

To appear in *Quantitative Finance*, Vol. 00, No. 00, Month 20XX, 1–31

Calibration and advanced simulation schemes for the Wishart Stochastic Volatility model

G. LA BUA[†] and D. MARAZZINA^{*†}

[†] Department of Mathematics, Politecnico di Milano, 32 p.zza Leonardo da Vinci, I-20133
Milan, Italy.

(Received 00 Month 20XX; in final form 00 Month 20XX)

In this article we deal with calibration and Monte Carlo simulation of the Wishart Stochastic Volatility model. Despite the analytical tractability of the considered model, being of affine type, the implementation of Wishart-based stochastic volatility models poses non-trivial challenges from a numerical point of view. The goal of this article is to overcome these problems providing efficient numerical schemes for Monte Carlo simulations. Moreover, a fast and accurate calibration procedure is proposed.

Keywords: Wishart process; Calibration; Monte Carlo; Efficient Numerical Simulation Scheme

JEL Classification: C02, C63, G12, G13

1. Introduction

A Wishart process is a matrix valued continuous time stochastic process with a marginal Wishart distribution, i.e., a generalization to multiple dimensions of the chi-squared distribution, or, in the case of non-integer degrees of freedom, of the gamma distribution. The introduction of Wishart-based Stochastic Volatility models in finance is fully motivated by the need to describe the multidimensional structure of asset variances. There are, indeed, empirical evidences (see for example Christoffersen *et al.* (2009) and Cont and Da Fonseca (2002)) that the dynamics of the implied volatility surface is driven by several factors. This causes the standard one-factor stochastic volatility models not to be flexible enough to consistently price plain vanilla options and forward volatility sensitive derivatives (e.g. forward starting and cliquet options). It is well accepted that a multi-factor approach would be necessary to take into account the variability of the skew. Further, the pricing of derivatives written on more than a single underlying assets requires a sound modelling of the multivariate dependence structure. A large part of existing literature considered vector-valued stochastic processes to model the multidimensional stochastic evolution of asset(s) volatility. The choice of \mathbb{R}^d as state space for the volatility process, however, could lead to unsatisfactory dependence structures among variance factors. This is particularly true if we restrict ourselves to the case of affine processes. In the light of the above, it appears reasonable to consider more general multidimensional processes: recently an increasing attention has been devoted to applications of matrix-defined stochastic processes in derivatives pricing. In particular, stochastic processes defined on the cone of real positive semidefinite matrices $\mathcal{S}_d^+(\mathbb{R})$ can be seen as natural candidates to model the latent volatility factors.

*Corresponding author. Email: daniele.marazzina@polimi.it

In our analysis we focus on the so-called Wishart processes introduced in Bru (1991) as a matrix generalization of square-root processes. A remarkable feature is that the analytical tractability is fully preserved since these processes belong to the class of affine processes. Given the strict connection with the well-known CIR processes, Wishart processes have been used to define multi-factor (Da Fonseca *et al.* 2008) and multi-asset (Da Fonseca *et al.* 2007) extensions of the classic Heston model.

Despite the analytical tractability, the implementation of Wishart-based stochastic volatility models poses non-trivial challenges from a numerical point of view. To overcome these numerical problems, we propose efficient model approximations that alleviate the associated computational burden. These approximation schemes permit us to introduce efficient numerical schemes for Monte Carlo simulations. More in details, as far as we know, the simulation schemes proposed in literature face a severe trade-off in terms of consistency and computational complexity. For example in Gauthier and Possamai (2012) several schemes are proposed as generalization of already existing schemes for 1-factor models. While simple and easy to implement, they rely on parameters assumptions hardly met in realistic market conditions. An exact simulation scheme is proposed in Kang and Kang (2013): authors exploit the remarkable exact sampling method for the Wishart process devised in Ahdida and Alfonsi (2013). Despite its accuracy, the approach is rather involved and time consuming since it requires the computation of matrix-valued special functions to sample from the conditional distribution of the log-price given the realization of the Wishart process. Our simulation schemes embed the exact sampling scheme in Ahdida and Alfonsi (2013) and turn out to be accurate and simple to implement.

Moreover, in this article we deal with the model calibration, highlighting the constraints that need to be satisfied in order to get a well-defined Wishart process and their impact on pricing performances. In fact, as shown in Da Fonseca and Grasselli (2011), model calibration to market data turns out to be a delicate task due to the high dimensionality of the optimization problem to deal with. The issue is worsened by the fact that characteristic functions of log-asset prices, even if known in closed formula, are computationally intensive since they involve the evaluation of functions of matrix argument. As a consequence, standard calibration routines, that rely on transform-based techniques to price vanilla options, are proven to be inadequate in realistic applications. Unfortunately, the existing approximations available in literature seem to deal just with restrictive parameters settings (Gauthier and Possamai 2011) or provide reliable results only for options with moneyness close to the at-the-money level (Da Fonseca and Grasselli 2011). By exploiting the distributional properties of Wishart process, we construct efficient model approximations that mitigate the complexity of the calibration problem by replacing the Wishart-based characteristic function with those of simpler affine models. More precisely, we show that both Heston (1993) and the Bi-Heston of Christoffersen *et al.* (2009) models may provide a reliable approximation of the Wishart Stochastic Volatility model. We provide the analytical form of the gradient of calibration problem objective function with respect to Wishart-based parameters allowing for a further reduction in the computational burden. This methodology extends the remarkable result provided in Cui *et al.* (2017) where the gradient of a European call with respect to Heston model parameters is derived explicitly. Additionally, a thorough analysis of calibration outputs is carried out. The results, in line with other evidences in literature, show that optimal parameters do not satisfy the conditions to have a well defined Wishart process. Enforcing such conditions could have a significant impact on the pricing performance of these models. It seems, however, that the topic has not been developed in literature so far. For the first time, then, we study the effects produced by such constraints on the model induced implied volatility surfaces.

The article is organized as follows. In Section 2 we present the Wishart process and its basic properties. Moreover, a new approximation of the process is presented, which is the starting point for our main results. In Section 3 we deal with the Wishart Stochastic Volatility model, and we present a fast and accurate calibration procedure, based on the approximation of the Wishart process with a Bi-Heston process. Finally Section 4 presents two simulation schemes for the Wishart Stochastic Volatility model: the Stochastic Integral Approximation, and the Gaussian Variable Approximation.

2. Definition of Wishart process and basic properties

Given the importance that Wishart process plays in the proposed framework, we start our work with a brief review of its main properties.

DEFINITION 1 (*Wishart process*) Let $W(t)$ be a $d \times d$ Brownian motion (i.e. a matrix of $d \times d$ independent scalar Brownian motions) and $\mathcal{S}_d^+(\mathbb{R})$ the set of real $d \times d$ positive semidefinite matrices. We define the Wishart process as the solution on $\mathcal{S}_d^+(\mathbb{R})$ of the following stochastic differential equation (SDE):

$$\begin{aligned} d\Sigma(t) &= (\Omega\Omega^\top + M\Sigma(t) + \Sigma(t)M^\top)dt + \sqrt{\Sigma(t)} dW(t) Q + Q^\top dW^\top(t) \sqrt{\Sigma(t)}, \\ \Sigma(0) &= \Sigma_0 \in \mathcal{S}_d^+(\mathbb{R}), \end{aligned} \quad (1)$$

with $\Omega, Q, M \in \mathcal{M}_d(\mathbb{R})$ (the set of real $d \times d$ square matrices).¹

In the most general framework we can think of each element of all matrices appearing in (1) to be non null. In particular, in order to embed mean-reversion and stationarity, we consider matrix M to have only eigenvalues with negative real part (Gnoatto and Grasselli 2014). Furthermore, we relate the deterministic part of the drift in (1), $\Omega\Omega^\top$, to the expected long-term value of the process, denoted with Σ_∞ , by means of the equation

$$-\Omega\Omega^\top = M\Sigma_\infty + \Sigma_\infty M^\top. \quad (2)$$

From (1) and (2) we can easily sketch the close connection existing between Wishart and CIR processes. Indeed, Wishart processes have been firstly introduced in Bru (1991) as a multidimensional extension of classic square root process: not surprisingly if we set $d = 1$ in (1), we end up with a scalar CIR process defined by the SDE

$$dv(t) = \kappa(\theta - v(t))dt + \eta\sqrt{v(t)}dw_v(t), \quad v(0) = v_0, \quad (3)$$

with κ, θ , and η strictly positive parameters, $v_0 \geq 0$ and $w_v(t)$ a scalar Brownian motion.

A remarkable feature of Wishart process is that it entails a non-trivial dependence structure among its elements. Indeed, it holds that

$$d[\Sigma_{ij}(t), \Sigma_{kl}(t)] = (\Sigma_{ik}(t)Q_{jl}^* + \Sigma_{il}(t)Q_{jk}^* + \Sigma_{jk}(t)Q_{il}^* + \Sigma_{jl}(t)Q_{ik}^*) dt, \quad (4)$$

where the notation $[\cdot, \cdot]$ refers as usual to the quadratic covariation of two stochastic processes and $Q^* = Q^\top Q$. As we will see in the following, this property constitutes an *unicuum* among the affine generalizations of (3).

2.1. Existence and uniqueness conditions

Wishart processes are a particular case of affine processes defined on the cone $\mathcal{S}_d^+(\mathbb{R})$ for which general results about existence and uniqueness of solutions are provided in Cuchiero *et al.* (2011) and Mayerhofer *et al.* (2011). We report the main result (as formulated in Ahdida and Alfonsi (2013)) since it plays a crucial role in our framework.

¹We further use the notation $\Sigma_{i,j}(t)$ to indicate the element in the i -th row and j -th column of $\Sigma(t)$. If the time dependence is omitted, we assume to refer to time $t = 0$, i.e., $\Sigma_{i,j} = \Sigma_{i,j}(0)$.

PROPOSITION 2.1 (*Affine models in $\mathcal{S}_d^+(\mathbb{R})$* Cuchiero *et al.* (2011), Mayerhofer *et al.* (2011)) *Let $X(t)$ be a generic affine process with continuous trajectories defined in $\mathcal{S}_d^+(\mathbb{R})$ by the following SDE*

$$X(t) = X(0) + \int_0^t (D_X + \mathcal{L}[X(s)]) ds + \int_0^t \left(\sqrt{X(s)} dW(s) C_X + C_X^\top dW^\top(s) \sqrt{X(s)} \right), \quad (5)$$

where $X(0), D_X \in \mathcal{S}_d^+(\mathbb{R})$, $C_X \in \mathcal{M}_d(\mathbb{R})$, $\mathcal{L} : \mathcal{S}_d^+(\mathbb{R}) \rightarrow \mathcal{S}_d^+(\mathbb{R})$ is a linear transformation. Such process admits a unique weak solution in $\mathcal{S}_d^+(\mathbb{R})$ if

- a) $D_X - (d-1)C_X^\top C_X \in \mathcal{S}_d^+(\mathbb{R})$,
- b) $\forall P_1, P_2 \in \mathcal{S}_d^+(\mathbb{R})$ s.t. $\text{Tr}[P_1 P_2] = 0 \Rightarrow \text{Tr}[\mathcal{L}(P_1) P_2] \geq 0$, where $\text{Tr}[\cdot]$ is the trace of a square matrix (i.e., the sum of the elements on the main diagonal).

If $X(0)$ is in the set of real positive definite matrices $\mathcal{S}_d^{++}(\mathbb{R})$ and condition a) is replaced by the stronger requirement

- c) $D_X - (d+1)C_X^\top C_X \in \mathcal{S}_d^+(\mathbb{R})$,

then there exist a unique strong solution to (5) in $\mathcal{S}_d^{++}(\mathbb{R})$.

A direct comparison shows that we can get the Wishart SDE (1) from (5) by setting $D_X = \Omega\Omega^\top$, $C_X = Q$, and $\mathcal{L}[P_0] = MP_0 + P_0M^\top$. Moreover, if we assume a more restrictive parametrization for the deterministic part of the drift

$$\Omega\Omega^\top = \beta Q^\top Q, \quad (6)$$

conditions a) and c) of Proposition 2.1 are satisfied as soon as

$$\beta \geq d-1, \quad (7)$$

and

$$\beta \geq d+1, \quad (8)$$

respectively, where the real positive parameter β plays the role of Feller's condition in the univariate case. Additionally if condition a) is not met the whole process is not well defined.

For the rest of the paper we consider a Wishart process defined by (1) and (6) as usually done in financial literature. A significant constraint has thus to be imposed on parameter β . In the case $d=2$, for example, we must require $\beta \geq 1$. As we are going to see, this condition is not usually met when we perform a straight calibration of Wishart-based pricing models to market prices of plain vanilla options.

2.2. Distribution of Wishart process and related results

In this section we present more insights about the analogy between Wishart and CIR (Cox *et al.* 1985) processes by formalizing some distributional features of the Wishart process.

Exploiting the affine nature of the Wishart process, we have that its characteristic function is an exponential affine transformation of the initial state as shown in the following proposition:

PROPOSITION 2.2 (*Characteristic function of Wishart process*) *Let Λ be a real symmetric $d \times d$ matrix, $t \geq 0$ and $T-t = \tau > 0$. The (conditional) characteristic function of the Wishart process defined by (1) and (6) is*

$$\phi_\Sigma(\Lambda, \tau) = \mathbb{E}[\exp(\iota \text{Tr}[\Lambda \Sigma(T)]) | \Sigma(t)] = \exp(\text{Tr}[A_\Sigma(\Lambda, \tau) \Sigma(t)] + b_\Sigma(\Lambda, \tau)), \quad (9)$$

where ι is the imaginary unit (i.e. $\iota = \sqrt{-1}$) and matrix $A_\Sigma(\Lambda, \tau)$ and scalar function $b_\Sigma(\Lambda, \tau)$ are such that

$$\begin{aligned}\text{Tr}[A_\Sigma(\Lambda, \tau)\Sigma(t)] &= \text{Tr}\left[\iota\Lambda(\mathbb{I}_d - 2\iota\Theta(\tau)\Lambda)^{-1}\Gamma(\tau)\right], \\ b_\Sigma(\Lambda, \tau) &= -\frac{\beta}{2}\text{Tr}\left[\log\left((\mathbb{I}_d - 2\iota\Theta(\tau)\Lambda)\exp\left(\tau M^\top\right)\right) - \tau M\right].\end{aligned}$$

The additional matrix functions appearing in the above equations are given by

$$\begin{aligned}\Gamma(\tau) &= \exp(\tau M)\Sigma(t)\exp(\tau M^\top), \\ \Theta(\tau) &= \int_0^\tau \exp(uM)Q^\top Q\exp(uM^\top)du.\end{aligned}$$

Proof. Due to affinity, it is possible to show that $A_\Sigma(\Lambda, \tau)$ is solution of a matrix Riccati ODE and $b_\Sigma(\Lambda, \tau)$ is obtained by direct integration. We refer to Grasselli and Tebaldi (2008) for a general discussion of the solving technique and Ahdida and Alfonsi (2013) for the explicit derivation of the result. Here we simply prove that the matrix $\mathcal{G} = (\mathbb{I}_d - 2\iota\Theta(\tau)\Lambda)$ is invertible, i.e. $\det[\mathcal{G}] \neq 0$. Matrix $\Theta(\tau)$ is trivially positive semi-definite¹ (since it is the integral of a positive semi-definite matrix) and let $\mathcal{H}\mathcal{H}^\top = \Theta(\tau)$. We then have

$$\det[\mathcal{G}] = \det[\mathbb{I}_d - 2\iota\mathcal{H}\mathcal{H}^\top\Lambda] = \det[\mathbb{I}_d - 2\iota\mathcal{H}^\top\Lambda\mathcal{H}] = \det[\mathbb{I}_d - 2\iota P\mathcal{D}P^\top] = \det[\mathbb{I}_d - 2\iota\mathcal{D}],$$

where the second equality is a direct application of the Sylvester's determinant identity and the third follows from the spectral theorem with \mathcal{D} the diagonal matrix whose entries are the eigenvalues of the symmetric real matrix $\mathcal{H}^\top\Lambda\mathcal{H} = P\mathcal{D}P^\top$. By recognizing that $\det[\mathbb{I}_d - 2\iota\mathcal{D}]$ is the characteristic polynomial of the diagonal imaginary matrix $2\iota\mathcal{D}$ evaluated at 1 we immediately get the invertibility of \mathcal{G} . \square

As a consequence of the analytical tractability of Wishart process and of the knowledge of its characteristic function, we are able to present an additional result regarding the distribution of the trace of the Wishart process, i.e. the process $\mathcal{V}(t) = \text{Tr}[\Sigma(t)]$. In Da Fonseca *et al.* (2008) it is shown that via direct application of Itô's lemma the dynamics of $\mathcal{V}(t)$ is given by

$$d\mathcal{V}(t) = \left(\text{Tr}\left[\beta Q^\top Q\right] + 2\text{Tr}\left[M\Sigma(t)\right]\right)dt + 2\text{Tr}\left[\sqrt{\Sigma(t)}dW(t)Q\right]. \quad (10)$$

The following result, firstly appeared in Kourouklis and Moschopoulos (1985) for a generic non-central Wishart distribution, allows to express the trace of Wishart process as a weighted sum of d independent non-central chi-square random variables:

COROLLARY 2.3 (*Distribution of the Trace of Wishart process*) *Let $\mathcal{V}(t)$ be the sum of the elements of the main diagonal of $\Sigma(t)$ and $\chi_\nu^2(\delta)$ be a non-central chi-square random variable with ν degrees of freedom and non-centrality parameter δ . For a fixed $T > t$, let \mathcal{Q} be the orthogonal matrix that diagonalizes $\Theta(\tau)$, that is $\mathcal{Q}^\top\Theta(\tau)\mathcal{Q} = \mathcal{E} = \text{diag}[\varepsilon_1, \varepsilon_2, \dots, \varepsilon_d]$. Then $\mathcal{V}(T)$ admits the representation*

$$\mathcal{V}(T) = \sum_{i=1}^d \varepsilon_i \chi_\beta^2\left(\frac{\zeta_i}{\varepsilon_i}\right) \quad (11)$$

¹Furthermore, in the following we assume it to be non singular.

where ζ_i is the i -th diagonal element of $\mathcal{Q}^\top \Gamma(\tau) \mathcal{Q}$.

