Quantitative Finance

ISSN: 1469-7688 (Print) 1469-7696 (Online) Journal homepage: https://www.tandfonline.com/loi/rquf20

Riding on the smiles

José da Fonseca & Martino Grasselli

To cite this article: José da Fonseca & Martino Grasselli (2011) Riding on the smiles, Quantitative Finance, 11:11, 1609-1632, DOI: 10.1080/14697688.2011.615218

To link to this article: https://doi.org/10.1080/14697688.2011.615218

Quantitative Finance, Vol. 11, No. 11, November 2011, 1609–1632

Routledge
Taylor & Francis Group

Riding on the smiles

JOSÉ DA FONSECA*† and MARTINO GRASSELLI‡§

†Department of Finance, Auckland University of Technology, Private Bag 92006, 1142 Auckland, New Zealand
‡Dipartimento di Matematica, Università degli Studi di Padova, Via Trieste 63, Padova, Italy
§Département Mathématiques et Ingénierie Financière, Ecole Supérieure d'Ingénieurs Léonard de Vinci, 92916 Paris La Défense, France

(*Received 25 October 2010; in final form 18 April 2011*)

Using a data set of vanilla options on the major indexes we investigate the calibration properties of several multi-factor stochastic volatility models by adopting the fast Fourier transform as the pricing methodology. We study the impact of the penalizing function on the calibration performance and how it affects the calibrated parameters. We consider single-asset as well as multiple-asset models, with particular emphasis on the single-asset Wishart Multidimensional Stochastic Volatility model and the Wishart Affine Stochastic Correlation model, which provides a natural framework for pricing basket options while keeping the stylized smile–skew effects on single-name vanillas. For all models we give some option price approximations that are very useful for speeding up the pricing process. In addition, these approximations allow us to compare different models by conveniently aggregating the parameters, and they highlight the ability of the Wishart-based models to control separately the smile and the skew effects. This is extremely important from a risk-management perspective of a book of derivatives that includes exotic as well as basket options.

Keywords: Calibration; Implied volatility; Wishart stochastic volatility; FFT; Stochastic skew

JEL Classification: C52, G12, G13

1. Introduction

The notion of implied volatility has been well known to traders since the celebrated Black–Scholes formula, which gives the price of an option by assuming that the volatility of the underlying asset remains constant. The implied volatility σ_{imp} is the number that one should replace in the Black–Scholes formula in order to obtain the quoted price of the option, that is

$$\underbrace{C_{\text{mkt}}(t,T,F_t,K)}_{\text{market price}} = \underbrace{c_{\text{B\&S}}(t,T,F_t,K,\sigma^2_{\text{imp}}(T-t))}_{\text{price in the Black-Scholes model}}.^{\P} \quad (1)$$

When we apply this methodology to a book of vanilla options we observe that the implied volatility is not constant over strikes and time to maturity, thus leading to the so-called implied volatility surface. This shows that the Black–Scholes model is not consistent with real data and explains the reason why the implied volatility is often referred to as the *wrong number in the wrong formula to get the correct price* (Rebonato 1999). The cross-section of the implied volatility surface is called the volatility smile effect and reveals typical decreasing asymmetric behavior with respect to the log-moneyness (defined as the logarithm of the ratio strike price/forward price), which is consistent with the leverage effect between the stock and its volatility (market returns are negatively correlated with their volatilities). The asymmetric volatility smile is referred to as the volatility smirk or skew effect and is typical of any equity market.

Figure 1 shows the volatility smiles for different maturities on the DAX index on August 28, 2008. We observe some stylized facts shared by all major equity indexes. First, we notice that, for the shortest maturity considered in the book (3 weeks), the smile is rather regular and displays strong convex behavior. Second, as the time to maturity increases the smile becomes almost linear. What is more, we have a term structure of skews that changes with the maturity.

*Corresponding author. Email: jose.dafonseca@aut.ac.nz

¶F_t denotes the forward price of the underlying asset at the evaluation time t, T is the maturity of the option and K the strike price.

Quantitative Finance
ISSN 1469–7688 print/ISSN 1469–7696 online © 2011 Taylor & Francis
http://www.tandfonline.com
http://dx.doi.org/10.1080/14697688.2011.615218

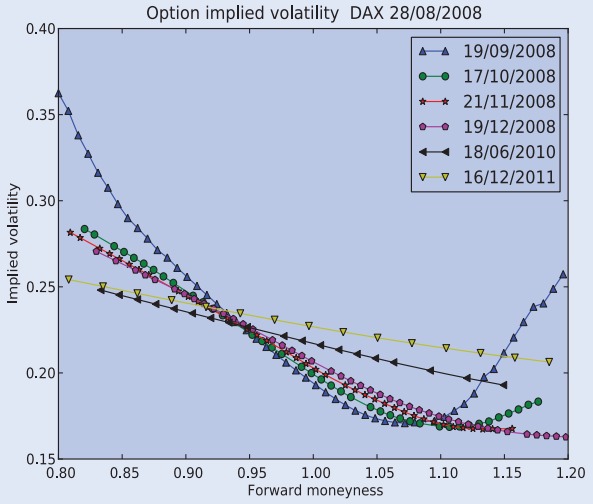

Figure 1. Option implied volatilities for the DAX on 28/08/2008.

During the last two decades the volatility smile has received much attention from academics and practitioners and many articles have been published on this subject.† The most celebrated stochastic volatility model, which constitutes the starting point of our research, is that proposed by Heston (1993). The reason for its success relies on its analyticity, since it allows the computation of option prices very efficiently. In addition, the fact that each parameter of this model has a clear financial interpretation is greatly appreciated by traders. A very useful parameter is that associated with the correlation between the asset's noise and the noise of the volatility process: in fact, this correlation is responsible for the volatility smirk, that is the slope of the implied volatility smile due to the leverage effect.

Much effort has been extended to generalize the Heston model in order to accommodate certain stylized effects observed in the option market that cannot be recovered in its original formulation. In particular, the pure diffusive nature of the Heston model cannot reproduce the highly convex smile for short-time-expiring options and a constant correlation parameter cannot generate a term structure for the volatility skew. Therefore, there have been two main streams of literature aiming at addressing these stylized facts. The first introduces jump components in order to capture the short-term implied volatility surface. For example, Bates (1996) extends the Heston model by adding jumps to the dynamics of the stock in order to be consistent with the stylized facts observed on time series and (more importantly) to reproduce the strong convexity of the short-term smile. Duffie *et al.* (2000) propose, within the affine framework, a stochastic volatility model with jumps both on the stock and the volatility.

The second stream investigates the multidimensional aspect of the volatility, that is the necessity of introducing other factors in order to describe the term structure of the implied volatility smile. Christoffersen *et al.* (2009) develop a two-factor (purely diffusive) stochastic volatility model built upon the square-root process. Two-factor models have much more flexibility in controlling the level and slope of the smile with respect to single-factor models. In fact, Christoffersen *et al.* (2009) describe the leverage by a stochastic process involving both factors. Bates (2000) extends Bates (1996) to a two-factor stochastic volatility model with jumps, arguing that an additional factor is needed to capture the term structure of implied volatilities. All these generalizations aim at producing a richer modeling for the skew term structure with respect to the flat structure in the original Heston model.‡ Da Fonseca *et al.* (2008) tackle the same objective using a new process recently introduced in finance by Gouriéroux and Sufana (2010), namely the Wishart process. This model extends that of Christoffersen *et al.* (2009) with the presence of an additional stochastic factor playing the role of covariance between the positive factors. This fact leads to greater flexibility of the Wishart-based models in terms of modeling the spot/volatility correlation structure. Gruber *et al.* (2010) study the model proposed by Da Fonseca *et al.* (2008) from a financial economics perspective, whilst our paper is written from a more quantitative finance/practitioner point of view. The differences dwell mainly in the implementation strategy, which affects the results.

We also recall an independent approach based on pure jump processes, which started in the 1990s, when Levy process-based models were systematically investigated (see Madan and Milne (1991), Carr *et al.* (2003), Carr and Wu (2004) and Cont and Tankov (2008), to name a few).

More recently, the academic literature has started to focus on multi-asset stochastic volatility models. Gouriéroux and Sufana (2010) proposed a model based on the continuous-time Wishart process, whilst Da Fonseca *et al.* (2007b) extended their model by introducing a correlation structure between the dynamics of the assets and their volatility. This aspect is of practical importance as it controls the leverage effect observed both in the time series and the smile. Another multi-asset model in the spirit of Da Fonseca *et al.* (2007b) and based on a matrix pure jump volatility process was developed by Barndorff-Nielsen and Stelzer (2011). The main feature of these models compared with the previous models is that they involve a matrix diffusion process that allows, at least from an analytical point of view, for richer dynamic properties.

The choice of the model is of course important, but it is not the only ingredient from a calibration perspective.

†We limit ourselves to considering the stochastic volatility approach and we mainly focus on the parametric models: in particular, we do not consider here the local volatility model approach introduced by Dupire and the subsequent extensions.

‡Recently, there has been an attempt to relax the flat term structure by assuming a piecewise constant correlation parameter in the Heston model as in Mikhailov and Nogel (2003), or a more general structure as in Benhamou *et al.* (2010). This apparently simple choice leads to non-trivial numerical issues in the pricing procedure and *a fortiori* in the calibration.

In fact, the implementation strategy also has a strong impact on the estimation of the parameters and, as a consequence, on the model properties. That is, the choice of the implementation methodology is part of the model itself, so that one should also deal with the *implementation risk*. In order to explain this concept let us consider the calibration issue on vanilla options. Suppose that we have a sample of several days of option prices. The market practice consists of performing one calibration per day, hence leading to a time series of calibrated parameters. The objective is to obtain a model explaining separately each calibration day. Of course, the drawback is that the parameters might change over time. This is the path followed, for example, by Bakshi *et al.* (1997) and Carr *et al.* (2003). Another possibility is to perform *one* calibration using all the available trading days. This is the strategy followed, for example, by Bates (2000), Carr and Wu (2007) and Christoffersen *et al.* (2009). This second implementation strategy often leads to numerical difficulties and ambiguous results which we will discuss in the next section. Lastly, the calibration procedure involves the minimization of a penalizing function that can depend either on option prices or on the implied volatility. We will show how this choice heavily affects the outcome of the calibration and plays a crucial role in the performance of the model.

The main contribution of this paper is to provide the first calibration of the Wishart Multidimensional Stochastic Volatility (WMSV) model introduced by Da Fonseca *et al.* (2008) and the Wishart Affine Stochastic Correlation model developed by Da Fonseca *et al.* (2007b). To provide a benchmark we also implement the Heston model and the BiHeston model of Christoffersen *et al.* (2009). All these models belong to the class of affine models, that is models for which the fast Fourier transform (FFT) methodology of Carr and Madan (1999) applies, thus leading to efficient pricing procedures. However, in the Wishart case the FFT requires non-trivial matrix calculus, so that some price approximations, in the spirit of Benabid *et al.* (2009), can be useful to speed up the calibration. In addition, we shall see that the implied volatility approximations are crucial in order to map and compare the different models by conveniently aggregating the parameters. This confirms both the intuition of Da Fonseca *et al.* (2007b, 2008) and the empirical findings of Da Fonseca *et al.* (2007a).

Using vanilla options on major indexes (DAX, FTSE and EuroStoxx50) we provide a complete set of calibrations that confirm the analytical results. According to intuition, multi-factor stochastic volatility models perform better for all the indexes. Also, they can describe the skew term structure of the implied volatility surface and give an idea about the number of factors needed to handle this aspect. Using short-term implied volatility expansions we underline the advantages of the Wishart Multidimensional Stochastic Volatility model in managing the skew risk with respect to the BiHeston model of Christoffersen *et al.* (2009). In fact, beyond the calibration performance we should also take into account how each model allows us to control the implied volatility surface movements.

In the multiple-asset framework we analyse the calibration performance of the WASC model for several index pairs. Here also the implied volatility expansion confirms the consistency of the calibrated parameters across the models. Moreover, the approximations suggest how a basket option priced using a WASC model (calibrated on two sets of vanilla options, one for each index) can be (roughly) hedged using two Heston models, each calibrated on a set of vanilla options. This result confirms that the implied volatility expansions are an interesting contribution of the paper. The resulting mapping procedure demonstrates the consistency of the information provided by the different models. It also raises the crucial question of managing several models for one asset and, more precisely, the aggregation of the hedging ratios. Depending on the product at hand, we will have to choose among different models, which implies that we will have to relate the information provided by several models. This is obviously connected to model risk (Cont 2006) and the robustness of the Black–Scholes formula (El Karoui *et al.* 1998).

The structure of the paper is as follows. In section 2 we review several issues related to the implementation of stochastic volatility models. In section 3 we develop the analytical tools needed to perform the pricing of vanilla options, the price approximations and the implied volatility expansions. In section 4 we present the calibration results and illustrate how the data confirm the analytical results. In section 5, some open issues are addressed along with the conclusion of the paper. We gather all the proofs in appendix A.

2. Stylized facts and implementation issues

Many models have been investigated and tested in the market using different implementations. However, it is important to understand that the implementation strategy has a strong impact on the estimation of the parameters and, as a consequence, on the model properties. In other words, the choice of the implementation methodology is part of the model itself.

To illustrate this point we will use the Heston (1993) stochastic volatility model. As mentioned before and widely discussed in the literature (see Bakshi *et al.* (1997) for a review), the Black–Scholes–Merton model exhibits many shortcomings and it cannot be used as an option pricing model. The most important problem comes from the assumption of a Normal distribution for the returns of the underlying asset and from the assumption of a constant volatility. On the contrary, the empirical probability distribution of underlying asset returns displays leptokurtosis (fat tails) and asymmetry, which is in contrast to the assumption of the Black–Scholes–Merton model under the risk-neutral measure. In his pioneering paper, Heston (1993) investigated the characteristics of a market where the dynamics of the volatility

are stochastic and he was able to obtain a quasi-analytical solution for the pricing of plain vanilla options.

Heston (1993) considers the following framework:

$$\mathrm{d}F_t = F_t\sqrt{v_t}\,\mathrm{d}W_t^1, \tag{2}$$

$$\mathrm{d}v_t = \kappa(\theta - v_t)\mathrm{d}t + \sigma\sqrt{v_t}\,\mathrm{d}W_t^2, \tag{3}$$

where F_t is the forward price of the underlying asset, v_t the variance of the underlying asset returns, W_t^1 and W_t^2 are two correlated Brownian motions under the risk-neutral pricing measure with $\mathrm{d}W^1\,\mathrm{d}W^2 = \rho\,\mathrm{d}t$, κ is the speed of mean reversion of the variance to its long-run mean θ and σ is the volatility of volatility.

The success of the Heston model relies on the clear meaning and simple financial interpretation of each parameter of the model. For example, if we want to obtain asymmetric distributions we will have to change the value of the parameter ρ, that is the correlation factor between the underlying asset returns and the volatility. Intuitively, if $\rho > 0$, high returns imply high levels of volatility and *vice versa*. This is translated into a right asymmetry on the probability distribution of returns. On the other hand, if $\rho < 0$, low returns imply high levels of volatility and *vice versa*. This produces a left asymmetry on the probability distribution of returns. What we observe in a real equity market is a negative correlation between returns and volatility: in fact, when the market produces highly negative returns the volatility is typically very high. This stylized fact is also known as the *leverage effect*.

A typical phenomenon that can be observed in the probability distribution of returns is the accentuate leptokurtosis. In other words, the probability of extreme events is greater than that implied for a Normal distribution. The Heston model can take into account this fact, allowing for a calibration of the parameter σ that governs the thickness of the tails. In fact, for different values of this parameter the distribution actually becomes sharper around the mean and tails become fatter. If $\sigma = 0$, then the volatility becomes constant and we are able to recover a Normal distribution of returns. On the contrary, if $\sigma > 0$, we will then produce a leptokurtic distribution.

Another stylized fact mentioned previously is the volatility clustering phenomenon. The new model is able to reproduce this fact, allowing the parameter κ to govern the speed of volatility mean reversion to its long-run mean. Intuitively, the greater the value of this parameter, the quicker the speed of mean reversion and the smaller the volatility clusters and *vice versa*. From this quick test of the potential of the Heston model, we deduce that if we conveniently calibrate parameters we are able to generate a wide variety of probability distributions of underlying asset returns. The new framework is therefore much more flexible than the Black–Scholes–Merton framework.

Last, but not least, thanks to the quasi-closed-form pricing formula for vanilla products, it is possible to efficiently calibrate the model using the FFT procedure, which we briefly recall. Let us consider the problem of pricing a contingent claim whose payoff is a function of the terminal log-forward price $x_T = \log(F_T)$, say $\Phi(x_T, T)$. From the usual risk-neutral argument, the price P_0 of such an option can be written as the risk-neutral expected value:†

$$P_0 = \mathrm{e}^{-rT}\mathbb{E}[\Phi(x_T, T)],$$

and by applying standard arguments (see, e.g., Carr and Madan (1999), Bakshi and Madan (2000), Duffie *et al.* (2000) and Sepp (2003)) it can be expressed in terms of the inverse Fourier transform of the payoff function and the characteristic function of the asset returns. In fact,

$$\begin{aligned} P_0 &= \mathrm{e}^{-rT}\mathbb{E}[\Phi(x_T, T)] \\ &= \mathrm{e}^{-rT}\frac{1}{2\pi}\int_{\mathcal{Z}} \phi_{x_0,v_0}(-i\lambda, 0, T)\widehat{\Phi}(\lambda)\mathrm{d}\lambda, \end{aligned} \tag{4}$$

where

$$\phi_{x_0,v_0}(\lambda, 0, T) = \mathbb{E}[\mathrm{e}^{i\lambda x_T}]$$

denotes the characteristic function of the asset's returns and

$$\widehat{\Phi}(\lambda) = \int_{\mathbb{R}} \mathrm{e}^{i\lambda x}\Phi(x, T)\mathrm{d}x$$

is the Fourier transform of the payoff function. Therefore, in order to solve the pricing problem analytically, one has to compute the characteristic function of the asset's returns and compute the Fourier transform of the specific payoff together with its admissible domain, i.e. the convergence set $\mathcal{Z} \subset \mathbb{C}$ for which the Fubini theorem implicitly used in (4) holds. For the Heston (1993) model (as well as for all the affine models considered in this paper) the computation of the characteristic function is explicit. We refer, for example, to Carr and Madan (1999) for further details.

