UNIVERSITÀ DEGLI STUDI DI ROMA 'LA SAPIENZA'

Dottorato di Ricerca in

Statistica Economica

XXVI Ciclo

New Tools for Volatility Models

SUPERVISOR

Roberto Renò

CANDIDATE

Manola Santilli

4 April 2014

Perché la vita è così. Procediamo a piccoli passi.
Rialziamo la testa e torniamo ad affrontare
il volto feroce e sorridente del mondo.
Pensiamo. Agiamo. Sentiamo.
Diamo il nostro piccolo contributo
alle maree del bene e del male,
che inondano e prosciugano la terra.
Trasciniamo le nostre croci ammantate d'ombra
nella speranza di una nuova notte.
Lanciamo i nostri cuori coraggiosi
nelle promesse di un nuovo giorno.
Con amore: l'appassionata ricerca
di una verità diversa dalla nostra.
Con struggimento: il puro,
ineffabile anelito di essere salvati.
Poiché fino a quando il destino ce lo consente,
continuiamo a vivere.
Che Dio ci aiuti. Che Dio ci perdoni.
Continuiamo a vivere.

G. D. ROBERTS, SHANTARAM

Ringraziamenti

Al raggiungimento di questo traguardo non indifferente, mi sembra doveroso ringraziare coloro i quali lo hanno reso possibile.
Ringrazio Roberto Renò, non solo per i suoi preziosi insegnamenti, ma soprattutto perchè la grinta e la passione che mette nel suo lavoro sono state per me fonte di ispirazione. Un altro Roberto che voglio ringraziare è Roberto Monte, perchè, iniziandomi alle meraviglie del mondo stocastico, è stato il primo a credere in me e nelle mie potenzialità.
Inoltre, ringrazio Eric Reynolds per averci fornito i dati e Claudio Pacati per gli utili suggerimenti.
Un sentito ringraziamento va alla mia famiglia, mamma Carla e papà Loreto, i miei due angeli custodi, e mia sister Vanessa, per avermi sempre lasciata libera di scegliere. Siete i migliori.
Ringrazio Guido, per avermi reso completa e per avermi dato la forza e la sicurezza quando spesso a me sono mancate. Infine ringrazio Lara, per esserci sempre stata.

Contents

Introduction

From the late 1990s, *stochastic volatility* has been one of the most active areas of research in empirical finance and time series econometrics. Given the importance of asset return variation on a great number of practical financial management decisions, there have been extensive efforts to provide good real time estimates and forecasts of current and future volatility. The development of such a field requires a highly multidisciplinary knowledge, from the concepts of financial economics and econometrics to the principles of probability theory, merging together to produce methods and models which have aided our understanding of the realistic option pricing, efficient asset allocation and accurate risk management. In particular, the dynamics of the *spot volatility*, also known as the *instantaneous volatility*, is extremely important for modeling financial series. However, the spot volatility itself is a latent variable, which is not directly observable. Therefore, during the last decades, there has been considerable attention on which was the best method, in terms of efficiency and consistency, to estimate a latent variable, such as the spot volatility. Since option prices reflect expectations and rumors of economic agents on future movements of the underlying asset, the implied volatility is widely believed to contain much more information than the historical volatility. As a consequence, it is considered a more efficient forecast for future realized volatility. Recently, Andersen, Fusari and Todorov [20] have developed consistent estimators of the risk-neutral parameters vector and the dynamic realization of the state vector which governs an option price dynamics, in a fully parametric framework.

The goal of the first part of this work is to propose a new estimation method of the spot volatility, based on a semi-nonparametric model, which employs the information content of a complete data set of European options, daily quoted in the market, under no arbitrage assumptions. The technique we propose is based on the idea of *model-free implied volatility*, developed by Britten-Jones and Neuberger [49] in 2000, and exploits the *observed* VIX term structure to make inference on the *unobserved* spot volatility. In 1993, the Chicago Board Options Exchange introduced the *volatility index*, known as VIX, the first index to measure the aggregate volatility of the US equity market. Nowadays, this index has become the premier benchmark for the stock market volatility. Often defined as the "investor fear gauge", the VIX measures market expectations of 30-day volatility implied by equity index option prices. Hence, it is considered one of the most issued financial indicators and it is widely followed by academics and quants, especially after the financial turmoil started in 2008.

Let $(\Omega, \mathfrak{E}, \mathbb{P}) \equiv \Omega$ be a probability space endowed with a filtration $(\mathfrak{F}_t)_{t\geq 0}$ and let $(S_t)_{t\geq 0} \equiv S_t$ be the logarithm of a financial equity asset price evolving in continuous time, whose quadratic variation $[S,S]_t$ is an adapted, increasing and càdlàg (i.e. with paths that are a.s. right continuous with left limits) process. Hence, following Todorov and Tauchen [154] approach, the VIX can be written as

$$VIX_t = \mathbf{E}^{\mathbb{Q}}\left[[S,S]_T - [S,S]_t \mid \mathfrak{F}_t\right], \qquad T > t, \tag{1}$$

where $T > 0$ is given and the expectation is taken under the risk-neutral distribution $\mathbb{Q}$, whose existence is guaranteed by no arbitrage assumptions. Then, if there are no jumps in the price process S_t, the VIX obviously equals the familiar *expected risk-neutral integrated variance*, given by

$$VIX_t = \mathbf{E}^{\mathbb{Q}}\left[\int_t^T \sigma_s^2 ds \;\middle|\; \mathfrak{F}_t\right]. \tag{2}$$

The theoretical foundations of the VIX refers to the concept of model-free implied volatility, according to which the expected risk-neutral integrated return variance between two arbitrary dates, say T_1 and T_2 with $T_1 < T_2$, is completely specified by the set of option prices expiring on the two dates, that is

$$\mathbf{E}^{\mathbb{Q}}\left[\int_{T_1}^{T_2}\left(\frac{dS_u}{S_u}\right)^2\right] = 2\int_0^\infty \frac{C(T_2,K) - C(T_1,K)}{K^2}\,dK, \tag{3}$$

where $\mathbf{E}^{\mathbb{Q}}[\cdot]$ denotes the expectation taken under the risk-neutral probability measure. Hence, unlike the traditional concept of implied volatility, the model-free implied volatility is not based on any specific option pricing model. Instead, it is derived entirely from no arbitrage conditions, disregarding the underlying dynamics. This allows us to define the VIX as the risk-neutral forecast of the instantaneous variance, whose value is equivalent to that of a portfolio of out-of-the-money options weighted inversely proportional to the square of their strike prices.

Let $r \in \mathbb{R}$ be a real constant and let $(\beta_t)_{t\geq 0}$ be a measurable and locally bounded càdlàg (i.e. with paths that are a.s. right continuous with left limits) process. We assume that the logarithm of a financial asset price $(S_t)_{t\geq 0} \equiv S_t$ is modeled by an Itô process and the instantaneous variance $(V_t)_{t\geq 0} \equiv V_t$ follows a *mean-reverting* process, that is

$$\begin{aligned} dS_t &= r dt + \sqrt{V_t} dW_t \\ dV_t &= \kappa(\omega - V_t)dt + \beta_t dZ_t \end{aligned}, \tag{4}$$

where $\kappa \in \mathbb{R}$ is the *rate of mean reversion*, $\omega \in \mathbb{R}$ is the *long-run mean level* of V_t, $(W_t)_{t\geq 0} \equiv W_t$ and $(Z_t)_{t\geq 0} \equiv Z_t$ are two correlated Wiener processes, with correlation coefficient equal to $\rho \in [-1,1]$, thus introducing an asymmetric return-variance relation into the asset price dynamics. Integrating the expected value $\mathbf{E}^{\mathbb{Q}}[V_t \mid \mathfrak{F}_t]$ with respect to time, we obtain

$$\mathbf{E}^{\mathbb{Q}}\left[\int_0^t V_s\, ds\right] = \frac{V_0}{\kappa}(1 - \mathrm{e}^{-\kappa t}) + \omega t - \frac{\omega}{\kappa}(1 - \mathrm{e}^{-\kappa t}). \tag{5}$$

The last equation represents our theoretical model, which is semi-nonparametric since it does not depend on β_t, but only on the drift parameters.

The inference from option prices can be complicated by the fact that these prices are often observed with some error. Moreover, a continuum of strikes is not available, since generally options are quoted in the market with discrete strikes. Thus, in order to take into account these sources of uncertainty, we applied a Generalized Method of Moments approach, considering an empirical approximation of the true value of the expected integrated risk-neutral variance, plus the contribution of a random variable which represents the observation error on option prices. Then, the VIX term structure obtained from market data can be fitted with our theoretical model by minimizing the mean squared error, to obtain estimates of the parameters (κ, ω, V_0).

Some advantages of this method can be outlined. Firstly, it is a very intuitive procedure, which does not require heavy computational algorithms. Second, it allows us to relax the strong parametric assumptions we find in the existent estimation methods (see Andersen, Fusari and Todorov [20]). Indeed, our theoretical model is completely semi-nonparametric, since it does not depend on β_t, and it allows us to obtain a valid estimate of the spot volatility parameter only by means of options market data. This approach should give more accurate estimates than those obtained with high-frequency data (see Andersen, Fusari and Todorov [20]). Daily market data are available in real time free from biases, in order to avoid any arbitrage opportunity. On the contrary, estimation methods based on high-frequency data often use larger data sets, composed by historical volatilities. As a consequence, even though they can reduce the variance of an estimator, they surely present a much higher bias, since historical volatilities are well away from the instantaneous volatility measure. Hence, this represents a clear theoretical advantage of an options based methodology. Moreover, the estimation procedure described above can be applied to a wide class of stochastic volatility models. For instance, the celebrate Heston [97] model can be obtained with $\beta_t = \sigma\sqrt{V_t}$, while the Luo and Zhang [120] model is a special case with stochastic long-run mean level θ.

In order to study the finite sample properties of our estimator, we perform 1000 Monte Carlo simulations based on the Heston [97] model. To construct the data set, we fix five maturities, from two weeks to one year. Then, for each maturity, we compute option prices considering four possible scenarios: a low number of quoted options, only 10, observed with low [resp. high] error, that is 1% [resp. 5%], and then a higher number of options, say 50, observed with same low and high errors. Simulation results clearly show that a larger sample of available options, observed with same error, reduces uncertainty, while with same number of options, the higher observation error increases the dispersion of the parameters distribution.

In the second part of this work, we focus on an option pricing problem. Despite the fact that the financial literature concerning stochastic volatility models is quite plentiful, in quantitative finance it is still open the long-lasting challenging problem to find a practical option pricing model which, at the same time,

1. is guaranteed to be free of arbitrage opportunities;

2. provides fast algorithms for prices and greeks calculation;

3. is able to fit the quoted volatility surfaces, across both maturities and strikes;

4. adequately describes price and volatility risk.

These four features are equally needful, but it is not possible to find an existent model which satisfies them all together. For instance, arbitrage-free volatility interpolation can be achieved with local volatility models (Derman and Kani [77] and Dupire [80]), as shown recently in Andreasen and Huge [22] and Kahalé [109], among others. However, local volatility models do not accommodate idiosyncratic volatility risk. On the other hand, the Heston [97] model includes volatility risk in a mathematically convenient way, but it is not flexible enough to fit the whole volatility surface. Similarly, the SABR (Hagan et al. [95]) model is almost arbitrage-free and has an analytical pricing formula, but it can hardly reproduce volatility surfaces, if parameters are not changed along with maturities. The goal of this chapter is to fill this gap, introducing the *Heston++* class of stochastic volatility models, which is compliant to the four pillars mentioned above.

Consider the price of a non dividend-paying risky equity asset $(S_t)_{t\geq 0} \equiv S_t$, evolving in continuous time. Under the risk-neutral probability measure $\mathbb{Q}$, we assume that S_t follows the general dynamics

$$dS_t = rS_t dt + \sqrt{V_t} S_t dW_t^S, \tag{6}$$

$$V_t = v_t + \phi_t, \tag{7}$$

$$dv_t = \kappa(\omega - v_t)dt + \Lambda\sqrt{v_t} dW_t^v, \tag{8}$$

where r, κ, ω and $\Lambda \in \mathbb{R}_+$ are constants, $\phi_t \geq 0$ is a smooth enough non-negative real function integrable on closed intervals with $\phi_0 = 0$, and $\left(W_t^S\right)_{t\geq 0} \equiv W_t^S$ and $(W_t^v)_{t\geq 0} \equiv W_t^v$ are correlated Wiener processes on Ω adapted to the filtration $\mathfrak{F}_t$, with time dependent instantaneous risk-neutral correlation given by

$$\mathrm{corr}^{\mathbb{Q}}\left(dW_t^S, dW_t^v\right) = \rho\sqrt{\frac{v_t}{v_t + \phi_t}}, \tag{9}$$

where $\rho \in [-1, 1]$ is an additional constant.

Then, the one-factor model is specified by the risk-free rate, five parameters $(\kappa, \omega, \Lambda, \rho, v_0)$ and the deterministic function ϕ_t. The Heston++ model is an extension of the standard Heston [97] model, obtained by adding a deterministic factor $\phi_t \neq 0$ in the volatility process, meant to fit the term structure of the at-the-money implied variances, without sacrificing the computational advantages of the Heston model. We also present a two-factor and two-factor with jumps Heston++ models, which we have developed by applying the idea of an additional deterministic volatility factor to the original model proposed in Christoffersen, Heston and Jacobs [67], Bates [37] and [38] and Duffie, Pan and Singleton [79].

The technique is borrowed from deterministic shift extension (Brigo and Mercurio [47]) and its application to short rate models. The preliminary fit of the ATM term structure strongly eases the fit of the whole volatility surface, since the model parameters are employed with their full flexibility. The straightforward application of our model allows a fast and accurate arbitrage-free interpolation, as well as a sound extrapolation, of the volatility surface, which can be used for market making or exotic pricing. Hence, the model we propose here unifies the advantages of local volatility models and those of stochastic volatility models. An approach similar to the Heston++ model is represented by the Heston model with time-dependent parameters, introduced by Mikhailov and Nögel [129] and then further deepened by Elices [83], in which the constant parameters are replaced by deterministic functions. However, our model has an inherently different specification and presents, in our view, several advantages. The difference consists in the fact that in the Heston++ model the function ϕ_t can be interpreted as the lower bound for the spot variance, as it is clear from Equation (4.25), while in the Heston model with time-dependent parameters the lower bound for the local variance is still equal to 0. This technical difference is harmless from a specification point of view, but it is crucial in making the pricing of European vanilla options straightforward. Furthermore, the Heston model with time-dependent parameters displays serious mathematical troubles, which are typically solved by virtue of approximations based on Taylor expansions or using deterministic functions which are piecewise constant[1]. Moreover, its extension to multi-factor models with jumps appears cumbersome. Conversely, the Heston++ model has no additional mathematical and implementation complexity with respect to the traditional Heston model, providing simple formulas and fast algorithms, and can be easily generalized to affine models, such as those in Bates [37], Bates [38], Christoffersen et al. [67] and Duffie et al. [79], or to Wishart models, in the sense of Da Fonseca et al. [72].

To illustrate the gain in terms of pricing, we calibrate the model on a time series of daily option panels on FX rate EUR/USD from 2005 to 2012, for strikes up to 10Δ and 10 maturities ranging from one week to two years. The Heston++ models are readily fitted, with no added computational cost with respect to the standard versions, and obtain an average root mean square error of 1.26%, in the best case.

This work has the following structure. The first chapter is a practical guide to the foreign exchange option market, illustrating the various quote styles, deltas definitions and market conventions, with plenty of examples. Then, we present the Vanna-Volga method and we propose an interpretation of the procedure in the context of European plain vanilla options. In the second chapter, we introduce the volatility index, presenting the generalized formula for its calculation, and we describe its economic and theoretical interpretation. In the third chapter, we show our new semi-nonparametric estimation method of the spot volatility, based on option prices. Finally, in the last chapter, we briefly review the main properties of the Heston [97] model. Next, we illustrate the proposed class of stochastic volatility models Heston++ and provide the corresponding option pricing formulae. Then, we show empirical results of the Heston++ model, suitably calibrated to the foreign exchange option market.

[1] See Benhamou et al. [39] for a detailed discussion.

Chapter 1

The Foreign Exchange Market

Nowadays, in the prevailing financial markets, options with different strikes or maturities are usually priced with different implied volatilities. This stylized fact, commonly known as *smile effect*, is often taken into account in specific models, with the goal of either pricing exotic derivatives or inferring implied volatilities for non quoted strikes or maturities. While the former task is typically achieved by introducing alternative dynamics for the underlying asset price, the latter is often tackled by means of statistical adjustments or interpolations.

In this part of our work, we deal with the latter topic, in order to analyze a possible solution in a *foreign exchange* (FX) option market setting. In such a market, three volatility quotes are available, namely the at-the-money, the risk reversal and the strangle, thus arising the issue of a consistent determination of the whole volatility surface. In particular, FX quants and market makers address this problem by using an empirical procedure, called the *Vanna-Volga* (VV) method, to construct the smile for a given maturity. Furthermore, volatilities are quoted in terms of option deltas, for ranges from 5Δ put to 5Δ call.

In the following, we attempt to provide a general description of the FX option market, illustrating the various quote styles and deltas definitions. Then, we extensively review market conventions, with plenty of examples. Finally, we present the Vanna-Volga method and we propose an interpretation of the procedure in the context of European plain vanilla options.

1.1 Introduction to the FX Option Market

A crucial ingredient to the Vanna-Volga method, that is often overlooked in the literature, is the correct handling of the market data. In FX markets, the precise meaning of broker quotes depends on the contract details. For instance, there are at least four different definitions of delta, such as spot, spot percentage, forward and forward percentage. Using the wrong definition can lead to significant errors in the construction of the smile. Therefore, before we begin to explore the effectiveness of the Vanna-Volga technique, we present some of the relevant FX conventions.

The *foreign exchange* (FX) market, also called *forex* or *currency* market, is the landmark for the global decentralized trading of international currencies. Its primary role is to determine relative values of different currencies, in order to assist the international trade and investments by enabling currency conversion.

During the 1970s, after three decades of government restrictions on foreign exchange transactions, countries gradually switched from the fixed exchange rate regime, previously established by the Bretton Woods system, to floating exchange rates.

Due to its specific characteristics, the FX market significantly differs from other financial markets. In particular, some of these features are:

- its huge trading volume, representing the largest asset class in the world;
- its geographical dispersion;
- its 24-hour activity, except weekends;
- its direct influence on the international trade pattern of every country;
- the low margins of relative profit, compared with other markets;
- the intensive use of leverage to improve profit and loss margins.

It is the most liquid over the counter (OTC) financial market in the world and, despite currency intervention by central banks, it is considered the closest market to the idea of perfect competition. At the same time, it is also the largest market for options. Traders include central banks, governments, large banks, institutional investors, currency speculators, corporations, exporters and importers, retail investors and other financial institutions.

The FX market comprises three different markets, with distinct functions, but still closely related: the spot market, the futures market and the derivatives market. Currencies for immediate delivery are traded on the *spot market*. For instance, the purchase of foreign currency at the 'bureau de change', is the simplest spot market transaction possible.

Definition 1.1. *For a currency pair quoted*[1] *as ccy1ccy2, the* spot rate S_t*, at time* t*, is the number of units of ccy2, known as the* domestic currency*, the* terms currency *or the* quote currency*, required to buy one unit of ccy1, the* foreign currency *or the* base currency*.*

Therefore, the spot rate is dimensionally equal to units of ccy2 per ccy1. The GBPUSD quote is for US dollars per pound sterling, so if GBPUSD is 1.6011, then one British pound can be bought for $1.6011 in the spot market. In other words, it is the cost of one pound sterling in US dollars. The actual exchange of the two currencies, which is called *settlement*, is handled through the banking system and occurs at the *spot date*, which is generally in a couple of days, although some trades, such as exchanges of US dollars for Canadian dollars, are settled more quickly.

To this type of transaction is associated the so-called *foreign exchange settlement risk*, which occurs when one of the two payments does not go through. This is also known as the *Herstatt risk*, from the German bank default on dollar payments in 1974.

[1]The use of abbreviations, such as ccy1 and ccy2, is standard among FX market practitioners.

On the other hand, the FX *forward market* allows participants to lock in the actual exchange rate in order to protect against *currency risk*, that is, the exposure to an unexpected variation of the currency rate. In this case, the forward rate is fixed today and funds are going to be transferred in the future. The prearranged date in which the settlement takes place is called the *delivery date*.

Nevertheless, most of the FX trading occurs in the *derivatives market*. Currently, the various traded products range from simple vanilla options and first-generation exotics, such as touch-like and barriers options, to second-generation exotics, that is options with a fixing-date structure or options with no available closed form value, and even third-generation exotics, hybrid products between different asset classes.

Definition 1.2. *A* FX option*, or* currency option*, is a financial derivative contract between two parties, the holder and the writer, which gives the holder the right, but not the obligation, to buy from (*call option*) or sell to (*put option*) the writer a currency pair at a given exchange rate, called* strike price*, within a predetermined date T, known as* maturity *or* expiry date*.*

Since the option gives the holder a right with no obligation, she has to pay some money, the so-called *premium*, on signing the contract. For instance, suppose that an economic agent is structurally long in pounds sterling. To protect against depreciation of the sterling amount until the delivery date, she can buy a put option on the GBPUSD exchange rate, that is, an option to buy USD, or equivalently sell GBP, at a predetermined strike price.

Unlike other asset classes, in the FX market there is no natural numeraire currency. As a result, there is no special reason to quote spot or forward rates for foreign currencies in any particular order. Hence, the choice of which way a particular currency pair should be quoted is purely market convention and seems to be quite arbitrary. For British pounds against US dollars, with ISO[2] codes of GBP and USD respectively, the market standard quote could be GBPUSD, the price of 1 GBP in USD, or USDGBP, the price of 1 USD in GBP. By market convention, the former is commonly used, as we can see in Figure 1.1, which shows some of the most frequently traded currency pair quotations. The following hierarchy can be useful in order to remember which currency tends to be considered the ccy1 most of times:

$$\text{EUR} > \text{GBP} > \text{AUD} > \text{NZD} > \text{USD} > \text{CAD} > \text{CHF} > \text{JPY}.$$

An interesting note is that the mainstream financial press, such as the *Financial Times* and the *Economist*, report all currencies in the same quote terms, such as the value of one US dollar in each of AUD, CAD, EUR, GBP, . . . , ZAR. While easier to understand, this is not the real way in which currency pairs are quoted in FX markets.

Furthermore, as FX quotations are made to finite precision, the least significant digit of the spot rate is called a *pip*. It represents the smallest usual price increment possible in the FX spot market. On the other hand, a *big figure* is invariably 100 pips. For example, if the spot rate for EURUSD is 1.4591, the big figure is 1.45 and there are 91 additional pips in the price.

[2] ISO 4217 code, from the International Organization for Standardization (www.iso.org).

Currency pair	Common trading floor jargon
EURUSD	Euro-dollar
USDJPY	Dollar-yen
EURJPY	Euro-yen
GBPUSD	Cable (from the late 1800s transatlantic telegraph cables)
EURGBP	Euro-sterling
USDCHF	Dollar-swiss
AUDUSD	Aussie-dollar
NZDUSD	Kiwi-dollar
USDCAD	Dollar-cad (or dollar-canada or, less commonly, dollar-loonie)
EURNOK	Euro-nokkie
EURSEK	Euro-stockie (from Stockholm)
EURDKK	Euro-danish
EURHUF	Euro-huff ('huff', not H-U-F)
EURPLN	Euro-polish
USDTRY	Dollar-turkey (or dollar-try, pron. 'try')
USDZAR	Dollar-rand (or dollar-zar, pron. 'zar')
USDMXN	Dollar-mex ('mex', not M-E-X)
USDBRL	Dollar-brazil
USDSGD	Dollar-sing

Figure 1.1: Currency Pair Quotation Conventions and Market Terminology

It is worth remembering that precious metals are quoted in the same way as currencies. In particular, gold and silver have ISO codes XAU and XAG, respectively. Similarly, we have XPT for platinum and XPD for palladium. Thus, a currency pair XAUUSD with spot rate equal to 1000.00 means that one ounce of gold is worth $1000.

1.2 Deltas and Market Conventions

Since FX volatility surfaces are specified in terms of volatilities corresponding to different deltas, market participants need a clear understanding of which deltas are available, and this is where the connection between deltas and market conventions arises. Price and volatility at different delta are the main specifications of the options quoted in the FX market, thus, it is necessary to analyze in detail the various types of FX delta.

Given the Black-Scholes [42] price for an option, it is possible to compute the variation of that price with respect to infinitesimal changes in the underlying spot or forward exchange rate. This clearly represents the concept of the option *delta*.

Definition 1.3. *The* delta *of an option is defined as the instantaneous sensitivity of the option price to infinitesimal changes in the underlying asset price.*

In most asset classes other than FX, the delta is perfectly straightforward. This is not true in FX markets, due to the existence of different quotation styles for prices, which leads to several types of delta. Since the entire concept of the FX volatility smile is based on parametrization with respect to delta, a deep study of the available deltas is of crucial importance. Therefore, in this section we introduce the fundamental market conventions commonly used in practice, with the aim of providing a comprehensive description of all different types of delta encountered. We then present the conditions that a volatility smile must satisfy in order to be consistent with market volatilities, given a fixed maturity T.

In currency markets, as opposed to equity markets, four relative quote styles are available for options, that is, *domestic per foreign* (d/f), *foreign per domestic* (f/d), *percentage foreign* (%f) and *percentage domestic* (%d). Since currencies have economic purchasing power in their respective home countries, whereas equities do not have this property anywhere, and since in a typical currency exchange there are two notionals, investors in the FX market can have one of two numeraires[3]. Therefore, a risk-neutral investor in the domestic currency can obtain a domestic per domestic price or a domestic per foreign price. Similarly, a risk-neutral investor in the foreign currency can obtain a foreign per foreign price or a foreign per domestic price.

Now, let S_0 be the current spot exchange rate and let r^d and r^f be the continuous domestic and foreign risk-free interest rates. Consider a European vanilla option, with maturity T and strike K, whose value is denoted by Q. The standard Black-Scholes [42] formula[4] for this option is given by

$$Q_{d/f} = \phi S_0 e^{-r^f T} N\left(\phi d_1\right) - \phi K e^{-r^d T} N\left(\phi d_2\right), \tag{1.1}$$

where

$$d_{1,2} = \frac{\ln \frac{S_0}{K} + \left(r^d - r^f \pm \frac{1}{2}\sigma^2\right) T}{\sigma\sqrt{T}}, \tag{1.2}$$

σ is the volatility process, $N\left(x\right)$ is the cumulative density function of the normal distribution and ϕ is the option indicator, equal to 1 for a call and to -1 for a put. Equation (1.1) represents the value of an option, whether a call or a put, expressed in the domestic currency, since it gives the right to exchange K units of domestic currency for 1 unit of foreign currency. Hence, we can refer to $Q_{d/f}$ as the *domestic/foreign price.* Sometimes, it is also called the ccy2/ccy1 price or the domestic pips price.

To obtain the *percentage foreign price*, we can merely take the d/f price above and convert the domestic currency value into foreign currency terms, using the current spot exchange rate S_0, that is

$$Q_{\%f} = \frac{Q_{d/f}}{S_0}. \tag{1.3}$$

Since $Q_{d/f}$ is the price in the domestic currency of an option with a unit notional in the foreign currency, equivalently, it also represents the price in the domestic currency of an option with a notional of K in the domestic currency. Hence, Equation (1.1) divided by K gives the *percentage domestic price*:

$$Q_{\%d} = \frac{Q_{d/f}}{K}. \tag{1.4}$$

[3] Note that true numeraires are money market accounts, in either currency. In accordance with the common short dated FX market practice, we assume that discount factors are deterministic and can be removed from expectations. Thus, we refer to ccy1 and ccy2 as numeraires, where we should really call it quasi-numeraires.

[4] The equivalent of the Black-Scholes model in the FX setting is the Garman-Kohlhagen [91] model, so whenever in this work we refer to the Black-Scholes model, we mean the Garman-Kohlhagen model.

From the last equation, the *foreign/domestic price* is obtained by converting the ccy2 value of the option into ccy1. Since 1 unit of the domestic currency is worth $1/S_0$ units of the foreign currency, the foreign/domestic price is equal to

$$Q_{f/d} = \frac{Q_{\%d}}{S_0} = \frac{Q_{d/f}}{S_0 K}. \tag{1.5}$$

The four prices presented above are relative to a notional of K in the domestic currency and to a unit notional in the foreign currency. Now, we can obtain two absolute prices Q_d (domestic) and Q_f (foreign), given actual notionals N_d and N_f in domestic and foreign currencies, respectively, with $N_d = K \times N_f$. We then have

$$Q_d = N_f \times Q_{d/f}, \tag{1.6}$$

$$Q_f = \frac{N_f}{S_0} \times Q_{d/f}. \tag{1.7}$$

All the possible quote styles are summarized below:

$$\begin{aligned} Q_{d/f} &= \phi \left[S_0 e^{-r^f T} N(\phi d_1) - K e^{-r^d T} N(\phi d_2) \right] \\ Q_{\%f} &= \frac{Q_{d/f}}{S_0} \\ Q_{\%d} &= \frac{Q_{d/f}}{K} \\ Q_{f/d} &= \frac{Q_{d/f}}{S_0 K} \\ Q_d &= N_f \times Q_{d/f} \\ Q_f &= \frac{N_f}{S_0} \times Q_{d/f}. \end{aligned} \tag{1.8}$$

It has to be stressed that this technique of constructing all these different quotation styles works only when there are two notionals, in foreign and domestic currencies, with a fixed relation between them known from the start. This is true for European and American vanilla options, even in the presence of barriers, but it is definitely not true for digital options. Consider a cash-or-nothing digital option, which pays one dollar if the EURUSD exchange rate is above a prearranged level at a certain maturity. Clearly, the digital has a unit notional in USD, so we can compute percentage domestic and foreign/domestic prices. However, there is no notional in the foreign currency, then, the other two quotes are meaningless.

According to the *law of one price*, investors must agree on the Black-Scholes [42] price of an FX option, regardless of which of the two currencies they consider domestic or foreign, as long as they agree on market components, such as spot rate, volatility and interest rates. Nonetheless, we have just seen above that there are several ways to quote prices of currency options. As a consequence, there must be different corresponding ways to quote the option delta. This is not surprising, because the delta is defined as the instantaneous derivative of the option price with respect to changes in the underlying price. Therefore, if there exist various quote styles for the option price, then, there are several ways to construct the delta. In market jargon, this is called the *law of many deltas*, opposed to the law of one price.

In order to derive deltas, first we recall some necessary technicalities:

$$\frac{\partial d_1}{\partial S_0} = \frac{\partial d_2}{\partial S_0} = \frac{1}{\sigma S_0 \sqrt{T}},$$

and then

$$\frac{\partial N(\phi d_k)}{\partial S_0} = \phi N'(\phi d_k) \frac{\partial d_k}{\partial S_0} = \frac{\phi}{\sigma S_0 \sqrt{T}} N'(\phi d_k), \qquad k = 1, 2.$$

Furthermore, the following result holds in general:

$$N'(\phi d_1) = N'(\phi d_2)\, e^{\phi^2 (r^f - r^d) T} \left(\frac{K}{S_0}\right)^{\phi^2}.$$

Proof. Since

$$N(\phi d_k) = \int_{-\infty}^{\phi d_k} f(u)\, du = \int_{-\infty}^{\phi d_k} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}\, du, \qquad k = 1, 2,$$

we have

$$N'(\phi d_1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(\phi d_1)^2}{2}},$$

and

$$\begin{aligned}
N'(\phi d_2) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{(\phi d_2)^2}{2}} \\
&= \frac{1}{\sqrt{2\pi}} e^{-\frac{\phi^2 (d_1 - \sigma\sqrt{T})^2}{2}} \\
&= \frac{1}{\sqrt{2\pi}} e^{-\frac{\phi^2 d_1}{2}} e^{-\frac{\phi^2 \sigma^2 T}{2}} e^{\phi^2 d_1 \sigma \sqrt{T}} \\
&= \frac{1}{\sqrt{2\pi}} e^{-\frac{\phi^2 d_1}{2}} e^{-\frac{\phi^2 \sigma^2 T}{2}} e^{\phi^2 \ln \frac{S_0}{K}} e^{\phi^2 \left(r^d - r^f + \frac{1}{2}\sigma^2\right) T} \\
&= \frac{1}{\sqrt{2\pi}} e^{-\frac{\phi^2 d_1}{2}} e^{\phi^2 (r^d - r^f) T} \left(\frac{S_0}{K}\right)^{\phi^2}.
\end{aligned}$$

Hence,

$$N'(\phi d_1) = N'(\phi d_2)\, e^{\phi^2 (r^f - r^d) T} \left(\frac{K}{S_0}\right)^{\phi^2},$$

which is the desired result. □

In particular, we are only interested in $|\phi| = 1$, for which $\phi^2 = 1$.

We start with the definition of the *domestic/foreign spot delta*, which is the ratio of the change in the option price to the infinitesimal variation in the spot rate, both expressed in domestic/foreign terms, that is

$$\Delta_{S;d/f} = \lim_{\Delta S_0 \to 0} \frac{\Delta Q_{d/f}}{\Delta S_0} = \frac{\partial Q_{d/f}}{\partial S_0}, \tag{1.9}$$

where $\Delta Q_x = Q_x(S_0 + \Delta S_0) - Q_x(S_0)$, for any quotation style x. Now, taking the derivative of $Q_{d/f}$ in (1.1) with respect to S_0, we have

$$\begin{aligned}\Delta_{S;d/f} &= \phi e^{-r^f T} N(\phi d_1) + \phi S_0 e^{-r^f T} \frac{\partial N(\phi d_1)}{\partial S_0} - \phi K e^{-r^d T} \frac{\partial N(\phi d_2)}{\partial S_0} \\ &= \phi e^{-r^f T} N(\phi d_1) + \frac{\phi^2}{\sigma\sqrt{T}} e^{-r^f T} N'(\phi d_1) - \frac{\phi^2 K}{\sigma S_0 \sqrt{T}} e^{-r^d T} N'(\phi d_2) \\ &= \phi e^{-r^f T} N(\phi d_1) + \frac{\phi^2}{\sigma S_0 \sqrt{T}} \left(S_0 e^{-r^f T} N'(\phi d_1) - K e^{-r^d T} N'(\phi d_2) \right) \\ &= \phi e^{-r^f T} N(\phi d_1). \end{aligned} \tag{1.10}$$

Hence, the domestic/foreign spot delta (1.10) is the standard Black-Scholes [42] delta, up to a factor $e^{-r^f T}$. It is also known as *pips spot delta*. To the end of hedging, the spot delta (1.10) is the number of units of foreign currency we need to hold in order to hedge an option with a unit notional of foreign currency and an equivalent notional of K units of domestic currency. Thus, the spot delta is expressed as a percentage of the foreign currency.

On the other hand, the *percentage spot delta* is the ratio of the change in the option value to the variation in the spot rate, both expressed in percentage foreign terms. We then have

$$\Delta_{S;\%} = \lim_{\Delta S_0 \to 0} \frac{\Delta Q_{\%f}}{\Delta S_0 / S_0} = S_0 \frac{\partial Q_{\%f}}{\partial S_0}. \tag{1.11}$$

Now, substituting (1.3) in (1.11), we obtain

$$\begin{aligned}\Delta_{S;\%} &= S_0 \frac{\partial \left(\frac{Q_{d/f}}{S_0} \right)}{\partial S_0} \\ &= S_0 \left(\frac{\partial Q_{d/f}}{\partial S_0} S_0 - Q_{d/f} \right) \frac{1}{S_0^2} \\ &= \frac{\partial Q_{d/f}}{\partial S_0} - \frac{Q_{d/f}}{S_0}. \end{aligned} \tag{1.12}$$

Hence, the percentage spot delta above represents the *premium-adjusted* pips spot delta, that is, the domestic/foreign spot delta in (1.10) adjusted with the percentage foreign price of the option:

$$\Delta_{S;\%} = \Delta_{S;d/f} - Q_{\%f}. \tag{1.13}$$

In order to understand why premium adjustment should be required, consider the case of a EURUSD call option. If the two counterparties are euro and US dollar based, respectively, then they agree on the price, as stated by the law of one price. However, this price is expressed and actually exchanged in only one of the two currencies, EUR or USD. Since the market convention favors a premium in US dollars, in this case the premium itself is riskless for the dollar investor, while it constitutes an extra source of currency risk for the euro investor. As a result, the euro based investor wants to premium adjust the delta. In other words, a euro based investor can either obtain premium-adjusted deltas for FX call options on the spot rate EURUSD or she can construct spot deltas, with no premium adjustment, for FX put options on the flipped spot rate USDEUR.

The *domestic/foreign forward delta* is the ratio of the change in the future value of the option to the infinitesimal variation in the forward exchange rate $F_{0,T} = S_0 e^{(r^d - r^f)T}$, both quoted in ccy2/ccy1 terms, that is

$$\begin{aligned} \Delta_{F;d/f} &= \lim_{\Delta F_{0,T} \to 0} \frac{\Delta \left(e^{r^d T} Q_{d/f} \right)}{\Delta F_{0,T}} \\ &= e^{r^d T} \frac{\partial Q_{d/f}}{\partial F_{0,T}} \\ &= e^{r^d T} \left(\frac{\partial F_{0,T}}{\partial S_0} \right)^{-1} \frac{\partial Q_{d/f}}{\partial S_0} \\ &= e^{r^f T} \Delta_{S;d/f} = \phi N \left(\phi d_1 \right). \end{aligned} \tag{1.14}$$

Now, if we consider the ratio of the change in the option future value to the change in the relevant forward, both expressed in percentage foreign terms, we obtain the *percentage forward delta*, given by

$$\begin{aligned} \Delta_{F;\%} &= \lim_{\Delta F_{0,T} \to 0} \frac{\Delta \left(e^{r^d T} Q_{\% f} \right)}{\Delta F_{0,T} / F_{0,T}} \\ &= e^{r^f T} \left(\Delta_{S;d/f} - Q_{\% f} \right) \\ &= e^{r^f T} \Delta_{S;\%}. \end{aligned} \tag{1.15}$$

Thus, substituting Equations (1.1) and (1.10) in (1.12), we obtain

$$\Delta_{S;\%} = \phi e^{-r^d T} \frac{K}{S_0} N(\phi d_2), \tag{1.16}$$

$$\Delta_{F;\%} = \phi \frac{K}{F_{0,T}} N \left(\phi d_2 \right). \tag{1.17}$$

The last delta we introduce is not a proper delta in the usual sense, but it is rather a simplified measure of moneyness, intermediate between $\phi N(\phi d_1)$ and $\phi N(\phi d_2)$, used for ease of computation instead of other deltas, when the construction of parametric functions of delta is needed. This *simple delta* is given by

$$\Delta_{simple} = \phi N \left(\phi d \right), \qquad d = \frac{\ln \left(F_{0,T} / K \right)}{\sigma \sqrt{T}}, \tag{1.18}$$

where d is the arithmetic average of d_1 and d_2 in (1.2).

Remark 1.1. *Note that the percentage spot delta (1.12) for an investor with ccy2 numeraire is related to the domestic/foreign spot delta (1.10) for an investor with ccy1 numeraire, by a negative constant multiple.*

Currency pair	ccy1	ccy2	Premium ccy	Δ convention
EURUSD	EUR	USD	USD	Pips
USDJPY	USD	JPY	USD	%
EURJPY	EUR	JPY	EUR	%
USDCHF	USD	CHF	USD	%
EURCHF	EUR	CHF	EUR	%
GBPUSD	GBP	USD	USD	Pips
EURGBP	EUR	GBP	EUR	%
AUDUSD	AUD	USD	USD	Pips
AUDJPY	AUD	JPY	AUD	%
USDCAD	USD	CAD	USD	%
USDBRL	USD	BRL	USD	%
USDMXN	USD	MXN	USD	%

Figure 1.2: Delta conventions for common currency pairs

Proof. For an investor whose numeraire is currency 1, the delta is equal to

$$\begin{aligned}
\Delta_{S;f/d} &= \frac{\partial Q_{f/d}}{\partial \hat{S}_0} = \frac{\partial Q_{f/d}}{\partial\,(1/S_0)} \\
&= \frac{1}{\partial\,(1/S_0)} \partial \left(\frac{Q_{d/f}}{S_0 K} \right) \\
&= -\frac{1}{1/S_0^2} \frac{1}{K} \left(\frac{\frac{\partial Q_{d/f}}{\partial S_0} S_0 - Q_{d/f}}{S_0^2} \right) \\
&= -\frac{S_0}{K} \left(\frac{\partial Q_{d/f}}{\partial S_0} - \frac{Q_{d/f}}{S_0} \right) \\
&= -\frac{S_0}{K} \Delta_{S;\%}.
\end{aligned}$$

□

Hence, due to the duality among the two currencies and the change between domestic and foreign risk-neutral measures, it is always possible to establish a clear relationship between the two styles of delta, that is, domestic per foreign and premium-adjusted.

Since we have introduced several definitions of deltas, the choice of which one has to be used depends on the circumstance. For instance, whether to apply the premium adjustment or not completely depends on which currency pair we are dealing with and is determined by the choice of the premium currency, see Table 1.2. Basically, if the premium currency is ccy2, then no premium adjustment is applied, whereas if the premium currency is ccy1, then percentage delta is considered. Typically, the premium currency is taken to be the more commonly traded currency of the two, with the exception of JPY, which is rarely the premium currency by market convention. Table 1.3 shows the choice of premium currency for various currency pairs.

Furthermore, since in 2008 the credit crunch has induced low levels of liquidity in short-term interest rate products, it became unfeasible for banks to agree on spot deltas, which include

	ccy2							
ccy1	USD	EUR	GBP	AUD	NZD	CAD	CHF	JPY
USD						USD	USD	USD
EUR	USD		EUR	EUR	EUR	EUR	EUR	EUR
GBP	USD			GBP	GBP	GBP	GBP	GBP
AUD	USD				AUD	AUD	AUD	AUD
NZD	USD					NZD	NZD	NZD
CAD							CAD	CAD
CHF								CHF
JPY								

Figure 1.3: Premium currency for major currency pairs

discount factors. As a consequence, market practice has largely shifted to the use of forward deltas in the construction of FX smiles. Events of 2008 show how market conditions can cause market conventions to evolve. Hence, it is always advisable to check current market practice, in particular during or after extreme market conditions.

Summary

$$\begin{aligned}
\Delta_{S;d/f} &= \phi e^{-r^f T} N\left(\phi d_1\right) \\
\Delta_{S;\%} &= \phi e^{-r^d T} \frac{K}{S_0} N(\phi d_2) \\
\Delta_{F;d/f} &= \phi N\left(\phi d_1\right) \\
\Delta_{F;\%} &= \phi \frac{K}{F_{0,T}} N\left(\phi d_2\right) \\
\Delta_{simple} &= \phi N\left(\phi d\right). \qquad (1.19)
\end{aligned}$$

1.3 Market Volatility Surfaces

For historical reasons, the FX option market uses the so-called *sticky delta rule* to build volatility smiles, which implies that volatilities are quoted as a function of delta rather than strikes value. Practically, this means that when the underlying asset price moves and the option delta changes accordingly, a different implied volatility has to be plugged into the pricing formula. The choice of delta as a parameter to describe the volatility smile is reasonable, otherwise a strike that might correspond to a considerably out-of-the-money option at a small maturity would be very close to the at-the-money for a larger maturity. Therefore, the parametrization of the smile as a function of deltas allows a better coverage of the volatility surface, given a known selection of deltas.

Since the Black-Scholes [42] formula is inadequate to price options of all strikes and maturities

EURUSD (spot reference 1.3940)					
Maturity	σ_{ATM}	σ_{25-RR}	σ_{25-MS}	σ_{10-RR}	σ_{25-MS}
1W	20.37	0.24	0.64	0.47	2.20
2W	20.86	0.30	0.67	0.54	2.27
1M	24.15	0.40	0.71	0.69	2.68
2M	24.12	0.31	0.76	0.52	2.89
3M	23.74	0.24	0.83	0.40	3.19
6M	21.52	0.11	0.90	0.19	3.41
9M	20.26	0.01	0.94	0.03	3.58
1Y	19.32	-0.07	1.00	-0.09	3.76
18M	18.70	-0.11	1.00	-0.13	3.69
2Y	18.19	-0.14	0.90	-0.17	3.48

Table 1.1: Sample market volatility surface for EURUSD on 2008, 22 December

consistently with the market, it is necessary to construct a volatility smile $\sigma(K)$ that assigns a volatility value to each strike K. Malz [123] provides a good description, in terms of market parameters, of the way in which market behavior deviates from the Black-Scholes prediction. However, he does not clarify neither which delta has to be used nor how to identify the at-the-money strike. Furthermore, the strangle definition he gives in his work refers to the *smile strangle*, which does not correspond to the typical strangle observed and traded in the market, known as the *market strangle*. In the rest of this chapter, we want to give a more complete and thorough discussion about these topics.

In the FX option market, 25Δ[5] call and 25Δ put are commonly traded, for a given maturity T. Thus, together with the at-the-money quote, they are considered benchmark strikes for the construction of the FX volatility surface. If we prescribe volatilities at these three strikes, we obtain the so-called *three-point smile*. Furthermore, a *five-point smile* can be constructed combining these three strikes with 10Δ put and 10Δ call quotations, which are less liquid but still available.

A particular property of the FX option market, which makes it different from other financial markets, consists in the fact that market volatility surfaces are described by the measurement of three components, for each maturity: the *at-the-money* (ATM), the *risk reversal* (RR) and the *market strangle* (MS). These three quotes represent separate and nonoverlapping constraints on the volatility smile, in order to be consistent with the market.

A clear example of a volatility surface in given by the market snapshot in Table 1.1, which is collected on 2008, 22 December and describes market volatility quotations, expressed in percentage values. Before starting the smile construction, it is important to analyze the exact characteristics of the quotes in Table 1.1. In particular, one has to identify first which delta type and which at-the-money convention are used. It is obvious that smiles can have very different shapes, in particular for out-of-the-money and in-the-money options. Misunderstanding the delta type, which the market data refers to, would lead to a wrong pricing of vanilla op-

[5]We drop the % sign after the level of the delta, in accordance with market jargon. Therefore, a 25Δ call [resp. put] is an option whose delta is 0.25 [resp. -0.25].

tions. For example, the quotes in the given market sample refer to a spot delta and delta neutral straddle quotation is used for the at-the-money strike. In the next subsections, we explain which information these quotes contain.

1.3.1 At-The-Money Conventions

As in the case of the delta, *at-the-money* (ATM) volatilities quoted by brokers can have various interpretations depending on currency pairs. In particular, we find two definitions of the ATM strike, denoted by K_{ATM}, used by practitioners.

The first is the *at-the-money-forward* (ATMF) strike, which is set equal to the forward price $F_{0,T} = S_0 e^{(r^d - r^f)T}$, for a given maturity T. We then have

$$K_{ATM} \equiv K_{ATMF} = F_{0,T}. \tag{1.20}$$

This convention is only used for currency pairs including a Latin American emerging market currency.

A more natural way to define K_{ATM} is the strike corresponding to a *delta-neutral straddle* (DNS). In this case, the ATM volatility is the value from the smile curve where the strike is such that the delta of a long call equals, in absolute value, that of a long put, for every given maturity. This has the advantage that no delta hedging is needed when the straddle[6] is traded. Furthermore, it is the purest way to buy volatility at a fairly central level of the strike price. For these reasons, the delta-neutral straddle has become the standard choice of ATM strike for many financial industries. Hence, the delta-neutral straddle strike $K_{ATM} \equiv K_{DNS}$ is given by the solution of equation

$$\Delta(1, T, K_{DNS}, \sigma_{ATM}) + \Delta(-1, T, K_{DNS}, \sigma_{ATM}) = 0. \tag{1.21}$$

Note that, in order to solve Equation (1.21), it is irrelevant using spot or forward delta, since the discount factor, if it appears, is just a constant. Consequently, only two deltas can be sensibly considered, that is, the forward delta $\Delta_{F;d/f}$ and the premium-adjusted forward delta $\Delta_{F;\%}$. Actually, we can solve (1.21) for K analytically in both cases.

The former implies solving

$$\Delta_{F;d/f}(1, T, K_{DNS}, \sigma_{ATM}) + \Delta_{F;d/f}(-1, T, K_{DNS}, \sigma_{ATM}) = 0, \tag{1.22}$$

the solution of which we denote by $K_{DNS;d/f}$. By (1.14), $\Delta_{F;d/f} = \phi N(\phi d_1)$. Then, we require that $N(d_1) = N(-d_1)$. From Equation (1.2), we have

$$d_1 = \frac{\ln \frac{F_{0,T}}{K} + \frac{1}{2}\sigma^2 T}{\sigma\sqrt{T}}.$$

[6]In finance, a *straddle* is a non-directional options trading strategy, which involves a simultaneous purchase (long straddle) or sale (short straddle) of both a call and a put, written on the same underlying asset and with same strike and maturity.

