables $X = \{X_i\}_{-\infty}^{\infty}$, for some fixed $\delta > 0$, define a series of transformations T_n, such that

$$(T_n X)_i = \frac{1}{n^\delta} \sum_{j=in}^{(i+1)n-1} X_j \ , n \geq 1$$

This is called re-normalisation group transformation, and $\{T_n\}$ is a semi-group and satisfies $T_n T_m = T_{mn}$.

Proposition 18 *An i.i.d. strictly α-stable random variable sequence is a fixed point for T_n at $\delta = \frac{1}{\alpha}$. For dependent stable sequences, it is still a fixed point for T_n, but the corresponding δ is a function of α and the dependence.*

Definition A.6.3 *X is stable if it has a domain attraction, that is, there exists i.i.d. random variables $Y_1, Y_2, ...$, $d_n > 0$, $a_n \in \mathbb{R}$, such that*

$$\frac{1}{d_n} \sum_{i=1}^{n} Y_i + a_n = X \tag{A.6.5}$$

If X is Gaussian and Y_i are i.i.d. with finite variance, then Equation (A.6.5) is just the Central Limit Theorem (CLT). When $d_n = n^1 \alpha$, then Y is said to belong to the normal domain of attraction X. In general, we have $d_n = n^1 \alpha h(n)$, where $h(x)$ is a slowly varying function at infinity [1] We are now going to define stable distributions in terms of characteric functions.

Definition A.6.4 *X is stable if there exists $0 < \alpha \leq 2$, $c \geq 0$, $-1 \leq \beta \leq 1$, and real number μ,such that the characteric function for X is*

$$E[e^{i\theta X}] = \begin{cases} e^{-c^{\alpha}|\theta|^{\alpha}\left(1 - i\beta(sign\theta)\tan\frac{\pi\alpha}{2}\right) + i\mu\theta} & \text{if } \alpha \neq 1 \\ e^{-c|\theta|\left(1 + i\beta\frac{2}{\pi}(sign\theta)\ln\theta\right) + i\mu\theta} & \text{if } \alpha = 1 \end{cases}$$

where α is the stability index, and c, β, μ are unique, except when $\alpha = 2$, β is arbitrary. We denote that $X \sim S(\alpha, \beta, c, \mu)$.

Many properties can be derived from the characteric functions.

Proposition 19 *If X_1 and X_2 are independent, and $X_i \sim S(\alpha, \beta_i, c_i, \mu_i)$, then $X_1 + X_2 \sim S(\alpha, \beta, c, \mu)$ with*

$$c = (c_1^{\alpha} + c_1^{\alpha})^{\frac{1}{\alpha}}, \ \beta = \frac{\beta_1 c_1^{\alpha} + \beta_2 c_2^{\alpha}}{c_1^{\alpha} + c_1^{\alpha}}, \ \mu = \mu_1 + \mu_2$$

Proposition 20 *If $X \sim S(\alpha, \beta, c, \mu)$, $a \in \mathbb{R}$, then $X + a \sim S(\alpha, \beta, c, a + \mu)$.*

Thus, μ is a location parameter.

Proposition 21 *If $0 < \alpha < 2$, and $X \sim S(\alpha, \beta, c, 0)$, then $-X \sim S(\alpha, -\beta, c, 0)$.*

which means that β is a skewness parameter.

Definition A.6.5 *$S(\alpha, \beta, c, \mu)$ is called skewed to the right (left) if $\beta > 0$ ($\beta < 0$), and it is called totally skewed to the right (left) if $\beta = 1$ ($\beta = -1$).*

Note, the skewness here does not refer to the support of the distribution, but to the parameters P and Q in the Levy measure.

Proposition 22 *$X \sim S(\alpha, \beta, c, \mu)$ is symmetric if and only if $\beta = \mu = 0$.*

Symmetric distributions have real characteristic functions.

Proposition 23 *$X \sim S(\alpha, \beta, c, \mu)$, with $\alpha \neq 1$, then X is strictly stable if and only if $\mu = 0$.*

Proposition 24 *$X \sim S(1, \beta, c, \mu)$, then X is strictly stable if and only if $\beta = 0$.*

[1] $\lim_{x \to \infty} \frac{h(ux)}{h(x)} = 1$, for any $u > 0$.

Proposition 25 *If $X_i \sim S(\alpha, \beta, c, \mu)$, then*

$$\sum_{i=1}^{n} X_i = \begin{cases} n^{\frac{1}{\alpha}} X_1 + \mu(n - n^{\frac{1}{\alpha}}) \, , \, \alpha \neq 1 \\ n X_1 + \frac{2}{\pi} c \beta n \ln n \, , \, \alpha = 1 \end{cases}$$

Proposition 26 *If X has distribution $S(\alpha, \beta, c, 0)$ with $\alpha < 2$, then there exists two i.i.d. random variables Y_1 and Y_2, with common distribution $S(\alpha, 1, c, 0)$, such that*

$$X = \begin{cases} (\frac{1+\beta}{2})^{\frac{1}{\alpha}} Y_1 + (\frac{1-\beta}{2})^{\frac{1}{\alpha}} Y_2 \text{ if } \alpha \neq 1 \\ (\frac{1+\beta}{2}) Y_1 + (\frac{1-\beta}{2}) Y_2 + c\big(\frac{1+\beta}{\pi} \ln \frac{1+\beta}{2} - \frac{1-\beta}{\pi} \ln \frac{1-\beta}{2} \big) \text{ if } \alpha = 1 \end{cases}$$

it means that those stable subordinators can be used as building blocks of more general stable random variables.

Proposition 27 *Let $X \sim S(\alpha, \beta, c, 0)$ with $0 < \alpha < 2$. Then*

$$\begin{cases} \lim_{\lambda \to \infty} \lambda^{\alpha} P(X > \lambda) = C_{\alpha} \frac{1+\beta}{2} c^{\alpha} \\ \lim_{\lambda \to \infty} \lambda^{\alpha} P(X < -\lambda) = C_{\alpha} \frac{1-\beta}{2} c^{\alpha} \end{cases}$$

where

$$C_{\alpha} = \big(\int_0^{\infty} x^{-\alpha} \sin x dx \big)^{-1} = \begin{cases} 1 - \alpha \Gamma(2 - \alpha) \cos \frac{\pi \alpha}{2} \text{ if } \alpha \neq 1 \\ \frac{2}{\pi} \text{ if } \alpha = 1 \end{cases}$$

Proposition 28 *Let $X \sim S(\alpha, \beta, c, 0)$ with $0 < \alpha < 2$. Then*

$$E[|X|^p] < \infty \text{ for any } 0 < p < \alpha$$
$$E[|X|^p] = \infty \text{ for any } p > \alpha$$

Proposition 29 *Let $X \sim S(\alpha, \beta, c, \mu)$ with $0 < \alpha \leq 2$. Then the location parameter μ equals to the mean.*

If X is symmetric α-stable (SαS), then $X \sim S(\alpha, 0, c, 0)$. Its characteristic function is

$$E[e^{i\theta X}] = e^{-c^{\alpha} |\theta|^{\alpha}}$$

When $c = 1$, X is called standard, and when $\alpha = 2$, a standard SαS distribution is $N(0, 2)$.

Proposition 30 *Let $X \sim S(\alpha^{'}, 0, c, 0)$ with $0 < \alpha^{'} \leq 2$. If $0 < \alpha < \alpha^{'}$, if we choose a random variable $A \sim S(\frac{\alpha}{\alpha^{'}}, 1, (\cos \frac{\pi \alpha}{2\alpha^{'}})^{\frac{\alpha^{'}}{\alpha}}, 0)$, and assume X and A are independent, then*

$$Z = A^{\frac{1}{\alpha^{'}}} X \sim S(\alpha, 0, c, 0)$$

In particular, if $\alpha^{'} = 2$, for a zero mean Gaussian random variable X, and if A is a positive $\frac{\alpha}{2}$-stable random variable independent of X, then $Z = A^{\frac{1}{\alpha^{'}}} X \sim S\alpha S$. That is, every S$\alpha$S random variable is conditionally Gaussian.

We can also express stable distributions in terms of random series. We let $N(t)$ be the number of arrivals within time $[0, t]$, the process $\{N_t\}_{t \geq 0}$ is a homogeneous Poisson process with rate λ, if the inter-arrival times $\tau_{i+1} - \tau_i$, $i \geq 1$ are independent exponentially distributed with mean $\frac{1}{\lambda}$.

Proposition 31 *Let* $\{\tau_i\}$ *denote the arrival times of a Poisson process with rate* 1*, and* $\{R_i\}$ *are i.i.d. random variables, independent of thesequence* $\{\tau_i\}$*. If the series*

$$\sum_{i=1}^{\infty} \tau^{-\frac{1}{\alpha}} R_i$$

concerges a.s., then it converges to a strictly α*-stable random variable.*

Appendix B

Equity miscellaneous

B.1 The Black-Scholes world

B.1.1 A different way of expressing the Black-Scholes formula

B.1.1.1 Setting the rate to zero

If we set the interest rate to zero $r = 0$ and shift the strike to $K^{'} = P(t,T)K$ in the Black-Scholes formula given in Equation (1.2.6), the call price becomes

$$C_{BS}(t,x,K^{'},T;r=0) = xRe(t,T)N(d_1(T-t,xR(t,T),K^{'})) - K^{'}N(d_2(T-t,xR(t,T),K^{'}))$$

with

$$\begin{aligned} d_2(T-t,xR(t,T),K^{'}) &= \frac{1}{\sigma\sqrt{T-t}}\log\frac{xR(t,T)}{K^{'}} - \frac{1}{2}\sigma\sqrt{T-t} \\ &= d_2(T-t,xC(t,T),K^{'}) \end{aligned}$$

Hence, both call prices are identical, that is

$$C_{BS}(t,S_t,K,T) = C_{BS}(t,S_t,K^{'},T,r=0) \tag{B.1.1}$$

B.1.1.2 Normalising the price

Given the forward moneyness $\eta = \frac{KP(t,T)}{xRe(t,T)}$, we let $\xi = \frac{1}{2}\sigma\sqrt{T-t}$ and $m = -\frac{1+\eta}{1-\eta}\log\eta$ such that

$$d_1(T-t,xC(t,T),K) = -\frac{\log\eta}{2\xi} + \xi$$

where $C(t,T) = \frac{Re(t,T)}{B(t,T)}$ such that the forward is $F(t,T) = xC(t,T)$. Following Liang et al. [2009], we set $d_3 = \frac{d_1+d_2}{2} = -\frac{\log\eta}{2\xi}$ and get

$$\begin{aligned} d_1(T-t,xC(t,T),K) &= d_3(T-t,xC(t,T),K) + \xi \\ d_2(T-t,xC(t,T),K) &= d_3(T-t,xC(t,T),K) - \xi \end{aligned}$$

Factorising the call price in Equation (1.2.6) with respect to $xRe(t,T)$ and dividing both side of the equation with $1-\eta$, we get

$$\frac{1}{1-\eta}\frac{C_{BS}(t,x,K,T)}{xRe(t,T)}=\frac{1}{1-\eta}\big(N(d_3(T-t,xC(t,T),K)+\xi)-\eta N(d_3(T-t,xC(t,T),K)-\xi)\big)$$

or in a more concise form

$$\widehat{C}_{BS}(t,x,K,T)=g_1(d_3(T-t,xC(t,T),K)) \tag{B.1.2}$$

B.1.2 Some Greeks

We describe a few Greeks in the Black-Scholes formula that will be used later on to devise our parametric model. The option in the Black-Scholes model is hedged with a portfolio containing

$$\Delta(t,S_t)=\partial_x C_{BS}(t,S_t,K,T)=Re(t,T)N(d_1(T-t,S_te^{(r-q)(T-t)},K)) \tag{B.1.3}$$

stocks, since $\frac{d}{dx}d_i=\frac{1}{x\sigma\sqrt{T-t}}$ for $i=1,2$. Note, $x=S_t$ and $Re(t,T)=e^{-q(T-t)}$. Similarly, the price of a put option is given by

$$\begin{aligned}P_{BS}(t,x,K,T) &= -xe^{-q(T-t)}N(-d_1(T-t,xe^{(r-q)(T-t)},K))+Ke^{-r(T-t)}N(-d_2(T-t,xe^{(r-q)(T-t)},K))\\ &= Ke^{-r(T-t)}N(d_1(T-t,K,xe^{(r-q)(T-t)}))-xe^{-q(T-t)}N(d_2(T-t,K,xe^{(r-q)(T-t)}))\end{aligned}$$

with delta

$$\begin{aligned}\Delta(t,S_t)=\partial_x P_{BS}(t,S_t,K,T) &= -Re(t,T)N(-d_1(T-t,S_te^{(r-q)(T-t)},K))\\ &= -Re(t,T)N(d_2(T-t,K,S_te^{(r-q)(T-t)}))\end{aligned}$$

since

$$d_2(t,y,x,\sigma^2)=-d_1(t,x,y,\sigma^2)$$

Further, the standard normal CDF, $N(\bullet)$, has 2-fold rotational symmetry around the point $(0,\frac{1}{2})$, such that $N(x)=1-N(-x)$. Hence we can rewrite the delta put, $\Delta_P(t,S_t)$, in terms of the delta call as

$$\Delta_P(t,S_t)=\Delta_C(t,S_t)-Re(t,T)$$

Now, the Gamma with respect to x is given by

$$\Gamma(t,S_t)=\frac{\partial^2}{\partial x^2}C_{BS}(t,x,K,T)=Re(t,T)\frac{\partial}{\partial x}N(d_1(T-t,x,K))$$

Using the chain rule, for $x=S_t$ we get the gamma

$$\Gamma(t,S_t)=Re(t,T)\frac{\partial}{\partial d_1}N(d_1(T-t,x,K))\frac{dd_1}{dx}=n(d_1(T-t,x,K))\frac{Re(t,T)}{x\sigma\sqrt{T-t}}$$

and we see that $\Gamma(t,S_t)>0$. The vega in the Black-Scholes model is

$$Vega=xRe(t,T)\sqrt{T-t}n(d_1(T-t,x,K)) \tag{B.1.4}$$

where $n(x)=\frac{\partial}{\partial x}N(x)$ and $N^{''}(x)=n^{'}(x)=-xn(x)$, with the density given by

$$n(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

Further, we have

$$KP(t,T)n(d_2(T-t,x,K)) = xRe(t,T)n(d_1(T-t,x,K)) \tag{B.1.5}$$

so that the vega can be expressed as

$$Vega = KP(t,T)\sqrt{T-t}n(d_2(T-t,x,K))$$

Also, we have the limit cases

$$\begin{aligned}
\lim_{(T-t)\to 0} Vega &= 0 \\
\lim_{(T-t)\to \infty} Vega &= 0
\end{aligned}$$

Thus, the vega is positive or equal to zero. We can express the vega in terms of the gamma as follow

$$Vega = x^2\sigma(T-t)\Gamma(t,S_t) \tag{B.1.6}$$

Differentiating the vega with respect to the stock, we get the Vanna as

$$\frac{\partial}{\partial x} Vega = Vanna = \big(\frac{1}{x} - \frac{d_1}{x\sigma\sqrt{T-t}}\big) Vega \tag{B.1.7}$$