The first issue one has to deal with while calibrating a model is the choice of the penalizing function. In fact, calibrating means minimizing with respect to the model parameters the distance between the market prices and the model values. It is well known (see, e.g., Christoffersen and Jacobs (2004)) that the choice of the distance has a substantial impact on the calibrated parameters.

The criterion that is most widely adopted in the academic literature aims at minimizing (usually the square of) the difference $C_{\mathrm{model}}(t, F_t, T, K) - C_{\mathrm{mkt}}(t, F_t, T, K)$, that is

$$\min \frac{1}{N}\sum_{i=1}^{N} \omega_i(C_{\mathrm{model}}(t, T_i, K_i) - C_{\mathrm{mkt}}(t, T_i, K_i))^2. \tag{5}$$

This choice can be referred to as the norm-in-price criterion when $\omega_i = 1$ and is used by Bates (1996), Bakshi *et al.* (1997), Duffie *et al.* (2000) and Carr *et al.* (2003) (of course, this list is far from exhaustive).

†r is the (not necessarily constant) interest rate.

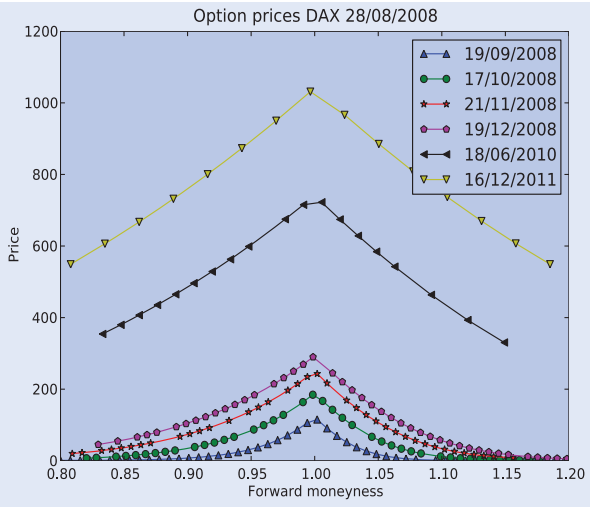

Figure 2. Option prices for the DAX on 28/08/2008.

In order to improve the calibration of the short-term implied volatility surface, one can put more weight on short maturity options. Choosing for ω_i the inverse on the Vega (that is, taking $\omega_i = (\partial_\sigma C^{\text{bs}}(t, T_i, K_i))^{-2}$) allows us to achieve this goal and it is indeed the market practice. In fact, dividing by the Vega will put more weight on OTM options with shorter maturity. This is the norm adopted by Carr and Wu (2007) and Christoffersen *et al.* (2009). Using a Taylor expansion it is possible to see that this leads to a norm that is equivalent to that based on implied volatility given by

$$\min \frac{1}{N} \sum_{i=1}^{N} (\sigma^{\text{imp}}_{\text{model}}(t, T_i, K_i) - \sigma^{\text{imp}}_{\text{mkt}}(t, T_i, K_i))^2. \qquad (6)$$

In order to gain intuition into the implications deriving from the choice of the criterion, let us consider a book of OTM vanilla call and put prices provided by Datastream on the DAX index quoted on August 28, 2008.† In this sample there are maturities between three weeks and four years. Following standard procedures, we keep only OTM option quotes because they tend to be more liquid and present no intrinsic value effect in their behavior. We further consider strikes leading to a moneyness ranging from 0.8 to 1.2, which gives a sample of 349 option prices. In figure 1 we plot the implied volatilities for different maturities, and figure 2 reports the prices.

The main point to remark upon in figure 2 is the size of the short-term option prices and, more precisely, the one-month maturity options. They are much smaller than the mid-to-long-term options, which means that they will have a small, if any, influence on the outcome of a calibration based on a norm-in-price criterion. To illustrate this point we carry out a calibration using this criterion and two data sets, one of which does not contain the short-term options. We report the results in table 1. We can see that the error, given by the column 'Error', is mostly unaffected by the presence of short-term options and the same remark also applies for the calibrated parameters.

If we use the implied volatility criterion (6) the result is significantly different, as can be seen from table 1. In fact, if the short-term options are used in the calibration procedure, then the calibration error is multiplied by a factor of 10. Furthermore, we observe that, in this case, we obtain larger values for the correlation parameter ρ and the volatility of volatility σ. This might explain, to some extent, the 'internal inconsistency' found by Bakshi *et al.* (1997), who used the norm-in-price criterion. Lastly, from figure 1 we see that discarding the first maturity results in a much more regular skew and simpler term structure for the smile. This point will be crucial later when we will deal with multi-factor stochastic volatility models.

Let us now turn our attention to another important issue, namely the time aggregation of the calibration error. When we calibrate a model, we have to specify if we choose a set of options quoted at a fixed day or a time series of option prices. In the former case we look for the parameters that minimize the distance between the model values and the market prices for that specific day. In the latter case we minimize the distance for the whole time window considered. Ideally, we would like to obtain the same parameters whatever the set of options and criterion used. Of course, this is often not the case. On the one hand, the calibration error we can achieve with a procedure based on a fixed day is likely to be smaller on that day than that obtained by calibrating the model using a set of options running over several days.

The market practice is to achieve a calibration per day and this is the path followed by Bakshi *et al.* (1997) and Carr *et al.* (2003). In that case we obtain a time series of calibrated parameters and the evolution of the volatility.

If we perform a calibration on a time series of option prices we face one major difficulty. As the volatility is not observable, it should be considered as a parameter and then estimated along with the other parameters. This strategy leads to the optimization of a function with respect to a large number of variables, which can become too difficult numerically and can give odd solutions.‡ To avoid numerical difficulties, Christoffersen *et al.* (2009) propose a two-step procedure: in the first step, for a given set of parameters, a volatility estimation is performed for each day, and, in the second step, an optimization with respect to the parameters is carried out. The algorithm loops over the two steps until convergence is reached. As such, the calibration involves the computation of several optimization problems, each of reasonable size. Nevertheless, it is rather difficult to determine if the best possible optimization is indeed achieved because the

†The results are qualitatively similar by choosing other indexes (such as FTSE, EuroStoxx50) and are available upon request.
‡For example, using this strategy, Bates (2000) estimates a two-factor stochastic volatility model using two square-root processes. Once calibrated, it turns out that one of them has a mean-reverting parameter equal to zero, therefore the dynamics of this diffusion is degenerate.

Table 1. Calibration on August 20, 2008 on DAX with the Heston (1993) model for different norms.

Norm	MinTime	Error	Error vol	κ	σ	ρ	θ	v_0
Price	0.06	2.27E−07	2.35E−4	0.8455	0.4639	−0.7120	0.0818	0.0423
Vol	0.06	2.33E−06	1.12E−4	1.3925	0.9217	−0.5473	0.0846	0.0414
Price	0.14	2.06E−07	1.57E−05	0.8265	0.4700	−0.7012	0.0826	0.0426
Vol	0.14	1.04E−06	1.25E−05	1.1007	0.6178	−0.6193	0.0819	0.0414

MinTime is the shortest maturity in the sample (0.06 = 3 weeks). Error denotes the MSE in price (normalized by the forward price). Error vol is the MSE in implied volatility. We used for the forward moneyness the range [0.8 1.2] for all maturities.

Table 2. Some relevant contributions to calibration in the literature.

Authors	Model	Data set	Norm	One day	Period
Bates (2000)	2SVJ	Daily S&P futures options	Price	No	1988–1993
Bakshi *et al.* (1997)	SV, SVJ	Daily S&P500	Price	Yes	1988–1991
Carr *et al.* (2003)	VG-LVG	Weekly S&P500	Price	Yes	2000–2001
Carr and Wu (2007)	SV, SVJ, SSM	Weekly JPYUSD, GBPUSD	Price norm. Vega	No	1996–2204
Christoffersen *et al.* (2009)	2SV	Weekly S&P500	Implied vol.	No	1990–1995
Huang and Wu (2004)	TCL-SVJSI	Daily S&P500	Price+opt. weights	No	April–May 1996

DM, Deutsche Mark foreign currency options; SV, Heston model; 2SV, double Heston pure diffusion model; CEV-SV, stochastic volatility with CEV volatility process; SVJ, Heston model with jumps in the stock's returns; SVJJ, Heston model with jumps in the stock's returns and in volatility; SVJSI, Heston model with jumps in the stock's returns with stochastic intensity; TCL-SVJSI, time-changed Lévy processes in the jumps of SVJSI; DTIG, discrete time inverse Gaussian pure diffusion model; P-B&S, practitioner B&S model (B&S model with deterministic volatility fitted on polynomials); VG-LVG, variance Gamma and general Levy–variance Gamma; NP-StochVar, non-parametric stochastic variance; DVSDJ, dynamic volatility with separate dynamic jumps.

dependency of the objective function with respect to the parameters and the volatility process is not straightforward.† Another solution to the problem related to the volatility was developed by Carr and Wu (2007), who used the Kalman algorithm to filter out the process. Here also, the optimization procedure remains tedious. They calibrate a two-factor stochastic volatility with jumps model based on two time-changed Levy processes and they also find a dynamics that is degenerate.‡ In conclusion, a desirable feature for an implementation strategy is that the calibration involves the optimization of a function that should be as simple as possible.

In table 2 we summarize some relevant contributions in the literature on calibration. We classify the papers according to the different choices of the model adopted, the data set, the norm used in the penalizing function, the sample period and according to the single day or time series calibration. In light of these results and taking into account the fact that the models at hand have many parameters, in this paper we restrict ourselves to a one-day calibration. In addition, we follow the market practice and we choose the implied volatility norm (6).

3. Multi-factor stochastic volatility models

There is empirical evidence that the dynamics of the implied volatility surface is driven by several factors (see, e.g., Skiadopoulos *et al.* (1999) and Cont and Da Fonseca (2002)). Several authors have attempted to deduce information on the dynamics of the spot volatility from the implied volatility (see, e.g., Schonbucher (1999) and Schweizer and Wissel (2008) and references therein). However, this procedure seems to be quite involved and difficult to exploit since the starting point is usually the model dynamics in terms of the spot volatility. This leads to the extension of the original Heston (1993) stochastic volatility model to a multiple-factor setting.

As shown in our calibration exercise with the Heston (1993) model, we know that the skew effect is controlled via the correlation between stock returns and the volatility process. In order to manage the term structure of the (possibly time-dependent or even stochastic) skew, it is then natural to look for a model where the correlation has a richer structure than the Heston (1993) model.

3.1. *The BiHeston model of Christoffersen* et al. *(2009)*

Following the previous intuition, Christoffersen *et al.* (2009) introduced a model where the diffusion term of the asset is described by a combination of two square-root processes. This specification is also referred to as the BiHeston model:

$$\mathrm{d}F_t = F_t\left(\sqrt{v_t^0}\,\mathrm{d}Z_t^0 + \sqrt{v_t^1}\,\mathrm{d}Z_t^1\right), \tag{7}$$

$$\mathrm{d}v_t^0 = \kappa_0\left(\theta_0 - v_t^0\right)\mathrm{d}t + \sigma_0\sqrt{v_t^0}\,\mathrm{d}W_t^0, \tag{8}$$

†For example, Christoffersen *et al.* (2009) calibrate using a sample of one year option prices and find for the Heston model a MSE based on criterion (6) of 3.98E−4, which is far above what can be obtained when calibrating using only one day (and renders the model useless for trading purposes).

‡See table 4 of Carr and Wu (2007) for the currency pair GBP–USD, where the correlation is found to be equal to −1.

$$\mathrm{d}v_t^1 = \kappa_1\left(\theta_1 - v_t^1\right)\mathrm{d}t + \sigma_1\sqrt{v_t^1}\,\mathrm{d}W_t^1, \tag{9}$$

with $\mathrm{d}Z_t^0\mathrm{d}W_t^0 = \rho_0\mathrm{d}t$ and $\mathrm{d}Z_t^1\mathrm{d}W_t^1 = \rho_1\mathrm{d}t$, while all the other correlations are set to zero in order to give the model analytical tractability.†

The appealing feature of the BiHeston model is the possibility of allowing for stochastic skew, which is a most desired feature of a model in order to take into account the empirical evidence (see, e.g., Carr and Wu (2007) for the FOREX market). In fact, contrary to the Heston (1993) case in which the instantaneous correlation between the asset returns and the volatility process is constant, Christoffersen *et al.* (2009) obtain

$$\mathrm{d}Corr_t(\ln(F_t), vol(F_t)) = \frac{\rho^1\sigma^1 v^1 + \rho^2\sigma^2 v^2}{\sqrt{v^1 + v^2}\sqrt{\sigma^1 v^1 + \sigma^2 v^2}}\mathrm{d}t, \tag{10}$$

where the leverage effect is driven by both volatility factors, which leads to a richer structure for the skew than that of the Heston (1993) model.

As explained by Christoffersen *et al.* (2009), a single-factor stochastic volatility model can generate steep smirks or flat smirks at a given volatility level, but cannot generate both for a given parameterization. On the contrary, the BiHeston model has much more flexibility in controlling the level and slope of the smile. An additional advantage is that two-factor models also provide more flexibility to model the volatility term structure.

3.2. *The Wishart multidimensional stochastic volatility model of Da Fonseca* et al. *(2008)*

Da Fonseca *et al.* (2008) proposed the Wishart multi-dimensional stochastic volatility model (WMSV), which is also an extension of the original Heston (1993) model, but is based on the Wishart process Bru (1991). Within the WMSV model the dynamics for the forward price is given by

$$\mathrm{d}F_t = F_t\,\mathrm{Tr}\left[\sqrt{\Sigma_t}(\mathrm{d}W_t R^\top + \mathrm{d}B_t\sqrt{\mathbb{I} - RR^\top})\right], \tag{11}$$

where Tr is the trace operator, $W_t, B_t \in M_n$ (the set of square matrices) are composed of n^2 independent Brownian motions under the risk-neutral measure (B_t and W_t are independent), $R \in M_n$ represents the correlation matrix and Σ_t belongs to the set of symmetric $n \times n$ positive-definite matrices. In this specification the volatility is multi-dimensional and depends on the elements of the matrix process Σ_t, which is assumed to satisfy the following dynamics:

$$\begin{aligned}\mathrm{d}\Sigma_t = {} & (\Omega\Omega^\top + M\Sigma_t + \Sigma_t M^\top)\mathrm{d}t \\ & + \sqrt{\Sigma_t}\mathrm{d}W_t Q + Q^\top(\mathrm{d}W_t)^\top\sqrt{\Sigma_t},\end{aligned} \tag{12}$$

with $\Omega, M, Q \in M_n$. Equation (12) characterizes the Wishart process investigated by Bru (1991) and recently introduced in finance by Gouriéroux and Sufana (2010). It represents the matrix analogue of the square-root mean-reverting process. In order to grant the strict positivity and the typical mean-reverting feature of the volatility, the matrix M is assumed to be negative definite, while Ω satisfies $\Omega\Omega^\top = \beta Q^\top Q$ with the real parameter $\beta > n-1$ (Bru 1991, p. 747).

Apart from the technicalities that can be found in Da Fonseca *et al.* (2008), it is important to recall the affine property of the WMSV model. This means that the pricing problem can be solved analytically through the usual FFT methodology exactly as in the classic Heston (1993) or BiHeston models. However, the Wishart object is quite different because the factors live in a space that is not the linear orthant as in the usual affine models. Intuitively, the positive definite property of the Wishart matrix involves the (nonlinear) notion of a determinant. The nonlinear state space domain of positive definite symmetric matrices represents a special case of a symmetric cone, a more general state space domain in which one can define linear-affine models (see Grasselli and Tebaldi (2008) for details). In addition, the Wishart technology opens the door to the possibility of allowing for stochastic covariances among the volatility factors.

Da Fonseca *et al.* (2008) showed that the instantaneous correlation between the asset returns and the volatility process is also stochastic as in the BiHeston model. In fact, they obtained

$$\mathrm{d}Corr_t(\ln(F_t), vol(F_t)) = \frac{\mathrm{Tr}[RQ\Sigma_t]}{\sqrt{\mathrm{Tr}[\Sigma_t]}\sqrt{\mathrm{Tr}[Q^T Q\Sigma_t]}}\mathrm{d}t. \tag{13}$$

In order to illustrate the crucial role of the off-diagonal element Σ_t^{12} in the story, let us consider the case where $n=2$ and both matrices Q and M are diagonal. The previous equality reads

$$\begin{aligned}\mathrm{d}Corr_t(\ln(F_t), vol(F_t)) = {} & \frac{R_{11}Q_{11}\Sigma_t^{11} + R_{22}Q_{22}\Sigma_t^{22}}{\sqrt{\Sigma_t^{11} + \Sigma_t^{22}}\sqrt{Q_{11}^2\Sigma_t^{11} + Q_{22}^2\Sigma_t^{22}}}\mathrm{d}t \\ & + \frac{Q_{22}R_{12}\Sigma_t^{12}}{\sqrt{\Sigma_t^{11} + \Sigma_t^{22}}\sqrt{Q_{11}^2\Sigma_t^{11} + Q_{22}^2\Sigma_t^{22}}}\mathrm{d}t.\end{aligned} \tag{14}$$

That is, in the Wishart specification there is an additional degree of freedom w.r.t. the BiHeston model in order to control the stochasticity of (the correlation and, in turn, of) the leverage effect once the pure volatility terms Σ_t^{11} and Σ_t^{22} are fixed (e.g., by fitting the short-term and long-term implied volatility levels). This means that, contrary to the BiHeston model, in the Wishart model one can fit the volatility levels and the skew *separately*. Note that the volatility factor Σ_t^{12} is not constrained to be positive: this means that the slope of the skew may change sign, which is particularly appealing in a FOREX perspective (Carr and Wu 2007).

