Report on

Multivariate Modeling for Efficient Pricing and Hedging of Multi-asset Derivatives with HPC

Andrea Barletta
Department of Economics and Business Economics at Aarhus BSS
Aarhus University
E-mail: abarletta (at) econ.au.dk

Report on RP 14 - Deliverable 5.3
February 2016

This deliverable summarizes the results achieved within Research Project 14 of the 4-year multi-site Marie Curie Initial Training Network named 'HPCFinance'. The research leading to this deliverable has received funding from the European Union Seventh Framework Programme ([FP7/2007-2013] [FP7/2007-2011]) under grant agreement No. 289032. The results reported in this deliverable are included in three working papers whose full text is accessible on request to the authors.
SUBPROJECT 1
Orthogonal Expansions for Volatility Derivatives
Under Affine Jump Diffusion

Andrea Barletta

Abstract
In this work we derive new closed-form pricing formulas for common volatility derivatives such as realized variance and VIX options. Our approach is based on the classic methodology of approximating a density function with an orthogonal expansion, the latter consisting of a finite sum of polynomials weighted by a kernel. Orthogonal expansions on the Gaussian distribution, such as Edgeworth or Gram-Charlier expansions, have been successfully employed by a number of authors in the context of equity options. However, these expansions are not quite suitable for volatility/variance densities as they inherently assign positive mass to the negative real line. Here we approximate option prices via expansions that instead are based on kernels defined on the positive real line. Specifically, we consider a rather flexible family of distributions, which generalizes the Gamma kernel associated with the classic Laguerre expansions. The method can be employed whenever the moments of the underlying variance distribution are known. It provides fast and accurate price-computations and therefore it represents a valid and possibly more robust alternative to pricing techniques based on Laplace transform inversions. We analyze the accuracy of the developed approximations for volatility options in the Heston (1993) model as well as for the jump-diffusion SVJJ model proposed by (Duffie et al., 1999).

Keywords
volatility derivatives — orthogonal expansions — Laguerre polynomials — jump-diffusion

Department of Economics and Business Economics, Aarhus University, Denmark
*E-Mail: abarletta (AT) econ.au.dk