Differentiating one more time the Vega with respect to volatility which is called the Volga, and using the relations $\frac{dd_1}{d\sigma} = -\frac{1}{\sigma}d_2$ or $\frac{dd_2}{d\sigma} = -\frac{1}{\sigma}d_1$, we get

$$\frac{\partial}{\partial \sigma} Vega = Volga = xRe(t,T)\sqrt{T-t}\frac{1}{\sigma}d_1 d_2 n(d_1) = \frac{1}{\sigma}d_1 d_2 Vega \tag{B.1.8}$$

Again, differentiating the Volga with respect to volatility, we get

$$\frac{\partial}{\partial \sigma} Volga = \big(-d_1 d_2 - d_2^2 - d_1^2 + d_1^2 d_2^2\big)\frac{1}{\sigma^2} Vega$$

Since $\frac{d}{dK}d_i = -\frac{1}{K\sigma\sqrt{T-t}}$ for $i = 1, 2$, we differentiate the Vega with respect to the strike K, getting

$$\frac{\partial}{\partial K} Vega = \frac{d_1}{K\sigma\sqrt{T-t}} Vega$$

which we differentiate one more time with respect to the strike, getting

$$\frac{\partial^2}{\partial K^2} Vega = \frac{d_1 d_2}{K^2\sigma^2(T-t)} Vega - \frac{1}{K^2\sigma^2(T-t)} Vega = \frac{(d_1 d_2 - 1)}{K^2\sigma^2(T-t)} Vega$$

To conclude, we need to differentiate the Vega with respect to maturity T, getting

$$\frac{\partial}{\partial T} Vega = -q(T)Vega + \frac{1}{2(T-t)} Vega + d_1\big(\frac{1}{2(T-t)}d_2 - \frac{(r(T)-q(T))}{\sigma\sqrt{T-t}}\big) Vega$$

We now differentiate the call price with respect to the strike

$$\frac{\partial}{\partial K} C_{BS}(t,x,K,T) = -P(t,T)N(d_2(T-t,x,K))$$

Differentiating the call price twice with respect to the strike we get

$$\frac{\partial^2}{\partial K^2} C_{BS}(t,x,K,T) = \frac{P(t,T)}{K\sigma\sqrt{T-t}} n(d_2(T-t,x,k))$$

Using Equation (B.1.5), it can be rewritten as

$$\frac{\partial^2}{\partial K^2} C_{BS}(t,x,K,T) = \frac{xRe(t,T)}{K^2\sigma\sqrt{T-t}} n(d_1(T-t,x,k)) = \frac{1}{K^2\sigma(T-t)} Vega(K,T;\sigma) \tag{B.1.9}$$

Setting $x = S_t$, we differentiate the call price with respect to maturity T getting

$$\frac{\partial}{\partial T} C_{BS}(t,S_t,K,T) = \frac{\sigma}{2(T-t)} Vega + r_T KP(t,T)N(d_2) - q_T xRe(t,T)N(d_1)$$

We define $X_{BS}(t,S_t,K,T)$ as

$$X_{BS}(t,S_t,K,T) = \frac{\sigma}{2(T-t)} Vega + r_T KP(t,T)N(d_2)$$

which is always positive, and rewrite the above partial derivative as

$$\frac{\partial}{\partial T} C_{BS}(t,S_t,K,T) = X_{BS}(t,S_t,K,T) - q_T xRe(t,T)N(d_1)$$

Similarly, the partial derivative of the put price with respect to maturity T is

$$\frac{\partial}{\partial T} P_{BS}(t,S_t,K,T) = -\frac{\sigma}{2(T-t)} Vega + r_T KP(t,T)N(d_1) - q_T xRe(t,T)N(d_2)$$

Since $\frac{d\tau}{dt} = -1$ where $\tau = T-t$, the partial derivative of the price with respect to time t is given by $\frac{\partial C_{BS}(t,S_t,K,T)}{\partial \tau}\frac{d\tau}{dt}$. From the relation

$$r_t S_t \Delta(t,S_t) - r_t C_{BS}(t,S_t,K,T) = r_t KP(t,T)N(d_2)$$

the theta of call option becomes

$$\frac{\partial}{\partial t} C_{BS}(t,S_t,K,T) = -\frac{\sigma}{2(T-t)} Vega - r_t x\Delta(t,S_t) + r_t C_{BS}(t,S_t,K,T) + q_t xRe(t,T)N(d_1)$$

From the Equation (B.1.6) relating the vega with the gamma, we can rewrite the theta as

$$\frac{\partial}{\partial t} C_{BS}(t,S_t,K,T) = -\frac{1}{2}x^2\sigma^2\Gamma(t,S_t) - r_t x\Delta(t,S_t) + r_t C_{BS}(t,S_t,K,T) + q_t xRe(t,T)N(d_1) \tag{B.1.10}$$

Note, we can write the theta as

$$\frac{\partial}{\partial t} C_{BS}(t,S_t,K,T) = -\frac{1}{2}x^2\sigma^2\Gamma(t,S_t) + q_t xRe(t,T)N(d_1) - r_t KP(t,T)N(d_2)$$

and study the ratio

$$R = \frac{q_t Re(t,T)N(d_1) - r_t \frac{K}{x} P(t,T)N(d_2)}{\frac{1}{2}x\sigma^2\Gamma(t,S_t)}$$

Since $N(d_1)$ and $N(d_2)$ take value in $[0,1]$, when $xRe(t,T) = KP(t,T)$ then $d_1 = \frac{1}{2}\sigma\sqrt{T-t}$ and $d_2 = -\frac{1}{2}\sigma\sqrt{T-t}$, and the gamma is at its maximum. In that case the ratio simplifies to

$$R=\frac{q_tRe(t,T)N(\frac{1}{2}\sigma\sqrt{T-t})-r_tRe(t,T)N(-\frac{1}{2}\sigma\sqrt{T-t})}{\frac{1}{2}x\sigma^2\Gamma(t,S_t)}$$

Further assuming that $q_t=r_t$, then $Re(t,T)=P(t,T)$ and the ratio simplifies to

$$R=\frac{q_tRe(t,T)\big(N(\frac{1}{2}\sigma\sqrt{T-t})-N(-\frac{1}{2}\sigma\sqrt{T-t})\big)}{\frac{1}{2}x\sigma^2n(d_1)\frac{Re(t,T)}{x\sigma\sqrt{T-t}}}=\frac{q_tRe(t,T)0.4\sigma\sqrt{T-t}}{\frac{1}{2}x\sigma n(d_1)\frac{Re(t,T)}{\sqrt{T-t}}}=\frac{q_t(T-t)}{\frac{1}{2}Ke^{-0.5d_1^2}}$$

and the numerator is smaller than the denominator. However, when $N(d_1)$ and $N(d_2)$ are both close to 0 or close to 1 the gamma tends to zero. The question being to know which of the numerator or the denominator goes to zero faster. We can then assume that $q_txRe(t,T)N(d_1)-r_tKP(t,T)N(d_2)$ is small compared to $\frac{1}{2}x^2\sigma^2\Gamma(t,S_t)$ so that the theta can be approximated as

$$\frac{\partial}{\partial t}C_{BS}(t,S_t,K,T)\approx-\frac{1}{2}x^2\sigma^2\Gamma(t,S_t)$$

B.2 Modelling discrete dividends

B.2.1 An overview

In presence of discrete dividends, it is natural to assume that the stock price S_t jumps down with the amount of dividend d_i paid at time t_i, while in between dividend dates it follows a continuous process (eg a local volatility model). However, stock prices are no-longer lognormally distributed and there is no convenient closed-form solution for European options. One can either apply a finite difference method scheme or run Monte Carlo simulations in this case, or directly make some adjustments to the Black-Scholes formula. Frishling [2002] gives an overview of the different model assumptions, among which are

- The escrowed model which assumes that the asset price minus the present value of all dividends to be paid until the maturity of the option follows a geometric Brownian motion.
- The forward model which assumes that the asset price plus the forward value of all dividends (from past dividend dates to today) follows a geometric Brownian motion.
- The piecewise lognormal model which assumes that the stock price shows a jump downwards at dividend dates and follows a geometric Brownian motion in between those dates.

Several approaches use the idea of an escrowed dividend process where the stock price process is decomposed into two parts, a risky part and an escrowed dividend part. The idea being to make a simple adjustment to use the BS-formula by replacing the stock price S_0 by the stock price minus the present value of the dividend $S_0-P(0,t_d)d$, where d is the cash dividend at time t_d. The main advantage of the escrowed model and the forward model is that we can use the BS-formula with an adjusted value of the current stock price or strike, respectively. Moreover, for American options it is easy to construct binomial trees which recombines. However, the escrowed approach undervalue call options and lead to arbitrage opportunities (for details see Bos et al. [2003], Haug et al. [2003]). One solution is to consider a modified escrowed model where we assume that the asset price minus the present value of all dividends to be paid in the future follows a geometric Brownian motion. However, the option prices in that setting will depend on the dividends paid after the options have expired. Modifying both the escrowed model and the forward model, Bos et al. [2002] defined a mixture model where one part of the dividends is incorporated in a modified asset price, and the other part in a modified strike price. From a theoretical point of view, the piecewise lognormal model should be preferred even if it is computationaly slower. Considering a piecewise lognormal model, Wilmott et al. [1993] proposed a method based on interpolation steps within a partial differential equation (PDE). Using conditional expectation, Haug et al. [2003]) replaced a multiple integration by a succession of single integrations over Black-Scholes like approximating functions.

Following the same approach as Wilmott et al. [1993], but using a stock-price dependent dividend, Vellekoop et al. [2006] constructed a recombing tree. In any case, when modelling discrete dividends we must face a trade-off between accuracy and time. We beriefly describe some popular models for discrete dividends where a closed-form solution can be derived for European options. For notation purpose, we consider the definition of the discrete dividends given in Section (5.1.1.2).

B.2.2 The HW model

B.2.2.1 The HW spot model

In the Hull-White model the stock price is decomposed into a risky component Z and a riskless one D that correspond to the known dividends during the life of the option. That is, Z excludes all future dividends. It is therefore dependent on the maturity T of the option and can only be used for a single option. In any case, the stock price process should not depend upon the option maturity considered.

***Remark* B.2.1** *It can be used to infer a distribution but one should not diffuse S for exotic pricing but only Z because the local volatility is shifted downwards as maturity passes an ex-dividend dates.*

In that model, the stock price process $(S_t)_{t\in[0,T]}$ is given by the transformation

$$S_t = Z_t + D_{PV}(t;t,T)$$

where, for $T = N\Delta t$, the displaced coefficient is $D_{PV}(t;t,T) = \sum_{t<t_i\leq T} \frac{P(t,t_i)}{Re(t,t_i)} d_{t_i}$. At time $t_0 = 0$, the trajectory of the process Z_t starts at S_0 to which we substract the present value of all the expected future dividend payments up to maturity. The process ends its trajectory at maturity where it is equal the stock price, that is, $Z_T = S_T$. Note, we can not pay a constant dividend d at arbitrarily low stock prices as we can get negative stock prices leading to arbitrage issues. This said, we get the present value of all the future dividend payments up to maturity T described in Section (5.1.1.2), differentiate that function with respect to time and get its dynamics as

$$d\widehat{D}_{PV}(t;t,T) = \mu_t D_{PV}(t;t,T)dt - D_t^{'}dt$$

It implies that at maturity we have $Z_T = S_T$ and we can write the price of a call option as

$$C(t,S_t,K,T) = P(t,T)E_t[(S_T-K)^+] = P(t,T)E_t[(Z_T-K)^+]$$

This is similar to using the transformation in Equation (5.1.1) with $a(t) = 1$ and $b(t) = D_{PV}(t;t,T)$ such that the strike of the call option is $k = K$ since $D_{PV}(T;T,T) = 0$. Assuming that the risky process is lognormally distributed with dynamics given by

$$\frac{dZ_t}{Z_t} = \mu dt + \sigma_Z dW_t$$

the price of the call option becomes

$$C(t,S_t,K,T) = Z_t N(d_1(t,\frac{Z_t}{P(t,T)},K)) - KP(t,T)N(d_2(t,\frac{Z_t}{P(t,T)},K))$$

with volatility σ_Z. Beneder et al. [2001] showed that arbitrage opportunities exist if the volatility surface is continuously interpolated around ex-dividend dates. It was further investigated by Bos et al. [2003] who showed that these interpolations could lead to significant mispricing for American options. In addition, the stock price is no-longer log-normally distributed and has the dynamics

$$dS_t = \mu S_t dt - \overline{D}_t dt + \sigma_Z(S_t - D_{PV}(t;t,T))dW_t \quad \text{(B.2.11)}$$

with local volatility

$$\sigma_S(x,t) = \sigma_Z \frac{x - D_{PV}(t;t,T)}{x} = \sigma_Z(1 - \frac{D_{PV}(t;t,T)}{x})$$

In case where σ_Z is flat then the corresponding stock volatility σ_S is no-longer flat. Since $D_{PV}(t;t,T) \geq 0$ and usually $D_{PV}(t;t,T) \leq S_t$ then $\sigma_S < \sigma_{LN}$ so that the average volatility of the stock S will be lower than the volatility of a process log-normally distributed between dividend dates. Moreover, the entire local volatility surface is shifted downwards as maturity passes an ex-dividend date. As explained by Bos et al. [2003], this behaviour is especially undesirable for path-dependent options. Note, the volatility of the stock converge to that of the risky process when time to maturity is reaching zero, $\lim_{(T-t)\to 0} \sigma_S = \sigma_Z$. Obviously when this model is calibrated to market prices one retrieve the volatility σ_Z of the risky process.