†In other words, $\mathrm{d}W_t^0\mathrm{d}W_t^1 = \mathrm{d}Z_t^0\mathrm{d}Z_t^1 = \mathrm{d}W_t^0\mathrm{d}Z_t^1 = \mathrm{d}W_t^1\mathrm{d}Z_t^0 = 0$ in order to grant affinity of the infinitesimal generator (see, e.g., Da Fonseca *et al.* (2008).

3.3. The Wishart affine stochastic correlation model of Da Fonseca et al. (2007b)

The typical problem that traders face every day concerns the risk management of the entire book in terms of Greeks and VaR. Of course, in the book there are vanilla options but also exotic, basket and any sort of structured multiple-asset products.

The single-asset stochastic volatility models introduced so far have different calibration properties and traders typically choose one model rather than another according to the particular exotic product they have to price and hedge. For example, for some products, one may consider that the Heston model is good enough, while for other, more sophisticated products that are highly sensitive with respect to the term structure of the skew, one may prefer, for example, the BiHeston or the Wishart model. Basically, for any problem we look for the simplest model that solves that problem.

However, books also contain basket products, so that we need multiple-asset models that have to be consistent with the smile and skew effects on the single-name options. If a multiple-asset model could be calibrated using only vanilla options it would be a very interesting feature as correlation products are illiquid. What is more, we would like to hedge the basket products by building up hedging strategies involving only vanilla options (written on different underlying assets) whenever possible. Once a structured basket product has been priced, we want to know the risk in volatility and skew associated with each vanilla option used in the hedging strategy. However, management of a book of vanilla options written on a single stock should be carried out using a single-asset model. Therefore, it has to be possible to relate the multi-asset model to a single-asset model in order to manage efficiently the hedging associated with vanilla options.

When we look at the entire book and we think in terms of global risk, it follows that we face an aggregation problem for the hedging ratios computed with different models. Thus, we need a general multi-asset stochastic volatility model that can be related to a single-asset model so as to map the information given by the former into the latter.

In this subsection we introduce the Wishart Affine Stochastic Correlation (WASC) model of Da Fonseca *et al.* (2007b), which satisfies all these requirements. In fact, the WASC model is consistent with the smile and skew effects of plain vanilla option prices. Secondly, it is analytically tractable. What is more, the WASC model allows for non-trivial stochastic volatility of asset returns and stochastic correlation of cross-sectional asset returns. Thirdly, we show in the following subsections via an asymptotic argument that we will be able to aggregate risks within the WASC model in order to quantify the hedging ratios on vanilla options. Thus, the WASC model seems to provide a natural framework in order to aggregate risks in view of global portfolio management.

The WASC model of Da Fonseca *et al.* (2007b) consists of an n-dimensional risky asset $F_t = (F_t^1, \ldots, F_t^n)^\top$ whose dynamics is given by

$$\mathrm{d}F_t = \mathrm{diag}[F_t][\sqrt{\Sigma_t}\mathrm{d}Z_t], \tag{15}$$

where $Z_t \in \mathbb{R}^n$ is a vector Brownian motion, while the returns' variance–covariance matrix Σ_t evolves stochastically, according to the Wishart dynamics (12).

The leverage effects and the asymmetric correlation effects are modeled by introducing the following correlation structure among the Brownian motions:

$$\mathrm{d}Z_t = \sqrt{1 - \rho^\top \rho}\, \mathrm{d}B_t + \mathrm{d}W_t \rho,$$

where ρ is a vector of size n, with $\rho \in [-1, 1]^n$ and $\rho^\top \rho \leq 1$ (B_t is a vector Brownian motion under the risk-neutral measure and is independent of W_t). This represents the leverage effect in the multi-asset framework. It should be mentioned that it remains a very parsimonious way of handling it. What is more, with such a correlation structure, the model belongs to the affine model class, thus granting complete analyticity of the model (see Da Fonseca *et al.* (2007b) for further information). We also refer to Da Fonseca *et al.* (2007a), Da Fonseca *et al.* (2011) for intuition on the parameters and basic properties of the WASC model. Here we limit ourselves to mentioning two basic results. First, the stochastic covariation among the returns is associated with the off-diagonal terms of the Wishart matrix, that is

$$\mathrm{d}\langle \ln F_t^i, \ln F_t^j \rangle = \Sigma_t^{ij} \mathrm{d}t. \tag{16}$$

Secondly, the leverage effect on single assets is constant and is modeled by a cocktail of parameters including the correlation vector ρ. For $n = 2$ we obtain

$$\mathrm{d}Corr_t\big(\ln(F_t^1), vol(F_t^1)\big) = \frac{Q_{11}\rho_1 + Q_{21}\rho_2}{\sqrt{Q_{11}^2 + Q_{21}^2}}\mathrm{d}t, \tag{17}$$

which means that, for the skew of vanilla options, the WASC model has the same properties as the Heston (1993) model. It is possible to combine two WMSV models to build a multi-asset framework that is affine and allows for stochastic skewness (Branger and Muck 2009). Obviously, the calibration will be very difficult numerically.

3.4. Option price approximations and implied volatility dynamics

In this subsection we derive some useful approximations of the price of vanilla options based on a volatility-of-volatility expansion for all the multi-factor stochastic volatility models introduced so far.

Using the approach proposed by Benabid *et al.* (2009) and Bensusan (2010) we extend to a multi-dimensional framework (both multi-asset and multi-factor) the approach of Lewis (2000) (see also Medvedev and Scaillet (2007)), based on the characteristic function, which is explicitly known for the affine models. This is

very useful in order to derive our computations and to obtain the asymptotic approximations. Our approach differs from that of Benabid *et al*. (2009) and Bensusan (2010), who consider a development based on high-frequency and low-frequency volatility factors as in Fouque *et al*. (2003) for the Wishart-based model of Da Fonseca *et al*. (2008). In addition, we can write explicitly the joint characteristic function of the asset returns, the variance process and the integrated variance. This function will be used in proposition 3.8, but can also be used for the computation of volatility derivatives, for example. This emphasizes the usefulness of the linearization presented by Grasselli and Tebaldi (2008) in developing analytical aspects of Wishart-based models.

The starting point of our analysis consists of finding the relevant characteristic functions for the different models. The following propositions are quite technical and their proofs are given in appendix A.

3.4.1. Characteristic function and vol-of-vol expansion in the BiHeston model. Consider first the BiHeston model (7), where for convenience we express the underlying in terms of the log-forward price $x_t := \ln(F_t)$, so that (7) becomes

$$\mathrm{d}x_t = -\frac{v_t^0 + v_t^1}{2}\mathrm{d}t + \sqrt{v_t^0}\mathrm{d}Z_t^0 + \sqrt{v_t^1}\mathrm{d}Z_t^1. \tag{18}$$

The (conditional) joint characteristic function of the asset returns, the variance process $V_T = (v_T^0 + v_T^1)$ and the integrated variance process $\int_t^T V_u \mathrm{d}u = \int_t^T (v_u^0 + v_u^1)\mathrm{d}u$ is given by

$$\phi(\lambda_x, \lambda_{V^0}, \lambda_{V^1}, \lambda_{I^0}, \lambda_{I^1}, t, T) = \mathbb{E}_t\left[\mathrm{e}^{i\lambda_x x_T + i(\lambda_{V^0} v_T^0 + \lambda_{V^1} v_T^1) + i\int_t^T (\lambda_{I^0} v_u^0 + \lambda_{I^1} v_u^1)\mathrm{d}u}\right].$$

Since the model is affine, it is natural to look for an exponentially affine form for the characteristic function. This is indeed the case, as shown in the following proposition whose standard proof is omitted.

Proposition 3.1:

$$\begin{aligned}\phi(\lambda_x, \lambda_{V^0}, \lambda_{V^1}, \lambda_{I^0}, \lambda_{I^1}, t, T) &= \mathbb{E}_t\left[\mathrm{e}^{i\lambda_x x_T + i(\lambda_{V^0} v_T^0 + \lambda_{V^1} v_T^1) + i\int_t^T (\lambda_{I^0} v_u^0 + \lambda_{I^1} v_u^1)\mathrm{d}u}\right] \\ &= \mathrm{e}^{i\lambda_x x_t + A^0(t,T)v_t^0 + B^0(t,T) + A^1(t,T)v_t^1 + B^1(t,T)},\end{aligned} \tag{19}$$

where the deterministic functions $A^j(t,T)$ *and* $B^j(t,T)$, $j = 0, 1$, *are given by*

$$A^j(t,T) = \frac{\eta_j \lambda_+^j \mathrm{e}^{-\sqrt{\Gamma_j}(T-t)} + \lambda_-^j}{(\sigma_j^2/2)(\eta_j \mathrm{e}^{-\sqrt{\Gamma_j}(T-t)} + 1)}, \tag{20}$$

$$B^j(t,T) = -2\frac{\kappa_j \theta_j}{\sigma_j^2}\log\left(\frac{\eta_j \mathrm{e}^{-\sqrt{\Gamma_j}(T-t)} + 1}{1 + \eta_j}\right), \tag{21}$$

with

$$\lambda_\pm^j = \frac{(\kappa_j - i\rho_j\sigma_j\lambda_x) \pm \sqrt{\Gamma_j}}{2}, \tag{22}$$

$$\Gamma_j = (\kappa_j - i\rho_j\sigma_j\lambda_x)^2 - 2i\sigma_j^2\lambda_{I^j}, \tag{23}$$

$$\eta_j = -\frac{i\sigma_j^2\lambda_{V^j} - 2\lambda_-^j}{i\sigma_j^2\lambda_{V^j} - 2\lambda_+^j}. \tag{24}$$

Expression (19) is very useful: we recover the characteristic function of the asset returns simply by keeping $\lambda_V = \lambda_I = 0$, while if we take $\lambda_x = 0$ we obtain the relevant characteristic functions for pricing pure volatility products (such as variance swaps).

Using standard techniques (see, e.g., Lewis (2000)) it is easy to find a vol-of-vol approximation of the implied volatility in the BiHeston model. We perturb the volatility of volatility with a scale factor α and we can show the following result.

Proposition 3.2: *For a short time to maturity the implied volatility expansion in terms of the vol-of-vol scale factor* α *in the BiHeston model is given by*

$$\begin{aligned}\sigma_{\mathrm{imp}}^2 = v^0 + v^1 &+ \frac{\rho_0\sigma_0 v^0 + \rho_1\sigma_1 v^1}{2(v^0 + v^1)} m_f\alpha + \frac{m_f^2\alpha^2}{12(v^0 + v^1)^2} \\ &\times \left(\sigma_0^2 v^0 + \sigma_1^2 v^1 + 2(\rho_0^2\sigma_0^2 v^0 + \rho_1^2\sigma_1^2 v^1)\right. \\ &\left. - \frac{15}{4}\frac{(\rho_0\sigma_0 v^0 + \rho_1\sigma_1 v^1)^2}{(v^0 + v^1)}\right),\end{aligned} \tag{25}$$

where $m_f := \ln(K/F_t)$ *denotes the log-forward moneyness.*

Note that this development links the slope of the implied volatility to the correlation structure. What is more, the volatility factors v^0 and v^1 impact both the level and the slope of the smile. Finally, the smile is not convex for all correlation structures, which is well known in the Heston model (see, e.g., Lewis (2000)). For convenience, we recall below the expansion for the Heston model that can be obtained from (25) by simplifying one of the stochastic processes:

$$\sigma_{\mathrm{imp}}^2 = v^0 + \frac{\rho_0\sigma_0}{2} m_f + \frac{m_f^2\sigma_0^2}{12(v^0)^2}(4 - 7\rho_0^2). \tag{26}$$

We illustrate the accuracy of this formula in figure 4.

3.4.2. The joint characteristic function for the WMSV model. We express as before the underlying in terms of the log-forward price $x_t := \ln(F_t)$, so that (11) becomes

$$\mathrm{d}x_t = -\frac{\mathrm{Tr}[\Sigma_t]}{2}\mathrm{d}t + \mathrm{Tr}[\sqrt{\Sigma_t}(\mathrm{d}W_t R^\top + \mathrm{d}B_t\sqrt{\mathbb{I} - RR^\top})]. \tag{27}$$

The (conditional) joint characteristic function of the asset returns, the (Wishart) variance process Σ_T and the integrated variance process $\int_t^T \Sigma_u \mathrm{d}u$ is parametrized by

a scalar λ_x and two square (symmetric) matrices Λ_V and Λ_I:

$$\phi(\lambda_x, \Lambda_V, \Lambda_I, t, T) = \mathbb{E}_t\left[\mathrm{e}^{i\lambda_x x_T + i\,\mathrm{Tr}[\Lambda_V \Sigma_T] + i\int_t^T \mathrm{Tr}[\Lambda_I \Sigma_u]\mathrm{d}u}\right].$$

Using the Feynman–Kac methodology we can prove the following.

Proposition 3.3:

$$\phi(\lambda_x, \Lambda_V, \Lambda_I, t, T) = \mathrm{e}^{i\lambda_x x_t + \mathrm{Tr}[A(t,T)\Sigma_t] + b(t,T)}, \tag{28}$$

where the deterministic matrix $A(t,T)$ and the scalar function $b(t,T)$ satisfy the following ODE:

$$\begin{aligned}-\frac{\partial}{\partial t}A(t,T) &= A(t,T)(M + i\lambda_x Q^\top R^\top) \\ &\quad + (M + i\lambda_x Q^\top R^\top)^\top A(t,T) \\ &\quad + 2A(t,T)Q^\top Q A(t,T) + \frac{i\lambda_x(i\lambda_x - 1)}{2}\mathbb{I} + i\Lambda_I,\end{aligned} \tag{29}$$

$$-\frac{\partial}{\partial t}b(t,T) = \mathrm{Tr}[\Omega\Omega^\top A(t,T)], \tag{30}$$

with boundary conditions $A(T,T) = i\Lambda_V$ and $b(T,T) = 0$, whose solution is given by ($\tau = T - t$)

$$A(\tau) = (i\Lambda_V A_{12}(\tau) + A_{22}(\tau))^{-1}(i\Lambda_V A_{11}(\tau) + A_{21}(\tau)), \tag{31}$$

$$b(\tau) = -\frac{\beta}{2}\mathrm{Tr}[\log(i\Lambda_V A_{12}(\tau) + A_{22}(\tau)) + \tau(M + i\lambda_x Q^\top R^\top)], \tag{32}$$

with

$$\begin{aligned}&\begin{pmatrix} A_{11}(\tau) & A_{12}(\tau) \\ A_{21}(\tau) & A_{22}(\tau)\end{pmatrix} \\ &\quad = \exp\tau\begin{pmatrix} M + i\lambda_x Q^\top R^\top & -2Q^\top Q \\ \dfrac{i\lambda_x(i\lambda_x - 1)}{2}\mathbb{I}_n + i\Lambda_I & -(M + i\lambda_x Q^\top R^\top)^\top\end{pmatrix}.\end{aligned} \tag{33}$$

Proof: See appendix A. □

3.4.3. The joint characteristic function for the WASC model. Now let us turn our attention to the model of Da Fonseca *et al.* (2007b), whose vector of assets returns is given by

$$\mathrm{d}X_t = -\frac{1}{2}\mathrm{Vec}[\Sigma_t^{jj}]\mathrm{d}t + \sqrt{\Sigma_t}\mathrm{d}Z_t,$$

where $\mathrm{Vec}[\Sigma_t^{jj}]$ is a vector containing the diagonal elements of the matrix Σ.

The (conditional) joint characteristic function of the assets returns X, the (Wishart) variance process Σ_T and the integrated variance process $\int_t^T \Sigma_u \mathrm{d}u$ is parametrized by a vector Λ_x and two square (symmetric) matrices Λ_V and Λ_I:

$$\phi(\Lambda_x, \Lambda_V, \Lambda_I, t, T) = \mathbb{E}_t\left[\mathrm{e}^{i\langle\Lambda_x, X_T\rangle + i\,\mathrm{Tr}[\Lambda_V\Sigma_T] + i\int_t^T \mathrm{Tr}[\Lambda_I\Sigma_u]\mathrm{d}u}\right],$$

where $\langle .,. \rangle$ denotes the scalar product in $\mathbb{R}^n$. Using the same Feynman–Kac-based methodology as before we can prove the following version of the result presented by Da Fonseca *et al.* (2007b).