1 Introduction

Trading volatility derivatives enables investors to implement speculation or hedging strategies by taking a direct position on the risk associated with a certain asset, without investing on the stock itself. Since volatility derivatives were introduced in organized exchanges, the financial literature has dedicated significant attention to stochastic modeling of these products. Part of this literature focuses on stochastic volatility models, which allow the realized variance to be written as the quadratic variation of the underlying process. The advantage of these models over other approaches is that the underlying process and its volatility are modeled together and therefore, in principle, pricing can be done jointly on both assets. However not every stochastic volatility model is suited to price volatility derivatives. For example, it is known that the Heston (1993) model, while enjoying good mathematical tractability, is not capable of matching the upward sloping skew of VIX implied volatilities together with their level. A good compromise between mathematical tractability and consistency with empirical observation can be reached by considering generalizations of the Heston model where the affine structure in the drift of the volatility process is preserved. One way to extend the Heston model without increasing the number of Brownian motions is to add jumps in the dynamics of both stock and volatility (see e.g. Duffie et al. (1999), Sepp (2008a), and Sepp (2012)) or only in the volatility (see e.g. Sepp (2008b)). For this class of models closed-form expressions are unavailable for either the joint or the marginal distributions of the underlying processes. However, due to the affine structure of the volatility drift coefficient, variance swaps and consequently the VIX, can be expressed as elementary functions of the volatility process. Moreover, since the diffusion coefficient is a square-root, the joint characteristic function of stock and volatility processes can be written in closed-form. Consequently, numerical approximations of volatility derivatives may be obtained based on inversion of the characteristic function (see e.g. Lian and Zhu (2011) and cited bibliography). A known drawback of Laplace inversion techniques is that they normally depend on one or more parameters to which they exhibit great instability. Moreover, they might entail slow execution times as they are used for calibration purposes. In this paper we propose a different methodology that only requires a finite number of moments of the underlying process to be computed. This technique, as compared to Laplace inversion, requires that the Laplace transform be “well-behaved” only in a neighborhood of the origin. The core idea behind this methodology is that the density of the underlying process can be approximated by a simpler probability density function (kernel) corrected by a linear combination of polynomials. These polynomials only depend on the kernel density, and therefore all the information on the underlying density is embedded in the coefficients determining the linear combination.
Approximate pricing formulas for derivatives written on a generic asset can thus be obtained based on these polynomial expansions and are still linear functions of the expansion coefficients. A remarkable feature of this class of expansions is that the polynomials can be chosen to be "orthogonal" with respect to the kernel density. When orthogonal polynomials are used, it is possible to characterize the expansion as the solution of a truncated moment problem. Related examples within this framework are Madan and Milne (1994), Abken et al. (1996), and Coutant et al. (2001), whereas later contributions are Zhang et al. (2011), Níguez and Perote (2012), Xiu (2014), and Lin et al. (2015). A common feature of all these works is that proposed expansions are always based on a Gaussian kernel (Hermite expansions), which allows for interpreting the resulting pricing formulas as corrections of the well-known Black and Scholes (1973) formula, accounting for features related to higher moments. However, in all the cases mentioned above, these expansions are applied to an underlying that may reasonably be proxied by a Gaussian kernel, such as the log-return process of a stock. Within the financial applications and in particular for the class of models considered in this paper, the logarithm of the underlying process (the volatility) does not preserve the same mathematical tractability of the underlying itself, which as aforementioned admits exact calculation of moments, up to differentiating its characteristic function. For this reason, we deviate from the Gaussian kernel and derive expansions based on kernels that are supported over the positive real axis to be applied directly to the instantaneous volatility process. Specifically, we consider a family of kernels that embeds the Gamma, the GIG, and the Weibull distributions. The resulting expansions can be seen as generalizations of Laguerre expansions, obtained when the kernel is a Gamma density. We show that generalizing the Gamma kernel significantly improves the accuracy of low orders approximations. Alternatives to Hermite expansions have successfully been adopted in other fields, such as physics, astrophysics, and meteorology, see e.g. Hazut et al. (2015), Gaztanaga et al. (2000), and Morrissey and Greene (2012). In the context of quantitative finance, alternative kernels have been explored for example by Filipovic et al. (2013), where a bivariate orthogonal expansion is applied to pricing of equity options based on a kernel obtained as the product of a Gamma and a bilateral Gamma density, while classic Laguerre expansions are applied to pricing of CDOs. Interestingly, the application to credit risk considered in Filipovic et al. (2013) involves the same class of stochastic jump-diffusion models considered in the present paper. Differently from Filipovic et al. (2013), we consider a generalization of Laguerre expansions, as previously mentioned. Additionally, since the orthogonal polynomials related to our expansions involve special functions of the kernel parameters, we outline an efficient iterative procedure for their numerical computation. The paper proceeds as follows: in the first section we describe the methodology and provide general consistency conditions; in the second section we introduce the family of kernels considered in this paper and derive convergence conditions in terms of the characteristic function of the underlying process to be approximated; in the third section we derive approximate pricing formulas for VIX and realized variance options under the AJD model of Duffie et al. (1999) and we provide convergence conditions in terms of relations between kernel and model parameters; in the last section we illustrate and discuss numerical tests.

2 Methodology

Denoting by $U$ an underlying process and by $f$ its density, the price of a European option with maturity $T$ and payoff $P$ is given by

$$E^Q \{ P(U_T) \} = \int_0^{+\infty} P(x) f(x) dx. \quad (2.1)$$

Since there is no closed-form for $f$, we expand it in a finite sum of orthogonal polynomials, that is

$$f \approx \phi(x) \sum_{i=0}^{n} c_i \rho_i(x), \quad (2.2)$$

where $\phi$ is a fixed measurable function (kernel) with support on $[0, +\infty)$ and possessing all moments, $(\rho_k)_{k \in \mathbb{N}}$ is a family of orthogonal polynomials with respect to $\phi$, and $(c_k)_{k \in \mathbb{N}}$ are constant coefficients carrying all the information on $f$. Once $\phi$ has been fixed, the related orthogonal polynomials can be determined efficiently by the following iterative algorithm

$$h_k(x) = 1, \quad h_1(x) = \frac{x - \mu_1}{\sqrt{\mu_2 - \mu_1^2}}, \quad (2.3)$$

$$h_n(x) = \frac{1}{C_n} (x - \zeta_n) h_{n-1}(x) - \zeta_n h_{n-2}(x), \quad n \leq 2,$$

with

$$\zeta_n = \int_{-\infty}^{+\infty} h_{k-1}(x)^2 x \phi(x) dx,$$

$$\sigma_n = \int_{-\infty}^{+\infty} h_{k-1}(x) h_{k-2}(x) x \phi(x) dx,$$

$$C_n = \sqrt{\int_{-\infty}^{+\infty} h_k^2(x) \phi(x) dx}. \quad (2.4)$$

The expansion coefficients are determined in terms of linear combinations of moments of $U$, as follows