Proof. We propose a new simple proof of the result by exploiting the characteristic function of the Wishart process. We use the fact that $\exp(\text{Tr}[\log(G)]) = \det[G]$ for any matrix invertible matrix¹ G to write

$$\exp(b_\Sigma(\Lambda, \tau)) = \det \left[(\mathbb{I}_d - 2i\Theta(\tau)\Lambda) \exp(\tau M^\top) \right]^{-\frac{\beta}{2}} \exp \left(\frac{\beta}{2} \text{Tr}[M] \tau \right) = \det [\mathbb{I}_d - 2i\Theta(\tau)\Lambda]^{-\frac{\beta}{2}},$$

and (9) becomes

$$\phi_\Sigma(\Lambda, \tau) = \det [\mathbb{I}_d - 2i\Theta(\tau)\Lambda]^{-\frac{\beta}{2}} \exp \left(\text{Tr} \left[i\Lambda (\mathbb{I}_d - 2i\Theta(\tau)\Lambda)^{-1} \Gamma(\tau) \right] \right). \quad (12)$$

Let λ be a real variable, then by setting $\Lambda_\nu = \lambda \mathbb{I}_d$, the characteristic function of $\mathcal{V}(T)$ is

$$\begin{aligned} \phi_\nu(\lambda, \tau) &= \phi_\Sigma(\Lambda_\nu, \tau) = \det [\mathbb{I}_d - 2i\lambda\Theta(\tau)]^{-\frac{\beta}{2}} \exp \left(i\lambda \text{Tr} \left[(\mathbb{I}_d - 2i\lambda\Theta(\tau))^{-1} \Gamma(\tau) \right] \right) \\ &= \det [\mathbb{I}_d - 2i\lambda\mathcal{Q}\mathcal{E}\mathcal{Q}^\top]^{-\frac{\beta}{2}} \exp \left(i\lambda \text{Tr} \left[(\mathbb{I}_d - 2i\lambda\mathcal{Q}\mathcal{E}\mathcal{Q}^\top)^{-1} \Gamma(\tau) \right] \right) \\ &= \det [\mathbb{I}_d - 2i\lambda\mathcal{E}]^{-\frac{\beta}{2}} \exp \left(i\lambda \text{Tr} \left[\mathcal{Q}^\top \Gamma(\tau) \mathcal{Q} (\mathbb{I}_d - 2i\lambda\mathcal{E})^{-1} \right] \right) \\ &= \prod_{i=1}^d (1 - 2i\lambda\varepsilon_i)^{-\frac{\beta}{2}} \exp \left(\frac{i\lambda\zeta_i}{1 - 2i\lambda\varepsilon_i} \right) \end{aligned} \quad (13)$$

where the last equality comes from the properties of determinant and trace. \square

From the knowledge of the characteristic function, we can also easily compute the (conditional) moments² of \mathcal{V}_T :

$$\mathbb{E}[\mathcal{V}(T)|\Sigma(t)] = \text{Tr}[\Gamma(\tau) + \beta\Theta(\tau)], \quad (14)$$

$$\text{Var}[\mathcal{V}(T)|\Sigma(t)] = 2 \text{Tr}[(2\Gamma(\tau) + \beta\Theta(\tau))\Theta(\tau)], \quad (15)$$

and so on, with higher moments that get more and more involved.

The representation (11) allows to consider $\mathcal{V}(T)$ in terms of non-negative definite quadratic forms in non-central normal variables (see for example Imhof (1961) and Kotz *et al.* (1967)) typically arising in statistical applications. As we will see in the next section, it could be convenient to dispose of some approximations for the distribution of $\mathcal{V}(T)$. A feasible tool is proposed in Liu *et al.* (2009) where the distribution of a non-negative quadratic form is approximated by means of an affine transformation of a non-central chi-square random variable. The unknown parameters are then chosen so that the first three cumulants of the quadratic form are matched and the difference in kurtosis is minimized. In our case this means that

$$\Pr[\mathcal{V}(T) > \varrho | \Sigma(t)] \approx \Pr[\alpha_1 \chi_{\nu^*}^2(\delta^*) + \alpha_0 > \varrho] \quad (16)$$

¹Here we exploit the invertibility of $(\mathbb{I}_d - 2i\Theta(\tau)\Lambda) \exp(\tau M^\top)$ which is given by the invertibility of both \mathcal{G} (proved in the previous Proposition) and $\exp(\tau M^\top)$ (by definition of matrix exponential).

²Let \mathcal{X} be a random variable with characteristic function $\phi_{\mathcal{X}}(\lambda) = \mathbb{E}[\exp(i\lambda\mathcal{X})]$, the computation relies on the standard formula $\mathbb{E}[\mathcal{X}^n] = i^{-n} \phi_{\mathcal{X}}^{(n)}(0)$. Additionally, we exploit the fact that for a matrix $U = U(\lambda)$ it holds that $\frac{d}{d\lambda} \det[U] = \det[U] \text{Tr} \left[U^{-1} \frac{dU}{d\lambda} \right]$ and $\frac{dU^{-1}}{d\lambda} = -U^{-1} \frac{dU}{d\lambda} U^{-1}$. We leave the details of the computation to the reader.

where ν^* and δ^* are, respectively, the optimal degrees of freedom and non-centrality parameters while $\alpha_1 = \sqrt{\frac{\text{Var}[\mathcal{V}(T)|\Sigma(t)]}{2(\nu^* + 2\delta^*)}}$ and $\alpha_0 = \mathbb{E}[\mathcal{V}(T)|\Sigma(t)] - \alpha_1(\nu^* + \delta^*)$.

2.3. A new approximation of $\mathcal{V}(t)$

Even if appealing from a theoretical point of view, the approximation (16) turns out to be inadequate in those cases in which $\mathcal{V}(t)$ describes the instantaneous variance of an asset returns. We here develop an alternative approximation inspired by the idea to exploit a non-central chi-square random variable to approximate $\mathcal{V}(T)$. Instead of considering an affine transformation, however, we just use a scaled non-central chi-square random variable with parameters fitted to match (14) and (15). In other words, we aim at approximating \mathcal{V} with a scalar CIR process as given in the following proposition:

PROPOSITION 2.4 (*Moment-matching CIR process approximation of $\mathcal{V}(T)$*) *Let $v(t)$ be a CIR process defined by (3). For a fixed $T > t$, $v(t)$ is a 2-moment matching approximation of $\mathcal{V}(T)$ provided that*

$$v(t) = \mathcal{V}(t), \quad (17)$$

$$\theta = \text{Tr}[\Sigma_\infty] > 0, \quad (18)$$

$$\kappa = -\frac{1}{\tau} \log \left(\frac{\mathbb{E}[\mathcal{V}(T)|\Sigma(t)] - \theta}{\mathcal{V}(t) - \theta} \right), \quad (19)$$

$$\eta^2 = \frac{\kappa \text{Var}[\mathcal{V}(T)|\Sigma(t)]}{a \left((1-a)\mathcal{V}(t) + \frac{a}{2}\theta \right)}, \quad (20)$$

with $a = (1 - \exp(-\kappa\tau))$ and provided that $\mathcal{V}(t) \neq \theta$. A sufficient condition for κ to be well-defined (i.e. to be a positive real number) is to have the matrix

$$\mathcal{F} := M\Sigma(t) + \Sigma(t)M^\top + \beta Q^\top Q$$

(positive or negative) definite.

Proof. Given the distributional properties of CIR process, the process $\tilde{v}(t)$ with parameters (17)-(20) has the same first two T -conditional moments of $\mathcal{V}(T)$. Furthermore, given the stability¹ of M , we have that $\theta \geq 0$, with the equality that holds only in the degenerate case $Q = 0_d$ (the zero matrix of order d). Indeed, for M stable, the solution of the Lyapunov equation (2) admits the integral representation

$$\Sigma_\infty = \beta \int_0^\infty \exp(uM) Q^\top Q \exp(uM^\top) du,$$

which is positive semidefinite and then its trace is non-negative. Additionally it is possible to show that $\text{Tr}[\Sigma_\infty] = 0$ if and only if $Q = 0_d$. To see this, let us write

$$\theta = \beta \text{Tr} \left[\int_0^\infty \exp(uM) Q^\top Q \exp(uM^\top) du \right] = \beta \text{Tr} \left[Q^\top Q \int_0^\infty \exp(uM) \exp(uM^\top) du \right].$$

¹A square matrix is called stable if all its eigenvalues have strictly negative real part.

We now have to prove that θ is strictly positive. The claim follows² by using the fact that θ is the trace of the product of positive semi-definite matrix $\beta Q^\top Q$ and the positive definite matrix $\int_0^\infty \exp(uM) \exp(uM^\top) du$.

From (19) we see that $\kappa \in \mathbb{R}_{>0}$ as soon as $0 < \frac{\mathbb{E}[\mathcal{V}(T)|\Sigma(t)] - \theta}{\mathcal{V}(t) - \theta} < 1$, with 2 possible cases:

- if $\mathcal{V}(t) - \theta > 0$, it must hold $\theta < \text{Tr}[\Gamma(\tau) + \beta\Theta(\tau)] < \mathcal{V}(t)$,
- otherwise $\theta > \text{Tr}[\Gamma(\tau) + \beta\Theta(\tau)] > \mathcal{V}(t) \geq 0$.

In order for the condition to be fulfilled for any $T > t$, we can require the function

$$\mathbf{g} : \tau \mapsto \text{Tr}[\Gamma(\tau) + \beta\Theta(\tau)]$$

to be monotone in $[0, \infty)$. We indeed have that $\mathcal{V}(t) = \text{Tr}[\Gamma(0) + \beta\Theta(0)]$ and $\theta = \lim_{\tau \rightarrow +\infty} \text{Tr}[\Gamma(\tau) + \beta\Theta(\tau)]$. It is easy to check that $E(\tau) = \mathbb{E}[\Sigma(T)|\Sigma(t)] = \Gamma(\tau) + \beta\Theta(\tau)$ is solution of the first order, linear inhomogeneous matrix ODE

$$\dot{E}(\tau) = ME(\tau) + E(\tau)M^\top + \beta Q^\top Q \quad (21)$$

with $E(0) = \Sigma(t)$. Differentiating (21) we obtain

$$\ddot{E}(\tau) = M\dot{E}(\tau) + \dot{E}(\tau)M^\top$$

with initial condition $\dot{E}(0) = M\Sigma(t) + \Sigma(t)M^\top + \beta Q^\top Q = \mathcal{F}$. Following standard variation of constants arguments, we have

$$\dot{E}(\tau) = \exp(\tau M) \mathcal{F} \exp(\tau M^\top). \quad (22)$$

This in turns means that the derivative of function \mathbf{g} can be written as

$$\frac{d}{d\tau} \mathbf{g}(\tau) = \text{Tr}[\dot{E}(\tau)] = \text{Tr}[\mathcal{F} \exp(\tau M^\top) \exp(\tau M)]. \quad (23)$$

As shown in Wang *et al.* (1986), for any real symmetric matrix A and $B \in \mathcal{S}_d^+(\mathbb{R})$, the following inequality holds

$$a_d \text{Tr}[B] \leq \text{Tr}[AB] \leq a_1 \text{Tr}[B]$$

where a_i is the i -th largest eigenvalue of A . Applying this result to (23) we get the sufficient condition for the monotonicity of $\mathbf{g}(\tau)$ in terms of definiteness of \mathcal{F} . We see, indeed, that if \mathcal{F} is (positive or negative) definite, the derivative of \mathbf{g} is bounded to be always positive or negative - depending on the sign of the eigenvalues of \mathcal{F} - on the interval $[0, \infty)$. \square

Unfortunately, since we deal with traces of matrix products, obtaining sharp conditions for the monotonicity of \mathbf{g} is not an easy task. We are able just to provide the sufficient condition claimed, based on the known (loose) bounds for the trace of the product of two matrices. However, for a wide range of realistic parameter values we could have that κ is still well-defined even if the condition

²Specifically, there is no nilpotent matrix (i.e. with all the eigenvalues equal to zero), except the zero matrix, that can be obtained as the product of 2 non-negative matrices. Moreover we exploit the fact that if A and B are two non-zero square matrices such that $AB = 0$, then both A and B must be singular. Given that $\int_0^\infty \exp(uM) \exp(uM^\top) du$ is positive definite (indeed, it is the integral of a positive definite matrix), we get the result.

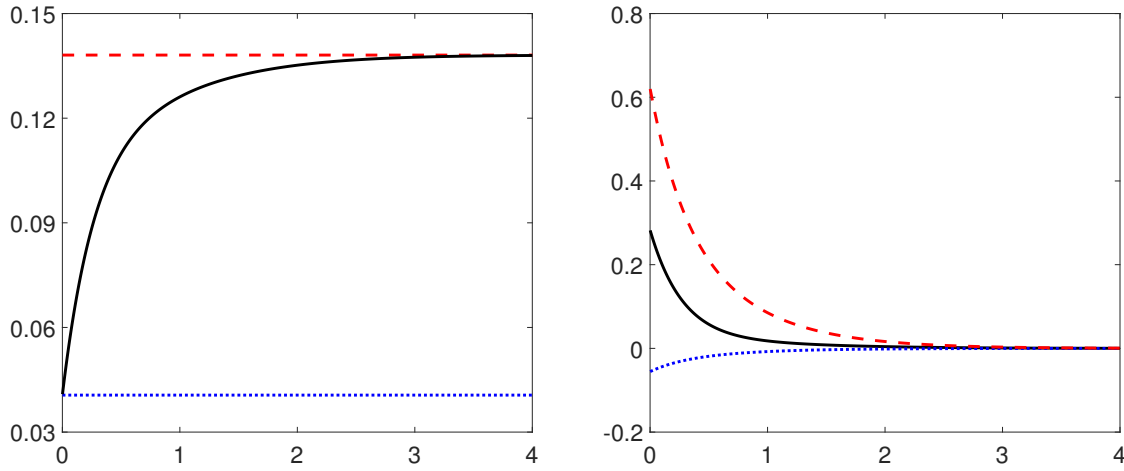


Figure 1.: Left panel: $g(\tau)$ in the interval $[0, 4]$ (solid line). It departs from $\text{Tr}[\Sigma_0]$ (dotted line) for $\tau = 0$ and converges towards θ (dashed line) as $\tau \rightarrow +\infty$. Right panel: derivative of $g(\tau)$ (solid line) and its bounds.

on \mathcal{F} is not fulfilled. As an example, we consider the parameters set:

$$\begin{aligned} \beta &= 1.1, & \Sigma_0 &= \begin{bmatrix} 0.0298 & 0.0119 \\ 0.0119 & 0.0108 \end{bmatrix}, \\ M &= \begin{bmatrix} -1.2479 & -0.8985 \\ -0.0820 & -1.1433 \end{bmatrix}, & Q &= \begin{bmatrix} 0.3417 & 0.3493 \\ 0.1848 & 0.3090 \end{bmatrix}. \end{aligned} \quad (24)$$

The parameters are taken from Da Fonseca and Grasselli (2011) with the exception of β that we set higher than 1 in order to deal with a well-defined Wishart process. Interestingly, this is the same set used in Kang and Kang (2013). As shown in Figure 1, the function $g(\tau)$ is monotonic even if \mathcal{F} is indefinite (its eigenvalues are 0.3106 and -0.0278). From the rightmost panel of Figure 1, it is also evident that for short time frames, the bounds for the trace of $E(\tau)$ are not really tight. For $d = 2$, this means that if the non-dominant eigenvalue has different sign with respect to the dominant one but its value is enough close to zero, the sign of the derivative of $g(\tau)$ is likely not to change.