Proposition 3.4:

$$\phi(\Lambda_x, \Lambda_V, \Lambda_I, t, T) = \mathrm{e}^{i\langle\Lambda_x X_t\rangle + \mathrm{Tr}[A(t,T)\Sigma_t] + b(t,T)}, \tag{34}$$

where the deterministic matrix $A(t,T)$ and the scalar function $b(t,T)$ satisfy the following ODE:

$$\begin{aligned}-\partial_t A(t,T) &= A(t,T)(M + iQ^\top\rho\Lambda_x^\top) \\ &\quad + (M + iQ^\top\rho\Lambda_x^\top)^\top A(t,T) \\ &\quad + 2A(t,T)Q^\top Q A(t,T) - \frac{1}{2}\Lambda_x\Lambda_x^\top \\ &\quad - \frac{1}{2}\sum_{j=1}^n \Lambda_x^j \mathrm{e}^{jj} + i\Lambda_I,\end{aligned} \tag{35}$$

$$-\partial_t b(t,T) = \mathrm{Tr}[\Omega\Omega^\top A(t,T)], \tag{36}$$

with boundary conditions $A(T,T) = i\Lambda_V$ and $b(T,T) = 0$, whose solution is given by ($\tau = T - t$)

$$A(\tau) = (i\Lambda_V A_{12}(\tau) + A_{22}(\tau))^{-1}(i\Lambda_V A_{11}(\tau) + A_{21}(\tau)), \tag{37}$$

$$b(\tau) = -\frac{\beta}{2}\mathrm{Tr}[\log(i\Lambda_V A_{12}(\tau) + A_{22}(\tau)) + \tau(M + iQ^\top\rho\Lambda_x^\top)^\top], \tag{38}$$

with

$$\begin{aligned}&\begin{pmatrix} A_{11}(\tau) & A_{12}(\tau) \\ A_{21}(\tau) & A_{22}(\tau)\end{pmatrix} \\ &\quad = \exp\tau\begin{pmatrix} M + iQ^\top\rho\Lambda_x^\top & -2Q^\top Q \\ -\dfrac{1}{2}(\Lambda_x\Lambda_x^\top + \sum_{j=1}^n \Lambda_x^j \mathrm{e}^{jj}) + i\Lambda_I & -(M + iQ^\top\rho\Lambda_x^\top)^\top\end{pmatrix}.\end{aligned} \tag{39}$$

Proof: We apply the proof of proposition 3.3 to Da Fonseca *et al.* (2007b) and immediately obtain the result. □

3.5. Approximations in the Wishart model of Da Fonseca et al. (2008)

3.5.1. Price approximation. We consider a perturbation of the volatility-of-volatility matrix Q by a scale factor $\alpha \in \mathbb{R}$. We will show, in full analogy with the Heston model (Lewis 2000), that it is possible to obtain an approximation of the vanilla option price in terms of the vol-of-vol scale factor by simply differentiating the classic Black–Scholes formula.

We denote by $c_{\mathrm{B\&S}}(t,T,x,K,\xi)$ the Black–Scholes price when $x = \log(S_t \mathrm{e}^{r(T-t)})$ is the log-forward price and $\xi = \sigma^2(T-t)$ is the integrated volatility. Subscripts denote partial derivatives, e.g. $\partial^2_{\xi^2} c_{\mathrm{B\&S}}(t,T,x,K,\xi)$ stands for $(\partial^2/\partial\xi^2)c_{\mathrm{B\&S}}(t,T,x,K,\xi)$, etc.

Using the affine property of the Wishart model we arrive at the following second-order expansion of the price.

Proposition 3.5: *The call price* $Call(t, T, x, K)$ *in the model of Da Fonseca* et al. (2008) *can be approximated in terms of the vol-of-vol scale factor* α *by differentiating the Black–Scholes formula*:

$$\begin{aligned} &Call(t, T, x, K) \\ &= c_{\mathrm{B\&S}}(t, T, x, K, \xi_0) + \alpha(\operatorname{Tr}[\tilde{A}^1(\tau)\Sigma_t] \\ &\quad + \tilde{b}^1(\tau))\partial^2_{x\xi} c_{\mathrm{B\&S}}(t, T, x, K, \xi_0) + \alpha^2(\operatorname{Tr}[\tilde{A}^{20}(\tau)\Sigma_t] \\ &\quad + \tilde{b}^{20}(\tau))\partial^2_{\xi^2} c_{\mathrm{B\&S}}(t, T, x, K, \xi_0) + \alpha^2(\operatorname{Tr}[\tilde{A}^{21}(\tau)\Sigma_t] \\ &\quad + \tilde{b}^{21}(\tau))\partial^3_{x^2\xi} c_{\mathrm{B\&S}}(t, T, x, K, \xi_0) + \frac{\alpha^2}{2}(\operatorname{Tr}[\tilde{A}^1(\tau)\Sigma_t] \\ &\quad + \tilde{b}^1(\tau))^2\partial^4_{x^2\xi^2} c_{\mathrm{B\&S}}(t, T, x, K, \xi_0) + o(\alpha^2), \end{aligned} \tag{40}$$

where τ *denotes the time to maturity and* $\xi_0 = (\operatorname{Tr}[\tilde{A}^0(T-t)\Sigma_t] + \tilde{b}^0)$ *is the integrated variance. The real deterministic matrices* $\tilde{A}^0$, $\tilde{A}^1$, $\tilde{A}^{20}$ *and* $\tilde{A}^{21}$ *are given by*

$$\tilde{A}^0(\tau) = \int_0^\tau e^{(\tau-u)M^\top} e^{(\tau-u)M} du, \tag{41}$$

$$\begin{aligned} \tilde{A}^1(\tau) = \int_0^\tau e^{(\tau-u)M^\top} &\{\tilde{A}^0(u)(Q^\top R^\top) \\ &+ (Q^\top R^\top)^\top \tilde{A}^0(u)\} e^{(\tau-u)M} du, \end{aligned} \tag{42}$$

$$\tilde{A}^{20}(\tau) = \int_0^\tau e^{(\tau-u)M^\top} \{2\tilde{A}^0(u)Q^\top Q\tilde{A}^0(u)\} e^{(\tau-u)M} du, \tag{43}$$

$$\begin{aligned} \tilde{A}^{21}(\tau) = \int_0^\tau e^{(\tau-u)M^\top} &\{\tilde{A}^1(u)(Q^\top R^\top) \\ &+ (Q^\top R^\top)^\top \tilde{A}^1(u)\} e^{(\tau-u)M} du, \end{aligned} \tag{44}$$

and the scalar functions $\tilde{b}^0$, $\tilde{b}^1$, $\tilde{b}^{20}$ *and* $\tilde{b}^{21}$ *are given by*

$$\begin{aligned} \tilde{b}^0(\tau) &= \operatorname{Tr}\left[\Omega\Omega^\top \int_0^\tau \tilde{A}^0(u) du\right], \\ \tilde{b}^1(\tau) &= \operatorname{Tr}\left[\Omega\Omega^\top \int_0^\tau \tilde{A}^1(u) du\right], \\ \tilde{b}^{20}(\tau) &= \operatorname{Tr}\left[\Omega\Omega^\top \int_0^\tau \tilde{A}^{20}(u) du\right], \\ \tilde{b}^{21}(\tau) &= \operatorname{Tr}\left[\Omega\Omega^\top \int_0^\tau \tilde{A}^{21}(u) du\right]. \end{aligned}$$

Proof: See appendix A. □

Thus, in order to find a price approximation it is sufficient to differentiate the B&S formula and compute its derivatives with the argument $\xi_0 = (\operatorname{Tr}[\tilde{A}^0\Sigma_t] + \tilde{b}^0)$ (the integrated variance), which can be done very efficiently.†

3.5.2. Implied volatility expansion for short time to maturity. Price approximations are usually considered interesting because they lead very quickly to a proxy for the true price, or because they give good starting points for the calibration procedure. Unfortunately, as in the Heston model, it turns out that the formulas work well for a very narrow interval around the ATM, thus they cannot be used for a true calibration scope. Nevertheless, they indeed remain interesting in order to obtain intuition concerning the relation between model parameters and implied volatility dynamics. In fact, using an expansion in vol of vol for a short time to maturity, we are able to recover explicitly the relation between the parameters and the slope and curvature of the implied volatility smile/skew. This fact is interesting in itself since it represents a useful tool for comparing different models by aggregating parameters.

By definition, the implied volatility is such that

$$c_{\mathrm{B\&S}}(t, T, x, K, \sigma^2_{\mathrm{imp}}\tau) = Call(t, T, x, K), \tag{45}$$

where the right-hand side is the option price computed within the WMSV model.

Proposition 3.6: *For a short time to maturity the implied volatility expansion in terms of the vol-of-vol scale factor* α *in the model of Da Fonseca* et al. (2008) *is given by*

$$\begin{aligned} \sigma^2_{\mathrm{imp}} = \Sigma_t^{11} + \Sigma_t^{22} &+ \alpha\frac{\operatorname{Tr}[RQ\Sigma_t]}{\operatorname{Tr}[\Sigma_t]} m_f + \alpha^2 \frac{m_f^2}{(\operatorname{Tr}[\Sigma_t])^2} \\ &\times \left(\frac{1}{3}\operatorname{Tr}[Q^\top Q\Sigma] + \frac{1}{3}\operatorname{Tr}[RQ(Q^\top R^\top \right. \\ &\left. + RQ)\Sigma] - \frac{5}{4}\frac{(\operatorname{Tr}[RQ\Sigma])^2}{\operatorname{Tr}[\Sigma_t]}\right), \end{aligned} \tag{46}$$

where $m_f := \ln(K/F_t)$ *denotes the log-forward moneyness.*

Proof: See appendix A. □

The first-order term in the development is linked to the slope of the implied volatility and is controlled by the correlation structure between the asset returns and the volatility noise. By comparing this expansion with that in proposition 3.2 of the BiHeston model, we see that, here, the slope of the implied volatility depends on the off-diagonal element Σ_t^{12}, which does not affect the level of the smile. That is, the Wishart model allows us to control the level and the slope of the implied volatility separately: this is a nice property that is not shared, for example, by the BiHeston model, where the volatility factors impact both the level and the skew of the implied volatility.

Finally, we note that, in analogy with the Heston model (see, e.g., Lewis (2000) or formula (26)) and the BiHeston model (see formula (25)), the smile is not a convex function of the moneyness for all correlation structures. This is quite disappointing if we reason from a calibration perspective.

†See also Gauthier and Possamai (2009b) for another approximation and Gauthier and Possamai (2009a), Cuchiero *et al.* (2010) and Mayerhofer *et al.* (2010) for theoretical and numerical aspects of the Wishart and related processes.

Table 3. Parameters given by the approximation of the marginals for the EuroStoxx50/DAX for the WASC model calibrated on August 20, 2008.

	EuroStoxx50			DAX		
Maturity:	1 month	2 months	3 months	1 month	2 months	3 months
κ	2.1730	2.1638	2.1544	2.5925	2.5928	2.5930
σ	0.9580	0.9550	0.9520	0.9810	0.9807	0.9805
θ	0.0774	0.0773	0.0771	0.0680	0.0680	0.0680
v_0	0.0446	0.0446	0.0446	0.0424	0.0424	0.0424
ρ	−0.5844	−0.5844	−0.5844	−0.5423	−0.5423	−0.5423

Maturity is for the horizon for which the law of the process is approximated.

3.6. *Approximations in the WASC model of Da Fonseca* et al. *(2007b)*

3.6.1. Implied volatility approximation. We focus on the characteristic function of the first asset and we choose w.l.o.g. $n=2$, therefore $\lambda_x=(\lambda_1,0)^\top$ with $\lambda_1\in\mathbb{C}$. We repeat the reasoning of the previous subsection and arrive at the following proposition.

Proposition 3.7: *For a short time to maturity the implied volatility expansion in terms of the vol-of-vol scale factor α in the model of Da Fonseca* et al. (2007b) *is given by*

$$\sigma^2_{\text{imp}} = \Sigma_t^{11} + \alpha m_f(\rho_1 Q_{11} + \rho_2 Q_{21}) + \frac{1}{2}\alpha^2 m_f^2\left[\frac{4(Q_{11}^2+Q_{21}^2) - 7(\rho_1 Q_{11}+\rho_2 Q_{21})^2}{6\Sigma_t^{11}}\right]. \quad (47)$$

Proof: See appendix A. □

We remark that if we take into account (17) and the quadratic variation of the process Σ_t^{11} given by $\mathrm{d}\langle\Sigma^{11},\Sigma^{11}\rangle_t = 4\Sigma_t^{11}(Q_{11}^2+Q_{21}^2)\mathrm{d}t$ (see Da Fonseca *et al.* (2007a, 2008) for basic computations), then the expansion of proposition 3.7 can be rewritten exactly as (26).

3.6.2. Approximation of the marginals. The previous approximations help us to understand the impact of the parameters on the smile as well as speed up the pricing of options, hence leading to a faster calibration procedure. Unfortunately, they provide an accurate approximation only for a narrow moneyness range, smaller than what is required in practice. This means that we need to mix the exact pricing formula, based on the FFT, and the approximations. It is possible to take advantage of the fact that the law of the volatilities Σ_t^{ii} is known explicitly and allows us to build a synthetic Heston model that leads to accurate price approximations. From a computational point of view we can replace the computation of the characteristic function of the WASC, which involves functions of matrices, by the computation of the characteristic function of the Heston model, which involves only simple functions.

In the Heston model, which is given by equations (7) and (8) and setting $v_t^1=0$, we know that $(v_T^0/L)|v_t^0 \sim \chi^2(\delta,\tilde{\zeta})$† with $L=(\sigma_0^2/\kappa_0)(1-\mathrm{e}^{-\kappa_0(T-t)})$, $\delta = 4\kappa_0\theta_0/\sigma_0^2$ and $\tilde{\zeta}=(v_t^0/L)\mathrm{e}^{-\kappa_0(T-t)}$. Within the WASC model the volatility of the first asset is given by Σ_t^{11} whose law is known in closed form (see, for example, Letac and Massam (2004)) and is given in the following.

Proposition 3.8: *There exist a_{11}, β and N_2^{11}, three scalars depending on the model parameters such that* $(2\Sigma_T^{11}/a_{11})|\Sigma_t^{11} \sim \chi^2(\beta,N_2^{11})$.

Proof: See appendix A. □

Combining this proposition and the results for the Heston model we can state the following equalities: $L=a_{11}/2$, $\delta=\beta$ and $\tilde{\zeta}=N_2^{11}$. From the first we can deduce σ_0 the volatility of volatility, from the second we deduce θ_0 and the last gives κ_0. The starting value of the volatility process is obviously given by $v_t^0=\Sigma_t^{11}$. We still need to specify the correlation ρ_0 that controls the leverage effect and using relation (17) it is natural to choose $\rho_0=(Q_{11}\rho_1+Q_{21}\rho_2)/\sqrt{Q_{11}^2+Q_{21}^2}$.

We show in figure 5 for the calibrated parameters a comparison between the vanilla prices produced by the WASC model and the synthetic Heston models. We can check that the accuracy is good for a larger range of forward moneyness with much less computational cost. Also, table 3 reports the associated parameters for the WASC model for the index pair EuroStoxx50/DAX, where we can check that, for each maturity, we have a different Heston model.

4. Calibration analysis and model mapping

We now analyse the calibration performance of the stochastic volatility models introduced so far. As expected, the Heston model leads to the largest calibration error irrespective of the underlying index (see table 4). The parameters are consistent with the existing findings of Bakshi *et al.* (1997). The correlation is negative, implying the leverage effect and the smirk shape of the smile. The volatility of volatility is rather high compared with estimates based on time series. The initial volatility is close to the ATM implied volatility, while the long-term spot volatility θ is consistent with the

†$\chi^2(k,\lambda)$ is the non-central chi-square distribution with degrees of freedom k and non-centrality parameter λ.

Table 4. Calibration on August 20, 2008 with the Heston (1993) model.

Parameter	DAX	FTSE	EuroStoxx50
κ	1.4078	1.7545	1.5532
σ	0.9319	0.8484	0.7439
ρ	−0.5409	−0.5770	−0.5986
θ	0.0838	0.0823	0.0797
v_0	0.0414	0.0376	0.0416
Error vol	1.09E−4	8.45E−05	1.00E−04
Error price	2.68E−06	2.16E−06	4.33E−06

MinTime is the shortest maturity in the sample (0.06 = 3 weeks). Error vol denotes the MSE in implied volatility. Error price is the MSE in price (normalized by the forward price).

Table 5. Calibration on August 20, 2008 with the BiHeston model.

Parameter	DAX	FTSE	EuroStoxx50
κ_0	1.3080	1.3123	1.1636
σ_0	1.1202	0.9736	0.8036
ρ_0	−0.3884	−0.4194	−0.8815
θ_0	0.0281	0.0299	0.0454
v_t^0	0.0187	0.0168	0.0287
κ_1	1.4134	1.4919	1.9936
σ_1	0.4822	0.5159	0.8388
ρ_1	−0.8395	−0.8862	−0.2896
θ_1	0.0485	0.0504	0.0399
v_t^1	0.0229	0.0211	0.0159
Error vol	7.60E−05	7.56E−05	8.42E−05
Error price	1.06E−06	2.15E−06	4.48E−06

MinTime is the shortest maturity in the sample (0.06 = 3 weeks). Error vol denotes the MSE in implied volatility. Error price is the MSE in price (normalized by the forward price).

long-term implied volatility. We also note that the Feller condition is not satisfied.†

The BiHeston model performs systematically better than the Heston model and this is not a surprise (see table 5). The salient result is the two-regime property of the calibrated dynamics. We have a low correlation (in absolute value terms) and a high volatility-of-volatility regime that we associate with the short-term smile. We have a high correlation (in absolute value terms) and low volatility-of-volatility regime that we associate with the long-term smile. This interpretation relies on the results found for the Heston model and reported in table 1. The short-term regime allows us to better capture the short-term smile, which is more convex. This is the main advantage of the multi-factor feature of this model.

In order to relate the calibrated parameters to those obtained for the Heston model we use the implied volatility expansion (25) and we report in tables 8, 9 and 10 the corresponding numerical values. For all the indexes the two models provide the same implied volatility expansions. This allows us to ascertain the consistency between the two calibrated models. When conveniently aggregated the parameters of the BiHeston model lead to those of the Heston model.