Hence,

$$N\left(\frac{\ln F_{0,T}-\ln K+\frac{1}{2}\sigma^2 T}{\sigma\sqrt{T}}\right)=N\left(\frac{\ln K-\ln F_{0,T}-\frac{1}{2}\sigma^2 T}{\sigma\sqrt{T}}\right).$$

Since $N(\cdot)$ is monotonically increasing, we need K such that

$$\frac{\ln F_{0,T}-\ln K+\frac{1}{2}\sigma^2 T}{\sigma\sqrt{T}}=\frac{\ln K-\ln F_{0,T}-\frac{1}{2}\sigma^2 T}{\sigma\sqrt{T}}.$$

Then, we obtain

$$2\ln K=2\ln F_{0,T}+\sigma^2 T,$$

which has the exact solution

$$K_{DNS;d/f}=F_{0,T}e^{\frac{1}{2}\sigma^2 T}. \tag{1.23}$$

On the other hand, the latter case implies the resolution for K of equation

$$\Delta_{F;\%}(1,T,K_{DNS},\sigma_{ATM})+\Delta_{F;\%}(-1,T,K_{DNS},\sigma_{ATM})=0. \tag{1.24}$$

By (1.15), we have $\Delta_{F;\%}=\phi\frac{K}{F_{0,T}}N(\phi d_2)$ and then, after canceling out $\frac{K}{F_{0,T}}$, we require that $N(d_2)=N(-d_2)$. Now, the algebra is almost the same as above, from which it easily follows that the desired delta-neutral straddle strike is given by

$$K_{DNS;\%}=F_{0,T}e^{-\frac{1}{2}\sigma^2 T}. \tag{1.25}$$

Proposition 1.1. *The at-the-money-forward strike in (1.20) is actually delta-neutral with respect to the simple delta.*

Proof. Suppose we try to find the strike $K_{DNS;simple}$ which is solution of (1.21), with respect to the simple delta (1.18). We require that $N(d)=N(-d)$, which gives $\ln F_{0,T}-\ln K=\ln K-\ln F_{0,T}$. Then, we have

$$K_{DNS;simple}=K_{ATMF}=F_{0,T}.$$ □

Proposition 1.2 (Rule of thumb)**.** *For any currency pair, if the ATM strike is above [resp. below] the forward price, the market convention states that deltas for that currency pair must be quoted as domestic/foreign [resp. premium-adjusted].*

Independently of the choice of $\sigma(K)$, it has to be ensured that the volatility for the at-the-money strike is σ_{ATM}. Consequently, the construction procedure for $\sigma(K)$ has to guarantee that equation

$$\sigma(K_{ATM})=\sigma_{ATM}, \tag{1.26}$$

holds.

1.3.2 Market Strangle

Suppose that we neglect the effect of volatility skew and assume that volatility is basically symmetric, while not constant across strikes. The concept of *market strangle* implies that we can buy an out-of-the-money put and an out-of-the-money call with strikes placed at a similar distance away from the ATM strike, in moneyness terms. For the sake of clarity, we consider a 25Δ market strangle, though same conclusions can be drawn for any delta.

The idea underneath a market strangle is that, knowing nothing about the actual volatility smile, we can estimate the volatility needed to price a market strangle instrument consistently with the market, by adding a strangle (S) premium σ_{25-S} to the ATM volatility, that is

$$\sigma_{25-MS} = \sigma_{ATM} + \sigma_{25-S}. \tag{1.27}$$

Given this single volatility, we can extract a call strike K_{25C-MS} and a put strike K_{25P-MS} which, using σ_{25-MS} as volatility value, yields a delta of 0.25 and -0.25, respectively. The resulting strikes will then fulfill

$$\begin{aligned} \Delta\left(1, T, K_{25C-MS}, \sigma_{25-MS}\right) &= 0.25, \\ \Delta\left(-1, T, K_{25P-MS}, \sigma_{25-MS}\right) &= -0.25. \end{aligned} \tag{1.28}$$

Given strikes K_{25C-MS} and K_{25P-MS}, one can compute the price of an option position, which includes a long call with strike K_{25C-MS} and volatility σ_{25-MS} and a long put with strike K_{25P-MS} and the same volatility. The resulting price Q_{25-MS} is

$$Q_{25-MS} = Q\left(1, T, K_{25C-MS}, \sigma_{25-MS}\right) + Q\left(-1, T, K_{25P-MS}, \sigma_{25-MS}\right). \tag{1.29}$$

Figure 1.4 shows the market strangle as a spread to the ATM volatility. In the graph, the long horizontal line represents the ATM volatility level, while the other two shorter horizontal lines, at the same vertical level, show the market strangle volatility.

The important point to be stressed here is that, even if the volatility is not symmetric and neither the call with strike K_{25C-MS} nor the put with strike K_{25P-MS} can actually be priced with same volatility σ_{25-MS}, the corresponding option prices at these strikes must add up to Q_{25-MS}. In other words, a market consistent volatility function $\sigma(K)$ can have volatilities at strikes K_{25C-MS} and K_{25P-MS} different from σ_{25-MS}, but should guarantee that the aggregate price of the two options is equal to Q_{25-MS}. To summarize,

$$Q_{25-MS} = Q\left(1, T, K_{25C-MS}, \sigma(K_{25C-MS})\right) + Q\left(-1, T, K_{25P-MS}, \sigma(K_{25P-MS})\right). \tag{1.30}$$

1.3.3 Risk Reversal

At-the-money and market strangle volatilities provide two degrees of freedom, which allow a volatility smile to be described, but without a skew. In order to consider a market parameter that describes the skew, we introduced the *risk reversal* (RR), as defined by Malz [123].

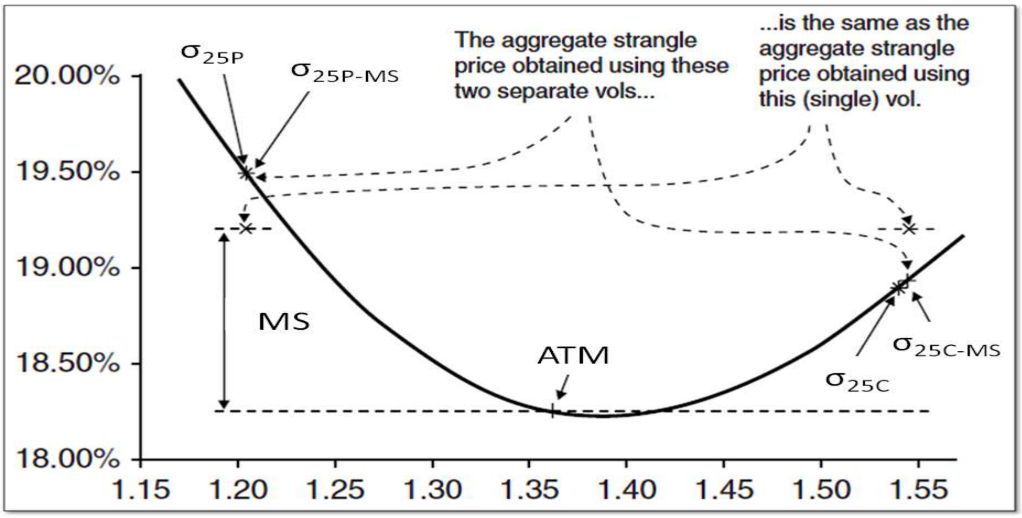

Figure 1.4: Schematic illustration of the market strangle

Assume that a volatility smile of the form $\sigma(K)$ exists and that each European option with different strike K can be priced using this volatility smile $\sigma(K)$, such that conditions (1.26) and (1.30) are fulfilled. Now, given $\sigma(K)$, our aim is to find 25Δ strike-volatility pairs for a call and a put, which yield a delta of 0.25 and -0.25, respectively. Therefore, we need to solve[7] for K

$$\begin{aligned} \Delta\left(1, T, K_{25C}, \sigma\left(K_{25C}\right)\right) &= 0.25, \\ \Delta\left(-1, T, K_{25P}, \sigma\left(K_{25P}\right)\right) &= -0.25. \end{aligned} \tag{1.31}$$

Then, we set

$$\begin{aligned} \sigma_{25C} &= \sigma\left(K_{25C}\right), \\ \sigma_{25P} &= \sigma\left(K_{25P}\right). \end{aligned}$$

Finally, the 25Δ risk reversal is quoted as the difference between these two implied volatilities:

$$\sigma_{25-RR} = \sigma_{25C} - \sigma_{25P}. \tag{1.32}$$

Positive values of the risk reversal mean that calls are favored with respect to puts, that is, the market attributes higher implied volatilities to calls than to puts, while for negative risk reversals the opposite is true.

Equations (1.26), (1.30) and (1.32) provide us with necessary conditions for $\sigma(K)$ to be consistent with the market, as parametrized by σ_{ATM}, σ_{25-MS} and σ_{25-RR}. If other deltas are quoted on the smile, most often 10Δ, then conditions (1.30) and (1.32) are applied in the same way as above, except using σ_{10-MS} and σ_{10-RR}.

[7]This can be achieved by using a standard root search algorithm.

1.3.4 A Simplified Formula

Frequently, a simplified formula is stated in the literature to allow an easy calculation of the 25Δ volatilities, given market quotes.

Let σ_{25C} be the call volatility corresponding to a delta of 0.25 and σ_{25P} the -0.25 delta put volatility. Let K_{25C} and K_{25P} denote the corresponding strikes. The simplified formula states that

$$\begin{aligned} \sigma_{25C} &= \sigma_{ATM} + \sigma_{25-S} + \tfrac{1}{2}\sigma_{25-RR}, \\ \sigma_{25P} &= \sigma_{ATM} + \sigma_{25-S} - \tfrac{1}{2}\sigma_{25-RR}. \end{aligned} \tag{1.33}$$

Including the at-the-money volatility would result in a smile with three anchor points, which can then be interpolated. In this case, no calibration procedure is needed. Note that

$$\sigma_{25C} - \sigma_{25P} = \sigma_{25-RR},$$

such that the 25Δ volatility difference automatically matches the quoted risk reversal.

The simplified formula can be reformulated to calculate σ_{25-S}, given σ_{25C}, σ_{25P} and σ_{ATM} quotes. This yields

$$\sigma_{25-S} = \frac{\sigma_{25C} + \sigma_{25P}}{2} - \sigma_{ATM}, \tag{1.34}$$

which presents the strangle as a convexity parameter. However, the problem arises in the matching of the market strangle, as given in Equation (1.29). Interpolating the smile from the three anchor points given by the simplified formula and calculating the market strangle with the corresponding volatilities at K_{25P-MS} and K_{25C-MS} does not necessary lead to the matching of Q_{25-MS}. If $\sigma_{25-RR} = 0$, then by (1.32) we have $\sigma_{25C} = \sigma_{25P}$, hence $\sigma_{25-S} = \sigma_{25C} - \sigma_{ATM} = \sigma_{25P} - \sigma_{ATM}$. As a consequence, $\sigma_{25-MS} = \sigma_{25-S}$ only in the case in which $\sigma_{25-RR} = 0$, otherwise the equality does not hold true. The discrepancy is most relevant for currency pairs where the risk reversal is large in absolute value, typically greater than 1%.

Furthermore, the σ_{25-S} value is in general not directly observable in FX markets. Instead, quants usually communicate the σ_{25-MS}. The difference between σ_{25-S} and σ_{25-MS} can be at times confusing. Often for convenience one sets $\sigma_{25-MS} = \sigma_{25-S}$, as this greatly simplifies the procedure to build up a smile curve. However, it leads to errors when applied to a steeply skewed market.

Remark 1.2. *Using the simplified smile construction procedure yields a market strangle consistent smile setup only in case of a zero risk reversal. The other market matching requirements are met by default. In any other case, the strangle price might not be matched, which leads to a non market consistent setup of the volatility smile.*

Nonetheless, the simplified formula can still be useful, even for large risk reversals, if σ_{25-S} is replaced by another parameter, known in literature as the *smile strangle* (SS). Assume that volatilities $\sigma(K_{25C})$ and $\sigma(K_{25C})$ are given by the calibrated smile function $\sigma(K)$. Then, we can

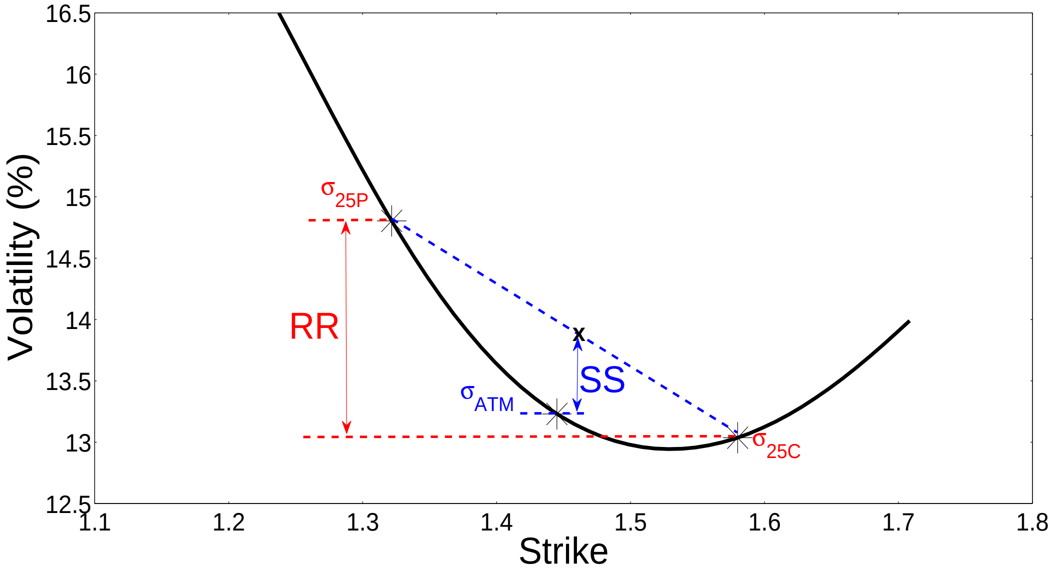

Figure 1.5: Schematic illustration of the risk reversal and the smile strangle

compute the smile strangle via

$$\sigma_{25-SS} = \frac{\sigma\left(K_{25C}\right) + \sigma\left(K_{25P}\right)}{2} - \sigma(K_{ATM}). \tag{1.35}$$

Thus, the smile strangle measures the convexity of the calibrated smile function.

Figure 1.5 presents a smile volatility $\sigma(K)$, obtained from the fit of 1Y EURUSD data seen in Table 1.1. The 25Δ put and call strikes are shown with respect to their corresponding volatilities on the smile, same as for the at-the-money strike. The horizontal dotted lines indicate volatility levels at these three strikes. The graph depict the risk reversal as the difference between σ_{25C} and σ_{25P}. On the other hand, the oblique dotted segment, which connects σ_{25C} and σ_{25P} points, is bisected, so that the midpoint has a volatility equal to their average. Hence, the vertical distance of that midpoint from the ATM level represents the smile strangle.

Note that the smile strangle construction is similar to Equation (1.34), but in (1.35) we are using out-of-the-money volatilities obtained from the calibrated smile and not from the simplified formula. Given σ_{25-SS}, the simplified formula (1.33) can still be used, if the quoted strangle volatility σ_{25-S} is replaced by the smile strangle volatility σ_{25-SS}. Clearly, σ_{25-SS} is not known a priori, but is obtained with the calibration of the smile function.

Figure 1.6 shows that, when the risk reversal is large, as is typically the case for the currency pair USDJPY, strikes can be markedly different. This discrepancy between smile strangle and market strangle strikes increases for large absolute value of the risk reversal.

To close this section, let us clarify another enigmatic concept of FX markets often used by practitioners, the so-called *vega-weighted strangle* quote. This is an approximation for the value

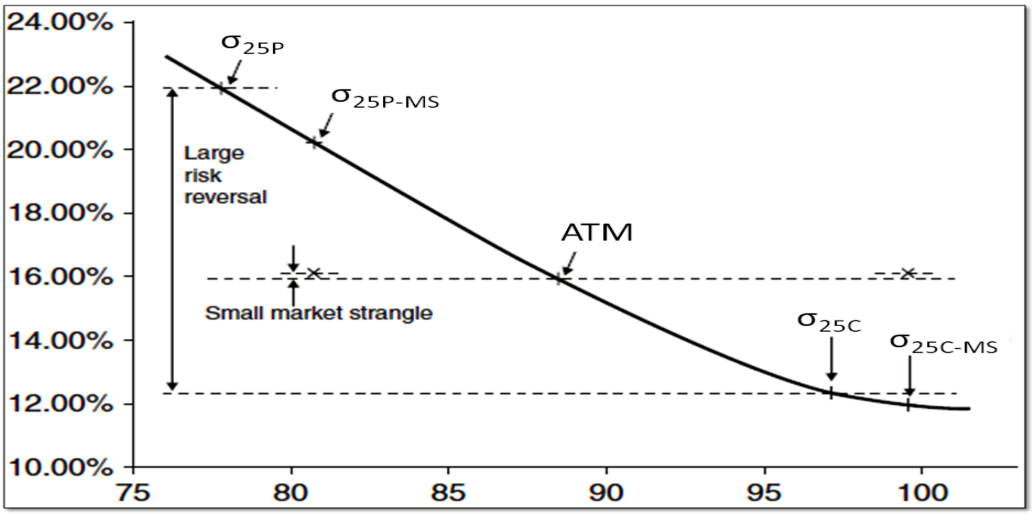

Figure 1.6: Highly skewed smile for the currency pair USDJPY

of σ_{25-MS}. To show this, we start from equality

$$Q\left(1,T,K_{25C},\sigma_{25-MS}\right)+Q\left(-1,T,K_{25P},\sigma_{25-MS}\right)= \\ Q\left(1,T,K_{25C},\sigma\left(K_{25C}\right)\right)+Q\left(-1,T,K_{25P},\sigma\left(K_{25P}\right)\right),$$

and we develop both sides in a first order Taylor expansion in σ around σ_{ATM}. Hence, we have

$$\begin{aligned} &Q\left(1,T,K_{25C},\sigma_{ATM}\right)+Q\left(-1,T,K_{25P},\sigma_{ATM}\right)+ \\ &\quad+\mathcal{V}\left(1,T,K_{25C},\sigma_{ATM}\right)\left(\sigma_{25MS}-\sigma_{ATM}\right)+ \\ &\qquad+\mathcal{V}\left(-1,T,K_{25P},\sigma_{ATM}\right)\left(\sigma_{25-MS}-\sigma_{ATM}\right)\approx \\ &Q\left(1,T,K_{25C},\sigma_{ATM}\right)+Q\left(-1,T,K_{25P},\sigma_{ATM}\right)+ \\ &\quad+\mathcal{V}\left(1,T,K_{25C},\sigma_{ATM}\right)\left(\sigma\left(K_{25C}\right)-\sigma_{ATM}\right)+ \\ &\qquad+\mathcal{V}\left(-1,T,K_{25P},\sigma_{ATM}\right)\left(\sigma\left(K_{25P}\right)-\sigma_{ATM}\right), \end{aligned}$$

where $\mathcal{V}\left(\phi,T,K,\sigma(K)\right)$ represents the Vega of the option, namely the sensitivity of the option price with respect to a change of the implied volatility. Deleting repeating terms on the left and right-hand side, we obtain

$$\sigma_{25-MS}\left[\mathcal{V}\left(1,T,K_{25C},\sigma_{ATM}\right)+\mathcal{V}\left(-1,T,K_{25P},\sigma_{ATM}\right)\right]\approx \\ \mathcal{V}\left(1,T,K_{25C},\sigma_{ATM}\right)\sigma\left(K_{25C}\right)+\mathcal{V}\left(-1,T,K_{25P},\sigma_{ATM}\right)\sigma\left(K_{25P}\right).$$

Solving for σ_{25-MS} yields

$$\sigma_{25-MS}\approx\frac{\mathcal{V}\left(1,T,K_{25C},\sigma_{ATM}\right)\sigma\left(K_{25C}\right)+\mathcal{V}\left(-1,T,K_{25P},\sigma_{ATM}\right)\sigma\left(K_{25P}\right)}{\mathcal{V}\left(1,T,K_{25C},\sigma_{ATM}\right)+\mathcal{V}\left(-1,T,K_{25P},\sigma_{ATM}\right)},$$

that is

$$\sigma_{25-MS} \approx \frac{\mathcal{V}(1,T,K_{25C},\sigma_{ATM})\,\sigma_{25C} + \mathcal{V}(-1,T,K_{25P},\sigma_{ATM})\,\sigma_{25P}}{\mathcal{V}(1,T,K_{25C},\sigma_{ATM}) + \mathcal{V}(-1,T,K_{25P},\sigma_{ATM})},$$

which corresponds to the average of the call and put implied volatilities, weighted by Vega. According to Castagna and Mercurio [60], practitioners also use the term *Vega-weighted butterfly* for a structure where a strangle is bought and an amount of ATM straddle is sold, such that the overall Vega of the strategy is zero.

1.4 The Vanna-Volga Method

While the Black-Scholes [42] model is an important market benchmark, it is widely recognized that it presents several deficiencies when it comes to describe modern market phenomena. In particular, the assumptions of constant interest rates and a single volatility, sufficient to price options with different maturities and strikes, are inadequate to describe the more realistic and highly volatile world of FX derivatives. Nowadays, the Black-Scholes [42] theoretical value is only used as a reference quotation and provides a starting point for the development of advanced models.

More realistic models should assume that the foreign/domestic interest rates and the FX volatility follow stochastic processes. The choice of the stochastic process depends on empirical observations, among other factors. For example, for long-dated options the effect of the interest rate volatility can become as significant as that of the FX volatility. On the other hand, for short-dated options, typically less than 1 year, assuming constant interest rates does not normally lead to significant mispricing. Although stochastic volatility models can explain much more complex markets, their main drawback is that they are computationally demanding and, in most cases, they require a relatively great quantity of market data, which is not always readily available.

This has led FX brokers and market makers to introduce alternative empirical procedures, that give fast results and are simpler to implement. Certainly, one useful approach is the *Vanna-Volga* (VV) method, which allows to infer an implied volatility smile from three available market quotes, for a given maturity. The terms *Vanna* and *Volga* are commonly used by practitioners to denote partial derivatives of an option Vega, with respect to spot and volatility, respectively. In technical terms, we have

$$Vanna = \frac{\partial Vega}{\partial Spot}, \qquad Volga = \frac{\partial Vega}{\partial Vol}. \tag{1.36}$$

This technique is based on the construction of a locally replicating portfolio, which is Vega-neutral in the Black-Scholes world. In this way, hedging costs associated to this portfolio are added to the corresponding Black-Scholes prices, in order to derive smile consistent values. Besides being intuitive and easy to implement, this procedure has a clear financial interpretation, which further supports its use in practice.

In the financial literature, the Vanna-Volga method seems to appear first in Lipton and McGhee [118], who apply the adjustment of the Black-Scholes price with an hedging portfolio to double-no-touch options, and successively in Wystup [158], who describes its application to the valuation of one-touch options. However, their analyses are rather informal and mostly based on numerical examples. The first systematic formulation of the Vanna-Volga method was proposed by Castagna and Mercurio [60], who derive some important results concerning the tractability of the method and its robustness. In particular, they show that it can be used as a smile interpolation tool to obtain a value of volatility for a given strike, while reproducing exactly market quoted volatilities. Furthermore, Bossens et al. [43] and Fisher [87] introduce a number of corrections to handle the pricing inconsistencies of the first-generation exotics. Finally, a more rigorous and theoretical review is given by Shkolnikov [148], where the method is extended to include interest rate risk, among other directions.

Now, consider the three basic options quoted in the FX market for a given maturity T, that is, the 25Δ put, the ATM and the 25Δ call. Denote the corresponding strikes by K_i, $i = 1, 2, 3$, such that $K_1 < K_2 < K_3$, and set $\mathbb{K} = (K_1, K_2, K_3)$[8]. Market implied volatilities associated to K_i are denoted by σ_i, $i = 1, 2, 3$. Moreover, market option prices $Q^{MKT}(T, K_1)$, $Q^{MKT}(T, K_2)$ and $Q^{MKT}(T, K_3)$ are assumed to satisfy standard no arbitrage conditions.

The VV method is used to define an implied volatility smile, which is consistent with the basic volatilities σ_i. The rationale behind it stems from a replication argument in a flat-smile world, where the level of the constant implied volatility varies stochastically over time.

It is well known that in the Black-Scholes [42] model the payoff of a European call, with maturity T and strike K, can be replicated by a dynamic delta-hedging strategy, whose initial value matches, at every time $0 \leq t < T$, the option price $Q_t^{BS}(T, K)$, given by

$$Q_t^{BS}(T, K) = S_t e^{-r^f \tau} N\left(\frac{\ln \frac{S_t}{K} + \left(r^d - r^f + \frac{1}{2}\sigma^2\right)\tau}{\sigma\sqrt{\tau}}\right) + \\ - K e^{-r^d \tau} N\left(\frac{\ln \frac{S_t}{K} + \left(r^d - r^f - \frac{1}{2}\sigma^2\right)\tau}{\sigma\sqrt{\tau}}\right), \quad (1.37)$$

where $\tau = T - t$, S_t is the FX rate at time t, r^d and r^f are the domestic and foreign, respectively, risk-free interest rates and σ is the constant implied volatility. In real financial markets, however, volatility is stochastic and traders hedge the associated risk by constructing portfolios that are Vega-neutral in a Black-Scholes world.

Denote by Π_t the value of a risk-neutral replicating portfolio, composed of a long position in a European call option with maturity T and strike K, whose price is $Q_t^{BS}(T, K)$ as in (1.37), two short positions in Δ_t units of the underlying asset S_t and x_i units of three European calls with strikes K_i and prices $Q_t^{BS}(T, K_i)$, $i = 1, 2, 3$. In a small interval of time dt, the value of the portfolio Π_t is subject to a variation, given by

$$d\Pi_t = dQ_t^{BS}(T, K) - \Delta_t dS_t - \sum_{i=1}^{3} x_i dQ_t^{BS}(T, K_i). \quad (1.38)$$

[8] K_1, K_2 and K_3 replace K_{25P}, K_{ATM} and K_{25C}, respectively.

Under diffusion dynamics for both S_t and $\sigma = \sigma_t$, by Itô's lemma we have

$$
\begin{aligned}
d\Pi_t = & \left[\frac{\partial Q_t^{BS}(T,K)}{\partial t} - \sum_{i=1}^{3} x_i \frac{\partial Q_t^{BS}(T,K_i)}{\partial t}\right] dt \\
& + \left[\frac{\partial Q_t^{BS}(T,K)}{\partial S_t} - \Delta_t - \sum_{i=1}^{3} x_i \frac{\partial Q_t^{BS}(T,K_i)}{\partial S_t}\right] dS_t \\
& + \left[\frac{\partial Q_t^{BS}(T,K)}{\partial \sigma_t} - \sum_{i=1}^{3} x_i \frac{\partial Q_t^{BS}(T,K_i)}{\partial \sigma_t}\right] d\sigma_t \\
& + \left[\frac{\partial^2 Q_t^{BS}(T,K)}{\partial S_t \partial \sigma_t} - \sum_{i=1}^{3} x_i \frac{\partial^2 Q_t^{BS}(T,K_i)}{\partial S_t \partial \sigma_t}\right] dS_t d\sigma_t \\
& + \frac{1}{2}\left[\frac{\partial^2 Q_t^{BS}(T,K)}{\partial S_t^2} - \sum_{i=1}^{3} x_i \frac{\partial^2 Q_t^{BS}(T,K_i)}{\partial S_t^2}\right] (dS_t)^2 \\
& + \frac{1}{2}\left[\frac{\partial^2 Q_t^{BS}(T,K)}{\partial \sigma_t^2} - \sum_{i=1}^{3} x_i \frac{\partial^2 Q_t^{BS}(T,K_i)}{\partial \sigma_t^2}\right] (d\sigma_t)^2 . \qquad (1.39)
\end{aligned}
$$

By virtue of stochastic calculus rules, coefficients of the terms $(dt)^2$, $dt d\sigma_t$ and $dt dS_t$ vanish. On the other hand, the following equation holds:

$$d\Pi_t = r^d \Pi_t dt, \qquad (1.40)$$

based on the no arbitrage assumption. In order to obtain a locally hedging portfolio, we choose Δ_t and x_i so as to zero out the coefficients of dS_t, $d\sigma_t$, $(d\sigma_t)^2$ and $dS_t d\sigma_t$[9]. By construction of the hedging portfolio Π_t, we get rid of the risk associated with the fluctuations of the spot price S_t and of the volatility σ_t. Hence, our portfolio is now locally risk-free at time t, that is, no stochastic terms are involved in its differential.

Then, the amount Δ_t of the underlying asset is given by

$$\Delta_t = \frac{\partial Q_t^{BS}(T,K)}{\partial S_t} - \sum_{i=1}^{3} x_i \frac{\partial Q_t^{BS}(T,K_i)}{\partial S_t}. \qquad (1.41)$$

Now, from Equation (1.39), we want to find time-t weights $x_i(K)$, $i = 1, 2, 3$, for which the resulting portfolio of European call options with maturity T and strikes K_1, K_2 and K_3, respectively, hedges the price variations of the call with maturity T and strike K, up to the second order in the underlying and the volatility. Assuming a delta-hedged position and given that, in the Black-Scholes world, those portfolios of plain vanilla options with the same maturity that are Vega neutral are also Gamma neutral, then, the weights $x_1(K)$, $x_2(K)$, and $x_3(K)$ can be found by imposing that the replicating portfolio has the same Vega, Vanna and Volga as the

[9] The coefficient of $(dS_t)^2$ will be automatically zeroed out, due to the relation linking an option Gamma and Vega in the Black-Scholes world.

call with strike K, that is

$$\begin{aligned}\frac{\partial Q_t^{BS}(T,K)}{\partial\sigma_t} &= \sum_{i=1}^{3} x_i(K)\,\frac{\partial Q_t^{BS}(T,K_i)}{\partial\sigma_t}\\ \frac{\partial^2 Q_t^{BS}(T,K)}{\partial S_t\partial\sigma_t} &= \sum_{i=1}^{3} x_i(K)\,\frac{\partial^2 Q_t^{BS}(T,K_i)}{\partial S_t\partial\sigma_t}\\ \frac{\partial^2 Q_t^{BS}(T,K)}{\partial\sigma_t^2} &= \sum_{i=1}^{3} x_i(K)\,\frac{\partial^2 Q_t^{BS}(T,K_i)}{\partial\sigma_t^2}\end{aligned}\quad. \tag{1.42}$$

Denoting by $\mathcal{V}_t(K)$ the Vega at time t of an option, with maturity T and strike K, we can write (1.42) in a more compact matricial form as

$$\mathbf{Vx} = \mathbf{A}, \tag{1.43}$$

where

$$\mathbf{x} = \begin{bmatrix} x_1(K)\\ x_2(K)\\ x_3(K)\end{bmatrix}, \tag{1.44}$$

$$\mathbf{A} = \begin{bmatrix} \mathcal{V}_t(K)\\ \frac{\partial\mathcal{V}_t(K)}{\partial S_t}\\ \frac{\partial\mathcal{V}_t(K)}{\partial\sigma_t}\end{bmatrix} = \begin{bmatrix} \frac{\partial Q_t^{BS}(T,K)}{\partial\sigma_t}\\ \frac{\partial^2 Q_t^{BS}(T,K)}{\partial S_t\partial\sigma_t}\\ \frac{\partial^2 Q_t^{BS}(T,K)}{\partial\sigma_t^2}\end{bmatrix}, \tag{1.45}$$

and

$$\mathbf{V} = \begin{bmatrix} \mathcal{V}_t(K_1) & \mathcal{V}_t(K_2) & \mathcal{V}_t(K_3)\\ \frac{\partial\mathcal{V}_t(K_1)}{\partial S_t} & \frac{\partial\mathcal{V}_t(K_2)}{\partial S_t} & \frac{\partial\mathcal{V}_t(K_3)}{\partial S_t}\\ \frac{\partial\mathcal{V}_t(K_1)}{\partial\sigma_t} & \frac{\partial\mathcal{V}_t(K_2)}{\partial\sigma_t} & \frac{\partial\mathcal{V}_t(K_3)}{\partial\sigma_t}\end{bmatrix} = \begin{bmatrix} \frac{\partial Q_t^{BS}(T,K_1)}{\partial\sigma_t} & \frac{\partial Q_t^{BS}(T,K_2)}{\partial\sigma_t} & \frac{\partial Q_t^{BS}(T,K_3)}{\partial\sigma_t}\\ \frac{\partial^2 Q_t^{BS}(T,K_1)}{\partial S_t\partial\sigma_t} & \frac{\partial^2 Q_t^{BS}(T,K_2)}{\partial S_t\partial\sigma_t} & \frac{\partial^2 Q_t^{BS}(T,K_3)}{\partial S_t\partial\sigma_t}\\ \frac{\partial^2 Q_t^{BS}(T,K_1)}{\partial\sigma_t^2} & \frac{\partial^2 Q_t^{BS}(T,K_2)}{\partial\sigma_t^2} & \frac{\partial^2 Q_t^{BS}(T,K_3)}{\partial\sigma_t^2}\end{bmatrix}. \tag{1.46}$$

Proposition 1.3. *The system (1.43) admits a unique solution, which is given by*

$$\begin{aligned} x_1(K) &= \frac{\mathcal{V}_t(K)}{\mathcal{V}_t(K_1)}\frac{\ln\frac{K_2}{K}\ln\frac{K_3}{K}}{\ln\frac{K_2}{K_1}\ln\frac{K_3}{K_1}}\\ x_2(K) &= \frac{\mathcal{V}_t(K)}{\mathcal{V}_t(K_2)}\frac{\ln\frac{K}{K_1}\ln\frac{K_3}{K}}{\ln\frac{K_2}{K_1}\ln\frac{K_3}{K_2}}\\ x_3(K) &= \frac{\mathcal{V}_t(K)}{\mathcal{V}_t(K_3)}\frac{\ln\frac{K}{K_1}\ln\frac{K}{K_2}}{\ln\frac{K_3}{K_1}\ln\frac{K_3}{K_2}}\end{aligned}\quad. \tag{1.47}$$

In particular, if $K = K_j$, then $x_i(K) = 1$ for $i = j$ and zero otherwise.

Proof. We have

$$
\begin{aligned}
\mathcal{V}_t(K) &= \frac{\partial Q_t^{BS}(T,K)}{\partial \sigma_t} \\
&= S_t e^{-r^f\tau} N'(d_1) \frac{\partial d_1}{\partial \sigma_t} - K e^{-r^d\tau} N'(d_2) \frac{\partial d_2}{\partial \sigma_t} \\
&= S_t e^{-r^f\tau} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}d_1^2} \left\{ \frac{\sigma_t^2 \tau\sqrt{\tau} - \left[\ln\frac{S_t}{K} + \left(r^d - r^f + \frac{1}{2}\sigma_t^2\right)\tau\right]\sqrt{\tau}}{\sigma_t^2\tau} \right\} + \\
&\quad - K e^{-r^d\tau} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}d_1^2} \frac{S_t}{K} e^{\left(r^d - r^f\right)\tau} \left\{ \frac{-\sigma_t^2 \tau\sqrt{\tau} - \left[\ln\frac{S_t}{K} + \left(r^d - r^f - \frac{1}{2}\sigma_t^2\right)\tau\right]\sqrt{\tau}}{\sigma_t^2\tau} \right\} \\
&= S_t e^{-r^f\tau} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}d_1^2} \left(\frac{\sigma_t^2\tau\sqrt{\tau}}{\sigma_t^2\tau} \right) \\
&= S_t e^{-r^f\tau} \sqrt{\tau} N'(d_1).
\end{aligned}
$$

Computing the second order derivatives, we obtain

$$
\begin{aligned}
Vanna &= \frac{\partial \mathcal{V}_t(K)}{\partial S_t} \\
&= e^{-r^f\tau}\sqrt{\tau}\left[N'(d_1) - S_t \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}d_1^2} d_1 \frac{\partial d_1}{\partial S_t}\right] \\
&= e^{-r^f\tau}\sqrt{\tau}\left[N'(d_1) - S_t \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}d_1^2} d_1 \frac{1}{S_t\sigma_t\sqrt{\tau}}\right] \\
&= -e^{-r^f\tau}\sqrt{\tau} N'(d_1) \frac{d_2}{\sigma_t\sqrt{\tau}} \\
&= -\frac{\mathcal{V}_t(K)}{S_t} \frac{d_2}{\sigma_t\sqrt{\tau}},
\end{aligned}
$$

and

$$
\begin{aligned}
Volga &= \frac{\partial \mathcal{V}_t(K)}{\partial \sigma_t} \\
&= -S_t e^{-r^f\tau}\sqrt{\tau}\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}d_1^2} d_1 \frac{\partial d_1}{\partial \sigma_t} \\
&= -\mathcal{V}_t(K)\, d_1 \left\{ \frac{\sigma_t^2\tau\sqrt{\tau} - \left[\ln\frac{S_t}{K} + \left(r^d - r^f + \frac{1}{2}\sigma_t^2\right)\tau\right]\sqrt{\tau}}{\sigma_t^2\tau} \right\} \\
&= \frac{\mathcal{V}_t(K)}{\sigma_t} d_1 \left[\frac{\ln\frac{S_t}{K} + \left(r^d - r^f + \frac{1}{2}\sigma_t^2\right)\tau}{\sigma_t\sqrt{\tau}} - \sigma_t\sqrt{\tau} \right] \\
&= \frac{\mathcal{V}_t(K)}{\sigma_t} d_1 \left(d_1 - \sigma_t\sqrt{\tau}\right) \\
&= \frac{\mathcal{V}_t(K)}{\sigma_t} d_1 d_2.
\end{aligned}
$$

From the system (1.43), straightforward algebra leads to

$$
\begin{aligned}
\det\left(\mathbf{V}\right) = {} & \mathcal{V}_t\left(K_1\right)\frac{\partial\mathcal{V}_t\left(K_2\right)}{\partial S_t}\frac{\partial\mathcal{V}_t\left(K_3\right)}{\partial\sigma_t}+\mathcal{V}_t\left(K_2\right)\frac{\partial\mathcal{V}_t\left(K_3\right)}{\partial S_t}\frac{\partial\mathcal{V}_t\left(K_1\right)}{\partial\sigma_t}+\\
& +\mathcal{V}_t\left(K_3\right)\frac{\partial\mathcal{V}_t\left(K_1\right)}{\partial S_t}\frac{\partial\mathcal{V}_t\left(K_2\right)}{\partial\sigma_t}-\left[\mathcal{V}_t\left(K_3\right)\frac{\partial\mathcal{V}_t\left(K_2\right)}{\partial S_t}\frac{\partial\mathcal{V}_t\left(K_1\right)}{\partial\sigma_t}+\right.\\
& \left.+\mathcal{V}_t\left(K_1\right)\frac{\partial\mathcal{V}_t\left(K_3\right)}{\partial S_t}\frac{\partial\mathcal{V}_t\left(K_2\right)}{\partial\sigma_t}+\mathcal{V}_t\left(K_2\right)\frac{\partial\mathcal{V}_t\left(K_1\right)}{\partial S_t}\frac{\partial\mathcal{V}_t\left(K_3\right)}{\partial\sigma_t}\right]\\
= {} & -\mathcal{V}_t\left(K_1\right)\frac{\mathcal{V}_t\left(K_2\right)}{S_t\sigma_t\sqrt{\tau}}d_2\left(K_2\right)\frac{\mathcal{V}_t\left(K_3\right)}{\sigma_t}d_1\left(K_3\right)d_2\left(K_3\right)+\\
& -\mathcal{V}_t\left(K_2\right)\frac{\mathcal{V}_t\left(K_3\right)}{S_t\sigma_t\sqrt{\tau}}d_2\left(K_3\right)\frac{\mathcal{V}_t\left(K_1\right)}{\sigma_t}d_1\left(K_1\right)d_2\left(K_1\right)+\\
& -\mathcal{V}_t\left(K_3\right)\frac{\mathcal{V}_t\left(K_1\right)}{S_t\sigma_t\sqrt{\tau}}d_2\left(K_1\right)\frac{\mathcal{V}_t\left(K_2\right)}{\sigma_t}d_1\left(K_2\right)d_2\left(K_2\right)+\\
& +\mathcal{V}_t\left(K_3\right)\frac{\mathcal{V}_t\left(K_2\right)}{S_t\sigma_t\sqrt{\tau}}d_2\left(K_2\right)\frac{\mathcal{V}_t\left(K_1\right)}{\sigma_t}d_1\left(K_1\right)d_2\left(K_1\right)+\\
& +\mathcal{V}_t\left(K_1\right)\frac{\mathcal{V}_t\left(K_3\right)}{S_t\sigma_t\sqrt{\tau}}d_2\left(K_3\right)\frac{\mathcal{V}_t\left(K_2\right)}{\sigma_t}d_1\left(K_2\right)d_2\left(K_2\right)+\\
& +\mathcal{V}_t\left(K_2\right)\frac{\mathcal{V}_t\left(K_1\right)}{S_t\sigma_t\sqrt{\tau}}d_2\left(K_1\right)\frac{\mathcal{V}_t\left(K_3\right)}{\sigma_t}d_1\left(K_3\right)d_2\left(K_3\right)\\
= {} & \frac{\mathcal{V}_t\left(K_1\right)\mathcal{V}_t\left(K_2\right)\mathcal{V}_t\left(K_3\right)}{S_t\sigma_t^2\sqrt{\tau}}\left[d_2\left(K_2\right)d_1\left(K_1\right)d_2\left(K_1\right)+d_2\left(K_3\right)d_1\left(K_2\right)d_2\left(K_2\right)+\right.\\
& \qquad +d_2\left(K_1\right)d_1\left(K_3\right)d_2\left(K_3\right)-d_2\left(K_2\right)d_1\left(K_3\right)d_2\left(K_3\right)+\\
& \qquad \left.-d_2\left(K_3\right)d_1\left(K_1\right)d_2\left(K_1\right)-d_2\left(K_1\right)d_1\left(K_2\right)d_2\left(K_2\right)\right]\\
= {} & \frac{\mathcal{V}_t\left(K_1\right)\mathcal{V}_t\left(K_2\right)\mathcal{V}_t\left(K_3\right)}{S_t\sigma_t^5\tau^2}\ln\frac{K_2}{K_1}\ln\frac{K_3}{K_1}\ln\frac{K_3}{K_2},
\end{aligned}
$$

which is strictly positive since $K_1 < K_2 < K_3$. Hence, the system (1.43) admits a unique solution and (1.47) follows from Cramer's rule. □

Therefore, when volatility is stochastic and options are valued with the Black-Scholes [42] formula, we can still have a locally perfect hedge, provided that we hold suitable amounts of three more options, in order to rule out the model risk. Note that the hedging strategy is irrespective of the true asset and volatility dynamics, under the assumption of no jumps.

Remark 1.3. *The validity of the previous replication argument may be questioned because no stochastic volatility model can produce implied volatilities that are flat and stochastic at the same time. The simultaneous presence of these features, though inconsistent from a theoretical point of view, can be justified on empirical grounds. In fact, the practical advantages of the Black-Scholes [42] paradigm are so clear that many FX option traders implement hedging strategies based on a Black-Scholes flat smile model, with the ATM volatility being continuously updated to the actual market level.*

Delta	2Y	
25Δ Put	1.2714	11.73%
ATM	1.4112	10.93%
25Δ Call	1.5663	10.93%

Table 1.2: Strikes and volatilities corresponding to the three main quotes, as in December 25, 2008

1.4.1 The Vanna-Volga Option Price

Now, we can proceed to the definition of the option price resulting from the Vanna-Volga method, which is consistent with market prices of the basic options.

The above replication argument shows that a portfolio, with x_i units of an option with strike K_i and Δ_t units of the underlying asset, gives a local perfect hedge in a Black-Scholes world. The hedging strategy, however, has to be implemented at prevailing market prices. This generates an extra cost with respect to the Black-Scholes portfolio value. Thus, such a cost has to be added to the Black-Scholes price (1.37), in order to satisfy no arbitrage conditions.

Hence, a smile consistent price for a call option, with strike K and maturity T, is obtained by adding to the Black-Scholes price the cost of implementing the hedging strategy at prevailing market prices. In order to lighten the notation, we consider $t = 0$, so that we can remove the subscript for time t. We then have

$$Q^{VV}(T,K) = Q^{BS}(T,K) + \sum_{i=1}^{3} x_i \left[Q^{MKT}(T,K_i) - Q^{BS}(T,K_i) \right]. \tag{1.48}$$

The quantity $Q^{VV}(T,K)$ in (1.48) is defined as the *Vanna-Volga option price*, implicitly assuming that the replication error is also negligible for longer maturities. Moreover, the term

$$\widetilde{Q} = \sum_{i=1}^{3} x_i \left[Q^{MKT}(T,K_i) - Q^{BS}(T,K_i) \right], \tag{1.49}$$

is known as the *Vanna-Volga correction*, or *adjustment*, or even *overhedge*, since it represents the additional cost of the hedging portfolio, induced by market implied volatilities with respect to the constant volatility[10].

It is worth noting that calls and puts can be considered interchangeably, due to the put-call parity. Using puts instead of calls only changes the value of Δ_t, but does not affect x_i values.

When $K = K_j$, we clearly have that $Q^{VV}(T,K_j) = Q^{MKT}(T,K_j)$, since $x_i(K) = 1$ for $i = j$ and zero otherwise. Therefore, Equation (1.48) defines a rule for either interpolating or extrapolating prices from the three option quotes $Q^{MKT}(T,K_1)$, $Q^{MKT}(T,K_2)$ and $Q^{MKT}(T,K_3)$.

A market implied volatility curve can then be constructed by inverting (1.48) through the Black-Scholes [42] formula, for each K considered. An example of such a curve is provided in Figure 1.7, where we plot implied volatilities against strikes, and the three market principal

[10]The Vanna-Volga price $Q^{VV}(T,K)$ depends on the volatility parameter σ. In practice, the typical choice is to set $\sigma = \sigma_{ATM}$.

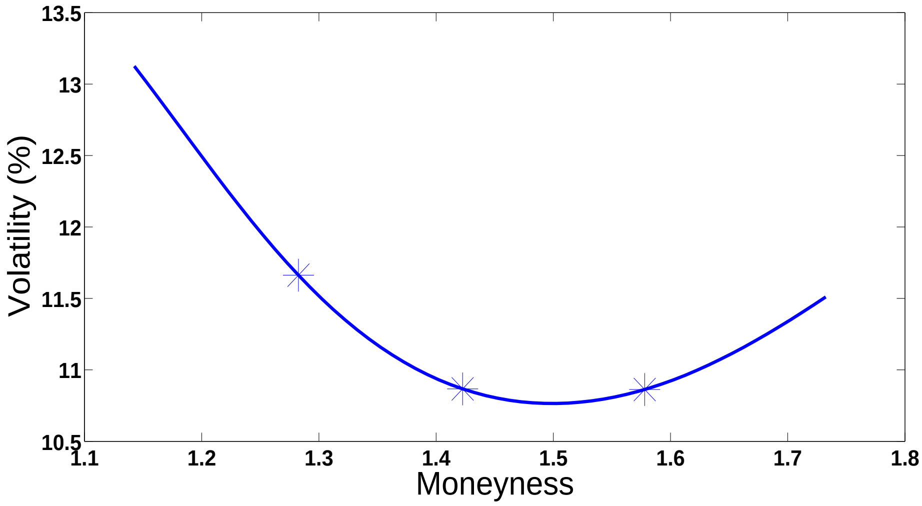

Figure 1.7: EURUSD implied volatilities plotted against strikes as in September 10^{th}, 2008

quotes are highlighted. Table 1.2 shows the data used.

For any given maturity T, the option price $Q^{VV}(T,K)$, as a function of the strike K, satisfies the following no arbitrage conditions:

1. $Q^{VV}(T,K) \in \mathcal{C}^2(0,+\infty)$;
2. $\lim_{K\to 0^+} Q^{VV}(T,K) = S_0 e^{-r^f T}, \quad \lim_{K\to +\infty} Q^{VV}(T,K) = 0$;
3. $\lim_{K\to 0^+} \frac{\partial Q^{VV}(T,K)}{\partial K} = -e^{-r^d T}, \quad \lim_{K\to +\infty} K \frac{\partial Q^{VV}(T,K)}{\partial K} = 0$.

The second and third properties, which are trivially satisfied by $Q^{BS}(T,K)$, follow from the fact that, for each i, both $x_i(K)$ and $\frac{\partial x_i(K)}{\partial K}$ go to zero for $K \to 0^+$ or $K \to +\infty$.