We now deal with the distribution of the square root of $\mathcal{V}(T)$ that can be used, as shown in the next section, to model the stochastic volatility in a multi-factor extension of the Heston model.

It is possible to recover the probability density of $\sqrt{\mathcal{V}(T)}$ by applying the formula

$$p_{\sqrt{\mathcal{V}(T)}}(x) = 2x p_{\mathcal{V}(T)}(x^2) \quad \forall x \in \mathbb{R}_{\geq 0},$$

where the probability density function of $\mathcal{V}(T)$, $p_{\mathcal{V}(T)}$, can be obtained via FFT or quadrature methods from (13). In Figure 2 we also consider the distribution induced by the two approximations taken into account. It is worth noting that for the affine transformation approach, there is a discontinuity at the point $\sqrt{\alpha_0}$ such that for lower values the probability density is zero: the far left tail of the exact distribution is unattainable and compensated by a corresponding peak in a (positive) neighborhood of the discontinuity. Our moment-matching approximation, on the other hand, does not suffer from this drawback by construction (since it has no displacement term) and it is still capable of reproducing the distributional properties of $\sqrt{\mathcal{V}(T)}$. In particular, our approach gives an accurate approximation in the right tail of the distribution (comparable to the alternative affine approximation) while for intermediate values it fails in some extent in reproducing correctly the peak in the exact distribution. Considering, however, that this is an approximation based simply on the matching of the first 2 moments of $\mathcal{V}(T)$, a lack of accuracy is somehow expected. Finally, we

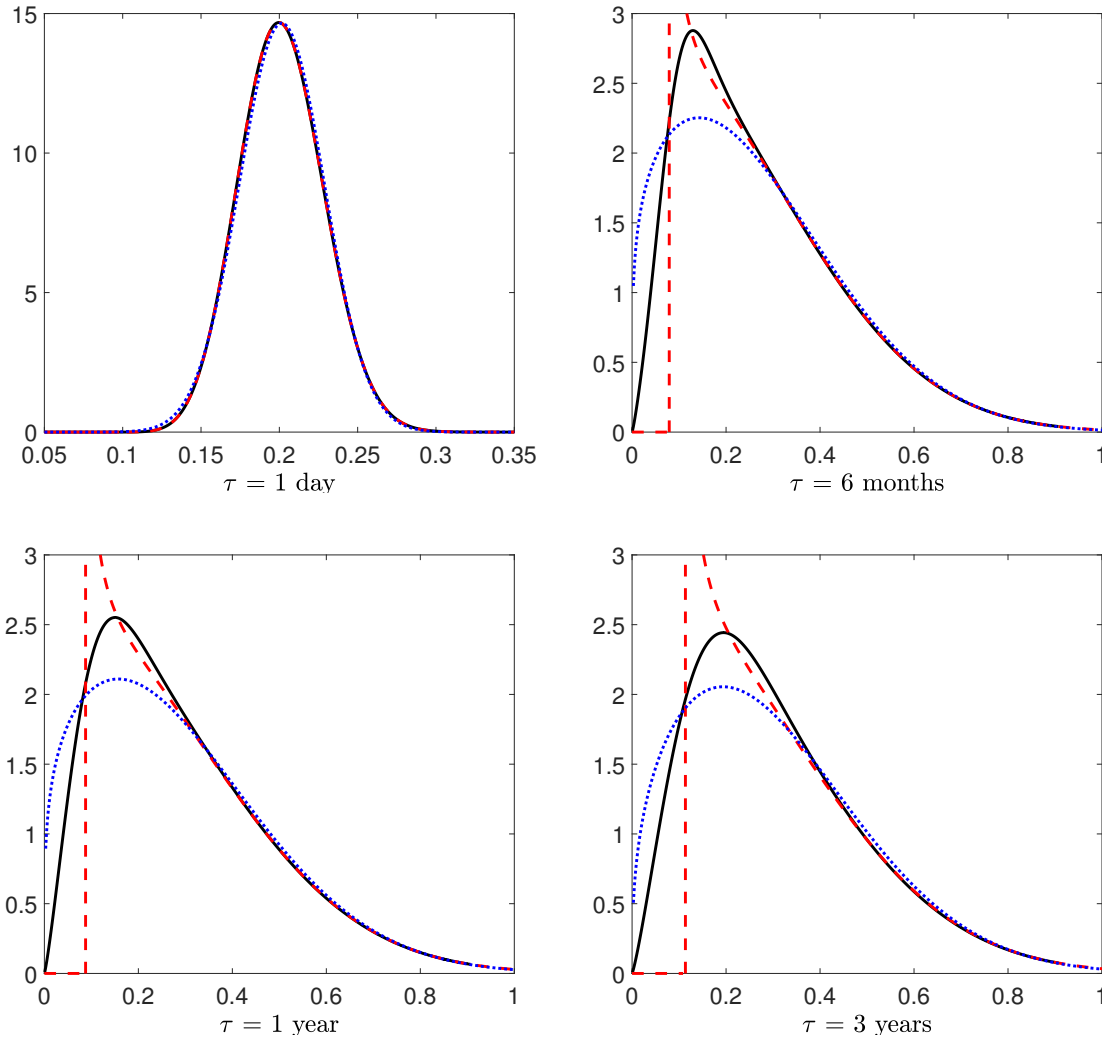


Figure 2.: Comparison among exact probability density function of $\sqrt{\mathcal{V}(T)}$ (solid line), approximation via affine transformation of non-central chi-square random variables (16) (dashed line) and our moment-matching approximation (dotted line) for different time horizons.

want to stress that the approximation proposed in Proposition 2.4 is not meant to strictly describe the distribution of $\mathcal{V}(T)$ but rather to provide a simple and reliable tool to map the parameters that drive the trace of $\Sigma(t)$ to those of a simpler scalar CIR process. In the next section we use such mapping to develop efficient numerical techniques for models that use $\sqrt{\mathcal{V}(t)}$ to describe the multi-factor dynamics of asset volatility. Since our main concern in this paper is the pricing of contingent claims, from now on we assume to operate under a risk-neutral measure as defined in the standard way.

3. The Wishart Stochastic Volatility Model and its calibration

In order to overcome inherent limitations of 1-factor SV models in describing the term structure of volatility skew, as documented for example in Christoffersen *et al.* (2009) and Cont and Da Fonseca (2002), a matrix generalization of Heston model is proposed in Da Fonseca *et al.* (2008). This is the Wishart Stochastic Volatility model (WSVM) where the dynamics of the forward-price of an

equity asset is given by

$$df(t) = f(t) \operatorname{Tr} \left[\sqrt{\Sigma(t)} dB(t) \right], \quad f(0) > 0, \quad (25)$$

where $B(t)$ is a d -dimensional matrix Brownian such that

$$B(t) = W(t)R^\top + Z(t) \sqrt{\mathbb{I}_d - RR^\top}, \quad (26)$$

with $Z(t)$ another matrix of Brownian motions independent of $W(t)$ and $R \in \mathcal{M}_d(\mathbb{R})$ that fulfills the condition $\mathbb{I}_d - RR^\top \in \mathcal{S}_d^+(\mathbb{R})$. This correlation structure is required in order to preserve the analytical tractability of the model: the WSVM so defined, indeed, belongs to the class of affine models in the sense of Duffie *et al.* (2003).

From now on we consider the WSVM dynamics (25) with the underling assumption of $d = 2$. This is the most common parametrization appeared in literature and constitutes an adequate balance between parsimony and flexibility.

In the original paper Christoffersen *et al.* (2009), the authors show that the WSVM can be expressed in a scalar form that, as we are going to see, turns out to be extremely useful in our framework. Before stating this result, we recall an auxiliary Lemma:

LEMMA 3.1 (*Lemma 4.6 in Mayerhofer et al. (2011)*) *Let $X(t)$ be a continuous stochastic process on $\mathcal{S}_d^+(\mathbb{R})$ and let $h : \mathcal{M}_d(\mathbb{R}) \rightarrow \mathcal{M}_d(\mathbb{R})$. Then there exists a scalar Brownian motion $w_h(t)$ such that*

$$\operatorname{Tr} \left[\int_0^T h(X(u)) dW(u) \right] = \int_0^T \sqrt{\operatorname{Tr} [h(X(u))^\top h(X(u))]} dw_h(u)$$

holds true.

PROPOSITION 3.2 (*Scalar version of WSVM dynamics*) *Let $y(t) = \log(f(t))$ be the asset log-price, then its dynamics in the WSVM can be written as*

$$dy(t) = -\frac{1}{2} \mathcal{V}(t) dt + \sqrt{\mathcal{V}(t)} db(t), \quad (27)$$

$$d\mathcal{V}(t) = \left(\operatorname{Tr} [\beta Q^\top Q] + 2 \operatorname{Tr} [M\Sigma(t)] \right) dt + 2\sqrt{\operatorname{Tr} [\Sigma(t)Q^\top Q]} dw(t), \quad (28)$$

where $b(t)$ and $w(t)$ are two scalar Brownian motions with stochastic correlation given by

$$\rho_w(t) = \frac{\operatorname{Tr} [RQ\Sigma(t)]}{\sqrt{\operatorname{Tr} [\Sigma(t)]} \sqrt{\operatorname{Tr} [Q^\top Q\Sigma(t)]}}. \quad (29)$$

Proof. See (Da Fonseca *et al.* 2008, Appendix A). Notice that standard Itô's lemma immediately implies that

$$dy(t) = -\frac{1}{2} \mathcal{V}(t) dt + \operatorname{Tr} \left[\sqrt{\Sigma(t)} dB(t) \right], \quad (30)$$

and applying Lemma 3.1 we get

$$\operatorname{Tr} \left[\sqrt{\Sigma(t)} dB(t) \right] = \sqrt{\operatorname{Tr} [\Sigma(t)]} db(t) = \sqrt{\mathcal{V}(t)} db(t).$$

□

This result highlights also the appealing property that, in the WSVM framework, the correlation between stock returns and volatility is stochastic.

An important peculiarity of WSVM is that, differently than other multifactor extensions of the Heston model, like the Bi-Heston model in Christoffersen *et al.* (2009), matrix specification of volatility factors allows to separately manage and calibrate implied volatility levels and skew thanks to the presence of non null off-diagonal elements in Σ_0 . Let us suppose to fix the elements on the diagonal of Σ_0 to match the term structure of implied volatility. We still have the possibility to fit the skew thanks to the residual elements in Σ_0 . This is also confirmed by the analysis performed in Da Fonseca and Grasselli (2011) where the expansion of model implied volatility for short times to maturity is found to be

$$\begin{aligned} \sigma_{imp}^2 = & \text{Tr}[\Sigma_0] + \frac{\text{Tr}[RQ\Sigma_0]}{\text{Tr}[\Sigma_0]}mf \\ & + \frac{mf^2}{(\text{Tr}[\Sigma_0])^2} \left(\frac{1}{3} \text{Tr}[Q^\top Q\Sigma_0] + \frac{1}{3} \text{Tr}[RQ(Q^\top R^\top + RQ)\Sigma_0] - \frac{5}{4} \frac{(\text{Tr}[RQ\Sigma_0])^2}{\text{Tr}[\Sigma_0]} \right) \end{aligned} \quad (31)$$

where $mf = \log \frac{K}{f(t)}$ denotes the log-forward moneyness. From (31), we can appreciate that the off-diagonal element Σ_{12} does not affect the level of the smile but it has a relevant impact on the slope of implied volatility.

3.1. Connection with Heston and Bi-Heston model

In this section we study the relationship between WSVM and other (simpler) affine stochastic volatility models, namely the Heston (1993) and Bi-Heston of Christoffersen *et al.* (2009) ones. The idea comes from the distributional properties of the trace of Wishart process and its role in describing the asset volatility. This will turn out to be extremely useful in devising efficient calibration algorithms. In Section 2 we propose to approximate the conditional distribution of $\mathcal{V}(T)$ by means of a scaled non-central χ^2 random variable and we derive the (T -specific) parameters of the corresponding CIR process. For a fixed time horizon T , then, the WSVM dynamics can be approximated by the Heston one

$$\begin{aligned} df(t) &= f(t)\sqrt{v(t)}db(t) \\ dv(t) &= \kappa(\theta - v(t))dt + \eta\sqrt{v(t)}dw(t), \quad v(0) = v_0 \geq 0 \end{aligned}$$

with parameters v_0 , κ , θ and η defined in Proposition 2.4 (here we assume $t = 0$). The asset-volatility correlation is driven by a constant parameter ρ (in the sense that $db(t)dw(t) = \rho dt$) that we set as

$$\rho = \frac{\text{Tr}[RQ\Sigma_0]}{\sqrt{\text{Tr}[\Sigma_0]}\sqrt{\text{Tr}[Q^\top Q\Sigma_0]}}, \quad (32)$$

where the right-hand side of (32) is the initial value of the process $\rho_W(t)$. This means that given a WSVM parameters set π_W and a time horizon T , we can construct a model approximation by mapping WSVM parameters into Heston ones. In other words, we define a function $g^{H-W} : \mathbb{R}^{N_W} \times \mathbb{R}_{>0} \rightarrow \mathbb{R}^5$ with N_W the number of parameters in the chosen configuration of WSVM,¹

¹For example, by considering M , Q and R to be full matrices we have $\pi_W = [\beta, \Sigma_{11}, \Sigma_{12}, \Sigma_{22}, M_{11}, M_{12}, M_{21}, M_{22}, Q_{11}, Q_{12}, Q_{21}, Q_{22}, R_{11}, R_{12}, R_{21}, R_{22}]^\top$ and $N_W = 16$.

such that $g^{H-W}(\pi_W, T) = \pi_H = [v_0, \kappa, \theta, \eta, \rho]^\top$. We refer to function g^{H-W} as the *WSVM-Heston* mapping.

From Corollary 2.3 we know that the asset variance in WSVM is described by a linear combination of 2 independent non-central χ^2 random variables. This is the same distributional assumption underlying the Bi-Heston model proposed in Christoffersen *et al.* (2009), where the asset dynamics is the following:

$$\begin{aligned} df(t) &= f(t) \left(\sqrt{v_1(t)} db_1(t) + \sqrt{v_2(t)} db_2(t) \right) \\ dv_1(t) &= \kappa_1 (\theta_1 - v_1(t)) dt + \eta_1 \sqrt{v_1(t)} dw_1(t), \quad v_1(0) = v_{0,1} \geq 0 \\ dv_2(t) &= \kappa_2 (\theta_2 - v_2(t)) dt + \eta_2 \sqrt{v_2(t)} dw_2(t), \quad v_2(0) = v_{0,2} \geq 0 \end{aligned}$$

with $db_i(t)dw_i(t) = \rho_i dt$ for $i = 1, 2$ and all other correlation are equal to zero to preserve the affinity of the model. As shown in Christoffersen *et al.* (2009), in this model the variance of log-asset price is the sum of the 2 independent CIR processes

$$\text{Var}[d \log(f(t))] = (v_1(t) + v_2(t)) dt = v_{BH}(t) dt. \quad (33)$$

Furthermore, the model presents a stochastic asset-variance correlation given by

$$\text{Corr}[d \log(f(t)), dv_{BH}(t)] = \frac{\eta_1 \rho_1 v_1(t) + \eta_2 \rho_2 v_2(t)}{\sqrt{\eta_1^2 v_1(t) + \eta_2^2 v_2(t)} \sqrt{v_1(t) + v_2(t)}} dt. \quad (34)$$

In the lights of the analogy between (11) and (33), it could be interesting to find a suitable parameters set for the Bi-Heston model such that it represents a close approximation of WSVM.