$$c_k = \int_0^{+\infty} f(x) p_k(x) dx.$$

However, since computing higher moments might be numerically tricky, it is desirable to optimize the expansion to get reasonable accuracy at the lowest possible order. To this purpose, we generalize Laguerre expansions by adding two extra
parameters to the Gamma kernel. Specifically, we introduce
the following four-parameter family of kernels
\[ \phi(x) = x^{\alpha-1}e^{-((\beta x + \xi) - x^{-1})}1_{[0, +\infty)}(x), \quad \beta > 0, \xi \geq 0, 0 < \rho \leq 1. \]

(2.5)

Other than the Gamma kernel, this family includes the GIG kernel (for \( p = 1 \)) and the Weibull kernel (for \( \xi = 0 \) and \( \alpha > 0 \)). Approximate pricing formulas are obtained by replacing \( f \) with its approximation (2.2.5) in formula (2.1)
\[ E^Q[P(U_T)] = \sum_{k=0}^{n} \int_{0}^{\infty} p_k(x) \phi(x) P(x) \, dx. \]

(2.6)

Convergence conditions for (2.6) can be determined based on Laplace transforms of both the underlying process and its inverse.

**Proposition 2.1** (Convergence conditions). Let \( U \) be a random variable with density function \( f \) and denote by \( g \) the density function associated with \( Y = U^{-1} \). If Laplace transforms \( L(f)(z) \) and \( L(g)(z) \) of \( f \) and \( g \) are respectively defined for real values \( z > 0 \) and \( \lambda > 0 \), then (2.6) converges for all \( \alpha > 0 \), \( 0 < \beta < \rho \) and \( 0 \leq \xi < \lambda \). Condition on \( \beta \) is unnecessary if \( \rho < 1 \).

When the kernel is a Gamma density and \( f \) is bounded we get the following relaxed convergence conditions.

**Proposition 2.2** (Convergence of Laguerre expansions). Assume that \( \xi = 0, p = 1 \) in (2.5). If \( f \) is bounded and its Laplace transform \( L(f)(z) \) is defined for a real value \( z > 0 \), then (2.6) converges for all \( 0 < \alpha < 2 \) and \( 0 < \beta < \rho \).

### 3 Theoretical framework

In this work we consider an affine jump diffusion model for the log-price \( X \) and the volatility \( \nu \). Specifically, we consider the following SVJJ model, conceived by Duffie et al. (1999)
\[
\begin{align*}
\text{d}X_t &= r - \lambda \bar{\nu} - T^{1/2} W^{(1)}_t + dZ^{(1)}_t, \\
\text{d}\nu_t &= k(\theta - \nu_t) dt + \sigma \sqrt{\nu_t} dZ^{(2)}_t + dZ^{(3)}_t,
\end{align*}
\]

(3.1)

where \( W = (W^{(1)}, W^{(2)}) \) is a correlated Brownian motion with constant correlation \( \rho \), and \( Z = (Z^{(1)}, Z^{(2)}) \) is a bi-dimensional compound Poisson process with intensity \( \lambda \) and correlated jump size processes \( J^{(1)}, J^{(2)} \), that is
\[
Z^{(1)}(\omega) = \sum_{i=1}^{N^{(1)}(\omega)} J^{(1)}(\omega), \quad Z^{(2)}(\omega) = \sum_{i=1}^{N^{(2)}(\omega)} J^{(2)}(\omega),
\]

where for every \( J^{(i)} \sim \text{Exp}(\mu_i) \) are \( J^{(1)} \sim N(\mu_S + \rho \nu z, \sigma^2_Z) \) conditionally to the event \( J^{(i)} = z \). Finally
\[
\bar{\nu} = e^{\kappa z + \frac{1}{2} \sigma^2_Z} - 1 + \rho f \nu \nu - 1
\]
is the compensator related to the jump component in \( X \).