Furthermore, in order to avoid arbitrage opportunities, the option price $Q^{VV}(T,K)$ should also be a convex function of the strike K, that is

$$\frac{\partial^2 Q^{VV}(T,K)}{\partial K^2} > 0, \qquad \forall K > 0.$$

This property, which is not true in general, holds for typical market parameters, so that Equation (1.48) leads to prices that are arbitrage free in practice.

Finally, if we consider all available maturities we can obtain the entire implied volatility surface, as we can see in Figure 1.8.

Contrary to other interpolation schemes proposed in the financial literature, the Vanna-Volga pricing formula (1.48) has several advantages:

1. its definition has a clear financial rationale supporting the analytical derivation, which is based on the hedging argument;
2. since it is an explicit function of the main market quotes, it allows for an automatic calibration to the daily volatility data;

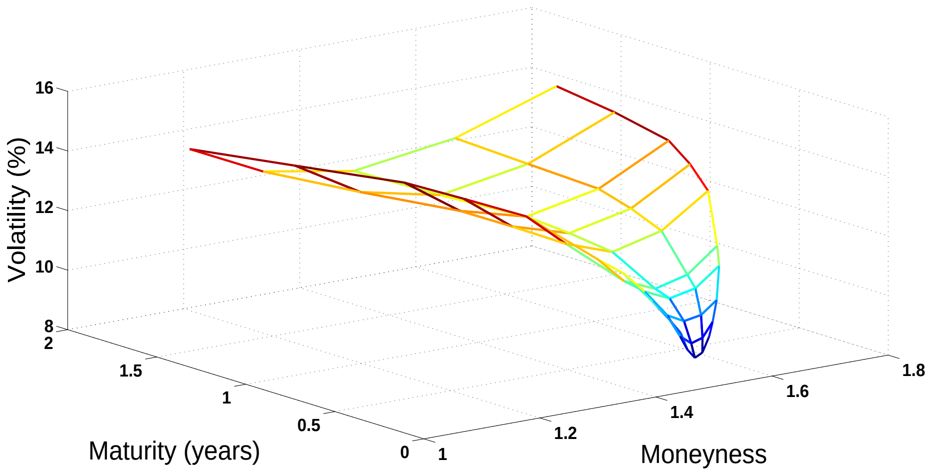

Figure 1.8: EURUSD implied volatility surface observed on September 10^{th}, 2008.

3. it can be extended to any European style derivative.
4. furthermore, it allows to detect the entire volatility surface, simply starting from the three basic market quotes.

Remark 1.4. *It has to be stressed that, once obtained the whole volatility surface by using the Vanna-Volga method, it is possible to price any option with any delta or strike, even those not quoted in the market.*

Note that, compared with the second order polynomial function in delta proposed by Malz [123], the interpolation (1.48) perfectly fits the three basic quotes, but it boosts the volatility value both for low and high extreme deltas, in accordance with typical market quotes.

Hence, the Vanna-Volga pricing formula (1.48) yields a very good approximation of the smile induced, after calibration to strikes K_i, by the most celebrate stochastic volatility models in the financial literature, in particular within the strikes range $\mathbb{K} = (K_1, K_2, K_3)$. This is not surprising, since the three strikes provide information on the second, third and fourth moments of the marginal distribution of the underlying asset. Therefore, models agreeing on these three points are likely to produce very similar smiles. As a confirmation of this statement, Castagna and Mercurio [60] show a comparison between the Vanna-Volga implied volatility smile and the one obtained from the SABR functional form of Hagan et al. [95], which is considered a standard in the market, as far as the modeling of implied volatilities is concerned. The SABR and VV curves tend to agree well in the range set by the two 10Δ options, departing from each other only for very illiquid strikes. The advantage of the VV method over the SABR model, tough, is that no calibration procedure is involved, since $\sigma(K_1)$, $\sigma(K_2)$ and $\sigma(K_3)$ are direct inputs of (1.48).

Chapter 2

The Volatility Term Structure

In 1993, the Chicago Board Options Exchange (CBOE) introduced the *volatility index*, known as VIX, the first index to measure the aggregate volatility of the US equity market. Nowadays, this index has become the premier benchmark for the stock market volatility. Often defined as the "investor fear gauge", the VIX measures market expectations of 30-day volatility implied by equity index option prices. Hence, it is considered one of the most issued financial indicators and it is widely followed by academics and quants, especially after the financial turmoil started in 2008.

The VIX was originally based on the Black-Scholes [42] implied volatilities of at-the-money options written on the S&P 100 index (OEX). Nonetheless, it is widely recognized that the constant volatility assumption of the Black-Scholes [42] model is no longer sufficient to capture modern market phenomena[1]. For this reason, on September 22, 2003, the CBOE updated the VIX definition. The new VIX is computed directly from prices of out-of-the-money call and put options, regardless of any model specification.

Given the explicit economic meaning of the new VIX and its direct link to a portfolio of options, the launch of derivatives on this index becomes the natural step ahead. On March 26, 2004, the CBOE introduced the first exchange traded VIX futures. Two years later, on February 24, 2006, contracts on VIX options started market trading. The negative correlation of volatility to equity market returns is well documented in the financial literature and suggests a diversification advantage of including volatility in an investment portfolio. Hence, VIX derivatives are designed to deliver pure volatility exposure in a single and more efficient package. Nowadays, VIX options and futures are the most actively traded contracts at the CBOE and the CBOE Futures Exchange (CFE).

In derivatives market, European options are normally quoted at different strike prices and, most importantly, at different maturities. Since the current VIX measures expected implied volatility over next 30 days only, it is evident the lack of an equivalent measure of market expectations concerning investments with time to maturity other than one month. Even though

[1]Various empirical studies illustrate that the Black-Scholes [42] model has well-known pricing biases, see Bakshi, Cao and Chen [27], Black and Scholes [41], Chance [61], Galai [90], Lauterbach and Schultz [113], MacBeth and Merville [121], Rubinstein [141] and Shastri and Tandon [147], among others.

the CBOE has recently launched a new SPX three-month volatility index[2], by now it is widely recognized that modeling the term structure of the VIX is a primary requirement. In this chapter, our goal is to construct such a term structure of the VIX, starting from option prices quoted in the market and applying the general CBOE formula for the calculation of the VIX to any available maturity.

In recent years, the popularity of the VIX has generated a rapidly growing literature on the importance of volatility trading in derivatives market. A recent attempt to study the VIX index and VIX futures is Zhang and Zhu [161], who at a later stage of their work have extended their model by allowing the long-term mean variance to be time-dependent, see Zhang and Zhu [162]. Brenner, Shu and Zhang [46] provide a comprehensive analysis of VIX futures market. Chen et al. [63] show that investors can use VIX futures to improve their equity portfolio performance, whereas Hilal et al. [98] use VIX futures to hedge the black swan risk. Lin [116] introduces an affine jump-diffusion model with jumps in both index and volatility processes. Recently, Lian and Zhu [115] provide an analytical formula for VIX futures. On the other hand, Albanese et al. [9], Chung et al. [68], Cont and Kokholm [69], Daigler and Wang [73], Li [114], Lin and Chang [117] and Sepp [145] focus on VIX options. In particular, Sepp [146] studies options on realized volatility. Carr and Lee [54] provide a complete overview of the volatility derivatives market, including volatility swaps and VIX futures and options.

Note that all this extensive literature only concerns the VIX with a single fixed 30-day maturity. Some notable works concerning the volatility term structure are those of Stein [151], who documents overreactions of long-term option prices to changes in short-term volatility, Taylor and Xu [153], who describe movements in the term structure of implied volatilities of foreign exchange options traded on the Philadelphia Stock Exchange, Campa and Chang [52], who test the expectations hypothesis in the volatility term structure of foreign exchange options and Huskaj and Nossman [102], who propose a term structure model for VIX futures.

In this chapter, we investigate features of the European options implied volatility along the time to maturity dimension, which allows us to construct a daily VIX term structure. One related study is that of Luo and Zhang [120], who analyze the term structure of the VIX, up to 15 months, and provide a two-factor model for the instantaneous squared VIX. Our theoretical model covers a wide set of stochastic volatility models, including the Heston [97] model, affine jump-models, one-factor and two-factor models, and also the Luo and Zhang [120] model, as a special case with stochastic long-run mean level.

The rest of this chapter is organized as follows. In Section 2.1, we present the generalized formula for the VIX calculation and we describe its economic and theoretical interpretation. In Section 2.2, we show applications for one-factor and two-factor models. Finally, in the last section, we draw some considerations about implementation problems.

[2] On November 12, 2007, the CBOE launched the SPX three-month volatility with ticker symbol “VXV”, which employs the same method used to calculate the VIX, but including options with a constant maturity of 93 calendar days.

2.1 The VIX Index

The idea of a volatility index and financial instruments based on such an index was first developed and described by Brenner and Galai [45] in 1989. The VIX was originally computed as an average of the Black-Scholes [42] implied volatilities of near-the-money OEX American option prices. In 2003, the CBOE updated the VIX definition and revised its methodology of calculation, following the theoretical results of Carr and Madan [55] and Demeterfi et al. [75], who proposed the original idea of replicating the realized variance by a static portfolio of out-of-the-money call and put options.

The main differences between the two indices[3] are that the new VIX is model-free and it is based on the SPX European options. Furthermore, since the new VIX is an average of the weighted prices of out-of-the-money SPX options, it is able to incorporate information from the volatility smile, by using a wider range of strikes. This new methodology transformed the VIX from an abstract concept into a practical standard for trading and hedging volatility, by supplying the opportunity to replicate volatility exposure with a portfolio of SPX options.

In contrast to equity indices, such as the SPX, which are calculated using prices of their component stocks, the VIX is a volatility index computed with options rather than stocks, with the price of each option reflecting the expectation of the market participants as for the future volatility. Following conventional indices, though, the VIX employs rules for selecting component options and a formula to compute index values.

Given a panel of European options with common maturity $T > 0$, the generalized formula used to compute the VIX at time $0 \leq t < T$ is

$$v_t^2 = \frac{2}{T} \sum_i \frac{\Delta K_i}{K_i^2} \mathrm{e}^{rT} Q(T, K_i) - \frac{1}{T} \left[\frac{F}{K_0} - 1 \right]^2, \tag{2.1}$$

where K_i is the strike of the i^{th} out-of-the-money option considered, ΔK_i denotes the interval between strike prices, defined as

$$\Delta K_i = \frac{(K_{i+1} - K_{i-1})}{2},$$

that is, half the difference between the strike on either side[4] of K_i, r is the risk-free interest rate, $Q(T, K_i)$ is the price of the out-of-the-money option with maturity T and strike K_i, F is the forward index level derived from index option prices and K_0 is the first strike immediately below F.

To determine the forward index level F, let $C(T, \hat{K}_i)$ and $P(T, \hat{K}_i)$ be two options with same strike $\hat{K}_i$, whose price difference is the smallest. Then, via the put-call parity relation, we have

$$F = \mathrm{e}^{rT} \left[C(T, \hat{K}_i) - P(T, \hat{K}_i) \right] + \hat{K}_i. \tag{2.2}$$

[3]See Carr and Wu [57] for a detailed comparison between the two indices.

[4]At the extreme edges of any given strip of options, ΔK_i is simply the difference between K_i and the adjacent strike price.

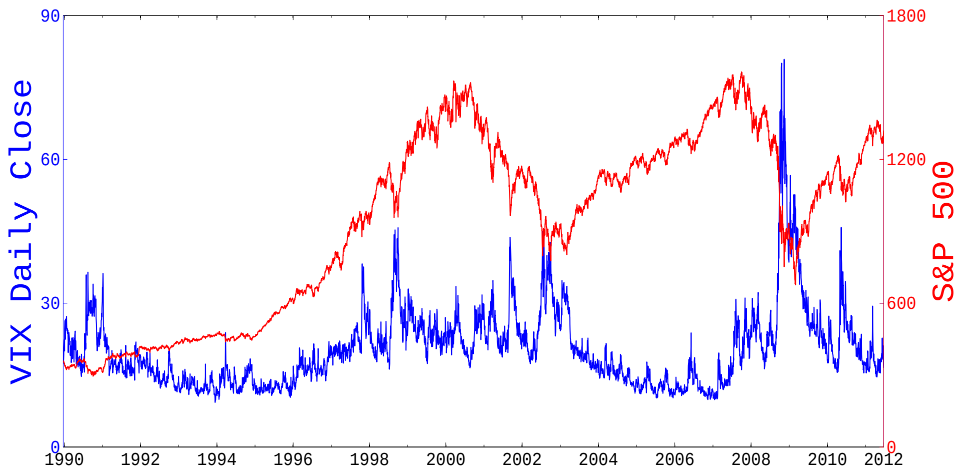

Figure 2.1: VIX and S&P 500 Indices from January 2, 1990 to July 31, 2012

After choosing K_0 as the first strike below F, the calculation involves all available out-of-the-money puts with strikes $K_i < K_0$, all out-of-the-money calls with strikes $K_i > K_0$ and both the put and the call with strike K_0. Note that two options are selected at K_0, while a single option, either a put or a call, is used for every other strike price. Furthermore, since $K_0 < F$, the mid-price $Q(T, K_0)$ implies that one in-the-money call with strike K_0 is used in the computation. For this reason, the last term in Equation (2.1) represents the adjustment needed to convert this in-the-money call into an out-of-the-money put, by virtue of the put-call parity relation. It is worth noting that, as volatility rises and falls, the strikes range considered tends to expand and contract. As a result, the number of options used to compute the VIX may vary from day to day.

The CBOE uses Equation (2.1) to calculate v_t^2 at the two closest available maturities before and after 30 days, which are then interpolated by taking the weighted average. Finally, the quoted VIX is given by

$$VIX_t = 100 \times v_t, \tag{2.3}$$

and it represents the annualized volatility percentage over the upcoming 30 calendar days[5].

2.1.1 Economic and Theoretical Interpretation

One of the most valuable features of the VIX is the existence of more than 20 years of historical prices. This extensive data set provides investors with a useful perspective of how option prices have behaved in response to a variety of market conditions. For instance, Figure 2.1 shows the historical behavior of both VIX and SPX from January 2, 1990 to July 31, 2012. The highest closing value of the VIX, equal to 80.86, was reached in November 20, 2008, in conjunction with the uncertainty of the markets, affected by the recent tough financial crisis. From the graph, some considerations are possible. Firstly, we can see that the VIX is more volatile that the SPX.

[5] All details of the computation are available at http://www.cboe.com/micro/vix/vixwhite.pdf.

This means that the variability of the volatility itself, namely the *vol of vol*, is higher than the volatility exhibited by the SPX. Furthermore, in certain periods the two indices tend to move in opposite directions. In particular, we can see this inverse correlation in 2001, when the market collapsed after the attack on the World Trade Center, and then in 2008, when quotations have sharply fallen due to the global financial crisis.

Although the VIX is often called the "fear index", a high VIX is not necessarily bearish for stocks. Instead, the VIX is a measure of the volatility perceived by the market in either direction. In practical terms, when investors anticipate large upside volatility, they are unwilling to sell call options, unless they receive a large premium. On the other hand, call option buyers are willing to pay such high premium only if their forecasts confirm an upside market movement. As a consequence, the resulting aggregate increase in call option prices raises the VIX value. In a similar way, the aggregate growth in the premiums of put options, that occurs when option buyers and sellers anticipate a likely sharp downside market movement, has the same increasing effect on the VIX index. When the market is believed as likely to soar as to plummet, writing a call or a put option may look equally risky in the event of a sudden large movement in either direction. Hence, high VIX values occur when investors anticipate that huge market moves are likely to happen, whether downward or upward.

Theoretically, the VIX is based on a portfolio of short-maturity out-of-the-money SPX options over a continuum of strike prices, whose value equals that of a *variance swap*, see Britten-Jones and Neuberger [49], Carr and Wu [58] and Jiang and Tian [107]. The latter is defined as a forward contract, which allows to speculate on, or hedge, risks associated with the total quadratic volatility of some underlying product, such as an exchange rate, an interest rate or a stock index, over a fixed interval of time. In practice, at maturity the long side of the variance swap contract receives the realized variance and pays a fixed variance rate, which is the variance swap rate.

Let $(\Omega, \mathfrak{E}, \mathbb{P}) \equiv \Omega$ be a probability space endowed with a filtration $(\mathfrak{F}_t)_{t\geq 0}$ and let $(S_t)_{t\geq 0} \equiv S_t$ be the logarithm of a financial equity asset price evolving in continuous time, whose quadratic variation $[S,S]_t$ is an adapted, increasing and càdlàg (i.e. with paths that are a.s. right continuous with left limits) process. Hence, following Todorov and Tauchen [154] approach, the VIX can be written as

$$VIX_t = \mathbf{E}^{\mathbb{Q}}\left[[S,S]_T - [S,S]_t \mid \mathfrak{F}_t\right], \qquad T > t, \tag{2.4}$$

where $T > 0$ is given and the expectation is taken under the risk-neutral distribution $\mathbb{Q}$, whose existence is guaranteed by no arbitrage assumptions. The quadratic variation process $[S,S]_t$ can be split into two components:

$$[S,S]_t = [S,S]_t^c + [S,S]_t^d, \tag{2.5}$$

which correspond to the quadratic variation of the continuous and discontinuous parts, respectively, of the price process S_t. We assume absolute continuity of $[S,S]_t^c$, which is a standard assumption in finance, that is

$$[S,S]_t^c = \int_0^t \sigma_s^2 \, ds, \tag{2.6}$$

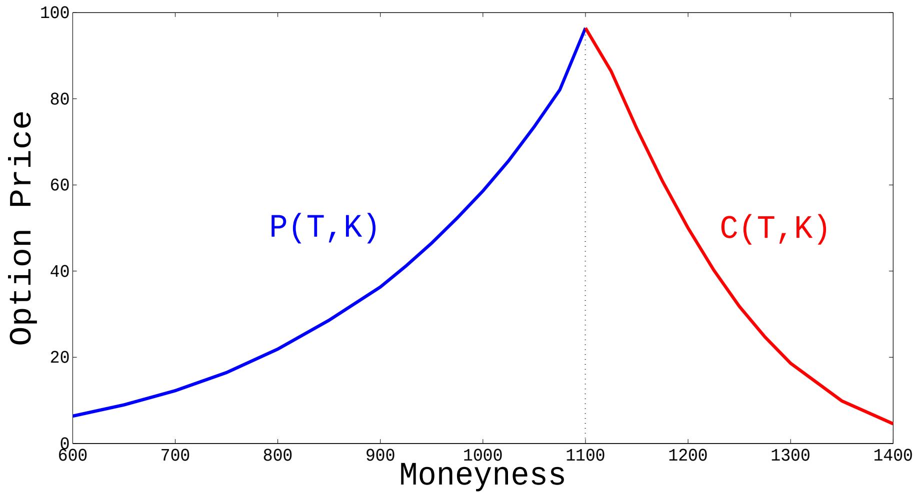

Figure 2.2: Moneyness and Option Prices

where σ_t^2 is the *spot variance* of S_t, also known as *instantaneous variance*[6].

Definition 2.1. *The* spot variance σ_t^2 *is defined as the instantaneous increment to the quadratic variation of the continuous martingale component of the asset price process* S_t*.*

Then, if there are no jumps in the price process S_t, the VIX obviously equals the familiar *expected risk-neutral integrated variance*, given by

$$VIX_t = \mathbf{E}^{\mathbb{Q}} \left[\int_t^T \sigma_s^2 ds \,\middle|\, \mathfrak{F}_t \right]. \tag{2.7}$$

The last expression is a well-known result and serves as a theoretical basis for the VIX index, see Britten-Jones and Neuberger [49] and Jiang and Tian [107].

It is important to bear in mind the fundamental distinction between the *observed* VIX and the *unobserved* spot variance. The observed VIX is the CBOE measurement of the 30-day expected integrated volatility in (2.7). We use these observations to make inference about important features of the random process for the spot variance. The inference is complicated by the fact that the VIX is forward looking and its increments are generated by movements in variables that influence the conditional expectation in (2.7). Moreover, we can only observe discrete realizations of the VIX, meaning that we have to take into account some implementation errors as we work in a continuous-time framework. This issue is deeply studied in the last section of this chapter. Now, consider a portfolio Π of European call and put options, weighted inversely proportional to the square of their strike prices. The value of Π at time t is given by

$$\Pi_t (T, K_0) = \int_0^{K_0} \frac{P(T,K)}{K^2} dK + \int_{K_0}^{\infty} \frac{C(T,K)}{K^2} dK, \tag{2.8}$$

[6]We follow the convention of using the term *variance* for squared quantities and the term *volatility* for measures of standard deviation. Of course, variance is easier to manage with mathematical instruments, while volatility is easier to interpret because it is expressed in the same units as the data itself.

where $K_0 \in (0, \infty)$ is a real number, $P(T, K)$ and $C(T, K)$ are prices of put and call options, respectively, at time t, with strike K and maturity $T > t$. Figure 2.2 shows that, relating all the available strikes and option prices for a given maturity, the VIX value represents the area under the curve of prices, which increases for put out-of-the-money options until K_0 and then decreases for out-of-the-money calls. In the following of this chapter, we present the model-free relation between the VIX and option prices observed in the market.

2.1.2 Discontinuities in the Asset Price

The importance of considering jumps has long been stressed in the empirical option pricing literature, such as Bakshi, Cao and Chen [27], Bates [38], Chernov and Ghysel [64], Duffie, Pan and Singleton [79], Eraker [84] and Pan [134]. In light of this issue, we release the assumption of absence of jumps in the price process S_t. Thus, the VIX expression (2.7) can be rewritten as

$$VIX_t = \mathbf{E}^{\mathbb{Q}} \left[\int_t^T \sigma_s^2 \, ds \;\middle|\; \mathfrak{F}_t \right] + \mathbf{E}^{\mathbb{Q}} \left[[S,S]_T^d - [S,S]_t^d \mid \mathfrak{F}_t \right], \tag{2.9}$$

where the first term is the risk-neutral expectation of the forward integrated variance previously introduced, while the second is the risk-neutral expected contribution of the price jumps.

Furthermore, the spot volatility process itself can also be split into continuous and discontinuous parts, that is

$$\sigma_t = \sigma_t^c + \sigma_t^d.$$

Let σ_{t-} be the left-hand limit of the function σ_s as s approaches t from the left, that is, $\sigma_{t-} = \lim_{s \to t} \sigma_s$. A jump discontinuity $\sigma_t^d - \sigma_{t-}^d$ influences the entire trajectory $\mathbf{E}^{\mathbb{Q}} [\sigma_{t+s} \mid \mathfrak{F}_t]$, with $s \geq 0$, and thereby induces a jump discontinuity in the VIX, given Equation (2.9). Historically, stochastic volatility models have assumed continuous spot volatility, that is $\sigma_t \equiv \sigma_t^c$. Recently, there has been interest in pure jump stochastic volatility models, where $\sigma_t \equiv \sigma_t^d$, see Barndorff-Nielsen and Shephard [31]. Moreover, Duffie, Pan and Singleton [79] show that the two opposite cases can be combined, as in the double-jump affine model.

Indeed, a joint consideration of both volatility and jump risk factors is expected to better perform the dynamics of equity returns, as it could explain the observed phenomena of volatility smile and smirk, and, as a result, improves the valuation of options. However, empirical results in the literature appear to be mixed. For instance, Andersen, Benzoni and Lund [14] and Eraker, Johannes and Polson [85] conclude that allowing jumps in prices can improve the fitting for the time series of equity returns. On the other hand, Bakshi, Cao and Chen [27], Bates [38], Eraker [84] and Pan [134] offer different and inconsistent results in terms of improvement on option pricing. In their recent work, Luo and Zhang [120] have demonstrated that the jump component in the dynamics of the SPX is negligible in modeling the VIX. Hence, in most cases there is no joint significance in volatility and jump risk premium estimates.

2.1.3 Model-free Implied Volatility

Now, we examine in depth the theoretical foundations of the VIX, within the broader context of the *model-free implied volatility.* In other terms, we study the information content of the options market, showing an alternative measure of implied volatility, other than the celebrate Black-Scholes [42] one, which is independent of any option pricing model.

The idea underneath the model-free implied volatility stems from the development of variance swaps and it is derived by Britten-Jones and Neuberger [49], building on the pioneering work of Breeden and Litzenberger [44]. It is further refined by Carr and Madan [55], Carr and Wu [57], Demeterfi et al. [76] and Jiang and Tian [107]. Unlike the traditional concept of implied volatility, the model-free implied volatility is not based on any specific option pricing model. Instead, it is derived entirely from no arbitrage conditions. In the following, we show that the volatility measure underlying the VIX is theoretically equivalent to the model-free implied volatility formulated in Britten-Jones and Neuberger [49].

Assume the existence of a complete set of European call options, with a continuum of strikes $K \geq 0$ and a continuum of maturities $T \geq 0$. The price of an option at time $0 \leq t < T$ is denoted by $C(T, K)$. Let $(S_t)_{t \geq 0} \equiv S_t$ be the price of an underlying equity asset, following a diffusion process with time-varying volatility. Under the risk-neutral probability measure $\mathbb{Q}$, S_t is a positive martingale and it can be expressed as

$$\frac{dS_t}{S_t} = \sigma\left(t, \cdot\right) dW_t, \tag{2.10}$$

where $(W_t)_{t \geq 0} \equiv W_t$ is a standard Wiener process and the instantaneous volatility $\sigma(t, \cdot)$ may depend on S_t, the history of S_t or other unspecified state variables. For simplicity, but without loss of generality, it is further assumed that the underlying asset does not make any interim payments, such as dividends or interest, and the risk-free interest rate is zero. Under this general setting, the following Proposition holds.

Proposition 2.1 (Britten-Jones and Neuberger [49])**.** *The expected risk-neutral integrated return variance between two arbitrary dates, say T_1 and T_2 with $T_1 < T_2$, is completely specified by the set of option prices expiring on the two dates, that is*

$$\mathbf{E}^{\mathbb{Q}}\left[\int_{T_1}^{T_2} \left(\frac{dS_u}{S_u}\right)^2\right] = 2\int_0^{\infty} \frac{C(T_2, K) - C(T_1, K)}{K^2}\, dK, \tag{2.11}$$

where $\mathbf{E}^{\mathbb{Q}}\left[\cdot\right]$ denotes the expectation taken under the risk-neutral probability measure.

Proof. For the diffusion process (2.10) to support prices $C(T, K)$, it is required that

$$C(T, K) = \mathbf{E}^{\mathbb{Q}}\left[(S_T - K)^+\right], \qquad T, K \geq 0,$$

where $(S_T - K)^+ = \max\left(S_T - K, 0\right)$ is the option payoff function at maturity T. Let $\phi\left(S_t\right)$ be the risk-neutral density of the asset price S_t, such as a log-normal distribution in the Black-

Scholes [42] model. Then, we can write the above condition as

$$C(T,K) = \int_K^\infty (S_T - K)\,\phi(S_T)\,dS_T, \qquad T,K \geq 0.$$

Differentiating with respect to K yields

$$\frac{\partial C(T,K)}{\partial K} = \int_K^\infty -\phi(S_T)\,dS_T, \qquad T,K \geq 0,$$

and differentiating again, we have

$$\frac{\partial^2 C(T,K)}{\partial K^2} = \phi(K), \qquad T,K \geq 0.$$

Thus, we can obtain the density by differentiating twice the call price with respect to the strike. Now, consider the derivative of the call price with respect to maturity T, given by[7]

$$\frac{\partial C(T,K)}{\partial T} = \frac{\mathbf{E}^{\mathbb{Q}}\left[\partial (S_T - K)^+\right]}{\partial T} = \frac{\mathbf{E}^{\mathbb{Q}}\left[\mathbf{1}_{S_T \geq K}\,dS_T + \frac{1}{2}\delta(S_T - K)(dS_T)^2\right]}{\partial T}.$$

Since S_T is a martingale, the previous expression simplifies to

$$\frac{\partial C(T,K)}{\partial T} = \frac{1}{2}\mathbf{E}^{\mathbb{Q}}\left[\delta(S_T - K)\,\sigma^2(T,\cdot)S_T^2\right],$$

which can be rewritten as

$$\frac{\partial C(T,K)}{\partial T} = \frac{1}{2}\int_0^\infty \int_0^\infty \delta(S_T - K)\,\sigma^2(T,\cdot)S_T^2\,\phi(S_T)\,\phi(\sigma(T,\cdot) \mid S_T)\;dS_T\,d\sigma(T,\cdot),$$

where $\phi(S_T)\,\phi(\sigma(T,\cdot) \mid S_T)$ is the joint distribution of S_T and $\sigma(T,\cdot)$, given by the product between the marginal distribution of S_T and the distribution of $\sigma(T,\cdot)$ conditional on S_T. By virtue of the property of the Dirac delta function

$$\int_{-\infty}^{+\infty} \delta(x - x_0)\,f(x)\,dx = f(x_0),$$

we can evaluate the inner integral and we obtain

$$\frac{\partial C(T,K)}{\partial T} = \frac{1}{2}K^2\phi(K)\int_0^\infty \sigma^2(T,\cdot)\,\phi(\sigma(T,\cdot) \mid S_T = K)\;d\sigma(T,\cdot).$$

The last integral represents the expectation of the squared volatility conditional on the asset

[7] Thanks to the generalized Itô formula applied to the option payoff function $(S_T - K)^+$, we have

$$d(S_T - K)^+ = \mathbf{1}_{S_T \geq K}\,dS_T + \frac{1}{2}\delta(S_T - K)(dS_T)^2,$$

where the first derivative of the payoff function with respect to S_T is the indicator function $\mathbf{1}_{S_T \geq K}$ and the second derivative again with respect to S_T is the Dirac delta function $\delta(S_T - K)$.

price $S_T = K$, that is

$$\frac{\partial C(T,K)}{\partial T} = \frac{1}{2} K^2 \phi(K)\, \mathbf{E}^{\mathbb{Q}} \left[\sigma^2(T,\cdot) \mid S_T = K \right].$$

Thus, inverting the previous equation, we can have a formal expression for the conditional expected variance, given by

$$\mathbf{E}^{\mathbb{Q}} \left[\sigma^2(t,\cdot) \mid S_t = K \right] = \frac{2}{K^2} \frac{1}{\phi(K)} \frac{\partial C(t,K)}{\partial t}. \tag{2.12}$$

In order to obtain an unconditional expectation of the instantaneous squared volatility, we need to integrate across strikes the previous equation. Hence, we obtain

$$\mathbf{E}^{\mathbb{Q}} \left[\sigma^2(T,\cdot) \right] = \int_0^\infty \mathbf{E}^{\mathbb{Q}} \left[\sigma^2(T,\cdot) \mid S_T = K \right] \phi(K)\ dK.$$

Substituting (2.12), we have

$$\begin{aligned} \mathbf{E}^{\mathbb{Q}} \left[\sigma^2(T,\cdot) \right] &= \int_0^\infty \frac{2}{K^2} \frac{1}{\phi(K)} \frac{\partial C(T,K)}{\partial T} \phi(K)\ dK \\ &= 2 \int_0^\infty \frac{1}{K^2} \frac{\partial C(T,K)}{\partial T}\ dK. \end{aligned}$$

Finally, we integrate with respect to time and we have

$$\mathbf{E}^{\mathbb{Q}} \left[\int_{T_1}^{T_2} \sigma^2(s,\cdot)\, ds \right] = 2 \int_0^\infty \frac{C(T_2,K) - C(T_1,K)}{K^2}\, dK,$$

which is the desired result. □

Proposition 2.1 implies that the risk-neutral equity return squared volatility can be obtained from a set of option prices with different maturities, observed at a single point in time. Since Equation (2.11) is derived without any specific assumption on the underlying stochastic process, the risk-neutral return variance is totally model-free. As a consequence, the right-hand-side of Equation (2.11) [resp. its square root] will be referred to as the *model-free implied variance* [resp. *model-free implied volatility*]. It is worth noting that this method is internally consistent, prices of options at all strikes are used in a theoretically coherent manner and nothing is assumed about the asset price process except from continuity.

Since we are mostly interested in forecasting volatility from the present $t = 0$ to some future date $T > 0$, a more useful special case of Equation (2.11) is given by

$$\mathbf{E}^{\mathbb{Q}} \left[\int_0^T \left(\frac{dS_u}{S_u} \right)^2 \right] = 2 \int_0^\infty \frac{C(T,K) - \max(S_0 - K, 0)}{K^2}\, dK. \tag{2.13}$$

In other words, the expression above provides the risk-neutral volatility forecast between 0 and T only by means of the set of call options with maturity T.

Equation (2.13) is stated in spot prices. When applications require the use of forward prices some appropriate adjustments are needed. Consider the assumption of non zero deterministic interest rates and let $F_t = S_t \mathrm{e}^{r(T-t)}$ and $C^F(T,K) = C(T,K)\mathrm{e}^{r(T-t)}$ be the forward asset price and the forward option price at time $t \in [0,T]$, respectively. Now, substituting expressions for F_t and $C^F(T,K)$ in (2.13), we have

$$\mathbf{E}^{\mathbb{F}}\left[\int_0^T \left(\frac{dF_u}{F_u}\right)^2\right] = 2\int_0^\infty \frac{C^F(T,K) - \max(F_0 - K, 0)}{K^2}\,dK, \tag{2.14}$$

where $\mathbf{E}^{\mathbb{F}}[\cdot]$ is the expected value under the forward probability measure. The equation above provides an alternative definition of the model-free implied squared volatility in forward prices.

A variance definition closely related to the model-free implied volatility is that of *fair value of future variance*, the *DDKZ variance* hereafter, from Demeterfi, Derman, Kamal and Zou [75], who developed this concept in 1999. Unlike the traditional ex-ante volatility measure, such as the Black-Scholes [42] implied volatility, the DDKZ variance is determined from market observables, usually option prices and interest rates, with no reference to any specific option pricing model. In particular, the price of the underlying asset may follow any diffusion process.

Definition 2.2. *The DDKZ variance is defined as*

$$V_{DDKZ} = \frac{2}{T}\left[rT - \left(\frac{S_0}{\hat{S}}\mathrm{e}^{rT} - 1\right) - \ln\frac{\hat{S}}{S_0} + \mathrm{e}^{rT}\int_0^{\hat{S}} \frac{P(T,K)}{K^2}\,dK + \mathrm{e}^{rT}\int_{\hat{S}}^\infty \frac{C(T,K)}{K^2}\,dK\right], \tag{2.15}$$

where S_0 is the current asset price, $\hat{S}$ is an arbitrary equity price, typically chosen close to the forward price, $P(T,K)$ and $C(T,K)$ are put and call option prices, respectively, with maturity T and strike K, and r is the risk-free interest rate.

Now, we show that the DDKZ variance is conceptually equal to the model-free variance of Britten-Jones and Neuberger [49].

Proposition 2.2. *The* fair value of future variance V_{DDKZ} *developed by Demeterfi et al. [75] is equivalent to the* model-free implied variance *introduced by Britten-Jones and Neuberger [49], which is equal to*

$$V_{BJN} = \frac{2}{T}\int_0^\infty \frac{C^F(T,K) - \max(F_0 - K, 0)}{K^2}\,dK. \tag{2.16}$$

Proof. From Equation (2.16), partitioning the integral into two segments at the forward price $F_0 = S_0\mathrm{e}^{rT}$, we have

$$\begin{aligned} V_{BJN} &= \mathbf{E}^{\mathbb{F}}\left[\frac{1}{T}\int_0^T \left(\frac{dF_u}{F_u}\right)^2\right] \\ &= \frac{2}{T}\mathrm{e}^{rT}\left[\int_0^{F_0} \frac{C(T,K) - S_0 + K\mathrm{e}^{-rT}}{K^2}\,dK + \int_{F_0}^\infty \frac{C(T,K)}{K^2}\,dK\right]. \end{aligned}$$

Then, by virtue of the put-call parity, a more compact expression of the model-free implied variance is given by

$$V_{BJN} = \frac{2}{T}\mathrm{e}^{rT}\left[\int_0^{F_0} \frac{P(T,K)}{K^2}\,dK + \int_{F_0}^{\infty} \frac{C(T,K)}{K^2}\,dK\right]. \tag{2.17}$$

The above equation can be rewritten as

$$V_{BJN} = \frac{2}{T}\mathrm{e}^{rT}\left[\int_0^{\hat{S}} \frac{P(T,K)}{K^2}\,dK + \int_{\hat{S}}^{\infty} \frac{C(T,K)}{K^2}\,dK + \int_{\hat{S}}^{F_0} \frac{P(T,K)-C(T,K)}{K^2}\,dK\right],$$

where $\hat{S}$ is an arbitrary equity price, such that $\hat{S} < F_0$. Using the put-call parity once again, we have

$$V_{BJN} = \frac{2}{T}\mathrm{e}^{rT}\left[\int_0^{\hat{S}} \frac{P(T,K)}{K^2}\,dK + \int_{\hat{S}}^{\infty} \frac{C(T,K)}{K^2}\,dK + \int_{\hat{S}}^{F_0} \frac{K\mathrm{e}^{-rT}-S_0}{K^2}\,dK\right].$$

Finally, solving the last integral inside the brackets, it holds

$$V_{BJN} = \frac{2}{T}\left[rT - \left(\frac{S_0}{\hat{S}}\mathrm{e}^{rT}-1\right) - \ln\frac{\hat{S}}{S_0} + \mathrm{e}^{rT}\int_0^{\hat{S}} \frac{P(T,K)}{K^2}\,dK + \mathrm{e}^{rT}\int_{\hat{S}}^{\infty} \frac{C(T,K)}{K^2}\,dK\right].$$

This alternative expression for the model-free implied variance is exactly equal to the DDKZ variance in Equation (2.15), which ends the proof. □

Proposition 2.2 is an important result for two reasons. First, it provides a direct connection between two concepts of variance, which have been developed separately for different purposes. In fact, it merges key results in the studies concerning variance swaps and implied distributions. Hence, the theoretical foundation of the VIX is embedded in the broader context of the model-free implied volatility. Second, Proposition 2.2 also implies that empirical findings on the information content of the model-free implied volatility can be directly applied to the concept of the DDKZ variance. As shown by Jiang and Tian [107], the model-free implied variance subsumes all the information contained in the Black-Scholes [42] implied volatility as well as in the historical variance. Thus, it is a more efficient forecast for the future realized volatility. Since the VIX is based on the model-free implied squared volatility, it should be considered a suitable indicator of the expectations of economic agents.

Proposition 2.3. *The VIX represents the risk-neutral forecast of the instantaneous variance and its value is equivalent to that of a portfolio of out-of-the-money options weighted inversely proportional to the square of their strike prices. In mathematical terms, we have*

$$VIX_t = \mathbf{E}^{\mathbb{Q}}\left[\int_t^T \sigma_s^2\,ds \;\middle|\; \mathfrak{F}_t\right] = 2\int_0^{\infty} \frac{C(T,K)-C(t,K)}{K^2}\,dK. \tag{2.18}$$

Clearly, the above Proposition provides an immediate link between the VIX and option prices available in the market.

2.2 Semi-nonparametric Model for the VIX Term Structure

So far, we have discussed the VIX calculation regardless of any specification about the variance dynamics. Although the study of the VIX with a model-independent approach has its advantages, now it is better to consider a specific model, in order to illustrate the analysis of the VIX term structure. Hence, here we present a general semi-nonparametric stochastic volatility model, at first introducing only one factor for the diffusion component and then extending the model to two factors. The idea underneath this model highlights the importance of the forecasting power and information content of option prices.

2.2.1 One-factor Model

Let $(\Omega, \mathfrak{E}, \mathbb{P}) \equiv \Omega$ be a probability space endowed with a filtration $(\mathfrak{F}_t)_{t\geq 0}$, let $r \in \mathbb{R}$ be a real constant and let $(\beta_t)_{t\geq 0}$ be a measurable and locally bounded càdlàg (i.e. with paths that are a.s. right continuous with left limits) process. We assume that the logarithm of a financial asset price $(S_t)_{t\geq 0} \equiv S_t$ is modeled by an Itô process and the instantaneous variance $(V_t)_{t\geq 0} \equiv V_t$ follows a *mean-reverting* process, that is

$$\begin{aligned} dS_t &= r dt + \sqrt{V_t} dW_t \\ dV_t &= \kappa(\omega - V_t) dt + \beta_t dZ_t \end{aligned} \quad , \tag{2.19}$$

where $\kappa \in \mathbb{R}$ is the *rate of mean reversion*, $\omega \in \mathbb{R}$ is the *long-run mean level* of V_t, $(W_t)_{t\geq 0} \equiv W_t$ and $(Z_t)_{t\geq 0} \equiv Z_t$ are two correlated Wiener processes, with correlation coefficient equal to $\rho \in [-1, 1]$, thus introducing an asymmetric return-variance relation into the asset price dynamics.

The mean-reverting process is one of several approaches used to model interest rates, currency exchange rates, prices of commodities and, in particular, stochastic volatility, due to the presence of its interesting characteristic known as *mean reversion*, that is, the tendency of a stochastic process to remain near, or tend to return over time, to a long-run average value. From a financial modeling perspective, mean-reverting refers to a linear pull-back term in the drift of the stochastic process itself, or in the drift of some underlying of which the process is a function.

The Markov process V_t, given by

$$V_t = V_0 \mathrm{e}^{-\kappa t} + \omega(1 - \mathrm{e}^{-\kappa t}) + \mathrm{e}^{-\kappa t} \int_0^t \beta_s \mathrm{e}^{\kappa s} dZ_s, \tag{2.20}$$

obtained from Equation (2.19) by virtue of the Itô's lemma, is normally distributed, with conditional expected value and variance equal to

$$\begin{aligned} \mathbf{E}^{\mathbb{Q}}\left[V_t \mid \mathfrak{F}_t\right] &= V_0 \mathrm{e}^{-\kappa t} + \omega(1 - \mathrm{e}^{-\kappa t}), \\ \mathbf{Var}^{\mathbb{Q}}\left[V_t \mid \mathfrak{F}_t\right] &= \mathrm{e}^{-2\kappa t} \int_0^t \beta_s^2 \mathrm{e}^{2\kappa s}\, ds. \end{aligned}$$

Integrating the expected value $\mathbf{E}^{\mathbb{Q}}\left[V_t \mid \mathfrak{F}_t\right]$ with respect to time, we obtain

$$\mathbf{E}^{\mathbb{Q}}\left[\int_0^t V_s\, ds\right] = \frac{V_0}{\kappa}(1 - \mathrm{e}^{-\kappa t}) + \omega t - \frac{\omega}{\kappa}(1 - \mathrm{e}^{-\kappa t}). \tag{2.21}$$

The last equation represents our theoretical model, which is semi-nonparametric since it does not depend on β_t, but only on the drift parameters.

Presence of Jumps

Several works examine equity price models with jumps in returns and stochastic volatility, see Andersen, Benzoni, and Lund [14], Bakshi, Cao, and Chen [27], Bates [38] and Pan [134]. While it is clear that both stochastic volatility and jumps in returns are important components, Bakshi, Cao, and Chen [27], Bates [38] and Pan [134] find strong evidence of significant misspecification in terms of internal consistency of a stochastic volatility model. In particular, Bakshi, Cao, and Chen [27] demonstrate that each stochastic volatility model requires highly implausible levels of return-volatility correlation and of volatility variation, in order to rationalize the negative skewness and excess kurtosis implicit in option prices. Additionally, Bates [38] and Pan [134] show that the higher moments of volatility changes are inconsistent with the diffusion specification. These results point out the presence of an additional sharply moving factor driving conditional volatility, which has a persistent component, unlike jumps in returns. Indeed, jumps in volatility could provide such a factor.

As seen in Equation (2.9), when the dynamics of the underlying asset follows a jump-diffusion process, a jump discontinuity contributes to the total variance. In order to identify the effect of the presence of jumps in the asset price, we have to compare VIX expressions in the two different settings, that is whether or not jumps are added into the dynamics of the underlying. Consider a general jump-diffusion process for the equity asset price S_t under the risk-neutral measure $\mathbb{Q}$, given by

$$\frac{dS_t}{S_{t-}} = rdt + \sqrt{V_t}dW_t + \left(\mathrm{e}^x - 1\right)dN_t - \lambda \mathbf{E}^{\mathbb{Q}}\left(\mathrm{e}^x - 1\right)dt, \tag{2.22}$$

where S_{t-} is the value of S_t before a possible jump occurs, N_t is a pure jump process,assumed to be independent of W_t, with intensity λ, x is the jump size of $\ln S_t$ and the last term $\lambda \mathbf{E}^{\mathbb{Q}}\left(\mathrm{e}^x - 1\right)dt$ is used to compensate the jump innovation. Applying the Itô lemma with jumps to Equation (2.22) gives the logarithmic process of S_t, that is

$$d\,\ln S_t = \left[r - \frac{1}{2}V_t - \lambda \mathbf{E}^{\mathbb{Q}}\left(\mathrm{e}^x - 1\right)\right]dt + \sqrt{V_t}dW_t + xdN_t. \tag{2.23}$$

Since the jump component also affects the variance of S_t, the instantaneous total variance becomes

$$\sigma_t^2 = V_t + \mathbf{E}^{\mathbb{Q}}\left(\lambda x^2\right), \tag{2.24}$$

where the first term is diffusion variance and the second term is jump variance. Then, according to Equation(2.7), the VIX is given by

$$\begin{aligned}\widehat{VIX}_t &= \mathbf{E}^{\mathbb{Q}}\left[\frac{1}{T}\int_t^T \sigma_s^2\, ds\right] \\ &= \frac{1}{T}\int_t^T \mathbf{E}^{\mathbb{Q}}\left[\sigma_s^2\right] ds \\ &= \frac{1}{T}\int_t^T \mathbf{E}^{\mathbb{Q}}\left[V_s + \lambda x^2\right] ds \\ &= \frac{1}{T}\int_t^T \mathbf{E}^{\mathbb{Q}}\left[V_s\right] ds + \mathbf{E}^{\mathbb{Q}}\left[\lambda x^2\right], \end{aligned} \tag{2.25}$$

where we use property of iterated expectations. On the other hand, as shown by Brenner, Shu and Zhang [46], the VIX can also be computed using Equations (2.22) and (2.23), that is

$$\begin{aligned}\widetilde{VIX}_t &= \frac{2}{T}\mathbf{E}^{\mathbb{Q}}\left[\int_t^T \frac{dS_u}{S_u} - d\left(\ln S_u\right)\right] \\ &= \frac{2}{T}\mathbf{E}^{\mathbb{Q}}\left[\int_t^T \left(\frac{1}{2}V_u + \lambda\left(\mathrm{e}^x - 1 - x\right)\right) du\right] \\ &= \frac{1}{T}\int_t^T \mathbf{E}^{\mathbb{Q}}\left[V_u\right]\, du + \mathbf{E}^{\mathbb{Q}}\left[2\lambda\left(\mathrm{e}^x - 1 - x\right)\right]. \end{aligned} \tag{2.26}$$

Therefore, the difference between the two formulas (2.25) and (2.26) is given by

$$\begin{aligned}\Delta &= \mathbf{E}^{\mathbb{Q}}\left[2\lambda\left(\mathrm{e}^x - 1 - x\right) - \lambda x^2\right] \\ &\approx \lambda\mathbf{E}^{\mathbb{Q}}\left[\frac{1}{3}x^3\right]. \end{aligned} \tag{2.27}$$

This difference is small for general jump parameter values. For instance, with $\lambda = 0.4845$ and $x = -0.0789$, which are obtained in Chang, Zhang and Zhao [62], we have $\Delta = -0.00008$. Thus, for a general value of the VIX at 20, we have $\widehat{VIX}_t = 20.06$ and $\widetilde{VIX}_t = 20$, which corresponds to 0.3% overvalue by using Equation (2.25) instead of (2.26). In other words, the difference between the jump components in (2.25) and (2.26) is negligible, and then both are useful formulas to compute the VIX in the presence of jumps in returns.

Alternative models for the VIX are Brenner, Shu and Zhang [46], Huang and Zhang [99] and Zhang and Zhu [161], where the dynamics of the asset price is given by a diffusion process. Moreover, Broadie and Jain [51] consider the effect of jumps when the jump size is assumed to be normally distributed.

So far, we discuss the effect of the presence of jumps by concentrating on the underlying asset process, without any specification of the variance dynamics. Now, we analyze the contribution of both price and variance jumps, with the support of our theoretical model (2.19). Let $Y_t^{(S)}$ and $Y_t^{(V)}$ be two i.i.d. random variables taking values in $\mathbb{R}$, with common distribution $\gamma_{Y_t^{(S)}} = \gamma_{Y_t^{(V)}} = \gamma$ and let $\left(N_t^{(S)}\right)_{t\geq 0} \equiv N_t^{(S)}$ and $\left(N_t^{(V)}\right)_{t\geq 0} \equiv N_t^{(V)}$ be two Poisson processes of constant intensities $\lambda^{(S)}$ and $\lambda^{(V)}$, respectively, independent of $Y_t^{(S)}$ and $Y_t^{(V)}$.