For a fixed $T > 0$, we propose to approximate the WSVM dynamics (25) by means of a Bi-Heston model whose CIR processes parameters are (for $i = 1, 2$)

$$v_{0,i} = \tilde{v}_i, \quad (35)$$

$$\kappa_i = -\frac{1}{T} \log \left(\frac{\zeta_i}{v_{0,i}} \right), \quad \eta_i = 2 \sqrt{\frac{\varepsilon_i \kappa_i}{(1 - e^{-\kappa_i T})}}, \quad \theta_i = \frac{\beta \eta_i^2}{4 \kappa_i}, \quad (36)$$

where \tilde{v}_i and ζ_i are, respectively, the i -th diagonal elements of matrices $\mathcal{Q}^\top \Sigma_0 \mathcal{Q}$ and $\mathcal{Q}^\top \Gamma(T) \mathcal{Q}$. Parameters κ_i , θ_i , and η_i directly follow from the representation of the distribution of $\mathcal{V}(T)$ as formulated in Corollary 2.3. Reasonably, we set the initial values of variance processes (35) to be equal to the diagonal elements of the matrix $\mathcal{Q}^\top \Sigma_0 \mathcal{Q}$, i.e. the matrix obtained applying the change of basis that diagonalizes $\Theta(T)$.

The choice of coefficients ρ_i is the most problematic. Not only because we want to "map" the effect of the 2×2 matrix R into just 2 parameters, but also because there seems not to be an immediate way to link (29) and (34). In Da Fonseca *et al.* (2008) it is shown that the covariation between Σ_i and log-asset price y induced by the WSVM is $d[y(t), \Sigma_i(t)] = 2 \sum_{k,h=1}^2 \Sigma_{ih}(t) Q_{ki} R_{hk} dt = 2 D_{y\Sigma_i}(t) dt$.

A tempting solution could be, then, to set ρ_i as the stochastic correlation between Σ_i and y valued at $t = 0$:

$$\rho_i = \frac{D_{y\Sigma_i}(0)}{\sqrt{\text{Tr}[\Sigma_0]} \sqrt{\Sigma_i Q_{ii}^*}}$$

where we make use of (4) to obtain $d[\Sigma_i(t), \Sigma_i(t)] = 4 \Sigma_i(t) Q_{ii}^* dt$. However, extensive numerical

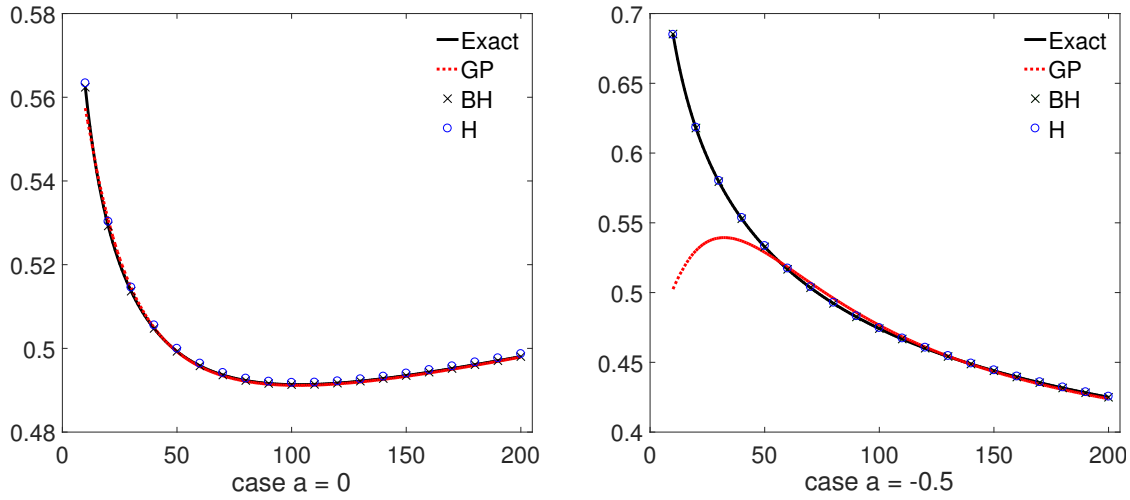


Figure 3.: Implied variance smiles for the WSVM (Exact), the second-order expansion in Gauthier and Possamaï (2011) (GP), our Bi-Heston (BH) and Heston (H) approximations. Left panel $a = 0$, right panel $a = -0.5$.

experiments provide evidences in favour of the following alternative formulation

$$\rho_i = \frac{D_{y\Sigma_i}(0)}{\Sigma_i \sqrt{Q_{ii}^*}}. \quad (37)$$

As done for the Heston model, we can define the *WSVM-BiHeston* mapping $g^{BH-W}(\pi_W, T) = \pi_{BH} = [v_{0,1}, \kappa_1, \theta_1, \eta_1, \rho_1, v_{0,2}, \kappa_2, \theta_2, \eta_2, \rho_2]^\top$.

We now want to test the approximations proposed and assess their accuracy. In the first numerical experiment, we compare the performance of our new methodologies with the second-order price expansion developed in Gauthier and Possamaï (2011). The parameters are those considered in Gauthier and Possamaï (2011):

$$\beta = 4, \quad \Sigma_0 = \begin{bmatrix} 0.4 & 0 \\ 0 & 0.35 \end{bmatrix}, \quad M = \begin{bmatrix} -1 & 0 \\ 0 & -0.8 \end{bmatrix}, \quad Q = \begin{bmatrix} 0.3 & 0 \\ 0 & 0.2 \end{bmatrix}, \quad R = a\mathbb{I}_2$$

for two different choices of a : 0 and -0.5 . The constraint of R to be a multiple of the identity matrix is, indeed, a strict requirement for the derivation of Theorem 1 in Gauthier and Possamaï (2011). The results corresponding to $T = 1$ and $r = 3\%$ are shown in Figure 3. In the simple case of zero asset-volatility correlation (leftmost panel), all the methodologies considered work extremely well for any level of moneyness. Introducing a negative correlation, though, we notice a deterioration in the accuracy of the second-order expansion for deep in-the-money options (rightmost panel). On the other hand, the Heston and Bi-Heston approximations are still hardly distinguishable from the true values. Another advantage of the new approaches proposed is their extreme simplicity: we just need to price options in well-known affine models, while the formula in Gauthier and Possamaï (2011) requires to compute several integrals of (products of) exponential and hyperbolic functions of matrix argument. Additionally, if used to tackle the calibration problem, our techniques lead to the explicit computation of the gradient of the objective function. This topic will be further extended in the next section. Moving towards more challenging parametrizations, we investigate the dataset (24) coupled with matrix

$$R = \begin{bmatrix} -0.2243 & -0.1244 \\ -0.2545 & -0.7230 \end{bmatrix}$$

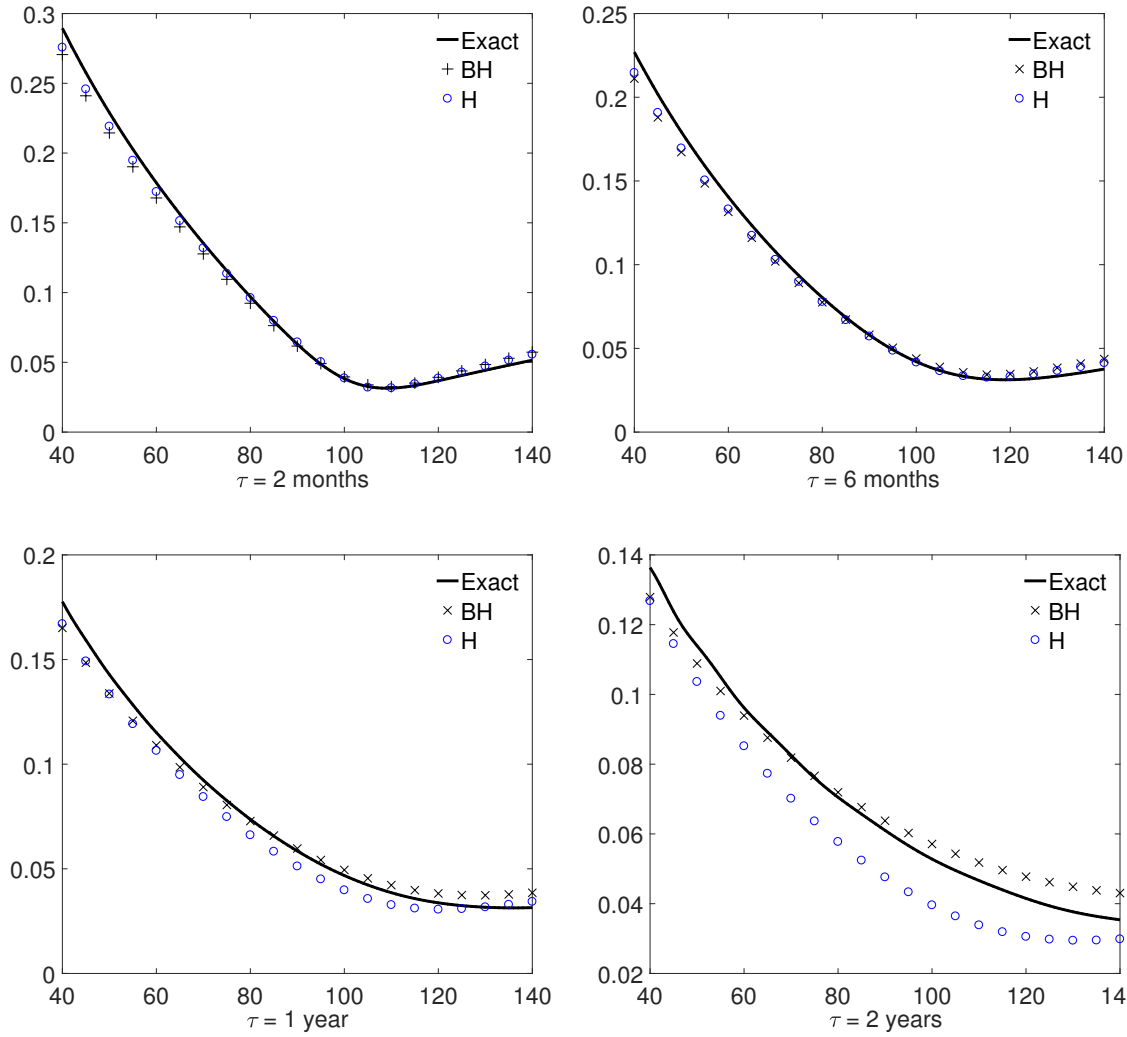


Figure 4.: Implied variance smiles for the WSVM (Exact), Bi-Heston (BH) and Heston (H) approximations for different time horizons.

as originally calibrated in Da Fonseca and Grasselli (2011). In this case, the approximation in Gauthier and Possamaï (2011) is no longer applicable since $R \neq a\mathbb{I}_2$. On the contrary, our methodologies do not suffer from this limitation and we report the results in Figure 4 for different time horizons. As we consider longer maturities, not surprisingly, we experience a worsening in the accuracy of both approximations. The (full) matrix structure of WSVM parameters is too complex to be entirely captured. Notwithstanding, the smile generated by the Bi-Heston approach is still reasonably in line with the WSVM values: in the worst-performing case ($T = 2$ years) the mean absolute error in volatility terms over the range of moneyness $[40\% - 140\%]$ is 0.9%.

3.2. Characteristic function and calibration to market prices

Given the affinity of WSVM, it is possible to express the characteristic function of log-prices $y(T)$ as the exponential of an affine combination of state variables $y(t)$ and $\Sigma(t)$. In particular the closed formula for the characteristic function is derived in Da Fonseca *et al.* (2008) where the resulting matrix Riccati equation is solved via linearization technique. Without entering into technical details (for which we refer the interested reader to the original paper (Da Fonseca *et al.* 2008)) we focus our attention on the formula as presented in the following proposition.

PROPOSITION 3.3 (*Characteristic function of log-price in WSVM* (Da Fonseca *et al.* 2008)) *Let the log-forward price $y(t)$ be described by (30) and λ be an auxiliary real variable (different from those used above). Then for $T > t$, the WSVM (conditional) characteristic function of $y(T)$ admits the following closed formula representation*

$$\phi_y^W(\lambda, \tau) = \mathbb{E}[\exp(\iota \lambda y(T)) | y(t)] = \exp(\iota \lambda y(t) + \text{Tr}[A_y(\tau) \Sigma(t)] + b_y(\tau)), \quad (38)$$

with the deterministic matrix $A_y(\tau)$ and the scalar function $b_y(\tau)$ given by

$$A_y(\tau) = A_{22}(\tau)^{-1} A_{21}(\tau), \quad b_y(\tau) = -\frac{\beta}{2} \text{Tr} \left[\log(A_{22}(\tau)) + \tau(M + \iota \lambda Q^\top R^\top) \right],$$

and

$$\begin{bmatrix} A_{11}(\tau) & A_{12}(\tau) \\ A_{21}(\tau) & A_{22}(\tau) \end{bmatrix} = \exp \left(\tau \begin{bmatrix} M + \iota \lambda Q^\top R^\top & -2Q^\top Q \\ \frac{\iota \lambda (\lambda - 1)}{2} \mathbb{I}_d & -(M + \iota \lambda Q^\top R^\top)^\top \end{bmatrix} \right).$$

Given the availability of a closed formula for the characteristic function of $y(T)$, we can price plain vanilla options through efficient numerical techniques. However, from a computational point of view, a direct application of (30) for calibration purposes could turn out to be highly cumbersome. Firstly because we are required to compute several functions of matrix argument each time, and secondly because the corresponding optimization problem would present a multiplicity of local minima and a strong dependence on the starting point. It is also important to point out that, as stated in Da Fonseca and Grasselli (2011), the use of existing approximations (both in terms of price and implied volatility) is likely to be restricted to narrow ranges around the at-the-money level. In the light of the above, then, efficient calibration of WSVM is still an open problem that could severely limit the real-world application of the model.

To overcome such limitations, we propose a fast and accurate calibration procedure that relies on the model approximations developed in Section 3.1. We start by illustrating the properties of the optimization problem. Let $C(K, T)$ be the market price of a call option struck at K with maturity in T years and $C^{Model}(\boldsymbol{\pi}, K, T)$ be the corresponding price obtained via the chosen model with (model-specific) parameters set $\boldsymbol{\pi}$. We formulate the calibration problem as an inverse problem of the form

$$\min_{\boldsymbol{\pi} \in \mathbb{R}^{N_\pi}} \frac{1}{2} f_{obj}(\boldsymbol{\pi}) \quad (39)$$

where N_π is the dimension of the parameters set and f_{obj} is the so-called objective function, i.e. the metric that defines the distance between market and model values. Several specifications for f_{obj} can be used and it is well understood that they can lead to quite different results. In our approach we follow Da Fonseca and Grasselli (2011) and define

$$f_{obj}(\boldsymbol{\pi}) = \sum_{s=1}^{N_s} \left(\frac{C^{Model}(\boldsymbol{\pi}, K_s, T_s) - C(K_s, T_s)}{\omega_s} \right)^2, \quad (40)$$

that is, we consider the sum of the square weighted difference between model and market values over a set of N_s quoted instruments. The weights ω_s can be chosen in order to put more emphasis on a certain subset of options (in terms, for example, of liquidity, moneyness or time to maturity). In the following we consider ω_s to be the inverse of (squared) Black-Scholes vega computed with respect to the s -th market option price. This choice of weights put more emphasis on short dated OTM options that otherwise would have almost no influence in the calibration procedure.

Defining the (weighted) residuals $\tilde{r}_s(\boldsymbol{\pi}) = (C^{Model}(\boldsymbol{\pi}, K_s, T_s) - C(K_s, T_s)) / \omega_s$, the calibration problem can be written as

$$\min_{\boldsymbol{\pi} \in \mathbb{R}^{N_\pi}} \frac{1}{2} f_{obj}(\boldsymbol{\pi}) = \min_{\boldsymbol{\pi} \in \mathbb{R}^{N_\pi}} \frac{1}{2} \tilde{\mathbf{r}}(\boldsymbol{\pi})^\top \tilde{\mathbf{r}}(\boldsymbol{\pi}) \quad (41)$$

with $\tilde{\mathbf{r}}(\boldsymbol{\pi}) \in \mathbb{R}^{N_s}$. Furthermore, let $J^{Model}(\boldsymbol{\pi}) = \nabla \tilde{\mathbf{r}}(\boldsymbol{\pi})^\top \in \mathbb{R}^{N_\pi \times N_s}$ be the Jacobian matrix of the residuals vector $\tilde{\mathbf{r}}(\boldsymbol{\pi})$ with elements

$$j_{r,s} = \frac{\partial}{\partial \pi_r} \tilde{r}_s(\boldsymbol{\pi}) = \frac{1}{\omega_s} \frac{\partial}{\partial \pi_r} C^{Model}(\boldsymbol{\pi}, K_s, T_s), \quad (42)$$

then the gradient of the objective function is given by $\nabla f_{obj} = J^{Model}(\boldsymbol{\pi}) \tilde{\mathbf{r}}(\boldsymbol{\pi})$. Usually, for many financial models the gradient of f_{obj} is computed numerically by, for example, finite differences, thus requiring a large number of function evaluations. There is, however, a noticeable exception: Cui *et al.* (2017) exploit an alternative representation of the Heston model characteristic function to obtain the analytical gradient of the price of a vanilla option with respect to model parameters (with our notation, this means that $J^H(\boldsymbol{\pi}_H)$ can be computed explicitly, where H stands for Heston model). The resulting calibration algorithm is extremely fast and robust.