#### 3.0.1 The VIX index

Under affine jump diffusion for the instantaneous variance \( \nu \), the VIX index defined in CBOE (2009) can be written as the square root of an affine function of \( \nu \). Specifically, for the SVJJ model, we have the following result
\[
\text{VIX}_T = 10000 \cdot \sqrt{\text{var} + b + c}
\]

(3.2)

where
\[
a = \frac{1 - e^{\kappa T}}{k T}, \quad \tau = \frac{30 \mu}{365},
\]

(3.3)

\[
b = \left( \theta + \frac{\lambda \mu_S}{k} \right) (1 - a), \quad c = 2\lambda (\mu - \mu_S - \rho \mu). \]

Therefore, it is possible to price derivatives on the VIX by considering \( \nu \) as option underlying. The advantage of this substitution is given by the fact that a closed-form is available for its Laplace transform, as shown by the following proposition.

**Proposition 3.1.** Denote by \( L_{\nu_T} \) the Laplace transform of \( \nu_T \), then, for every \( z \in \mathcal{D}, z < \min[z_1, z_2] \) with
\[
\begin{align*}
z_1 &= \frac{2k}{(1 - e^{k T})e^z + 2\mu_k e^{k T}}, \\
z_2 &= \frac{1}{\mu}
\end{align*}
\]
we have
\[
L_{\nu_T}(z) = h_1(z, T; k, \theta, \epsilon) e^{h_2(z, T, k, \theta, \epsilon) \nu_0}
\]
where
\[
\begin{align*}
h_1(z, T; k, \theta, \epsilon) &= \frac{z (e^{-k T - 1}) (e^z - 2\mu_k e^{k T}) + 1}{2 k (1 - 2 \mu_k)} + \frac{2 k \epsilon}{e^{k T}} (2 k - e^z + z e^2), \\
h_2(z, T; k, \theta, \epsilon) &= \frac{2 k z}{e^{k T} (2 k - e^z) + z e^2}.
\end{align*}
\]

\[\square\]

#### 3.0.2 Realized variance

To price options on the realized variance we shall work under the assumption that jumps only occur in the volatility, meaning that we obtain an SVJJ model. By removing jumps in the log-price the realized variance \( V \) can be written in terms of integrated normalized variance, that is
\[
V_T = \frac{1}{T} \int_{0}^{T} v_T \, dt.
\]

(3.4)

Also in this case, due to the affine structure of the model we get a closed form expression for the Laplace transform of \( V_T \).

**Proposition 3.2.** Denote by \( L(V_T) \) the Laplace transform of \( V_T \), that is
\[
L(V_T)(z) = E[e^{z V_T}] = \int_{0}^{\infty} e^{z x f(x)} \, dx.
\]
Define

\[ \gamma = \gamma(z; k, \epsilon) = \sqrt{k^2 - 2z\epsilon^2} \]

then, for every \( z \in \mathcal{D} \) such that \( z < \min \left\{ \frac{k^2}{2\epsilon^2}, \frac{k}{\sqrt{2\epsilon}} \right\} \), we have

\[ \mathcal{L}(V_T)(zT) = e^{h_1(z, T, k, \theta, \epsilon) + h_1(z, T, k, \theta, \epsilon)zT} \]

where

\[ h_1(z, T, k, \theta, \epsilon) = -\frac{2k\theta}{\sigma^2} \log \left( \frac{c_1 + d_1 e^{-\gamma T}}{c_1 + d_1} \right) \]

\[ + \left( \frac{\lambda}{c_2} - \lambda + \frac{k\theta}{c_1} \right)T + \lambda \left( \frac{d_1}{c_2} - \frac{d_2}{c_2} \right) \log \left( \frac{c_2 + d_2 e^{-\gamma T}}{c_2 + d_2} \right) \]

\[ h_2(z, T, k, \theta, \epsilon) = \frac{1 - e^{-\gamma T}}{c_1 + d_1 e^{-\gamma T}}, \]

and

\[ c_1 = c_1(z; k, \epsilon) = \frac{k + \gamma}{2z}, \quad c_2 = c_2(z; k, \epsilon) = 1 - \frac{\mu_\gamma}{c_1}, \]

\[ d_1 = d_1(z; k, \epsilon) = -\frac{k + \gamma}{2z}, \quad d_2 = d_2(z; k, \epsilon) = \frac{d_1 + \mu_\gamma}{c_1}. \]

\[ \square \]

4 Pricing volatility derivatives

Consider the orthogonal polynomials \((p_k)_{k \in \mathbb{N}}\) introduced in 2.3 and 2.4. We denote by \( a_{ki} \), the \( i \)-th coefficient of \( p_k \), that is

\[ p_k(x) = \sum_{i=0}^{k} a_{ki} x^i. \]

Moreover we denote by \( \mu_i \), the incomplete moment of \( \phi \) defined by

\[ \mu_i(K) = \int_0^{+\infty} x^i \phi(x)dx. \]