Allowing for jumps both in returns and volatility, our dynamics (2.19) becomes

$$\begin{aligned} dS_t &= r\,dt + \sqrt{V_t}dW_t + Y_t^{(S)}dN_t^{(S)} - \lambda^{(S)}\gamma dt \\ dV_t &= \kappa(\omega - V_t)dt + \beta_t dZ_t + Y_t^{(V)}dN_t^{(V)} - \lambda^{(V)}\gamma_Y\,dt \end{aligned}\quad , \tag{2.28}$$

where r, κ and $\omega \in \mathbb{R}$, $(W_t)_{t\geq 0} \equiv W_t$ and $(Z_t)_{t\geq 0} \equiv Z_t$ are two correlated Wiener processes and $Y_t^{(S)}$ and $Y_t^{(V)}$ represent the jump sizes in returns and volatility, respectively. The terms $\lambda^{(S)}\gamma dt$ and $\lambda^{(V)}\gamma dt$ are used to center Poisson innovations, so that the compounded and compensated Poisson process

$$\tilde{J}_t^{(V)} \equiv \sum_{i=1}^{N_t^{(V)}} Y_i^{(V)} - \lambda^{(V)}\gamma t, \tag{2.29}$$

has zero mean. It is worth noting that the presence of jumps both in returns and volatility leaves Equation (2.21) unchanged.

Proof. Differentiating the stochastic differential equation

$$dV_t = \kappa(\omega - V_t)dt + \beta_t dZ_t + Y_t^{(V)}dN_t^{(V)} - \lambda^{(V)}\gamma dt,$$

we obtain

$$\begin{aligned} V_t &= \mathrm{e}^{-\kappa t}\left(V_0 + \int_0^t \kappa\omega \mathrm{e}^{\kappa} ds + \int_0^t \beta_t \mathrm{e}^{\kappa s} dZ_s\right) + \int_0^t Y_t^{(V)} dN_t^{(V)} - \int_0^t \lambda^{(V)}\gamma ds, \\ V_t &= V_0\mathrm{e}^{-\kappa t} + \omega(1 - \mathrm{e}^{-\kappa t}) + \mathrm{e}^{-\kappa t}\int_0^t \beta_t \mathrm{e}^{\kappa s} dW_s + \sum_{i=1}^{N_t^{(V)}} Y_i^{(V)} - \lambda^{(V)}\gamma t, \\ V_t &= V_0\mathrm{e}^{-\kappa t} + \omega(1 - \mathrm{e}^{-\kappa t}) + \mathrm{e}^{-\kappa t}\int_0^t \beta_t \mathrm{e}^{\kappa s} dW_s + \tilde{J}_t^{(V)}. \end{aligned}$$

Now, it is straightforward to compute the expectation of V_t, which is given by

$$\mathbf{E}^{\mathbb{Q}}\left[V_t \mid \mathfrak{F}_t\right] = V_0\mathrm{e}^{-\kappa t} + \omega(1 - \mathrm{e}^{-\kappa t}) + \mathbf{E}^{\mathbb{Q}}\left[\tilde{J}_t^{(V)} \mid \mathfrak{F}_t\right],$$

where $\tilde{J}_t^{(V)}$ is a compounded and compensated Poisson process, whose mean is equal to zero. Thus, we have

$$\mathbf{E}^{\mathbb{Q}}\left[V_t \mid \mathfrak{F}_t\right] = V_0\mathrm{e}^{-\kappa t} + \omega(1 - \mathrm{e}^{-\kappa t}),$$

and, as a consequence,

$$\mathbf{E}^{\mathbb{Q}}\left[\int_0^t V_s\,ds \,\middle|\, \mathfrak{F}_t\right] = \frac{V_0}{\kappa}(1 - \mathrm{e}^{-\kappa t}) + \omega t - \frac{\omega}{\kappa}(1 - \mathrm{e}^{-\kappa t}),$$

which is the desired result. □

2.2.2 Two-factor Model

Now, we consider a more general framework, where the return total variance is the sum of two independent factors, modeled by mean-reverting processes. In particular, under the risk-neutral probability measure $\mathbb{Q}$, our two-factor dynamics is given by

$$\begin{array}{l} dS_t = r\,dt + \sqrt{V_t^{(1)}}dW_t^{(1)} + \sqrt{V_t^{(2)}}dW_t^{(2)} \\ dV_t^{(1)} = \kappa_1(\omega_1 - V_t^{(1)})dt + \beta_t^{(1)}dZ_t^{(1)} \\ dV_t^{(2)} = \kappa_2(\omega_2 - V_t^{(2)})dt + \beta_t^{(2)}dZ_t^{(2)} \end{array}, \tag{2.30}$$

where $\left(W_t^{(i)}\right)_{t\geq 0} \equiv W_t^{(i)}$ and $\left(Z_t^{(i)}\right)_{t\geq 0} \equiv Z_t^{(i)}$, $i = 1, 2$, are Wiener processes with correlation defined as

$$d\left\langle W_t^{(1)}, W_t^{(2)}\right\rangle = d\left\langle Z_t^{(1)}, Z_t^{(2)}\right\rangle = 0 \quad \text{and} \quad d\left\langle W_t^{(i)}, Z_t^{(i)}\right\rangle = \rho_i\,dt, \qquad i = 1, 2.$$

Note that the return total variance is composed of the sum of the two diffusion processes, that is

$$\sigma_t^2 = V_t^{(1)} + V_t^{(2)}. \tag{2.31}$$

Hence, the VIX can be computed as stated in the following Proposition.

Proposition 2.4. *Under the general dynamics in Equation (2.30), the VIX is given by*

$$\begin{aligned} \mathbf{E}^{\mathbb{Q}}\left[\int_0^t \sigma_s^2\,ds \;\middle|\; \mathfrak{F}_t\right] = \frac{V_0^{(1)}}{\kappa_1}(1 - \mathrm{e}^{-\kappa_1 t}) + \omega_1 t - \frac{\omega_1}{\kappa_1}(1 - \mathrm{e}^{-\kappa_1 t}) + \\ + \frac{V_0^{(2)}}{\kappa_2}(1 - \mathrm{e}^{-\kappa_2 t}) + \omega_2 t - \frac{\omega_2}{\kappa_2}(1 - \mathrm{e}^{-\kappa_2 t}). \end{aligned} \tag{2.32}$$

Proof. Since the return total variance is defined as (2.31), it holds that

$$\begin{aligned} \mathbf{E}^{\mathbb{Q}}\left[\int_0^t \sigma_s^2\,ds \;\middle|\; \mathfrak{F}_t\right] &= \mathbf{E}^{\mathbb{Q}}\left[\int_0^t (V_s^{(1)} + V_s^{(2)})\,ds \;\middle|\; \mathfrak{F}_t\right] \\ &= \mathbf{E}^{\mathbb{Q}}\left[\int_0^t V_s^{(1)}\,ds \;\middle|\; \mathfrak{F}_t\right] + \mathbf{E}^{\mathbb{Q}}\left[\int_0^t V_s^{(2)}\,ds \;\middle|\; \mathfrak{F}_t\right]. \end{aligned}$$

From the dynamics described in Equation (2.30), we have

$$\mathbf{E}^{\mathbb{Q}}\left[\int_0^t V_s^{(i)}\,ds \;\middle|\; \mathfrak{F}_t\right] = \frac{V_0^{(i)}}{\kappa_i}(1 - \mathrm{e}^{-\kappa_i t}) + \omega_i t - \frac{\omega_i}{\kappa_i}(1 - \mathrm{e}^{-\kappa_i t}), \qquad i = 1, 2.$$

Hence, the sum of $\mathbf{E}^{\mathbb{Q}}\left[\int_0^t V_s^{(i)}\,ds \mid \mathfrak{F}_t\right]$ for $i = 1, 2$ fulfills Equation (2.32). □

Remark 2.1. *As shown in the previous section, the presence of compensated volatility jumps does not affect the expression of our theoretical model (2.21). It holds true also in the case of a two-factor model.*

Multifactor models have been extensively used in the option pricing literature, see among others Adrian and Rosenberg [1], Bates [38], Brigo and Mercurio [47], Christoffersen et al. [67], Taylor and Xu [153], Todorov and Tauchen [154] and Luo and Zhang [120]. While there is no explicit confidence to support a multifactor stock index option model, evidence from currency options shows that two-factor models can do better than one-factor models in explaining the term structure of implied volatilities over time. Nonetheless, one risk is that considering too many factors may overfit options data. Consequently, the appropriate number of factors should be judged not only by how good is the fit of option prices but by other criteria, such as whether using more factors modifies implicit conditional distributions or whether implicit factors appear excessively noisy over time.

2.3 Implementation

The new VIX is conceptually more interesting than its predecessor with regard to the economic interpretation, but, at the same time, the actual construction of the index is more complex. This is clear from Equations (2.15) and (2.16), where the model-free implied variance is defined as an integral of weighted option prices over an infinite range of strikes. Hence, some steps are necessary in order to minimize the implementation errors, which are due to the major complexity and the stochastic environment we work with. Nonetheless, as pointed out by Carr and Wu [58] and Jiang and Tian [107], the actual procedure adopted by the CBOE for the construction of the VIX does not seem to take these steps into account and, thus, it may lead to substantial biases in the computed index values.

Clearly, direct economic consequences may result from any existent bias in the VIX value. For instance, an underestimation [resp. overestimation] of the VIX leads to a corresponding underpricing [resp. overpricing] of the variance swaps. In other words, the bias in the VIX is translated into pricing errors of the variance swaps, consequently raising the possibility of arbitrage opportunities. In addition, a bias in the VIX may also produce a misestimation of the volatility risk premium, which measures the differences between the risk-neutral variance and the realized variance. In their work, Carr and Wu [58] highlight that a correctly estimated volatility risk premium is important for pricing and hedging derivative securities, written on underlying assets with stochastic volatility. If the VIX is used as a proxy for the risk-neutral variance, computation errors in the VIX are subsequently embedded in the volatility risk premium, which will lead to biased empirical results.

Furthermore, the general VIX formula seen in Equation (2.1) may contain several approximation errors, including truncation, discretization, observation, expansion and interpolation errors. Some of these errors, such as the expansion error, can be considered negligible, since they do not affect the robustness of the empirical results. On the other hand, it is important to understand their impact on the estimated variance. In the following of this section, we identify some of the inaccuracies present in the CBOE procedure to construct the new VIX and in its implementation of the model-free implied volatility.

The model-free implied volatility is defined by the integral of option prices over an infinite range of strikes, as shown in Equation (2.14). If option prices are available for all strikes, the required integral is straightforward to compute by means of numerical integration methods. However, only a finite number of strike prices are actually traded in the market place.

Let $[K_{\min}, K_{\max}]$ be the range of all strikes available for a given maturity. We require that $0 < K_{\min} < F_0 < K_{\max} < \infty$ to avoid trivial cases. We further assume that all strike prices in the interval are available. Let $I(K)$ be the integrand of Equation (2.14), that is

$$I(K) \equiv \frac{C^F(T,K) - \max(F_0 - K, 0)}{K^2}. \tag{2.33}$$

Truncation errors are present when we ignore the tails of the distribution and approximate the right-hand side of Equation (2.14) by the following integral

$$\int_0^\infty I(K)\,dK \approx \int_{K_{\min}}^{K_{\max}} I(K)\,dK. \tag{2.34}$$

Hence, the size of the truncation error is given by

$$\begin{aligned}\delta_{tr} &= \int_0^\infty I(K)\,dK - \int_{K_{\min}}^{K_{\max}} I(K)\,dK \\ &= \int_0^{K_{\min}} I(K)\,dK + \int_{K_{\max}}^\infty I(K)\,dK. \end{aligned} \tag{2.35}$$

Proposition 2.5. *Beyond the strikes range $[K_{\min}, K_{\max}]$, the right and left truncation errors have the following upper bounds:*

$$\int_{K_{\max}}^\infty I(K)\,dK \le \mathbf{E}^{\mathbb{Q}}\left[\left(\frac{F_T - K_{\max}}{K_{\max}}\right)^2 \,\middle|\, F_T > K_{\max}\right], \tag{2.36}$$

and

$$\int_0^{K_{\min}} I(K)\,dK \le \mathbf{E}^{\mathbb{Q}}\left[\left(\frac{F_T - K_{\min}}{K_{\min}}\right)^2 \,\middle|\, F_T < K_{\min}\right], \tag{2.37}$$

respectively.

Proof. We first consider the right truncation error (2.36). Applying integration by parts to the left-hand side of Equation (2.36), we have

$$\begin{aligned}\int_{K_{\max}}^\infty I(K)\,dK &= -\frac{C^F(T,K) - \max(F_0 - K, 0)}{K}\bigg|_{K_{\max}}^\infty + \int_{K_{\max}}^\infty \left(\frac{\partial C^F(T,K)}{\partial K} + \mathbf{1}_{F_0>K}\right)\frac{1}{K}dK \\ &= \frac{C^F(T,K_{\max}) - \max(F_0 - K_{\max}, 0)}{K_{\max}} + \int_{K_{\max}}^\infty \left(\frac{\partial C^F(T,K)}{\partial K} + \mathbf{1}_{F_0>K}\right) d\ln K,\end{aligned}$$

and, since

$$\frac{\partial C^F(T,K)}{\partial K} = \mathbf{E}^{\mathbb{Q}}\left[\frac{\partial \max(F_T - K, 0)}{\partial K}\right] = -\mathbf{E}^{\mathbb{Q}}\left[\mathbf{1}_{F_T>K}\right],$$

the previous expression simplifies to

$$\int_{K_{\max}}^{\infty} I(K)\,dK = \frac{C^F(T, K_{\max}) - \max(F_0 - K_{\max}, 0)}{K_{\max}} + \\ - \mathbf{E}^{\mathbb{Q}}\left[\max\left(\ln\frac{F_T}{K_{\max}}, 0\right)\right] + \max\left(\ln\frac{F_0}{K_{\max}}, 0\right).$$

With $K_{\max} > F_0$, we have

$$\int_{K_{\max}}^{\infty} I(K)\,dK = \int_{K_{\max}}^{\infty}\left[\frac{F_T - K_{\max}}{K_{\max}} - \ln\left(1 + \frac{F_T - K_{\max}}{K_{\max}}\right)\right]\phi(F_T)\,dF_T,$$

where $\phi(F_T)$ is the density of the forward price. By virtue of the properties of the Taylor series expansion for the logarithm function, we obtain

$$\int_{K_{\max}}^{\infty} I(K)\,dK \leq \int_{K_{\max}}^{\infty}\left(\frac{F_T - K_{\max}}{K_{\max}}\right)^2\phi(F_T)\,dF_T,$$

which is the upper bound for the right truncation error. The right-hand side of the last equation can be rewritten as

$$\int_{K_{\max}}^{\infty}\left(\frac{F_T - K_{\max}}{K_{\max}}\right)^2\phi(F_T)\,dF_T = \mathbf{E}^{\mathbb{Q}}\left[\left(\frac{F_T - K_{\max}}{K_{\max}}\right)^2 \,\middle|\, F_T > K_{\max}\right],$$

which provides Equation (2.36). The left truncation error (2.37) can be derived in a similar way. □

Both upper bounds are quite intuitive, as they reflect the local variations in the tails of the equity asset return distribution. Although these upper bounds are shown subsequently to be quite tight, they are not model-free. It is straightforward to derive model-free upper bounds based on available option prices. In fact, invoking the monotonicity and convexity of option prices, we have

$$\begin{aligned} \int_{K_{\max}}^{\infty} I(K)\,dK &\leq \frac{C^F(T, K_{\max})}{K_{\max}}, \\ \int_{K_{\min}}^{\infty} I(K)\,dK &\leq \frac{P^F(T, K_{\min})}{K_{\min}}\ln\frac{K_{\min}}{K^*} + \varepsilon(K^*), \end{aligned}$$

where $P^F(T, K)$ is the forward put option price and K^* is a sufficiently small positive number, such that the quantity $\varepsilon(K^*)$, given by

$$\varepsilon(K^*) = \int_0^{K^*} I(K)\,dK,$$

is negligible. However, the model-free upper bounds are not as tight as the ones in Equations (2.36) and (2.37).

In the analysis of Jiang and Tian [107], it is shown that the truncation error declines monotonically and diminishes as the truncation points, $K_{\min}$ and $K_{\max}$, move away from the spot price S_0. In particular, the left and right truncation errors are of similar magnitude and exhibit nearly identical convergence properties when the return distribution is symmetric. In contrast, the left truncation error is larger than the corresponding right truncation error when the return distribution is skewed to the left. In this case, the left truncation error converges at a slower rate than the right truncation error. With a fatter left tail, a larger range of strike prices is needed on the left of S_0 if we wish to have identical truncation errors from both sides. Nevertheless, the differences between left and right truncation errors are relatively small and convergence is rapid in all cases.

In general, the truncation error is negligible if the truncation points are more than two standard deviations from S_0. Furthermore, noting that the integrated variance is an increasing function of maturity, hence, a larger range of strike prices is needed to control the truncation errors for longer maturities.

A second source of approximation is represented by the *discretization error*, due to numerical integration. The definition of the model-free implied volatility requires a continuum of strikes. Since strike prices are listed for trading only in discrete increments, the integral in Equation (2.14) is approximated by a discrete sum, that is

$$\int_0^\infty \frac{C^F(T,K) - \max(F_0 - K, 0)}{K^2} dK \approx \sum_i \frac{\Delta K_i}{K_i^2} \mathrm{e}^{rT} Q(T, K_i). \tag{2.38}$$

Hence, the discretization error essentially derives from using a numerical integration method to approximate an integral. The size of this error is given by

$$\delta_{di} = \sum_i \frac{\Delta K_i}{K^2} \mathrm{e}^{rT} Q\left(T, K_i\right) - \int_{K_{\min}}^{K_{\max}} \frac{C^F(T,K) - \max(F_0 - K, 0)}{K^2} dK. \tag{2.39}$$

Although the size and direction of the discretization error may depend on the shape of the implied volatility function[8], the practice of using the mid-price of call and put options tends to induce a positive discretization error. Note that the discretization error can be easily minimized, if we consider an extremely thin partition of the available strike prices.

A third approximation error is the *expansion error*, which is related to the last term in Equation (2.1), that is

$$-\frac{1}{T}\left(\frac{F_0}{K_0} - 1\right)^2.$$

This term derives from the Taylor series expansion of the logarithm function preceding the integrals in Equation (2.15), given by

$$\frac{2}{T}\left[rT - \left(\frac{F_0}{K_0} - 1\right) - \log\left(\frac{K_0}{S_0}\right)\right], \tag{2.40}$$

[8] More precisely, the direction and magnitude of the discretization error depend on the shape of the integrand function $\mathrm{e}^{rT} Q\left(T, K\right)/K^2$.

which can be rewritten as

$$\frac{2}{T}\left[rT-\left(\frac{F_0}{K_0}-1\right)+\log\left(\frac{F_0\mathrm{e}^{-rT}}{K_0}\right)\right]=\frac{2}{T}\left[\log\left(\frac{F_0}{K_0}\right)-\left(\frac{F_0}{K_0}-1\right)\right]. \tag{2.41}$$

Applying the Taylor series expansion of the logarithm function and ignoring terms higher than the second order, we have

$$\log\left(\frac{F_0}{K_0}\right)\simeq\left(\frac{F_0}{K_0}-1\right)-\frac{1}{2}\left(\frac{F_0}{K_0}-1\right)^2.$$

Thus, substituting the previous expression into Equation (2.41), the following approximation holds

$$\frac{2}{T}\left[rT-\left(\frac{F_0}{K_0}-1\right)-\log\left(\frac{K_0}{S_0}\right)\right]\simeq-\frac{1}{T}\left(\frac{F_0}{K_0}-1\right)^2.$$

Clearly, the size of this error is in the order of $\left(\frac{F_0}{K_0}-1\right)^3$ and it is in general negligible, as long as K_0 is chosen to be the strike closest to F_0, with $K_0 \leq F_0$.

Another different approximation error, known as *interpolation error*, is due to the interpolation of maturities used in the CBOE procedure. The VIX is based on the model-free implied volatility with a fixed 30-day maturity, but in general there are no options that expire exactly in 30 calendar days. The solution is to consider two maturities, T_1 and T_2, which are closest to the required 30-day maturity. The volatility measure is first computed for these two selected maturities. Then, the model-free implied volatility with the 30-day maturity is linearly interpolated between the corresponding volatilities at the two selected maturities. Nevertheless, as documented in the financial literature (see Taylor and Xu [153]), the implied volatility term structure in neither linear nor monotonic in option maturity. Hence, linear interpolation leads to a positive [resp. negative] error if the model-free implied volatility is a convex [resp. concave] function of maturity.

The final approximation error is the *observation error*, whose presence is linked to the fact that option prices, quoted in terms of the Black-Scholes [42] implied volatility, are observed with error, that is

$$\hat{C}(T,K)=C(T,K)+\epsilon. \tag{2.42}$$

If a limited set of options is included in the analysis, then, inference procedure for the model parameters is only feasible under strict parametric assumptions regarding the error distribution. This is problematic as we have little evidence pertaining to the nature of these price errors. In contrast, a large cross section allows us to average out the errors and remain fully nonparametric regarding their distribution. However, this error diversification only works if we can ensure that the effect of the option price errors vanishes in a suitable manner.

Let Ψ be the difference between the true value of the VIX and its measure obtained from the data, that is

$$\Psi = \frac{2}{T}\int_0^{\infty} \frac{C^F(T,K) - \max(F_0 - K, 0)}{K^2} dK - \frac{2}{T}\int_{K_{\min}}^{K_{\max}} \frac{\hat{C}^F(T,K) - \max(F_0 - K, 0)}{K^2} dK. \tag{2.43}$$

Now, we denote by $I(K)$ and $\hat{I}(K)$ the integrand of the first and second integrals, respectively. Thus, adding and subtracting the same term from Equation (2.43), we have

$$\Psi = \frac{2}{T}\underbrace{\left[\int_0^{\infty} I\left(K\right) dK - \int_{K_{\min}}^{K_{\max}} I\left(K\right) dK\right]}_{\textit{Truncation Error}} + \frac{2}{T}\underbrace{\left[\int_{K_{\min}}^{K_{\max}} I\left(K\right) dK - \int_0^{\infty} \hat{I}\left(K\right) dK\right]}_{\textit{Observation Error}}.$$

From the previous expression, we can see that the two error components, that substantially affect the VIX estimation, are the truncation and the observation errors.

Chapter 3

Spot Volatility Estimation from Option Prices

The dynamics of the *spot volatility* is extremely important for modeling financial series. However, the spot volatility itself is a latent variable, which is not directly observable. Therefore, during the last decades, there has been considerable attention on which was the best method, in terms of efficiency and consistency, to estimate a latent variable, such as the spot volatility. Since option prices reflect expectations and rumors of economic agents on future movements of the underlying asset, the implied volatility is widely believed to contain much more information than the historical volatility. As a consequence, it is considered a more efficient forecast for future realized volatility. This is supported by a wide branch of the financial literature, such as Christensen, Hansen and Prabhala [65], Christensen and Prabhala [66], Ederington and Guan [81], Fleming [88], Jiang and Tian [107], Pong et al. [136] and Yu, Lui and Wang [160], among others. On the contrary, Canina and Figlewski [53] find that the implied volatility of the S&P 100 index options is a poor forecast of the subsequent realized volatility and does not incorporate the information contained in historical volatility. Following this literature, Luo and Zhang [120] investigate the information content of the VIX term structure and they find that the VIX contains more information than historical volatility, consistently with previous studies.

The goal of this chapter is to propose a new estimation method of the spot volatility from a panel of options available in the market, which relies on the concept of model-free implied volatility previously introduced. This technique exploits the *observed* VIX term structure to make inference on the *unobserved* spot volatility. This can be achieved by simply using the information content of a complete set of European options daily quoted in the market, with a significant range of strikes and maturities, which is a natural assumption for active and liquid derivatives markets. We use this panel of options to obtain a daily volatility term structure, whose representative shape can be seen in Figure 3.1.

Consider the semi-nonparametric theoretical model in (2.21), which conveys the VIX as a functional form of the option maturity, the spot volatility and the mean-reverting parameters. We fit this model on the daily VIX term structure and we obtain the spot volatility estimation

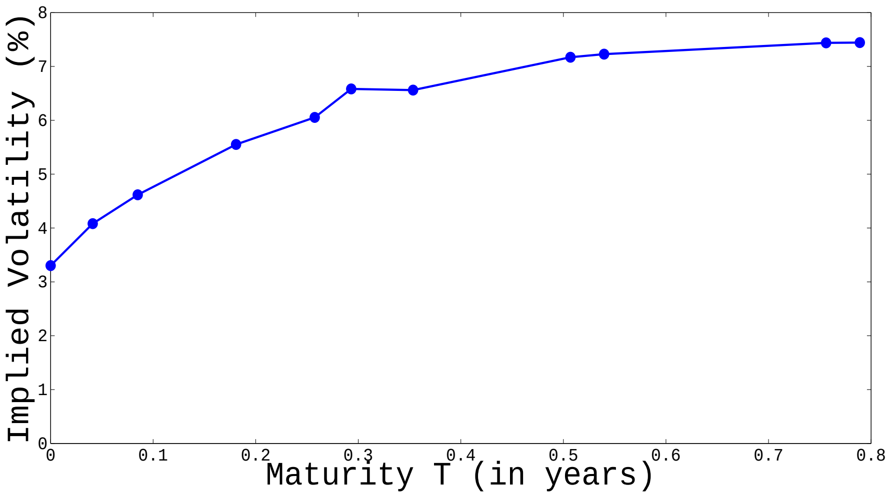

Figure 3.1: Representative Daily VIX Term Structure

as solution of a non-linear optimization problem. This allows us to relax the hypothesis of parametric estimation, thus obtaining a semi-nonparametric estimation procedure, which can be applied to a wide variety of stochastic volatility models.

Proposition 3.1. *Under the general dynamics described in Equation (2.19), the annualized VIX at the present time $t = 0$, referred to maturity T, is given by*

$$VIX = \frac{1}{T}\left[\frac{V_0}{\kappa}\left(1 - e^{-\kappa T}\right) + \omega T - \frac{\omega}{\kappa}\left(1 - e^{-\kappa T}\right)\right]. \tag{3.1}$$

Proof. Since the Markov process V_t is given by Equation (2.20), we have

$$\mathbf{E}^{\mathbb{Q}}\left[V_t \mid \mathfrak{F}_t\right] = V_0 e^{-\kappa t} + \omega(1 - e^{-\kappa t}).$$

By definition, the VIX equals the expected risk-neutral integrated variance over the interval $[0, T]$. Thus, we have

$$\begin{aligned} VIX &= \mathbf{E}^{\mathbb{Q}}\left[\frac{1}{T}\int_0^T V_s\, ds\right] = \frac{1}{T}\int_0^T \mathbf{E}^{\mathbb{Q}}\left[V_s\right]\, ds \\ &= \frac{1}{T}\int_0^T \left[V_0 e^{-\kappa s} + \omega\left(1 - e^{-\kappa s}\right)\right]\, ds \\ &= \frac{1}{T}\left[\frac{V_0}{\kappa}\left(1 - e^{-\kappa T}\right) + \omega T - \frac{\omega}{\kappa}\left(1 - e^{-\kappa T}\right)\right]. \end{aligned}$$ □

Remark 3.1. *Our semi-nonparametric theoretical model (3.1) represents a formal expression of the VIX as a function of maturity T, spot variance V_0 and mean-reverting parameters κ and ω. Thus, given the vector of parameters $\Theta = (\kappa, \omega, V_0)$, we have $VIX = f(T, \Theta)$.*

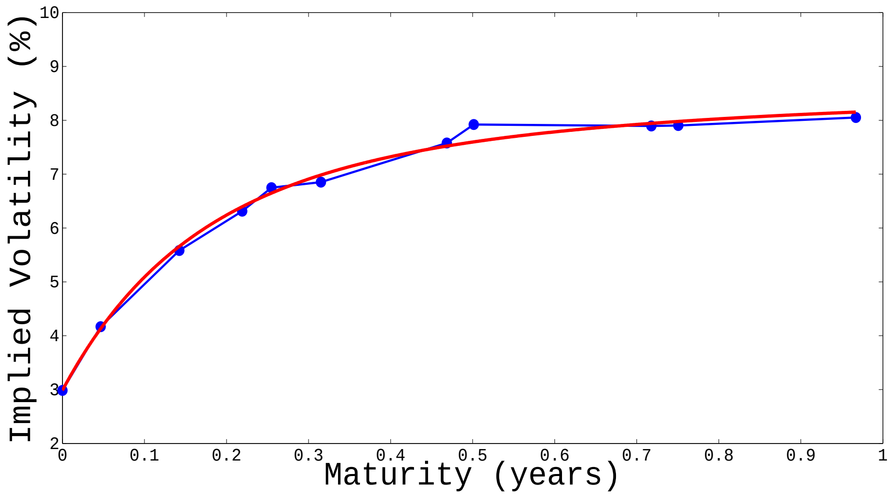

Figure 3.2: Example of Model Fit

Hence, we fit this closed-form expression to obtain an estimation $\hat{\Theta}$ of Θ, which is given by

$$\hat{\Theta} = \arg\min_{\Theta} \sum_{i=1}^{N} \left[VIX_{T_i}^{MKT} - f\left(T_i, \Theta\right) \right]^2, \tag{3.2}$$

where N is the number of available maturities, $VIX_{T_i}^{MKT}$ is the market value of the VIX with maturity T_i computed as in (2.14) and $f\left(T_i, \Theta\right)$ is the corresponding theoretical value, given by Equation (3.1). In other words, we have to solve a non-linear optimization problem of the form

$$\min_{\Theta} \sum_{i=1}^{N} \left\{ \frac{2}{T_i} \int_0^{\infty} \frac{C^F(t,K) - \max\left(F_0 - K, 0\right)}{K^2} \, dK + \right. \\ \left. - \frac{1}{T_i} \left[\frac{V_0}{\kappa} \left(1 - e^{-\kappa t}\right) + \omega t - \frac{\omega}{\kappa} \left(1 - e^{-\kappa t}\right) \right] \right\}. \tag{3.3}$$

The optimization algorithm yields the value of Θ, which guarantees the theoretical function $f\left(T, \Theta\right)$ as close as possible to the observed values $VIX_{T_i}^{MKT}$. Clearly, the parameter V_0 represents the *spot variance* estimation. Figure 3.2 shows an example of the theoretical model fit with respect to the market VIX values.

Some advantages of this method can be outlined. Firstly, it is a very intuitive procedure, which does not require heavy computational issues. Second, it allows us to relax parametric assumptions. Indeed, our theoretical model is completely semi-nonparametric, since it does not depend on β_t, and it allows us to obtain a valid estimate of the spot volatility parameter only by means of options market data. This approach should give more accurate estimates than those obtained with high-frequency data (see Andersen, Fusari and Todorov [20]). Daily market data are available in real time free from biases, in order to avoid any arbitrage opportunity. On the contrary, estimation methods based on high-frequency data often use larger data sets, composed by historical volatilities. As a consequence, even though they can reduce the variance of an estimator, they surely present a much higher bias, since historical volatilities are well away from the instantaneous volatility measure. Hence, this represents a clear theoretical advantage of an options based methodology. Moreover, the estimation procedure described above can be applied to a wide class of stochastic volatility models. For instance, the celebrate Heston [97] model can be obtained with $\beta_t = \sigma\sqrt{V_t}$, while the Luo and Zhang [120] model is a special case with stochastic long-run mean level θ.

3.1 Testing Spot Volatility Estimation: GMM

The nonlinear optimization problem (3.3) is an intuitive estimation method easy to implement. Nevertheless, it does not take into account a way to model the uncertainty, induced by the fact that option prices can be observed with error or that a continuum of strikes is generally not available, but we can only manage discrete realizations. These sources of uncertainty can entail the occurrence of possible estimation errors, which must be inferred. Hence, now we analyze a slightly different approach with respect to the previous one, based on the so-called *Generalized Method of Moments* (GMM).

Given a probability space $(\Omega, \mathfrak{E}, \mathbb{P}) \equiv \Omega$ endowed with a filtration $(\mathfrak{F}_t)_{t\geq 0}$, let $(S_t)_{t\geq 0}$ be the spot price of a financial equity asset. We consider a panel of options market data, with N maturities $(T_i)_{i=1,\dots,N} \equiv T \in \mathbb{R}_+^N$, J strikes $(K_j)_{j=1,\dots,J} \equiv K \in \mathbb{R}_+^J$ for each maturity and option prices $C(T, K) \in \mathbb{R}_+^{N\times J}$. Let us define, for every $i \in [1, N]$:

$$Y_i^* = \frac{2}{T_i}\int_0^\infty \frac{C(T_i, K) - \max(S_0 - K, 0)}{K^2} dK, \tag{3.4}$$

and

$$Y_i = \frac{2}{T_i}\sum_{j=1}^{J} \frac{\Delta K_j}{K_j^2} C(T_i, K_j), \tag{3.5}$$

such that

$$Y_i = Y_i^* + \varepsilon_i. \tag{3.6}$$

Furthermore, we assume that

1. $\mathbf{E}[\varepsilon_i|\mathfrak{F}_t] = 0$,
2. $\mathbf{E}[\varepsilon_i\varepsilon_i^{\intercal}|\mathfrak{F}_t] = \sigma_\varepsilon^2 > 0$,
3. $\mathbf{E}\left[\varepsilon_i\varepsilon_l^{\intercal}|\mathfrak{F}_t\right] = 0$, whenever $i \neq l$, with $i, l \in [1, N]$,
4. $\mathbf{E}\left[|\varepsilon_i|^4|\mathfrak{F}_t\right] < \infty$, almost surely.

In other words, $(Y_i^*)_{i=1,\dots,N} \equiv Y^*$ is the real value of the expected integrated risk-neutral variance, as seen in Equation (2.13), while $(Y_i)_{i=1,\dots,N} \equiv Y$ is its empirical approximation, in light of the fact that we can observe only discrete realizations of the strike prices K and then we cannot compute the exact integral from 0 to ∞ in (3.4). However, the discretization of Y^* using (3.5) implies that the observed expected risk-neutral integrated variance is equal to its real value plus a computational error, as shown by Equation (3.6). Note that (3.5) is the general formula (2.1) used by the CBOE to compute the VIX, where we consider a free-risk interest rate equal to zero and we omit the last term $\frac{1}{T}\left(\frac{F}{K_0}-1\right)^2$, which is negligible for the upcoming analysis.

Given a compact parameter space $\Theta \subseteq \mathbb{R}^3$, let be $\theta_0 \in \Theta$ the true parameter value. Consider the mean-reverting instantaneous variance process $(V_t)_{t\geq 0} \equiv V_t$ in the general dynamics (2.19), which we report here for sake of clarity:

$$dV_t = \kappa(\omega - V_t)dt + \beta_t dZ_t. \tag{3.7}$$

Then, we can write N unconditional moment restrictions $g\left(Y^*, \theta_0\right)$, that is

$$\begin{aligned} g\left(Y^*, \theta_0\right) &= \begin{bmatrix} g\left(Y_1^*, \theta_0\right) \\ \vdots \\ g\left(Y_N^*, \theta_0\right) \end{bmatrix} \\ &= \begin{bmatrix} \frac{2}{T_1}\int_0^\infty \frac{C(T_1,K)-\max(S_0-K,0)}{K^2}\,dK - \frac{1}{T_1}\int_0^{T_1} V_s\,ds \\ \vdots \\ \frac{2}{T_N}\int_0^\infty \frac{C(T_N,K)-\max(S_0-K,0)}{K^2}\,dK - \frac{1}{T_N}\int_0^{T_N} V_s\,ds \end{bmatrix}, \end{aligned} \tag{3.8}$$

such that

$$\mathbf{E}^{\mathbb{Q}}\left[g\left(Y^*, \theta_0\right)\right] = 0. \tag{3.9}$$

Hence, our theoretical model (2.21) implies that

$$\begin{aligned} g\left(Y, \theta\right) &= \begin{bmatrix} g\left(Y_1, \theta\right) \\ \vdots \\ g\left(Y_N, \theta\right) \end{bmatrix} \\ &= \begin{bmatrix} \frac{2}{T_1}\sum_{j=1}^J \frac{\Delta K_j}{K_j^2} C\left(T_1, K_j\right) - \frac{1}{T_1}\left[\frac{V_0}{\kappa}\left(1-e^{-\kappa T_1}\right) + \omega T_1 - \frac{\omega}{\kappa}\left(1-e^{-\kappa T_1}\right)\right] \\ \vdots \\ \frac{2}{T_N}\sum_{j=1}^J \frac{\Delta K_j}{K_j^2} C\left(T_N, K_j\right) - \frac{1}{T_N}\left[\frac{V_0}{\kappa}\left(1-e^{-\kappa T_N}\right) + \omega T_N - \frac{\omega}{\kappa}\left(1-e^{-\kappa T_N}\right)\right] \end{bmatrix}, \end{aligned} \tag{3.10}$$

for admissible $\theta = [\kappa, \omega, V_0] \in \Theta$. Denoting by

$$b_i = \frac{1-e^{-\kappa T_i}}{\kappa T_i}, \qquad i \in [1, N],$$

we can rewrite the last equation in a more compact way, that is

$$g\left(Y, \theta\right) = \begin{bmatrix} \frac{2}{T_1}\sum_{j=1}^J \frac{\Delta K_j}{K_j^2} C\left(T_1, K_j\right) - \left[b_1 V_0 + \left(1-b_1\right)\omega\right] \\ \vdots \\ \frac{2}{T_N}\sum_{j=1}^J \frac{\Delta K_j}{K_j^2} C\left(T_N, K_j\right) - \left[b_N V_0 + \left(1-b_N\right)\omega\right] \end{bmatrix}. \tag{3.11}$$

Intuitively, if our theoretical model is right, then Y and the vector $[bV_0 + (1-b)\,\omega]$ should be close to each other on average and, as a consequence, the parameter estimation errors should tend to zero. In particular, this is true when $J \to \infty$, that is, when the number of strikes available for each maturity approaches infinity, since Y approximates the true Y^*. Before testing whether this is the case, a value for θ has to be chosen. It seems natural to select the value of θ that would give the model the best chance. The GMM estimates θ by minimizing the empirical function $g(Y,\theta)$. Hence, given an estimate for θ, say $\hat{\theta}_{GMM}$, we can evaluate the size of the estimation error.

Denote by $g_J(\theta) \equiv g(Y,\theta)$, in order to stress the fact that a good approximation of the integral in (3.4) mostly depends on the number of strikes J. Now, let W_J be a positive definite weighting matrix, such that $W_J \xrightarrow{p} W$, where W is positive definite. Then, the *GMM estimator* $\hat{\theta}_{GMM}$ is the choice of θ that minimizes the scalar quadratic objective function $Q_J(\theta) = g_J(\theta)^\intercal W_J g_J(\theta)$, that is

$$\hat{\theta}_{GMM} = \arg\min_{\theta\in\Theta} g_J(\theta)^\intercal W_J g_J(\theta)\,. \tag{3.12}$$

Under regularity conditions,

$$g_J(\theta) \xrightarrow{p} 0. \tag{3.13}$$

Proof. For a given maturity T_i, we have

$$\begin{aligned}\mathbf{E}\left[g_J(\theta) - g(Y_i^*,\theta_0)\right] &= \frac{2}{T_i}\sum_{j=1}^{J}\frac{\Delta K_j}{K_j^2}C(T_i,K_j) - \frac{2}{T_i}\int_0^\infty \frac{C(T_i,K) - \max(S_0-K,0)}{K^2}dK \\ &= Y_i - Y_i^* = \varepsilon_i.\end{aligned}$$

As $J \to \infty$,

$$\lim_{J\to\infty}\varepsilon_i = 0. \qquad \square$$

3.1.1 Consistency and Asymptotic Normality of the GMM Estimator

Assume N moment conditions and d parameters, with $N \geq d$. In addition, assume there is no dependence in the data. By virtue of the Taylor's expansion, we have

$$\frac{\partial Q_J\left(\hat{\theta}_{GMM}\right)}{\partial\theta} - \frac{\partial Q_J(\theta_0)}{\partial\theta} = \frac{\partial^2 Q_J(\theta_0)}{\partial\theta\partial\theta^\intercal}\left(\hat{\theta}_{GMM} - \theta_0\right).$$

Note that $\frac{\partial Q_J(\hat{\theta}_{GMM})}{\partial\theta} = 0$, since $\hat{\theta}_{GMM}$ is the solution of the problem (3.12) and then it minimizes the objective function $Q_J(\theta)$. Hence, it holds

$$\hat{\theta}_{GMM} - \theta_0 = -\left(\frac{\partial^2 Q_J(\theta_0)}{\partial\theta\partial\theta^\intercal}\right)^{-1}\frac{\partial Q_J(\theta_0)}{\partial\theta},$$

or

$$\sqrt{J}\left(\hat{\theta}_{GMM}-\theta_0\right)=-\underbrace{\left(\frac{\partial^2 Q_J(\theta_0)}{\partial\theta\partial\theta^{\intercal}}\right)^{-1}}_{B}\underbrace{\left(\sqrt{J}\frac{\partial Q_J(\theta_0)}{\partial\theta}\right)}_{A}. \tag{3.14}$$

Let us focus on A term first. We have

$$\sqrt{J}\frac{\partial Q_J(\theta_0)}{\partial\theta}=\sqrt{J}2\left(\frac{\partial g_J(\theta_0)}{\partial\theta^{\intercal}}\right)^{\intercal}W_J g_J(\theta_0). \tag{3.15}$$

By virtue of the *strong law of large numbers*, we can state that

$$\frac{\partial g_J(\theta_0)}{\partial\theta^{\intercal}}\overset{a.s.}{\to}\mathbf{E}\left[\frac{\partial g(Y^*,\theta_0)}{\partial\theta^{\intercal}}\right]. \tag{3.16}$$

Proof. Computing the derivatives of (3.8) and (3.10) with respect to parameters, we obtain

$$\begin{aligned}\frac{\partial g(Y^*,\theta_0)}{\partial\theta^{\intercal}}&=\begin{bmatrix}\frac{\partial g(Y^*,\theta_0)}{\partial\kappa} & \frac{\partial g(Y^*,\theta_0)}{\partial\omega} & \frac{\partial g(Y^*,\theta_0)}{\partial V_0}\end{bmatrix}\\
&=\begin{bmatrix}\frac{\partial}{\partial\kappa}\left(-\frac{1}{T_1}\int_0^{T_1}V_s ds\right) & \frac{\partial}{\partial\omega}\left(-\frac{1}{T_1}\int_0^{T_1}V_s ds\right) & \frac{\partial}{\partial V_0}\left(-\frac{1}{T_1}\int_0^{T_1}V_s ds\right)\\ \vdots & \vdots & \vdots\\ \frac{\partial}{\partial\kappa}\left(-\frac{1}{T_N}\int_0^{T_N}V_s ds\right) & \frac{\partial}{\partial\omega}\left(-\frac{1}{T_N}\int_0^{T_N}V_s ds\right) & \frac{\partial}{\partial V_0}\left(-\frac{1}{T_N}\int_0^{T_N}V_s ds\right)\end{bmatrix}\\
&=\begin{bmatrix}-(V_0-\omega)\frac{e^{-\kappa T_1}-b_1}{\kappa}-\frac{1}{T_1}\frac{\partial}{\partial\kappa}\left(\int_0^{T_1}e^{-\kappa s}\int_0^s\beta_u e^{\kappa u}dW_u ds\right) & b_1-1 & -b_1\\ \vdots & \vdots & \vdots\\ -(V_0-\omega)\frac{e^{-\kappa T_N}-b_N}{\kappa}-\frac{1}{T_N}\frac{\partial}{\partial\kappa}\left(\int_0^{T_N}e^{-\kappa s}\int_0^s\beta_u e^{\kappa u}dW_u ds\right) & b_N-1 & -b_N\end{bmatrix},\end{aligned}$$

and

$$\begin{aligned}\frac{\partial g_J(\theta_0)}{\partial\theta^{\intercal}}&=\begin{bmatrix}\frac{\partial g_J(\theta_0)}{\partial\kappa} & \frac{\partial g_J(\theta_0)}{\partial\omega} & \frac{\partial g_J(\theta_0)}{\partial V_0}\end{bmatrix}\\
&=\begin{bmatrix}-(V_0-\omega)\frac{e^{-\kappa T_1}-b_1}{\kappa} & b_1-1 & -b_1\\ \vdots & \vdots & \vdots\\ -(V_0-\omega)\frac{e^{-\kappa T_N}-b_N}{\kappa} & b_N-1 & -b_N\end{bmatrix}.\end{aligned}$$

Details 1. *Consider a general maturity T, where we omit the subscript i for sake of simplicity.*

We can compute the following derivatives:

$$\begin{aligned}
\frac{\partial}{\partial\kappa}\left(-\frac{1}{T}\int_0^T V_s ds\right) &= -\frac{1}{T}\frac{\partial}{\partial\kappa}\int_0^T V_s ds\\
&= -\frac{1}{T}\frac{\partial}{\partial\kappa}\int_0^T\left(V_0e^{-\kappa s}+\omega-\omega e^{-\kappa s}+e^{-\kappa s}\int_0^s \beta_u e^{\kappa u}dW_u\right)ds\\
&= -\frac{1}{T}\frac{\partial}{\partial\kappa}\left[(V_0-\omega)\frac{1-e^{-\kappa T}}{\kappa}+\omega T\right]+\\
&\qquad -\frac{1}{T}\frac{\partial}{\partial\kappa}\underbrace{\left(\int_0^T e^{-\kappa s}\int_0^s \beta_u e^{\kappa u}dW_u ds\right)}_{\text{it depends on } \kappa}\\
&= -\frac{1}{T}(V_0-\omega)\frac{T\kappa e^{-\kappa T}-\left(1-e^{-\kappa T}\right)}{\kappa^2}+\\
&\qquad -\frac{1}{T}\frac{\partial}{\partial\kappa}\left(\int_0^T e^{-\kappa s}\int_0^s \beta_u e^{\kappa u}dW_u ds\right)\\
&= -(V_0-\omega)\frac{e^{-\kappa T}-b}{\kappa}-\frac{1}{T}\frac{\partial}{\partial\kappa}\left(\int_0^T e^{-\kappa s}\int_0^s \beta_u e^{\kappa u}dW_u ds\right),
\end{aligned}$$

then

$$\begin{aligned}
\frac{\partial}{\partial\omega}\left(-\frac{1}{T}\int_0^T V_s ds\right) &= -\frac{1}{T}\frac{\partial}{\partial\omega}\int_0^T V_s ds\\
&= -\frac{1}{T}\frac{\partial}{\partial\omega}\left[(V_0-\omega)\frac{1-e^{-\kappa T}}{\kappa}+\omega T\right]+\\
&\qquad -\frac{1}{T}\frac{\partial}{\partial\omega}\underbrace{\left(\int_0^T e^{-\kappa s}\int_0^s \beta_u e^{\kappa u}dW_u ds\right)}_{\text{it does not depend on } \omega}\\
&= -\frac{1}{T}\left(-\frac{1-e^{-\kappa T}}{\kappa}+T\right)=b-1,
\end{aligned}$$

and finally

$$\begin{aligned}
\frac{\partial}{\partial V_0}\left(-\frac{1}{T}\int_0^T V_s ds\right) &= -\frac{1}{T}\frac{\partial}{\partial V_0}\int_0^T V_s ds\\
&= -\frac{1}{T}\frac{\partial}{\partial V_0}\left[(V_0-\omega)\frac{1-e^{-\kappa T}}{\kappa}+\omega T\right]+\\
&\qquad -\frac{1}{T}\frac{\partial}{\partial V_0}\underbrace{\left(\int_0^T e^{-\kappa s}\int_0^s \beta_u e^{\kappa u}dW_u ds\right)}_{\text{it does not depends on } V_0}\\
&= -\frac{1-e^{-\kappa T}}{T\kappa}=-b.
\end{aligned}$$

Since

$$\mathbf{E}\left[\frac{\partial}{\partial\kappa}\left(\int_0^T e^{-\kappa s}\int_0^s \beta_u e^{\kappa u} dW_u ds\right)\right] = \frac{\partial}{\partial\kappa}\left(\int_0^T e^{-\kappa s}\mathbf{E}\left[\int_0^s \beta_u e^{\kappa u} dW_u\right] ds\right) = 0,$$

the expected value of $\frac{\partial g(Y^*,\theta_0)}{\partial\theta^\intercal}$ is completely deterministic and exactly equal to $\frac{\partial g_J(\theta_0)}{\partial\theta^\intercal}$. Thus, we have

$$\frac{\partial g_J(\theta_0)}{\partial\theta^\intercal} \xrightarrow{a.s} \mathbf{E}\left[\frac{\partial g(Y^*,\theta_0)}{\partial\theta^\intercal}\right],$$

which ends the proof. □

Moreover, from an application of the *Central Limit Theorem* (CLT), we have

$$\sqrt{J} g_J(\theta_0) \xrightarrow{d} \mathcal{N}(0,\Phi_0), \tag{3.17}$$

where

$$\Phi_0 = \mathbf{E}\left[g(Y^*,\theta_0)\, g(Y^*,\theta_0)^\intercal\right] + \sigma_\varepsilon^2.$$

Proof. The expected value of $g_J(\theta_0)$ is equal to

$$\begin{aligned}\mathbf{E}[g_J(\theta_0)] &= \mathbf{E}[Y_i - (b_i V_0 + (1-b_i)\,\omega)] \\ &= \mathbf{E}[Y_i^* + \varepsilon_i - (b_i V_0 + (1-b_i)\,\omega)] \\ &= \mathbf{E}[Y_i^* - (b_i V_0 + (1-b_i)\,\omega)] + \mathbf{E}[\varepsilon_i],\end{aligned}$$

with $i \in [1,N]$. By assumption 1, $\mathbf{E}[\varepsilon_i] = 0$, then we have

$$\begin{aligned}\mathbf{E}[g_J(\theta_0)] &= \mathbf{E}[Y_i^* - (b_i V_0 + (1-b_i)\,\omega)] \\ &= \mathbf{E}[g(Y_i^*,\theta_0)] = 0,\end{aligned}$$

from (3.9).