We propose to combine the algorithm in Cui *et al.* (2017) with the model approximations developed in section 3.1. In other words, we transform, by parameters mapping, the WSVM calibration problem in a simpler one for which we are able to compute the gradient of the objective function in closed formula. The resulting procedure has two inherent advantages from a computational point of view: firstly we replace the WSVM characteristic function with a less computational demanding one; secondly we reduce the number of function evaluations (we do not need to approximate the gradient via finite differences). More importantly, by letting the exact gradient to drive the optimization routine, we are able to rapidly identify a suitable parameters set in the (high dimensional) search space.

We show how to compute the gradient of the objective function. Let us consider firstly the WSVM-Heston case: the (transformed or approximated) calibration problem can be written as

$$\min_{\boldsymbol{\pi}_W \in \mathbb{R}^{N_W}} \frac{1}{2} f_{obj}(\boldsymbol{\pi}_W) = \min_{\boldsymbol{\pi}_W \in \mathbb{R}^{N_W}} \frac{1}{2} \tilde{\mathbf{r}}(g^{H-W}(\boldsymbol{\pi}_W))^\top \tilde{\mathbf{r}}(g^{H-W}(\boldsymbol{\pi}_W)),$$

where, for simplicity, we suppress the time dependency.¹ By chain rule, we have that the Jacobian matrix of residuals with respect to WSVM parameters is given by

$$J^W(\boldsymbol{\pi}_W) = J^{H-W}(\boldsymbol{\pi}_W) J^H(g^{H-W}(\boldsymbol{\pi}_W)), \quad (43)$$

where J^H is known thanks to Cui *et al.* (2017). The matrix $J^{H-W}(\boldsymbol{\pi}_W) = \nabla g^{H-W}(\boldsymbol{\pi}_W) \in \mathbb{R}^{N_W \times 5}$ is the Jacobian matrix of function g^{H-W} with respect to WSVM parameters, with elements

$$j_{q,r}^{H-W} = \frac{\partial}{\partial \pi_{W,q}} g_r^{H-W}(\boldsymbol{\pi}_W), \quad (44)$$

that, as shown in Appendix A.2, can be computed in closed formula. Analogously to the general case, finally, it holds that $\nabla f_{obj} = J^W(\boldsymbol{\pi}_W) \tilde{\mathbf{r}}(g^{H-W}(\boldsymbol{\pi}_W))$.

Straightforwardly, we can apply the same methodology in conjunction with the Bi-Heston approximation. In Appendix A.3 we show how to extend the approach in Cui *et al.* (2017) to compute the

¹We remark, however, that the computation of function h must be performed for any maturity in the calibration basket.

Table 1.: Calibration on February, 3 2016 with the WSVM over a full set of DAX European call options.

Parameter	$\beta \geq 0$		$\beta \geq 1$	
	1 st step	2 nd step	1 st step	2 nd step
β	0.3612	0.3287	1.0403	1.0405
Σ_{11}	0.0639	0.0653	0.0789	0.0794
Σ_{12}	0	0.0105	0.0056	0.0038
Σ_{22}	0.0237	0.0213	0.0004	0.0003
M_{11}	-0.9879	-1.0793	-0.6561	-0.7020
M_{12}	0	-0.8468	0.0904	0.0893
M_{22}	-1.3541	-1.4760	-0.9001	-0.9895
Q_{11}	0.4639	0.4060	0.2647	0.2703
Q_{12}	0	0.1623	-0.0220	-0.0198
Q_{21}	0	0.4097	0.0351	0.0317
Q_{22}	0.4590	0.4763	0.0976	0.0879
R_{11}	-0.6618	-0.7280	-0.7267	-0.7056
R_{12}	0	-0.1718	-0.0082	-0.0090
R_{21}	0	0.6232	-0.0308	-0.0277
R_{22}	-0.6272	-0.5645	-0.5881	-0.5293
Error Price	4.52E-06	2.45E-06	6.34E-06	3.19E-06
Error Vol	6.16E-05	2.85E-05	2.32E-04	2.14E-04
Time (s)	1.93	24.33	3.01	13.42

As in Da Fonseca and Grasselli (2011) Error Price stands for the Mean Squared Error (MSE) in price normalized by the forward price and Error Vol denotes the MSE in implied volatility. The part $\beta \geq 0$ shows calibrated parameters for the case with no constraints on parameter β , while the part $\beta \geq 1$ reports values of parameters obtained imposing the condition of existence and uniqueness of a weak solution to SDE (1). Columns "1st step" refer to intermediate results obtained via Bi-Heston approximation, while columns "2nd step" report the final output of calibration routine computed via WSVM characteristic function.

gradient of Bi-Heston call options with respect to model parameters.²We then couple the matrix J^{BH} with the map g^{BH-W} to write, as in the previous case:

$$J^W(\pi_W) = J^{BH-W}(\pi_W)J^{BH}(g^{BH-W}(\pi_W, T)), \quad (45)$$

where, once more, J^{BH-W} can be obtained in closed formula (see Appendix A.4).

In the light of the evidences shown in the previous section, we decide to implement the Bi-Heston approximation when it comes to calibrate WSVM to market data. Nonetheless, we are confident that both methodologies can be helpful in understanding the role of WSVM parameters and performing models comparisons in the spirit of Da Fonseca and Grasselli (2011). We devise a 2 steps calibration procedure: initially we make use of WSVM-BiHeston mapping to tackle the simplified calibration problem and obtain a robust guess of the optimal parameters. In the second step, we consider this parameters set as the starting point of the standard calibration algorithm based on the pricing of vanilla options via WSVM characteristic function (38). This can be interpreted as the "fine-tuning" phase in order to further improve the accuracy achieved. By starting from a robust initial guess, we expect the algorithm to converge after few function evaluations. In this step, we implement the so-called COS method (Fang and Oosterlee 2008). This approach basically exploits the relation between the characteristic function and the coefficients of the Fourier cosine series expansion of the corresponding probability density function.

In Table 1 we test the proposed algorithm on the calibration of WSVM parameters to market prices of 182 European call options written on DAX index as of February 3, 2016. We here consider matrix

²As far as we know, this is the first time that such a result is derived. It provides an efficient tool to calibrate Bi-Heston parameters to market data.

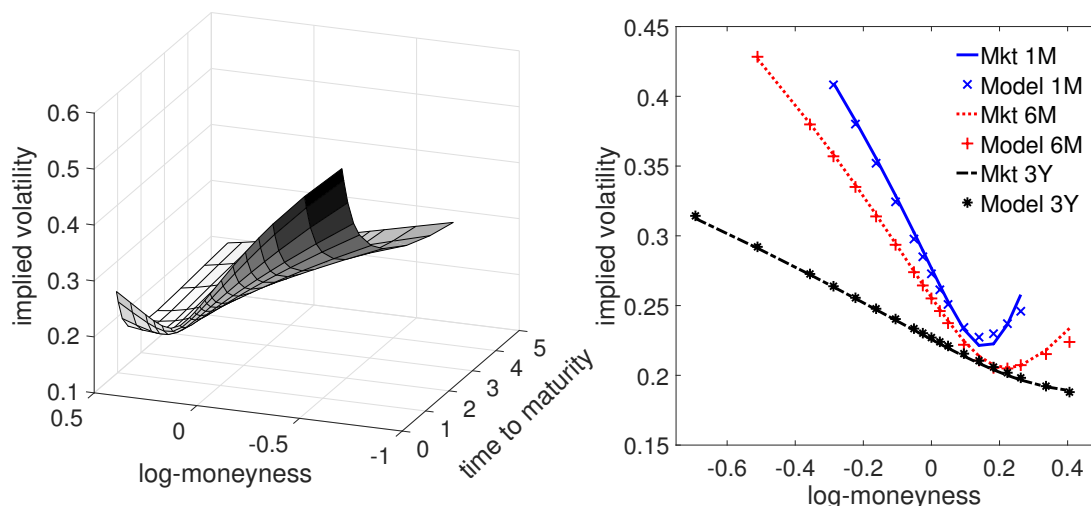


Figure 5.: Calibration results for WSVM. Resulting value for parameter β is 0.3287. Implied volatility surface obtained with the set of calibrated parameters (left) and comparison with market implied volatility for selected tenors (right).

M to be symmetric: this choice is motivated from the fact that the resulting WSVM model embeds an higher degree of tractability. As shown in Gauthier and Possamaï (2011), for example, in this setting it is possible to compute explicitly the expected value of integrated variance, thus opening the way to consistent pricing of volatility derivatives.

Calibrated parameters are shown in Table 1 - case $\beta > 0$. As we can see, the Bi-Heston step takes less than 2 seconds.¹ It is worthwhile noting that the resulting parameters matrices are all diagonal. This is, however, not the standard case. In the second step, the fine-tuning effect is quite relevant: the diagonal elements are very close to those found with the Bi-Heston mapping while the off-diagonal ones are set so that the accuracy is significantly improved (both mean squared errors in price and volatility terms are roughly halved). The efficiency of the overall procedure can be also inferred from the fact that the second step takes only 24.33 seconds. In comparison, the calibration procedure fully based on the WSVM characteristic function would take roughly 250 seconds. Figure 5 shows results of calibration procedure in terms of corresponding implied volatility surface and absolute error with respect to market values.

We see that apart from very short dated far from the money options the WSVM can fit the market surface quite well. According to the evidences in Da Fonseca and Grasselli (2011) the resulting value of parameter β is strictly lower than 1. In light of what we have shown in Proposition 2.1 the conditions for the existence and uniqueness of solution to SDE (1) are consequently not satisfied. This is a crucial restriction specifically when we use calibrated WSVM to price derivatives for which we need to rely on simulation methods (which is of course the main interest for a structuring team dealing with realistic applications). Even if we use an unbiased (or yet exact) method to simulate the Wishart process, we would not have any chance to consider the corresponding process as an approximation of the original SDE (1).

If our goal is to use WSVM in a Monte Carlo framework, then, we need to impose at least condition (7) that in the case $d = 2$ corresponds to $\beta \geq 1$. We then perform the calibration of WSVM over the same market set with the additional constraint on parameter β . As far as we know this is the first attempt to deal with such a constrained problem in a real market context. Once more, we implement the two steps procedure described above. The conclusions about the algorithm efficiency are the same: the combined approach significantly reduces the computational time required: the overall procedure takes now just 16 seconds. The mechanics is also very similar: the first step fastly

¹The algorithms are implemented via Matlab code on a laptop PC with an Intel Core i7 CPU and 8 GB RAM.

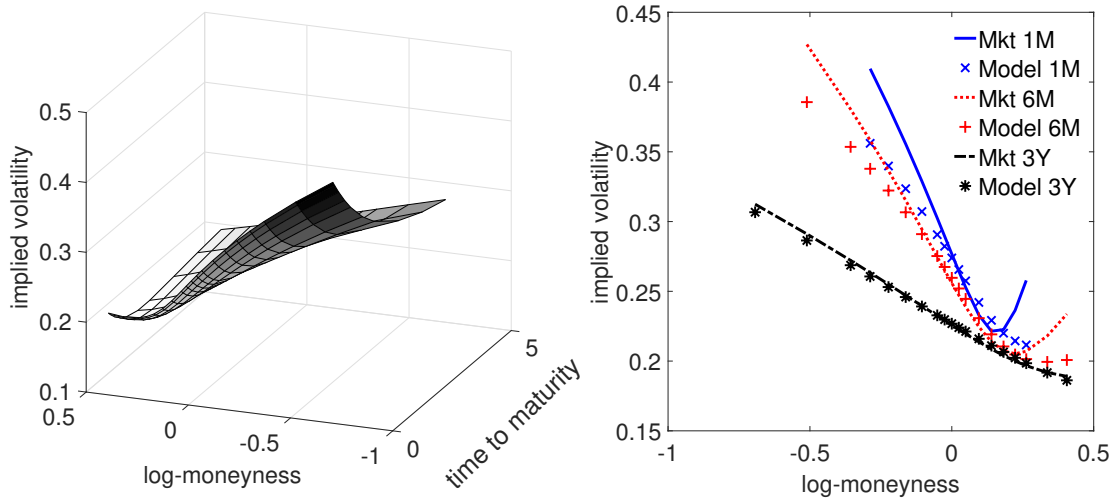


Figure 6.: Calibration results for WSVM when $\beta \geq 1$. Resulting value for parameter β is 1.0405. Implied volatility surface obtained with the set of calibrated parameters (left) and comparison with market implied volatility for selected tenors (right).

produce a robust guess while the final step act as a targeted adjustment devoted to the enhancement of calibration accuracy. Also in this case, indeed, the calibrated parameters are very close to the outcomes of the first step. The output of the calibration routine is shown in Figure 6. By a direct comparison with Figure 5 we can easily see that the additional constraint on parameter β has not negligible impact. In particular the accuracy for short dated far-from-the-money options is quite worsened.

4. New simulation schemes for the WSVM

In this section we propose two new simulation algorithms for the WSVM. The basic idea of our new schemes is to exploit the exact simulation method for the Wishart process developed in Ahdida and Alfonsi (2013) to sample the WSVM log-price process (25). From now on, indeed, we assume to have a collection of N simulated trajectories of the Wishart process over the time grid $0 = t_0 < t_1 < \dots < t_{M_T} = T$ obtained via the exact scheme in Ahdida and Alfonsi (2013). For simplicity, we also consider a uniform time step $\Delta = T/M_T = t_{m+1} - t_m$ for $m = 0, 1, \dots, M_T - 1$. The extension of such a scheme to the WSVM case is not however a trivial task since there is not an immediate way to reconstruct the correlation structure between the Wishart process and the asset price. We are not able then to substitute (26) into (25). This is a direct consequence of how the sampling procedure works: rather than simulating directly the desired Wishart process, the procedure is based on a sort of "bottom-up" approach. Basically the infinitesimal generator of a canonical Wishart process is found to be given by the sum of commuting operators associated to simple SDEs. Then any admissible Wishart process is linked to the canonical one through a law identity. This means that we need to simulate just the SDEs whose generators are linked to the one of the canonical Wishart process. As a consequence, we do not have a matrix of Gaussian variables \widehat{W} + used in (26) to discretize the trajectory of the asset price.

A method to circumvent this problem is proposed in Kang and Kang (2013) where the Wishart process is sampled exactly with the aforementioned technique and then the log-price is sampled from the conditional distribution $F_{y(T)|\Sigma(T)}(y; s) = \Pr[y(T) \leq y | \Sigma(T) = s]$ that is retrieved numerically from the conditional characteristic function of y . However, even if formally correct, this approach turns out to be unfeasible in our case and in general when the construction of an entire path of y is needed. It requires, indeed, the evaluation of special functions of matrix argument at each step

of the sampling scheme.

We instead start from the intuition that the dynamics of $\mathcal{V}(t)$ itself can be used to link (25) and (1) in a proper way. Let us combine (27) and (29) to rewrite the scalar dynamics of $y(t)$ as

$$dy(t) = -\frac{1}{2}\mathcal{V}(t)dt + \sqrt{\mathcal{V}(t)} \left(\rho_W(t)dw(t) + \sqrt{1 - \rho_W^2(t)}dz(t) \right) \quad (46)$$

with $z(t)$ a scalar Brownian motion independent on $w(t)$. This will be the starting point of the two new simulation schemes we propose.