To relate this to the existing literature we refer to Christoffersen *et al.* (2009), who perform a calibration on a time series of option prices on the S&P500. The calibration is based on a one-year window and a norm in price normalized by the Vega. If we compare our results with the results of Christoffersen *et al.* (2009), albeit using a different index, we find the following differences. From 1997 to 2004, they find a correlation parameter value for one of the volatility processes to be nearly equal to -0.99. One of the mean-reverting parameters is very low (around 0.15)‡, in contrast to our values, which are around 1.3 for all the processes and indexes. While our results for the Heston and BiHeston models are consistent, it seems to us that this is not the case with theirs. This might be related to the numerical complexity of their calibration objective function, which involves a large number of optimization problems. We point out that we were unable to calibrate the BiHeston model (as well as the subsequent WMSV model) using a norm in price and the reason for this is obvious. As previously mentioned, this norm places little weight on short-term options and discarding them leads to a smile with a simpler skew term structure for which the Heston model performs very well. Adding another volatility process makes the problem numerically unstable. Using a two-factor stochastic volatility Lévy process model, Carr and Wu (2007) performed a calibration on a time series of FX option prices. They advocate a multi-factor volatility model to take into account the stochastic nature of the skew in this market. They rightfully argue that the Heston model cannot handle this property, and the implied volatility expansion (26) confirms this fact. Their model, which is quite similar to a BiHeston model, leads to a correlation parameter ρ_1 equal to -1 when calibrated on the GBPUSD market. As such, the dynamics is degenerate. It is not clear whether this result means the model is not adapted to this options data set or if it is due to the numerical complexity of the objective function which jeopardizes the calibration. As a final remark, from table 5 we conclude that the Feller condition is not satisfied for this model.

For the Wishart Multidimensional Stochastic Volatility model we obtain a calibration performance slightly better than the BiHeston model and obviously much better than the Heston model (see table 6 and figure 3). The matrix M is negative definite, which implies mean-reverting behavior for the Wishart process. What is more, the matrices R and Q are non-diagonal and according to equation (14) entail an impact of Σ_t^{12}, the off-diagonal term of the Wishart process, on the dynamics of the correlation. To further analyse the calibrated parameters it is better to rely on the short-term implied volatility expansion presented in proposition 3.6. We report in table 8 the implied volatility expansion for the three models given by equations (26), (25) and (46) for the calibrated parameters for the DAX. The models lead to the same expansion up to the first-order development, which means that they induce for the implied volatility smile the same short-term

†Feller's condition ensures that the square-root process does not reach zero. It is given by $2\theta\kappa > \sigma^2$.
‡This result is very similar to Bates (2000) where a value equal to zero is found.

Table 6. Calibration on August 20, 2008 with the WMSV model.

Parameter	DAX	FTSE	EuroStoxx50
Σ_t^{11}	0.0298	0.0285	0.0327
Σ_t^{12}	0.0119	0.0092	0.0069
Σ_t^{22}	0.0108	0.0098	0.0089
β	0.5776	0.6705	0.6229
M_{11}	−1.2479	−1.0270	−0.9858
M_{12}	−0.8985	−0.3764	−0.5224
M_{21}	−0.0820	0.0860	−0.1288
M_{22}	−1.1433	−0.6330	−0.9746
Q_{11}	0.3417	0.3365	0.3193
Q_{12}	0.3493	0.1703	0.2590
Q_{21}	0.1848	0.2332	0.2899
Q_{22}	0.3090	0.2121	0.2469
R_{11}	−0.2243	−0.3829	−0.2116
R_{12}	−0.1244	−0.1408	−0.4428
R_{21}	−0.2545	−0.1631	−0.2113
R_{22}	−0.7230	−0.7509	−0.5921
Error vol	7.44E−05	6.95E−05	5.89E−05
Error price	1.34E−06	1.33E−06	2.28E−06

MinTime is the shortest maturity in the sample (0.06 = 3 weeks). Error vol denotes the MSE in implied volatility. Error price is the MSE in price (normalized by the forward price).

Table 7. Calibration on August 20, 2008 with the WASC model.

Parameter	EuroStoxx50/ DAX	FTSE/ DAX	FTSE/ EuroStoxx50
Σ_t^{11}	0.0446	0.0387	0.0382
Σ_t^{12}	0.0366	0.0320	0.0304
Σ_t^{22}	0.0424	0.0422	0.0441
β	0.7332	0.7055	0.8247
M_{11}	−0.7820	−0.6487	−0.9886
M_{12}	−0.3772	−0.3600	−0.3631
M_{21}	−0.0539	−0.0567	−0.4464
M_{22}	−1.2497	−1.2240	−0.7599
Q_{11}	0.3898	0.3932	0.3296
Q_{12}	0.3573	0.3603	0.2866
Q_{21}	0.2809	0.2552	0.3446
Q_{22}	0.3362	0.3437	0.3524
ρ_1	−0.6407	−0.5305	−0.2675
ρ_2	−0.1105	−0.2228	−0.5496
Error vol	7.47E−05	7.99E−05	7.81E−05
Error vol	9.08E−05	9.05E−05	7.20E−05
Error price	3.64E−06	1.77E−06	1.85E−06
Error price	2.74E−06	2.70E−06	3.64E−06

MinTime is the shortest maturity in the sample (0.06 = 3 weeks). Error vol denotes the MSE in implied volatility. Error price is the MSE in price (normalized by the forward price). For the index pair A/B, the first 'Error vol' is for A and the second 'Error vol' for B. The same convention applies to 'Error price'.

value and slope. We obtain similar results for the FTSE (table 9) and EuroStoxx50 (table 10), hence the robustness of the conclusions. For the second-order term, whilst the Heston and BiHeston give the same coefficient, the WMSV leads systematically to a larger coefficient. Those expansions emphasize how vanilla options aggregate the

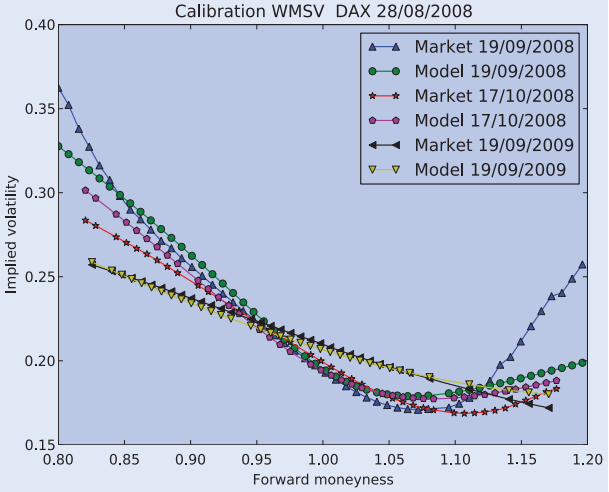

Figure 3. Calibration for the WMSV for the DAX on 28/08/2008.

model parameters and confirm the consistency of the values. Even though we obtain similar values numerically, the expansions contain more interesting information. The BiHeston model possesses the stochastic skew property, as can be deduced from equation (25), hence the standard affine Duffie–Kan process provides a convenient framework for modeling this feature. For the WMSV we have an additional striking property, namely the factor Σ_t^{12} that is specific to the slope. These two models allow for a more subtle description of the smile shape and dynamics. We will show below how this turns out to be an advantage.

When dealing with the Wishart process, the analog of the Feller condition is called the Gindikin condition and is related to the value of β, which in the two-dimensional case should be greater than one. From table 6 we deduce that this condition is not satisfied.† In conclusion, whether we use a Heston model, a BiHeston model or a WSMV model we face the same problem, namely the process might reach zero (or the boundary of the cone of symmetric positive definite matrices in the case of the Wishart process). We are not surprised by this property, although it might be a problem in some practical situations. For example, there are derivatives whose payoff depends on the time spent by the volatility process close to zero. Those products will be mispriced by the above models and might justify the market practice of using processes for the volatility that are positive by *construction*. If we want to gauge the mean-reverting parameter, then it is tempting to look along a volatility path (we discard the fact that the volatility is not observable). Unfortunately, vanilla options contain integrated volatility and the integration produces a smoothing effect with respect to this parameter. As such, options can only lead to a small value for the mean-reverting parameter. This problem can become even worse if we use

†This fact is consistent with the feedback we have from practitioners. In this respect we wish to thank Olivier Croissant from Natixis for useful remarks.

Table 8. Aggregated parameters of the short-term implied volatility expansion on August 20, 2008 for the DAX.

				WASC	
Parameter	Heston	BiHeston	WMSV	EuroStoxx50/DAX	DAX/FTSE
Order 0	0.0413	0.0417	0.0405	0.0423	0.0422
Order 1	−0.2520	−0.2092	−0.2554	−0.2660	−0.2677
Order 2	0.8537	0.7724	1.7919	0.9185	0.9667

MinTime is the shortest maturity in the sample (0.06 = 3 weeks). Error vol denotes the MSE in implied volatility. Error price is the MSE in price (normalized by the forward price).

Table 9. Aggregated parameters of the short-term implied volatility expansion on August 20, 2008 for the FTSE.

				WASC	
Parameter	Heston	BiHeston	WMSV	DAX/FTSE	FTSE/EuroStoxx50
Order 0	0.0376	0.0379	0.0383	0.0386	0.0381
Order 1	−0.2447	−0.2177	−0.2461	−0.2654	−0.2775
Order 2	0.6649	0.5208	1.0409	0.8310	0.8086

MinTime is the shortest maturity in the sample (0.06 = 3 weeks). Error vol denotes the MSE in implied volatility. Error price is the MSE in price (normalized by the forward price).

Table 10. Aggregated parameters of the short-term implied volatility expansion on August 20, 2008 for the EuroStoxx50.

				WASC	
Parameter	Heston	BiHeston	WMSV	DAX/EuroStoxx50	EuroStoxx50/FTSE
Order 0	0.0416	0.0446	0.0416	0.0446	0.0441
Order 1	−0.2226	−0.2710	−0.2637	−0.2807	−0.2703
Order 2	0.4134	0.4697	1.3379	0.6936	0.5920

MinTime is the shortest maturity in the sample (0.06 = 3 weeks). Error vol denotes the MSE in implied volatility. Error price is the MSE in price (normalized by the forward price).

a norm in price in the calibration procedure. As previously mentioned, this norm places little weight on short-term options and discarding them leads to a smile with a flat skew term structure (along the time-to-maturity axis). In that case, calibration of the Heston model leads to smaller values for the mean-reverting parameter (see table 1). Furthermore, if we are interested in the market price of volatility risk (for example, for portfolio management purposes), then we need to estimate the mean-reverting parameter under both the risk-neutral probability and the historical probability. Under the latter, the parameter is estimated using a volatility path with the realized volatility as a proxy for the volatility.† In that case, there is no smoothing effect related to the integration. The resulting mean-reverting parameter is typically much larger, thereby leading to a large market price of volatility risk with, as a consequence, large hedging ratios (Da Fonseca *et al.* 2011, Egloff *et al.* 2010).

We turn our attention to the WASC model. We carried out the calibration using two sets of vanilla options, one for each index and report the results in Table 7. The first result is that the error on each set is lower than that achieved using the Heston model. This is clearly not surprising since the WASC model is a multi-factor stochastic volatility model. For a given index the calibration error does not depend on which other index is involved in the estimation procedure. For all the indexes, the matrix M is negative definite, hence ensuring mean-reverting behavior for the Wishart process. To further understand the estimated parameters we use the implied volatility expansion (47) and, depending on the index, we report in tables 8, 9, and 10 the numerical results. The WASC model gives the same implied volatility expansion up to first order as the three previous models for both indexes. The short-term skews produced by the models are identical. Discrepancies appear for the second-order coefficient, the WASC leading to a value between the Heston model and the WMSV model. Once again, this confirms the interpretation based on the aggregation of the parameters developed by Da Fonseca *et al.* (2007a). In this latter work we estimated the WASC model using time

†A Kalman filter can also be used, but this leads to the same problem: with respect to the realized volatility we face serious difficulties. In fact, we tried to use the realized volatility computed with high-frequency data in the calibration procedure; we obtained inconsistencies between the information content of the option prices and the statistical properties of the high-frequency data. We will delve more deeply into this problem in future work.

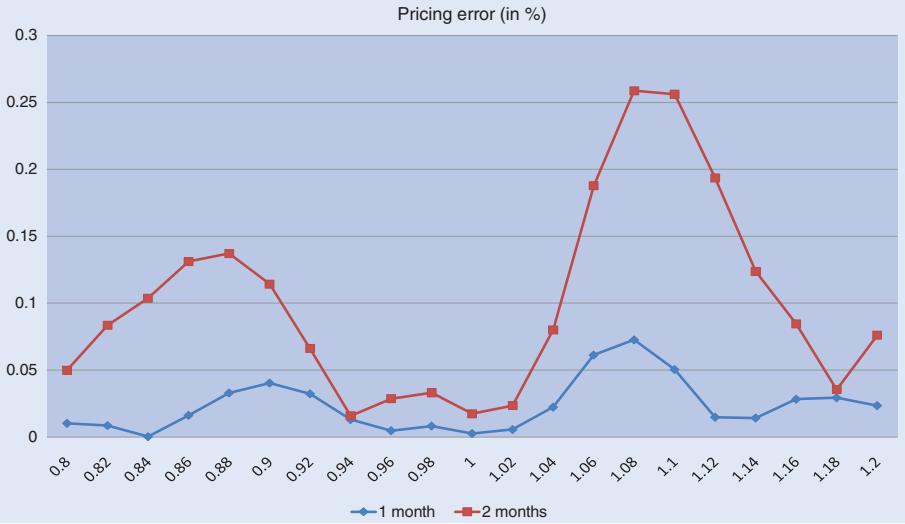

Figure 4. Absolute pricing error between the exact formula (FFT) and the second-order approximation for the Heston model and the calibrated parameters for the DAX on 28/08/2008 ($r=0$, $S_t=1$).

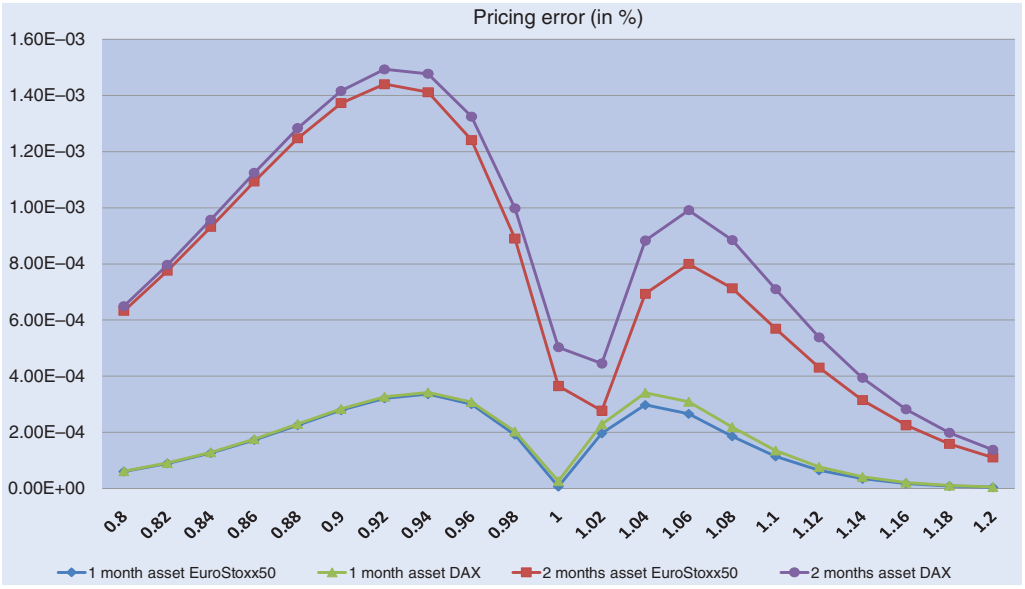

Figure 5. Absolute pricing error between the exact formula (FFT) and the approximation of the marginals for the WASC model and the calibrated parameters for the EuroStoxx50/DAX on 28/08/2008 ($r=0$, $S_t=1$).

series of indexes and the empirical characteristic function methodology proposed by Carrasco *et al.* (2007) and by aggregating the parameters we computed the Heston equivalent volatility of volatility.† If we carry out the same computations with the calibrated parameters we obtain a result in agreement with Bakshi *et al.* (1997): the calibrated values are typically larger than the estimated values. This is consistent with the market practice that suggests increasing the option volatility in order to increase the option price. It is a way of taking into account transaction costs and/or preventing model risk (see, e.g., Leland (1985) and El Karoui *et al.* (1998)).

One striking consequence of our results is that we can consider vanilla options as basket products. Recall that the volatility of the first asset is Σ_t^{11}, which alone does not follow a Markovian dynamics. Therefore, the volatility risk on an option on the first asset can be partially hedged using options on the other asset. A model based on the standard affine Duffie–Kan framework would lead to similar conclusions. For example, if we add to equations (7), (8) and (9) the following dynamics:

$$\mathrm{d}F_t^1 = F_t^1\left(\sqrt{v_t^0}\mathrm{d}Z_t^0 + \sqrt{v_t^2}\mathrm{d}Z_t^2\right), \tag{48}$$

$$\mathrm{d}v_t^2 = \kappa_2(\theta_2 - v_t^2)\mathrm{d}t + \sigma_2\sqrt{v_t^2}\mathrm{d}W_t^2, \tag{49}$$

†As Σ_t^{11} is the volatility of the first asset, the equivalent Heston volatility of volatility is given by $2\sqrt{Q_{11}^2 + Q_{21}^2}$ because we have $\mathrm{d}\langle\Sigma^{11}, \Sigma^{11}\rangle_t = 4\Sigma_t^{11}(Q_{11}^2 + Q_{21}^2)\mathrm{d}t$.

with some constraints on the correlation among the different Brownian motions,† the resulting model shares some characteristics with the WASC model. The process v_t^0 is naturally considered as the main eigenvalue of the principal component analysis of the assets' covariance matrix. In fact, using the spectral decomposition proposed by Benabid *et al.* (2009), the dynamics of the eigenvalues (and eigenvectors) of the Wishart process is known explicitly and the main eigenvalue follows a dynamics similar to a square-root process. Nevertheless, considering v_t^0 as the main eigenvalue is problematic as nothing ensures that v_t^0 dominates v_t^1 and v_t^2, whereas this property is automatically granted for the Wishart process (this is the *non-colliding property of the eigenvalues*) (Bru 1991). An interesting point of the WASC model dwells in the fact that the assets' covariance matrix is a state variable, whereas in the BiHeston model we have a factor model for the covariance matrix. This eases the estimation procedure to a certain extent. We will illustrate this point in forthcoming work.