The variance of $g_J(\theta_0)$ is

$$\begin{aligned}\operatorname{var}(g_J(\theta_0)) &= \mathbf{E}[g_J(\theta_0)\, g_J(\theta_0)^\intercal] \\ &= \mathbf{E}[(Y_i - (b_i V_0 + (1-b_i)\,\omega))\,(Y_i - (b_i V_0 + (1-b_i)\,\omega))^\intercal] \\ &= \mathbf{E}[(Y_i^* + \varepsilon_i - (b_i V_0 + (1-b_i)\,\omega))\,(Y_i^* + \varepsilon_i - (b_i V_0 + (1-b_i)\,\omega))^\intercal] \\ &= \mathbf{E}[Y_i^*\,(Y_i^*)^\intercal] + \mathbf{E}[Y_i^*\varepsilon_i^\intercal] - \mathbf{E}[Y_i^*\,(b_i V_0 + (1-b_i)\,\omega)^\intercal] + \\ &\quad + \mathbf{E}[\varepsilon_i\,(Y_i^*)^\intercal] + \mathbf{E}[\varepsilon_i\varepsilon_i^\intercal] - \mathbf{E}[\varepsilon_i\,(b_i V_0 + (1-b_i)\,\omega)^\intercal] + \\ &\quad - \mathbf{E}[(b_i V_0 + (1-b_i)\,\omega)\,(Y_i^*)^\intercal] - \mathbf{E}[(b_i V_0 + (1-b_i)\,\omega)\,\varepsilon_i^\intercal] + \\ &\quad + \mathbf{E}[(b_i V_0 + (1-b_i)\,\omega)\,(b_i V_0 + (1-b_i)\,\omega)^\intercal] \\ &= \mathbf{E}[Y_i^*\,(Y_i^*)^\intercal] - \mathbf{E}[Y_i^*\,(b_i V_0 + (1-b_i)\,\omega)^\intercal] + \sigma_\varepsilon^2 + \\ &\quad - \mathbf{E}[(b_i V_0 + (1-b_i)\,\omega)\,(Y_i^*)^\intercal] + \mathbf{E}[(b_i V_0 + (1-b_i)\,\omega)\,(b_i V_0 + (1-b_i)\,\omega)^\intercal],\end{aligned}$$

on account of assumptions 1-2. Hence,

$$\begin{aligned}\operatorname{var}\left(g_J\left(\theta_0\right)\right) &= \sigma_\varepsilon^2 + \mathbf{E}\left[Y_i^*\left(Y_i^* - \left(b_i V_0 + \left(1-b_i\right)\omega\right)\right)^\intercal\right] + \\ &\quad - \mathbf{E}\left[\left(b_i V_0 + \left(1-b_i\right)\omega\right)\left(Y_i^* - \left(b_i V_0 + \left(1-b_i\right)\omega\right)\right)^\intercal\right] \\ &= \sigma_\varepsilon^2 + \mathbf{E}\left[\left(Y_i^* - \left(b_i V_0 + \left(1-b_i\right)\omega\right)\right)\left(Y_i^* - \left(b_i V_0 + \left(1-b_i\right)\omega\right)\right)^\intercal\right] \\ &= \sigma_\varepsilon^2 + \mathbf{E}\left[g\left(Y_i^*, \theta_0\right) g\left(Y_i^*, \theta_0\right)^\intercal\right],\end{aligned}$$

which is the desired result. □

Remark 3.2. *We have*

$$\begin{aligned}\sqrt{J}\frac{\partial Q_J\left(\theta_0\right)}{\partial\theta} &= 2\left(\frac{\partial g_J\left(\theta_0\right)}{\partial\theta^\intercal}\right)^\intercal W_J \sqrt{J} g_J\left(\theta_0\right) \\ &\xrightarrow{d} \mathcal{N}\left(0, 4\Gamma_0^\intercal W \Phi_0 W \Gamma_0\right),\end{aligned} \tag{3.18}$$

where

$$\Gamma_0 = \mathbf{E}\left[\frac{\partial g\left(Y^*, \theta_0\right)}{\partial\theta^\intercal}\right], \tag{3.19}$$

$$\Phi_0 = \mathbf{E}\left[g\left(Y^*, \theta_0\right) g\left(Y^*, \theta_0\right)^\intercal\right] + \sigma_\varepsilon^2. \tag{3.20}$$

Proof. From the properties of a linear transformation of random variables normally distributed[1], we have

$$\sqrt{J}\frac{\partial Q_J\left(\theta_0\right)}{\partial\theta} \xrightarrow{d} \mathcal{N}\left(0, \left(2\Gamma_0^\intercal W\right)\Phi_0\left(2\Gamma_0^\intercal W\right)^\intercal\right),$$

and, recalling that $W^\intercal = W$ since W is symmetric, we obtain

$$\sqrt{J}\frac{\partial Q_J\left(\theta_0\right)}{\partial\theta} \xrightarrow{d} \mathcal{N}\left(0, 4\Gamma_0^\intercal W \Phi_0 W \Gamma_0\right).$$ □

Now, we need to consider the additional term, namely $\frac{\partial^2 Q_J(\theta_0)}{\partial\theta\partial\theta^\intercal}$. Let $(\cdot)_{mn}$ be the $mn-th$ element of the Hessian matrix, given by

$$\left(\frac{\partial^2 Q_J\left(\theta_0\right)}{\partial\theta\partial\theta^\intercal}\right)_{mn} = 2\left(\frac{\partial g_J\left(\theta_0\right)}{\partial\theta_m}\right)^\intercal W_J \frac{\partial g_J\left(\theta_0\right)}{\partial\theta_n} + 2\left(\frac{\partial^2 g_J\left(\theta_0\right)}{\partial\theta_m\partial\theta_n}\right)^\intercal W_J g_J\left(\theta_0\right),$$

for $m, n = 1, \ldots, d$. Note that the last term converges to zero by (3.13). Hence, by virtue of (3.16), we have

$$\begin{aligned}\frac{\partial^2 Q_J\left(\theta_0\right)}{\partial\theta\partial\theta^\intercal} &\xrightarrow{p} 2\mathbf{E}\left[\frac{\partial g\left(Y^*, \theta_0\right)}{\partial\theta^\intercal}\right]^\intercal W \mathbf{E}\left[\frac{\partial g\left(Y^*, \theta_0\right)}{\partial\theta^\intercal}\right] \\ &= 2\Gamma_0^\intercal W \Gamma_0.\end{aligned} \tag{3.21}$$

[1]Let x be a random variable, given by $x = Ay + b$, where $y \backsim \mathcal{N}\left(\mu, S\right)$. Then, x is normal distributed, with $x \backsim \mathcal{N}\left(A\mu + b, ASA^\intercal\right)$.

Now, we can put (3.18) and (3.21) together and we obtain

$$\sqrt{J}\left(\hat{\theta}_{GMM}-\theta_0\right) = -\left(\frac{\partial^2 Q_J(\theta_0)}{\partial\theta\partial\theta^\intercal}\right)^{-1}\sqrt{J}\frac{\partial Q_J(\theta_0)}{\partial\theta}$$
$$\xrightarrow{d} \mathcal{N}\left(0, \left(\Gamma_0^\intercal W \Gamma_0\right)^{-1}\Gamma_0^\intercal W \Phi_0 W \Gamma_0 \left(\Gamma_0^\intercal W \Gamma_0\right)^{-1}\right), \tag{3.22}$$

where Γ_0 and Φ_0 are as in (3.19) and (3.20), respectively.

Proof. Exploiting the same property used for the proof of (3.18), we have

$$\sqrt{J}\left(\hat{\theta}_{GMM}-\theta_0\right) \xrightarrow{d} \mathcal{N}\left(0, \left(-2\Gamma_0^\intercal W \Gamma_0\right)^{-1} 4\Gamma_0^\intercal W \Phi_0 W \Gamma_0 \left(\left(-2\Gamma_0^\intercal W \Gamma_0\right)^{-1}\right)^\intercal\right),$$

that is

$$\sqrt{J}\left(\hat{\theta}_{GMM}-\theta_0\right) \xrightarrow{d} \mathcal{N}\left(0, \left(\Gamma_0^\intercal W \Gamma_0\right)^{-1}\Gamma_0^\intercal W \Phi_0 W \Gamma_0 \left(\left(\Gamma_0^\intercal W \Gamma_0\right)^{-1}\right)^\intercal\right).$$

Since $\left(\left(\Gamma_0^\intercal W \Gamma_0\right)^{-1}\right)^\intercal = \left(\left(\Gamma_0^\intercal W \Gamma_0\right)^\intercal\right)^{-1}$ and noting that $\left(\Gamma_0^\intercal W \Gamma_0\right)^\intercal = \Gamma_0^\intercal W \Gamma_0$, we then obtain

$$\sqrt{J}\left(\hat{\theta}_{GMM}-\theta_0\right) \xrightarrow{d} \mathcal{N}\left(0, \left(\Gamma_0^\intercal W \Gamma_0\right)^{-1}\Gamma_0^\intercal W \Phi_0 W \Gamma_0 \left(\Gamma_0^\intercal W \Gamma_0\right)^{-1}\right),$$

which ends the proof. □

Proposition 3.2. *The GMM estimates are consistent, with $\hat{\theta}_{GMM}$ converging to the true value θ_0 at speed $\sqrt{J}$. Moreover, the precision of our estimates can be quantified by evaluating the asymptotic variance-covariance matrix.*

An important issue in the GMM estimation is the exact identification of the model. If the number of moment conditions N is the same as the number of parameters d, then Γ_0 becomes a non-singular square matrix. Thus, the asymptotic variance-covariance matrix in (3.22) can be rewritten as

$$\begin{aligned}\left(\Gamma_0^\intercal W \Gamma_0\right)^{-1}\Gamma_0^\intercal W \Phi_0 W \Gamma_0 \left(\Gamma_0^\intercal W \Gamma_0\right)^{-1} &= \Gamma_0^{-1} W^{-1}\left(\Gamma_0^\intercal\right)^{-1}\Gamma_0^\intercal W \Phi_0 W \Gamma_0 \Gamma_0^{-1} W^{-1}\left(\Gamma_0^\intercal\right)^{-1} \\ &= \Gamma_0^{-1}\Phi_0\left(\Gamma_0^\intercal\right)^{-1} \\ &= \left(\Gamma_0^\intercal \Phi_0^{-1}\Gamma_0\right)^{-1}.\end{aligned}$$

This means that, when $N = d$, it does not really matter which weighting matrix W we are using, since it does not affect the asymptotic distribution of our parameter estimates. Hence, in this fairly standard case, we have

$$\sqrt{J}\left(\hat{\theta}_{GMM}-\theta_0\right) \xrightarrow{d} \mathcal{N}\left(0, \left(\Gamma_0^\intercal \Phi_0^{-1}\Gamma_0\right)^{-1}\right). \tag{3.23}$$

On the other hand, if the number of moment conditions is not the same as the number of parameters, which is the case concerning this work, then the optimal weighting matrix matters, even asymptotically.

Proposition 3.3. *If $W = \Phi_0^{-1}$, then the asymptotic variance-covariance matrix of the GMM estimator $\hat{\theta}_{GMM}$ is minimal and equal to $\left(\Gamma_0^\intercal \Phi_0^{-1} \Gamma_0\right)^{-1}$, which corresponds to the value obtained for the standard case of exact identification.*

Proof. By substituting $W = \Phi_0^{-1}$ into (3.22), we have

$$\begin{aligned}
\left(\Gamma_0^\intercal W \Gamma_0\right)^{-1} \Gamma_0^\intercal W \Phi_0 W \Gamma_0 \left(\Gamma_0^\intercal W \Gamma_0\right)^{-1} &= \left(\Gamma_0^\intercal \Phi_0^{-1} \Gamma_0\right)^{-1} \Gamma_0^\intercal \Phi_0^{-1} \Phi_0 \Phi_0^{-1} \Gamma_0 \left(\Gamma_0^\intercal \Phi_0^{-1} \Gamma_0\right)^{-1} \\
&= \Gamma_0^{-1} \Phi_0 \left(\Gamma_0^\intercal\right)^{-1} \Gamma_0^\intercal \Phi_0^{-1} \Phi_0 \Phi_0^{-1} \Gamma_0 \Gamma_0^{-1} \Phi_0 \left(\Gamma_0^\intercal\right)^{-1} \\
&= \Gamma_0^{-1} \Phi_0 \left(\Gamma_0^\intercal\right)^{-1} \\
&= \left(\Gamma_0^\intercal \Phi_0^{-1} \Gamma_0\right)^{-1}.
\end{aligned}$$

□

Remark 3.3. *It holds*

$$\left(\Gamma_0^\intercal W \Gamma_0\right)^{-1} \Gamma_0^\intercal W \Phi_0 W \Gamma_0 \left(\Gamma_0^\intercal W \Gamma_0\right)^{-1} - \left(\Gamma_0^\intercal \Phi_0^{-1} \Gamma_0\right)^{-1} \geq 0.$$

Proof. Given two generic matrices, namely A and B, if $B - A \geq 0$, then $A^{-1} - B^{-1} \geq 0$. Hence, we want to show that

$$\Gamma_0^\intercal \Phi_0^{-1} \Gamma_0 - \Gamma_0^\intercal W \Gamma_0 \left(\Gamma_0^\intercal W \Phi_0 W \Gamma_0\right)^{-1} \Gamma_0^\intercal W \Gamma_0 \geq 0.$$

Gathering the term $\Gamma_0^\intercal \Phi_0^{-1} \Gamma_0$ and denoting by $H = \Gamma_0^\intercal \Phi_0^{-1/2}$, we can write

$$H \left[I - \Phi_0^{1/2} W \Gamma_0 \left(\Gamma_0^\intercal W \Phi_0 W \Gamma_0\right)^{-1} \Gamma_0^\intercal W \Phi_0^{1/2}\right] H^\intercal = HPH^\intercal,$$

where P is a symmetric and idempotent matrix. Then,

$$HPPH^\intercal = HPP^\intercal H^\intercal = HP \left(HP\right)^\intercal.$$

The last expression is necessarily positive semidefinite, since

$$z^\intercal \left(HP\right) \left(HP\right)^\intercal z \geq 0, \qquad \forall z \in \mathbb{R}^d.$$

□

Remark 3.4. *Finally, we can conclude that the optimal weighting matrix is given by*

$$W = \Phi_0^{-1} = \left(\mathbf{E}\left[g\left(Y^*, \theta_0\right) g\left(Y^*, \theta_0\right)^\intercal\right] + \sigma_\varepsilon^2\right)^{-1}, \tag{3.24}$$

and the asymptotic variance of the GMM estimator is equal to

$$\begin{aligned}
\text{asyvar}\left(\hat{\theta}_{GMM}\right) &= \left(\Gamma_0^\intercal \Phi_0^{-1} \Gamma_0\right)^{-1} \\
&= \left\{\mathbf{E}\left[\frac{\partial g\left(Y^*, \theta_0\right)}{\partial \theta^\intercal}\right]^\intercal \left(\mathbf{E}\left[g\left(Y^*, \theta_0\right) g\left(Y^*, \theta_0\right)^\intercal\right] + \sigma_\varepsilon^2\right)^{-1} \mathbf{E}\left[\frac{\partial g\left(Y^*, \theta_0\right)}{\partial \theta^\intercal}\right]\right\}^{-1}.
\end{aligned} \tag{3.25}$$

It is worth noting that, in our model the most efficient choice of W is represented by the inverse of the sum between the variance of the moment restriction function and an additional term, given by the error variance. This could sound in contrast with the standard GMM theory, in which σ_ε^2 does not appear. As a consequence, we can state that the computation of Y^* (3.4) through its discrete approximation Y (3.5) allows us to obtain consistent and asymptotically normal, although not efficient, spot volatility estimations.

When $N > d$, since the optimal weighting matrix (3.24) depends on θ_0, a two stage procedure is needed, in order to estimate θ_0 by using a preliminary estimate $\hat{\theta}_{GMM(1)}$, which is given by

$$\hat{\theta}_{GMM(1)} = \arg\min_{\theta\in\Theta} g_J(\theta)^\intercal I_N g_J(\theta), \tag{3.26}$$

where I_N is the N-dimensional identity matrix. Now, the optimal weighting matrix can be computed by means of

$$W_J = \hat{\Phi}_0^{-1} = \left[g_J\left(\hat{\theta}_{GMM(1)}\right) g_J\left(\hat{\theta}_{GMM(1)}\right)^\intercal + \sigma_\varepsilon^2\right]^{-1}. \tag{3.27}$$

Finally, in the second step we can implement

$$\hat{\theta}_{GMM(2)} = \arg\min_{\theta\in\Theta} g_J(\theta)^\intercal W_J g_J(\theta), \tag{3.28}$$

where W_J is given by (3.27). Furthermore, GMM inference on the parameter values requires the estimation of the asymptotic variance-covariance matrix. Hence, from Equation (3.25), we have

$$\begin{aligned}&\widehat{\text{asyvar}}\left(\hat{\theta}_{GMM}\right)\\ &= \left[\left(\frac{\partial g_J\left(\hat{\theta}_{GMM(2)}\right)}{\partial\theta^\intercal}\right)^\intercal \left(g_J\left(\hat{\theta}_{GMM(2)}\right) g_J\left(\hat{\theta}_{GMM(2)}\right)^\intercal + \sigma_\varepsilon^2\right)^{-1} \frac{\partial g_J\left(\hat{\theta}_{GMM(2)}\right)}{\partial\theta^\intercal}\right]^{-1}.\end{aligned}$$

By the strong law of large numbers (3.16) and the convergence in probability of $\hat{\theta}_{GMM}$ to $\hat{\theta}_0$, $\widehat{\text{asyvar}}\left(\hat{\theta}_{GMM}\right) \xrightarrow{p} \text{asyvar}\left(\hat{\theta}_0\right)$.

3.1.2 Test of Overidentifying Restrictions

Once obtained the parameters estimates, we can do some inference on them and, also, we can evaluate whether all the moment conditions are close enough to zero. To this goal, the so-called *Hansen test* is available, that is

$$JQ_J\left(\hat{\theta}_{GMM}\right) = g_J\left(\hat{\theta}_{GMM}\right)^\intercal \hat{\Phi}_0^{-1} g_J\left(\hat{\theta}_{GMM}\right) \xrightarrow{d} \chi^2_{N-d}. \tag{3.29}$$

Proof. By Taylor expansion, we have

$$g_J\left(\hat{\theta}_{GMM}\right) = g_J(\theta_0) + \frac{\partial g_J(\theta_0)}{\partial\theta^\intercal}\left(\hat{\theta}_{GMM} - \theta_0\right),$$

and

$$\sqrt{J}g_J\left(\hat{\theta}_{GMM}\right) \approx \sqrt{J}g_J\left(\theta_0\right) + \Gamma_0\sqrt{J}\left(\hat{\theta}_{GMM} - \theta_0\right).$$

Now, recalling (3.14), from Equations (3.15), (3.16) and (3.21) we can write

$$\sqrt{J}\left(\hat{\theta}_{GMM} - \theta_0\right) \approx -\left(\Gamma_0^\intercal\Phi_0^{-1}\Gamma_0\right)^{-1}\Gamma_0^\intercal\Phi_0^{-1}\sqrt{J}g_J\left(\theta_0\right),$$

and then

$$\sqrt{J}g_J\left(\hat{\theta}_{GMM}\right) \approx \sqrt{J}g_J\left(\theta_0\right) - \Gamma_0\left(\Gamma_0^\intercal\Phi_0^{-1}\Gamma_0\right)^{-1}\Gamma_0^\intercal\Phi_0^{-1}\sqrt{J}g_J\left(\theta_0\right).$$

Denote by $\Psi_0 = \sqrt{J}g_J\left(\theta_0\right)$, the previous expression becomes

$$\sqrt{J}g_J\left(\hat{\theta}_{GMM}\right) \approx \left(I_N - \Gamma_0\left(\Gamma_0^\intercal\Phi_0^{-1}\Gamma_0\right)^{-1}\Gamma_0^\intercal\Phi_0^{-1}\right)\Psi_0.$$

Premultiplying $\sqrt{J}g_J\left(\hat{\theta}_{GMM}\right)$ by $\Phi_0^{-1/2}$, after some computations we obtain

$$\begin{aligned}
\Phi_0^{-1/2}\sqrt{J}g_J\left(\hat{\theta}_{GMM}\right) &\approx \Phi_0^{-1/2}\left(I_N - \Gamma_0\left(\Gamma_0^\intercal\Phi_0^{-1}\Gamma_0\right)^{-1}\Gamma_0^\intercal\Phi_0^{-1}\right)\Psi_0 \\
&\approx \left(I_N - \Phi_0^{-1/2}\Gamma_0\left(\Gamma_0^\intercal\Phi_0^{-1}\Gamma_0\right)^{-1}\Gamma_0^\intercal\Phi_0^{-1/2}\right)\Phi_0^{-1/2}\Psi_0 \\
&\approx P\Phi_0^{-1/2}\Psi_0,
\end{aligned}$$

where P is symmetric and idempotent. Hence,

$$\begin{aligned}
JQ_J\left(\hat{\theta}_{GMM}\right) &= g_J\left(\hat{\theta}_{GMM}\right)^\intercal\hat{\Phi}_0^{-1}g_J\left(\hat{\theta}_{GMM}\right) \\
&\approx \Psi_0^\intercal\Phi_0^{-1/2}P^\intercal P\Phi_0^{-1/2}\Psi_0 \\
&\approx \Psi_0^\intercal\Phi_0^{-1/2}P\Phi_0^{-1/2}\Psi_0.
\end{aligned}$$

Consider the trace of P, given by

$$\begin{aligned}
tr\left(I_N - \Phi_0^{-1/2}\Gamma_0\left(\Gamma_0^\intercal\Phi_0^{-1}\Gamma_0\right)^{-1}\Gamma_0^\intercal\Phi_0^{-1/2}\right) &= tr\left(I_N\right) - tr\left(\Phi_0^{-1/2}\Gamma_0\left(\Gamma_0^\intercal\Phi_0^{-1}\Gamma_0\right)^{-1}\Gamma_0^\intercal\Phi_0^{-1/2}\right) \\
&= N - tr\left(\Phi_0^{-1/2}\Gamma_0\left(\Gamma_0^\intercal\Phi_0^{-1}\Gamma_0\right)^{-1}\Gamma_0^\intercal\Phi_0^{-1/2}\right) \\
&= N - tr\left(\left(\Gamma_0^\intercal\Phi_0^{-1}\Gamma_0\right)^{-1}\Gamma_0^\intercal\Phi_0^{-1}\Gamma_0\right) \\
&= N - tr\left(I_d\right) = N - d.
\end{aligned}$$

Thus,

$$\begin{aligned}
JQ_J\left(\hat{\theta}_{GMM}\right) &= \Psi_0^\intercal\Phi_0^{-1/2}P\Phi_0^{-1/2}\Psi_0 \\
&= \Psi_0^\intercal\Phi_0^{-1/2}Q\Lambda Q^\intercal\Phi_0^{-1/2}\Psi_0,
\end{aligned}$$

Parameter	Value
κ	4.0320
ω	0.0144
Λ	0.20
ρ	-0.46
V_0	0.003

Table 3.1: Risk-neutral parameters setting for the numerical experiment.

where Λ is a matrix of ones and zeros, with $N-d$ ones. Denoting by $z = Q^\intercal {\Phi_0}^{-1/2} \Psi_0$, we have

$$JQ_J\left(\hat{\theta}_{GMM}\right) = z^\intercal \Lambda z = \sum_{i=1}^{N-d} z_i^2.$$

By virtue of (3.17), z is normally distributed with

$$z \sim \mathcal{N}\left(0, Q^\intercal {\Phi_0}^{-1/2} \Phi_0^{-1} {\Phi_0}^{-1/2} Q\right),$$

or

$$z \sim \mathcal{N}\left(0, I_{N-d}\right),$$

since P is symmetric and, then, $Q^\intercal Q = I_{N-d}$. As a consequence, z has independent components $\mathcal{N}(0,1)$-distributed. Finally,

$$JQ_J\left(\hat{\theta}_{GMM}\right) = \sum_{i=1}^{N-d} z_i^2 \xrightarrow{d} \chi^2_{N-d},$$

which ends the proof. □

3.2 Monte Carlo Simulations

We now investigate the finite-sample properties of the proposed estimator by virtue of several Monte Carlo simulations, based on the Heston [97] diffusion model. We fix the parameters to the consensus values from the literature, provided by Broadie et al. [50]. The full set of initial parameters values is reported in Table 3.1. We apply our inference procedures on a total of 1,000 replications.

The panel of options is constructed as follows. We set 5 maturities, ranging from two weeks to one year, which resemble the available maturities in the actual data. Firstly, we simulate only 10 strikes and, for each maturity, we compute the corresponding option prices. Then, we apply an observation error to the computed prices, in light of Equation (2.42) introduced in the previous chapter. This error is of two types: a 1% low error and an higher one equal to 5%. Secondly, we consider the case in which 50 options are available in the market and we repeat the same simulation procedure. Thus, the simulation covers four possible cases in a day: 10 options with 1% of pricing error; 10 options with 5% of pricing error; 50 options with 1% of pricing error; and finally, 50 options with 1% of pricing error.

In Tables 3.2 and 3.3, we report the results of the Monte Carlo simulation for the two panels of options. These results support our intuitions. Indeed, the observation error on option prices influences the accuracy of the estimates. In particular, Figures 3.3 and 3.4 show that the distribution of estimated parameters is highly concentrated around the mean value when the observation error is low, while parameters dispersion is more than doubled in case of an higher error. On the other hand, the effect of the option prices observation error on the estimated parameters decreases dramatically as the number of options considered increases. Our estimator bias with respect to the vector of true values is attributable to the fact that we are performing a semi-nonparametric estimation, so that a fully unbiased estimation is not achievable. Nevertheless, our estimator is able to fit almost perfectly the generated VIX term structure, as shown in Figure 3.5.

An alternative method for the estimation of spot volatility, pretty much used and appreciated in the literature, is based on the observation of high-frequency data on the underlying asset price. Such an estimator can be readily derived substituting the integrated variance with its consistent estimator, that is, the realized variance (see Andersen et al. [18], Andersen and Benzoni [12] and Barndorff-Nielsen and Shephard [32], among others, for reviews on the topic). Clearly, using intra-daily data, instead of daily observations, has the advantage of shrinking the variance of estimates, since the data set available is larger than a daily option panel. On the other hand, in order to obtain consistent estimation values, the time series considered have to be much longer. Then, high-frequency estimates have higher bias than low-frequency ones. As a consequence, spot volatility estimation methods based on realized variance present an evident trade off between the variance and the bias of the estimates. For instance, assume the availability of observations on the underlying for a year, discretely sampled every trading day, corresponding to a total of 252 daily quotes. For each such day, intra-daily returns are recorded every 5 minutes between 8.30 a.m. to 3.15 p.m.. Then, every day we would have a total of 81 price returns to estimate daily spot volatility Hence, realized volatility is a substantially more demanding estimator than our.

	True Value	Mean	Median	Min	Max	StdDev	RMSE
(a)							
κ	4.0320	2.4239	2.4256	2.2962	2.4936	0.0280	0.5080
ω	0.0144	0.2546	0.2545	0.2526	0.2582	0.0008	0.0759
v_0	0.0030	0.0926	0.0926	0.0920	0.0939	0.0003	0.0283
(b)							
κ	4.0320	2.4251	2.4236	2.1715	2.6893	0.0809	0.5086
ω	0.0144	0.2546	0.2546	0.2478	0.2623	0.0023	0.0760
v_0	0.0030	0.0926	0.0926	0.0904	0.0950	0.0007	0.0283

Table 3.2: Monte Carlo results of the risk-neutral parameters estimation for a panel of 10 options with 1% (a) and 5% (b) of observation error.

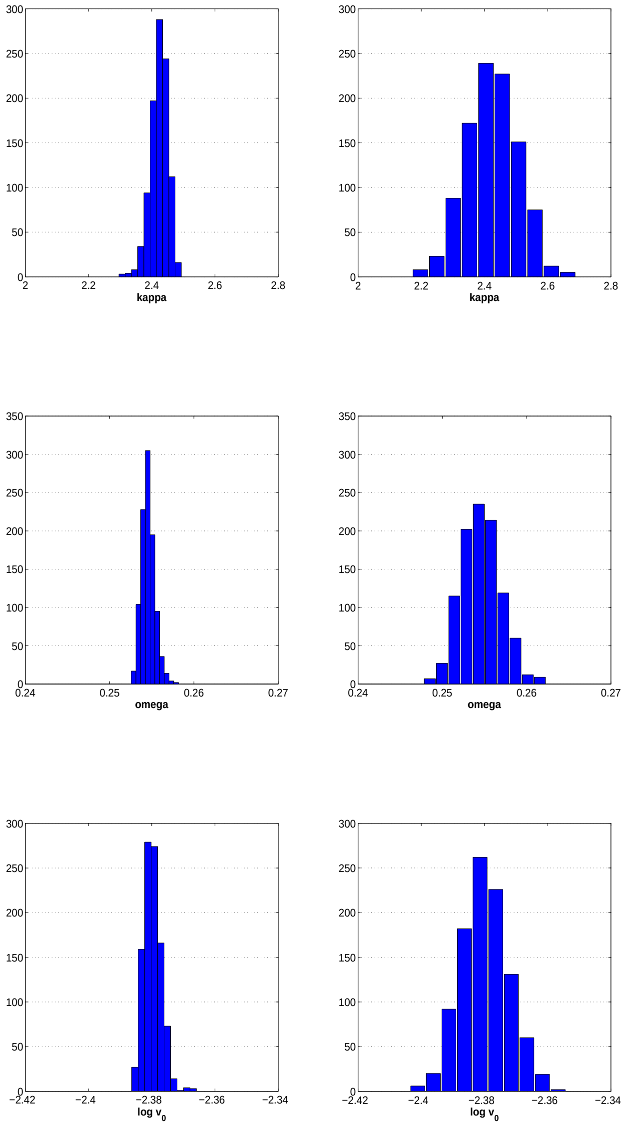

Figure 3.3: Finite-sample parameters distribution for a panel of 10 options with 1% (*left*) and 5% (*right*) of observation error.

	True Value	Mean	Median	Min	Max	StdDev	RMSE
(a)							
κ	4.0320	2.5225	2.5233	2.4703	2.5506	0.0108	0.4771
ω	0.0144	0.2463	0.2462	0.2456	0.2475	0.0003	0.0733
v_0	0.0030	0.0915	0.0915	0.0913	0.0920	0.0001	0.0280
(b)							
κ	4.0320	2.5228	2.5231	2.3717	2.6416	0.0416	0.4771
ω	0.0144	0.2463	0.2462	0.2432	0.2502	0.0011	0.0733
v_0	0.0030	0.0915	0.0915	0.0904	0.0929	0.0004	0.0280

Table 3.3: Monte Carlo results of the risk-neutral parameters estimation for a panel of 50 options with 1% (a) and 5% (b) of observation error.

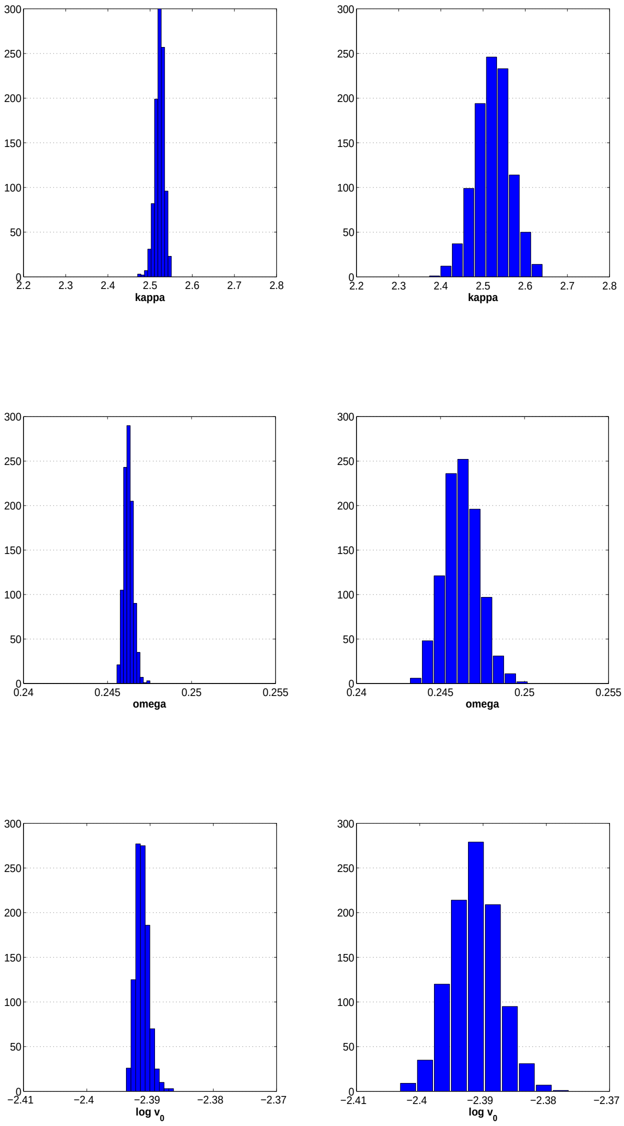

Figure 3.4: Finite-sample parameters distribution for a panel of 50 options with 1% (*left*) and 5% (*right*) of observation error.

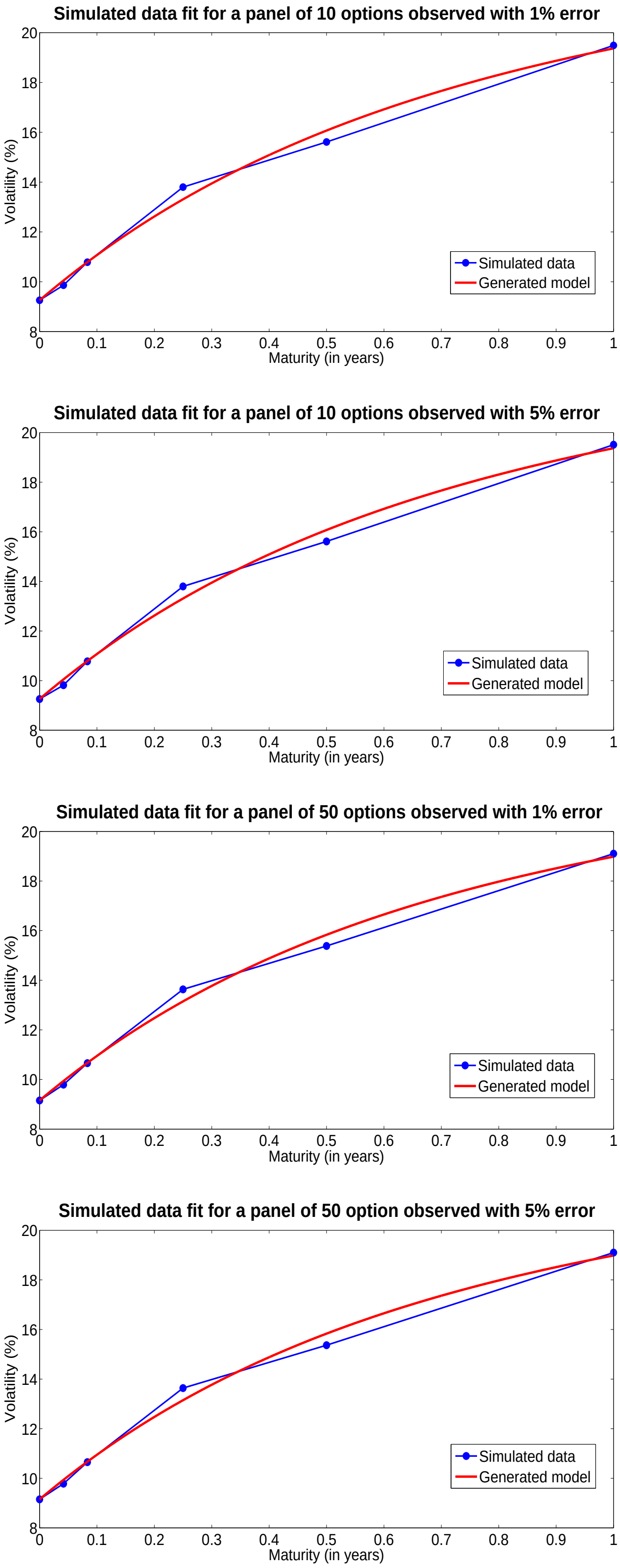

Figure 3.5: Simulated data of the VIX term structure compared to the theoretical model generated by the estimated parameters.

Chapter 4

Heston++

Nowadays, the well-known Black-Scholes [42] formula is still considered by academics and quants a universal benchmark for option pricing and hedging. On the other hand, it is widely recognized that the plainness of the Black-Scholes [42] model is no longer sufficient to capture modern market phenomena, especially since the 1987 crash. In fact, as rigorously demonstrated by the works of Bates [36] and Rubinstein [142], among others, prior to the 1987 market crash the Black-Scholes formula prices options quite accurately, whereas after the crash it systematically underprices out-of-the-money equity index put contracts.

Furthermore, empirical studies have shown some results contrary to the unrealistic assumptions of the Black-Scholes [42] model, which have been accepted as stylized facts about financial returns, such as

- the *volatility smile*, or *smirk*, which characterizes the behavior of implied volatilities with different strike prices, for a given maturity;
- the *volatility persistence*, which is the property that the autocorrelation function of squared returns is significantly positive with a range of several months, implying that in financial markets periods of low volatility are followed by periods of high volatility;
- the existence of *excess kurtosis* larger than three in the asset price returns distribution, with skewness close to zero;
- the *leverage effect*, which is the phenomenon of negative correlation between volatility and returns, or in other terms, negative returns have a greater impact on future variance than positive returns.

Hence, during the last four decades, researchers and scholars have tried to find extensions of the standard Black-Scholes [42] model, in order to fix its failures. One of the attempt to tackle the limits of the Black-Scholes model is Merton [126], who suggests that the volatility can be considered as a deterministic function of time. This approach explains the presence of a term structure for the implied volatility at different maturities, but it still does not consider the smile effect. In 1994, a series of contributions developed by Derman and Kani [77], Dupire [80] and Rubinstein [142] introduces the idea of allowing both time and state variables dependence of the volatility coefficient. These studies, concerning deterministic volatility approaches, yield complete market models, thus letting the local volatility surface to be fitted. However, they cannot explain the persistence of the smile shape even for longer maturities. Furthermore, these models do not allow to efficiently price exotic derivatives, whose trading volume is rapidly growing in modern financial markets.

In light of the random features observed in the behavior of equity price returns, *stochastic volatility models* in continuous time represent the most powerful and realistic overcoming of the Black-Scholes [42] model, able to describe a much more complex market. A list of some remarkable works which have led to the development of stochastic volatility models includes Hull and White [101], Johnson and Shanno [108], Melino and Turnbull [124], Scott [144], Stein [151], Stein and Stein [152] and Wiggings [156] (for reviews see Fouque, Papanicolau and Sircar [89] and Gatheral [92]). Nonetheless, these early efforts to identify a more realistic model for the underlying return are slowed by the analytical and computational complexity of the option pricing problem. In fact, unlike the Black-Scholes model, the first stochastic volatility specifications do not admit closed-form solutions. Thus, the evaluation of the option price requires time consuming computations through simulation method or numerical solution of the pricing partial differential equation by finite difference methods. Further, the presence of a latent factor, such as the volatility, and the lack of closed-form expressions for the likelihood function complicate the estimation problem.

Later studies have introduced some restrictions in the distribution of the underlying return process, which are consistent with the empirical evidences from the data, with the aim of allowing for closed-form, or semi-closed-form, solutions. To this goal, the *affine class* of continuous time stochastic volatility models are particularly useful in providing a flexible setting, yet preserving analytical tractability. In particular, *affine jump diffusions* are characterized by the fact that the drift term, the conditional covariance and the jump intensity are all a linear function of the state variable. The Vasicek [155] bond valuation model and the Cox, Ingersoll and Ross [70] intertemporal asset pricing model provide considerable examples of the advantages of the affine paradigm.

Let $(\Omega, \mathfrak{C}, \mathbb{P}) \equiv \Omega$ be a probability space endowed with a filtration $(\mathfrak{F}_t)_{t\geq 0}$. Under the risk-neutral measure $\mathbb{Q}$, consider the general dynamics

$$dS_t = rdt + \sqrt{V_t}dW_t^S + Y_t^S dN_t, \tag{4.1}$$

$$dV_t = \kappa\left(\omega - V_t\right)dt + \Lambda\sqrt{V_t}dW_t^V + Y_t^V dN_t, \tag{4.2}$$

where r, κ, ω and $\Lambda \in \mathbb{R}$ are positive constants, $\left(W_t^S\right)_{t\geq 0} \equiv W_t^S$ and $\left(W_t^V\right)_{t\geq 0} \equiv W_t^V$ are correlated Wiener processes, with correlation coefficient $\rho \in [-1, 1]$, $(N_t)_{t\geq 0} \equiv N_t$ is a Poisson process, uncorrelated with W_t^S and W_t^V, whose jump intensity is given by

$$\lambda_t = \lambda_0 + \lambda_1 V_t, \tag{4.3}$$

that is, $\Pr(dN_t = 1) = \lambda_t dt$. From this general setting of an affine jump diffusion model, we can see that volatility is not only stochastic, but also subject to jumps, which occur simultaneously with jumps in returns. The Black-Scholes [42] model is a special case of (4.1)-(4.3), with constant volatility and absence of jumps, that is

$$V_t = \sigma^2 \qquad \text{and} \qquad \lambda_t = 0, \qquad 0 \leq t \leq T.$$

In the same way, the Merton [127] model arises from (4.1)-(4.3) if volatility is constant but we allow for jumps in returns.

An important special case of (4.1)-(4.3), with stochastic volatility and no jumps neither in returns nor in the volatility process, is represented by the celebrate Heston [97] model. Using the characteristic function method, he derives a closed-form European option pricing formula, which can be readily evaluated by virtue of simple numerical integration. Among all the known techniques for option pricing, the Heston model is by far the most studied and used in practice mainly for two reasons. Firstly, the volatility process is non-negative and mean-reverting to its long-run mean level, which is consistent with the properties of equity index returns observed in financial markets. Secondly, its computational efficiency becomes crucial when the model needs to be calibrated to market prices. The latter is considered by most the greatest advantage of the Heston model over other stochastic volatility models, even though potentially more realistic. Its popularity also stems from the fact that it is one of the first models to be able to explain the implied volatility smile and, at the same time, to allow a front office implementation and evaluation of many exotic derivatives. Moreover, the Heston model assumes a negative correlation $\rho < 0$ between W_t^S and W_t^V, allowing for the leverage effect. Finally, a fatter left tail in the return distribution results in a higher cost for crash insurance and thus makes out-of-the-money put options more expensive. This is qualitatively consistent with patterns observed after the 1987 market crash discussed above.

Despite this profusion of literature concerning stochastic volatility models, in quantitative finance it is still open the long-lasting challenging problem to find a practical option pricing model which, at the same time,

1. is guaranteed to be free of arbitrage opportunities;
2. provides fast algorithms for prices and greeks calculation;
3. is able to fit the quoted volatility surfaces, across both maturities and strikes;
4. adequately describes price and volatility risk.

These four features are equally needful, but it is very hard to find an existent model which satisfies them all together. For instance, arbitrage-free volatility interpolation can be achieved with local volatility models (Derman and Kani [77] and Dupire [80]), as shown recently in Andreasen and Huge [22] and Kahalé [109], among others. However, local volatility models do not accommodate idiosyncratic volatility risk. On the other hand, the Heston [97] model includes volatility risk in a mathematically convenient way, but it is not flexible enough to fit the whole volatility surface. Similarly, the SABR (Hagan et al. [95]) model is almost arbitrage-free and has an analytical pricing formula, but it can hardly reproduce volatility surfaces, if parameters are not changed along with maturities.

The goal of this chapter is to fill this gap, introducing the *Heston++* class of stochastic volatility models, which is compliant to the four pillars mentioned above. We extend multi-factor affine models, with the possible inclusion of jumps, by adding a deterministic volatility factor meant to fit the term structure of the at-the-money (ATM) implied variances. The technique is borrowed from deterministic shift extension (Brigo and Mercurio [47]) and its application to short rate models. The preliminary fit of the ATM term structure strongly eases the fit of the whole volatility surface, since the model parameters are employed with their full flexibility. The straightforward application of our model allows a fast and accurate arbitrage-free interpolation, as well as a sound extrapolation, of the volatility surface, which can be used for market making or exotic pricing. Hence, the model we propose here unifies the advantages of local volatility models and those of stochastic volatility models.

To illustrate the gain in terms of pricing, we calibrate the model on a time series of daily option panels on FX rate EUR/USD from 2005 to 2012, for strikes up to 10Δ and 10 maturities ranging from one week to two years. The Heston++ models are readily fitted, with no added computational cost with respect to the standard versions, and obtain an average root mean square error of 1.26%, in the best case.

This chapter has the following structure. In Section 4.1 we briefly review the main properties of the Heston [97] model, with the goal of making as clearer as possible the rest of the exposition. Next, in Section 4.2 we illustrate the proposed model and provide the corresponding option pricing formulae. Then, in Section 4.3 we show empirical results of the Heston++ model, suitably calibrated to the FX option market.

4.1 Brief Review of the Heston Model

Let $(S_t)_{t\geq 0} \equiv S_t$ be the spot price of a tradable equity asset and let $(V_t)_{t\geq 0} \equiv V_t$ be the corresponding instantaneous variance, driven by a mean-reverting square root process. Heston [97] proposes a stochastic variance model, as an extension of the Black-Scholes [42] model, which follows the risk-neutral dynamics

$$dS_t = rS_t dt + \sqrt{V_t} S_t dW_t^S, \tag{4.4}$$

$$dV_t = \kappa\left(\omega - V_t\right) dt + \Lambda\sqrt{V_t} dW_t^V, \tag{4.5}$$

where r, κ, ω and $\Lambda \in \mathbb{R}_+$ are constants and $\left(W_t^S\right)_{t\geq 0} \equiv W_t^S$ and $\left(W_t^V\right)_{t\geq 0} \equiv W_t^V$ are two correlated Wiener processes, with correlation coefficient equal to $\rho \in [-1, 1]$.

By virtue of the usual change of variable $s_t = \ln S_t$, we can express the Heston model in

Drift	r
Rate of mean reversion	κ
Long-run variance	ω
Vol of vol	Λ
Return/variance correlation	ρ
Initial variance	V_0

Table 4.1: Parameters of the Heston model

terms of the logarithm of the spot price $(s_t)_{t\geq 0} \equiv s_t$ and we obtain

$$ds_t = \left(r - \frac{1}{2}V_t\right)dt + \sqrt{V_t}dW_t^S. \tag{4.6}$$

As we can see in Table 4.1, this model is specified by several constant parameters, such as, the *drift* r, the *rate of mean reversion* κ, the *long-run variance* ω, the *vol of vol* Λ, the *return/variance correlation* coefficient ρ and the *initial variance* V_0.

Note that Equation (4.5) is the same that we find in Cox, Ingersoll and Ross [71] for the short rate process. In particular, the mean-reverting square root process was originally mentioned in the literature by Feller [86] in 1951.