4.1. Stochastic Integral Approximation

For the first approach, that we call the *Stochastic Integral Approximation* (SIA) scheme, we focus on the integral form of (46),

$$\begin{aligned} y(t + \Delta) = y(t) &- \frac{1}{2} \int_t^{t+\Delta} \mathcal{V}(s)ds + \int_t^{t+\Delta} \rho_W(s) \sqrt{\mathcal{V}(s)}dw(s) \\ &+ \int_t^{t+\Delta} \sqrt{(1 - \rho_W^2(s))\mathcal{V}(s)}dz(s). \end{aligned} \quad (47)$$

In the discretization of (47), the time integral that involves \mathcal{V} can be approximated as in Andersen (2008) by

$$\int_t^{t+\Delta} \mathcal{V}(s)ds \approx \Delta (\mu_1 \mathcal{V}(t) + \mu_2 \mathcal{V}(t + \Delta))$$

where we can set, for example, $\mu_1 = \mu_2 = 0.5$. By Itô's isometry we also have

$$\int_t^{t+\Delta} \sqrt{(1 - \rho_W^2(s))\mathcal{V}(s)}dz(s) \sim N \left(0, \int_t^{t+\Delta} (1 - \rho_W^2(s))\mathcal{V}(s)ds \right) \quad (48)$$

where the variance of the Gaussian random variable can be approximated as above. The only non-trivial term in (47) is the stochastic integral with respect to w . Unfortunately, there is no possible way to formulate this integral in terms of the dynamics of \mathcal{V} . This is due to the fact that the stochastic component in (28) involves the trace of a matrix product of $\Sigma(t)$ that cannot be decomposed as the product of the traces. However we can make use of the CIR approximation we devised in the previous section: if we write

$$d\mathcal{V}(t) \approx \kappa(\theta - \mathcal{V}(t))dt + \eta\sqrt{\mathcal{V}(t)}dw(t) \quad (49)$$

with parameters κ , θ and η given in Proposition 2.4, we obtain

$$\int_t^{t+\Delta} \sqrt{\mathcal{V}(s)}dw(s) \approx \frac{1}{\eta} \left(\mathcal{V}(t + \Delta) - \mathcal{V}(t) - \kappa\theta\Delta + \kappa \int_t^{t+\Delta} \mathcal{V}(s)ds \right). \quad (50)$$

Let $\widehat{\mathcal{V}}(t)$ and $\widehat{\mathcal{V}}(t + \Delta)$ be the realizations of the trace of Wishart process for two adjacent time points as sampled with the scheme devised in Ahdida and Alfonsi (2013), we propose the following

Algorithm 1 The *Stochastic Integral Approximation* (SIA) scheme for the WSVM

```

1: Set  $T > 0$ , compute parameters  $\kappa, \theta$  and  $\eta$  as in Proposition 2.4 conditionally on  $\Sigma$ 
   with  $\tau = T$ 
2: for each simulation trial  $n, n = 1, \dots, N$  do
3:   Initialize  $\widehat{\Sigma}(0) = \Sigma_0$ 
4:   Initialize  $\widehat{y}(0) = \log(f(0))$ 
5:   for each time-step  $t_m, m = 0, \dots, M_T - 1$  do
6:     Compute  $\widehat{\rho}_W(t_m)$  as given by (26)
7:     Sample the Wishart process for a time step of  $\Delta$  with initial state  $\widehat{\Sigma}(t_m)$ 
       using the scheme in Ahdida and Alfonsi (2013)
8:     Compute  $\widehat{\rho}_W(t_m + \Delta)$  as given by (26)
9:     Compute coefficients  $K_0(t_m), \dots, K_4(t_m)$  as in (53)-(57)
10:    Draw  $\widehat{z} \sim N(0, 1)$ 
11:    Discretize (47) by means of (52)
12:  end for
13: end for

```

discretization scheme for the log-price in WSVM:

$$\widehat{y}(t + \Delta) = \widehat{y}(t) + K_0(t) + K_1(t) \widehat{\mathcal{V}}(t) + K_2(t) \widehat{\mathcal{V}}(t + \Delta) \quad (51)$$

$$+ \sqrt{K_3(t) \widehat{\mathcal{V}}(t) + K_4(t) \widehat{\mathcal{V}}(t + \Delta)} \widehat{z} \quad (52)$$

with \widehat{z} a random number sampled from a $N(0, 1)$ and coefficients

$$K_0(t) = -\frac{\kappa\theta\Delta}{\eta} \widehat{\rho}_W(t), \quad (53)$$

$$K_1(t) = \mu_1\Delta \left(\frac{\kappa}{\eta} \widehat{\rho}_W(t) - \frac{1}{2} \right) - \frac{\widehat{\rho}_W(t)}{\eta}, \quad (54)$$

$$K_2(t) = \mu_2\Delta \left(\frac{\kappa}{\eta} \widehat{\rho}_W(t) - \frac{1}{2} \right) + \frac{\widehat{\rho}_W(t)}{\eta}, \quad (55)$$

$$K_3(t) = \mu_1\Delta (1 - \widehat{\rho}_W^2(t)), \quad (56)$$

$$K_4(t) = \mu_2\Delta (1 - \widehat{\rho}_W^2(t + \Delta)), \quad (57)$$

where we perform the *freezing* of $\rho_W(s)$ to write

$$\int_t^{t+\Delta} \rho_W(s) \sqrt{\mathcal{V}(s)} dw(s) \approx \rho_W(t) \int_t^{t+\Delta} \sqrt{\mathcal{V}(s)} dw(s). \quad (58)$$

This can be justified, as we will see, from the fact that the proposed scheme is mainly devoted to sampling the trajectories of y over fine time grids. The complete scheme is illustrated in Algorithm 1.

4.2. Gaussian Variable Approximation

We also present a simpler simulation technique, indicated as *Gaussian Variable Approximation* (GVA) scheme, for the discretization of (46). Given a discrete trajectory of \mathcal{V} , the Euler approxi-

Algorithm 2 The *Gaussian Variable Approximation* (GVA) scheme for the WSVM

```

1: for each simulation trial  $n$ ,  $n = 1, \dots, N$  do
2:   Initialize  $\hat{\Sigma}(0) = \Sigma_0$ 
3:   Initialize  $\hat{y}(0) = \log(f(0))$ 
4:   for each time-step  $t_m$ ,  $m = 0, \dots, M_T - 1$  do
5:     Compute  $\hat{\rho}_W(t_m)$  as given by (26)
6:     Sample the Wishart process for a time step of  $\Delta$  with initial state  $\hat{\Sigma}(t_m)$ 
       using the scheme in Ahdida and Alfonsi (2013)
7:     Approximate the Gaussian random variable  $\tilde{w}$  by means of (60)
8:     Draw  $\tilde{z} \sim N(0, \Delta)$ 
9:     Discretize (46) by means of (59)
10:  end for
11: end for

```

mation of (46) reads

$$\hat{y}(t + \Delta) = \hat{y}(t) - \frac{1}{2} \hat{\mathcal{V}}(t) \Delta + \hat{\rho}_W(t) \sqrt{\hat{\mathcal{V}}(t)} \tilde{w} + \sqrt{(1 - \hat{\rho}_W^2(t)) \hat{\mathcal{V}}(t)} \tilde{z} \quad (59)$$

where \tilde{w} is the Gaussian random variable that would drive the discretization of (28) and \tilde{z} is a Gaussian random variable with variance Δ and independent on \tilde{w} . By exploiting the sampling technique in Ahdida and Alfonsi (2013) to get $\hat{\Sigma}(t)$ and $\hat{\Sigma}(t + \Delta)$, we can use (28) to obtain an approximation of \tilde{w} . Indeed, it holds that

$$\hat{\mathcal{V}}(t + \Delta) - \hat{\mathcal{V}}(t) \approx \left(\text{Tr} [\beta Q^\top Q] + 2 \text{Tr} [M \hat{\Sigma}(t)] \right) \Delta + 2 \sqrt{\text{Tr} [\hat{\Sigma}(t) Q^\top Q]} \tilde{w}, \quad (60)$$

from which we can easily retrieve the value of \tilde{w} . Finally we can plug \tilde{w} into (59) to discretize the path of y . The complete procedure is summarized in Algorithm 2.

The resulting scheme is easy to implement and well suited for any application that requires to discretize the price process over a tiny time grid.¹ The two schemes proposed share the remarkable property that, thanks to the scalar representation, we can avoid the generation of $d^2 - 1$ additional random variables for the dynamics of y with an evident reduction of complexity burden.

4.3. Numerical comparison

We perform extensive numerical tests using the set of calibrated parameters obtained imposing $\beta \geq 1$ and reported in the rightmost column of Table 1. The exercise considered is the pricing of European call options with moneyness in the range $\{70\%, 100\%, 130\%\}$ and maturities varying from 6 months up to 3 years. Additional data: spot price equal to 100 and $r = q = 0$. In Tables 2-4 we report the prices of European call options obtained with the new simulation schemes proposed for the WSVM. We set the number of simulations to 2×10^5 and we exploit the antithetic variates as a variance reduction technique. Reference values are computed via COS method. The asterisk means that the corresponding reference value lies outside of the 95% confidence interval.

Despite the limited number of simulated trajectories, the proposed algorithms generate estimates in strict accordance with true prices. The only noticeable inconsistencies concern the SIA scheme for long-dated options ($T = 3$). We would like to stress that the SIA scheme exploits the CIR approximation (49), therefore it is not an exact simulation scheme.

¹As for example exotic options embedded in structured products with daily (or continuous) monitoring of underlying asset price.

Table 2.: SIA and GVA schemes, $T = 0.5$

Strike	Reference Value	No. of time steps	MC estimates	Confidence Interval (95%)	Error
SIA					
$K = 70$	30.6457	5	30.6724	30.6229 - 30.7218	0.09%
		10	30.6646	30.6149 - 30.7142	0.06%
		50	30.6662	30.6167 - 30.7158	0.07%
		100	30.6590	30.6094 - 30.7086	0.04%
$K = 100$	7.1533	5	7.1378	7.1130 - 7.1626	-0.22%
		10	7.1431	7.1182 - 7.1679	-0.14%
		50	7.1468	7.1220 - 7.1716	-0.09%
		100	7.1384	7.1135 - 7.1632	-0.21%
$K = 130$	0.1879	5	0.1934*	0.1885 - 0.1984	2.94%
		10	0.1932*	0.1883 - 0.1982	2.83%
		50	0.1928	0.1878 - 0.1976	2.63%
		100	0.1902	0.1853 - 0.1951	1.22%
GVA					
$K = 70$	30.6457	5	30.3723*	30.3245 - 30.4201	-0.89%
		10	30.4949*	30.4460 - 30.5438	-0.49%
		50	30.6256	30.5759 - 30.6752	-0.07%
		100	30.6425	30.5928 - 30.6922	-0.01%
$K = 100$	7.1533	5	6.9397*	6.9149 - 6.9646	-2.99%
		10	7.0432*	7.0182 - 7.0682	-1.54%
		50	7.1290	7.1039 - 7.1541	-0.34%
		100	7.1449	7.1198 - 7.1701	-0.12%
$K = 130$	0.1879	5	0.2211*	0.2158 - 0.2264	17.66%
		10	0.2069*	0.2018 - 0.2119	10.09%
		50	0.1955*	0.1906 - 0.2005	4.03%
		100	0.1919	0.1869 - 0.1968	2.12%

Table 3.: SIA and GVA schemes, $T = 1$

Strike	Reference Value	No. of time steps	MC estimates	Confidence Interval (95%)	Error
SIA					
$K = 70$	31.7060	5	31.6804	31.6207 - 31.7401	-0.08%
		10	31.6918	31.6319 - 31.7518	-0.04%
		100	31.7411	31.6811 - 31.8011	0.11%
		200	31.7176	31.6575 - 31.7776	0.04%
$K = 100$	9.5468	5	9.4650*	9.4318 - 9.4981	-0.86%
		10	9.4833*	9.4499 - 9.5166	-0.67%
		100	9.5265	9.4930 - 9.5598	-0.21%
		200	9.5170	9.4835 - 9.5504	-0.31%
$K = 130$	0.8632	5	0.8633	0.8508 - 0.8756	0.01%
		10	0.8600	0.8476 - 0.8724	-0.37%
		100	0.8670	0.8546 - 0.8794	0.44%
		200	0.8667	0.8542 - 0.8791	0.40%
GVA					
$K = 70$	31.7060	5	31.0741*	31.0163 - 31.1319	-1.99%
		10	31.4158*	31.3564 - 31.4753	-0.92%
		100	31.6605	31.6000 - 31.7209	-0.14%
		200	31.7077	31.6472 - 31.7682	0.01%
$K = 100$	9.5468	5	9.1286*	9.0950 - 9.1621	-4.38%
		10	9.3591*	9.3252 - 9.3929	-1.97%
		100	9.4983*	9.4644 - 9.5321	-0.51%
		200	9.5413	9.5074 - 9.5751	-0.06%
$K = 130$	0.8632	5	0.9294*	0.9163 - 0.9425	7.66%
		10	0.9113*	0.8984 - 0.9241	5.57%
		100	0.8677	0.8553 - 0.8801	0.52%
		200	0.8726	0.8602 - 0.8850	1.09%

By comparing the accuracy of the schemes, two facts clearly arise: when we consider coarse time grids, the SIA scheme significantly outperforms the GVA scheme. This is particularly true for shorter time horizons (i.e., the 6 months and 1 year case). On the other hand, as we shrink the temporal mesh size, the latter converges faster to the true price, especially for in-the-money and at-the-money options. In the lights of these evidences, the choice of the optimal scheme should be made by taking

Table 4.: SIA and GVA schemes, $T = 3$

Strike	Reference Value	No. of time steps	MC estimates	Confidence Interval (95%)	Error
SIA					
$K = 70$	34.8315	10	34.6550*	34.5661 - 34.7438	-0.51%
		50	34.6935*	34.6042 - 34.7827	-0.40%
		100	34.6752*	34.5858 - 34.7645	-0.45%
		300	34.7400*	34.6505 - 34.8294	-0.26%
$K = 100$	15.5618	10	15.3455*	15.2837 - 15.4074	-1.39%
		50	15.3723*	15.3103 - 15.4342	-1.22%
		100	15.3645*	15.3026 - 15.4264	-1.27%
		300	15.4141*	15.3521 - 15.4761	-0.95%
$K = 130$	5.0151	10	4.8924*	4.8531 - 4.9317	-2.45%
		50	4.8928*	4.8538 - 4.9319	-2.44%
		100	4.8828*	4.8439 - 4.9217	-2.64%
		300	4.9053*	4.8663 - 4.9442	-2.19%
GVA					
$K = 70$	34.8315	10	34.3504*	34.2595 - 34.4414	-1.38%
		50	34.7785	34.6872 - 34.8697	-0.15%
		100	34.7414	34.6502 - 34.8326	-0.26%
		300	34.8476	34.7564 - 34.9387	0.05%
$K = 100$	15.5618	10	15.2894*	15.2255 - 15.3534	-1.75%
		50	15.5349	15.4713 - 15.5986	-0.17%
		100	15.5441	15.4809 - 15.6073	-0.11%
		300	15.5508	15.4875 - 15.6141	-0.07%
$K = 130$	5.0151	10	5.0217	4.9814 - 5.0619	0.13%
		50	5.0407	5.0009 - 5.0804	0.51%
		100	4.9960	4.9567 - 5.0353	-0.38%
		300	5.0144	4.9750 - 5.0538	-0.01%

Table 5.: Average computational time as function of the number of time steps.

Time steps	5	10	50	100	200	300
SIA (s)	1.02	1.96	10.45	21.01	42.24	63.77
GVA (s)	0.79	1.57	8.50	16.99	35.01	51.39

The number of simulated paths is fixed to 2×10^5 .

into account also the designated time framework (both in terms of overall horizon and step width). From a computational point of view, both schemes allow to deal with the discretization of WSVM asset price very efficiently: as shown in Table 5, for example, we are able to simulate 2×10^5 price trajectories with a time step of $T/100$ in no more than 21 seconds.² With such a time frame we get an error lower than 1% in all but 4 cases. A comparison between the two rows of Table 5 also reveals that the SIA scheme is, on average, 20% – 30% slower than the GVA one.

To conclude, both SIA and GVA schemes turn out to be efficient, fast, and simple to implement. Thanks to the underlying Wishart process sampling Ahdida and Alfonsi (2013), moreover, they do not pose any restriction on variance process parameters. Therefore these simulation schemes can be consider to price, as an example, European and exotic derivatives via Monte Carlo simulation in the WSVM framework.

5. Concluding Remarks

The matrix structure of Wishart-based stochastic volatility models provides a remarkable degree of flexibility in describing the evolution of asset(s) volatility. Realistic implementations, though,

²All tests have been carried out on a laptop PC with an Intel Core i7 CPU and 8 GB RAM. Algorithms are written in Matlab code and then compiled as MEX files to achieve better performances.

require the development of specific numerical techniques in order to deal with the inherent level of complexity. In this article we have shown, leveraging on a thorough analysis of distributional properties of Wishart process, some possible solutions intended to make this class of model more suitable for real market applications. Accordingly, we hope that our contribution will increase the interest of researchers and practitioners towards matrix-variate stochastic volatility dynamics.