B.2.2.2 The HW forward model

To circumvent the problem of the option maturity in the dynamics of the stock price process, Hull and White considered modelling the forward price. Given the risky process $(Z_t)_{t\in[0,T]}$ defined in Appendix (B.2.2.1), the forward price becomes

$$F(t,T) = Z_t \frac{Re(t,T)}{P(t,T)} = (S_t - D_{PV}(t;t,T))C(t,T) = S_t C(t,T) - D(t,T)$$

Assuming that the forward price is log-normally distributed with dynamics

$$\frac{dF(t,T)}{F(t,T)} = \sigma_F dW_t$$

one can then apply Black-Scholes formula to get the price of a call option

$$C(t,S_t,K,T) = P(t,T)F(t,T)N(d_1(t,F(t,T),K)) - KP(t,T)N(d_2(t,F(t,T),K))$$

B.2.3 The Spot model

An alternative approach described by Bos et al. [2003] is to adjust the stock price and the strike. We assume that the dynamics of the stock price $(S_t)_{0\leq t\leq T}$ in the spot model, and under the risk-neutral measure $\mathbb{Q}$, are given by

$$\begin{aligned} S_t &= C(t_0,t)Y_t - D_t \\ dY_t &= Y_t \sigma_Y(t,Y)dW_Y(t) \\ Y_{t_0} &= S_{t_0} \end{aligned}$$

with solution

$$Y_t = Y_{t_0}\mathcal{E}_Y(t_0,t) \text{ and } \mathcal{E}_Y(t_0,t) = e^{-\frac{1}{2}\int_{t_0}^t \sigma_Y^2(s,Y)ds + \int_{t_0}^t \sigma_Y(s,Y)dW_Y(s)}$$

In that setting the price of a European call option seen at time t_0 with strike K and maturity T is

$$C_S(t_0;K,T) = P(t_0,T)E_{t_0}[(C(t_0,T)Y_T - K^{'})^+]$$

where $K^{'} = K + D_T$. Note, for all time $t < T$ the stock price at maturity is

$$S_T = S_t C(t,T)\mathcal{E}_Y(t,T) - D_T + C(t,T)D_t\mathcal{E}_Y(t,T) = C(t,T)Y(t,T) - D_T + C(t,T)D_t\mathcal{E}_Y(t,T)$$

with $Y(t,T) = S_t \mathcal{E}_Y(t,T)$. Hence, the forward price is given by

$$F(t,T) = E_t[S_T] = C(t,T)Y_t - D(t,T) = C(t,T)\big(Y_t - D_{PV}(t;t,T)\big)$$

since $E_t[\mathcal{E}_Y(t,T)] = 1$ and $D_{PV}(t;t,T) = \frac{D(t,T)}{C(t,T)}$. Hence, we approximate the stock price seen at time t with fixed maturity T by

$$S_T = C(t,T)Y(t,T) - D(t,T)$$

This model is popular mainly due to the fact that when adding the already paid dividends to the strike and considering the special case where the volatility σ_Y is a consant, the call option price becomes

$$C_S(t;K,T) = P(t,T)E_t[(C(t,T)Y_T - K^{'})^+] = S_t Re(t,T)N(d_1) - K^{'}P(t,T)N(d_2)$$

Note, since the spot price follows the transformation in Equation (5.1.1), we can rewrite the price of the call option as

$$C_S(t;K,T) = P(t,T)C(t,T)C_Y(t;K^{''},T) = Re(t,T)E_t[(Y_T - K^{''})^+] = C(t,Y_t,K^{''},T) \quad \text{(B.2.12)}$$

where $K^{''} = \frac{K^{'}}{C(t,T)} = \frac{P(t,T)(K+D(t,T))}{Re(t,T)}$. Hence, to price a call option with strike K in the S-space domain, we need the strike $k = \frac{K^{'}}{C(t,T)}$ in the Y-space domain. Reversing the process, we get $K = C(t,T)K^{''} - D(t,T)$ and the call option price becomes

$$C_Y(t;k,T) = \frac{1}{Re(t,T)} C_S(t; C(t,T)k - D(t,T), T)$$

We can also rewrite the price of the call option as

$$C_S(t;K,T) = E_t[(Re(t,T)Y_T - \widehat{K})^+] = E_t[(Z_T - \widehat{K})^+]$$

where $\widehat{K} = P(t,T)(K + D(t,T))$ and $Z_t = Re(t_0,t)Y_t$.

B.3 Approximating the implied volatility surface in the MixVol model

B.3.1 An approximation to infer the implied volatility

B.3.1.1 The implied volatility

For simplicity of exposition, we let $C_{BS}(t,x,K,T;\sigma)$ be the Black-Scholes call price with volatility σ, and we assume that the market price of an option for strike K and maturity T can be decomposed into a Black-Scoles price around ATM forward plus some extra term $\alpha(K,T)$. That is,

$$C(t,x,K,T) = C_{BS}(t,x,K,T;\Sigma(T,K)) + \alpha(K,T) \quad \text{(B.3.13)}$$

where the volatility $\Sigma(T,K)$ is a function of strike and maturity. The function $\alpha(K,T)$ is a modification of the Black-Scholes formula capable of generating a skew, or a curvature, or both. It must be a convex function with respect to the strike K. For notation purpose, we let the modified price be expressed as

$$y = y(0) + \alpha$$

where $y(0) = C_{BS}(t,x,K,T;\sigma)$ is the Black-Scholes price around-the-money forward. Given the definition of the implied volatility in Section (1.4.1.2), we let $y = C(t,x,K,T)$ be the call price for strike K and maturity T, and we let $g(y) = f^{-1}(y)$ be the inverse function.

***Assumption* B.3.1** *We assume that $y(0)$ dominates α, that is, α is small with respect to y.*

As a result from Assumption (B.3.1), we can approximate the function $g(y)$ with a Taylor series around $y(0)$, getting

$$g(y) = g(y(0)) + (y - y(0))g^{'}(y(0)) + \frac{1}{2}(y - y(0))^2 g^{''}(y(0)) + ... + \frac{1}{n!}(y - y(0))^n g^{(n)}(y(0)) + R_n$$

where R_n is the remainder term. The series becomes

$$g(y) = g(y(0)) + \alpha g^{'}(y(0)) + \frac{1}{2}\alpha^2 g^{''}(y(0)) + ... + \frac{1}{n!}\alpha^n g^{(n)}(y(0)) + R_n$$

Note, this approximation may not satisfies some market conditions for deep in-or-out of the money options. Defining the implied volatility for the price $y(0)$ as

$$\sigma_0 : C_{BS}(K,T;\sigma(K,T)) - y(0) = 0$$

we get the ATM forward volatility $g(y(0)) = \sigma_0$. Consequently, from the derivatives of inverse functions

$$\begin{aligned}
(f^{-1})^{'}(y) &= \frac{1}{f^{'}(f^{-1}(y))} \\
(f^{-1})^{''}(y) &= -\frac{f^{''}(f^{-1}(y))}{[f^{'}(f^{-1}(y))]^3} \\
(f^{-1})^{'''}(y) &= -\frac{\big(f^{'''}(f^{-1}(y)) - 3[(f^{-1})^{''}(y)]^2[f^{'}(f^{-1}(y))]^5\big)}{[f^{'}(f^{-1}(y))]^4} \\
(f^{-1})^{''''}(y) &= -\frac{\big\{f^{''''}(f^{-1}(y)) - 10(f^{-1})^{'''}(y)(f^{-1})^{''}(y)[f^{'}(f^{-1}(y))]^6 + 15[(f^{-1})^{''}(y)]^3[f^{'}(f^{-1}(y))]^7\big\}}{[f^{'}(f^{-1}(y))]^5}
\end{aligned}$$

we can compute the functions

$$\begin{aligned}
g^{'}(y(0)) &= \frac{d}{dy}f^{-1}(y)|_{y=y(0)} = \frac{1}{Vega(\sigma_0)} \\
g^{''}(y(0)) &= \frac{d^2}{dy^2}f^{-1}(y)|_{y=y(0)} = -\frac{d_1 d_2 Vega(\sigma_0)}{\sigma_0}\frac{1}{[Vega(\sigma_0)]^3} = -\frac{d_1 d_2}{\sigma_0}\frac{1}{[Vega(\sigma_0)]^2} \\
g^{'''}(y(0)) &= -\frac{\big(-d_1 d_2 - d_2^2 - d_1^2 + d_1^2 d_2^2\big)\frac{1}{\sigma_0^2}Vega(\sigma_0) - 3\frac{d_1^2 d_2^2}{\sigma_0^2}Vega(\sigma_0)}{[Vega(\sigma_0)]^4} \\
&= \frac{\big(d_1 d_2 + d_2^2 + d_1^2 + 2d_1^2 d_2^2\big)\frac{1}{\sigma_0^2}}{[Vega(\sigma_0)]^3} \\
g^{''''}(y(0)) &= -\frac{\big(6d_1 d_2 + 3d_2^2 + 3d_1^2 - 3d_1 d_2^3 - 3d_1^3 d_2 - 3d_1^2 d_2^2 + d_1^3 d_2^3\big)\frac{1}{\sigma^3}Vega(\sigma_0)}{[Vega(\sigma_0)]^5} \\
&- \frac{10\big(d_1^2 d_2^2 + d_1 d_2^3 + d_1^3 d_2 + 2d_1^3 d_2^3\big)\frac{1}{\sigma_0^3}Vega(\sigma_0)}{[Vega(\sigma_0)]^5} + 15\frac{\frac{d_1^3 d_2^3}{\sigma_0^3}Vega(\sigma_0)}{[Vega(\sigma_0)]^5} \\
&= -\frac{\big(6d_1 d_2 + 3d_2^2 + 3d_1^2 + 7d_1 d_2^3 + 7d_1^3 d_2 + 7d_1^2 d_2^2 + 6d_1^3 d_2^3\big)\frac{1}{\sigma_0^3}}{[Vega(\sigma_0)]^4}
\end{aligned}$$

The Taylor series simplifies to

$$
\begin{aligned}
g(y) &= g(y(0)) + \alpha \frac{1}{Vega(\sigma_0)} - \frac{1}{2}\alpha^2 \frac{d_1(\sigma_0)d_2(\sigma_0)}{\sigma_0} \frac{1}{[Vega(\sigma_0)]^2} + \frac{1}{6}\alpha^3 \frac{\left(d_1 d_2 + d_2^2 + d_1^2 + 2d_1^2 d_2^2\right)\frac{1}{\sigma_0^2}}{[Vega(\sigma_0)]^3} \\
&- \frac{1}{24}\alpha^4 \frac{\left(6d_1 d_2 + 3d_2^2 + 3d_1^2 + 7d_1 d_2^3 + 7d_1^3 d_2 + 7d_1^2 d_2^2 + 6d_1^3 d_2^3\right)\frac{1}{\sigma_0^3}}{[Vega(\sigma_0)]^4} + ... +
\end{aligned}
$$

where $Vega(\sigma_0) = xRe(0,T)\sqrt{T}N'(\widehat{d}_1)$ with $\widehat{d}_1 = \frac{\log\frac{x}{K} + (r-q)T}{\sigma_0\sqrt{T}} + \frac{1}{2}\sigma_0\sqrt{T}$. Replacing the terms in the $g(y)$ expansion with the appropriate Black-Scholes formula, we get the approximated implied volatility

$$
\begin{aligned}
\Sigma(K,T) &\approx \sigma_0 + \frac{1}{Vega(\sigma_0)}\alpha - \frac{1}{2}\frac{d_1(\sigma_0)d_2(\sigma_0)}{\sigma_0}\frac{1}{[Vega(\sigma_0)]^2}\alpha^2 \\
&+ \frac{1}{6}\frac{\left(d_1(\sigma_0)d_2(\sigma_0) + d_2^2(\sigma_0) + d_1^2(\sigma_0) + 2d_1^2(\sigma_0)d_2^2(\sigma_0)\right)\frac{1}{\sigma_0^2}}{[Vega(\sigma_0)]^3}\alpha^3 + ... +
\end{aligned} \tag{B.3.14}
$$

As all the derivatives of inverse functions depend on $f^{-1}(y)$, the Taylor series around $y(0)$ depend crutially on the computation of the ATM forward volatility σ_0.

B.3.1.2 An example

We are going to use the technique described in Appendix (B.3.1.1) to infer the implied volatility from the call option price in Equation (3.1.4). As the first term of the parametric model generates a smile, we are first going to infer its volatility. Given the price $\widehat{y} = (1 + \mu_T)C_{BS}(t_0, S_{t_0}, K, T; \sigma)$ we get the decomposition

$$
\widehat{y} = C_{BS}(t_0, S_{t_0}, K, T; \sigma) + \mu_T C_{BS}(t_0, S_{t_0}, K, T; \sigma)
$$

with $\widehat{\alpha}(K,T) = \mu_T C_{BS}(t_0, S_{t_0}, K, T; \sigma)$. So, given Equation (B.3.14) the implied volatility becomes

$$
\widehat{\Sigma}(K,T) \approx \sigma + \frac{1}{Vega(\sigma)}\mu_T C_{BS}(t_0, S_{t_0}, K, T; \sigma) + ... +
$$

Now, given the price in Equation (3.1.4), we get $y(0) = \widehat{y}$ and $\alpha(K,T)$ is defined in Equation (3.1.5). Again, given the volatility expansion in Equation (B.3.14), the approximated implied volatility becomes

$$
\Sigma(K,T) \approx \widehat{\Sigma}(K,T) + \frac{1}{Vega(\widehat{\Sigma})}\alpha(K,T) + ... +
$$

B.3.2 A simplified Mixture model

B.3.2.1 The model

For simplicity of exposition, we let $C_{BS}(t,x,K,T;\sigma)$ be the Black-Scholes call price with volatility σ and assume that the market price can be decomposed into a linear combination of modified Black-Scholes prices by

$$
C(t,x,K,T) = \sum_{i=1}^{n} a_i C_{BS}(t,x,\overline{K},T;\Sigma_i(K,T)) \tag{B.3.15}
$$

where $\overline{K} = K(1 + \mu_i(T))$ and $\Sigma_i(K,T)$ is a function of strike and maturity. In addition, the weights $a_i > 0$ for $i = 1,..,n$ are positive and such that $\sum_{i=1}^{n} a_i = 1$. That is, the function $C_{BS}(t,x,\overline{K},T;\Sigma_i(K,T))$ is a modification of the Black-Scholes formula capable of generating a skew and curvature. For instance, considering the modified Black-Scholes price in Equation (B.3.13), we can rewrite the market price as

$$C(t,x,K,T)=\sum_{i=1}^{n}a_i\big(C_{BS}(t,x,K,T;\Sigma_i(K,T))+\alpha_i(K,T)\big)$$

Following Appendix (B.3.1.1), we let each modified Black-Scholes price be approximated by

$$y_i=y_i(0)+\alpha_i\ ,\ i=1,...,n$$

where $y_i(0)=C_{BS}(t,x,K,T;\sigma_i)$ is a price around-the-money forward, and such that α_i accounts respectively for the skew and the curvature. Given the definition of the implied volatility in Section (1.4.1.2), we let $y=C(t,x,K,T)$ be the call price and $g(y)=f^{-1}(y)$ be the inverse function. Hence, in our framework we can write

$$y(0)=\sum_{i=1}^{n}a_iy_i(0)\ ,\ \alpha=\sum_{i=1}^{n}a_i\alpha_i$$

such that the call price is

$$y=y(0)+\alpha$$

Note, Assumption (B.3.1) holds.