Proposition 4.1 (The Feller condition)**.** *Given the mean-reverting square root process V_t defined in (4.5), we have*

$$if \quad V_0 > 0 \cap 2\kappa\omega \geq \Lambda^2 \quad \Rightarrow \quad V_t > 0, \qquad \forall t > 0. \tag{4.7}$$

Hence, one supposed advantage of modeling V_t by a mean-reverting process is that, if the initial variance V_0 is strictly positive, which of course it will be the case for any realistic market with any measure of uncertainty, and if the ratio $2\kappa\omega/\Lambda^2$ is greater than or equal to one, then the instantaneous variance V_t remains always strictly positive with probability one. Conversely, if the Feller condition is not satisfied, then there exists a finite probability that the variance goes to zero, for some $t > 0$. Unfortunately, in the FX market the Feller condition is hardly ever satisfied, due to the particular convexities of the volatility surface typically encountered in practice. In fact, for the Feller condition to be satisfied both the mean reversion parameters κ and ω must be nonzero and, at the same time, the vol of vol Λ cannot be too large with respect to them. Furthermore, the correlation coefficient ρ is not considered in the Feller condition, but in order to introduce a volatility skew, the vol of vol needs to be large enough, such that the correlated part of W_t^V results in a volatility skew for S_t. Therefore, the calibration of the Heston model to either a volatility skew or a volatility smile will require a positive Λ and, since typical FX market conditions entail this parameter to be large enough, the Feller condition is often violated in practice.

It is worth noting that a Heston model which is always compliant with the Feller condition provides a much higher convexity of the implied volatility surface, presenting a trade-off between a real fit of market data and the respect of the Feller condition. Thus, it is crucial that any numerical schemes we develop are able to deal with the violation of the Feller condition and the variance absorption at the $V_t = 0$ boundary. This is discussed in the context of backward partial differential equation (PDE) schemes in Duffie [78] and in the original paper of Heston [97], in the context of forward Fokker-Planck PDE schemes in Lucic [119] and in Gatheral [92] for Monte Carlo methods.

We can transform the process for the variance into a process for the volatility, as suggested by Zhu [163]. Consider $\upsilon_t = \sqrt{V_t} = f(V_t)$, for which it holds $f'(V_t) = \frac{1}{2}V_t^{-1/2}$ and $f''(V_t) = -\frac{1}{4}V_t^{-3/2}$. By virtue of the Itô's lemma, we then have

$$\begin{aligned}
d\upsilon_t &= f'(V_t)\, dV_t + \frac{1}{2} f''(V_t) \langle dV_t, dV_t \rangle \\
&= \frac{1}{2}V_t^{-1/2} \left[\kappa(\omega - V_t)\, dt + \Lambda \sqrt{V_t} dW_t^S \right] - \frac{1}{8}V_t^{-3/2}\Lambda^2 V_t dt \\
&= \kappa(\omega - V_t) V_t^{-1/2} dt - \frac{1}{4}\Lambda^2 V_t^{-1/2} dt + \Lambda dW_t^S \\
&= \left(\frac{\kappa\omega}{\upsilon_t} - \kappa\upsilon_t - \frac{\Lambda^2}{4\upsilon_t} \right) dt + \Lambda dW_t^S \\
&= \kappa(\hat{\omega} - \upsilon_t)\, dt + \Lambda dW_t^S,
\end{aligned}$$

where $\hat{\omega} = \left(\omega - \frac{\Lambda^2}{4\kappa}\right)\frac{1}{\upsilon_t}$. This means that the Heston [97] model for the variance is equivalent to an Ornstein-Uhlenbeck process for the volatility, whose long-run mean level is nonstationary and state-dependent. In order to have a non negative long-run mean reversion level, we require that $\omega - \frac{\Lambda^2}{4\kappa} > 0$, that is $\frac{2\kappa\omega}{\Lambda^2} > \frac{1}{2}$. Hence, the fulfillment of the Feller condition (4.7) in the original Heston model is sufficient but not necessary for the positivity of the long-run mean reversion level.

Now, consider the value function of a general contingent claim $U(t, S_t, V_t)$ with $t \geq 0$, which pays $U(T, S_T, V_T)$ at maturity $T > t$. It can be shown[1] that $U(t, S_t, V_t)$ fulfills the PDE (functional dependencies omitted)

$$\begin{aligned}
\frac{\partial U}{\partial t} + \left(r^d - r^f\right) S_t \frac{\partial U}{\partial S_t} + \left[\kappa(\omega - V_t) - \lambda(t, S_t, V_t)\right] \frac{\partial U}{\partial V_t} + \\
+ \frac{1}{2} V_t S_t^2 \frac{\partial^2 U}{\partial S_t^2} + \rho\Lambda V_t S_t \frac{\partial^2 U}{\partial S_t \partial V_t} + \frac{1}{2}\Lambda^2 V_t \frac{\partial^2 U}{\partial V_t^2} - r^d U = 0, \quad (4.8)
\end{aligned}$$

where the term $\lambda(t, S_t, V_t)$ is the *market price of volatility risk.* Heston [97] assumes linearity of $\lambda(t, S_t, V_t)$ with respect to the instantaneous variance, that is $\lambda(t, S_t, V_t) = \lambda V_t$, in order to preserve the form of Equation (4.8) in the risk-neutral setting. A European FX option, with strike K in domestic currency units and maturity T, satisfies the PDE (4.8), subject to the

[1]For details on the derivation in the foreign exchange setting see Hakala and Wystup [96].

following boundary conditions:

$$\begin{aligned}
U\left(T, S_T, V_T\right) &= \phi\left(S_T - K\right)^+, \\
U\left(t, 0, V_t\right) &= \frac{1-\phi}{2} K \mathrm{e}^{-r^d (T-t)}, \\
\frac{\partial U}{\partial S_t}\left(t, \infty, V_t\right) &= \frac{1+\phi}{2} \mathrm{e}^{-r^f (T-t)}, \\
r^d U\left(t, S_t, 0\right) &= \left(r^d - r^f\right) S_t \frac{\partial U}{\partial S_t}\left(t, S_t, 0\right) + \kappa\omega \frac{\partial U}{\partial V_t}\left(t, S_t, 0\right) + \frac{\partial U}{\partial t}\left(t, S_t, 0\right), \\
U\left(t, S_t, \infty\right) &= \begin{cases} S_t \mathrm{e}^{-r^f (T-t)} & \text{if } \phi = 1 \\ K \mathrm{e}^{-r^d (T-t)} & \text{if } \phi = -1 \end{cases},
\end{aligned}$$

where ϕ is the option indicator, which is equal to 1 for a call and to -1 for a put.

Heston [97] solves the PDE (4.8) analytically by virtue of the characteristic function method. Hence, the pricing formula for a European FX option is given by

$$U\left(t, S_t, V_t\right) = \phi\left[\mathrm{e}^{-r^f (T-t)} S_t Q_+\left(\phi\right) - K \mathrm{e}^{-r^d (T-t)} Q_-\left(\phi\right)\right], \tag{4.9}$$

where

$$Q_+\left(\phi\right) = \frac{1-\phi}{2} + \phi Q_1\left(t, s, v, k\right), \tag{4.10}$$

$$Q_-\left(\phi\right) = \frac{1-\phi}{2} + \phi Q_2\left(t, s, v, k\right), \tag{4.11}$$

and functions $Q_j\left(t, s, v, k\right)$, $j = 1, 2$, are the cumulative distribution functions, in the variable $k = \ln K$, of the log-spot price, starting at $\ln S_t = s$ and $V_t = v$. We do not have closed form solutions for $Q_1\left(t, s, v, k\right)$ and $Q_2\left(t, s, v, k\right)$, but only for their Fourier transforms $f_1\left(t, s, v, z\right)$ and $f_2\left(t, s, v, z\right)$, respectively. Then, it holds

$$f_j\left(t, s, v, z\right) = \mathrm{e}^{A_j(\tau, z) + B_j(\tau, z) v + izs}, \qquad j = 1, 2, \tag{4.12}$$

where $\tau = T - t$, $i = \sqrt{-1}$ and

$$A_j\left(\tau, z\right) = \left(r^d - r^f\right) iz\tau + \frac{\kappa\omega}{\Lambda^2}\left[\left(c_j + d_j\right)\tau - 2\ln\left(\frac{1 - g_j \mathrm{e}^{d_j \tau}}{1 - g_j}\right)\right], \tag{4.13}$$

$$B_j\left(\tau, z\right) = \frac{c_j + d_j}{\Lambda^2}\left(\frac{1 - \mathrm{e}^{d_j \tau}}{1 - g_j \mathrm{e}^{d_j \tau}}\right), \tag{4.14}$$

and finally

$$\begin{aligned}
c_j &= b_j - \rho\Lambda iz, \qquad d_j = \sqrt{c_j^2 - \Lambda^2 z\left(ia_j - z\right)}, \qquad g_j = \frac{c_j + d_j}{c_j - d_j}, \\
b_1 &= \kappa + \lambda - \rho\Lambda, \qquad b_2 = \kappa + \lambda, \qquad a_1 = 1, \qquad a_2 = -1.
\end{aligned}$$

Hence, the two probabilities in (4.9) can be obtained numerically by the inverse transforms

$$Q_j\left(t,s,v,k\right) = \frac{1}{2} + \frac{1}{\pi}\int_0^{\infty} \Re\left[\frac{\mathrm{e}^{-izk} f_j\left(t,s,v,z\right)}{iz}\right] dz, \qquad j=1,2. \tag{4.15}$$

It is interesting to evaluate the Greeks, by taking the appropriate derivatives. In the Heston [97] model the *spot delta* and the so-called *dual delta* are given by

$$\Delta = \frac{\partial U\left(t,S_t,V_t\right)}{\partial S_t} = \phi \mathrm{e}^{-r^f(T-t)} Q_+(\phi) \quad \text{and} \quad \frac{\partial U\left(t,S_t,V_t\right)}{\partial K} = -\phi \mathrm{e}^{-r^d(T-t)} Q_-(\phi), \tag{4.16}$$

respectively.

The *Gamma*, which measures the sensitivity of delta with respect to the underlying, is equal to

$$\Gamma = \frac{\partial \Delta}{\partial S_t} = \frac{\mathrm{e}^{-r^f(T-t)}}{S_t} q_j\left(t,s,v,k\right), \tag{4.17}$$

where

$$q_j\left(t,s,v,k\right) = \frac{1}{\pi}\int_0^{\infty} \Re\left[\mathrm{e}^{-izk} f_j\left(t,s,v,z\right)\right] dz, \qquad j=1,2,$$

are the densities corresponding to the cumulative distribution functions Q_j (4.15).

The time sensitivity *theta*, $\Theta = \frac{\partial U(t,S_t,V_t)}{\partial t}$, can be easily computed from (4.8), while for *rho* we have

$$\frac{\partial U\left(t,S_t,V_t\right)}{\partial r^d} = \phi K \mathrm{e}^{-r^d(T-t)}(T-t)Q_-(\phi), \tag{4.18}$$

$$\frac{\partial U\left(t,S_t,V_t\right)}{\partial r^f} = -\phi S_t \mathrm{e}^{-r^f(T-t)}(T-t)Q_+(\phi). \tag{4.19}$$

Note that in the foreign exchange setting there are two expressions for rho, one is the derivative of the option price with respect to the domestic interest rate and the other is the derivative with respect to the foreign interest rate.

Finally, *vega*, *vanna* and *volga* are given by

$$\frac{\partial U\left(t,S_t,V_t\right)}{\partial V_t} = \mathrm{e}^{-r^f(T-t)} S_t \frac{\partial Q_1(t,s,v,k)}{\partial v} - K\mathrm{e}^{-r^d(T-t)} \frac{\partial Q_2(t,s,v,k)}{\partial v}, \tag{4.20}$$

$$\frac{\partial^2 U\left(t,S_t,V_t\right)}{\partial S_t \partial V_t} = \mathrm{e}^{-r^f(T-t)} \frac{\partial Q_1(t,s,v,k)}{\partial v} + \mathrm{e}^{-r^f(T-t)} S_t \frac{\partial^2 Q_1(t,s,v,k)}{\partial s \partial v} + - K\mathrm{e}^{-r^d(T-t)} \frac{\partial^2 Q_2(t,s,v,k)}{\partial s \partial v}, \tag{4.21}$$

$$\frac{\partial^2 U\left(t,S_t,V_t\right)}{\partial V_t^2} = \mathrm{e}^{-r^f(T-t)} S_t \frac{\partial^2 Q_1(t,s,v,k)}{\partial v^2} - K\mathrm{e}^{-r^d(T-t)} \frac{\partial^2 Q_2(t,s,v,k)}{\partial v^2}, \tag{4.22}$$

respectively, where

$$\begin{aligned}\frac{\partial Q_j(t,s,v,k)}{\partial v} &= \frac{1}{\pi}\int_0^\infty \Re\left[\frac{B(\tau,z)\,\mathrm{e}^{-izk} f_j(t,s,v,z)}{iz}\right] dz,\\ \frac{\partial^2 Q_j(t,s,v,k)}{\partial s \partial v} &= \frac{1}{S_t}\frac{1}{\pi}\int_0^\infty \Re\left[B(\tau,z)\,\mathrm{e}^{-izk} f_j(t,s,v,z)\right] dz,\\ \frac{\partial^2 Q_j(t,s,v,k)}{\partial v^2} &= \frac{1}{\pi}\int_0^\infty \Re\left[\frac{B^2(\tau,z)\,\mathrm{e}^{-izk} f_j(t,s,v,z)}{iz}\right] dz,\end{aligned}$$

with $j = 1, 2$.

The Heston solution is actually semi-analytical. Equations (4.10) and (4.11) require to integrate functions f_j, which are typically of oscillatory nature. Hence, different numerical approaches can be used to determine the price of a European FX option. These include finite difference and finite element methods (see Apel, Winkler and Wystup [23]), Monte Carlo simulations and the general Fast Fourier Transform of Carr and Madan [56]. Hakala and Wystup [96] propose to perform the integration in (4.15) with the Gauss-Laguerre quadrature, using 100 for ∞ and 100 abscissas. Jäckel and Kahl [104] suggest using the Gauss-Lobatto quadrature, (e.g. the *quadl.m* function in Matlab) and transform the original integral boundaries $[0, +\infty)$ to the finite interval $[0, 1]$. Nonetheless, as a number of authors have recently reported (Albrecher et al. [10], Gatheral [92] and Jäckel and Kahl [104]), the real problem starts when the functions f_j are evaluated as part of the quadrature scheme. In particular, the computation of the complex logarithm in (4.13) is prone to numerical instabilities. It turns out that taking the principal value of the logarithm causes A_j to jump discontinuously each time the imaginary part of the logarithm argument crosses the negative real axis. One solution is to keep track of the winding number in the integration (4.15), but it is difficult to implement since standard numerical integration routines cannot be used. In our work, we provide a practical solution to this problem, evaluating the integral in (4.15) by virtue of the adaptive Gauss-Kronrod quadrature, that is the *quadgk.m* function in Matlab, which is faster and most efficient with respect to the Gauss-Lobatto quadrature for oscillatory integrands and any smooth integrand at high accuracies.

4.2 The Heston++ Model

Let $(\Omega, \mathfrak{E}, \mathbb{P}) \equiv \Omega$ be a probability space endowed with a filtration $(\mathfrak{F}_t)_{t\geq 0}$ and let $(B_t)_{t\geq 0} \equiv B_t$ be the price of a riskless bond, satisfying the ordinary differential equation (ODE)

$$dB_t = rB_t dt, \tag{4.23}$$

where $r \in \mathbb{R}_+$ is the instantaneous constant risk-free interest rate for lending or borrowing money. Setting $B_0 = 1$, we have

$$B_t = \mathrm{e}^{rt}, \qquad t \geq 0.$$

Finally, let $(S_t)_{t\geq 0} \equiv S_t$ be the price process of a non dividend-paying risky equity asset, evolving in continuous time. Under the risk-neutral probability measure $\mathbb{Q}$, we assume that S_t follows the general dynamics

$$dS_t = rS_t dt + \sqrt{V_t} S_t dW_t^S, \tag{4.24}$$

$$V_t = v_t + \phi_t, \tag{4.25}$$

$$dv_t = \kappa(\omega - v_t)dt + \Lambda\sqrt{v_t}dW_t^v, \tag{4.26}$$

where r, κ, ω and $\Lambda \in \mathbb{R}_+$ are constants, $\phi_t \geq 0$ is a smooth enough non-negative real function integrable on closed intervals with $\phi_0 = 0$, and $\left(W_t^S\right)_{t\geq 0} \equiv W_t^S$ and $(W_t^v)_{t\geq 0} \equiv W_t^v$ are correlated Wiener processes on Ω adapted to the filtration $\mathfrak{F}_t$, with time dependent instantaneous risk-neutral correlation given by

$$\operatorname{corr}^{\mathbb{Q}}\left(dW_t^S, dW_t^v\right) = \rho\sqrt{\frac{v_t}{v_t + \phi_t}}, \tag{4.27}$$

where $\rho \in [-1, 1]$ is an additional constant. Moreover, we assume that the Feller condition (4.7) is satisfied, since $2\kappa\omega \geq \Lambda^2$, in order to ensure that the process $(v_t)_{t\geq 0} \equiv v_t$ remains strictly positive. Note that the form of the correlation in (4.27) guarantees the linearity of the pricing PDE associated with the model.

Remark 4.1. *We have*

$$\left|\operatorname{corr}^{\mathbb{Q}}\left(dW_t^S, dW_t^v\right)\right| \leq 1 \qquad \Longleftrightarrow \qquad \rho^2 \leq 1 + \frac{\phi_t}{v_t}, \tag{4.28}$$

and since we assume $|\rho| \leq 1$, $\phi_t \geq 0$ and $v_t > 0$, the condition (4.28) is met. Furthermore,

$$\operatorname{cov}^{\mathbb{Q}}\left(d\ln S_t, dV_t\right) = \rho\Lambda v_t dt \ \leq \ \rho\Lambda V_t dt,$$

where $\rho\Lambda V_t dt$ is the instantaneous covariance between the driving Wiener processes in the original Heston [97] model.

Then, the model is specified by the risk-free rate, five parameters $(\kappa, \omega, \Lambda, \rho, v_0)$ and the deterministic function ϕ_t.

An approach similar to the Heston++ model is represented by the Heston model with time-dependent parameters, introduced by Mikhailov and Nögel [129] and then further deepened by Elices [83], in which the constant parameters are replaced by deterministic functions. However, our model has an inherently different specification and presents, in our view, several advantages. The difference consists in the fact that in the Heston++ model the function ϕ_t can be interpreted as the lower bound for the spot variance, as it is clear from Equation (4.25), while in the Heston model with time-dependent parameters the lower bound for the local variance is still equal to 0. This technical difference is harmless from a specification point of view, but it is crucial in making the pricing of European vanilla options straightforward. Furthermore, the Heston model with time-dependent parameters displays serious mathematical troubles, which are typically solved by virtue of approximations based on Taylor expansions or using deterministic functions which are piecewise constant[2]. Moreover, its extension to multi-factor models with jumps appears cumbersome. Conversely, the Heston++ model has no additional mathematical and implementation complexity with respect to the traditional Heston model, providing simple formulas and fast algorithms, and can be easily generalized to affine models, such as those in Bates [37], Bates [38], Christoffersen et al. [67] and Duffie et al. [79], or to Wishart models, in the sense of Da Fonseca et al. [72].

Now, we provide the pricing formulas of European vanilla options in an univariate case, in a bivariate case and in a bivariate case with jumps. In the sequel, in order to compare the Heston++ model with the standard Heston model, all quantities related to the Heston model will have an H as superscript.

4.2.1 One factor

Let $s_t = \ln S_t$ be the logarithm of the risky equity asset price. By virtue of the Itô's lemma, we have

$$ds_t = \left[r - \frac{1}{2}(v_t + \phi_t)\right] dt + \sqrt{v_t + \phi_t}\, dW_t^S. \tag{4.29}$$

Furthermore, denote by

$$I_\phi(t,T) = \int_t^T \phi_u\, du, \qquad 0 \le t \le T. \tag{4.30}$$

Consider a European call option, with maturity $T > 0$ and strike $K > 0$, and set $k = \ln K$. Using Geman, El Karoui and Rochet [93][3], the call option pricing function at time $t \in [0,T]$ can be written in the form

$$C(t, s_t, v_t, k) = \mathrm{e}^{s_t} Q_1(t,s,v,k) - \mathrm{e}^{k - r(T-t)} Q_2(t,s,v,k), \tag{4.31}$$

where, for every $s, v > 0$,

$$Q_1(t,s,v,k) = \mathbb{Q}^S\left[s_T \ge k \mid s_t = s, v_t = v\right] \tag{4.32}$$

$$Q_2(t,s,v,k) = \mathbb{Q}^T\left[s_T \ge k \mid s_t = s, v_t = v\right] \tag{4.33}$$

are the probabilities of the call option exercise, computed with respect to the martingale probability measure $\mathbb{Q}^S$ for the numéraire S and to the T-forward probability measure $\mathbb{Q}^T$, respectively. Note that, since we assume a deterministic short interest rate, we have $\mathbb{Q}^T = \mathbb{Q}$, and then $Q_2(t,s,v,k)$ is the standard risk-neutral probability of the call option exercise.

As in the original Heston [97] model, in the Heston++ model we do not have closed form solutions for $Q_1(t,s,v,k)$ and $Q_2(t,s,v,k)$, but only for their Fourier transforms $f_1(t,s,v,z)$

[2] See Benhamou et al. [39] for a detailed discussion.

[3] Geman, El Karoui and Rochet [93] introduce the general formal framework for the change of numéraire technique. See Brigo and Mercurio [48] for a change of numéraire toolkit.

and $f_2(t,s,v,z)$, respectively. Hence, it holds

$$f_j(t,s,v,z) = f_j^H(t,s,v,z)\,\mathrm{e}^{\frac{1}{2}z(a_j i - z)I_\phi(t,T)}, \qquad j=1,2, \tag{4.34}$$

where $a_1 = 1$, $a_2 = -1$, $i = \sqrt{-1}$ and $f_j^H(t,s,v,z)$, $j=1,2$, are the transforms of Q_j^H in the Heston model, as defined in Equation (4.12).

Proof. We know that the call option pricing function $C(t,s_t,v_t,k)$ is the solution of a PDE with two space dimensions $s_t = s$ and $v_t = v$, given by (functional dependencies omitted)

$$\frac{\partial C}{\partial t} + \left[r - \frac{1}{2}(v+\phi_t)\right]\frac{\partial C}{\partial s} + \kappa(\omega - v)\frac{\partial C}{\partial v} + \\ + \rho\Lambda v\frac{\partial^2 C}{\partial s\partial v} + \frac{1}{2}(v+\phi_t)\frac{\partial^2 C}{\partial v^2} + \frac{1}{2}\Lambda^2 v\frac{\partial^2 C}{\partial v^2} - rC = 0, \tag{4.35}$$

with terminal condition

$$C(T,s_T,v_T,k) = \left(\mathrm{e}^{s_T} - \mathrm{e}^k\right)^+.$$

Note that, as mentioned before, linearity of the PDE (4.35) is guaranteed by the condition (4.27). The original Heston [97] argument, seen in the previous section, extends to our Heston++ model. Hence, the two probabilities $Q_j(t,s,v,k)$, $j=1,2$, in (4.31) are the solutions of PDEs

$$\frac{\partial Q_j}{\partial t} + \left[r + \frac{1}{2}a_j(v+\phi_t)\right]\frac{\partial Q_j}{\partial s} + \left[\kappa\omega - (\kappa - b_j\rho\Lambda)v\right]\frac{\partial Q_j}{\partial v} + \\ + \rho\Lambda v\frac{\partial^2 Q_j}{\partial s\partial v} + \frac{1}{2}(v+\phi_t)\frac{\partial^2 Q_j}{\partial s^2} + \frac{1}{2}\Lambda^2 v\frac{\partial^2 Q_j}{\partial s^2} = 0, \tag{4.36}$$

$j=1,2$, with terminal conditions

$$Q_j(T,s_T,v_T,k) = \mathbf{1}_{\{s_T \geq k\}}, \qquad j=1,2,$$

where $a_1 = 1$, $a_2 = -1$, $b_1 = 1$, $b_2 = 0$ and $\mathbf{1}_{\{s_T \geq k\}}$ is the indicator function of $s_T \geq k$.

Details 2. *Let us compute the relevant derivatives of (4.31). We then have*

$$\begin{aligned}
\frac{\partial C}{\partial t} &= \mathrm{e}^s\frac{\partial Q_1}{\partial t} - r\mathrm{e}^{k-r(T-t)}Q_2 - \mathrm{e}^{k-r(T-t)}\frac{\partial Q_2}{\partial t}\\
\frac{\partial C}{\partial s} &= \mathrm{e}^s Q_1 + \mathrm{e}^s\frac{\partial Q_1}{\partial s} - \mathrm{e}^{k-r(T-t)}\frac{\partial Q_2}{\partial s}\\
\frac{\partial C}{\partial v} &= \mathrm{e}^s\frac{\partial Q_1}{\partial v} - \mathrm{e}^{k-r(T-t)}\frac{\partial Q_2}{\partial v}\\
\frac{\partial^2 C}{\partial s\partial v} &= \mathrm{e}^s\frac{\partial Q_1}{\partial v} + \mathrm{e}^s\frac{\partial^2 Q_1}{\partial s\partial v} - \mathrm{e}^{k-r(T-t)}\frac{\partial^2 Q_2}{\partial s\partial v}\\
\frac{\partial^2 C}{\partial s^2} &= \mathrm{e}^s Q_1 + 2\mathrm{e}^s\frac{\partial Q_1}{\partial s} + \mathrm{e}^s\frac{\partial^2 Q_1}{\partial s^2} - \mathrm{e}^{k-r(T-t)}\frac{\partial^2 Q_2}{\partial s^2}\\
\frac{\partial^2 C}{\partial v^2} &= \mathrm{e}^s\frac{\partial^2 Q_1}{\partial v^2} - \mathrm{e}^{k-r(T-t)}\frac{\partial^2 Q_2}{\partial v^2}.
\end{aligned}$$

Substituting the computed derivatives and (4.31) into (4.35), we obtain

$$
\begin{aligned}
&\mathrm{e}^s \frac{\partial Q_1}{\partial t} - r\mathrm{e}^{k-r(T-t)} Q_2 - \mathrm{e}^{k-r(T-t)} \frac{\partial Q_2}{\partial t} + \\
&\quad + \left[r - \frac{1}{2}(v+\phi_t)\right]\left[\mathrm{e}^s Q_1 + \mathrm{e}^s \frac{\partial Q_1}{\partial s} - \mathrm{e}^{k-r(T-t)} \frac{\partial Q_2}{\partial s}\right] + \\
&\quad\quad + \kappa(\omega - v)\left[\mathrm{e}^s \frac{\partial Q_1}{\partial v} - \mathrm{e}^{k-r(T-t)} \frac{\partial Q_2}{\partial v}\right] + \\
&\quad\quad\quad + \rho\Lambda v \left[\mathrm{e}^s \frac{\partial Q_1}{\partial v} + \mathrm{e}^s \frac{\partial^2 Q_1}{\partial s \partial v} - \mathrm{e}^{k-r(T-t)} \frac{\partial^2 Q_2}{\partial s \partial v}\right] + \\
&\quad\quad\quad\quad + \frac{1}{2}(v+\phi_t)\left[\mathrm{e}^s Q_1 + 2\mathrm{e}^s \frac{\partial Q_1}{\partial s} + \mathrm{e}^s \frac{\partial^2 Q_1}{\partial s^2} - \mathrm{e}^{k-r(T-t)} \frac{\partial^2 Q_2}{\partial s^2}\right] + \\
&\quad\quad\quad\quad\quad + \frac{1}{2}\Lambda^2 v \left[\mathrm{e}^s \frac{\partial^2 Q_1}{\partial v^2} - \mathrm{e}^{k-r(T-t)} \frac{\partial^2 Q_2}{\partial v^2}\right] - r\mathrm{e}^s Q_1 + r\mathrm{e}^{k-r(T-t)} Q_2 = 0,
\end{aligned}
$$

that is,

$$
\begin{aligned}
&\mathrm{e}^s \frac{\partial Q_1}{\partial t} - \mathrm{e}^{k-r(T-t)} \frac{\partial Q_2}{\partial t} + r\left[\mathrm{e}^s \frac{\partial Q_1}{\partial s} - \mathrm{e}^{k-r(T-t)} \frac{\partial Q_2}{\partial s}\right] + \\
&\quad + \frac{1}{2}(v+\phi_t)\left[\mathrm{e}^s \frac{\partial Q_1}{\partial s} + \mathrm{e}^{k-r(T-t)} \frac{\partial Q_2}{\partial s} + \mathrm{e}^s \frac{\partial^2 Q_1}{\partial s^2} - \mathrm{e}^{k-r(T-t)} \frac{\partial^2 Q_2}{\partial s^2}\right] + \\
&\quad\quad + \kappa(\omega - v)\left[\mathrm{e}^s \frac{\partial Q_1}{\partial v} - \mathrm{e}^{k-r(T-t)} \frac{\partial Q_2}{\partial v}\right] + \\
&\quad\quad\quad + \rho\Lambda v \left[\mathrm{e}^s \frac{\partial Q_1}{\partial v} + \mathrm{e}^s \frac{\partial^2 Q_1}{\partial s \partial v} - \mathrm{e}^{k-r(T-t)} \frac{\partial^2 Q_2}{\partial s \partial v}\right] + \\
&\quad\quad\quad\quad + \frac{1}{2}\Lambda^2 v \left[\mathrm{e}^s \frac{\partial^2 Q_1}{\partial v^2} - \mathrm{e}^{k-r(T-t)} \frac{\partial^2 Q_2}{\partial v^2}\right] = 0,
\end{aligned}
$$

and then

$$
\begin{aligned}
&\mathrm{e}^s \left\{ \frac{\partial Q_1}{\partial t} + \left[r + \frac{1}{2}(v+\phi_t)\right] \frac{\partial Q_1}{\partial s} + \left[\kappa\omega - \kappa v + \rho\Lambda v\right] \frac{\partial Q_1}{\partial v} + \right. \\
&\quad \left. + \rho\Lambda v \frac{\partial^2 Q_1}{\partial s \partial v} + \frac{1}{2}(v+\phi_t) \frac{\partial^2 Q_1}{\partial s^2} + \frac{1}{2}\Lambda^2 v \frac{\partial^2 Q_1}{\partial v^2} \right\} + \\
&\quad\quad - \mathrm{e}^{k-r(T-t)} \left\{ \frac{\partial Q_2}{\partial t} + \left[r - \frac{1}{2}(v+\phi_t)\right] \frac{\partial Q_2}{\partial s} + \kappa(\omega - v) \frac{\partial Q_2}{\partial v} + \right. \\
&\quad\quad\quad \left. + \rho\Lambda v \frac{\partial^2 Q_2}{\partial s \partial v} + \frac{1}{2}(v+\phi_t) \frac{\partial^2 Q_2}{\partial s^2} + \frac{1}{2}\Lambda^2 v \frac{\partial^2 Q_2}{\partial v^2} \right\} = 0.
\end{aligned}
$$

A standard separation argument decomposes the last PDE into a pair of PDEs for Q_1 and Q_2, both of the form (4.36).

Also the characteristic functions of the two probabilities $f_j(t, s, v, z)$, $j = 1, 2$, are solutions of the same PDEs, but with terminal conditions given by

$$f_j\left(T,s_T,v_T,z\right)=\mathrm{e}^{izs_T},\qquad j=1,2. \tag{4.37}$$

Since PDEs (4.36) are linear in the state variable v and do not contain the state variable s, while the terminal condition (4.37) is log-linear in s and does not contain v, we have

$$f_j\left(t,s,v,z\right)=\mathrm{e}^{A_j(\tau,z)+B_j(\tau,z)v+izs},\qquad j=1,2, \tag{4.38}$$

where $\tau=T-t$ and functions $t\mapsto A_j\left(\tau,z\right)$ and $t\mapsto B_j\left(\tau,z\right)$, $j=1,2$, are solutions of the ODEs

$$\begin{cases}\dfrac{\partial A_j(\tau,z)}{\partial t}=-riz-\kappa\omega B_j(\tau,z)-\dfrac{1}{2}z\left(a_ji-z\right)\phi_t\\ A\left(0,z\right)=0\end{cases}, \tag{4.39}$$

and

$$\begin{cases}\dfrac{\partial B_j(\tau,z)}{\partial t}=-\dfrac{1}{2}z\left(a_ji-z\right)+\left[\kappa-\left(b_j+iz\right)\rho\Lambda\right]B_j(\tau,z)-\dfrac{1}{2}\Lambda^2B_j^2(\tau,z)\\ B\left(0,z\right)=0\end{cases}, \tag{4.40}$$

for fixed $\tau\geq0$ and z.

The ODE (4.40) is the same Riccati equation that we find in the original Heston model. Therefore, it is known that its solution is given by Equation (4.14), that we report here for the reader's convenience:

$$B_j\left(\tau,z\right)=B_j^H\left(\tau,z\right)=\frac{c_j+d_j}{\Lambda^2}\frac{1-\mathrm{e}^{d_j\tau}}{1-g_j\mathrm{e}^{d_j\tau}},\qquad j=1,2, \tag{4.41}$$

where

$$c_j=\kappa-\left(b_j+iz\right)\rho\Lambda,\qquad d_j=\sqrt{c_j^2-\left(a_ji-z\right)z\Lambda^2},\qquad g_j=\frac{c_j+d_j}{c_j-d_j}.$$

Once obtained Equation (4.41), the ODE (4.39) can be solved by direct integration, that is

$$A_j\left(\tau,z\right)=\int_t^T\left[izr+\kappa\omega B_j\left(u,z\right)\right]du+\frac{1}{2}z\left(a_ji-z\right)\int_t^T\phi_u du,\qquad j=1,2. \tag{4.42}$$

The first integral also appears in the derivation of the standard Heston model, thus we have for $j=1,2$

$$A_j\left(\tau,z\right)=A_j^H\left(\tau,z\right)+\frac{1}{2}z\left(a_ji-z\right)\int_t^T\phi_u du,$$

or

$$A_j\left(\tau,z\right)=riz\tau+\frac{\kappa\omega}{\Lambda^2}\left[\left(c_j+d_j\right)\tau-2\ln\left(\frac{1-g_j\mathrm{e}^{d_j\tau}}{1-g_j}\right)\right]+\frac{1}{2}z\left(a_ji-z\right)\int_t^T\phi_u du.$$

Details 3. *Computing the relevant partial derivatives of (4.38), we have*

$$\frac{\partial f_j}{\partial t} = \left(\frac{\partial A_j}{\partial t} + \frac{\partial B_j}{\partial t} v\right) f_j, \qquad \frac{\partial f_j}{\partial s} = iz f_j, \qquad \frac{\partial f_j}{\partial v} = B_j f_j,$$
$$\frac{\partial^2 f_j}{\partial s \partial v} = iz B_j f_j, \qquad \frac{\partial^2 f_j}{\partial s^2} = -z^2 f_j, \qquad \frac{\partial^2 f_j}{\partial v^2} = B_j^2 f_j,$$

with $j = 1, 2$. Substituting the computed derivatives into the PDE, we obtain

$$\left(\frac{\partial A_j}{\partial t} + \frac{\partial B_j}{\partial t} v\right) f_j + iz\left[r + \frac{1}{2} a_j (v + \phi_t)\right] f_j + \left[\kappa\omega - (\kappa - b_j \rho \Lambda) v\right] B_j f_j +$$
$$+ i\rho\Lambda v z B_j f_j - \frac{1}{2}(v + \phi_t) z^2 f_j + \frac{1}{2}\Lambda^2 v B_j^2 f_j = 0,$$

that is,

$$\frac{\partial A_j}{\partial t} + izr + \frac{1}{2} iz a_j \phi_t + \kappa\omega B_j - \frac{1}{2}\phi_t z^2 +$$
$$+ \left[\frac{\partial B_j}{\partial t} + \frac{1}{2} iz a_j - (\kappa - b_j \rho\Lambda) B_j + i\rho\Lambda z B_j - \frac{1}{2} z^2 + \frac{1}{2}\Lambda^2 B_j^2\right] v = 0.$$

A standard separation argument decomposes the above equation into a pair of ODEs of the form (4.39) and (4.40). □

Finally, the two probabilities in Equation (4.31) and, consequently, the option price can be obtained numerically by the inverse transforms

$$Q_j(t, s, v, k) = \frac{1}{2} + \frac{1}{\pi}\int_0^\infty \Re\left[\frac{\mathrm{e}^{-izk} f_j(t, s, v, z)}{iz}\right] dz, \qquad j = 1, 2. \tag{4.43}$$

As said before, the Heston++ model is an extension of the classic Heston [97] model, which is recovered when $\phi_t = 0$. However, the condition $\phi_t \neq 0$ can be exploited to fit the volatility surface without sacrificing the computational advantages of the Heston model. Indeed, in the Heston++ model, the risk-neutral expected integrated variance in the time interval $[T_1, T_2]$, conditional to time $t < T_1$, is given by

$$\begin{aligned} IV_t(T_1, T_2) &= \frac{1}{T_2 - T_1}\mathbf{E}^{\mathbb{Q}}\left[\int_{T_1}^{T_2} V_u du \,\middle|\, V_t\right] \\ &= \frac{1}{T_2 - T_1}\mathbf{E}^{\mathbb{Q}}\left[\int_{T_1}^{T_2} v_u du \,\middle|\, v_t\right] + \frac{1}{T_2 - T_1} I_\phi(T_1, T_2), \end{aligned}$$

that is

$$IV_t(T_1, T_2) = IV_t^H(T_1, T_2) + \frac{1}{T_2 - T_1} I_\phi(T_1, T_2). \tag{4.44}$$

Therefore, in the extended model, the integrated variance decomposes into the sum of the standard Heston integrated variance IV^H plus the contribution of the deterministic part of the variance, given by the average of ϕ_t over $[T_1, T_2]$.

4.2.2 Two factors

In this section, we present a *two-factor Heston++ model*, which we have developed by applying the idea of an additional deterministic volatility factor to the original model proposed in Christoffersen, Heston and Jacobs [67].

To this goal, assume now the existence of a second stochastic volatility factor, defined on the probability space Ω, given by

$$dv_{2,t} = \kappa_2 \left(\omega_2 - v_{2,t}\right) dt + \Lambda_2 \sqrt{v_{2,t}} dW^v_{2,t}, \tag{4.45}$$

where κ_2, ω_2 and $\Lambda_2 \in \mathbb{R}_+$ are constants, with $2\kappa_2\omega_2 \geq \Lambda_2^2$, and $\left(W^v_{2,t}\right)_{t\geq 0} \equiv W^v_{2,t}$ is a standard Wiener process adapted to the filtration $\mathfrak{F}_t$. Then, the general dynamics (4.24)-(4.26) becomes

$$\begin{aligned} dS_t &= rS_t dt + \sqrt{V_{1,t}} S_t dW^S_{1,t} + \sqrt{v_{2,t}} S_t dW^S_{2,t}, && (4.46)\\ V_{1,t} &= v_{1,t} + \phi_t, && (4.47)\\ dv_{1,t} &= \kappa_1(\omega_1 - v_{1,t})dt + \Lambda_1 \sqrt{v_{1,t}} dW^v_{1,t}, && (4.48)\\ dv_{2,t} &= \kappa_2 \left(\omega_2 - v_{2,t}\right) dt + \Lambda_2 \sqrt{v_{2,t}} dW^v_{2,t}, && (4.49) \end{aligned}$$

where $\left(W^S_{1,t}\right)_{t\geq 0} \equiv W^S_{1,t}$, $\left(W^S_{2,t}\right)_{t\geq 0} \equiv W^S_{2,t}$, $\left(W^v_{1,t}\right)_{t\geq 0} \equiv W^v_{1,t}$ and $\left(W^v_{2,t}\right)_{t\geq 0} \equiv W^v_{2,t}$ are Wiener processes, whose correlations are settled according to (4.27) and

$$\begin{gathered} \mathrm{corr}^{\mathbb{Q}}\left(dW^S_{1,t}, dW^S_{2,t}\right) = \mathrm{corr}^{\mathbb{Q}}\left(dW^v_{1,t}, dW^v_{2,t}\right) = 0,\\ \mathrm{corr}^{\mathbb{Q}}\left(dW^S_{2,t}, dW^v_{1,t}\right) = \mathrm{corr}^{\mathbb{Q}}\left(dW^S_{1,t}, dW^v_{2,t}\right) = 0,\\ \mathrm{corr}^{\mathbb{Q}}\left(dW^S_{2,t}, dW^v_{2,t}\right) = \rho_2, \end{gathered} \tag{4.50}$$

with $\rho_2 \in [-1, 1]$. Moreover, the SDE (4.29) associated to the logarithm of the spot price s_t becomes

$$ds_t = \left[r - \frac{1}{2}\left(v_{1,t} + v_{2,t} + \phi_t\right)\right] dt + \sqrt{v_{1,t} + \phi_t} dW^S_{1,t} + \sqrt{v_{2,t}} dW^S_{2,t}. \tag{4.51}$$

Hereafter, all quantities related to the original Christoffersen, Heston and Jacobs [67] model will be marked with a H_2 superscript.

The option pricing formula for the two-factor Heston++ model can be obtained by virtue of the same argument previously introduced for the one factor case. Hence, we have

$$C\left(t, s_t, v_{1,t}, v_{2,t}, k\right) = \mathrm{e}^{s_t} Q_1\left(t, s, v_1, v_2, k\right) - \mathrm{e}^{k - r(T-t)} Q_2\left(t, s, v_1, v_2, k\right). \tag{4.52}$$

Note that the two probabilities $Q_1\left(t, s, v_1, v_2, k\right)$ and $Q_2\left(t, s, v_1, v_2, k\right)$ have the same financial interpretation as in (4.33) and their characteristic functions are given by

$$f_j\left(t, s, v_1, v_2, z\right) = f_j^{H_2}\left(t, s, v_1, v_2, z\right) \mathrm{e}^{\frac{1}{2} z (a_j i - z) I_\phi(t,T)}, \qquad j = 1, 2, \tag{4.53}$$

where $I_\phi\left(t, T\right)$ is defined as in (4.30), $a_1 = 1$, $a_2 = -1$, $i = \sqrt{-1}$ and $f_j^{H_2}\left(t, s, v_1, v_2, z\right)$, with

$j = 1, 2$, are the transforms of $Q_j^{H_2}$ in the original Christoffersen, Heston and Jacobs [67] model.

Proof. In the two-factor model, the PDE (4.35) becomes (functional dependencies omitted)

$$\frac{\partial C}{\partial t} + \left[r - \frac{1}{2}\left(\upsilon_1 + \upsilon_2 + \phi_t\right)\right]\frac{\partial C}{\partial s} + \kappa_1\left(\omega_1 - \upsilon_1\right)\frac{\partial C}{\partial \upsilon_1} + \kappa_2\left(\omega_2 - \upsilon_2\right)\frac{\partial C}{\partial \upsilon_2} + \rho_1\Lambda_1\upsilon_1\frac{\partial^2 C}{\partial s\partial \upsilon_1} + \\ + \rho_2\Lambda_2\upsilon_2\frac{\partial^2 C}{\partial s\partial \upsilon_2} + \frac{1}{2}\left(\upsilon_1 + \upsilon_2 + \phi_t\right)\frac{\partial^2 C}{\partial s^2} + \frac{1}{2}\Lambda_1^2\upsilon_1\frac{\partial^2 C}{\partial \upsilon_1^2} + \frac{1}{2}\Lambda_2^2\upsilon_2\frac{\partial^2 C}{\partial \upsilon_2^2} - rC = 0, \quad (4.54)$$

with terminal condition

$$C\left(T, s_T, \upsilon_{1,T}, \upsilon_{2,T}, k\right) = \left(\mathrm{e}^s - \mathrm{e}^k\right)^+ .$$

As in the one factor case, the two probabilities $Q_j\left(t, s, \upsilon_1, \upsilon_2, k\right)$, $j = 1, 2$, in (4.52) are solutions of the PDEs

$$\frac{\partial Q_j}{\partial t} + \left[r - \frac{1}{2}a_j\left(\upsilon_1 + \upsilon_2 + \phi_t\right)\right]\frac{\partial Q_j}{\partial s} + \left[\kappa_1\omega_1 - \left(\kappa_1 - b_j\rho_1\Lambda_1\right)\upsilon_1\right]\frac{\partial Q_j}{\partial \upsilon_1} + \\ + \left[\kappa_2\omega_2 - \left(\kappa_2 - b_j\rho_2\Lambda_2\right)\upsilon_2\right]\frac{\partial Q_j}{\partial \upsilon_2} + \rho_1\Lambda_1\upsilon_1\frac{\partial^2 Q_j}{\partial s\partial \upsilon_1} + \rho_2\Lambda_2\upsilon_2\frac{\partial^2 Q_j}{\partial s\partial \upsilon_2} + \\ + \frac{1}{2}\left(\upsilon_1 + \upsilon_2 + \phi_t\right)\frac{\partial^2 Q_j}{\partial s^2} + \frac{1}{2}\Lambda_1^2\upsilon_1\frac{\partial^2 Q_j}{\partial \upsilon_1^2} + \frac{1}{2}\Lambda_2^2\upsilon_2\frac{\partial^2 Q_j}{\partial \upsilon_2^2} = 0, \quad (4.55)$$

with terminal conditions

$$Q_j\left(T, s_T, \upsilon_{1,T}, \upsilon_{2,T}, k\right) = \mathbf{1}_{\{s_T \geq k\}}, \qquad j = 1, 2,$$

where $a_1 = 1$, $a_2 = -1$, $b_1 = 1$, $b_2 = 0$ and $\mathbf{1}_{\{s_T \geq k\}}$ is the indicator function of $s_T \geq k$.