References

- Ahdida, A. and Alfonsi, A., Exact and high-order discretization schemes for Wishart processes and their affine extensions. *The Annals of Applied Probability*, 2013, **23**, 1025–1073.
- Andersen, L.B.G., Efficient simulation of the Heston stochastic volatility model. *Journal of Computational Finance*, 2008, **11**, 1–42.
- Bru, M., Wishart processes. *Journal of Theoretical Probability*, 1991, **4**, 725–751.
- Christoffersen, P., Heston, S. and Jacobs, K., The shape and term structure of the index option smirk: Why multifactor stochastic volatility models work so well. *Management Science*, 2009, **55**, 1914–1932.
- Cont, R. and Da Fonseca, J., Dynamics of implied volatility surfaces. *Quantitative Finance*, 2002, **2**, 45–60.
- Cox, J., Ingersoll, J. and Ross, S., A theory of the term structure of interest rates. *Econometrica: Journal of the Econometric Society*, 1985, pp. 385–407.
- Cuchiero, C., Filipovic, D., Mayerhofer, E. and Teichmann, J., Affine processes on positive semidefinite matrices. *Annals of Applied Probability*, 2011, **21**, 397–463.
- Cui, Y., Rollin, S.d.B.n. and Germano, G., Full and fast calibration of the Heston stochastic volatility model. *European Journal of Operational Research*, 2017, **263**, 625–638.
- Da Fonseca, J. and Grasselli, M., Riding on the smiles. *Quantitative Finance*, 2011, **11**, 1609–1632.
- Da Fonseca, J., Grasselli, M. and Tebaldi, C., Option pricing when correlations are stochastic: an analytical framework. *Review of Derivatives Research*, 2007, **10**, 151–180.
- Da Fonseca, J., Grasselli, M. and Tebaldi, C., A multifactor volatility Heston model. *Quantitative Finance*, 2008, **8**, 591–604.
- Duffie, D., Filipović, D. and Schachermayer, W., Affine processes and applications in finance. *Annals of Applied Probability*, 2003, **13**, 984–1053.
- Fang, F. and Oosterlee, C., A novel pricing method for European options based on Fourier-cosine series expansions. *SIAM Journal on Scientific Computing*, 2008, **31**, 826–848.
- Gauthier, P. and Possamaï, D., Prices expansion in the Wishart model. *The IUP Journal of Computational Mathematics*, 2011, **4**, 44–71.
- Gauthier, P. and Possamaï, D., Efficient Simulation of the Wishart Model. *The IUP Journal of Computational Mathematics*, 2012, **5**, 14–58.
- Gnoatto, A. and Grasselli, M., The explicit Laplace transform for the Wishart process. *Journal of Applied Probability*, 2014, **51**, 640–656.
- Grasselli, M. and Tebaldi, C., Solvable affine term structure models. *Mathematical Finance*, 2008, **18**, 135–153.
- Heston, S., A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, 1993, **6**, 327–343.
- Imhof, J., Computing the distribution of quadratic forms in normal variables. *Biometrika*, 1961, **48**, 419–426.
- Kang, C. and Kang, W., Exact Simulation of Wishart Multidimensional Stochastic Volatility Model. *preprint arXiv:1309.0557*, 2013.
- Kotz, S., Johnson, N. and Boyd, D., Series representations of distributions of quadratic forms in normal variables II. Non-central case. *The Annals of Mathematical Statistics*, 1967, **38**, 838–848.
- Kourouklis, S. and Moschopoulos, P., On the distribution of the trace of a noncentral Wishart matrix. *Metron*, 1985, **43**, 85–92.
- Liu, H., Tang, Y. and Zhang, H., A new chi-square approximation to the distribution of non-negative definite quadratic forms in non-central normal variables. *Computational Statistics & Data Analysis*, 2009, **53**, 853–856.
- Mayerhofer, E., Pfaffel, O. and Stelzer, R., On strong solutions for positive definite jump diffusions. *Stochastic Processes and their Applications*, 2011, **121**, 2072–2086.
- Van Loan, C., Computing integrals involving the matrix exponential. *IEEE Transactions on Automatic Control*, 1978, **23**, 395–404.

Wang, S., Kuo, T. and Hsu, C., Trace bounds on the solution of the algebraic matrix Riccati and Lyapunov equation. *IEEE Transactions on Automatic Control*, 1986, **31**, 654–656.

Wilcox, R.M., Exponential operators and parameter differentiation in quantum physics. *Journal of Mathematical Physics*, 1967, **8**, 962–982.

Appendix A: Heston and Bi-Heston approximation of WSVM

In this section we present the auxiliary results needed to obtain the (semi-)analytic formulation of the gradient of call option prices when the WSVM is approximated by the Heston or the Bi-Heston model.

A.1. The Heston model case: the approach of Cui et al. Cui et al. (2017)

The characteristic function of log-asset in the Heston model with parameters set $\boldsymbol{\pi}_H = [v_0, \kappa, \theta, \sigma, \rho]^\top$ is known to be (as originally derived in Heston (1993))

$$\begin{aligned} \phi_y^H(\lambda, T) = \exp & \left(\iota \lambda y(0) + \frac{\kappa \theta}{\sigma^2} \left((e_H + d_H) T - 2 \log \left(\frac{1 - l_H e^{d_H T}}{1 - l_H} \right) \right) \right. \\ & \left. + \frac{v_0}{\sigma^2} (e_H + d_H) \frac{1 - e^{d_H T}}{1 - l_H e^{d_H T}} \right), \end{aligned} \quad (\text{A1})$$

where we set, for simplicity, interest rates and dividends to zero. Further, we have $y(0) = \log(f(0))$ and

$$e_H = \kappa - \sigma \rho \iota \lambda, \quad d_H = \sqrt{e_H^2 + \sigma^2(\lambda^2 + \iota \lambda)}, \quad l_H = \frac{e_H + d_H}{e_H - d_H}. \quad (\text{A2})$$

In Cui et al. (2017), the authors show that (A1) can be written as

$$\phi_y^H(\lambda, T) = \exp \left(\iota \lambda y(0) - \frac{\kappa \theta \rho}{\sigma} \iota \lambda T - v_0 A + \frac{2 \kappa \theta}{\sigma^2} D \right) \quad (\text{A3})$$

with

$$\begin{aligned} A &= \frac{A_1}{A_2}, \quad A_1 = (\lambda^2 + \iota \lambda) \sinh \left(\frac{d_H T}{2} \right), \quad A_2 = d_H \cosh \left(\frac{d_H T}{2} \right) + e_T \sinh \left(\frac{d_H T}{2} \right), \\ B &= \frac{d_H e^{\frac{\kappa T}{2}}}{A_2}, \quad D = \log(B). \end{aligned}$$

The main advantage of the new representation, apart from the improved stability and lack of discontinuities with respect to alternative formulations, is that it is easily differentiable. In particular, it is possible to compute (semi-)analytically the derivatives of call options model prices $C^H(\boldsymbol{\pi}_H, K, T)$ with respect to model parameters. Let $\nabla C^H(\boldsymbol{\pi}_H, K, T)$ be the gradient of $C^H(\boldsymbol{\pi}_H, K, T)$ with respect to $\boldsymbol{\pi}_H$, then the following holds

$$\nabla C^H(\boldsymbol{\pi}_H, K, T) = \frac{1}{\pi} \left(\int_0^\infty \Re \left(\frac{K^{-\iota \lambda}}{\iota \lambda} \nabla \phi_y^H(\lambda - \iota, T) \right) d\lambda - K \int_0^\infty \Re \left(\frac{K^{-\iota \lambda}}{\iota \lambda} \nabla \phi_y^H(\lambda, T) \right) d\lambda \right) \quad (\text{A4})$$

where $\nabla \phi_y^H(\lambda, T) = \phi_y^H(\lambda, T) \mathbf{h}(\lambda)$, $\mathbf{h}(\lambda) = [h_1(\lambda), \dots, h_5(\lambda)]^\top$ with elements

$$h_1(\lambda) = -A, \quad (\text{A5})$$

$$h_2(\lambda) = \frac{v_0}{\sigma \iota \lambda} \frac{\partial A}{\partial \rho} + \frac{2\theta}{\sigma^2} D + \frac{2\kappa\theta}{\sigma^2 B} \frac{\partial B}{\partial \kappa} - \frac{\theta \rho \iota \lambda T}{\sigma}, \quad (\text{A6})$$

$$h_3(\lambda) = \frac{2\kappa}{\sigma^2} D - \frac{\kappa \rho \iota \lambda T}{\sigma}, \quad (\text{A7})$$

$$h_4(\lambda) = -v_0 \frac{\partial A}{\partial \sigma} - \frac{4\kappa\theta}{\sigma^3} D + \frac{2\kappa\theta}{\sigma^2 d_H} \left(\frac{\partial d_H}{\partial \sigma} - \frac{d_H}{A_2} \frac{\partial A_2}{\partial \sigma} \right) + \frac{\kappa \theta \rho \iota \lambda T}{\sigma^2}, \quad (\text{A8})$$

$$h_5(\lambda) = -v_0 \frac{\partial A}{\partial \rho} + \frac{2\kappa\theta}{\sigma^2 d_H} \left(\frac{\partial d_H}{\partial \rho} - \frac{d_H}{A_2} \frac{\partial A_2}{\partial \rho} \right) - \frac{\kappa \theta \iota \lambda T}{\sigma}. \quad (\text{A9})$$

The derivatives appearing in (A5)-(A9) can be easily obtained from the definition of involved quantities. We refer to the original paper (Cui *et al.* 2017) for their exact formulation.

A.2. Jacobian matrix of mapping between WSVM and Heston parameters

Let $g^{H-W}(\boldsymbol{\pi}_W, T)$ be the mapping presented in Section 3.1 between the WSVM and Heston parameters. Without loss of generality we consider $d = 2$ and assume matrix M to be symmetric. For a fixed $T > 0$ we have (from Proposition 2.4 and (32)):

$$v_0 = \text{Tr}[\Sigma_0], \quad \theta = -\frac{\beta}{2} \text{Tr}[M^{-1}Q^\top Q], \quad \kappa = -\frac{1}{\tau} \log \left(\frac{\mathbb{E}[\mathcal{V}(T)|\Sigma(t)] - \theta}{\mathcal{V}(t) - \theta} \right),$$

$$\eta = \sqrt{\frac{\text{Var}[\mathcal{V}(T)|\Sigma(t)]}{a\kappa((1-a)\mathcal{V}(t) + \frac{a}{2}\theta)}}, \quad \rho = \frac{\text{Tr}[RQ\Sigma_0]}{\sqrt{\text{Tr}[\Sigma_0]}\sqrt{\text{Tr}[Q^\top Q\Sigma_0]}},$$

where we simplify the formula for θ by exploiting the cyclic property of the trace, the symmetry of M and the fact that $\frac{1}{\alpha} M^{-1} e^{\alpha s M}$ is a primitive of $e^{\alpha s M}$.

We now show how to compute the elements of matrix $J^{H-W}(\boldsymbol{\pi}_W)$ defined by (44). Firstly, we recall a useful Lemma about the differentiation of matrix exponentials with respect to scalar parameters:

LEMMA A.1 (Wilcox's formula (Wilcox 1967)) *Let $\mathcal{A} = (\alpha_{i,j})$ be a square matrix of arbitrary order whose elements are functions of (at least) one scalar parameter p . It holds that*

$$\frac{\partial}{\partial p} \exp(\tau \mathcal{A}) = \int_0^\tau \exp((\tau - u)\mathcal{A}) \mathcal{A}'_p \exp(u\mathcal{A}) du \quad (\text{A10})$$

where $\mathcal{A}'_p = \left(\frac{\partial \alpha_{i,j}}{\partial p} \right)$.

From Van Loan (1978), we know that integrals like (A10) can be computed via matrix exponentiation: let \mathcal{C} be the block matrix $\mathcal{C} = \begin{pmatrix} \mathcal{A} & \mathcal{A}'_p \\ 0 & \mathcal{A} \end{pmatrix}$ and $\mathcal{D} = \exp(\tau \mathcal{C}) = \begin{pmatrix} D_{1,1} & D_{1,2} \\ 0 & D_{2,2} \end{pmatrix}$, then integral in (A10) is given by $D_{1,2}$.

Let us now consider the matrices $A_W = \begin{pmatrix} -M & Q^\top Q \\ 0 & M^\top \end{pmatrix}$ and $B_W = \exp(\tau A_W) = \begin{pmatrix} B_{1,1} & B_{1,2} \\ 0 & B_{2,2} \end{pmatrix}$. From Van Loan (1978), it holds that $\Gamma(\tau) = B_{2,2}^\top \Sigma B_{2,2}$ and $\Theta(\tau) = B_{2,2}^\top B_{1,2}$. For the q -th parameter in $\boldsymbol{\pi}_W$ we set $\mathcal{C}_W = \begin{pmatrix} A_W & \frac{\partial A_W}{\partial \pi_{W,q}} \\ 0 & A_W \end{pmatrix}$ and $\mathcal{D}_W = \exp(\tau \mathcal{C}_W) = \begin{pmatrix} D_{1,1}^W & D_{1,2}^W \\ 0 & D_{2,2}^W \end{pmatrix}$, then from (A10) and Van Loan

(1978), we have that

$$\frac{\partial}{\partial p} \exp(\tau A_W) = D_{1,2}^W = \begin{pmatrix} \frac{\partial B_{1,1}}{\partial \pi_{W,q}} & \frac{\partial B_{1,2}}{\partial \pi_{W,q}} \\ 0 & \frac{\partial B_{2,2}}{\partial \pi_{W,q}} \end{pmatrix}. \quad (\text{A11})$$

Let $e_{i,j} = (\mathbb{1}_{k=i,\ell=j})_{1 \leq k,\ell \leq d}$ for $1 \leq i, j \leq d$ form the standard basis of $\mathcal{M}_d(\mathbb{R})$. We can finally write

$$\begin{aligned} \frac{\partial \Gamma(\tau)}{\partial \Sigma_{i,j}} &= B_{2,2}^\top e_{i,j} B_{2,2}, \quad \frac{\partial \Gamma(\tau)}{\partial M_{i,j}} = \frac{\partial B_{2,2}^\top}{\partial M_{i,j}} \Sigma B_{2,2} + B_{2,2}^\top \Sigma \frac{\partial B_{2,2}}{\partial M_{i,j}}, \\ \frac{\partial \Theta(\tau)}{\partial M_{i,j}} &= \frac{\partial B_{2,2}^\top}{\partial M_{i,j}} B_{1,2} + B_{2,2}^\top \frac{\partial B_{1,2}}{\partial M_{i,j}}, \quad \frac{\partial \Theta(\tau)}{\partial Q_{i,j}} = B_{2,2}^\top \frac{\partial B_{1,2}}{\partial Q_{i,j}}, \end{aligned}$$

since $B_{2,2}^\top = \exp(\tau M^\top)$.