So far, we have introduced several stochastic volatility models that differ from each other with respect to calibration performance and flexibility, so a natural question arises: which one should we use? If our purpose is to price a basket option, then the WASC model is a natural candidate, but if we are interested in the pricing of a single stock exotic derivative, the calibration performance might not be sufficiently satisfactory to justify that choice. From a hedging point of view, Bakshi *et al.* (1997) developed an extensive empirical test and concluded that the Delta–Vega hedge‡ of an OTM vanilla option can be achieved equally well by the Black–Scholes model, the Heston model and the Bates model. Carrying out this test with the BiHeston or the WSMV models does not change this conclusion. Suppose that we need to hedge an exotic derivative that is sensitive to the shape of the skew. The implied volatility expansions can help us to decide which model should be used. Starting with the Heston model, the standard hedging strategy involves the sensitivity with respect to v_t^0, the initial value of the volatility process. From the implied volatility expansion we deduce that we hedge against a change of the implied volatility level. If our aim is to hedge the skew risk, then in light of (26) we should control the overall sensitivity with respect to the correlation parameter ρ despite the fact that it is inconsistent with the model hypothesis. The choice of this parameter is obviously justified by the impact of this coefficient on the smile. It is equivalent to the market practice of the Delta–Vega hedge in the Black and Scholes model. The implied volatility expansion also suggests that a hedge using the sensitivity with respect to the volatility of volatility could be used. If we consider the BiHeston model, then we have several possibilities. We can either control the sensitivity with respect to the correlation coefficient or to the volatility-of-volatility coefficient as in the Heston model, or we can also use the initial volatility processes. However, this latter strategy has a drawback. The parameter has an impact both on the level of the implied volatility (through the zero-order term) and the slope of the implied volatility level (through the first-order term). From a practical point of view it is always advisable to have a limited effect associated with a parameter: this eases the understanding of the model and the implementation of the hedging strategy. If we turn our attention to the WMSV model, we have a natural way of handling the problem. The implied volatility expansion suggests the use of Σ_t^{12} as the factor to control the skew risk. It does not impact the smile level (the zero-order term), but only the slope (the first-order term). The WMSV is a three-factor model and this should lead to an advantage with respect to the BiHeston model. It is remarkable how this additional factor affects the dynamics and the prices. We were unable to build a model based on the standard affine Duffie–Kan process, which shares this property. In conclusion, we have the possibility of choosing among the different stochastic volatility models the one with good properties. However, this choice is not based uniquely upon the calibration performance, but mostly on a hedging argument according to the product and the risks that have to be managed.

5. Jumps, parameter stability and numerical issues

In this section we briefly discuss some important aspects that are beyond the scope of this paper.

5.1. *Adding jumps*

To improve the calibration of the short-term smile and to be consistent with the empirical properties of time series, Bates (1996, 2000) recommends adding jumps on the stock dynamics. Duffie *et al.* (2000) develop a general extension. Within the standard affine framework the authors compute, in closed form, the characteristic function for a stock whose dynamics allows for jumps both on the asset and its volatility. We also tested whether the model of Bates (1996) leads to a better fit of the *very* short-term implied volatility surface with respect to the Heston, BiHeston as well as any other pure diffusive model.§

However, what is gained in terms of fit by jumps can be lost in a hedging perspective, as we will illustrate. If we perform the same exercise as in section 2 for the Bates (1996) jump-diffusion model, we see that the calibrated parameters are more stable (in particular, the correlation parameter ρ) for volatility norms with respect to different maturities. In fact, part of the skew dimension is handled through the jumps. This loss of sensitivity with respect to the term structure of the implied volatility surface may present a problem from a skew management perspective. For example, if we are interested in the skew risk, then

†To grant affinity of the model we require $\mathrm{d}Z_t^2\mathrm{d}W_t^2 = \rho_2\mathrm{d}t$, $\mathrm{d}Z_t^0\mathrm{d}W_t^2 = 0$, $\mathrm{d}Z_t^1\mathrm{d}W_t^2 = 0$ and $\mathrm{d}Z_t^1\mathrm{d}W_t^2 = 0$.
‡The Delta–Vega hedge is the hedging of an option using the stock, a discount bond and another option.
§Calibration results on different indexes are available upon request.

adding jumps only on the stock implies that the hedge of this risk must be carried out through the Delta. Adding jumps to the volatility will also be problematic. As suggested by the implied volatility expansion, jumps will affect the level of the smile and eventually the slope, hence there is a mixing effect that makes skew hedging problematic. As a consequence, the calibration performance is only part of the problem, in the sense that one also has to take into account how the model reacts to changes of the implied volatility surface.

The idea of managing the risks in terms of the sensitivity of the implied volatility surface is market practice. The usual procedure consists of humping the implied volatility surface and re-calibrating the parameters with the new surface. This can be done with different models such as CEV or displaced diffusions (Piterbarg 2003). For the latter, the typical way to manage the skew risk consists of re-calibrating the shifting parameter of the displaced diffusion after *ad hoc* perturbations of the implied volatility surface. This allows us to recover a term structure for the skew controlled by a time-dependent shifting parameter. Of course, this idea is equivalent to allowing the correlation parameter of the Heston model to be time-dependent as in Mikhailov and Nogel (2003) and Benhamou *et al*. (2010).

We conclude this subsection by mentioning a promising approach based on pure jump processes. In the literature on multi-asset models, only a few models are not straightforward extensions of the single-stock stochastic volatility models. Similar to the WASC model, there is the pure jump multi-asset model proposed by Barndorff-Nielsen and Stelzer (2011). It shares several features with the WASC model, one of which being that it involves a matrix diffusion process. Muhle-Karbe *et al*. (2009) propose a calibration of this model, but only under the very restrictive hypothesis of the matrix coefficients being diagonal. In light of our results it seems to us that this should be weakened.

5.2. Parameter stability

It is natural to question the stability of the calibrated parameters if the calibration is rolled over several days as in Bakshi *et al*. (1997). With the pure diffusion models presented in this paper we face the following problem: as the first maturity available converges to zero the corresponding implied volatility smile blows up. This is related to the fact that deep OTM options become worthless and inversion of the BS formula leads to large values. This pure calendar effect nevertheless strongly affects the shape of the smile. Using an implied volatility norm will magnify the problem and will compromise the optimization procedure. For the Heston model (but the same holds true for any other model), as the first time to maturity decreases to zero the MSE increases and drops once the first maturity slice is deleted from the calibration set.† Obviously, it will also affect the calibrated parameters. We are not sure about the rule we should use to limit the moneyness range for short-time-to-maturity options. The benefit of keeping our calibration strategy as simple as possible allows for a better comparison and understanding of our results. We leave to future work the study of parameter stability in a dynamic calibration perspective.

5.3. Numerical issues

It is necessary to clarify certain numerical aspects and difficulties that arise when implementing the FFT for the WMSV and WASC models. Essentially, we refer to Carr and Madan (1999) for a description of the FFT methodology. Since the option price is a real number we only need to integrate (4) over the positive axis. Discretization of the integral is done using 4096 points and a step of 0.18. To ensure integrability we need to specify a coefficient (denoted α by Carr and Madan (1999)) that we take equal to 1.2. The put options are priced using the call–put parity relation and all prices are normalized by the forward price and the discount factor. This implies that the characteristic function is computed on a grid of points whose values range from 0 to $4096 * 0.18$. As we compute the Fourier transform of an integrable function, the latter converges to zero, therefore we do not need to evaluate the function on all the points. We truncate once the integrand becomes smaller than $10\,e^{-20}$ and we take the same bound for all the maturities, but this is not optimal.‡ That is, we do not need to evaluate the characteristic function on 4096 points, which reduces the time needed for the pricing procedure. Moreover, this trick *partially* allows us to avoid some numerical problems that appear when computing the characteristic function for arguments with a large modulus. Essentially, this is related to the exponentiation in (33) and (39). Most of the standard algorithms do not preserve the block structure of the matrix and produce numerical instabilities. In that case, instead of using the matrix exponential argument to compute the characteristic function, we directly refer to the ODEs (29), (30) or (35), (36) and we apply a Runge–Kutta algorithm. Of course, in this case the pricing procedure slows down. We face this problem when dealing with options with a maturity greater than 2 years and it is not related to the failure of Gindikin's condition. We also suspect that the standard exponentiation can lead to numerical instabilities for long maturities, as typically encountered in interest rate problems. It is always possible to restrict ourselves to the simulation of the ODEs: on one hand, it slows down the pricing algorithm but, on the other, it avoids the advocated problems for the calibration.

The fractional Fourier transform, as proposed by Chourdakis (2004), or an exponentiation that takes into account the block structure of the matrix as in Paige and Van Loan (1981), may provide a solution to this problem. We will try to clarify this point in future work.

†These results are available upon request.
‡This bound can be increased.

To obtain an idea of the computational cost using the FFT methodology for the models considered so far, we need to specify the following fact. We use one FFT per option, and the gradient of the optimization procedure involves two computations of the objective function, hence the pricing of all options. We do not use any of the approximations given in this paper. Once a numerical problem arises during the computation of the characteristic function we use the ODEs given above directly. The implied volatility is computed using a dichotomy. We use a forward moneyness range of [0.8 1.2] for maturities of less than 2 years and [0.6 1.4] for maturities of between 2 and 5 years. With these constraints we end up with 370 options for the DAX and 290 for the EuroStoxx50. The calibration times for the DAX are: Heston 22.92 s, BiHeston 40.01 s and WMSV 8 min 35 s. The WASC is calibrated on DAX/EuroStoxx50 in 15 min 32 s.† Needless to say that using simple improvements such as mutualizing the computation of the objective function and the gradient, using only a subset of options, using price approximations (only accurate for near ATM options) and other refinements can significantly reduce the computational time.

6. Conclusions

In this study we have calibrated the Wishart Multidimensional Stochastic Volatility model introduced by Da Fonseca *et al.* (2008) and the Wishart Affine Stochastic Correlation model developed by Da Fonseca *et al.* (2007b). To provide a benchmark we implemented the Heston model and the BiHeston model of Christoffersen *et al.* (2009). We have also provided explicit short-time expansions of the implied volatility that turned out to be very useful for understanding the implied volatility surface of these models and we have shown how to map the different models by conveniently aggregating the parameters.

In the multiple-asset framework, the implied volatility expansion suggests how a basket option priced using a WASC model can be hedged using two Heston models, each calibrated on a set of vanilla options. More generally, the implied volatility expansions allow us to roughly map a multi-asset stochastic volatility model or a single-stock multi-factor stochastic volatility model into the Heston framework and to understand how vanilla option prices aggregate the parameters.

Having calibrated all these models, the natural question to ask is how should one choose the model? The main message of our paper is that, beyond the calibration performance, which is important, one should also consider how the model allows us to manage the implied volatility surface from a hedging perspective: a good stochastic volatility model should give clear outputs to shocks on the implied volatility surface. When the problem is simple, perhaps the Heston model (or even the Black–Scholes model) may be sufficient. This is no longer the case if, for example, the trader needs to hedge an exotic derivative that is sensitive to the shape of the skew. The implied volatility approximations help us to understand the ability of each model to hedge against movements of the volatility surface. In this respect, the WMSV model seems to be the best candidate for handling complex problems involving the shape of the skew. In fact, the implied volatility expansion indicates the factor that controls the pure skew risk, in the sense that it does not impact the smile level (the zero-order term) but only the slope (the first-order term) of the implied volatility.

If we choose a model according to a particular product and risk to hedge, it follows that we will manage different models at the same time when we allow for the entire book. As a consequence, we face a sort of model risk. Note, however, that here the point of view is quite different with respect to the classic approach (see, e.g., Kato and Yoshiba (2000)) in which model risk is referred as to the risk arising from the use of a model that cannot accurately evaluate market prices, or which is not a mainstream model in the market. It is also slightly different from the model risk presented by Cont (2006), where different models give the same market values for vanilla options, but lead to different prices for an exotic option. In fact, in our framework there is no 'right' model for the entire book, while any model we handle can be 'the' right model for a specific problem. Therefore, the question is how to quantify the risk induced by the presence of different models, how to unify and aggregate models and, more precisely, how to aggregate the hedging ratios in the presence of different models. Our approximation results are very helpful in this respect and provide a first simple answer to this interesting problem that will surely deserve much attention in future research.

Acknowledgements

The title of this paper obviously refers to the seminal work of Derman and Kani (1994). The study of different stochastic volatility models on different equity option markets led us to manage consistently several smiles through the use of the Wishart Affine Stochastic Correlation model. We were riding on the smiles. We are indebted to Alex Radicchi and Tommaso Gabbriellini from MPS Finance, Olivier Croissant from Natixis and Riadh Zaatour for fruitful discussions. As usual, the Zeliade Systems Team is greatly acknowledged for both computational support and stimulating discussions. We are also thankful to the participants of the Workshop on Spectral and Cubature Methods in Finance and Econometrics (Leicester, June 2009), the Workshop on Stochastic Analysis and Finance (City University of Hong Kong, June–July 2009), the 15th International Conference of Computing in Economics and Finance (University of Technology, Sydney, July 2009), the

†The computer has a 3 GHz processor and we use the Zeliade Quant Framework (ZQF) based on the programming language *C#*. The optimization algorithm is the Levenberg–Marquardt algorithm.

Finance and Stochastics Seminar (Department of Mathematics, Imperial College London, December 2009), the 3rd International Research Forum on Risk Dependencies (Paris, March 2010), and the 6th World Congress of the Bachelier Finance Society (Toronto, June 2010). All remaining errors are ours. This paper constitutes part of the Bruti-Liberati Lecture given at the QMF Conference (Sydney, 2010) by the second author, who gratefully acknowledges the Bruti-Liberati Foundation.

References

Bakshi, G.S., Cao, C. and Chen, Z., Empirical performance of alternative option pricing models. *J. Finance*, 1997, **52**, 2003–2049.

Bakshi, G.S. and Madan, D., Spanning and derivative-security valuation. *J. Financ. Econ.*, 2000, **55**, 205–238.

Barndorff-Nielsen, O.E. and Stelzer, R., The multivariate supOU stochastic volatility model. *Math. Finance*, 2011, DOI: 10.1111/j.1467-9965.2011.00494.x.

Bates, D.S., Jumps and stochastic volatility: exchange rate processes implicit in Deutsche Mark options. *Rev. Financ. Stud.*, 1996, **9**, 69–107.

Bates, D.S., Post-'87 crash fears in the S&P 500 futures option market. *J. Econometr.*, 2000, **14**(1/2), 181–238.

Benabid, A., Bensusan, H. and El Karoui, N., Wishart stochastic volatility: asymptotic smile and numerical framework. Working Paper, 2009. Available online at: http://hal.archives-ouvertes.fr/hal-00458014/fr

Benhamou, E., Gobet, E. and Miri, M., Time dependent Heston model. *SIAM J. Financ. Math.*, 2010, **1**, 289–325.

Bensusan, H., Risques de taux et de longévité: Modélisation dynamique et applications aux produits dérivés et à l'assurance-vie. PhD thesis, Ecole Polytechnique, 2010.

Branger, N. and Muck, M., Keep on smiling? Volatility surfaces and the pricing of quanto options when all covariances are stochastic. Working Paper, 2009. Available online at: http://www.wiwi.uni-muenster.de/fcm/fcm/organisation/details.php?weobjectID=162

Bru, M.F., Wishart processes. *J. Theor. Probab.*, 1991, **4**, 725–743.

Carr, P., Geman, H., Madan, D.P. and Yor, M., Stochastic volatility for Lévy processes. *Math. Finance*, 2003, **13**, 345–382.

Carr, P. and Madan, D., Option valuation using the FFT. *J. Comput. Finance*, 1999, **2**, 61–73.

Carr, P. and Wu, L., Time-changed Levy processes and option pricing. *J. Financ. Econ.*, 2004, **71**, 113–141.

Carr, P. and Wu, L., Stochastic skew in currency options. *J. Financ. Econ.*, 2007, **86**, 213–247.

Carrasco, M., Chernov, M., Florens, J.-P. and Ghysels, E., Efficient estimation of jump diffusions and general dynamic models with a continuum of moment conditions. *J. Econometr.*, 2007, **140**, 529–573.

Chourdakis, K., Option pricing using the fractional FFT. *J. Comput. Finance*, 2004, **8**(2), 1–18.

Christoffersen, P., Heston, S.L. and Jacobs, K., The shape and term structure of the index option smirk: why multifactor stochastic volatility models work so well. *Mgmt Sci.*, 2009, **55**, 1914–1932.

Christoffersen, P. and Jacobs, K., The importance of the loss function in option valuation. *J. Financ. Econ.*, 2004, **72**, 291–318.

Cont, R., Model uncertainty and its impact on the pricing of derivative instruments. *Math. Finance*, 2006, **16**, 519–547.

Cont, R. and Da Fonseca, J., Dynamics of implied volatility surfaces. *Quant. Finance*, 2002, **2**(1), 45–60.

Cont, R. and Tankov, P., *Financial Modelling with Jump Processes*, 2008 (Chapman & Hall/CRC Press: London).

Cuchiero, C., Filipovic, D., Mayerhofer, E. and Teichmann, J., Affine processes on positive semidefinite matrices. *Ann. Appl. Probab.*, 2010, **21**(2), 397–463.

Da Fonseca, J., Grasselli, M. and Ielpo, F., Estimating the Wishart affine stochastic correlation model using the empirical characteristic function. SSRN-1054721, 2007a.

Da Fonseca, J., Grasselli, M. and Tebaldi, C., Option pricing when correlations are stochastic: an analytical framework. *Rev. Deriv. Res.*, 2007b, **10**(2), 151–180.

Da Fonseca, J., Grasselli, M. and Tebaldi, C., A multifactor volatility Heston model. *Quant. Finance*, 2008, **8**(6), 591–604.