B.3.2.2 The approximated implied volatility

The first term $y(0)$ of the parametric model is a weighted sum of lognormal prices generating a symmetric smile. Similarly to the example in Appendix (B.3.1.2), we define the implied volatility for the price $y(0)$ as

$$\widehat{\sigma}:C_{BS}(K,T;\widehat{\sigma}(K,T))-y(0)=0$$

When the spot price S_0 is close to the at-the-money forward strike, we can use the property of linearity of the Black-Scholes formula with respect to the volatility (see Equation (1.2.12)) to get the approximated price

$$C_{BS}(t_0,t;\overline{\sigma})\approx\sum_{i=1}^{n}a_iC_{BS}(t_0,t;\sigma_i)$$

where $\overline{\sigma}=\sum_{i=1}^{n}a_i\sigma_i$. That is, the implied volatility $\widehat{\sigma}$ is no-longer a function of the strike, but only a function of time. For notational purpose we let $\overline{y}=C_{BS}(t_0,t;\overline{\sigma})$ and rewrite the model price as

$$y(0)=\overline{y}+\beta$$

where β is the error term

$$\beta=y(0)-\overline{y}=\sum_{i=1}^{n}a_iC_{BS}(t,x,K,T;\sigma_i)-C_{BS}(t_0,t;\overline{\sigma})$$

***Assumption* B.3.2** *We assume that $\overline{y}=C_{BS}(t_0,t;\overline{\sigma})$ dominates β, that is, β is small with respect to $y(0)$.*

So, given the volatility expansion in Equation (B.3.14), the approximated implied volatility for the price $y(0)$ satisfies

$$
\begin{aligned}
\widehat{\sigma}(K,T) &\approx \overline{\sigma} + \frac{1}{vega(\overline{\sigma})}\big(\sum_{i=1}^n a_i C_{BS}(t,x,K,T;\sigma_i) - C_{BS}(t_0,t;\overline{\sigma})\big) \\
&- \frac{1}{2}\frac{d_1(\overline{\sigma})d_2(\overline{\sigma})}{\overline{\sigma}}\frac{1}{[vega(\overline{\sigma})]^2}\big(\sum_{i=1}^n a_i C_{BS}(t,x,K,T;\sigma_i) - C_{BS}(t_0,t;\overline{\sigma})\big)^2 \\
&+ \frac{1}{6}\frac{\big(d_1(\overline{\sigma})d_2(\overline{\sigma}) + d_2^2(\overline{\sigma}) + d_1^2(\overline{\sigma}) + 2d_1^2(\overline{\sigma})d_2^2(\overline{\sigma})\big)\frac{1}{\overline{\sigma}^2}}{[vega(\overline{\sigma})]^3}\big(\sum_{i=1}^n a_i C_{BS}(t,x,K,T;\sigma_i) - C_{BS}(t_0,t;\overline{\sigma})\big)^3 + ... +
\end{aligned}
$$

Note, if the strike is at-the-money forward $K = F(t,T)$, which corresponds to $\eta = 1$, then the implied volatility simplifies to

$$
\widehat{\sigma}(K,T)\big|_{K=F(t,T)} \approx \overline{\sigma}
$$

Knowing the volatility $\widehat{\sigma}(K,T)$ for the price $y(0)$, and given the correction term $\alpha(K,T)$, we can then use Equation (B.3.14) to infer the implied volatility of the price y as

$$
\begin{aligned}
\Sigma(K,T) &\approx \widehat{\sigma}(K,T) + \frac{1}{vega(\widehat{\sigma})}\big(\sum_{i=1}^n a_i\alpha_i\big) - \frac{1}{2}\frac{d_1(\widehat{\sigma})d_2(\widehat{\sigma})}{\widehat{\sigma}}\frac{1}{[vega(\widehat{\sigma})]^2}\big(\sum_{i=1}^n a_i\alpha_i\big)^2 \\
&+ \frac{1}{6}\frac{\big(d_1(\widehat{\sigma})d_2(\widehat{\sigma}) + d_2^2(\widehat{\sigma}) + d_1^2(\widehat{\sigma}) + 2d_1^2(\widehat{\sigma})d_2^2(\widehat{\sigma})\big)\frac{1}{\widehat{\sigma}^2}}{[vega(\widehat{\sigma})]^3}\big(\sum_{i=1}^n a_i\alpha_i\big)^3 + ... +
\end{aligned} \tag{B.3.16}
$$

Given Equation (3.1.5), in the special case where the strike is ATM-forward, we can approximate the implied volatility with

$$
\Sigma(K,T)\big|_{K=F(T)} \approx \overline{\sigma} + \frac{1}{vega(\overline{\sigma})}\big(\sum_{i=1}^n a_i\alpha_i\big) \tag{B.3.17}
$$

B.4 Calculating the correlation term

B.4.1 Correlation : the constant case

The underlying stock price seen at the two different dates T_1 and T_2 is

$$
\begin{aligned}
S_1 &= F_1 e^{\sigma_1 W_1 - \frac{1}{2}\sigma_1^2 T_1} \\
S_2 &= F_2 e^{\sigma_2 W_2 - \frac{1}{2}\sigma_2^2 T_2}
\end{aligned}
$$

where $S_i = S(T_i)$. We only observe in the market the cumulative variance $\sigma_1^2 T_1$ up to T_1 and $\sigma_2^2 T_2$ up to T_2. We therefore need to make an assumption on the shape of the instantaneous variance $\nu(t)$ considering the market constraint on the cumulative variance, that is $\int_0^{T_2}\nu(s)ds = \sigma_2^2 T_2$. We assume piecewise constant instantaneous variance $\nu(t)$ and that the cumulative variance decomposes as

$$
\sigma_2^2 T_2 = \sigma_1^2 T_1 + \sigma_{1,2}^2(T_2 - T_1)
$$

with $\int_0^{T_1}\nu_1 ds = \sigma_1^2 T_1$ and $\int_{T_1}^{T_2}\nu_2 ds = \sigma_{1,2}^2(T_2 - T_1)$ for constant ν_1 and ν_2. Note, we need $\sigma_2^2 T_2 - \sigma_1^2 T_1 > 0$ so that ν_1 must be smaller or equal to $\nu_2 \frac{T_2}{T_1}$ to recover the total cumulative variance up to T_2. Also, there is no reason why we could not have chosen the instantaneous variance to be constant all the way to T_2, $\int_0^{T_2} \nu ds = \sigma_2^2 T_2$ so that we get $\sigma_2^2 T_2 = \sigma_2^2 T_1 + \sigma_2^2 (T_2 - T_1)$ as variance decomposition.

In the case of piecewise constant instantaneous variance, the value $\frac{S_2}{S_1}$ satisfy the equation

$$\frac{S_2}{S_1} = \frac{F_2}{F_1} e^{\sigma_{1,2}(W_2 - W_1) - \frac{1}{2}\sigma_{1,2}^2 (T_2 - T_1)}$$

Multiply by S_1, we get

$$\begin{aligned} S_2 &= \frac{S_2}{S_1} S_1 = \frac{F_2}{F_1} S_1 e^{\sigma_{1,2}(W_2 - W_1) - \frac{1}{2}\sigma_{1,2}^2 (T_2 - T_1)} \\ &= \frac{F_2}{F_1} F_1 e^{\sigma_1 W_1 - \frac{1}{2}\sigma_1^2 T_1} e^{\sigma_{1,2}(W_2 - W_1) - \frac{1}{2}\sigma_{1,2}^2 (T_2 - T_1)} \end{aligned}$$

Setting $\alpha = F_2 e^{-\frac{1}{2}\sigma_1^2 T_1 - \frac{1}{2}\sigma_{1,2}^2 (T_2 - T_1)}$, the stock S_2 simplifies to

$$S_2 = \alpha e^{\sigma_1 W_1 + \sigma_{1,2}(W_2 - W_1)}$$

In the same maner if we put $\beta = F_1 e^{-\frac{1}{2}\sigma_1^2 T_1}$, then S_1 becomes

$$S_1 = \beta e^{\sigma_1 W_1}$$

The correlation between two variables x and y is

$$\rho(x,y) = \frac{cov(x,y)}{\sqrt{var(x) var(y)}}$$

As we are only interested in the Brownians of the underlying stock, we assume independent increments between W_1 and $(W_2 - W_1)$. Hence, the variance and covariance of the logarithm of S_1 and S_2 are given by

$$\begin{aligned} cov(\sigma_1 W_1, \sigma_1 W_1 + \sigma_{1,2}(W_2 - W_1)) &= \sigma_1^2 T_1 \\ var(\sigma_1 W_1) &= \sigma_1^2 T_1 \\ var(\sigma_1 W_1 + \sigma_{1,2}(W_2 - W_1)) &= \sigma_1^2 T_1 + \sigma_{1,2}^2 (T_2 - T_1) \end{aligned}$$

Since we assumed a log-normal underlying stock price, we get

$$\sigma_2^2 T_2 = \sigma_1^2 T_1 + \sigma_{1,2}^2 (T_2 - T_1)$$

so that the correlation under the previous assumptions becomes

$$\rho = \frac{\sigma_1^2 T_1}{\sqrt{\sigma_1^2 T_1 \sigma_2^2 T_2}} = \sqrt{\frac{\sigma_1^2 T_1}{\sigma_2^2 T_2}} = \frac{\sigma_1}{\sigma_2} \sqrt{\frac{T_1}{T_2}}$$

Note, when the instantaneous variance $\nu(t)$ is time-depent this correlation is no longer true and it must be calculated with the integrals.

B.4.2 Correlation : the general case

B.4.2.1 Case 1

The underlying stock price at the two dates T_1 and T_2 satisfies

$$\begin{aligned}
S_1 &= F_1 e^{\int_0^{T_1} \nu(s)dW(s) - \frac{1}{2}\int_0^{T_1} \nu^2(s)ds} \\
S_2 &= F_2 e^{\int_0^{T_2} \nu(s)dW(s) - \frac{1}{2}\int_0^{T_2} \nu^2(s)ds}
\end{aligned}$$

In order to calculate the correlation term we need to compute

$$\begin{aligned}
cov\big(\int_0^{T_1} \nu(s)dW(s), \int_0^{T_2} \nu(s)dW(s)\big) &= \int_0^{T_1 \wedge T_2} \nu^2(s)ds \\
var(\int_0^{T_1} \nu(s)dW(s)) &= \int_0^{T_1} \nu^2(s)ds \\
var(\int_0^{T_2} \nu(s)dW(s)) &= \int_0^{T_2} \nu^2(s)ds
\end{aligned}$$

so that the correlation under the previous assumptions is

$$\rho = \frac{\int_0^{T_1 \wedge T_2} \nu^2(s)ds}{\sqrt{\int_0^{T_1} \nu^2(s)ds \int_0^{T_2} \nu^2(s)ds}} = \sqrt{\frac{\int_0^{T_1} \nu^2(s)ds}{\int_0^{T_2} \nu^2(s)ds}}$$

B.4.2.2 Case 2

The underlying stock price at the two dates T_1 and T_2 satisfies

$$\begin{aligned}
S_1 &= F_1 e^{\int_0^{T_1} \nu_1(s)dW(s) - \frac{1}{2}\int_0^{T_1} \nu_1^2(s)ds} \\
S_2 &= F_2 e^{\int_0^{T_2} \nu_2(s)dW(s) - \frac{1}{2}\int_0^{T_2} \nu_2^2(s)ds}
\end{aligned}$$

In order to calculate the correlation term we need to compute

$$\begin{aligned}
cov\big(\int_0^{T_1} \nu_1(s)dW(s), \int_0^{T_2} \nu_2(s)dW(s)\big) &= \int_0^{T_1 \wedge T_2} \nu_1(s)\nu_2(s)ds \\
var(\int_0^{T_1} \nu_1(s)dW(s)) &= \int_0^{T_1} \nu_1^2(s)ds \\
var(\int_0^{T_2} \nu_2(s)dW(s)) &= \int_0^{T_2} \nu_2^2(s)ds
\end{aligned}$$

so that the correlation under the previous assumptions is

$$\rho = \frac{\int_0^{T_1 \wedge T_2} \nu_1(s)\nu_2(s)ds}{\sqrt{\int_0^{T_1} \nu_1^2(s)ds \int_0^{T_2} \nu_2^2(s)ds}}$$

B.4.3 Decorrelating Brownian motions

Assuming non-degeneracy condition, the correlated Brownians increments $dZ_l(t)$ for $l = s, d, f$ are defined as a linear combination of three independent orthogonal increments dW_j

$$dZ_l(t) = \sum_{j=1}^{3} \alpha_{lj} dW_j \text{ for } l = s, d, f \tag{B.4.18}$$

We define the matrix A, the square root of the correlation matrix, as

$$A = \begin{pmatrix} \alpha_{f1} & \alpha_{f2} & \alpha_{f3} \\ \alpha_{d1} & \alpha_{d2} & \alpha_{d3} \\ \alpha_{s1} & \alpha_{s2} & \alpha_{s3} \end{pmatrix}$$

such that the correlation matrix is

$$AA^\top = \begin{pmatrix} 1 & \rho_{fd} & \rho_{fs} \\ \rho_{fd} & 1 & \rho_{sd} \\ \rho_{fs} & \rho_{sd} & 1 \end{pmatrix}$$

where the matrix $A^\top$ is the transpose of the matrix A.

Remark **B.4.1** *The matrix A is invertible. The symmetric, positive definite properties of the correlation matrix makes it special to factorise and the Cholesky decomposition can be applied.*

In particular, if we choose W_1, W_2 and W_3 such that $\alpha_{d1} = \alpha_{d2} = \alpha_{f1} = 0$, then [1]

$$A = \begin{pmatrix} 0 & \sqrt{1-\rho_{fd}^2} & \rho_{fd} \\ 0 & 0 & 1 \\ \sqrt{\frac{1-\rho_{fd}^2-\rho_{sd}^2-\rho_{fs}^2+2\rho_{fs}\rho_{sd}\rho_{fd}}{1-\rho_{fd}^2}} & \frac{\rho_{fs}-\rho_{fd}\rho_{sd}}{\sqrt{1-\rho_{fd}^2}} & \rho_{sd} \end{pmatrix}$$

B.5 Numerical tools and others

B.5.1 Approximating the normal functions

There is no closed-form solution to the cumulative standard normal integral function

$$N(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{y^2}{2}} dy$$

and a numerical approximation must be used. Common practice is to use an exponential and a fifth degree polynomial, see Abramowitz et al. [1974]. The problem of finding an accurate cumulative normal approximation appeared when one tried to find an accurate approximation to a n-variate cumulative function. This is because most approximations will use the $n-1$-variate. Hart [1968] proposed to use high degree rational functions obtaining double precision accuracy throughout the real line. An implementation is given by West [2004]. Also, Genz [2004] provides accurate algorithms for the computation of bivariate and trivariate normal and t probabilities for rectangles. The implementation is available from the author's website

`www.math.wsu.edu/faculty/genz/homepage`

[1] $\sum_{j=1}^{3} \alpha_{sj}\alpha_{dj} = \rho_{sd}$

where one can get the normal distribution probabilities accurate to $1e-15$ proposed by Schonfelder [1978]. Getting double precision accuracy on the cumulative normal approximation, we can refine to full machine precision the Moro approximation of the inverse normal cumulative distribution function. The Gaussian function being easily differentiable, one can use the Newton's second order method or the Halley's method (third order), see Acklam [2004].