Details 4. *The relevant derivatives of Equation (4.52) are given by*

$$
\begin{aligned}
\frac{\partial C}{\partial t} &= \mathrm{e}^s \frac{\partial Q_1}{\partial t} - r\mathrm{e}^{k-r(T-t)} Q_2 - \mathrm{e}^{k-r(T-t)} \frac{\partial Q_2}{\partial t} \\
\frac{\partial C}{\partial s} &= \mathrm{e}^s Q_1 + \mathrm{e}^s \frac{\partial Q_1}{\partial s} - \mathrm{e}^{k-r(T-t)} \frac{\partial Q_2}{\partial s} \\
\frac{\partial C}{\partial \upsilon_1} &= \mathrm{e}^s \frac{\partial Q_1}{\partial \upsilon_1} - \mathrm{e}^{k-r(T-t)} \frac{\partial Q_2}{\partial \upsilon_1} \\
\frac{\partial C}{\partial \upsilon_2} &= \mathrm{e}^s \frac{\partial Q_1}{\partial \upsilon_2} - \mathrm{e}^{k-r(T-t)} \frac{\partial Q_2}{\partial \upsilon_2} \\
\frac{\partial^2 C}{\partial s \partial \upsilon_1} &= \mathrm{e}^s \frac{\partial Q_1}{\partial \upsilon_1} + \mathrm{e}^s \frac{\partial^2 Q_1}{\partial s \partial \upsilon_1} - \mathrm{e}^{k-r(T-t)} \frac{\partial^2 Q_2}{\partial s \partial \upsilon_1} \\
\frac{\partial^2 C}{\partial s \partial \upsilon_2} &= \mathrm{e}^s \frac{\partial Q_1}{\partial \upsilon_2} + \mathrm{e}^s \frac{\partial^2 Q_1}{\partial s \partial \upsilon_2} - \mathrm{e}^{k-r(T-t)} \frac{\partial^2 Q_2}{\partial s \partial \upsilon_2} \\
\frac{\partial^2 C}{\partial s^2} &= \mathrm{e}^s Q_1 + 2\mathrm{e}^s \frac{\partial Q_1}{\partial s} + \mathrm{e}^s \frac{\partial^2 Q_1}{\partial s^2} - \mathrm{e}^{k-r(T-t)} \frac{\partial^2 Q_2}{\partial s^2} \\
\frac{\partial^2 C}{\partial \upsilon_1^2} &= \mathrm{e}^s \frac{\partial^2 Q_1}{\partial \upsilon_1^2} - \mathrm{e}^{k-r(T-t)} \frac{\partial^2 Q_2}{\partial \upsilon_1^2} \\
\frac{\partial^2 C}{\partial \upsilon_2^2} &= \mathrm{e}^s \frac{\partial^2 Q_1}{\partial \upsilon_2^2} - \mathrm{e}^{k-r(T-t)} \frac{\partial^2 Q_2}{\partial \upsilon_2^2}.
\end{aligned}
$$

Now, substituting the computed derivatives and (4.52) into (4.54), we obtain

$$
\begin{aligned}
&\mathrm{e}^s \frac{\partial Q_1}{\partial t} - r\mathrm{e}^{k-r(T-t)} Q_2 - \mathrm{e}^{k-r(T-t)} \frac{\partial Q_2}{\partial t} + \\
&+ \left[r - \frac{1}{2}(\upsilon_1 + \upsilon_2 + \phi_t)\right] \left[\mathrm{e}^s Q_1 + \mathrm{e}^s \frac{\partial Q_1}{\partial s} - \mathrm{e}^{k-r(T-t)} \frac{\partial Q_2}{\partial s}\right] + \\
&+ \kappa_1 (\omega_1 - \upsilon_1) \left[\mathrm{e}^s \frac{\partial Q_1}{\partial \upsilon_1} - \mathrm{e}^{k-r(T-t)} \frac{\partial Q_2}{\partial \upsilon_1}\right] + \\
&+ \kappa_2 (\omega_2 - \upsilon_2) \left[\mathrm{e}^s \frac{\partial Q_1}{\partial \upsilon_2} - \mathrm{e}^{k-r(T-t)} \frac{\partial Q_2}{\partial \upsilon_2}\right] + \\
&+ \rho_1 \Lambda_1 \upsilon_1 \left[\mathrm{e}^s \frac{\partial Q_1}{\partial \upsilon_1} + \mathrm{e}^s \frac{\partial^2 Q_1}{\partial s \partial \upsilon_1} - \mathrm{e}^{k-r(T-t)} \frac{\partial^2 Q_2}{\partial s \partial \upsilon_1}\right] + \\
&+ \rho_2 \Lambda_2 \upsilon_2 \left[\mathrm{e}^s \frac{\partial Q_1}{\partial \upsilon_2} + \mathrm{e}^s \frac{\partial^2 Q_1}{\partial s \partial \upsilon_2} - \mathrm{e}^{k-r(T-t)} \frac{\partial^2 Q_2}{\partial s \partial \upsilon_2}\right] + \\
&+ \frac{1}{2}(\upsilon_1 + \upsilon_2 + \phi_t) \left[\mathrm{e}^s Q_1 + 2\mathrm{e}^s \frac{\partial Q_1}{\partial s} + \mathrm{e}^s \frac{\partial^2 Q_1}{\partial s^2} - \mathrm{e}^{k-r(T-t)} \frac{\partial^2 Q_2}{\partial s^2}\right] + \\
&+ \frac{1}{2}\Lambda_1^2 \upsilon_1 \left[\mathrm{e}^s \frac{\partial^2 Q_1}{\partial \upsilon_1^2} - \mathrm{e}^{k-r(T-t)} \frac{\partial^2 Q_2}{\partial \upsilon_1^2}\right] + \\
&\frac{1}{2}\Lambda_2^2 \upsilon_2 \left[\mathrm{e}^s \frac{\partial^2 Q_1}{\partial \upsilon_2^2} - \mathrm{e}^{k-r(T-t)} \frac{\partial^2 Q_2}{\partial \upsilon_2^2}\right] - r\mathrm{e}^s Q_1 + r\mathrm{e}^{k-r(T-t)} Q_2 = 0,
\end{aligned}
$$

that is

$$
\begin{aligned}
&\mathrm{e}^{s}\frac{\partial Q_1}{\partial t}-\mathrm{e}^{k-r(T-t)}\frac{\partial Q_2}{\partial t}+r\left[\mathrm{e}^{s}\frac{\partial Q_1}{\partial s}-\mathrm{e}^{k-r(T-t)}\frac{\partial Q_2}{\partial s}\right]+\\
&\quad+\frac{1}{2}\left(\upsilon_1+\upsilon_2+\phi_t\right)\left[\mathrm{e}^{k-r(T-t)}\frac{\partial Q_2}{\partial s}+\mathrm{e}^{s}\frac{\partial Q_1}{\partial s}+\mathrm{e}^{s}\frac{\partial^2 Q_1}{\partial s^2}-\mathrm{e}^{k-r(T-t)}\frac{\partial^2 Q_2}{\partial s^2}\right]+\\
&\quad+\kappa_1\left(\omega_1-\upsilon_1\right)\left[\mathrm{e}^{s}\frac{\partial Q_1}{\partial \upsilon_1}-\mathrm{e}^{k-r(T-t)}\frac{\partial Q_2}{\partial \upsilon_1}\right]+\kappa_2\left(\omega_2-\upsilon_2\right)\left[\mathrm{e}^{s}\frac{\partial Q_1}{\partial \upsilon_2}-\mathrm{e}^{k-r(T-t)}\frac{\partial Q_2}{\partial \upsilon_2}\right]+\\
&\quad+\rho_1\Lambda_1\upsilon_1\left[\mathrm{e}^{s}\frac{\partial Q_1}{\partial \upsilon_1}+\mathrm{e}^{s}\frac{\partial^2 Q_1}{\partial s\partial \upsilon_1}-\mathrm{e}^{k-r(T-t)}\frac{\partial^2 Q_2}{\partial s\partial \upsilon_1}\right]+\\
&\quad+\rho_2\Lambda_2\upsilon_2\left[\mathrm{e}^{s}\frac{\partial Q_1}{\partial \upsilon_2}+\mathrm{e}^{s}\frac{\partial^2 Q_1}{\partial s\partial \upsilon_2}-\mathrm{e}^{k-r(T-t)}\frac{\partial^2 Q_2}{\partial s\partial \upsilon_2}\right]+\\
&\quad+\frac{1}{2}\Lambda_1^2\upsilon_1\left[\mathrm{e}^{s}\frac{\partial^2 Q_1}{\partial \upsilon_1^2}-\mathrm{e}^{k-r(T-t)}\frac{\partial^2 Q_2}{\partial \upsilon_1^2}\right]+\frac{1}{2}\Lambda_2^2\upsilon_2\left[\mathrm{e}^{s}\frac{\partial^2 Q_1}{\partial \upsilon_2^2}-\mathrm{e}^{k-r(T-t)}\frac{\partial^2 Q_2}{\partial \upsilon_2^2}\right]=0,
\end{aligned}
$$

and then

$$
\begin{aligned}
&\mathrm{e}^{s}\left\{\frac{\partial Q_1}{\partial t}+\left[r+\frac{1}{2}\left(\upsilon_1+\upsilon_2+\phi_t\right)\right]\frac{\partial Q_1}{\partial s}+\left[\kappa_1\left(\omega_1-\upsilon_1\right)+\rho_1\Lambda_1\upsilon_1\right]\frac{\partial Q_1}{\partial \upsilon_1}+\right.\\
&\quad+\left[\kappa_2\left(\omega_2-\upsilon_2\right)+\rho_2\Lambda_2\upsilon_2\right]\frac{\partial Q_1}{\partial \upsilon_2}+\rho_1\Lambda_1\upsilon_1\frac{\partial^2 Q_1}{\partial s\partial \upsilon_1}+\rho_2\Lambda_2\upsilon_2\frac{\partial^2 Q_1}{\partial s\partial \upsilon_2}+\\
&\quad\left.+\frac{1}{2}\left(\upsilon_1+\upsilon_2+\phi_t\right)\frac{\partial^2 Q_1}{\partial s^2}+\frac{1}{2}\Lambda_1^2\upsilon_1\frac{\partial^2 Q_1}{\partial \upsilon_1^2}+\frac{1}{2}\Lambda_2^2\upsilon_2\frac{\partial^2 Q_1}{\partial \upsilon_2^2}\right\}+\\
&\quad-\mathrm{e}^{k-r(T-t)}\left\{\frac{\partial Q_2}{\partial t}+\left[r-\frac{1}{2}\left(\upsilon_1+\upsilon_2+\phi_t\right)\right]\frac{\partial Q_2}{\partial s}+\kappa_1\left(\omega_1-\upsilon_1\right)\frac{\partial Q_2}{\partial \upsilon_1}+\right.\\
&\quad+\kappa_2\left(\omega_2-\upsilon_2\right)\frac{\partial Q_2}{\partial \upsilon_2}+\rho_1\Lambda_1\upsilon_1\frac{\partial^2 Q_2}{\partial s\partial \upsilon_1}+\rho_2\Lambda_2\upsilon_2\frac{\partial^2 Q_2}{\partial s\partial \upsilon_2}+\\
&\quad\left.+\frac{1}{2}\left(\upsilon_1+\upsilon_2+\phi_t\right)\frac{\partial^2 Q_2}{\partial s^2}+\frac{1}{2}\Lambda_1^2\upsilon_1\frac{\partial^2 Q_2}{\partial \upsilon_1^2}+\frac{1}{2}\Lambda_2^2\upsilon_2\frac{\partial^2 Q_2}{\partial \upsilon_2^2}\right\}=0.
\end{aligned}
$$

A standard separation argument decomposes the last PDE into a pair of PDEs for Q_1 and Q_2, both of the form (4.36).

Also the characteristic functions of the two probabilities $f_j\left(t,s,\upsilon_1,\upsilon_2,z\right)$, $j=1,2$, are solutions of the same PDEs, but with terminal conditions given by

$$f_j\left(T,s_T,\upsilon_{1,T},\upsilon_{2,T},z\right)=\mathrm{e}^{izs_T},\qquad j=1,2. \tag{4.56}$$

Since the PDEs (4.36) are linear in the state variables υ_1 and υ_2 and do not contain the state variable s, while the terminal conditions (4.56) are log-linear in s and do not depend on υ_1 nor υ_2, we have

$$f_j\left(t,s,\upsilon_1,\upsilon_2,z\right)=\mathrm{e}^{A_j(\tau,z)+B_j(\tau,z)\upsilon_1+C_j(\tau,z)\upsilon_2+izs},\qquad j=1,2, \tag{4.57}$$

where $\tau=T-t$. Then, the functions $\tau\to A_j\left(\tau,z\right)$, $\tau\to B_j\left(\tau,z\right)$ and $\tau\to C_j\left(\tau,z\right)$, $j=1,2$, are solutions of the ODEs

$$\begin{cases} \dfrac{\partial A_j(\tau,z)}{\partial t} = -izr - \kappa_1\omega_1 B_j - \kappa_2\omega_2 C_j - \dfrac{1}{2}z(ia_j - z)\phi_t \\ A(0,z) = 0 \end{cases}, \tag{4.58}$$

$$\begin{cases} \dfrac{\partial B_j(\tau,z)}{\partial t} = -\dfrac{1}{2}z(ia_j - z) + [\kappa_1 - (b_j + iz)\rho_1\Lambda_1]B_j - \dfrac{1}{2}\Lambda_1^2 B_j^2 \\ B(0,z) = 0 \end{cases}, \tag{4.59}$$

and

$$\begin{cases} \dfrac{\partial C_j(\tau,z)}{\partial t} = -\dfrac{1}{2}z(ia_j - z) + [\kappa_2 - (b_j + iz)\rho_2\Lambda_2]C_j - \dfrac{1}{2}\Lambda_2^2 C_j^2 \\ C(0,z) = 0 \end{cases}, \tag{4.60}$$

for fixed $\tau \geq 0$ and z. Once again, the ODE (4.59) is the same Riccati equation that we find in the original Heston [97] model and, from the previous section, we know that its solution is given by (4.41). Since the ODE (4.60) has the same form of (4.59), then its solution is

$$C_j(\tau,z) = \frac{c_{j,2} + d_{j,2}}{\Lambda_2^2}\frac{1 - \mathrm{e}^{d_{j,2}\tau}}{1 - g_{j,2}\mathrm{e}^{d_{j,2}\tau}}, \qquad j = 1,2, \tag{4.61}$$

where

$$c_{j,2} = \kappa_2 - (b_j + iz)\rho_2\Lambda_2, \qquad d_{j,2} = \sqrt{c_{j,2}^2 - z(ia_j - z)\Lambda_2^2}, \qquad g_{j,2} = \frac{c_{j,2} + d_{j,2}}{c_{j,2} - d_{j,2}}.$$

Substituting (4.41) and (4.61) into (4.39), we can solve it by direct integration, that is

$$A_j(\tau,z) = \int_t^T [riz + \kappa_1\omega_1 B_j(u,z)]\,du + \kappa_2\omega_2 \int_t^T C_j(u,z)\,du + \frac{1}{2}z(ia_j - z)\int_t^T \phi_u du.$$

The sum between the first and the third integral yields Equation (4.42), which we have already computed for the one-factor Heston++ model. On the other hand, the second integral gives a solution similar to the second addend of (4.42), apart from different coefficients. Hence, we have

$$\begin{aligned} A_j(\tau,z) = riz\tau &+ \frac{\kappa_1\omega_1}{\Lambda_1^2}\left[(c_{j,1} + d_{j,1})\tau - 2\ln\left(\frac{1 - g_{j,1}\mathrm{e}^{d_{j,1}\tau}}{1 - g_{j,1}}\right)\right] + \\ &+ \frac{\kappa_2\omega_2}{\Lambda_2^2}\left[(c_{j,2} + d_{j,2})\tau - 2\ln\left(\frac{1 - g_{j,2}\mathrm{e}^{d_{j,2}\tau}}{1 - g_{j,2}}\right)\right] + \frac{1}{2}z(ia_j - z)\int_t^T \phi_u du. \end{aligned} \tag{4.62}$$

Note that, if $\phi_t = 0$, we then obtain the original two-factor version of the Heston model.

Therefore, we can write

$$\begin{aligned}
A_j(\tau,z) &= A_j^{H_2}(\tau,z) + \frac{1}{2}z(ia_j - z)\int_t^T \phi_u du,\\
B_j(\tau,z) &= B_j^H(\tau,z) = B_j^{H_2}(\tau,z),\\
C_j(\tau,z) &= C_j^{H_2}(\tau,z).
\end{aligned}$$

Details 5. *The relevant partial derivatives of (4.57) are given by*

$$\begin{aligned}
&\frac{\partial f_j}{\partial t} = \left(\frac{\partial A_j}{\partial t} + \frac{\partial B_j}{\partial t}\upsilon_1 + \frac{\partial C_j}{\partial t}\upsilon_2\right) f_j, && \frac{\partial f_j}{\partial s} = izf_j, && \frac{\partial f_j}{\partial \upsilon_1} = B_j f_j,\\
&\frac{\partial f_j}{\partial \upsilon_2} = C_j f_j, && \frac{\partial^2 f_j}{\partial s \partial \upsilon_1} = izB_j f_j, && \frac{\partial^2 f_j}{\partial s \partial \upsilon_2} = izC_j f_j,\\
&\frac{\partial^2 f_j}{\partial s^2} = -z^2 f_j, && \frac{\partial^2 f_j}{\partial \upsilon_1^2} = B_j^2 f_j, && \frac{\partial^2 f_j}{\partial \upsilon_2^2} = C_j^2 f_j,
\end{aligned}$$

with $j = 1, 2$. Substituting them into the PDE, we obtain

$$\begin{aligned}
&\left(\frac{\partial A_j}{\partial t} + \frac{\partial B_j}{\partial t}\upsilon_1 + \frac{\partial C_j}{\partial t}\upsilon_2\right) f_j + \left[r + \frac{1}{2}a_j(\upsilon_1 + \upsilon_2 + \phi_t)\right] izf_j +\\
&\qquad + \left[\kappa_1\omega_1 - (\kappa_1 - b_j\rho_1\Lambda_1)\upsilon_1\right] B_j f_j + \left[\kappa_2\omega_2 - (\kappa_2 - b_j\rho_2\Lambda_2)\upsilon_2\right] C_j f_j +\\
&+ iz\rho_1\Lambda_1\upsilon_1 B_j f_j + iz\rho_2\Lambda_2\upsilon_2 C_j f_j - \frac{1}{2}(\upsilon_1 + \upsilon_2 + \phi_t) z^2 f_j + \frac{1}{2}\Lambda_1^2\upsilon_1 B_j^2 f_j + \frac{1}{2}\Lambda_2^2\upsilon_2 C_j^2 f_j = 0,
\end{aligned}$$

that is

$$\begin{aligned}
&\frac{\partial A_j}{\partial t} + izr + \frac{1}{2}iza_j\phi_t + \kappa_1\omega_1 B_j + \kappa_2\omega_2 C_j - \frac{1}{2}z^2\phi_t +\\
&\qquad + \left[\frac{\partial B_j}{\partial t} + \frac{1}{2}iza_j - (\kappa_1 - b_j\rho_1\Lambda_1) B_j + iz\rho_1\Lambda_1 B_j - \frac{1}{2}z^2 + \frac{1}{2}\Lambda_1^2 B_j^2\right]\upsilon_1 +\\
&\qquad\qquad + \left[\frac{\partial C_j}{\partial t} + \frac{1}{2}iza_j - (\kappa_2 - b_j\rho_2\Lambda_2) C_j + iz\rho_2\Lambda_2 C_j - \frac{1}{2}z^2 + \frac{1}{2}\Lambda_2^2 C_j^2\right]\upsilon_2 = 0.
\end{aligned}$$

A standard separation argument decomposes this equation into a triple of ODEs of the form (4.58), (4.59) and (4.60). □

The two probabilities in (4.52) and the corresponding option price can be obtained numerically by the inverse transforms

$$Q_j(t,s,\upsilon_1,\upsilon_2,k) = \frac{1}{2} + \frac{1}{\pi}\int_0^\infty \Re\left[\frac{\mathrm{e}^{-izk} f_j(t,s,\upsilon_1,\upsilon_2,z)}{iz}\right] dz, \qquad j = 1, 2. \tag{4.63}$$

Furthermore, the expected risk-neutral integrated variance in the two-factor Heston++ model can be rewritten as

$$IV_t(T_1,T_2) = IV_t^{H_2}(T_1,T_2) + \frac{1}{T_2 - T_1} I_\phi(T_1,T_2), \tag{4.64}$$

that is, the sum of the standard two-factor Heston integrated variance plus the contribution of the deterministic part of the variance, given by the average of ϕ_t over $[T_1, T_2]$.

4.2.3 Two factors with jumps

The same idea can be used with multi-factor affine models with jumps, as those proposed by Bates [37] and [38] and Duffie, Pan and Singleton [79]. For instance, if we need a two-factor model with jumps, the risk-neutral dynamics of the model would read

$$dS_t = \left(r - \bar{k}_J \lambda\right) S_t dt + \sqrt{v_{1,t} + \phi_t} S_t dW_{1,t}^S + \sqrt{v_{2,t}} S_t dW_{2,t}^S + k_J S_t dN_t, \tag{4.65}$$

$$dv_{j,t} = \kappa_j (\omega_j - v_{j,t}) dt + \Lambda_j \sqrt{v_{j,t}} dW_{j,t}^v, \qquad j = 1, 2, \tag{4.66}$$

where $\left(W_{1,t}^S\right)_{t \geq 0} \equiv W_{1,t}^S$, $\left(W_{2,t}^S\right)_{t \geq 0} \equiv W_{2,t}^S$, $\left(W_{1,t}^v\right)_{t \geq 0} \equiv W_{1,t}^v$ and $\left(W_{2,t}^v\right)_{t \geq 0} \equiv W_{2,t}^v$ are Wiener processes and $(N_t)_{t \geq 0} \equiv N_t$ is an independent Poisson process, with constant intensity λ and random jump size k_J, distributed as

$$\log\left(1 + k_J\right) \sim \mathcal{N}\left(\log\left(1 + \bar{k}_J\right) - \frac{1}{2}\sigma_J^2, \sigma_J^2\right). \tag{4.67}$$

Risk-neutral correlation between the driving Wiener processes are all zero, except

$$\operatorname{corr}^{\mathbb{Q}}\left(dW_{1,t}^S, dW_{1,t}^v\right) = \rho_1 \sqrt{\frac{v_{1,t}}{v_{1,t} + \phi_t}}, \qquad \operatorname{corr}^{\mathbb{Q}}\left(dW_{2,t}^S, dW_{2,t}^v\right) = \rho_2. \tag{4.68}$$

It is worth noting that with $\phi_t = 0$, the model (4.65)-(4.66) includes, as particular cases, both a simplified version of Bates [38] with constant λ and the model of Christoffersen, Heston and Jacobs [67] with jumps, as in Bates [37].

In the multi-factor model with jumps, the log-price s_t satisfies the SDE

$$ds_t = \left[r - \bar{k}_J \lambda - \frac{1}{2}\left(v_{1,t} + v_{2,t} + \phi_t\right)\right] dt + \sqrt{v_{1,t} + \phi_t} dW_{1,t}^S + \sqrt{v_{2,t}} dW_{2,t}^S + k_J dN_t. \tag{4.69}$$

Using affinity of the model, the option pricing formula is still of the form

$$C\left(t, s_t, v_{1,t}, v_{2,t}, k\right) = \mathrm{e}^{s_t} Q_1\left(t, s, v_1, v_2, k\right) - \mathrm{e}^{k - r(T-t)} Q_2\left(t, s, v_1, v_2, k\right), \tag{4.70}$$

where, for every $s, v_1, v_2 > 0$,

$$Q_1\left(t, s, v_1, v_2, k\right) = \mathbb{Q}^S\left[s_T \geq k \mid s_t = s, v_{1,t} = v_1, v_{2,t} = v_2\right]$$
$$Q_2\left(t, s, v_1, v_2, k\right) = \mathbb{Q}^T\left[s_T \geq k \mid s_t = s, v_{1,t} = v_1, v_{2,t} = v_2\right],$$

and the two probabilities have the same meaning as before. Of course, we have

$$Q_j\left(t, s, v_1, v_2, k\right) = \frac{1}{2} + \frac{1}{\pi} \int_0^\infty \Re\left[\frac{\mathrm{e}^{-izk} f_j\left(t, s, v_1, v_2, z\right)}{iz}\right] dz, \qquad j = 1, 2, \tag{4.71}$$

where $f_j(t, s, \upsilon_1, \upsilon_2, z)$, $j = 1, 2$, are the characteristic functions of the two probabilities. Due to the affinity of the model, their closed form expressions are obtained by the following steps[4]:

Step 1 Using independence between the continuous part of the model and its jump component, we first factor

$$f_j(t, s, \upsilon_1, \upsilon_2, z) = f_j^0(t, s, \upsilon_1, \upsilon_2, z)\, f_j^J(t, z), \qquad j = 1, 2, \tag{4.72}$$

where $f_j^0(t, s, \upsilon_1, \upsilon_2, z)$, $j = 1, 2$, are the characteristic functions in the no-jump case, that is $\lambda = 0$, and where

$$f_j^J(t, z) = \exp\left\{\lambda(T - t)\, h_j \left[\mathrm{e}^{i\left(\log(1+\bar{k}_J) + \frac{1}{2}a_j\sigma_J^2\right)z - \frac{1}{2}\sigma_J^2 z^2} - 1\right]\right\}, \qquad j = 1, 2,$$

with $h_1 = \log(1 + \bar{k}_J)$, $h_2 = 1$, $a_1 = 1$, $a_2 = -1$, are the contributions of the jump component.

Step 2 Due to the affinity of the model, the no-jump characteristic functions have the form

$$f_j^0(t, s, \upsilon_1, \upsilon_2, z) = \mathrm{e}^{A_j(\tau,z) + B_j(\tau,z)\upsilon_1 + C_j(\tau,z)\upsilon_2 + izs}, \qquad j = 1, 2, \tag{4.73}$$

where $\tau = T - t$. Then, the functions $\tau \to A_j(\tau, z)$, $\tau \to B_j(\tau, z)$ and $\tau \to C_j(\tau, z)$, $j = 1, 2$, are solutions of the ODEs, for fixed $\tau \geq 0$ and z,

$$\begin{cases} \dfrac{\partial A_j(\tau, z)}{\partial t} = -izr - \kappa_1\omega_1 B_j - \kappa_2\omega_2 C_j - \dfrac{1}{2}z(ia_j - z)\phi_t \\ A(0, z) = 0 \end{cases}, \tag{4.74}$$

$$\begin{cases} \dfrac{\partial B_j(\tau, z)}{\partial t} = -\dfrac{1}{2}z(ia_j - z) + [\kappa_1 - (b_j + iz)\rho_1\Lambda_1] B_j - \dfrac{1}{2}\Lambda_1^2 B_j^2 \\ B(0, z) = 0 \end{cases}, \tag{4.75}$$

and

$$\begin{cases} \dfrac{\partial C_j(\tau, z)}{\partial t} = -\dfrac{1}{2}z(ia_j - z) + [\kappa_2 - (b_j + iz)\rho_2\Lambda_2] C_j - \dfrac{1}{2}\Lambda_2^2 C_j^2 \\ C(0, z) = 0 \end{cases}, \tag{4.76}$$

where $a_1 = 1$, $a_2 = -1$, $b_1 = 1$, $b_2 = 0$.

Step 3 Since ODEs (4.75)-(4.76) have the same form of (4.59)-(4.60), we know that their solutions are given by Equations (4.41) and (4.61), respectively. Thus, substituting (4.41) and (4.61) into (4.74), we solve it by direct integration and we obtain

$$A_j(\tau, z) = \int_t^T [izr + \kappa_1\omega_1 B_j(u, z)]\, du + \kappa_2\omega_2 \int_t^T C_j(u, z)\, du + \frac{1}{2}z(ia_j - z)\int_t^T \phi_u du.$$

The sum of the first and the third integral yields a result similar to (4.42) we already

[4] See also Bates [38] and Christoffersen, Heston and Jacobs [67]

computed for the univariate Heston++ model, whereas the second integral is similar to the second addend of (4.42). Therefore, we have

$$A_j(\tau,z) = riz\tau + \frac{\kappa_1\omega_1}{\Lambda_1^2}\left[(c_{j,1}+d_{j,1})\tau - 2\ln\left(\frac{1-g_{j,1}\mathrm{e}^{d_{j,1}\tau}}{1-g_{j,1}}\right)\right] + \\ + \frac{\kappa_2\omega_2}{\Lambda_2^2}\left[(c_{j,2}+d_{j,2})\tau - 2\ln\left(\frac{1-g_{j,2}\mathrm{e}^{d_{j,2}\tau}}{1-g_{j,2}}\right)\right] + \frac{1}{2}z(ia_j - z)\int_t^T \phi_u du. \quad (4.77)$$

As a final remark, for $\phi_t = 0$ and no jumps, that is $\lambda = 0$, we obtain the Christoffersen, Heston and Jacobs [67] model and, then, we can write

$$\begin{aligned} A_j(\tau,z) &= A_j^{H_2}(\tau,z) + \frac{1}{2}z(ia_j - z)\int_t^T \phi_u du, \\ B_j(\tau,z) &= B_j^{H}(\tau,z) = B_j^{H_2}(\tau,z), \\ C_j(\tau,z) &= C_j^{H_2}(\tau,z), \end{aligned}$$

where the superscript H_2 is used to denote the corresponding functions of that model, which depend on t and T only trough their difference τ. Finally, we have

$$f_j^0(t,s,\upsilon_1,\upsilon_2,z) = f_j^{H_2}(t,s,\upsilon_1,\upsilon_2,z)\,\mathrm{e}^{\frac{1}{2}z(a_j i - z)I_\phi(t,T)}, \qquad j = 1,2.$$

4.3 Application to FX options

We calibrate the Heston++ model to a sample of FX option prices on the EUR/USD currency. To this end, we just need to replace the risk-free interest rate r with $r^d - r^f$ in Equation (4.24), where r^d is the domestic interest rate and r^f is the foreign rate (see Garman and Kohlhagen [91]), and then discount at the domestic rate, that is, the Euro in our case.

The data set is composed of daily mid-quotes from 12 December 2005 to 12 November 2012, for a total of 1,806 days. Each day, we have options quoted for *10 maturities*, ranging from one week to two years, and for *5 strikes*: at-the-money, in-the-money and out-of-the-money at 10Δ and 25Δ. Hence, in total our daily observed volatility surface consists of 50 observations. As a benchmark model, we fit the standard Heston [97] model its multivariate counterparts, obtained setting $\phi_t = 0$. Thus, we fit one-factor models, labelled as *1f + +* and *1f*, two-factor models, *2f + +* and *2f*, and two-factor models with jumps, *2fj + +* and *2fj*.

When fitting the Heston++ model, Equation (4.34) shows that what is actually needed are the integrals $I_\phi(t,T)$. We treat these integrals as additional parameters, one for each maturity, that is 10, in this empirical exercise. Furthermore, these additional parameters are constrained to be positive. An alternative choice could be imposing a parametric structure to the function $I_\phi(t,T)$, for example with the Nelson-Siegel-Svensson parameterization.

We fit the model by minimizing the weighted mean square error on volatilities. Weights are used to increase the relative importance of short maturities with respect to longer ones. With no weights, we would obtain a lower mean square error error but a higher maximum error. In

	first volatility factor					second volatility factor					jump component		
	κ_1	ω_1	Λ_1	ρ_1	$v_{0,1}$	κ_2	ω_2	Λ_2	ρ_2	$v_{0,2}$	$\bar{k}_J$	σ_J	λ
1f	18.213	0.022	0.921	-0.158	0.015	-	-	-	-	-	-	-	-
	(32.644)	(0.024)	(0.517)	(0.209)	(0.015)								
1f + +	15.174	0.020	1.277	-0.187	0.012	-	-	-	-	-	-	-	-
	(27.800)	(0.024)	(1.135)	(0.253)	(0.012)								
2f	3.748	0.018	0.729	-0.127	0.009	0.901	0.015	0.983	-0.367	0.006	-	-	-
	(4.047)	(0.061)	(1.401)	(0.373)	(0.011)	(0.963)	(0.038)	(1.166)	(0.415)	(0.006)			
2f + +	3.746	0.017	0.903	-0.149	0.008	0.882	0.018	1.217	-0.372	0.006	-	-	-
	(4.696)	(0.061)	(2.027)	(0.402)	(0.010)	(1.006)	(0.044)	(1.707)	(0.436)	(0.006)			
2fj	3.704	0.018	0.762	-0.087	0.009	0.873	0.014	0.994	-0.287	0.007	0.295	0.630	0.018
	(4.326)	(0.055)	(1.457)	(0.384)	(0.011)	(0.972)	(0.035)	(1.234)	(0.407)	(0.006)	(0.765)	(2.844)	(0.093)
2fj + +	3.695	0.016	0.884	-0.114	0.008	0.875	0.016	1.200	-0.288	0.006	0.281	0.695	0.017
	(4.621)	(0.054)	(1.592)	(0.406)	(0.010)	(1.308)	(0.037)	(1.709)	(0.422)	(0.006)	(0.774)	(3.217)	(0.086)

Table 4.2: Summary statistics for all six models parameters. Parameters are for yearly units time.

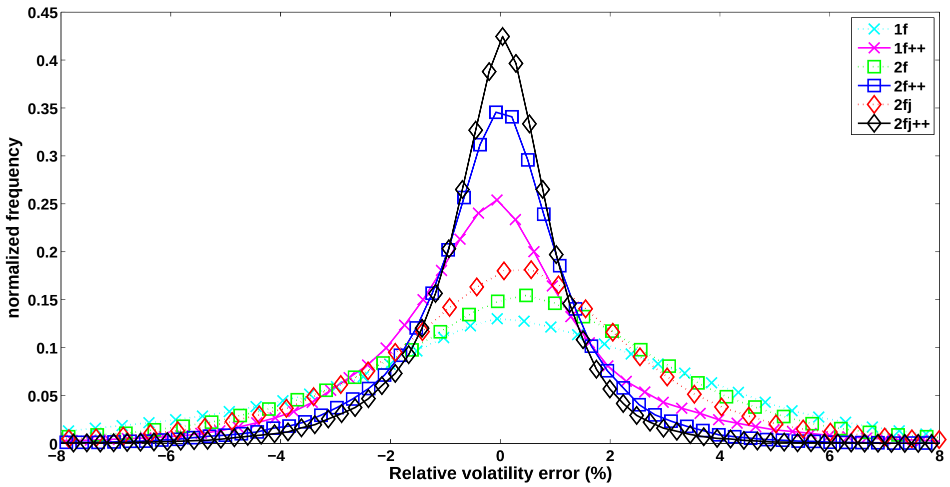

Figure 4.1: Daily relative fitting error of the competing models. Quotes are in terms of implied volatilities.

fact, if the goal of the optimization was the minimization of the pricing error alone, then the weight matrix would be the identity. In practice, the decision about the weights to use depends on specific needs. Summary statistics of daily estimated parameters is provided in Table 4.2.

Figure 4.1 shows the relative pricing error in terms of volatility on all $50 \times 1,806$ options. It is striking to note that even the 1f + + model performs better than the 2fj model: the deterministic shift extension increases the flexibility of affine models dramatically. When the 2fj ++ model is used, only 0.67% (resp. 0.05%) of the options have a relative pricing error greater than 5% (resp. 8%).

Figure 4.2 shows the relative, with respect to volatility, daily Root Mean Square Error (RMSE) when pricing market data, using the bivariate models with jumps. In terms of quoted volatility, Heston++ models can reduce the average calibration error substantially: from 3.60% of 1f to the 2.13% of 1f + +; from 3.01% of 2f to 1.51% of 2f + +; and from 2.64% of 2fj to 1.26% of 2fj + +. The advantage of the Heston++ is particularly pronounced in the middle of

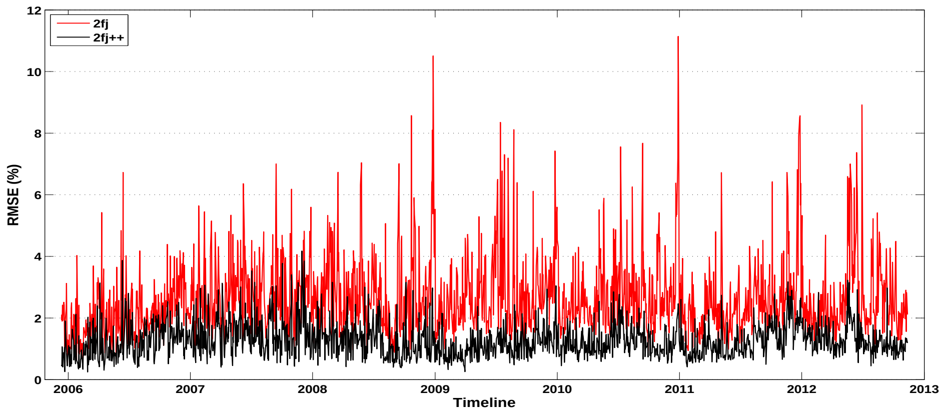

Figure 4.2: Daily relative Root Mean Square Error for 2fj + + and 2fj models.

the credit crunch crisis, as shown below.

Figure 4.3 shows a market day in which the term structure of at-the-money volatilities is humped. Standard affine models cannot reproduce this shape and, as a consequence, they fail in fitting the volatility surface. The fit across maturities is continuously above and below the observed market data, in the attempt of minimizing the overall pricing error. On the contrary, the Heston++ model with one factor, which is designed exactly to capture the term structure, provides an excellent fit across all maturities, with a negligible pricing error.

Conversely, in Figure 4.4 we show a day in which one factor is not enough. The error is almost totally driven by short-term options, which present no skew at all, in contrast to a prominent skew at higher maturity. As in the traditional Heston [97] model, in the Heston++ model the skew is controlled mainly by a single parameter, that is the leverage coefficient. Hence, if the skew, or the convexity, changes across maturities, the solution consists in adding a second stochastic factor and jumps.

Figure 4.5 presents the same day of Figure 4.4, but for the multifactor case. We can see that both Heston and Heston++ fit better than the previous case, thanks to the addition of a second stochastic factor. Roughly speaking, the Root Mean Square Error decreases from 16.23% to 11.40% for the standard Heston and from 5.64% to 3.82% for Heston++. Nevertheless, even with a second factor, the high skewness of the long-term maturities compromises the fit at lower maturities. Hence, this is a proper case in which jumps are needed to be considered.

Figure 4.6 shows the fit of the two-factor Heston++ model with jumps described in Section 4.2.3, together with the two-factor Heston model with jumps, that is, the Heston++ model with $\phi_t = 0$, used as a benchmark. Despite the fact that the two-factor Heston model, used for example by Bates [38], has 13 parameters, it still cannot fit the volatility surface, since it cannot account for the term structure of ATM volatilities. The Heston++ model, designed purposely to fit ATM volatilities automatically, manages to fit all strikes and maturities with a maximum relative error of 3.59%. Adding a deterministic shift extension to the multifactor models with jumps comes at no additional computational cost, as shown here: calibration on a volatility surface requires well less than one minute on a laptop bought in 2009.

Thus, the point we want to highlight here is the following.

Remark 4.2. *Whatever the number of purely stochastic factors you need to introduce in your model in order to account for time-varying skewness and/or time-varying convexity in a panel of options, adding a deterministic shift will always provide an automatic fit of the term structure at no additional computational cost, so that the factors or jumps added can be effectively used for what they are meant.*

Finally, Figures 4.7, 4.8, 4.9 and 4.10 show the daily estimates of all the parameters obtained from the calibration of the six models compared. Furthermore, Figure 4.11 display the daily estimates of the vector $I_\phi(t,T)$, as defined in Equation (4.30), for all the three specifications of the Heston++ model.

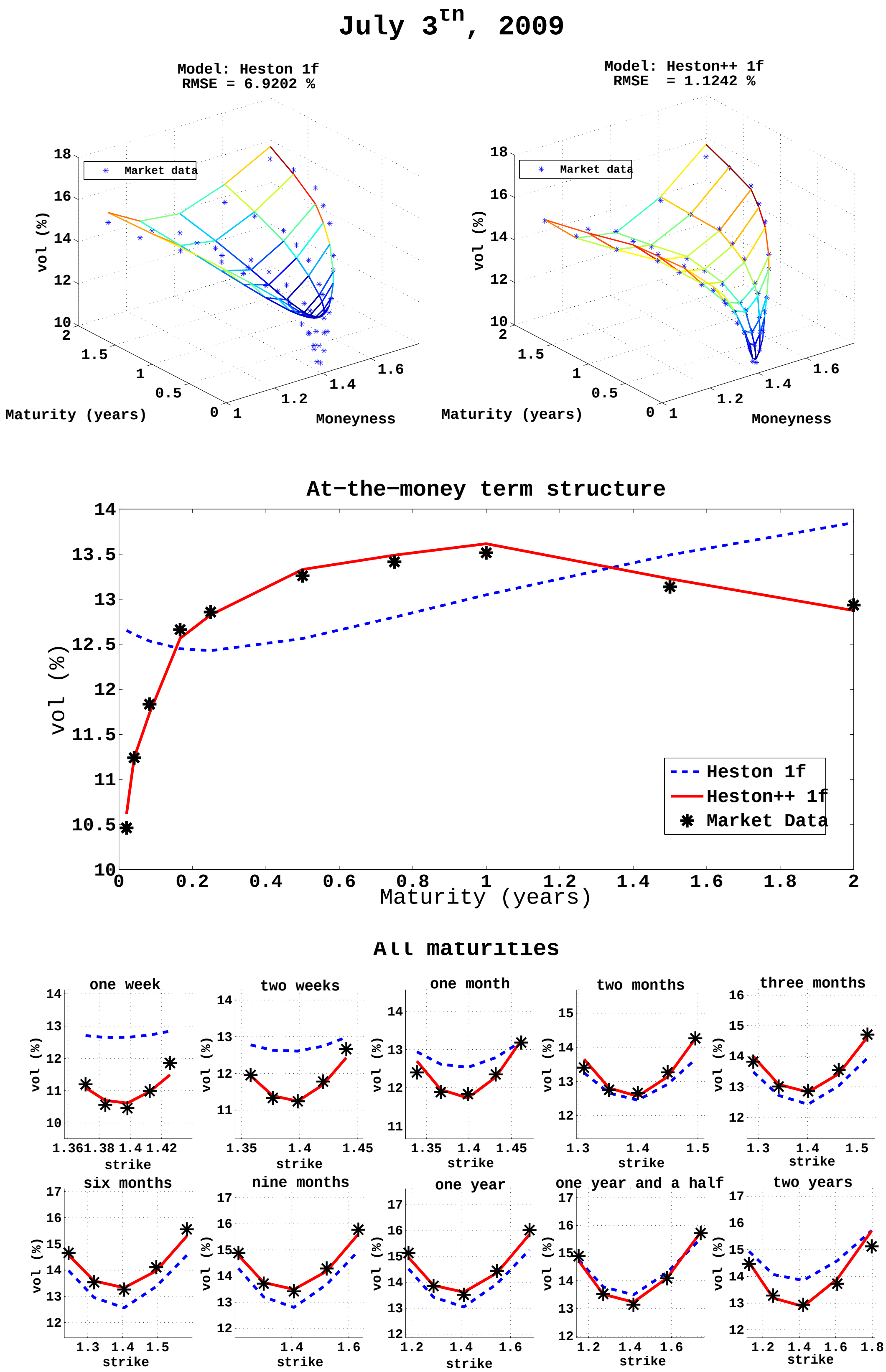

Figure 4.3: A comparison between Heston and Heston++ with one factor on July 3^{th}, 2009

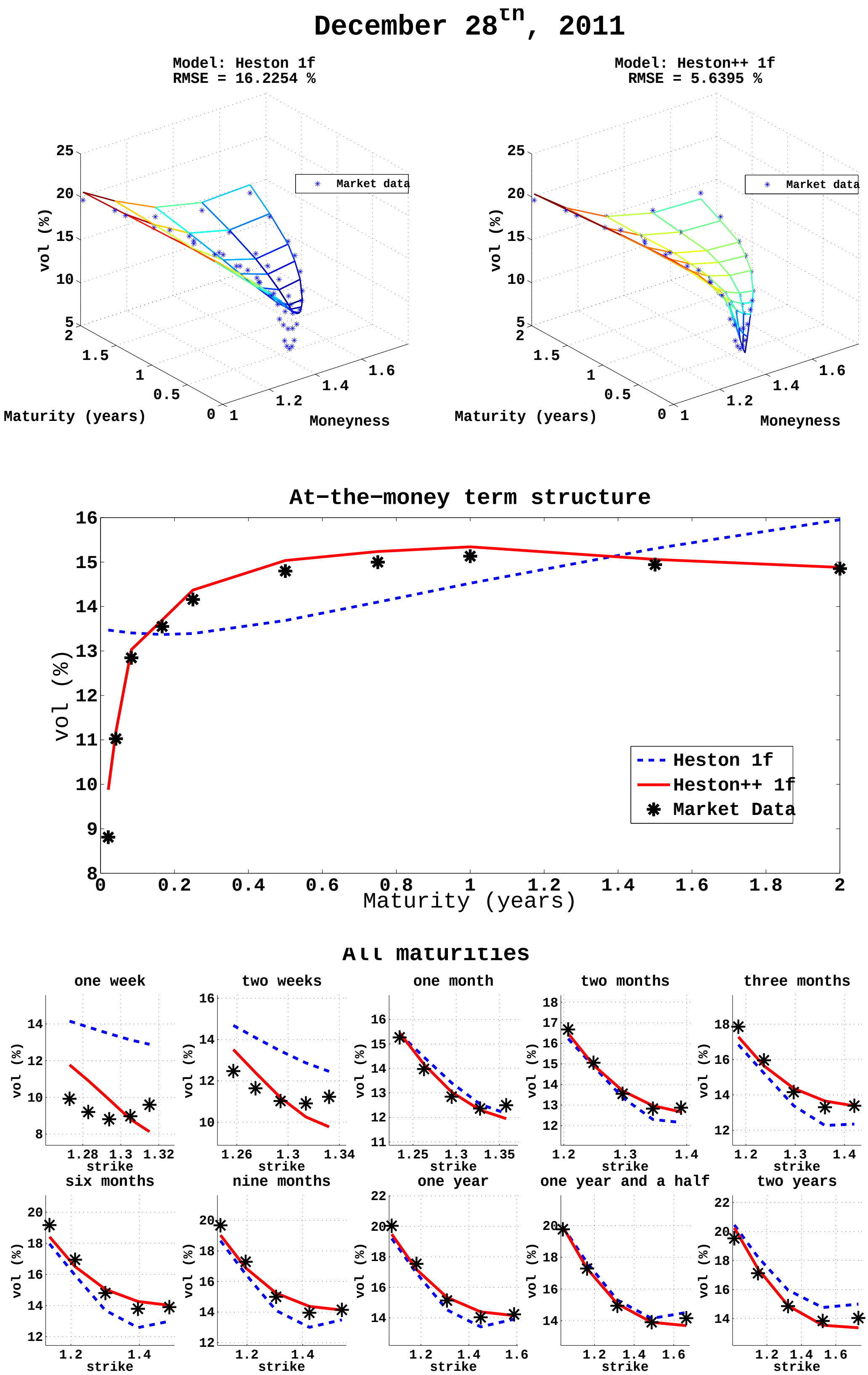

Figure 4.4: A comparison between Heston and Heston++ with one factor on December 28^{th}, 2011

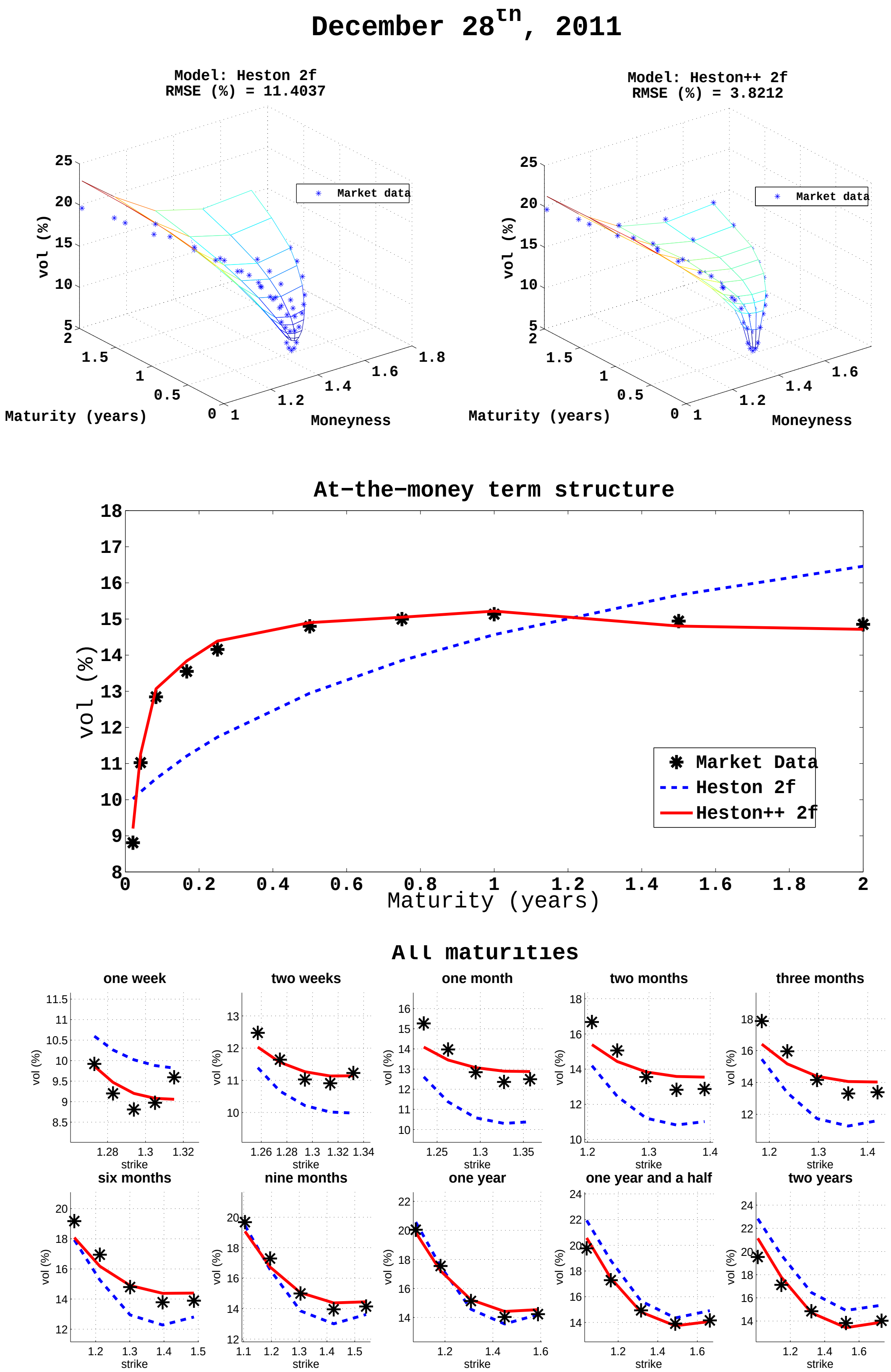

Figure 4.5: A comparison between multifactor Heston and Heston++ on December 28^{th}, 2011