Da Fonseca, J., Grasselli, M. and Ielpo, F., Heging (co)variance risk with variance swaps. *Int. J. Theor. Appl. Financ.*, 2011, forthcoming.

Derman, E. and Kani, I., Riding on a smile. *Risk*, 1994, **7**(2), 139–145.

Duffie, D., Pan, J. and Singleton, K., Transform analysis and asset pricing for affine jump-diffusion. *Econometrica*, 2000, **68**, 1343–1376.

Egloff, D., Leippold, M. and Wu, L., Variance risk dynamics, variance risk premia, and optimal variance swap investments. *J. Financ. Quant. Anal.*, 2010, **45**(5), 1279–1310.

El Karoui, N., Jeanblanc-Picque, M. and Shreve, S.E., Robustness of the Black and Scholes formula. *Math. Finance*, 1998, **8**, 93–126.

Fouque, J.P., Papanicolaou, G., Sircar, R. and Solna, K., Singular perturbations in option pricing. *SIAM J. Appl. Math.*, 2003, **63**, 1648–1665.

Gauthier, P. and Possamai, D., Efficient simulation of the Wishart model. SSRN-1474728, 2009a.

Gauthier, P. and Possamai, D., Prices expansion in the Wishart model, SSRN-1475153, 2009b.

Gouriéroux, C. and Sufana, R., Derivative pricing with multivariate stochastic volatility. *J. Business Econ. Statist.*, 2010, **28**(3), 438–451.

Grasselli, M. and Tebaldi, C., Solvable affine term structure models. *Math. Finance*, 2008, **18**, 135–153.

Gruber, P.H., Tebaldi, C. and Trojani, F., Three make a dynamic smile – Unspanned skewness and interacting volatility components in option valuation. SSRN eLibrary, 2010.

Heston, S.L., A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Rev. Financ. Stud.*, 1993, **6**(2), 327–343.

Huang, J. and Wu, L., Specification analysis of option pricing models based on time-changed Levy processes. *J. Finance*, 2004, **59**, 1405–1440.

Kato, T. and Yoshiba, T., Model risk and its control. *Monet. Econ. Stud.*, 2000, 129–157.

Leland, H.E., Option pricing and replication with transactions costs. *J. Finance*, 1985, **40**(5), 1283–1301.

Letac, G. and Massam, H., A tutorial on non central Wishart distributions. Working Paper, 2004.

Lewis, A.S., *Option Valuation Under Stochastic Volatility*, 2000 (Finance Press: Newport Beach).

Madan, D. and Milne, F., Option pricing with VG martingale components. *Math. Finance*, 1991, **1**, 39–56.

Mayerhofer, E., Pfaffel, O. and Stelzer, R., On strong solutions of positive definite jump-diffusions. 2010, arXiv:0910.1784v3.

Medvedev, A. and Scaillet, O., Approximation and calibration of short-term implied volatilities under jump-diffusion stochastic volatility. *Rev. Financ. Stud.*, 2007, **20**, 427–459.

Mikhailov, S. and Nogel, U., Heston's stochastic volatility model implementation, calibration and some extensions. *Wilmott Mag.*, 2003, 74–79.

Muhle-Karbe, J., Pfaffel, O. and Stelzer, R., Option pricing in multivariate stochastic volatility models of OU type, 2009. Preprint. Available online at: http://arxiv.org/abs/1001.3223

Paige, C. and Van Loan, C.F., A Schur decomposition for Hamiltonian matrices. *Linear Algebra Applic.*, 1981, **41**, 11–32.
Piterbarg, V., A stochastic volatility forward libor model with a term structure of volatility smiles, 2003. Available online at SSRN: http://ssrn.com/abstract=472061
Rebonato, R., *Volatility and Correlation*, 1999 (Wiley: Chichester).
Schonbucher, P., A market model for stochastic implied volatility. *Philos. Trans. R. Soc.*, 1999, **357**(1758), 2071–2092.
Schweizer, M. and Wissel, J., Term structures of implied volatilities: absence of arbitrage and existence results. *Math. Finance*, 2008, **18**(1), 77–114.
Sepp, A., Fourier transform for option pricing under affine jump-diffusions: an overview, Unpublished manuscript, 2003. Available online at: www.hot.ee/seppar
Skiadopoulos, G.S., Hodges, S.D. and Clewlow, L., The dynamics of the S&P 500 implied volatility surface. *Rev. Deriv. Res.*, 1999, **3**(3), 263–282.

Appendix A

A.1. Proof of proposition 3.3

Proof: By the Feynman–Kac argument, the characteristic function of $(x_T, \Sigma_T, \int_t^T \Sigma_u \mathrm{d}u)$ solves the following partial differential equation:

$$\frac{\partial}{\partial t}\phi + \mathcal{L}_{x,\Sigma}\phi + i\,\mathrm{Tr}[\Lambda_I \Sigma]\phi = 0,$$

where the infinitesimal generator of the couple (x, Σ) (Da Fonseca *et al.* 2008) is given by

$$\begin{aligned}\mathcal{L}_{x,\Sigma} = {} & \mathrm{Tr}[\Omega\Omega^\top D + M\Sigma D + \Sigma M^\top D + 2\Sigma D Q^\top Q D] \\ & - \frac{1}{2}\mathrm{Tr}[\Sigma]\partial_x + \frac{1}{2}\mathrm{Tr}[\Sigma]\partial_{x^2}^2 + 2\,\mathrm{Tr}[\Sigma R Q D]\partial_x,\end{aligned}$$

where D denotes the matrix differential operator with elements

$$D_{i,j} = \left(\frac{\partial}{\partial \Sigma_{i,j}}\right). \tag{A1}$$

Following Da Fonseca *et al.* (2008), we guess a solution of the form

$$\phi(\lambda_x, \Lambda_V, \Lambda_I, t, T) = \mathrm{e}^{i\lambda_x x_t + \mathrm{Tr}[A(\tau)\Sigma_t] + b(\tau)}. \tag{A2}$$

Plugging this guess into the previous PDE we obtain, after collecting the terms, the following matrix differential equation for A $(\tau := T - t)$:

$$\begin{aligned}\frac{\partial}{\partial \tau}A = {} & A(M + i\lambda_x Q^\top R^\top) + (M + i\lambda_x Q^\top R^\top)^\top A \\ & + 2AQ^\top Q A + \frac{i\lambda_x(i\lambda_x - 1)}{2}\mathbb{I} + i\Lambda_I,\end{aligned} \tag{A3}$$

$$\frac{\partial}{\partial \tau}b = \mathrm{Tr}[\Omega\Omega^\top A(\tau)], \tag{A4}$$

with boundary conditions $A(0) = i\Lambda_V$ and $b(0) = 0$.

Using standard techniques as presented by Grasselli and Tebaldi (2008) and Da Fonseca *et al.* (2008) we can transform the matrix Riccati ODE into a linear ODE whose solution is given by a matrix exponential. In fact, the flow of the Riccati ODE can be linearized by doubling the dimension of the problem due to the fact that the Riccati ODE belong to a quotient manifold (see Grasselli and Tebaldi (2008) for further details). For the sake of completeness we provide the complete procedure as in Da Fonseca *et al.* (2008).

Put

$$A(\tau) = F(\tau)^{-1}G(\tau), \tag{A5}$$

for $F(\tau) \in GL(n)$, $G(\tau) \in M_n$, then

$$\frac{\mathrm{d}}{\mathrm{d}\tau}[F(\tau)A(\tau)] - \frac{\mathrm{d}}{\mathrm{d}\tau}[F(\tau)]A(\tau) = F(\tau)\frac{\mathrm{d}}{\mathrm{d}\tau}A(\tau)$$

and

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\tau}G(\tau) - \frac{\mathrm{d}}{\mathrm{d}\tau}[F(\tau)]A(\tau) = {} & \left(\frac{i\lambda_x(i\lambda_x - 1)}{2} + i\Lambda_I\right)F(\tau) \\ & + G(\tau)(M + i\lambda_x Q^\top R^\top) \\ & + (F(\tau)(M + i\lambda_x Q^\top R^\top)^\top \\ & + 2G(\tau)Q^\top Q)A(\tau).\end{aligned}$$

The last ODE leads to the system of $(2n)$ *linear* equations

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\tau}G(\tau) &= \left(\frac{i\lambda_x(i\lambda_x - 1)}{2} + i\Lambda_I\right)F(\tau) + G(\tau)(M + i\lambda_x Q^\top R^\top), \\ \frac{\mathrm{d}}{\mathrm{d}\tau}F(\tau) &= -F(\tau)(M + i\lambda_x Q^\top R^\top)^\top - 2G(\tau)Q^\top Q,\end{aligned} \tag{A6}$$

which can be written as follows:

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}\tau}(G(\tau)F(\tau)) \\ &= (G(\tau)\ \ F(\tau))\begin{pmatrix} M + i\lambda_x Q^\top R^\top & -2Q^\top Q \\ \dfrac{i\lambda_x(i\lambda_x - 1)}{2} + i\Lambda_I & -(M + i\lambda_x Q^\top R^\top)^\top \end{pmatrix}.\end{aligned}$$

The solution is obtained through exponentiation:

$$\begin{aligned}&(G(\tau)\ F(\tau)) \\ &= (G(0)\ F(0))\exp\tau\begin{pmatrix} M + i\lambda_x Q^\top R^\top & -2Q^\top Q \\ \dfrac{i\lambda_x(i\lambda_x - 1)}{2} + i\Lambda_I & -(M + i\lambda_x Q^\top R^\top)^\top \end{pmatrix} \\ &= (A(0)\ \mathbb{I})\exp\tau\begin{pmatrix} M + i\lambda_x Q^\top R^\top & -2Q^\top Q \\ \dfrac{i\lambda_x(i\lambda_x - 1)}{2} + i\Lambda_I & -(M + i\lambda_x Q^\top R^\top)^\top \end{pmatrix} \\ &= (A(0)A_{11}(\tau) + A_{21}(\tau)\ A(0)A_{12}(\tau) + A_{22}(\tau)),\end{aligned}$$

where

$$\begin{aligned}&\begin{pmatrix} A_{11}(\tau) & A_{12}(\tau) \\ A_{21}(\tau) & A_{22}(\tau) \end{pmatrix} \\ &= \exp\tau\begin{pmatrix} M + i\lambda_x Q^\top R^\top & -2Q^\top Q \\ \dfrac{i\lambda_x(i\lambda_x - 1)}{2} + i\Lambda_I & -(M + i\lambda_x Q^\top R^\top)^\top \end{pmatrix}.\end{aligned} \tag{A7}$$

In conclusion, we obtain

$$A(\tau) = (i\Lambda_V A_{12}(\tau) + A_{22}(\tau))^{-1}(i\Lambda_V A_{11}(\tau) + A_{21}(\tau)),$$

which represents the closed-form solution of the matrix Riccati (A3).

Let us now turn our attention to equation (A4). From (A6) we obtain

$$G(\tau) = -\frac{1}{2}\left(\frac{\mathrm{d}}{\mathrm{d}\tau}F(\tau) + F(\tau)(M + i\lambda_x Q^\top R^\top)^\top\right)(Q^\top Q)^{-1},$$

and plugging this into (A5) and using the properties of the trace we obtain

$$\frac{\mathrm{d}}{\mathrm{d}\tau}c(\tau) = -\frac{\beta}{2}\,\mathrm{Tr}\left[F(\tau)^{-1}\frac{\mathrm{d}}{\mathrm{d}\tau}F(\tau) + (M + i\lambda_x Q^\top R^\top)^\top\right].$$

We can now integrate the last equation and obtain

$$c(\tau) = -\frac{\beta}{2}\,\mathrm{Tr}[\log F(\tau) + (M + i\lambda_x Q^\top R^\top)^\top \tau],$$

which leads to the result. □

A.2. Proof of proposition 3.5

Proof: The starting point for the approximations is the Riccati equation (29) associated with the characteristic function of the asset returns, i.e. formula (28) with $\Lambda_V = \Lambda_I = 0$ (also set $\lambda_x = \lambda$ for notational convenience). When we replace Q by αQ we obtain

$$\begin{aligned}-\frac{\partial}{\partial t}A(t,T) &= A(t,T)(M + i\alpha\lambda Q^\top R^\top) \\ &\quad + (M + i\alpha\lambda Q^\top R^\top)^\top A(t,T) \\ &\quad + 2\alpha^2 A(t,T)Q^\top Q A(t,T) + \frac{i\lambda(i\lambda - 1)}{2}\mathbb{I}. \quad \text{(A8)}\end{aligned}$$

Consider the expansion of the Riccati in terms of α, $A = A^0 + \alpha A^1 + \alpha^2 A^2$, and plug this solution into (A8). When collecting all terms we are lead to the system ($\tau = T - t$)

$$\partial_\tau A^0(\tau) = A^0(\tau)M + M^\top A^0(\tau) - \frac{(\lambda^2 + i\lambda)}{2}\mathbb{I}, \quad \text{(A9)}$$

$$\begin{aligned}\partial_\tau A^1(\tau) &= A^1(\tau)M + M^\top A^1(\tau) + A^0(\tau)(i\lambda Q^\top R^\top) \\ &\quad + (i\lambda Q^\top R^\top)^\top A^0(\tau), \quad \text{(A10)}\end{aligned}$$

$$\begin{aligned}\partial_\tau A^2(\tau) &= A^2(\tau)M + M^\top A^2(\tau) + A^1(\tau)(i\lambda Q^\top R^\top) \\ &\quad + (i\lambda Q^\top R^\top)^\top A^1(\tau) + 2A^0(\tau)Q^\top Q A^0(\tau). \quad \text{(A11)}\end{aligned}$$

We denote $\gamma_v = -(\lambda^2 + i\lambda)/2$ and $\gamma_x = i\lambda$, then it is straightforward to obtain

$$\begin{aligned}A^0(\tau) &= \gamma_v \int_0^\tau \mathrm{e}^{(\tau-u)M^\top}\mathbb{I}\mathrm{e}^{(\tau-u)M}\mathrm{d}u \\ &= \gamma_v \tilde{A}^0(\tau), \\ A^1(\tau) &= \gamma_v\gamma_x \int_0^\tau \mathrm{e}^{(\tau-u)M^\top}\{\tilde{A}^0(u)(Q^\top R^\top) \\ &\quad + (Q^\top R^\top)^\top \tilde{A}^0(u)\}\mathrm{e}^{(\tau-u)M}\mathrm{d}u \\ &= \gamma_v\gamma_x \tilde{A}^1(\tau),\end{aligned}$$

$$\begin{aligned}A^2(\tau) &= \gamma_v^2 \int_0^\tau \mathrm{e}^{(\tau-u)M^\top}\{2\tilde{A}^0(u)Q^\top Q\tilde{A}^0(u)\}\mathrm{e}^{(\tau-u)M}\mathrm{d}u \\ &\quad + \gamma_v\gamma_x^2 \int_0^\tau \mathrm{e}^{(\tau-u)M^\top}\{\tilde{A}^1(u)(Q^\top R^\top) \\ &\quad + (Q^\top R^\top)^\top \tilde{A}^1(u)\}\mathrm{e}^{(\tau-u)M}\mathrm{d}u \\ &= \gamma_v^2 \tilde{A}^{20}(\tau) + \gamma_v\gamma_x^2 \tilde{A}^{21}(\tau).\end{aligned}$$

The expansion of $b(t,T) = b(\tau)$ is obtained by direct integration:

$$\begin{aligned}b(\tau) &= \gamma_v\,\mathrm{Tr}\left[\Omega\Omega^\top \int_0^\tau \tilde{A}^0(u)\mathrm{d}u\right] \\ &\quad + \alpha\gamma_v\gamma_x\,\mathrm{Tr}\left[\Omega\Omega^\top \int_0^\tau \tilde{A}^1(u)\mathrm{d}u\right] \\ &\quad + \alpha^2\gamma_v^2\,\mathrm{Tr}\left[\Omega\Omega^\top \int_0^\tau \tilde{A}^{20}(u)\mathrm{d}u\right] \\ &\quad + \alpha^2\gamma_v\gamma_x^2\,\mathrm{Tr}\left[\Omega\Omega^\top \int_0^\tau \tilde{A}^{21}(u)\mathrm{d}u\right],\end{aligned}$$

where we denote by $\tilde{b}^0(\tau)$, $\tilde{b}^1(\tau)$, $\tilde{b}^{20}(\tau)$ and $\tilde{b}^{21}(\tau)$ the different traces of the right-hand side.

The key point is that the matrices $\tilde{A}^0(\tau)$, $\tilde{A}^1(\tau)$, $\tilde{A}^{20}(\tau)$ and $\tilde{A}^{21}(\tau)$ and the functions $\tilde{b}^0(\tau)$, $\tilde{b}^1(\tau)$, $\tilde{b}^{20}(\tau)$ and $\tilde{b}^{21}(\tau)$ do not depend on λ.