B.5.2 Pseudorandom number generators

In order to solve the non-linear programming problem under constraints with the DE algorithm, we need to generate pseudorandom numbers. We let the reader refer to L'Ecuyer [2007] where a review of the basic principles underlying the design of a uniform random number generators is done. For simplicity of use we recommand one of the random number generators introduced by L'Ecuyer [1998] such as

- MRG32ka, it has two components of order three with period length $\approx 2^{191}$
- MRG32k5a, it has two components of order five with period length $\approx 2^{319}$
- MRG63k3a, it is a 64-bit integer arithmetic with two components of order three with period length $\approx 2^{377}$

Alternatively, one can use a Linear Feedback Shift (LFSR) or Tausworthe random number generators which are based on linear recurrences modulo two with primitive characteristic polynomials. For details, see L'Ecuyer [1991] and Panneton et al. [2006]. The implementation is available from the author's website

`http://www.iro.umontreal.ca/~lecuyer`

L'Ecuyer [1996b] proposed ME-CF combined LFSR generators with length $L=32$ and $L=64$, whose components have recurrences with primitive trinomials of the form $P_j(z)=z^{k_j}-z^{q_j}-1$ with $0<2_{q_j}<k_j$, and with step size s_j satisfying $0<s_j\leq k_j-q_j<k_j\leq L$ and $gcd(s_j,2^{k_j}-1)$. Since a large number of generators had good properties, L'Ecuyer [1999b] performed extensive computer searches and introduced specific instances of such generators. For $L=32$, the generators have period lengths $(2^{31}-1)(2^{29}-1)(2^{28}-1)(2^{25}-1)\approx 2^{113}$ and characteristic polynomials of degree 113. The procedure LFSR113 has a period $\rho\approx 2^{113}$ and the procedure LFSR258 has a period $\rho\approx 2^{258}$.

Using the powers-of-2 decomposition method, L'Ecuyer et al. [2000b] proposed CMRGs that are faster for an equivalent statistical quality. Considering $a=\pm 2^q\pm 2^r$, the product of x by each power of 2 can be implemented by a left shift of the binary representation of x, and the product ax is computed by adding /or subtracting. Combining MRG with $J=2$ components of order $k=3$ with parameters defined such that each component has only two nonzero coefficients, one of the form $a_{ij}=2^q$ and the other one of the form $a_{ij}=2^q+1$, they obtained the MRG31k3p. It has two distinct cycles of length $\rho=\frac{m_1m_2}{2}\approx 2^{185}$ and provided very good results in terms of speed compared to the other MRGs having similar periods.

B.5.3 About continuous functions

Definition B.5.1 *Given $f:D\to\mathbb{R}$ a function. We say that f is continuous if it is continuous on all point of D.*

If f and g are continuous on D, then $f+g$ and fg are continuous on D, and $\frac{1}{f}$ is continuous wherever it is defined. The function $|f|$ is also continuous on D.

Theorem B.5.1 *de Rolle*
Given $f:[a,b]\to\mathbb{R}$ an application satisfying

1. *f is continuous on $[a,b]$.*

2. *f is differentiable on $]a, b[$.*

3. *$f(a) = f(b)$*

then there exists $c \in]a, b[$ such that $f^{'} = 0$.

Theorem B.5.2 *Finite increments*
Given $f : [a, b] \to \mathbb{R}$ a continuous application on $[a, b]$ and differentiable on $]a, b[$. Then

$$\exists c \in]a, b[\,,\; f^{'}(c) = \frac{f(b) - f(a)}{b - a}$$

Theorem B.5.3 *Given $f : I \to \mathbb{R}$ an application differentiable on I. Then the following statements are equivalent*

1. *f is convex.*

2. *$f^{'}$ is increasing.*

3. *the function f is above of its tangents.*

Theorem B.5.4 *Given $f(x)$ is a function twice continuously differentiable on an interval I. The function is*

- *convex if $f^{''}(x) > 0$ for all $x \in I$.*
- *concave if $f^{''}(x) < 0$ for all $x \in I$.*

The first derivative informs us of the slope of the tangent of a function, and second derivative shows us how it is curved. Given the function $f(x)$ and $x = x^*$ a static point of the function. We obtain a static point when $f^{'}(x) = 0$. Then, $f(x^*)$ is

- a local minimum if the function is convex at x^*, that is, $f^{''}(x^*) > 0$.
- a local maximum if the function is concave at x^*, that is, $f^{''}(x^*) < 0$.

Bibliography

[1974] Abramowitz M., Stegun I.A., Handbook of mathematical functions, with formulas, graphs and mathematical tables. Dover.

[2004] Acklam P.J., An algorithm for computing the inverse normal cumulative distribution function. http://home.online.no/ pjacklam, University of Oslo.

[2012] Ahdida A., Alfonsi A., A mean-reverting SDE on correlation matrices. Cornell University Library.

[2005] Ahmad R., Wilmott P., Which free lunch would you like today, sir?: Delta hedging, volatility arbitrage and optimal portfolios. *Wilmott Magazine*, **81**, pp 64–79.

[1998] Ait-Sahalia Y., Lo A.W., Nonparametric estimation of state-price densities implicit in financial asset prices. *Journal of Finance*, **53**, pp 499–547.

[2001] Alexander C., Market models - a guide to financial data analysis. John Wiley and Sons, New York.

[2003] Alexander C., Normal mixture diffusion with uncertain volatility: Modelling short- and long-term smile effects. ISMA Centre, School of Business, University of Reading, UK.

[2004] Alexander C., Nogueira L.M., Hedging with stochastic local volatility. ISMA, Centre Discussion Paper in Finance, University of Reading **2004-11**.

[2004b] Alexander C., Nogueira L.M., Bivariate normal mixture spread option valuation. *Quantitative Finance*, **4**, (6), pp 637–648.

[2010] Alexander C., Rubinov A., Kalepky M., Leontsinis S., Hedging with stochastic local volatility. ICMA, Centre Discussion Paper in Finance, University of Reading.

[2004] Alos E., A generalization of Hull and White formula and applications to option pricing approximation. Working Paper, Universitat Pompeu Fabra, February 4.

[2007] Alos E., Ewald CO., Malliavin differentiability of the Heston volatility and applications to option pricing. Working Paper, Universitat Pompeu Fabra, University of St.Andrews, May 14.

[1998] Andersen L., Andreasen J., Volatility skews and extensions of the libor market model. Working Paper, August.

[2002] Anh V.V., Heyde C.C., Leonenko N.N., Dynamic models of long memory processes driven by Levy noise. *Journal of Applied Probability*, **39**, pp 730–747.

[1993] Ansel J-P., Stricker C., Decomposition de Kunita-Watanabe. *Seminaire de probabilites*, **27**, pp 30–32, Springer-Verlag.

[2013] Arismendi J.C., Multivariate truncated moments. *Journal of Multivariate Analysis*, **117**, pp 41–75.

[2014] Arismendi J.C., A multi-asset option approximation for general stochastic processes. Working Paper, ICMA Centre, University of Reading.

[2014] Arismendi J.C., Prokopczuk M., An analytic approximation of the implied risk-neutral density of American multi-asset options. Working Paper, ICMA Centre, University of Reading.

[2002] Avellaneda M., Empirical aspects of dispersion trading in US equity markets. Petit Dejeuner de la Finance.

[2002b] Avellaneda M., Boyer-Olson D., Busca J., Friz P., Reconstructing volatility. *Risk*, **39**, pp 87–91.

[2005] Avellaneda M., From SABR to geodesics. Conference presentation slides.

[2004] Ayache E., Henrotte P., Nassar S., Wang X., Can anyone solve the smile problem? *Wilmott Magazine*.

[1900] Bachelier L., Théorie de la spéculation. *Ann. Sci, ENS*, **17**, pp 21–86.

[1997] Backus D., Foresi S., Li K., Wu L., Accounting for biases in Black-Scholes. Working Paper, Stern School of Business, New York University.

[2001] Bacry E., Delour J., Muzy J-F., Multifractal random walk. *Physical Review E*, **64**, 026103–026106.

[1997] Bahra B., Implied risk-neutral probability density functions from option prices: Theory and application. Working Paper, Bank of England.

[1996] Baillie R.T., Long memory processes and fractional integration in econometrics. *Journal of Econometrics*, **73**, pp 5–59.

[1997] Bakshi G., Cao C., Chen Z., Empirical performance of alternative option pricing models. *Journal of Finance*, **52**, (5), pp 2003–2049.

[2003a] Bakshi G., Kapadia N., Volatility risk premium embedded in individual equity options: Some new insights. *Journal of Derivatives*, pp 45–54.

[2003b] Bakshi G., Kapadia N., Madan D., Stock return characteristics, skew laws, and the differential pricing of individual equity options. *Review of Financial Studies*, **16**, pp 101–143.

[1991] Barabasi A-L., Vicsek T., Multifractality of self-affine fractals. *Physical Review A*, **44**, (4), pp 2730–2733.

[1987] Barone-Adesi G., Whaley R.E., Efficient analytic approximation of American option values. *Journal of Finance*, **42**, pp 301–320.

[2009] Barrieu P., El Karoui N., Indifference Pricing: Theory and Applications, Chapter 3: Pricing, Hedging, and Designing Derivatives with Risk Measures. *Princeton University Press*, ed. Carmona R., Cinlar E., pp 77–146.

[2011] Bartkiewicz K., Jakubowski A., Mikosch T., Wintenberger O., Stable limits for sums of dependent infinite variance random variables. *Probability Theory and Related Fields*, **150**, (3), pp 337–372.

[2012] Barunik J., Understanding the source of multifractality in financial markets. *Physica A*, **391**, (17), pp 4234–4251.

[2000] Bates D., Post-87 crash fears in the S & P 500 futures option market. *Journal of Econometrics*, **94**, pp 181–238.

[2001] Baude S., GRM smile model, Rapport de stage de fin d'etude de SUPELEC.

[2002] Baude S., Roubinet C-H., GRM smile model for the pricing of regular and exotic options on equities, Working Paper, GRM, Credit Lyonnais, October.

[2008] Bekiros S.D., Dimitris A.G., The extreme-value dependence of Asia-Pacific equity markets. *Journal of Multinational Financial Management*, **18**, pp 197–208.

[2000] Bellamy N., Jeanblanc M., Incompleteness of markets driven by a mixed diffusion. *Finance and Stochastics*, **4**, Number 2.

[2001] Beneder R., Vorst T., Options on dividends paying stocks. International Conference on Mathematical Finance, Shanghai.

[2001] Benhamou E., Delta equivalent asset flows. Swaps Strategy, Goldman Sachs International, London.

[1984] Bensoussan A., On the theory of option pricing. *Acta Applicandae Mathematicae*, **2**, pp 139–158.

[2002] Berestycki H., Busca J., Florent I., Asymptotics and calibration of local volatility models. *Quant. Finance*, **2**, pp 61–69.

[2004] Bergenthum J., Ruschendorf L., Comparison of option prices in semimartingale models. University of Freiburg, Eckerstr.

[1996] Bergman Y.Z., Grundy B.D., Wiener Z., General properties of option prices. *Journal of Finance*, **51**, pp 1573–1610.

[2006] Bermudez A., Buehler H., Ferraris A., Jordinson C., Lamnouar A., Overhaus M., Equity hybrid derivatives. Wiley.

[1713] Bernoulli J., Ars conjectandi, opus posthumum. Accedit Tractatus de seriebus infinitis, et epistola gallice scripta de ludo pilae reticularis, Basel: Thurneysen Brothers.

[1983] Bhattacharya R., Gupta V., Waymire E., The Hurst effect under trends. *Journal of Applied Probability*, **20**, 3, pp 649–662.

[2003] Bingham N., Kiesel R., Risk-neutral valuation, pricing and hedging of financial derivatives. Springer Finance Textbook, Second Edition.

[1993] Bjerksund P., Stensland G., Closed-form approximation of American options. *Scandinavian Journal of Management*, **9**, pp 87–99.

[2006] Bjerksund P., Stensland G., Closed form spread option valuation. NHH Dept. of Finance & Management Science Discussion Paper No. 2006/20.

[2001] Blacher G., A new approach for designing and calibrating stochastic volatility models for optimal delta-vega hedging of exotic options. Conference presentation at Global Derivatives and Risk Management, Juan-les-Pins.

[1973] Black F., Scholes M., The pricing of options and corporate liabilities. *Journal of Political Economy*, **81**, 3, pp 637–659.

[1975] Black F., Fact and fantasy in the use of options. *Financial Analysts Journal*, pp 36–72.

[2002] Bloch D., Miralles P., Statistical dynamics of the smile. Technical Report, DrKW.

[2007] Bloch D., Fast calibration of the Affine and quadratic models. Working Paper, Universite Paris 6 Pierre et Marie Curie.

[2008] Bloch D., Expanding forward starting options. Working Paper, Universite Paris 6 Pierre et Marie Curie.

[2010a] Bloch D., A note on calibration of Markov processes. *Wilmott Journal*, **2**, (2), pp 61–96.

[2010b] Bloch D., A practical guide to implied and local volatility. Working Paper, Mizuho Securities.

[2011] Bloch D., Coello Coello C.A., Smiling at evolution. *Applied Soft Computing*, **11**, (8), pp. 5724–5734.

[2012a] Bloch D., From implied to local volatility surface. Working Paper, University of Paris 6 Pierre et Marie Curie.

[2014] Bloch D., A practical guide to quantitative portfolio trading. Working Paper, University of Paris 6 Pierre et Marie Curie.

[2012a] Bloch D., From implied to local volatility surface. Working Paper, University of Paris 6 Pierre et Marie Curie.

[2004] Bollen N.P., Whaley R.E., Does net buying pressure affect the shape of implied volatility functions? *Journal of Finance*, **59**, pp 711–753.

[2003] Bondarenko O., Estimation of risk-neutral densities using positive convolution approximation. *Journal of Econometrics*, **116**, pp 85–112.

[2005] Borak S., Hardle W., Weron R., Stable distributions. SFB 649 Discussion Paper 2005-008, Humboldt-Universitat zu Berlin, Germany.

[2002] Bos R., Vandermark S., Finessing fixed dividends. *Risk*, **15**, (9), pp 157–158.

[2003] Bos R., Gairat A., Shepeleva A., Dealing with discrete dividends. *Risk*, **16**, (1), pp 109–112.

[2006] Bossu S., A new approach for modelling and pricing correlation swaps in equity derivatives. Global Derivatives and Risk Management.

[2010] Bouzoubaa M., Osseiran A., Exotic options and hybrids: A guide to structuring, pricing and trading. Wiley Finance, John Wiley & Sons.

[1988] Boyle P.P., A lattice framework for option pricing with two state variables. *Journal of Financial and Quantitative Analysis*, **23**, pp 1–12.

[1989] Boyle P.P., Evnine J., Gibbs S., Numerical evaluation of multivariate contingent claims. *Review of Financial Studies*, **2**, pp 241–250.

[1978] Breeden D, Litzenberger R., Prices of state contingent claims implicit in option prices. *Journal of Business*, **51**, pp 621–652.

[1987] Brenner M., Courtadon G., Subrahmanyan M.G., The valuation of stock index options. Working Paper, New York University.

[1988] Brenner M., Subrahmanyan M.G., A simple formula to compute the implied standard deviation. *Financial Analysts Journal*, **44**, No. 5, pp 80–83.

[2000] Brigo V, Mercurio F., A mixed-up smile. *Risk*, September, pp 123–126.

[2000b] Brigo V, Mercurio F., Rapisarda F., Lognormal-mixture dynamics and calibration to market volatility smiles. Product and Business Development Group, Banca IMI, San Paolo IMI Group, Milan, Italy.

[2004] Brigo V, Mercurio F., Rapisarda F., Scotti R., Approximated moment-matching dynamics for basket-options pricing. *Quantitative Finance*, **4**, pp 1–16.