4.0.1 Call options on the VIX index

As discussed above, to price VIX call options we consider the variance process \( v \) as underlying asset. Therefore, we consider a European option on \( v \) with payoff function \( P(x) = (\sqrt{ax + b + c - K})^+, \) with \( a, b, c \) as in 3.3. Thereby, the expansion is made on the density of \( v \) whose moments can be obtained via differentiation of Laplace transform. Then we get the following approximating formula for a VIX call option with strike \( K \) and maturity \( T \)

\[ \text{call}_{\text{VIX}}(K, T) = \lim_{n \to +\infty} \frac{1}{n} \sum_{k=0}^{n} c_k A_k, \quad (4.1) \]

with

\[ c_k = \sum_{i=0}^{k} a_{ki} m_i, \quad A_k = 10^4 \sum_{i=0}^{k} a_{ki} \left( I_i^{(K^*)} - K \mu_i^{(K^*)} \right), \]

where \( m_i^{(v)} \) is the \( i \)-th raw moment of \( v_T \) and \( I_i^{(K^*)} \) denotes the following integral

\[ \int_{K^*}^{+\infty} x^i \phi(x) \sqrt{ax + b} \, dx, \quad K^* = \frac{k^2 - b - c}{a}. \]

4.0.2 Call options on the realized variance

The following approximating formula holds for a call option with strike \( K \) and maturity \( T \) on the realized variance process \( V \) defined in 3.4

\[ \text{call}_{\text{RV}}(K, T) = \lim_{n \to +\infty} \frac{1}{n} \sum_{k=0}^{n} c_k A_k, \quad (4.2) \]

with

\[ c_k = \sum_{i=0}^{k} a_{ki} m_i, \quad A_k = \sum_{i=0}^{k} a_{ki} \left( \mu_{i+1}^{(K)} - K \mu_{i}^{(K)} \right), \]

where \( m_i^{(v)} \) is the \( i \)-th raw moment of \( V_T \).

5 Numerical illustrations

In this section we illustrate numerical examples of pricing of realized variance and VIX call options by means of orthogonal expansions based on three different kernels. To measure the accuracy of our results, we plot our approximations against true solutions obtained by Monte Carlo simulation of \( 10^6 \) sample paths on a discretized time mesh of \( 10^3 \) points/month. Parameter sets used in our tests are taken from Sepp (2008b) and Duffie et al. (1999) for the SVJJ model and from Sepp (2008a) for the SVJV model. Results plotted in Figures 1 and 2 demonstrate that low-order expansions based on GIG and Weibull kernels may prove themselves a valid alternative to higher order expansions based on the simpler Gamma kernel. We found that in the Weibull kernel best suits realized variance options, while the GIG kernel best suits VIX options. However, this technique is robust with respect to possible misspecification of the initial kernel, meaning that accuracy may always be recovered by increasing the expansion order, as shown in Figure 3. In particular, from Figure 3b it may be appreciated how the approximation performance is drastically enhanced by increasing the order up to 20. Increasing the order might be necessary to account for possible jump effects in the underlying density, which are not captured either by the kernel or by lower order expansions.
List of figures

**Figure 1.** Prices and implied volatilities of realized variance call options under SVJV model, parameters from Sepp (2008a). Comparison between MC (solid line) and expansions of order 5 based on Gamma (red asterisks) and Weibull (green asterisks) kernels.

**Figure 2.** Prices (left panel) and implied volatilities (right panel) of volatility options under SVJJ model. Comparison between MC (solid line) and expansions of order 5 based on Gamma (red asterisks) and GIG (green asterisks) kernels.

**Figure 3.** Prices (left panel) and implied volatilities (right panel) of VIX call options under SVJJ model. Comparison between MC (solid line) and expansions of order 20 based on GIG kernel (asterisks) kernels.
References


Subproject 2
The Short-time Behaviour of VIX Implied Volatilities in a Multifactor Stochastic Volatility Framework

Andrea Barletta¹*, Elisa Nicolato², StefanoPagliarani³

Abstract
We derive exact asymptotics for short-time VIX implied volatilities in a multifactor stochastic volatility framework which embeds the VAR model of Christoffersen et al. (2010), the double-mean reverting model of Gatheral (2008), the double Heston model of Christoffersen et al. (2009), and other extensions of the Heston (1993) model aiming at fitting simultaneously both SPX and VIX options. Our results are based on perturbation techniques applied to the infinitesimal generator of the underlying process. This methodology has been previously adopted to derive approximations of equity (SPX) options. However, the generalizations needed to cover the case of VIX options are by no means straightforward as the dynamics of the underlying VIX futures are not explicitly known. The obtained expansions are explicit, based on elementary functions and they neatly uncover how the VIX skew depends on the specific choice of the volatility and the vol-of-vol processes. We also provide error estimates as well as numerical implementations to illustrate the accuracy of our technique.