If we denote by $\widehat{F}(\lambda)$ the Fourier transform of the option payoff, from (4) we have the following Taylor expansion for the option price:

$$\begin{aligned}Price_t(\tau,x) &= \mathrm{e}^{-r\tau}\frac{1}{2\pi}\int_{\mathcal{Z}} \mathrm{e}^{i\lambda x - \frac{\lambda^2+i\lambda}{2}(\mathrm{Tr}[A^0(\tau)\Sigma_t] + \tilde{b}^0(\tau))}\widehat{F}(\lambda)\mathrm{d}\lambda \\ &\quad + \alpha(\mathrm{Tr}[\tilde{A}^1(\tau)\Sigma_t] + \tilde{b}^1(\tau))\frac{\mathrm{e}^{-r\tau}}{2\pi} \\ &\quad \int_{\mathcal{Z}} \gamma_v\gamma_x \mathrm{e}^{i\lambda x - \frac{\lambda^2+i\lambda}{2}(\mathrm{Tr}[A^0(\tau)\Sigma_t] + \tilde{b}^0(\tau))}\widehat{F}(\lambda)\mathrm{d}\lambda \\ &\quad + \alpha^2(\mathrm{Tr}[\tilde{A}^{20}(\tau)\Sigma_t] + \tilde{b}^{20}(\tau))\frac{\mathrm{e}^{-r\tau}}{2\pi} \\ &\quad \int_{\mathcal{Z}} \gamma_v^2 \mathrm{e}^{i\lambda x - \frac{\lambda^2+i\lambda}{2}(\mathrm{Tr}[A^0(\tau)\Sigma_t] + \tilde{b}^0(\tau))}\widehat{F}(\lambda)\mathrm{d}\lambda \\ &\quad + \alpha^2(\mathrm{Tr}[\tilde{A}^{21}(\tau)\Sigma_t] + \tilde{b}^{21}(\tau))\frac{\mathrm{e}^{-r\tau}}{2\pi} \\ &\quad \int_{\mathcal{Z}} \gamma_v\gamma_x^2 \mathrm{e}^{i\lambda x - \frac{\lambda^2+i\lambda}{2}(\mathrm{Tr}[A^0(\tau)\Sigma_t] + \tilde{b}^0(\tau))}\widehat{F}(\lambda)\mathrm{d}\lambda \\ &\quad + \frac{\alpha^2}{2}(\mathrm{Tr}[\tilde{A}^1(\tau)\Sigma_t] + \tilde{b}^1(\tau))^2\frac{\mathrm{e}^{-r\tau}}{2\pi} \\ &\quad \int_{\mathcal{Z}} (\gamma_v\gamma_x)^2 \mathrm{e}^{i\lambda x - \frac{\lambda^2+i\lambda}{2}(\mathrm{Tr}[A^0(\tau)\Sigma_t] + \tilde{b}^0(\tau))}\widehat{F}(\lambda)\mathrm{d}\lambda \\ &\quad + o(\alpha^2).\end{aligned}$$

Let us now consider the Black–Scholes price of the option, which can also be written in terms of the characteristic function

$$c_{\mathrm{B\&S}}(t,T,x,K,\xi) = \frac{\mathrm{e}^{-r\tau}}{2\pi}\int_{\mathcal{Z}} \mathrm{e}^{i\lambda x - \frac{(\lambda^2+i\lambda)\xi}{2}}\widehat{F}(\lambda)\mathrm{d}\lambda. \quad \text{(A12)}$$

Combining (A12) with the price expansion we conclude that

$$\begin{aligned}Price(\tau,T,x,K) = c_{B\&S}(t,T,x,K,\xi_0) &+ \alpha(\mathrm{Tr}[\tilde{A}^1\Sigma_t] \\ &+ \tilde{b}^1)\partial^2_{x\xi}c_{B\&S}(t,T,x,K,\xi_0) \\ &+ \alpha^2(\mathrm{Tr}[\tilde{A}^{20}\Sigma_t]+\tilde{b}^{20})\partial^2_{\xi^2}c_{B\&S}(t,T,x,K,\xi_0) \\ &+ \alpha^2(\mathrm{Tr}[\tilde{A}^{21}\Sigma_t]+\tilde{b}^{21})\partial^3_{x^2\xi}c_{B\&S}(t,T,x,K,\xi_0) \\ &+ \frac{\alpha^2}{2}(\mathrm{Tr}[\tilde{A}^1\Sigma_t]+\tilde{b}^1)^2\partial^4_{x^2\xi^2}c_{B\&S}(t,T,x,K,\xi_0),\end{aligned}$$

with $\xi_0 = (\mathrm{Tr}[\tilde{A}^0\Sigma_t] + \tilde{b}^0)$, which gives the result. □

A.3. Proof of proposition 3.6

Proof: If we assume for the implied volatility an expansion of the form $\sigma^2_{\mathrm{imp}}\tau = \zeta_0 + \alpha\zeta_1 + \alpha^2\zeta_2$, we deduce that

$$\begin{aligned}c_{B\&S}(t,T,x,\sigma^2_{\mathrm{imp}}\tau) = c_{B\&S}(t,T,x,\zeta_0) &+ \alpha\zeta_1\partial_\xi c_{B\&S}(t,T,x,\zeta_0) \\ &+ \frac{\alpha^2}{2}(2\zeta_2\partial_\xi c_{B\&S}(t,T,x,\zeta_0) \\ &+ \zeta_1^2\partial^2_{\xi\xi}c_{B\&S}(t,T,x,\zeta_0)).\end{aligned}$$

Comparison with (40) gives

$$\begin{aligned}\zeta_0 &= \xi_0, \\ \zeta_1 &= \frac{(\mathrm{Tr}[\tilde{A}^1\Sigma_t]+\tilde{b}^1)\partial^2_{x\xi}c_{B\&S}}{\partial_\xi c_{B\&S}}, \\ \zeta_2 &= \frac{-\zeta_1^2\partial^2_{\xi\xi}c_{B\&S} + 2(\mathrm{Tr}[\tilde{A}^{20}\Sigma_t]+\tilde{b}^{20})\partial^2_{\xi^2}c_{B\&S} + 2(\mathrm{Tr}[\tilde{A}^{21}\Sigma_t]+\tilde{b}^{21})\partial^3_{x^2\xi}c_{B\&S}}{2\partial_\xi c_{B\&S}} + \frac{(\mathrm{Tr}[\tilde{A}^1\Sigma_t]+\tilde{b}^1)^2\partial^4_{x^2\xi^2}c_{B\&S}}{2\partial_\xi c_{B\&S}},\end{aligned}$$

where the Black–Scholes formula $c_{B\&S}$ is evaluated at the point (x, ξ_0).

In order to find ζ_1 and ζ_2 we differentiate the integral relations (41)–(44), thus obtaining

$$\begin{aligned}\partial_\tau\tilde{A}^0 &= \tilde{A}^0M + M^\top\tilde{A}^0 + \mathbb{I}, \\ \partial_\tau\tilde{A}^1 &= \tilde{A}^1M + M^\top\tilde{A}^1 + \tilde{A}^0(Q^\top R^\top) + (Q^\top R^\top)^\top\tilde{A}^0, \\ \partial_\tau\tilde{A}^{20} &= \tilde{A}^{20}M + M^\top\tilde{A}^{20} + 2\tilde{A}^0Q^\top Q\tilde{A}^0, \\ \partial_\tau\tilde{A}^{21} &= \tilde{A}^{21}M + M^\top\tilde{A}^{21} + \tilde{A}^1(Q^\top R^\top) + (Q^\top R^\top)^\top\tilde{A}^1,\end{aligned}$$

and we take the Taylor expansion in terms of τ:

$$\begin{aligned}\tilde{A}^0 &= \tau\mathbb{I} + \frac{\tau^2}{2}(M+M^\top), \\ \tilde{A}^1 &= \frac{\tau^2}{2}(Q^\top R^\top + RQ) + \frac{\tau^3}{6}((Q^\top R^\top + RQ)M \\ &\quad + M^\top(Q^\top R^\top + RQ)) + \frac{\tau^3}{6}((M+M^\top)Q^\top R^\top \\ &\quad + RQ(M+M^\top)), \\ \tilde{A}^{20} &= \frac{\tau^3}{6}4Q^\top Q, \\ \tilde{A}^{21} &= \frac{\tau^3}{6}((Q^\top R^\top + RQ)Q^\top R^\top + RQ(Q^\top R^\top + RQ)).\end{aligned}$$

In conclusion, when τ is small we have

$$\begin{aligned}\mathrm{Tr}[\tilde{A}^0\Sigma_t] + \tilde{b}^0 &= \mathrm{Tr}[\Sigma]\tau + o(\tau), \\ \mathrm{Tr}[\tilde{A}^1\Sigma_t] + \tilde{b}^1 &= \mathrm{Tr}[RQ\Sigma]\tau^2 + o(\tau^2), \\ \mathrm{Tr}[\tilde{A}^{20}\Sigma_t] + \tilde{b}^{20} &= \frac{2}{3}\mathrm{Tr}[Q^\top Q\Sigma]\tau^3 + o(\tau^3), \\ \mathrm{Tr}[\tilde{A}^{21}\Sigma_t] + \tilde{b}^{21} &= \frac{1}{3}\mathrm{Tr}[RQ(Q^\top R^\top + RQ)\Sigma]\tau^3 + o(\tau^3),\end{aligned}$$

and the derivatives of the Black–Scholes formula are easily computed:

$$\begin{aligned}\frac{\partial^2_{x,\xi}c_{B\&S}(\tau,x,K,\mathrm{Tr}[\Sigma_t]\tau)}{\partial_\xi c_{B\&S}(\tau,x,K,\mathrm{Tr}[\Sigma_t]\tau)} &= \frac{1}{2} + \frac{m_f}{\mathrm{Tr}[\Sigma_t]\tau}, \\ \frac{\partial^2_{\xi^2}c_{B\&S}(\tau,x,K,\mathrm{Tr}[\Sigma_t]\tau)}{\partial_\xi c_{B\&S}(\tau,x,K,\mathrm{Tr}[\Sigma_t]\tau)} &= \frac{m_f^2}{2(\mathrm{Tr}[\Sigma_t]\tau)^2} - \frac{1}{2(\mathrm{Tr}[\Sigma_t]\tau)} - \frac{1}{8}, \\ \frac{\partial^3_{x^2,\xi}c_{B\&S}(\tau,x,K,\mathrm{Tr}[\Sigma_t]\tau)}{\partial_\xi c_{B\&S}(\tau,x,K,\mathrm{Tr}[\Sigma_t]\tau)} &= \frac{1}{4} + \frac{m_f-1}{\mathrm{Tr}[\Sigma_t]\tau} + \frac{m_f^2}{(\mathrm{Tr}[\Sigma_t]\tau)^2}, \\ \frac{\partial^4_{x^2,\xi^2}c_{B\&S}(\tau,x,K,\mathrm{Tr}[\Sigma_t]\tau)}{\partial_\xi c_{B\&S}(\tau,x,K,\mathrm{Tr}[\Sigma_t]\tau)} &= \frac{m_f^4}{2(\mathrm{Tr}[\Sigma_t]\tau)^4} + \frac{m_f^2(m_f-6)}{2(\mathrm{Tr}[\Sigma_t]\tau)^3},\end{aligned}$$

where m_f denotes the forward moneyness. Finally,

$$\begin{aligned}\zeta_0 &= \mathrm{Tr}[\Sigma_t]\tau, \\ \zeta_1 &= \frac{\mathrm{Tr}[RQ\Sigma_t]}{\mathrm{Tr}[\Sigma_t]}m_f\tau, \\ \zeta_2 &= \frac{\mathrm{Tr}[Q^\top Q\Sigma]}{3(\mathrm{Tr}[\Sigma_t])^2}m_f^2\tau + \frac{\mathrm{Tr}[RQ(Q^\top R^\top + RQ)\Sigma]}{3(\mathrm{Tr}[\Sigma_t])^2}m_f^2\tau \\ &\quad - \frac{5}{2}\frac{(\mathrm{Tr}[RQ\Sigma])^2}{(\mathrm{Tr}[\Sigma_t])^3}m_f^2\tau \\ &= \frac{m_f^2\tau}{(\mathrm{Tr}[\Sigma_t])^2}\left(\frac{1}{3}\mathrm{Tr}[Q^\top Q\Sigma] + \frac{1}{3}\mathrm{Tr}[RQ(Q^\top R^\top \right. \\ &\quad \left. + RQ)\Sigma] - \frac{5}{4}\frac{(\mathrm{Tr}[RQ\Sigma])^2}{\mathrm{Tr}[\Sigma_t]}\right),\end{aligned}$$

which gives the result. □

A.4. Proof of proposition 3.7

Proof: Let us consider the development $A = A^0 + \alpha A^1 + \alpha^2 A^2$ and plug this solution into (35).

When collecting all terms we obtain the system

$$\partial_\tau A^0 = A^0 M + M^\top A^0 - \frac{(\lambda_1^2 + i\lambda_1)}{2} e_{11}, \tag{A13}$$

$$\partial_\tau A^1 = A^1 M + M^\top A^1 + i\lambda_1 A^0 (Q^\top \rho e_1^\top) + i\lambda_1 (Q^\top \rho e_1^\top)^\top A^0, \tag{A14}$$

$$\partial_\tau A^2 = A^2 M + M^\top A^2 + i\lambda_1 A^1 (Q^\top \rho e_1^\top) + i\lambda_1 (Q^\top \rho e_1^\top)^\top A^1 + 2A^0 Q^\top Q A^0. \tag{A15}$$

Let $\gamma_v = -(\lambda_1^2 + i\lambda_1)/2$ and $\gamma_x = i\lambda_1$. By integration we obtain

$$A^0(\tau) = \gamma_v \int_0^\tau \mathrm{e}^{(\tau-u)M^\top} e_{11} \mathrm{e}^{(\tau-u)M} \mathrm{d}u \tag{A16}$$

$$= \gamma_v \tilde{A}^0(\tau), \tag{A17}$$

$$A^1(\tau) = \gamma_v \gamma_x \int_0^\tau \mathrm{e}^{(\tau-u)M^\top} \{\tilde{A}^0(u)(Q^\top \rho e_1^\top) + (Q^\top \rho e_1^\top)^\top \tilde{A}^0(u)\} \mathrm{e}^{(\tau-u)M} \mathrm{d}u \tag{A18}$$

$$= \gamma_v \gamma_x \tilde{A}^1(\tau), \tag{A19}$$

$$A^2(\tau) = \gamma_v^2 \int_0^\tau \mathrm{e}^{(\tau-u)M^\top} \{2\tilde{A}^0(u) Q^\top Q \tilde{A}^0(u)\} \mathrm{e}^{(\tau-u)M} \mathrm{d}u \tag{A20}$$

$$+ \gamma_v \gamma_x^2 \int_0^\tau \mathrm{e}^{(\tau-u)M^\top} \{\tilde{A}^1(u)(Q^\top \rho e_1^\top) + (Q^\top \rho e_1^\top)^\top \tilde{A}^1(u)\} \mathrm{e}^{(\tau-u)M} \mathrm{d}u \tag{A21}$$

$$= \gamma_v^2 \tilde{A}^{20}(\tau) + \gamma_v \gamma_x^2 \tilde{A}^{21}(\tau), \tag{A22}$$

with

$$\partial_\tau \tilde{A}^0 = \tilde{A}^0 M + M^\top \tilde{A}^0 + e_{11}, \tag{A23}$$

$$\partial_\tau \tilde{A}^1 = \tilde{A}^1 M + M^\top \tilde{A}^1 + \tilde{A}^0 (Q^\top \rho e_1^\top) + (Q^\top \rho e_1^\top)^\top \tilde{A}^0, \tag{A24}$$

$$\partial_\tau \tilde{A}^{20} = \tilde{A}^{20} M + M^\top \tilde{A}^{20} + 2\tilde{A}^0 Q^\top Q \tilde{A}^0, \tag{A25}$$

$$\partial_\tau \tilde{A}^{21} = \tilde{A}^{21} M + M^\top \tilde{A}^{21} + \tilde{A}^1 (Q^\top \rho e_1^\top) + (Q^\top \rho e_1^\top)^\top \tilde{A}^1. \tag{A26}$$

To compute the Taylor expansion of the functions $\tilde{A}^0$, $\tilde{A}^1$ and $\tilde{A}^2$ we derive the ODE and obtain the following expansion for the implied volatility:

$$\zeta_1 = \tau m_f (\rho_1 Q_{11} + \rho_2 Q_{21}), \tag{A27}$$

$$2\zeta_2 = \tau m_f^2 \left[\frac{4(Q_{11}^2 + Q_{21}^2) - 7(\rho_1 Q_{11} + \rho_2 Q_{21})^2}{6\Sigma_t^{11}} \right]. \tag{A28}$$

The smile is then of the form

$$\sigma^2 \tau = \Sigma_t^{11} \tau + \alpha \zeta_1 + \alpha^2 \zeta_2, \tag{A29}$$

which leads to the result once the relevant quantities are replaced. □

A.5. Proof of proposition 3.8

Proof: First we compute the function (34) for $\Lambda_x = 0$, $\Lambda_V = \lambda e_{11}$ and $\Lambda_I = 0$ with $\lambda \in R$ and e_{11} an element of the canonical basis. Using the fact that $A_{21}(\tau) = 0$, $A_{11}(\tau) = \mathrm{e}^{\tau M}$, $A_{22}(\tau) = \mathrm{e}^{-\tau M^\top}$, the symmetry of $A_{12}(\tau)$ and the property $\mathrm{e}^{\mathrm{Tr}(A)} = \det(\exp(A))$, after algebraic manipulation the characteristic function of the Σ_t^{11} process can be rewritten as

$$\phi(0, \Lambda_V, 0, t, T) = \det(I - i2N_1\Lambda_V)^{-\beta/2} \times \exp \mathrm{Tr}\left(\frac{N_2}{2} i2N_1 \Lambda_V (I - i2N_1\Lambda_V)^{-1} \right),$$

with

$$N_1 = -\frac{1}{2}(A_{22}^\top)^{-1} A_{12}^\top,$$
$$N_2 = -2(A_{22}^\top)^{-1} \Sigma_t A_{11}^\top (A_{12}^\top)^{-1} A_{22}^\top.$$

The matrix $2N_1 e_{11}$ is rank one, therefore there exists a basis such that $2N_1 e_{11}$ can be rewritten as

$$\begin{pmatrix} a_{11} & 0 \\ 0 & 0 \end{pmatrix},$$

which implies that

$$\phi(0, \Lambda_V, 0, t, T) = (1 - i\lambda a_{11})^{-\beta/2} \mathrm{e}^{(N_2^{11}/2)(i\lambda a_{11}/1 - i\lambda a_{11})}.$$

We recognize the characteristic function of a non-central chi-square distribution. □