[1997] Britten-Jones M, Schaefer S., Nonlinear Value-at-Risk: The distribution of a quadratic approximation to portfolio value. London Business School, London.

[1996] Broadie M., Detemple J., American option valuation: New bounds, approximations, and a comparison of existing methods. *Review of Financial Studies*, **9**, pp 1211–1250.

[2008] Buehler H., Volatility and dividends: Volatility modelling with cash dividends and simple credit risk. Working Paper, Institut fur Mathematik, MA 7-4, 10623 Berlin, Germany.

[2004] Cajueiro D.O., Tabak B.M., The Hurst exponent over time: Testing the assertion that emerging markets are becoming more efficient. *Physica A*, **336**, pp 521–537.

[2007] Cajueiro D.O., Tabak B.M., Long-range dependence and multifractality in the term structure of LIBOR interest rates. *Physica A*, **373**, pp 603–614.

[2001] Calvet L., Fisher A., Forecasting multifractal volatility. *Journal of Econometrics*, **105**, pp 27–58.

[2002] Calvet L., Fisher A., Multifractality in asset returns: Theory and evidence. *The Review of Economics and Statistics*, **84**, (3), pp 381–406.

[2004] Calvet L., Fisher A., Regime-switching and the estimation of multifractal processes. *Journal of Financial Econometrics*, **2**, pp 44–83.

[2004] Carbone A., Castelli G., Stanley H., Time-dependent Hurst exponents in financial time series. *Physica A*, **344**, pp 267–271.

[2003] Carmona R., Durrleman V., Pricing and hedging spread options. *SIAM Review*, **45**, pp 627–685.

[2009] Carmona R., Cinlar E., Indifference Pricing: Theory and Applications. *Princeton University Press*.

[2011] Carmona R., Sun Y., Implied and local correlations from spread options. *Princeton University Press*.

[1998a] Carr P., Madan D., Option valuation using the Fast Fourier Transform. Morgan Stanley, University of Maryland.

[1998b] Carr P., Madan D., Towards a theory of volatility trading. in R. A. Jarrow ed. *volatility : new estimation techniques for pricing derivatives*, chap. 29, pp 417–427.

[2005] Carr P., Verma A., Potential problems in market models of implied volatility. Working Paper.

[2005b] Carr P., Wu L., A note on sufficient conditions for no arbitrage. *Finance Research Letters*, **2**, pp 125–130.

[2014] Chance D.M., Hanson T.A., Li W., Muthuswamy J., A bias in the volatility smile. Working Paper.

[2004] Cherubini U., Luciano E., Vecchiato W., Copula methods in finance. The Wiley Finance Series, Wiley.

[1997] Chriss N., Black-Scholes and beyond: Option pricing models. Irwin Professional Publishing, Chicago.

[1973] Clark P.K., A subordinated stochastic process model with finite variance for speculative prices. *Econometrica*, **41**, (1), pp 135–155.

[2006] Clegg R.G., A practical guide to measuring the Hurst parameter. *International Journal of Simulation: Systems, Science and Technology*, **7**, (2), pp 3–14.

[1999] Clemen R., Reilly T., Correlations and copulas for decision and risk analysis. *Management Science*, **45**, pp 208–224.

[1999] Coles S., Hefferman J., Tawn J.A., Dependence measure for extreme value analyses. *Extremes*, **2**, pp 339–365.

[2002] Cont R., da Fonseca J., Dynamics of implied volatility surfaces. *Quantitative Finance*, **2**, pp 45–60.

[2003] Cont R., Tankov P., Financial modelling with jump processes. Chapman & Hall, CRC Press.

[2005] Cont R., Model uncertainty and its impact on the pricing of derivative instruments. CMAP, Ecole Polytechnique, F-91128 Palaiseau, France.

[2008] Cont R., Deguest R., Extracting implied correlation matrices from index option prices: A statistical approach. finance-innovation.org.

[1996] Corrado C., Su T., Skewness and kurtosis in $S\&P$ 500 index returns implied by option prices. *Journal of Financial Research*, **19**, (2), pp 175–192.

[1996b] Corrado C.J., Miller Jr. T.W., A note on a simple, accurate formula to compute implied standard deviations. *Journal of Banking and Finance*, **20**, pp 595–603.

[1996c] Corrado C., Su T., $S\&P$ 500 index option tests of Jarrow and Rudd's approximate option valuation formula. *Journal of Futures Markets*, **16**, (6), pp 611–629.

[1997] Corrado C.J., Su T., Implied volatility skews and stock index skewness and kurtosis implied by $S\&P$ 500 index option prices. *Journal of Derivatives*, **4**, pp 8–19.

[2007] Corrado C., The hidden martingale restriction in Gram-Charlier option prices. *Journal of Futures Markets*, **27**, (6), pp 517–534.

[2003] Costa R.L., Vasconcelos G.L., Long-range correlations and nonstationarity in the Brazilian stock market. *Physica A*, **329**, pp 231–248.

[1976] Cox J.C., Ross S., The valuation of options for alternative stochastic processes. *Journal of Financial Economics*, **3** (January-March), pp 145–166.

[1979] Cox J.C., Ross S.A., Rubinstein M., Option pricing: A simplified approach. *Journal of Financial Economics*, **7**, (3), pp 229–263.

[2005] Cox A., Hobson D., Local martingales, bubbles and option prices. *Finance and Stochastics*, **9**, (4), pp 477–492.

[1957] Cramer H., Mathematical methods of statistics. Princeton University Press, Princeton.

[2004] Crepey S., Delta-hedging vega risk? *Quantitative Finance*, **4**, pp 559–579.

[2011] Cui C., Frank D., Equity implied volatility surface. Quantitative Research and Development, Equities Team, Bloomberg.

[1988] Curtis C.E., Carriker G.L., Estimating implied volatility directly from "nearest-to-the-money" commodity option premiums. Working Paper 081588, Clemson University.

[2008] Czarnecki L., Grech D., Pamula G., Comparison study of global and local approches describing critical phenomena on the Polish stock exchange market. *Physica A: Statistical Mechanics and its Applications*, **387**, (29), pp 6601–6811.

[2001] Dacorogna M.M., Gencay R., Muller U.A., Olsen R.B., Pictet O.V., An introduction to high frequency finance. Academic Press, San Diego, CA.

[2006] Daglish T., Hull J., Suo W., Volatility surfaces: Theory, rules of thumb and empirical evidence. Working Paper, August.

[1995] Davis R.A., Hsing T., Point process and partial sum convergence for weakly dependent random variables with infinite variance. *Ann. Prob.*, **23**, pp 879–917.

[2007] Davis M.H.A., Hobson D.G., The range of traded option prices. *Mathematical Finance*, **17**, (1), pp 1–14.

[2002] Deelstra G., Liinev J., Vanmaele M., Option pricing: A simplified approach. Working Paper, Ghent University.

[2014] Delanoe P., Local correlation with local vol and stochastic vol. Working Paper, HSBC Equity Derivatives.

[2002] Delbaen F., Shirakawa H., An interest rate model with upper and lower bounds. *Asia-Pacific Financial Markets*, **9**, pp 191–209.

[2004] Delbaen F., Schachermayer W., What is a free lunch? *Notices of the AMS*, **51**, (5), pp 526–528.

[2008] Deng Q., Volatility dispersion trading. Working Paper, University of Illinois at Urbana-Champaign.

[1994] Derman E., Kani I., The volatility smile and its implied tree. Goldman Sachs Quantitative Strategies Reports.

[1999] Derman E., Regimes of volatility. *Risk*, **7**, (2).

[2008] Derman E., Lecture Notes 1–12.

[1993] Ding Z., Granger C.W.J., Engle R.F., A long memory property of stock market returns and a new model. *Journal of the Empirical Finance*, **1**, pp 83–106.

[2005] Driessen J., Maenhout P., Vilkov G., Option-implied correlations and the price of correlation risk. University of Amsterdam and INSEAD. Amsterdam and Fontainbleau.

[1996] Duffie D., Kan R., A yield factor model of interest rates. *Mathematical Finance*, **6**, (4), pp 379–406.

[1997] Duffie D., Pan J., An overview of Value at Risk. *Journal of Derivatives*, pp 7-49.

[2000] Duffie D., Pan J., Singleton K., Transform analysis and asset pricing for affine jump-diffusions. *Econometrica*, **68**, (6), pp 1343–1376.

[2001] Dufresne D., The integrated square-root process. Research Paper Number **90**, University of Montreal.

[1994] Dupire B., Pricing with a smile. *Risk*, **7**, pp 18–20.

[1996] Dupire B., A unified theory of volatility. Working Paper, January.

[2003] Dupire B., New Advances in Volatility Modelling and Trading. Risk Europe 2003, Paris.

[2003] Durrleman V., A note on initial volatility surface. Working Paper.

[2006] Ekstrom E., Tysk J., Convexity preserving jump-diffusion models for option pricing. *Journal of Mathematical Analysis and Applications*, **330**, pp 715–728.

[1989] El Karoui N., Rochet JC., A pricing formula for options on coupon-bonds. Cahier **8925**, October.

[1998] El Karoui N., Jeanblanc M., Shreve S., Robustness of the BS formula. *Mathematical Finance*, **8**, pp 93-126.

[1998] Embrechts P., Resnick S., Samorodnitsky G., Living on the edge. *Risk Magazine*, **11**, (1), pp 96-100.

[1999] Embrechts P., Resnick S., Samorodnitsky G., Extreme value theory as a risk management tool. *North American Actuarial Journal*, **26**, pp 30-41.

[1982] Engle R.F., Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. *Econometrica*, **50**, pp 987–1008.

[1965a] Fama E.F., The behavior of stock-market prices. *Journal of Business*, **38**, (1), pp 34–105.

[1970] Fama E.F., Efficient capital markets: A review of theory and empirical work. *Journal of Finance*, **25**, (2), pp 383–417.

[2005] Fengler M.R., Semiparametric modeling of implied volatility, *Springer Finance*, Springer-Verlag Berlin Heidelberg.

[2009] Fengler M.R., Herwartz H., Menn C., Werner C., Explaining the index skew via stochastic correlation models. Technical report, EEA-ESEM.

[2009] Fengler M.R., Arbitrage-free smoothing of the implied volatility surface. *Quantitative Finance*, **9**, (4), pp. 417–428.

[2005] Ferson S., Nelsen R., Hajagos J., Berleant D., Zhang J., Tucker W.T., Ginzburg L., Oberkampf W.L., Myths about correlations and dependencies and their implications for risk analysis. Working Paper.

[1985] Figlewski S., Arbitrage-based pricing of stock index options. Working Paper, New York University.

[1989] Figlewski S., What does an option pricing model tell us about option prices? *Financial Analysts Journal*, **45**, pp 12–15.

[2002] Flamouris D., Giamouridis D., Estimating implied pdfs from American options on futures: A new semiparametric approach. *Journal of Futures Markets*, **22**, (1), pp. 1–30.

[1990] Follmer H., Schweizer M., Hedging of contingent claims under incomplete information. In M.H.A. Davis and R.J. Elliott, editors, *Applied Stochastic Analysis*, pp 389–414, Gordon and Breach, London.

[2004] Follmer H., Schied A., Stochastic finance: An introduction in discrete time. second revised and extended ed., Walter de Gruyter & Co., Berlin.

[2003] Fouque J.P., Papanicolaou G., Sircar R., Solna K., Multiscale stochastic volatility asymptotics. *SIAM J. Multiscale Modeling and Simulation*, **2**, (1), pp 22–42.

[2002] Frishling V., A discrete question. *Risk Magazine*, **15**, pp 115–116.

[2004] Ganatra S., Implementation of variance swaps in dispersion trading strategies. Master Thesis, Cass Business School, London.

[2005] Garel B., d'Estampes L., Tjostheim D., Revealing some unexpected dependence properties of linear combinations of stable random variables using symmetric covariations. *Communications in Statistics, Theory and Methods*, **33**, (4), pp 769–786.

[2009] Garleanu N., Pedersen L.H., Poteshman A.M., Demand-based option pricing. *The Review of Financial Studies*, **22**, (10), pp 4259–4299.

[1999] Gatheral J., The volatility skew: Arbitrage constraints and asymptotic behaviour. Merrill Lynch.

[2004] Gatheral J., A parsimonious arbitrage-free implied volatility parameterization with application to the valuation of volatility derivatives. Presentation at Global Derivatives.

[2006] Gatheral J., The volatility surface: A practitioner's guide, John Wiley & Sons, Hoboken, New Jersey.

[2012] Gatheral J., Jacquier A., Arbitrage-free SVI volatility surfaces. Working Paper, Baruch College, CUNY, and Imperial College, London.

[1809] Gauss C.F., Theoria motvs corporvm coelestivm in sectionibvs conicis Solem ambientivm.

[1989] Geman H., The importance of the forward neutral probability in a stochastic approach of interest rates. Working paper, **7**, Essec.

[1995] Geman H., El Karoui N., Rochet J., Changes of numeraire, change of probability measure and option pricing. *Journal of Applied Probability*, **32**, pp 443–458.

[2004] Genz A., Numerical computation of rectangular bivariate and trivariate normal and t probabilities. *Statistics and Computing*, **14**, pp 151–160.

[2012] Gerhold S., Can there be an explicit formula for implied volatility? Working Paper, Id=2179924.

[1979] Geske R., A mote on an analytical formula for unprotected American call options on stocks with known dividends. *Journal of Financial Economics*, **7**, pp 375–380.

[1943] Gnedenko B.V., Sur la distribution limite du terme maximum d'une serie aleatoire. *Annals of Mathematics*, **44**, pp 423–453.

[1954] Gnedenko B.V., Kolmogorov A.N., Limit distributions for sums of independent random variables. Addison-Wesley, Cambridge.

[2004] Granger C., Hyng N., Occasional structural breaks and long memory with an application to the SP500 absolute stock returns. *Journal of Empirical Finance*, **11**, pp 399–421.

[2005] Grech D., Mazur Z., Statistical properties of old and new techniques in detrended analysis of time series. *Acta Physica Polonica B*, **36**, (8), pp 2403–2413.

[1992] Grimmett G.R., Stirzaker D.R., Probability and random processes. Oxford Science Publications, Second Edition.

[2010] Gu G-F., Zhou W-X., Detrending moving average algorithm for multifractals. *Physical Review E*, **82**, 011136, pp 1–12.

[2012] Guo G., Jacquier A., Martini C. and Neufcourt L., Generalised arbitrage-free SVI volatility surfaces. Working Paper.

[2011] Gurrieri S., A class of term structures for SVI implied volatility. Working Papaer, Risk Management Department, Mizuho Securities, Tokyo.

[2002] Gushchin A.A., Mordecki E., Bounds on option prices for semimartingale market models. *Proc. Steklov Inst. Math.*, **237**, pp 73–113.

[1986] Gyongy I., Mimicking the one-dimensional marginal distributions of processes having an Ito differential. *Probability Theory and Related Fields*, **71**, pp 501–516.

[1999] Hagan P.S., Woodward D.E., Equivalent Black volatilities. *App. Math. Finance*, **6**, pp 147–157.