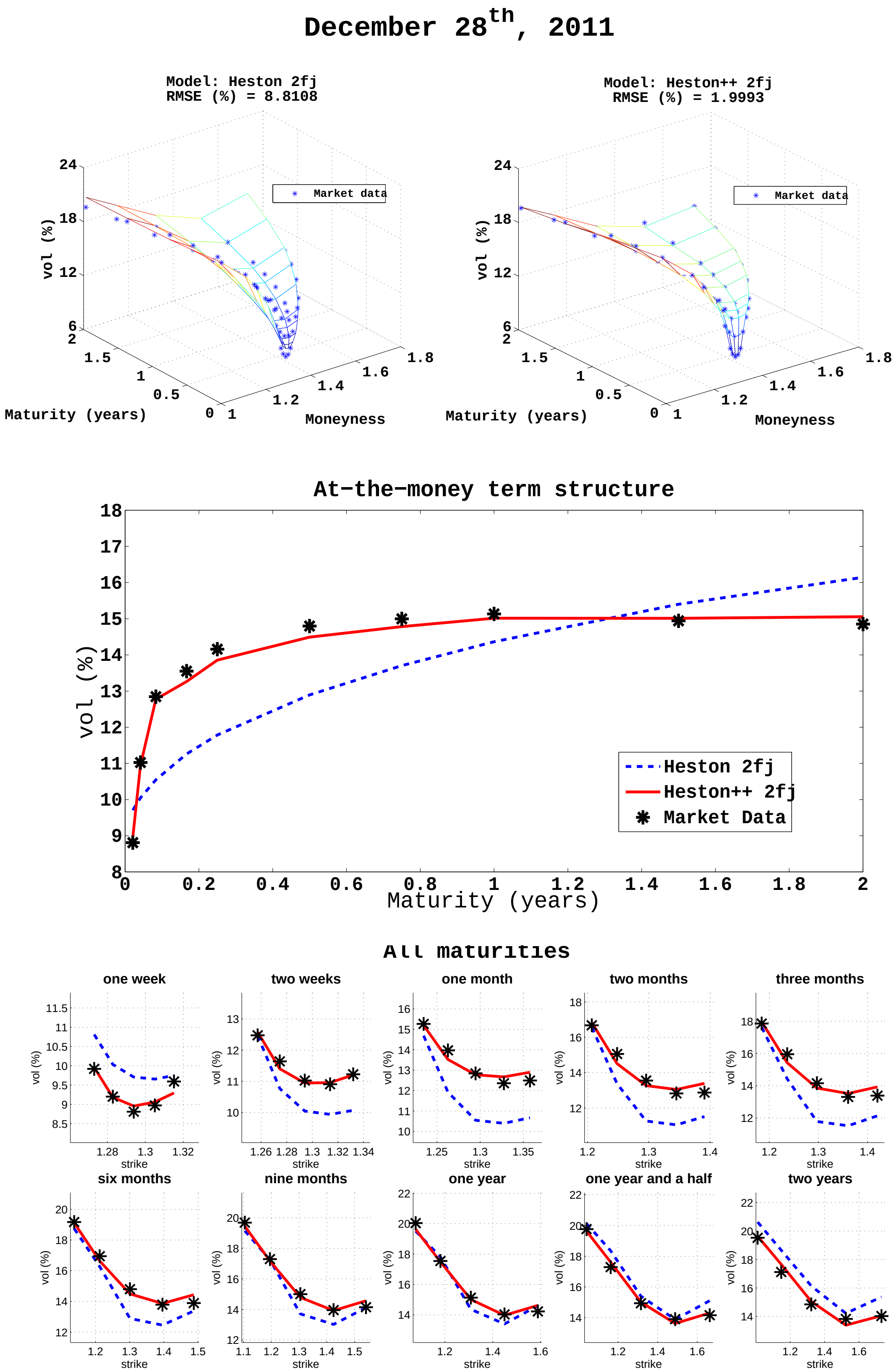

Figure 4.6: A comparison between multifactor Heston and Heston++ with jumps on December 28^{th}, 2011

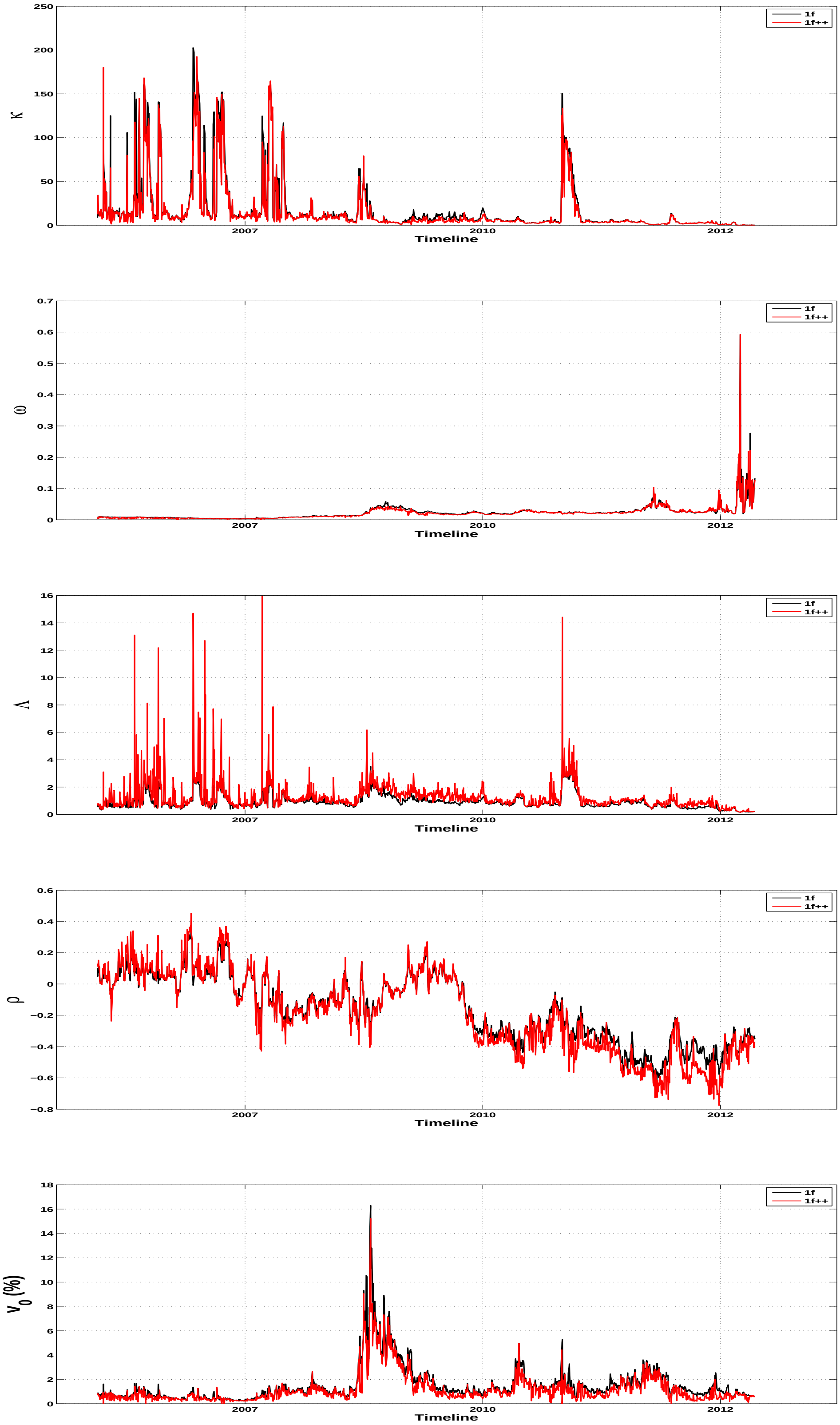

Figure 4.7: Daily parameters series for Heston and Heston++ models with one factor

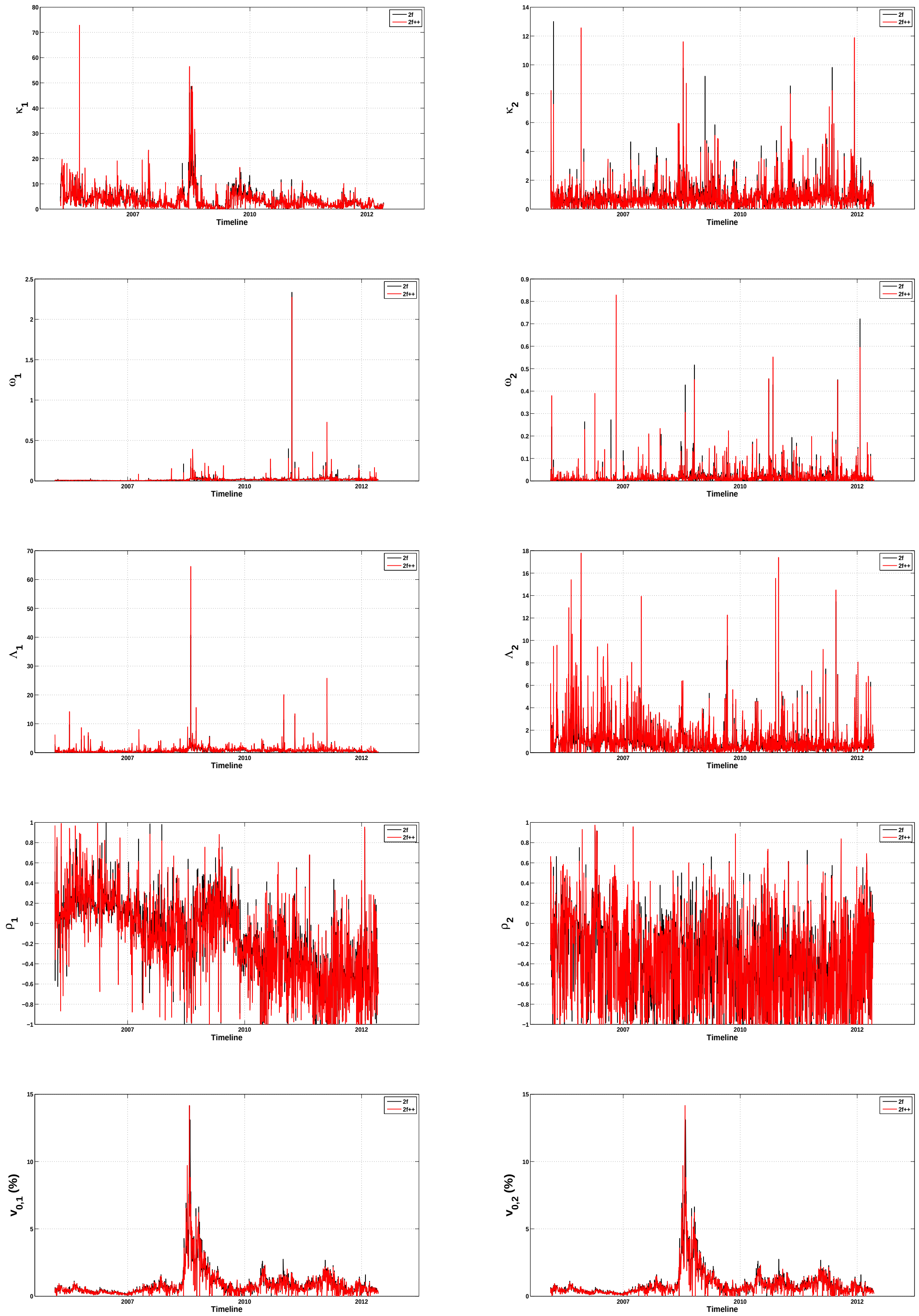

Figure 4.8: Daily parameters series of the first volatility factor (*left*) and second volatility factor (*right*) for multi-factor Heston and Heston++

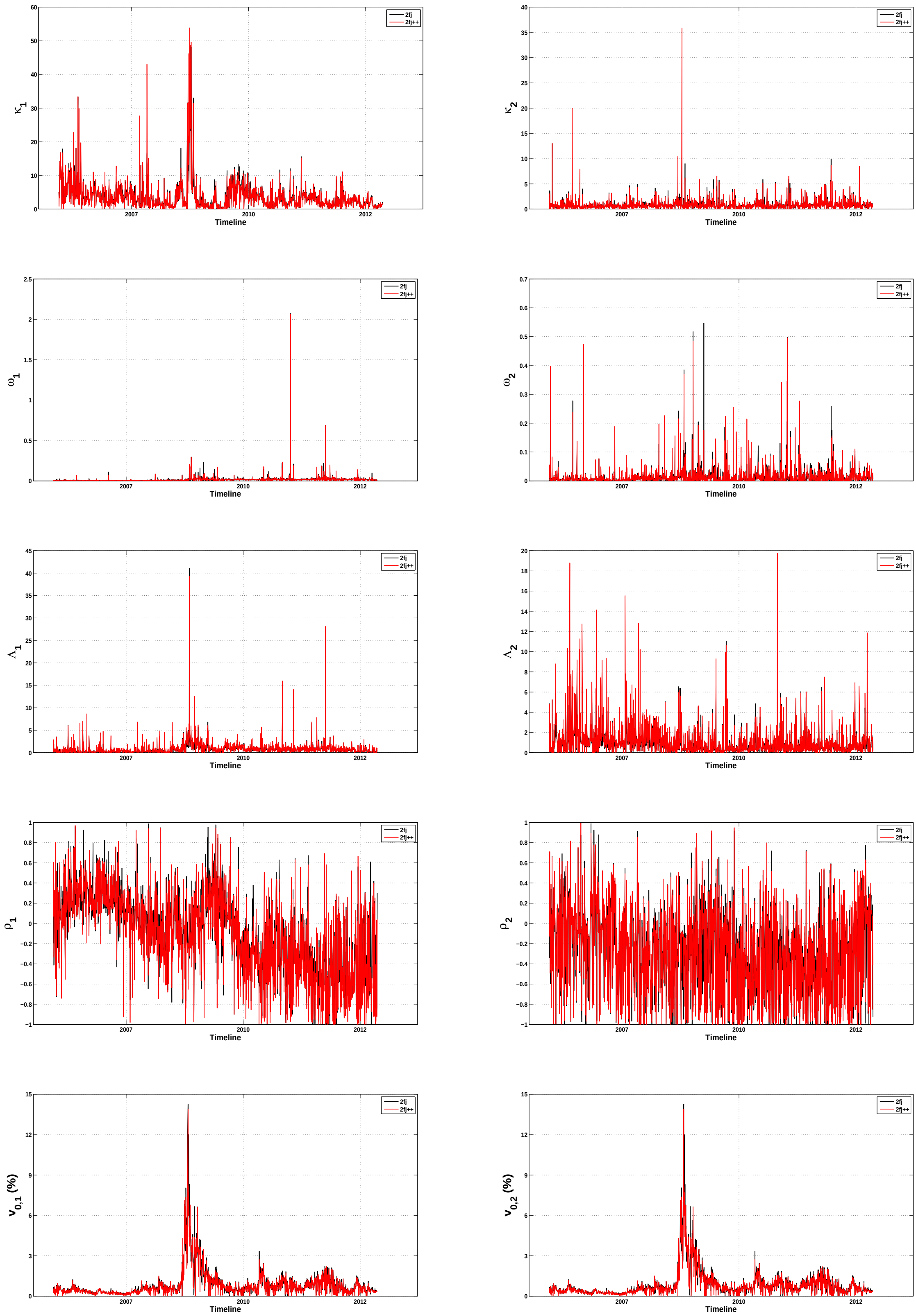

Figure 4.9: Daily parameters series of the first volatility factor (*left*) and second volatility factor (*right*) for affine multi-factor Heston and Heston++ with price jumps

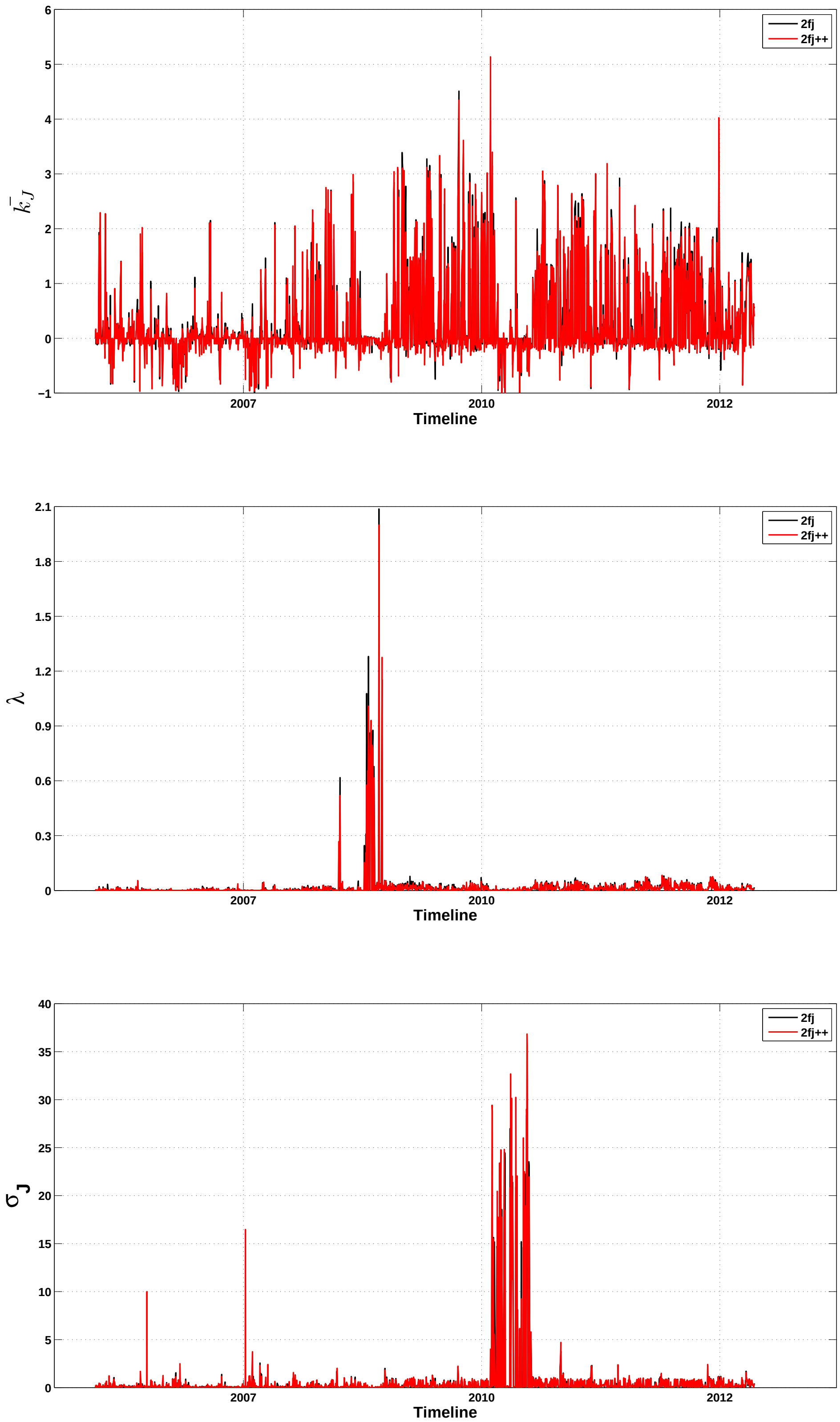

Figure 4.10: Daily parameters series of the jump component for affine multi-factor Heston and Heston++ with price jumps

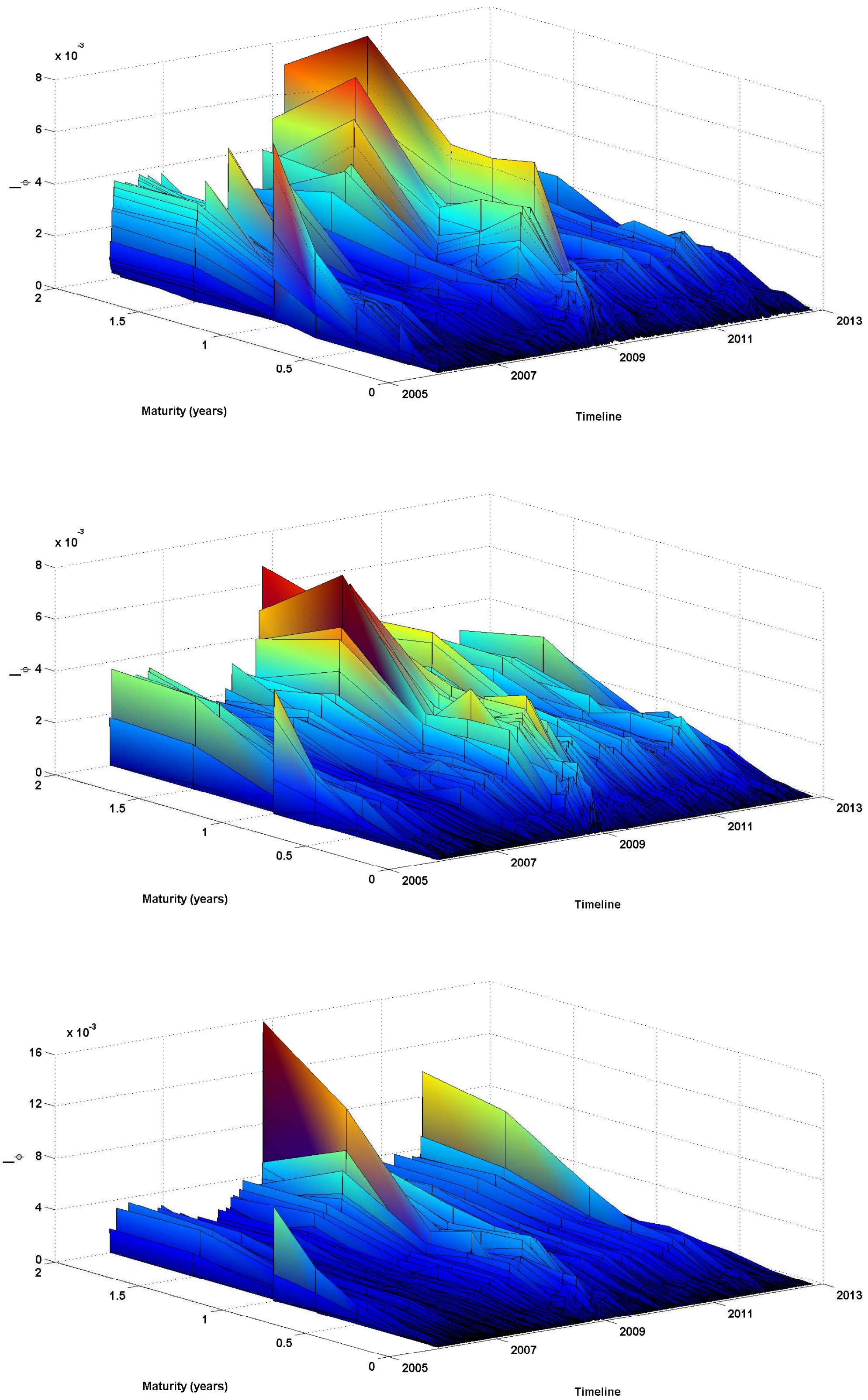

Figure 4.11: Daily estimates surfaces of the vector with components the integrals of ϕ_t for the one factor Heston++ (*top*), two-factor Heston++ (*middle*) and two-factor with price jumps Heston++ (*bottom*) models

Bibliography

[1] Adrian, T. & Rosenberg, J. (2008). "Stock returns and Volatility: Pricing the Short-run and Long-run Components of Market Risk". *Journal of Finance* 63, 2997-3030.

[2] Aït-Sahalia, Y. (1996). "Nonparametric Pricing of Interest Rate Derivative Securities". *Econometrica* 64, 527-560.

[3] Aït-Sahalia, Y. (2002). "Maximum-Likelihood Estimation of Discretely Sampled Diffusions: A Closed-Form Approximation Approach". *Econometrica* 70, 223-262.

[4] Aït-Sahalia, Y. (2008). "Closed-Form Likelihood Expansions for Multivariate Diffusions". *Annals of Statistics* 36, 906-937.

[5] Aït-Sahalia, Y. & Jacod, J. (2009). "Testing for Jumps in a Discretely Observed Process". *Annals of Statistics* 37, 184-222.

[6] Aït-Sahalia, Y., Karaman, M. & Mancini, L. (2012). "The Term Structure of Variance Swaps, Risk Premia and the Expectation Hypothesis". Working Paper.

[7] Aït-Sahalia, Y. & Kimmel, R. (2007). "Maximum Likelihood Estimation of Stochastic Volatility Models". *Journal of Financial Economics* 83, 413-452.

[8] Aït-Sahalia, Y., Mykland, P. & Zhang, L. (2005). "How Often to Sample a Continuous-time Process in the Presence of Microstructure Noise". *Review of Financial Studies* 18, 351-416.

[9] Albanese, C., Lo, H. & Mijatović, A. (2009). "Spectral Methods for Volatility Derivatives". *Quantitative Finance* 9, 663-692.

[10] Albrecher, H., Mayer, P., Schoutens, W. & Tistaert, J. (2006). "The Little Heston Trap". *Wilmott Magazine*, January: 83-92.

[11] Andersen, T. G. (1994). "Stochastic Autoregressive Volatility: A Framework for Volatility Modeling". *Mathematical Finance* 4, 75-102.

[12] Andersen, T. G. & Benzoni, L. (2008a). "Realized Volatility". In Andersen, T. G., Davis, R. A., Kreib, J.-P. & Mikosch, T. (Eds.) *Handbook of Financial Time Series*, New York: Springer-Verlag.

[13] Andersen, T. G. & Benzoni, L. (2008b). "Stochastic Volatility". In *Encyclopedia of Complexity and System Science*, New York: Springer-Verlag.

[14] Andersen, T. G., Benzoni, L. & Lund, J. (2002). "Towards an Empirical Foundation for Continuous-Time Equity Returns Models". *Journal of Finance* 57, 1239-1284.

[15] Andersen, T. G., Benzoni, L. & Lund, J. (2004). "Stochastic Volatility, Mean Drift and Jumps in the Short Term Interest Rate". Working Paper.

[16] Andersen, T. G. & Bollerslev, T. (1998). "Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts". *International Economic Review* 39, 885-905.

[17] Andersen, T., Bollerslev, T. & Diebold, F. X. (2010). "Parametric and Nonparametric Volatility Measurement". In Ait-Sahalia, Y. & Hansen, L. P. (Eds.) *Handbook of Financial Econometrics*, Amsterdam: North-Holland.

[18] Andersen, T. G., Bollerslev, T., Diebold, F. X., Labys, P. (2003). "Modeling and Forecasting Realized Volatility". *Econometrica* 71, 579-625.

[19] Andersen, T. G. & Bondarenko, O. (2007). "Construction and Interpretation of Model-Free Implied Volatility". Working Paper.

[20] Andersen, T. G., Fusari, N. & Todorov, V. (2012). "Parametric Inference and Dynamic State Recovery From Option Panels". National Bureau of Economic Research Working Paper.

[21] Andersen, T. G. & Lund, J. (1997). "Estimating Continuous-time Stochastic Volatility Models of the Short Term Interest Rate Diffusion". *Journal of Econometrics* 77, 343-377.

[22] Andreasen, J. & Huge, B. (2011). "Volatility Interpolation". *Risk*.

[23] Apel, T., Winkler, G. & Wystup, U. (2002). "Valuation of Options in Heston's Stochastic Volatility Model Using Finite Element Methods". In Hakala, J. & Wystup, U. (Eds.) *Foreign Exchange Risk*, London: Risk Books.

[24] Arnold, L. (1974). *Stochastic Differential Equations: Theory and Applications*. New York: Wiley.

[25] Bachelier, L. (1900). "Théorie de la Spéculation". *Annales Scientifiques de l'Ècole Normale Supérieure* 3, 21-86. English translation in Cootner, P.H. (ed.) (1964). *The Random Character of Stock Market Prices*. Cambridge, MA: MIT Press.

[26] Back, K. (1991). "Asset Prices for General Processes". *Journal of Mathematical Economics* 20, 371-395.

[27] Bakshi, G., Cao, C. & Chen, Z. (1997). "Empirical Performance of Alternative Option Pricing Models". *Journal of Finance* 52, 2003-2049.

[28] Bakshi, G., Cao, C. & Chen, Z. (2002). "Pricing and Hedging Long-term Options". *Journal of Econometrics* 94, 277-318.

[29] Bandi, F. M. & Phillips, P. C. B. (2003). "Fully Nonparametric Estimation of Scalar Diffusion Models". *Econometrica* 71, 241-283.

[30] Bandi, F. M. & Russell, J. (2006). "Separating Microstructure Noise from Volatility". *Journal of Financial Economics* 79, 655-692.

[31] Barndorff-Nielsen, O. E. & Shephard, N. (2001). "Non-Gaussian Ornstein-Uhlenbeck-based Models and Some of Their Uses in Financial Economics". *Journal of the Royal Statistical Society*, Series B 63, 167-241.

[32] Barndorff-Nielsen, O. E. & Shephard, N. (2002a). "Econometric Analysis of Realised Volatility and its Use in Estimating Stochastic Volatility Models". *Journal of the Royal Statistical Society*, Series B 64, 253-280.

[33] Barndorff-Nielsen, O. E. & Shephard, N. (2002b). "Estimating Quadratic Variation Using Realized Variance". *Journal of Applied Econometrics* 17, 457-477.

[34] Barucci, E. & Renò, R. (2002a). "On Measuring Volatility and the GARCH Forecasting Performance". *Journal of International Financial Markets, Institutions and Money* 12, 183-200.

[35] Barucci, E. & Renò, R. (2002b). "On Measuring Volatility of Diffusion Processes with High Frequency Data". *Economics Letters* 74, 371-378.

[36] Bates, D. S. (1991). "The Crash of '87: Was It Expected? The Evidence from Options Markets". *Journal of Finance* 46, 1009-1044.

[37] Bates, D. S. (1996). "Jumps and Stochastic Volatility: Exchange Rate Processes Implicit in Deutsche Mark Options". *Review of Financial Studies* 9, 69-107.

[38] Bates, D. S. (2000). "Post-'87 Crash Fears in the S&P 500 Futures Option Market". *Journal of Econometrics* 94, 181-238.

[39] Benhamou, E., Gobet, E. & Miri, M. (2010). "Time Dependent Heston Model". *Journal of Financial Mathematics* 1, 289-325.

[40] Benzoni, L. (2002). "Pricing Options under Stochastic Volatility: An Empirical Investigation". Working Paper.

[41] Black, F. & Scholes, M. (1972). "The Valuation of Option Contracts and a Test of Market Efficiency". *Journal of Finance* 27, 399-418.

[42] Black, F. & Scholes, M. (1973). "The Pricing of Options and Corporate Liabilities". *Journal of Political Economy* 81, 637-654.

[43] Bossens, F., Deelstra, G., Rayéè, G. & Skantzos, N.S. (2010). “Vanna-Volga Methods Applied to FX Derivatives: From Theory to Market Practice”. Working Paper.

[44] Breeden, D.T. & Litzenberger, R.H. (1978). “Prices of State-contingent Claims Implicit in Option Prices”. *Journal of Business* 51, 621-651.

[45] Brenner, M. & Galai, D. (1989). “New Financial Instruments for Hedging Changes in Volatility”. *Financial Analysts Journal* July/August.

[46] Brenner, M., Shu, J. & Zhang, J. E. (2010). “The New Market for Volatility Trading”. *Journal of Futures Markets* 30, 809-833.

[47] Brigo, D. & Mercurio, F. (2001). “A Deterministic-Shift Extension of Analytically Tractable and Time-Homogeneous Short-Rate Models”. *Finance & Stochastics* 5, 369-387.

[48] Brigo, D. & Mercurio, F. (2006). *Interest Rate Models - Theory and Practice: With Smile, Inflation and Credit*. Springer Finance.

[49] Britten-Jones, M. & Neuberger, A. (2000). “Option Prices, Implied Price Processes, and Stochastic Volatility”. *Journal of Finance* 55, 839-866.

[50] Broadie, M., Chernov, M. & Johannes, M. (2009). “Specification and Risk Premiums: The Information in S&P 500 Futures Options”. *Journal of Finance* 62, 1453-1490.

[51] Broadie, M. & Jain, A. (2008). “The Effect of Jumps and Discrete Sampling on Volatility and Variance Swaps”. *International Journal of Theoretical and Applied Finance* 11, 761-797.

[52] Campa, J. M. & Chang, P. H. K. (1995). “Testing the Expectations Hypothesis on the Term Structure of Volatilities in Foreign Exchange Options”. *Journal of Finance* 50, 529-547.

[53] Canina, L. & Figlewski, S. (1993). “The Information Content of Implied Volatility”. *Review of Financial Studies* 6, 659-681.

[54] Carr, P. & Lee, R. (2009). “Volatility Derivatives”. *Annual Review of Financial Economics* 1, 1-21.

[55] Carr, P. & Madan, D. (1998). “Towards a Theory of Volatility Trading”. In Jarrow, R. (Ed.) *Volatility Estimation Techniques for Pricing Derivatives*, London: Risk Books.

[56] Carr, P. & Madan, D. (1999). “Option Valuation Using the Fast Fourier Transform”. *Journal of Computational Finance* 2, 2003-2049.

[57] Carr, P. & Wu, L. (2006). “A Tale of Two Indices”. *Journal of Derivatives* 13, 13-29.

[58] Carr, P. & Wu, L. (2009). “Variance Risk Premia”. *Review of Financial Studies* 22, 1311-1341.

[59] Castagna, A. & Mercurio, F. (2006). “Consistent Pricing of FX Options”. Working Papers, Banca IMI, Milan.

[60] Castagna, A. & Mercurio, F. (2007). "The Vanna-Volga Method for Implied Volatilities". *Risk* January, 106-111.

[61] Chance, D.M. (1986). "Empirical Tests of the Pricing of Index Call Options". *Advances in Futures and Options Research* 1, 141-166.

[62] Chang, E. C., Zhang, J. E. & Zhao, H. (2012). "Equilibrium Asset and Option Pricing Under Jump Diffusion". *Mathematical Finance* 22, 538-568.

[63] Chen, H., Chung, S. & Ho, K. (2011). "The Diversification Effects of Volatility-Related Assets". *Journal of Banking & Finance*, 35, 1179-1189.

[64] Chernov, M. & Ghysel, E. (2000). "A Study Towards A Unified Approach to the Joint Estimation of Objective and Risk-Neutral Measures for the Purpose of Options Valuation". *Journal of Financial Economics* 56, 407-458.

[65] Christensen, B. J., Hansen, C. S. & Prabhala, N. R. (2001). "The Telescoping Overlap Problem in Option Data". Working Paper, University of Aarhus and University of Maryland.

[66] Christensen, B. J. & Prabhala, N. R. (1998). "The Relation Between Implied and Realized Volatility". *Journal of Financial Economics* 50, 125-150.

[67] Christoffersen, P. F., Heston, S. & Jacobs, K. (2009). "The Shape and Term Structure of the Index Option Smirk: Why Multifactor Stochastic Volatility Models Work So Well". *Management Science* 55, 1914-1932.

[68] Chung, S.-L., Tsai, W.-C., Wang, Y.-H. & Weng, P.-S. (2011). "The Information Content of the S&P 500 Index and VIX Options on the Dynamics of the S&P 500 Index". *Journal and Futures Markets* 31, 1170-1201.

[69] Cont, R. & Kokholm, T. (2012). "A Consistent Pricing Model for Index Options and Volatility Derivatives". *Mathematical Finance*, forthcoming.

[70] Cox, J.C., Ingersoll, J.E. & Ross, S.A. (1985a). "An Intertemporal General Equilibrium Model of Asset Prices". *Econometrica* 53, 363-384.

[71] Cox, J.C., Ingersoll, J.E. & Ross, S.A. (1985b). "A Theory of the Term Structure of Interest Rates". *Econometrica* 53, 385-407.

[72] Da Fonseca, J., Grasselli, M. & Tebaldi, C. (2008). "A Multifactor Volatility Heston Model". *Quantitative Finance* 8, 591-604.

[73] Daigler, R. T. & Wang Z.-G. (2011). "The Performance of VIX Option Pricing Models: Empirical Evidence Beyond Simulation". *Journal and Futures Markets* 31, 251-281.

[74] Das, S. R. and Sundaram, R.K. (1999). "Of Smiles and Smirks: A Term Structure Perspective". *Journal of Financial and Quantitative Analysis* 34, 211-239.

[75] Demeterfi, K., Derman, E., Kamal, M. & Zou, J. (1999a). "More Than You Ever Wanted To Know About Volatility Swaps". *Quantitative Strategies Research Notes*, Goldman Sachs.

[76] Demeterfi, K., Derman, E., Kamal, M. & Zou, J. (1999b). "A Guide to Volatility and Variance Swaps. *Journal of Derivatives* 6, 9-32.

[77] Derman, E. & Kani, I. (1994). "Riding on a Smile". *Risk* 7, 32-39.

[78] Duffie, D. (1996). *Dynamic Asset Pricing Theory* (2d ed.). Princeton: Princeton University Press.

[79] Duffie, D., Pan, J. & Singleton, K. (2000). "Transform Analysis and Asset Pricing for Affine Jump-Diffusions". *Econometrica* 68, 1343-1376.

[80] Dupire, B. (1994). "Pricing with a Smile". *Risk* 7, 18-20.

[81] Ederington, L. H. & Guan, W. (2002). "Is Implied Volatility Informationally Efficient and Effective Predictor of Future Volatility?". *Risk* 4, 29-46.

[82] Efron, B. (1979). "Bootstrap Method: Another Look at the Jackknife". *The Annals of Statistics* 7, 1-26.

[83] Elices, A. (2008). "Models with Time-dependent Parameters Using Transform Methods: Application to Heston's Model". arXiv:0708.2020v2.

[84] Eraker, B. (2004). "Do Equity Prices and Volatility Jump? Reconciling Evidence from Spot and Option Prices". *Journal of Finance* 59, 1367-1403.

[85] Eraker, B., Johannes, M.S. & Polson, N.G. (2003). "The Impact of Jumps in Returns and Volatility". *Journal of Finance* 53, 1269-1300.

[86] Feller, W. (1951). "Two Singular Diffusion Problems". *Annals of Mathematics* 54, 173-182.

[87] Fisher, T. (2007). "Variations on the Vanna-Volga Adjustment". Bloomberg Research Paper.

[88] Fleming, J. (1998). "The Quality of Market Volatility Forecasts Implied by S&P 100 Index Option Prices". *Journal of Empirical Finance* 5, 317-345.

[89] Fouque, J.- P., Papanicolaou, G. & Sircar, K. R. (2000). *Derivatives in Financial Markets with Stochastic Volatility*. Cambridge: Cambridge University Press.

[90] Galai, D. (1977). "Tests of Market Efficiency and the Chicago Board Options Exchange". *Journal of Business* 50, 167-197.

[91] Garman, M. B. & Kohlhagen, S. W. (1983). "Foreign currency option values". *Journal of International Monet & Finance* 2, 231-237.

[92] Gatheral, J. (2006). *The Volatility Surface: A Practitioner's Guide*. Wiley.

[93] Geman, H., El Karoui, N. & Rochet, J. C. (1995). "Changes of Numeraire, Changes of Probability Measures and Option Pricing. *Journal of Applied Probability* 32, 443-458.

[94] Greene, W. H. (2011). *Econometric Analysis* (7th ed.). N.J.: Prentice-Hall.

[95] Hagan, P., Kumar, D., Lesniewski, A. & Woodward, D. (2002). "Managing Smile Risk". *Wilmott Magazine* 1, 84-108.

[96] Hakala, J. & Wystup, U. (2002). "Heston's Stochastic Volatility Model Applied to Foreign Exchange Options". In *Foreign Exchange Risk.* London: Risk Books.

[97] Heston, S. (1993). "A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options". *Review of Financial Studies* 6, 327-343.

[98] Hilal, S., Poon, S.-H. & Tawn, J. (2011). "Hedging the Black Swan: Conditional Heteroskedasticity and Tail Dependence in S&P 500 and VIX". *Journal of Banking & Finance* 35, 2374-2387.

[99] Huang, Y. & Zhang, J. E. (2010). "The CBOE S&P 500 Three-month Variance Futures". *Journal of Futures Markets* 30, 48-70.

[100] Hull, J. (1993). *Options, Futures, and Other Derivative Securities* (2d ed.). Englewood Cliffs, N.J.: Prentice-Hall.

[101] Hull, J. & White, A. (1987). "The Pricing of Options on Assets with Stochastic Volatilities". *Journal of Finance* 42, 281-300.

[102] Huskaj, B. & Nossman, M. (2013). "A Term Structure Model for VIX Futures". *Journal of Futures Markets*, 33, 421-442.

[103] Ikeda, N. & Watanabe, S. (1981). *Stochastic Differential Equations and Diffusion Processes.* Amsterdam: North-Holland.

[104] Jäckel, P. & Kahl, C. (2005). "Not-So-Complex Logarithms in the Heston Model". *Willmott Magazine* 19, 94-103.

[105] Jarrow, R. (1988). *Finance Theory.* Englewood Cliffs, N.J.: Prentice-Hall.

[106] Jarrow, R. & Rudd, A. (1983). *Option Pricing.* Homewood, Ill.: Richard D. Irwin.

[107] Jiang, G. J. & Tian, Y. S. (2005). "The Model-Free Implied Volatility and Its Information Content". *Review of Financial Studies* 18, 1305-1342.

[108] Johnson, H. & Shanno, D. (1987). "Option Pricing when the Variance is Changing". *Journal of Financial and Quantitative Analysis* 22, 143-152.

[109] Kahalé, N. (2004). "An Arbitrage-free Interpolation of Volatilitites". *Risk.*

[110] Karatzas, I. & Shreve, S. (1988). *Brownian Motion and Stochastic Calculus*. New York: Springer-Verlag.

[111] Karatzas, I. & Shreve, S. (1995). "Methods of Mathematical Finance". Chapters 1-3, Draft Manuscript, Department of Mathematics, Carnegie-Mellon University, Pittsburgh.

[112] Kopp, P. (1984). *Martingales and Stochastic Integrals*. Cambridge: Cambridge University Press.

[113] Lauterbach, B. & P. Schultz. "Pricing Warrants: An Empirical Study of the Black-Scholes Model and its Alternatives". *Journal of Finance* 4, 1181-1210.

[114] Li, C.-X. (2010). "Efficient Valuation of Options on VIX under Gatheral's Double Lognormal Stochastic Volatility Model: An Asymptotic Expansion Approach". *Working Paper Series*. New York: Columbia University.

[115] Lian, G.-H. & Zhu, S.-P. (2012). "An Analytical Formula for VIX Futures and Its Applications". *Journal of Futures Markets* 32, 166-190.

[116] Lin, Y.-N. (2007). "Pricing VIX Futures: Evidence From Integrated Physical and Risk-Neutral Probability Measures". *Journal of Futures Markets* 27, 1175-1217.

[117] Lin, Y.-N. & Chang, C.-H. (2009). "VIX Option Pricing". *Journal of Futures Markets* 29, 523-543.

[118] Lipton, A. & McGhee, W. (2002). "Universal Barriers". *Risk* May, 81-85.

[119] Lucic, V. (2008). "Boundary Conditions for Computing Densities in Hybrid Models". Barclays Capital Working Paper Series.

[120] Luo, X. & Zhang, J. E. (2012). "The Term Structure of VIX". *Journal of Futures Market* 32, 1092-1123.

[121] MacBeth, J.D. & Merville, L. J. (1979). "An Empirical Examination of the Black-Scholes Call Option Pricing Model". *Journal of Finance* 34, 1173-1186.

[122] Malliaris, A. (1982). *Stochastic Methods in Economics and Finance*. Amsterdam: North-Holland.

[123] Malz, A. (1997). "Estimating the Probability Distribution of the Future Exchange Rate from Option Prices". *Journal of Derivatives* Winter, 20-36.

[124] Melino, A. & Turnbull, S. M. (1990). "Pricing Foreign Currency Options with Stochastic Volatility". *Journal of Econometrics* 45, 239-265.

[125] Merton, R. (1969). "Lifetime Portfolio Selection under Uncertainty: The Continuous Time Case". *Review of Economics and Statistics* 51, 247-257.

[126] Merton, R. (1973). "An Intertemporal Capital Asset Pricing Model". *Econometrica* 41, 867-887.

[127] Merton, R. (1976). "Option Pricing when Underlying Stock Returns are Discontinuous". *Journal of Financial Economics* 3, 125-144.

[128] Merton, R. (1990). *Continuous-Time Finance.* Oxford: Basil Blackwell.

[129] Mikhailov, S. & N'ogel, U. (2003). "Heston's Stochastic Volatility Model: Implementation, Calibration and Some Extensions". *Wilmott Magazine*, 74-79.

[130] Miller, R. G. (1974). "The Jackknife - A Review". *Biometrika* 61, 1-15.

[131] Neftci, S. N. (2000). *An Introduction to the Mathematics of Financial Derivatives* (2d ed.). New York: Academic Press.

[132] Nicolato, E. & Venardos, E. (2003). "Option Pricing in Stochastic Volatility Models of the Ornstein-Uhlenbeck Type". *Mathematical Finance* 13, 445-466.

[133] Øksendal, B. (1985). *Stochastic Differential Equations.* Berlin: Springer-Verlag.

[134] Pan, J. (2002). "The Jump-Risk Premia Implicit in Options: Evidence from an Integrated Time-Series Study". *Journal of Financial Economics* 63, 3-50.

[135] Pascucci, A. (2008). *Calcolo Stocastico per la Finanza.* Milano: Springer-Verlag.

[136] Pong, S.-Y., Shackleton, M. B., Taylor, S. J. & Xu, X.-Z. (2004). "Forecasting Sterling/Dollar Volatility: A Comparison of Implied Volatility and AR(FI)MA Models". *Journal of Banking & Finance* 28, 2541-2563.

[137] Protter, P. (1990). *Stochastic Integration and Differential Equations.* New York: Springer-Verlag.

[138] Quenouille, M. (1949). "Approximate tests of correlation in time series". *Journal of the Royal Statistical Society* 11, 18-44.

[139] Quenouille, M.H. (1956). "Notes on bias in estimation". *Biometrika* 61, 353-360.

[140] Revuz, D. & Yor, M. (1991). *Continuous Martingales and Brownian Motion.* New York: Springer-Verlag.

[141] Rubinstein, M. (1985). "Nonparametric Tests of Alternative Option Pricing Models Using All Reported Trades and Quotes on the 30 Most Active CBOE Option Classes from August 23, 1976 through August 31, 1978". *Journal of Finance* 40, 455-480.

[142] Rubinstein, M. (1994). "Implied Binomial Trees". *Journal of Finance* 49, 771-818.

[143] Schöbel, R. & Zhu, J. (1999). "Stochastic Volatility With an Ornstein-Uhlenbeck Process: An Extension". *European Finance Review* 3, 23-46.

[144] Scott, L. O. (1987). “Option Pricing when the Variance Changes Randomly: Theory, Estimation and an Application”. *Journal of Financial and Quantitative Analysis* 22, 419-438.

[145] Sepp, A. (2008a). “VIX Option Pricing in Jump-Diffusion Model”. *Risk* April 2008, 84-89.

[146] Sepp, A. (2008b). “Pricing Options on Realized Variance in Heston Model with Jumps in Returns and Volatility”. *Journal of Computational Finance* 11, 33-70.

[147] Shastri, K. & K. Tandon. “Valuation of Foreign Currency Options: Some Empirical Tests”. *Journal of Financial and Quantitative Analysis* 21, 57-74.

[148] Shkolnikov, Y. (2009). “Generalized Vanna-Volga Method and its Applications”. NumeriX Research Paper.

[149] Shreve, S. E. (2004). *Stochastic Calculus for Finance*, Vol. I. New York: Springer.

[150] Shreve, S. E. (2004). *Stochastic Calculus for Finance*, Vol. II. New York: Springer.

[151] Stein, J. C. (1989). “Overreactions in the Options Markets”. *Journal of Finance* 44, 1011-1023.

[152] Stein, E. M. & Stein, J. C. (1991). “Stock Price Distribution with Stochastic Volatility: An Analytic Approach”. *Review of Financial Studies* 4, 727-752.

[153] Taylor, S. J. & Xu, X. (1994). “The Term Structure of Volatility Implied by Foreign Exchange Options”. *Journal of Financial and Quantitative Analysis* 29, 57-74.

[154] Todorov, V. & Tauchen, G. (2011). “Volatility Jumps”. *Journal of Business and Economic Statistics* 29, 356-371.

[155] Vasicek, O. (1977). “An Equilibrium Characterization of the Term Structure”. *Journal of Financial Economics* 5, 177-188.

[156] Wiggings, J. B. (1987). “Option Values under Stochastic Volatility: Theory and Empirical Estimates”. *Journal of Financial Economics* 19, 351-372.

[157] Wilmott, P., Dewynne, J. & Howison, S. (1993). *Option Pricing: Mathematical Models and Computation*. Oxford: Oxford Financial Press.

[158] Wystup U. (2003). “The Market Price of One-touch Options in Foreign Exchange Markets”. *Derivatives Week* 12, 1-4.

[159] Wystup, U. (2006). *FX Options and Structured Products*. Wiley Finance.

[160] Yu, W. W., Liu, E. C. K. & Wang, J. W. (2010). “The Predictive Power of the Implied Volatility of Options Traded OTC and on Exchanges”. *Journal of Banking & Finance* 34, 1-11.

[161] Zhang, J. E. & Zhu, Y. (2006). “VIX Futures”. *Journal of Futures Markets* 26, 521-531.

[162] Zhang, J. E. & Zhu, Y. (2007). “Variance Term Structure and VIX Futures Pricing”. *Internationl Journal of Theoretical and Applied Finance* 10, 111-127.

[163] Zhu, J. (2008). “A Simple and Exact Simulation Approach to Heston Model”. Working Papers Series.