Let δ_{ij} be the Kronecker's delta, after some algebraic manipulations, the elements of $J^{H-W}(\pi_W)$ can be computed as follows:

$$\begin{aligned} \frac{\partial v_0}{\partial \Sigma_{i,j}} &= \delta_{i,j}, \quad \frac{\partial \theta}{\partial \beta} = -\frac{1}{2} \text{Tr} [M^{-1} Q^\top Q], \quad \frac{\partial \theta}{\partial M_{i,j}} = \frac{\beta}{2} \text{Tr} [M^{-1} e_{i,j} M^{-1} Q^\top Q], \\ \frac{\partial \theta}{\partial Q_{i,j}} &= -\frac{\beta}{2} \text{Tr} [M^{-1} (e_{j,i} Q + Q^\top e_{i,j})], \quad \frac{\partial \kappa}{\partial \beta} = -\frac{d_0 \text{Tr} [\Theta(\tau)] + d_1 \frac{\partial \theta}{\partial \beta}}{d_0 d_2 \tau}, \quad \frac{\partial \kappa}{\partial \Sigma_{i,j}} = -\frac{d_0 \text{Tr} [\frac{\partial \Gamma(\tau)}{\partial \Sigma_{i,j}}] - d_2 \delta_{i,j}}{d_0 d_2 \tau}, \\ \frac{\partial \kappa}{\partial M_{i,j}} &= -\frac{d_0 \text{Tr} [\frac{\partial \Gamma(\tau)}{\partial M_{i,j}}] + \beta \frac{\partial \Theta(\tau)}{\partial M_{i,j}} + d_1 \frac{\partial \theta}{\partial M_{i,j}}}{d_0 d_2 \tau}, \quad \frac{\partial \kappa}{\partial Q_{i,j}} = -\frac{\beta d_0 \text{Tr} [\frac{\partial \Theta(\tau)}{\partial Q_{i,j}}] + d_1 \frac{\partial \theta}{\partial Q_{i,j}}}{d_0 d_2 \tau}, \\ \frac{\partial \eta}{\partial \beta} &= \frac{2\bar{a} \text{Tr} [\Theta^2(\tau)] - \text{Var} [\mathcal{V}(T)|\Sigma(t)] \frac{\partial \bar{a}}{\partial \beta}}{2\eta \bar{a}^2}, \quad \frac{\partial \eta}{\partial \Sigma_{i,j}} = \frac{4\bar{a} \text{Tr} [\frac{\partial \Gamma(\tau)}{\partial \Sigma_{i,j}} \Theta(\tau)] - \text{Var} [\mathcal{V}(T)|\Sigma(t)] \frac{\partial \bar{a}}{\partial \Sigma_{i,j}}}{2\eta \bar{a}^2}, \\ \frac{\partial \eta}{\partial M_{i,j}} &= \frac{4\bar{a} \text{Tr} [(\Gamma(\tau) + \beta \Theta(\tau)) \frac{\partial \Gamma(\tau)}{\partial M_{i,j}} + \Theta(\tau) \frac{\partial \Gamma(\tau)}{\partial M_{i,j}}] - \text{Var} [\mathcal{V}(T)|\Sigma(t)] \frac{\partial \bar{a}}{\partial M_{i,j}}}{2\eta \bar{a}^2}, \\ \frac{\partial \eta}{\partial Q_{i,j}} &= \frac{4\bar{a} \text{Tr} [(\Gamma(\tau) + \beta \Theta(\tau)) \frac{\partial \Gamma(\tau)}{\partial Q_{i,j}}] - \text{Var} [\mathcal{V}(T)|\Sigma(t)] \frac{\partial \bar{a}}{\partial Q_{i,j}}}{2\eta \bar{a}^2}, \\ \frac{\partial \rho}{\partial \Sigma_{i,j}} &= \left(\frac{\text{Tr} [RQ e_{i,j}]}{\text{Tr} [RQ \Sigma]} - \frac{\text{Tr} [\Sigma] \text{Tr} [Q^\top Q e_{i,j}] + \text{Tr} [Q^\top Q \Sigma] \delta_{i,j}}{2 \text{Tr} [\Sigma] \text{Tr} [Q^\top Q \Sigma]} \right) \rho, \\ \frac{\partial \rho}{\partial Q_{i,j}} &= \frac{\text{Tr} [R e_{i,j} \Sigma]}{\text{Tr} [\Sigma] \text{Tr} [Q^\top Q \Sigma]} - \frac{\text{Tr} [RQ \Sigma] \text{Tr} [Q^\top e_{i,j} \Sigma]}{\sqrt{\text{Tr} [\Sigma]} \sqrt{\text{Tr} [Q^\top Q \Sigma]}^3} \frac{\partial \rho}{\partial R_{i,j}} = \frac{\text{Tr} [e_{i,j} Q \Sigma]}{\text{Tr} [RQ \Sigma]} \rho, \end{aligned}$$

with

$$d_0 = \mathcal{V}(0) - \theta, \quad d_1 = \mathbb{E} [\mathcal{V}(T)|\Sigma] - \mathcal{V}(0), \quad d_2 = \mathbb{E} [\mathcal{V}(T)|\Sigma] - \theta, \quad \bar{a} = \frac{a ((1-a)\mathcal{V}(t) + \frac{a}{2}\theta)}{\kappa}.$$

All the missing elements of $J^{H-W}(\pi_W)$ are equal to zero.

A.3. The Bi-Heston model case: new formulation of log-asset characteristic function

We now extend the approach in Cui *et al.* (2017) to the Bi-Heston model (Christoffersen *et al.* 2009) with parameters $\boldsymbol{\pi}_{BH} = [\boldsymbol{\pi}_1, \boldsymbol{\pi}_2]^\top$ and $\boldsymbol{\pi}_i = [v_{0,i}, \kappa_i, \theta_i, \sigma_i, \rho_i]$ for $i = 1, 2$. We simply need to recognize that the Bi-Heston characteristic function corresponds to

$$\begin{aligned} \phi_y^{BH}(\lambda, T) = \exp & \left(\iota \lambda y(0) + \sum_{i=1}^2 \frac{\kappa_i \theta_i}{\sigma_i^2} \left((e_{H,i} + d_{H,i})T - 2 \log \left(\frac{1 - l_{H,i} e^{d_{H,i}T}}{1 - l_{H,i}} \right) \right) + \right. \\ & \left. \sum_{i=1}^2 \frac{v_{0,i}}{\sigma_i^2} (e_{H,i} + d_{H,i}) \frac{1 - e^{d_{H,i}T}}{1 - l_{H,i} e^{d_{H,i}T}} \right) \end{aligned} \quad (\text{A12})$$

where the terms in (A12) are defined in (A2) provided that we use the corresponding i -th subset $\boldsymbol{\pi}_i$. Similarly to the Heston model, then, we can write ϕ_y^{BH} as following:

$$\phi_y^{BH}(\lambda, T) = \exp \left(\iota \lambda y(0) - \sum_{i=1}^2 \frac{\kappa_i \theta_i \rho_i}{\sigma_i} \iota \lambda T - v_{0,i} A_i + \frac{2 \kappa_i \theta_i}{\sigma_i^2} D_i \right) \quad (\text{A13})$$

where, once more, all the terms involved are defined as in the Heston case for a proper choice of $\boldsymbol{\pi}_i$. The last step, namely the computation of $\nabla C^{BH}(\boldsymbol{\pi}_{BH}, K, T)$, follows straightforwardly by noting that each component of the summation in (A13) does not depend on subset $\boldsymbol{\pi}_j$ for $j \neq i$. Following Cui *et al.* (2017), we have

$$\begin{aligned} \nabla C^{BH}(\boldsymbol{\pi}_{BH}, K, T) = & \frac{1}{\pi} \left(\int_0^\infty \Re \left(\frac{K^{-\iota \lambda}}{\iota \lambda} \nabla \phi_y^{BH}(\lambda - \iota, T) \right) d\lambda \right. \\ & \left. - K \int_0^\infty \Re \left(\frac{K^{-\iota \lambda}}{\iota \lambda} \nabla \phi_y^{BH}(\lambda, T) \right) d\lambda \right) \end{aligned} \quad (\text{A14})$$

where $\nabla \phi_y^{BH}(\lambda, T) = \phi_y^{BH}(\lambda, T) \mathbf{h}(\lambda)$, $\mathbf{h}(\lambda) = [\mathbf{h}_1(\lambda), \mathbf{h}_2(\lambda)]^\top$, with $\mathbf{h}_i(\lambda) = [h_{1,i}(\lambda), \dots, h_{5,i}(\lambda)]$ defined in (A5)-(A9) for $i = 1, 2$.

A.4. Jacobian matrix of mapping between WSVM and Bi-Heston parameters

Let $g^{BH-W}(\boldsymbol{\pi}_W, T)$ be the mapping presented in presented in Section 3.1 between the WSVM and Bi-Heston parameters. Here we show how to compute explicitly the elements of matrix $J^{BH-W}(\boldsymbol{\pi}_W)$.

As done for the Heston case, we consider $d = 2$ and assume matrix M to be symmetric. For a fixed T , Bi-Heston parameters are given in (35)-(37). Further, let \mathcal{Q} be the orthogonal matrix that diagonalizes $\Theta(T)$, that is $\mathcal{Q}^\top \Theta(T) \mathcal{Q} = \mathcal{E} = \text{diag}[\varepsilon_1, \varepsilon_2]$ with $\varepsilon_1 \geq \varepsilon_2$ (i.e. with eigenvalues sorted

in descending order). By direct computation the following holds for $i, j, k = 1, 2$:

$$\begin{aligned}
\frac{\partial \zeta_i}{\partial \Sigma_{j,k}} &= \text{Tr} \left[\mathcal{Q}^\top \frac{\partial \Gamma(\tau)}{\partial \Sigma_{j,k}} \mathcal{Q} e_{i,i} \right], \quad \frac{\partial \zeta_i}{\partial M_{j,k}} = \text{Tr} \left[\left(\frac{\partial \mathcal{Q}^\top}{\partial M_{j,k}} \Gamma(\tau) \mathcal{Q} + \mathcal{Q}^\top \frac{\partial \Gamma(\tau)}{\partial M_{j,k}} \mathcal{Q} + \mathcal{Q}^\top \Gamma(\tau) \frac{\partial \mathcal{Q}}{\partial M_{j,k}} \right) e_{i,i} \right], \\
\frac{\partial \zeta_i}{\partial Q_{j,k}} &= \text{Tr} \left[\left(\frac{\partial \mathcal{Q}^\top}{\partial Q_{j,k}} \Gamma(\tau) \mathcal{Q} + \mathcal{Q}^\top \Gamma(\tau) \frac{\partial \mathcal{Q}}{\partial Q_{j,k}} \right) e_{i,i} \right], \quad \frac{\partial v_{0,i}}{\partial \Sigma_{j,k}} = \text{Tr} \left[\mathcal{Q}^\top e_{j,k} \mathcal{Q} e_{i,i} \right], \\
\frac{\partial v_{0,i}}{\partial M_{j,k}} &= \text{Tr} \left[\left(\frac{\partial \mathcal{Q}^\top}{\partial M_{j,k}} \Sigma_0 \mathcal{Q} + \mathcal{Q}^\top \Sigma_0 \frac{\partial \mathcal{Q}}{\partial M_{j,k}} \right) e_{i,i} \right], \quad \frac{\partial v_{0,i}}{\partial Q_{j,k}} = \text{Tr} \left[\left(\frac{\partial \mathcal{Q}^\top}{\partial Q_{j,k}} \Sigma_0 \mathcal{Q} + \mathcal{Q}^\top \Sigma_0 \frac{\partial \mathcal{Q}}{\partial Q_{j,k}} \right) e_{i,i} \right], \\
\frac{\partial \kappa_i}{\partial \Sigma_{j,k}} &= \frac{1}{t} \left(\frac{1}{v_{0,i}} \frac{\partial v_{0,i}}{\partial \Sigma_{j,k}} - \frac{1}{\zeta_i} \frac{\partial \zeta_i}{\partial \Sigma_{j,k}} \right), \quad \frac{\partial \kappa_i}{\partial M_{j,k}} = \frac{1}{t} \left(\frac{1}{v_{0,i}} \frac{\partial v_{0,i}}{\partial M_{j,k}} - \frac{1}{\zeta_i} \frac{\partial \zeta_i}{\partial M_{j,k}} \right), \\
\frac{\partial \kappa_i}{\partial Q_{j,k}} &= \frac{1}{t} \left(\frac{1}{v_{0,i}} \frac{\partial v_{0,i}}{\partial Q_{j,k}} - \frac{1}{\zeta_i} \frac{\partial \zeta_i}{\partial Q_{j,k}} \right), \quad \frac{\partial \eta_i}{\partial \Sigma_{j,k}} = \frac{\eta_i (1 + t\kappa_i - e^{t\kappa_i})}{2\kappa_i (1 - e^{t\kappa_i})} \frac{\partial \kappa_i}{\partial \Sigma_{j,k}}, \\
\frac{\partial \eta_i}{\partial M_{j,k}} &= \frac{\eta_i}{2} \left(\frac{1}{\varepsilon_i} \frac{\partial \varepsilon_i}{\partial M_{j,k}} + \left(\frac{1}{\kappa_i} + \frac{t}{1 - e^{t\kappa_i}} \right) \frac{\partial \kappa_i}{\partial M_{j,k}} \right), \quad \frac{\partial \eta_i}{\partial Q_{j,k}} = \frac{\eta_i}{2} \left(\frac{1}{\varepsilon_i} \frac{\partial \varepsilon_i}{\partial Q_{j,k}} + \left(\frac{1}{\kappa_i} + \frac{t}{1 - e^{t\kappa_i}} \right) \frac{\partial \kappa_i}{\partial Q_{j,k}} \right), \\
\frac{\partial \theta_i}{\partial \Sigma_{j,k}} &= \frac{\theta_i}{\eta_i \kappa_i} \left(2\kappa_i \frac{\partial \eta_i}{\partial \Sigma_{j,k}} - \eta_i \frac{\partial \kappa_i}{\partial \Sigma_{j,k}} \right), \quad \frac{\partial \theta_i}{\partial \beta} = \frac{\theta_i}{\beta}, \\
\frac{\partial \theta_i}{\partial M_{j,k}} &= \frac{\theta_i}{\kappa_i \eta_i} \left(2\kappa_i \frac{\partial \eta_i}{\partial M_{j,k}} - \eta_i \frac{\partial \kappa_i}{\partial M_{j,k}} \right), \quad \frac{\partial \theta_i}{\partial Q_{j,k}} = \frac{\theta_i}{\kappa_i \eta_i} \left(2\kappa_i \frac{\partial \eta_i}{\partial Q_{j,k}} - \eta_i \frac{\partial \kappa_i}{\partial Q_{j,k}} \right), \\
\frac{\partial \rho_1}{\partial \Sigma_{1,1}} &= -\frac{(Q_{1,1}R_{2,1} + Q_{2,1}R_{2,2})\Sigma_{1,2}}{\sqrt{Q_{1,1}^2 + Q_{2,1}^2}\Sigma_{1,1}^2}, \quad \frac{\partial \rho_2}{\partial \Sigma_{2,2}} = -\frac{(Q_{1,2}R_{1,1} + Q_{2,2}R_{1,2})\Sigma_{1,2}}{\sqrt{Q_{1,2}^2 + Q_{2,2}^2}\Sigma_{2,2}^2}, \\
\frac{\partial \rho_i}{\partial \Sigma_{1,2}} &= \frac{\Sigma_{i,i}}{\Sigma_{1,2}} \frac{\partial \rho_i}{\partial \Sigma_{i,i}}, \quad \frac{\partial \rho_1}{\partial Q_{1,1}} = \frac{Q_{2,1}(Q_{2,1}(R_{1,1}\Sigma_{1,1} + R_{2,1}\Sigma_{1,2}) - Q_{1,1}(R_{1,2}\Sigma_{1,1} + R_{2,2}\Sigma_{1,2}))}{(Q_{1,1}^2 + Q_{2,1}^2)^{3/2}\Sigma_{1,1}}, \\
\frac{\partial \rho_1}{\partial Q_{2,1}} &= \frac{Q_{1,1}(Q_{1,1}(R_{1,2}\Sigma_{1,1} + R_{2,2}\Sigma_{1,2}) - Q_{2,1}(R_{1,1}\Sigma_{1,1} + R_{2,1}\Sigma_{1,2}))}{(Q_{1,1}^2 + Q_{2,1}^2)^{3/2}\Sigma_{1,1}}, \\
\frac{\partial \rho_2}{\partial Q_{1,2}} &= \frac{Q_{2,2}(Q_{2,2}(R_{1,1}\Sigma_{1,2} + R_{2,1}\Sigma_{2,2}) - Q_{1,2}(R_{1,2}\Sigma_{1,2} + R_{2,2}\Sigma_{2,2}))}{(Q_{1,2}^2 + Q_{2,2}^2)^{3/2}\Sigma_{2,2}}, \\
\frac{\partial \rho_2}{\partial Q_{2,2}} &= \frac{Q_{1,2}(Q_{1,2}(R_{1,2}\Sigma_{1,2} + R_{2,2}\Sigma_{2,2}) - Q_{2,2}(R_{1,1}\Sigma_{1,2} + R_{2,1}\Sigma_{2,2}))}{(Q_{1,2}^2 + Q_{2,2}^2)^{3/2}\Sigma_{2,2}}, \quad \frac{\partial \rho_i}{\partial R_{j,k}} = \frac{Q_{k,i}\Sigma_{i,j}}{\sqrt{Q_{1,i}^2 + Q_{2,i}^2}\Sigma_{i,i}}.
\end{aligned}$$

For the sake of simplicity, we left unspecified the derivatives of matrices \mathcal{E} and \mathcal{Q} with respect to WSVM parameters. These quantities can be easily computed from the definition of the matrices involved with the help of a computer algebra system like, for example, Mathematica. Even in this case, missing elements of $J^{BH-W}(\boldsymbol{\pi}_W)$ are equal to zero.