[2002] Hagan P.S., Kumar D., Lesniewski A.S., Woodward D.E., Managing smile risk. *Wilmott Magazine*, pp 84–108.

[1979] Harrison J.M., Kreps D., Martingale and arbitrage in multiperiods securities markets. *Journal of Economic Theory*, **20**, pp 381–408.

[1968] Hart J., Computer approximations. Wiley, Algorithm 5666 for the error function.

[2003] Haug E.G., Haug J., Lewis A., Back to basics: A new approach to the discrete dividend problem. *Wilmott Magazine*, pp 37–47.

[2009] Haugh M., Black-Scholes and the volatility surface. *Financial Engineering: Continuous-Time Models*, pp 1–17.

[1993] Heston S., A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, **6**, pp 327–343.

[1975] Hill B.M., A simple general approach to inference about the tail of a distribution. *Annals of Statistics*, **3**, pp 1163–1174.

[2004] Ho D-S., Lee C-K., Wang C-C., Chuang M., Scaling characteristics in the Taiwan stock market. *Physica A*, **332**, pp 448–460.

[1999] Hoeting J.A., Madigan D., Raftery A.E., Volinsky C.T., Bayesian model averaging: A tutorial. *Statistical Science*, **14**, (4), pp 382–417.

[1987] Hull J., White A., The pricing of options on assets with stochastic volatilities. *Journal of Finance*, **42**, (2), pp 281–300.

[1993] Hull J., White A., Efficient procedures for valuing european and american pathdependent options. *The Journal of Derivatives*, **1**, (1), pp 21–31.

[2012] Hull J.C., Options, futures and other derivatives. Prentice & Hall, Pearson Education, Eight Edition.

[1951] Hurst H.E., Long-term storage capacity of reservoirs. *Trans. Amer. Soc. Civil Engineers*, **116**, pp 770–799.

[1990] Hutchinson T.P., Lai C.D., Continuous bivariate distributions, emphasizing applications. Rumsby Scientific Publishing, Adelaide 5000, Australia.

[2011] Hutchinson D., Relative value concepts in volatility. CBOE Risk Management Conference, Dana Point, CA.

[2007] Isengildina-Massa O., Curtis Jr. C.E., Bridges Jr. W.C., Nian M., Accuracy of implied volatility approximations using "nearest-to-the-money" option premiums. Southern Agricultural Economics Association Meetings, Mobile, AL, February.

[1998] Ingersoll J.E., Approximating american options and other financial contracts using barrier derivatives. *Journal of Computational Finance*, **2** (Fall), pp 85–112.

[2000] Ingersoll J.E., Digital contracts: Simple tools for pricing complex derivatives. *The Journal of Business*, **73**, (1), pp 67–88.

[2006] Jackel P., By implication. *Wilmott Magazine*, November, pp 60–66.

[1996] Jackwerth J.C., Rubinstein M., Recovering probability distributions from option prices. *Journal of Finance*, **51**, (5), pp 1611–1631.

[1999] Jackwerth J.C., Option-implied risk-neutral distributions and implied binomial trees: A literature review. *Journal of Derivatives*, **7**, (2), pp 66–81.

[2004] Jacod J., Protter P., Probability essentials. Springer Second Edition, Universitext.

[2007] Jacquier A., Slaoui S., Variance dispersion and correlation swaps. School of Economics, Mathematics and Statistics, Birkbeck University of London.

[1993] Jakubowski A., Minimal conditions in p-stable limit theorems. *Stoch. Proc. Appl.*, **44**, pp 291–327.

[1997] Jakubowski A., Minimal conditions in p-stable limit theorems II. *Stoch. Proc. Appl.*, **68**, pp 1–20.

[1991] Jansen D.W., de Vries C.G., On the frequency of large stock market returns: Putting booms and busts into perspective. *Review of Economics and Statistics*, **73**, (1), pp 18–24.

[1982] Jarrow R.A., Rudd A., Approximate option valuation for arbitrary stochastic processes. *Journal of Financial Economics*, **10**, (3), pp 347–369.

[1983] Jarrow R.A., Rudd A., Option pricing. Homewood, Illinois, Richard Irwin, pp 183–188.

[1999] Jex M., Henderson R., Wang D., Pricing exotics under the smile. *Risk*, November, pp 72–75.

[2012] Jizba P., Korbel J., Methods and techniques for multifractal spectrum estimation in financial time series. FNSPE, Czech Technical University, Prague.

[2001] Jondeau E., Rockinger M., Gram-Charlier densities. *Journal of Economic Dynamics and Control*, **25**, pp 1457–1483.

[2002] Jones C.S., Wang T., The term structure of equity option implied volatility. Working Paper

[2004] Jones M.C., Families of distributions arising from distributions of order statistics. *Test*, **13**, pp 143.

[2009] Jones M.C., Kumaraswamy's distribution: A beta-type distribution with some tractability advantages. *Statistical Methodology*, **6**, (1), pp 70–81.

[1998] Ju N., Pricing American option by approximating its early exercise boundary as a multipiece exponential function. *Review of Financial Studies*, **11**, (3), pp 627–646.

[1999] Ju N., Zhong R., An approximate formula for pricing American options. *Journal of Derivatives*, **7**, (2), pp 31–40.

[2000] Kaas R., Dhaene J., Goovaerts M.J., Upper and lower bounds for sums of random variables. *Insurance: Mathematics and Economics*, **27**, pp 151–168.

[2004] Kahalé N., An arbitrage-free interpolation of volatilities. *Risk*, **17**, (5), pp. 102–106.

[2002] Kantelhardt J., Zschiegner S., Koscielny-Bunde E., Bunde A., Havlin S., Stanley H.E., Multifractal detrended fluctuation analysis of nonstationary time series. *Physica A*, **316**, (1-4), pp 87–114.

[1988] Karatzas I., On the pricing of American options. *Appl. Math. Optim*, **17**, pp 37–60.

[2004] Kim S-J., Choi J-S., Multifractal measures for the yen-dollar exchange rate. *Journal of the Korean Physical Society*, **44**, (3), pp 643–646.

[1995] Kirk E., Correlations in the energy markets. Managing energy price risk. *Risk Publications and Enron*, London, pp 71–78.

[1980] Koutrouvelis I.A., Regression-type estimation of the parameters of stable laws. *Journal of the American Statistical Association*, **75**, pp 918–928.

[2013] Kristoufek L., Vosvrda M., Measuring capital market efficiency: Global and local correlations structure. *Physica A*, **392**, (1), pp 184–193.

[2003] Kruse S., On the pricing of forward starting options under stochastic volatility. Institut Techno und Wirtschaftsmathematik.

[1980] Kumaraswamy P., A generalized probability density function for double-bounded random processes. *Journal of Hydrology*, **46**, (1-2), pp 79–88.

[1967] Kunita H., Watanabe S., On square integrable martingales. *Nagoya Math. Journal*, **30**, pp 209–245.

[1997] Kwiatkowski J., Current issues in Value at Risk for portfolio of derivatives. Global Risk Management, BWZ.

[2009] Ladokhin S., Volatility modeling in financial markets. Master Thesis, VU University Amsterdam.

[2009] Langnau A., Introduction into local correlation modelling. ArXiv Working Paper $090 : 3441v3$.

[2010] Langnau A., A dynamic model for correlation. *Risk*, pp 74–78.

[1774] Laplace P-S., Memoire sur la probabilite des causes par les evenements. *Memoires de l'Academie Royale des Sciences*, **6**, pp 621–656.

[1976] Latane H., Rendleman R., Standard deviations of stock price ratios implied in option prices. *The Journal of Finance*, **31**, pp 369–381.

[2014] Le T., Intrinsic prices of risk. Working Paper.

[1992] LeBaron B., Some relations between volatility and serial correlations in stock market returns. *Journal of Business*, **65**, pp 199–219.

[1991] L'Ecuyer P., Tables of maximality-equidistributed combined LFSR generators. Working Paper, DIRO, Universite de Montreal.

[1996b] L'Ecuyer P., Maximally equidistributed combined Tausworthe generators. *Mathematics of Computation*, **65**, (213), pp 203–213.

[1998] L'Ecuyer P., Good parameters and implementations for combined multiple recursive random number generators. Working Paper, DIRO, Universite de Montreal, May 4.

[1999b] L'Ecuyer P., Tables of maximally-equidistributed combined LFSR generators. *Mathematics of Computation*, **68**, (225), pp 261–269.

[2000b] L'Ecuyer P., Touzin R., Fast combined multiple recursive generators with multipliers of the form $a = \pm 2^q \pm 2^r$. in J.A. Joines, R.R. Barton, K. Kang, P.A. Fishwick, eds, *Proceedings of the 2000 Winter Simulation Conference*, pp 683–689.

[2007] L'Ecuyer P., Pseudorandom number generators. Working Paper, DIRO, Universite de Montreal, August 28.

[2001] Lee R.W., Implied and local volatilities under stochastic volatility. *International Journal of Theoretical and Applied Finance*, **4**, (1), pp 1178–1192.

[2004] Lee R.W., The moment formula for implied volatility at extreme strikes. *Mathematical Finance*, **14**, (3), pp 469–480.

[2005] Lee R.W., Implied volatility: Statistics, dynamics, and probabilistic interpretation. *Recent advances in applied probability*, Springer, New York, pp 241–268.

[2000] Li D., On default correlation: A copula function approach. *Journal of Fixed Income*, **9**, (4), pp 43–54.

[2009] Liang S., Tahara Y., A formula to compute implied volatility with error estimate. *Interdisciplinary Iinformation Sciences*, **15**, No. 2, pp 267–272.

[2002] Lipton A., The volatility smile problem. *Risk*, **2**, (2).

[2010] Lisauskas J., Dispersion trading in German option market. Master Thesis, Tilburg University.

[1991] Lo A.W., Long-term memory in stock market prices. *Econometrica*, **59**, (5), pp 1279–1313.

[2005] London J., Modeling derivatives in C++. John Wiley & Sons, Hoboken, New Jersey.

[1996] Longin F., The asymptotic distribution of extreme stock market returns. *Journal of Business*, **69**, pp 383–408.

[1994] Loretan M., Phillips P., Testing the covariance stationarity of heavy-tailed time series. *Journal of Empirical Finance*, **1**, pp 211–248.

[2005] Lozovaia T., Hizhniakova H., How to extend modern portfolio theory to make money from trading equity options. EGAR Technology.

[1996] Lux T., The stable Paretian hypothesis and the frequency of large returns: An examination of major German stocks. *Applied Economics Letters*, **6**, pp 463–475.

[2012] Lye C-T., Hooy C-W., Multifractality and efficiency: Evidence from Malaysian sectoral indices. *Int. Journal of Economics and Management*, **6**, (2), pp 278–294.

[2009a] Ma J., Pricing foreign equity options with stochastic correlation and volatility. *Annals of Economics and Finance*, **10**, (2), pp 303–327.

[2009b] Ma J., A stochastic correlation model with mean reversion for pricing multi-asset options. *Asia-pacific Financial Markets*, **66**, pp 97–109.

[1993] Maheswaran S., Sims C., Empirical implications of arbitrage-free asset markets. in P.C.B. Phillips, ed., *Models, Methods and Applications of Econometrics*, Cambridge, Basil Blackwell, pp 301–316.

[1997] Malz A., Estimating the probability distribution of the future exchange rate from option prices. *Journal of Derivatives*, Winter, pp 18–36.

[1982] Manaster S., Koehler G., The calculation of implied variances from the Black-Scholes-Merton model: A note. *The Journal of Finance*, **37**, pp 227–230.

[1963] Mandelbrot B.B., The variation of certain speculative prices. *Journal of Business*, **36**, (4), pp 394–419.

[1965] Mandelbrot B.B., Une classe de processus stochastiques homothetiques a soi: Application a loi climatologique de H.E. Hurst. *Comptes Rendus Academic Sciences Paris*, **240**, pp 3274–3277.

[1968] Mandelbrot B.B., van Ness J., Fractional Brownian motions, fractional noises and applications. *SIAM Review*, **10**, (4), pp 422–437.

[1968b] Mandelbrot B.B., Wallis J., Noah, Joseph and operational hydrology. *Water Resources Research*, **4**, pp 909–918.

[1969a] Mandelbrot B.B., Wallis J., Computer experiments with fractional Gaussian noises: Part 1, averages and variances. Water Resources Research 5.

[1969b] Mandelbrot B.B., Wallis J., Computer experiments with fractional Gaussian noises: Part 2, rescaled ranges and spectra. Water Resources Research 5.

[1974] Mandelbrot B.B., When can price be arbitraged efficiently? A limit to the validity of the random walk and martingale models. *Re. Econom. Statist.*, **53**, pp 225–236.

[1975] Mandelbrot B.B., Les objets fractals: forme, hasard et dimension. Paris, Flammarion.

[1979] Mandelbrot B.B., Taqqu M., Robust R/S analysis of long-run serial correlation. in *Proceedings of the 42nd Session of the International Statistical Institute*, , pp 69–104, Manila, Bulletin of the I.S.I..

[1997] Mandelbrot B.B., Fisher A., Calvet L., A multifractal model of asset returns. Cowles Foundation Discussion Paper No. 1164.

[2000] Mantegna R.N., Stanley H.E., An introduction to econophysics: Correlation and complexity in finance. Cambridge University Press, Cambridge.

[1978] Margrabe W., The value of an option to exchange one asset for another. *Journal of Finance*, **33**, pp 177–186.

[1952] Markowitz H.M., Portfolio selection. *Journal of Finance*, **7**, (1), pp 77–91.

[1999] Marris D., Financial option pricing and skewed volatility. MPhil Thesis, University of Cambridge.

[2008] Marshall C.M., Dispersion trading: Empirical evidence from US options markets. *Global Finance Journal*, (20), pp 289–301.

[2003] Matia K., Ashkenazy Y., Stanley H.E., Multifractal properties of price fluctuations of stocks and commodities. *Europhysics Letters*, **61**, (3), pp 422–428.

[2004] Matos J.A.O., Gama S.M.A., Ruskin H., Duarte J., An econophysics approach to the Portuguese stock index-psi-20. *Physica A*, **342**, (3-4), pp 665–676.

[2008] Matos J.A.O., Gama S.M.A., Ruskin H.J., Al Sharkasi A., Crane M., Time and scale Hurst exponent analysis for financial markets. *Physica A*, **387**, pp 3910–3915.

[1996] McCulloch J.H., Financial applications of stable distributions. in G.S. Maddala, C.R. Rao, eds., *Handbook of Statistics*, **14**, Elsevier, pp 393–425.

[1997] McCulloch J.H., Measuring tail thickness to estimate the stable index α: A critique. *Journal of Business & Economic Statistics*, **15**, (1), pp 74–81.

[1984] McDonald J.B., Some generalized functions for the size distribution of income. *Econometrica*, **52**, pp 647–663.

[1997] Melick W., Thomas C., Recovering an asset's implied pdf from option prices: An application to crude oil during the Gulf war. *Journal of Financial and Quantitative Analysis*, **32**, (1), pp 91–115.

[1973] Merton R.C., Theory of rational option pricing. *Bell J. of Econ. and Management Sci.*, **4**, (1), pp 141–183.

[1976] Merton R.C., Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics*, **3**, pp 125-144.

[2003] Mikosch T., Starica C., Long-range dependence effects and ARCH modeling. in *Theory and Applications of Long-range Dependence*, Birkhauser Boston, Boston, pp 439–459.

[1998] Milevsky M., Posner S., A closed-form approximation for valuing basket options. *The Journal of Derivatives*, **5**, (4), pp 54–61.

[1999] Mittnik S., Doganoglu T., Chenyao D., Computing the probability density function of the stable Paretian distribution. *Mathematical and Computer Modelling*, **29**, pp 235–240.

[1964] Moran P., On the range of cumulative sums. *Annals of the Institute of Statistical Mathematics*, **16**, pp 109-112.

[2003] Moskowitz T., An analysis of covariance risk and pricing anomalies. *Review of Financial Studies*, **16**, pp 417-457.

[2006] Moyano L.G., de Souza J., Duarte Queiros S.M., Multi-fractal structure of traded volume in financial markets. *Physica A*, **371**, pp 118–121.

[2005] Muck M., On the similarity between displaced diffusion and constant elasticity of variance market models of the term structure. WHU - Otto Beisheim Graduate School of Management.

[2010] Nelsen R.B., An introduction to Copulas. Springer.

[2003] Nielsen J.A., Sandmann K., Pricing bounds on Asian options. *Journal of Financial and Quantitative Analysis*, **38**, (2), pp 449–473.

[2013] Niere H.M., A multifractality measure of stock market efficiency in Asean region. *European Journal of Business and Management*, **5**, (22), pp 13–19.

[2001] Nolan J.P., Maximum likelihood estimation and diagnostics for stable distributions. in O.E. Barndorff-Nielsen, T. Mikosch, S. Resnick, eds., *Levy Processes*, Brikhauser, Boston.

[2014] Nolan J.P., Stable distributions: Models for heavy-tailed data. Math/Stat Department, American University.

[2005] Norouzzadeh P., Jafari G.R., Application of multifractal measures to Tehran price index. *Physica A*, **356**, pp 609–627.

[2006] Norouzzadeh P., Rahmani B., A multifractal detrended fluctuation description of Iranian rial-US dollar exchange rate. *Physica A*, **367**, pp 328–336.

[2012] Oh G., A multifractal analysis of Asian foreign exchange markets. *European Physical Journal B*, pp 85–214.

[1998] Oksendal V, Stochastic differential equations. Springer Fifth Edition.

[2006] Ouyang Z., Stable process. Working Notes.

[2002] Overhaus M., Ferraris A., Knudsen T., Milward R., Nguyen-Ngoc L., Schindlmayr G., Equity derivatives. Theory and Application John Wiley & Sons, New York.

[1995] Page M., Feng E., Measuring the risk of a portfolio of options on several exchange rates. Working Paper, Susquehanna Investment Group, Philadelphia.

[2006] Panneton F., L'Ecuyer P., Matsumoto M., Improved long-period generators based on linear recurrences modulo 2. *ACM Transactions on Mathematical Software*, **32**, (1), pp 1–16.

[2000] Pasquini M., Serva M., Clustering of volatility as a multiscale phenomenon. *European Physical Journal B*, **16**, (1), pp 195–201.

[1994] Peng C.K., Buldyrev S.V., Havlin S., Simons M., Stanley H.E., Goldberger A.L., Mosaic organization of DNA nucleotides. *Physical Review E*, **49**, (2), pp 1685–1689.

[1994] Peters E.E., Fractal market analysis: Applying chaos theory to investment and economics. John Wiley & Sons.

[2005] Piterbarg V., A multi-currency model with FX volatility skew. Working paper, February.

[2003] Poon S., Rockinger M., Tawn J.A., Modelling extreme-value dependence in international stock markets. *Statistica Sinica*, **13**, pp 929–953.

[2004] Poon S., Rockinger M., Tawn J.A., Extreme value dependence in financial markets: Diagnostic, models and financial implications. *Review of Financial Studies*, **17**, pp 581–610.

[1992] Press W.H., Teukolsky S.A., Vetterling W.T., Flannery B.P., Numerical recipes. 2nd ed. Cambridge, Cambridge University Press.

[2002] Rebonato R., Assigning future smile surfaces: Conditions for uniqueness and abscence of arbitrage. Working Paper

[2004] Rebonato R., Cardoso M.T., Unconstrained fitting of implied volatility surfaces using a mixture of normals, Working Paper, QUARC and Oxford University, July 6.

[2010] Reghai A., Breaking correlation breaks. *Risk*, pp 90–95.

[1997] Reiss R., Thomas M., Statistical analysis of extreme values with applications to insurance, finance, hydrology and other fields. Birkhauser, Basel.

[1999] Resnick S., Samorodnitsky G., Xue F., How misleading can sample ACFs of stable MAs be? (Very!). *Ann. Appl. Prob.*, **9**, pp 797–817.

[2000] Resnick S., Van Den Berg E., Sample correlation behavior for the heavy tailed general bilinear process. *Comm. Statist. Stochastic Models*, **16**, pp 233–258.

[1984] Rodgers J.L., Nicewander W.A., Toothaker L., Linearly independent, orthogonal, and uncorrelated variables. *The American Statistician*, **38**, (2), pp 133–134.

[1995] Rogers L.C.G., Shi Z., The value of an Asian option. *Journal of Applied Probability*, **32**, pp 1077–1088.

[2010] Rogers L.C.G., Tehranchi M.R., Can the implied volatility surface move by parallel shifts? *Finance and Stochastics*, **14**, (2), pp 235–248.

[2010] Roper M., Arbitrage free implied volatility surfaces. Working Papaer, School of Mathematics and Statistics, The University of Sydney, Australia.

[1998] Rosenberg J.V., Pricing multivariate contingent claims using estimated risk-neutral density functions. *Journal of International Money and Finance*, **17**, pp 229–247.

[2003] Rosenberg J.V., Nonparametric pricing of multivariate contingent claims. *Federal Reserve Bank of New York Staff Reports*, **162**.

[2006] Ross S., Private communication.

[2001] Routledge B.R., Zin S., Model uncertainty and liquidity. NBER Working Paper 8683.

[1983] Rubinstein M., Displaced diffusion option pricing. *Journal of Finance*, **38**, (1), pp 213–217.

[1994] Rubinstein M., Implied binomial trees. *Journal of Finance*, **49**, (3), pp 771–818.

[1998] Rubinstein M., Edgeworth binomial trees. *Journal of Derivatives*, **5**, (3), pp 20–27.

[1994] Samorodnitsky G., Taqqu M.S., Stable non-Gaussian random processes. Stochastic Modeling, Chapman & Hall, New York-London.

[1960] Samuelson P.A., Efficient paths of capital accumulation in terms of the calculus of variations. in K.J. Arrow, S. Karlin, P. Suppes *Mathematical Models in the Social Sciences*, 1959 Proceedings of the first Stanford Symposium, Stanford University Press, pp 77–88.

[1965] Samuelson P.A., Proof that properly anticipated prices fluctuate randomly. *Industrial Management Review*, **6**, pp 41–50.

[1965a] Samuelson P.A., Rational theory of warrant pricing. *Indus. Management Rev.*, **6**, pp 13–31. Reprinted in Cootner (1967), pp 506–532.

[1969] Samuelson P.A., Merton R.C., A complete model of warrant pricing that maximizes utility. *Industrial Management Review*, pp 17–46.

[2010] Schlogl E., Option pricing where the underlying assets follow a Gram/Charlier density of arbitrary order. Working Paper, School of Finance and Economics, University of Technology, Sydney.

[1999] Schmitt F., Schertzer D., Lovejoy S., Multifractal analysis of foreign exchange data. *Applied Stochastic Models and Data Analysis*, **15**, pp 29–53.

[1978] Schonfelder J.L., Chebyshev expansions for the error and related functions. *Mathematics of Computation*, **32**, No. 144, pp 1232–1240.

[1989] Schroder M., Computing the constant elasticity of variance option pricing formula. *Journal of Finance*, **44**, pp 211–219.

[1970] Sharpe W.F., Portfolio theory and capital markets. New York: McGraw-Hill.

[2003] Shreve S., Stochastic calculus for finance I. The binomial asset pricing model. Springer Finance.

[2004] Shreve S.E., Stochastic calculus for finance II, Continuous-time models. Springer Finance, Textbook, NY, USA.

[1955] Simon H.A., On a Class of Skew Distribution Functions. *Biometrika*, **42**, (3/4), pp 425–440.

[2000] Simon S., Goovaerts M.J., Dhaene J., An easy computable upper bound for the price of an arithmetic Asian option. *Insurance: Mathematics and Economics*, **26**, (2-3), pp 175–184.

[2003] Simpson J., Currency options and volatility. Goldman Sachs Working Paper, Chapter 11, Siegel M.J. at Stern NYU.

[1959] Sklar A., Fonctions de repartition a n dimensions et leurs marges. *Publications de l'Institut de Statistiques de l'Universite de Paris 8*, pp 229–231.

[1961] Sprenkle C., Warrant prices as indications of expectations. *Yale Econ. Essays 1*, pp 179–232. Reprinted in Cootner (1967), pp 412–474.

[2008] Staum J., Incomplete Markets. *Handbooks in OR & MS, Vol. 15: Financial Engineering*, ed. J. R. Birge and V. Linetsky, Elsevier, pp 511–563.

[1991] Stein E.M., Stein J.C., Stock price distribution with stochastic volatility: An analytic approach. *The Review of Financial Studies*, **4**, (4), pp 727–752.

[1989] Sterge A.J., On the distribution of financial futures price changes. *Financial Analysts Journal*, **45**, (3), pp 75–78.

[1969] Stoll H.R., The relationship between put and call option prices. *Journal of Finance*, **24**, pp 802–824.

[2014] Stosic D., Stosic D., Stosic T., Stanley H.E., Multifractal analysis of managed and independent float exchange rates. Department of Physics, Boston University.

[1999] Struzik Z.R., Local effective Holder exponent estimation on the wavelet transform maxima tree. in *Fractals: Theory and Applications in Engineering*, eds., M. Dekking, J. Levy Vehel, E. Lutton, C. Tricot, Springer Verlag, pp 93–112.

[2002] Struzik Z.R., Siebes A., Wavelet transform based multifractal formalism in outlier detection and localisation for financial time series. *Physica A*, **309**, pp 388–402.

[2007] Svoboda-Greenwood S., The displaced diffusion as an approximation of the CEV. Working paper, April.

[2000] Tankov P., GRM smile model. Rapport de stage de fin d'etude de l'X.

[2010] Tankov P., Touzi N., Calcul stochastique en finance. Ecole Polytechnique Paris, Departement de Mathematiques Appliquees.

[1995] Taqqu M.S., Teverovsky V., Willinger W., Estimators for long-range dependence: An empirical study. *Fractals*, **3**, (4), pp 785–798.

[2009] Tehranchi M.R., Asymptotics of implied volatility far from maturity. *Journal of Applied Probability*, **46**, (3), pp 629–650.

[1967] Thorp E.O., Kassouf S.T., Beat the market. New York: Random House.

[2010] Tian Y.S., Extracting risk-neutral density and its moments from American option prices. Schulich School of Business, York University, Toronto, Canada.

[1991] Trigeorgis L., A log-transformed binomial numerical analysis method for valuing complex multi-option investments. *Journal of Financial and Quantitative Analysis*, **26**, pp 309–326.

[2006] Turiel A., Perez-Vicente C.J., Grazzini J., Numerical methods for the estimation of multifractal singularity spectra on sampled data: A comparative study. *Journal of Computational Physics*, **216**, pp 362–390.

[1999] Uchaikin V.V., Zolotarev V.M., Chance and stability: Stable distributions and their applications. Modern Probability and Statistics, De Gruyter, The Netherlands.

[1997] Vandewalle N., Ausloos M., Coherent and random sequences in financial fluctuations. *Physica A*, **246**, (3), pp 454–459.

[1998] Vandewalle N., Ausloos M., Crossing of two mobile averages: A method for measuring the robustness exponent. *Phys. Rev.*, **58**, pp 177–188.

[1998b] Vandewalle N., Ausloos M., Multi-affine analysis of typical currency exchange rates. *European Physical Journal B.*, **4**, (2), pp 257–261.

[1998c] Vandewalle N., Ausloos M., Boveroux PH., Detrended fluctuation analysis of the foreign exchange market. Working Paper.

[2006] Vellekoop M.H., Nieuwenhuis J.W., Efficient pricing of derivatives on assets with discrete dividends. *Applied Mathematical Finance*, **13**, (3), pp 265–284.

[2004] Wang Y., Yin H., Qi L., No-arbitrage interpolation of the option price function and its reformation. *Journal of Optimization Theory and Applications*, **120**, (3), pp 627–649.

[2009] Wang Y., Liu L., Gu R., Analysis of efficiency for Shenzhen stock market based on multifractal detrended fluctuation analysis. *International Review of Financial Analysis*, **18**, pp 271–276.

[2011] Wang Y., Wei Y., Wu C., Analysis of the efficiency and multifractality of gold markets based on multifractal detrended fluctuation analysis. *Physica A: Statistical Mechanics and its Applications*, **390**, pp 817–827.

[2011b] Wang Y., Wu C., Pan Z., Multifractal detrending moving average analysis on the US Dollar exchange rates. *Physica A*, **390**, pp 3512–3523.

[1981] Whaley R.E., On the valuation of American call options on stocks with known dividends. *Journal of Financial Economics*, **9**, pp 207–211.

[2004] G. West, Better approximations to cumulative normal functions. Working Paper, University of the Witwatersrand.

[1991] Williams D., Statistical laboratory. DPMMS, Cambridge University, Probability with Martingales, Cambridge University Press.

[1993] Wilmott P., Option pricing: Mathematical models and computation. Oxford Financial Press, Oxford.

[2009] Xu S.J., Jin X.J., Predicting drastic drop in Chinese stock market with local Hurst exponent. *International Conference on Management Science and Engineering*, pp 1309–1315.

[2009] Yuan Y., Zhuang X-T., Jin X., Measuring multifractality of stock price fluctuation using multifractal detrended fluctuation analysis. *Physica A*, **388**, (11), pp 2189–2197.

[2009] Zeliade Systems, Quasi-explicit calibration of Gatheral's SVI model. Zeliade White Papaer, Zeliade Systems, Paris, France.

[2015] Zheng H., Avanced methods in derivatives pricing. MSc Mathematics and Finance, Imperial College London.

[2007] Zhu J., An extended libor market model with nested stochastic volatility dynamics. Working Paper, January 5.

[2007] Zunino L., Tabak B.M., Perez D.G., Garavaglia M., Rosso O.A., Inefficiency in Latin-American market indices. *The European Physical Journal B*, **60**, (1), pp 111–121.

[2008] Zunino L., Tabak B.M., Figlioa A., Perez D.G., Garavaglia M., Rosso O.A., A multifractal approach for stock market inefficiency. *Physica A*, **387**, pp 6558–6566.