Keywords
VIX options — multifactor stochastic volatility — asymptotic expansions

1 Introduction
The CBOE volatility index (VIX) extrapolates the investors’ expectation of significant future changes in the U.S. stock market on the basis of a discrete number of observed option prices related to the S&P500 (SPX), as specified in CBOE (2009). Under fairly general assumptions, the VIX can be related to the realized variance of the SPX and, specifically, it is known that the square of the VIX at time $t$ can be thought of as a no-arbitrage based attempt of replicating the fair strike of a 30-days variance swap. Due to the complex way that the VIX is defined, analytical tractability of VIX options is guaranteed only in a few exceptional cases. In particular, it is known that it is not possible to construct a modelling framework for VIX options that allows analytical tractability while remaining consistent with reality. As main contribution, in this work we derive short-time asymptotic results for the VIX implied volatilities in a general setting embedding most of the existing pure-diffusion modelling frameworks appearing in the financial literature of VIX option pricing. In our applications, we focus on a wide class of mean-reverting models for the spot variance including, as particular cases, the Heston model and many of its extensions, such as the the VAR model of Christoffersen et al. (2010), the double-mean reverting (DMR) model of Gatheral (2008), the double Heston (DH) model of Christoffersen et al. (2009), and the Chen (1996) model. However, our theoretical setting is rather general, meaning that our technique in principle can be employed in other frameworks. For example stochastic volatility models with non-affine drift could also be considered, such as the 3/2 model of Carr and Sun (2007) and its generalizations considered by Itkin (2013). Alternatively the technique is suited to models where dynamics are given for the entire forward structure of the volatility (see e.g. Buehler (2006) and Bergomi (2008)), provided that the randomness of this structure is determined by a finite number of factors. Finally, the technique also fits into frameworks where the VIX term structure is modelled (see e.g. Huskaj and Nossman (2013) and Badran and Goldys (2015)).

Our methodology consists in adopting a perturbation technique that allows for approximating option prices by expanding the infinitesimal generator of the underlying process. Approximated prices are proven to converge asymptotically in a region of time-to-maturity and log-moneyness that is sufficiently close to the origin. This methodology is pioneered in finance by Hagan and Woodward (1999) and Hagan et al. (2002) where heuristic expansions for the implied volatilities of equity options are formally derived in a local-stochastic volatility framework. Error estimates for this technique have recently been proven by Lorig et al. (2015) and hold in a more general local-stochastic volatility setting. Building on these estimates, asymptotics for short-time ATM implied volatili-
ties of equity options are obtained by Pagliarani and Pascucci (2015) within the same general setting. Differently from the framework of stock option pricing, in our context the underlying process is not included in the original model dynamics. What literature normally refers to as VIX options are indeed options on the VIX futures. Consequently, from a modelling perspective, the mathematical structure defining VIX implied volatilities becomes rather more complicated as compared to other frameworks. This paper, alongside the main contribution given to the literature of VIX option pricing, also yields a small contribution to the literature concerned with this class of perturbation techniques. Namely, we adapt the asymptotic results provided in Pagliarani and Pascucci (2015) to a context where the underlying is known to be a function of one or more processes solving a certain SDE, but whose dynamics are not known explicitly. In this view, it is worth mentioning that classic results obtained by Berestycki et al. (2002), Berestycki et al. (2004), and Durrleman (2010) could not be used for this purpose, as they provide asymptotic results for short-time maturities when the log-moneyness is fixed, which is not consistent with the fact that the ATM value for VIX options changes over time. Differently, the asymptotics provided in Pagliarani and Pascucci (2015) hold in a double regime, and specifically the limits of the implied volatilities and their sensitivities converge as time-to-maturity and log-moneyness jointly approach the origin in a parabolic region of the plane.

2 Modeling setting

In this work we focus on a class of stochastic volatility models all sharing the key feature that the spot volatility and its stochastic factors have an affine drift. Specifically, we consider the following dynamics for the log-price process $X$ and its volatility factors $Y$:

$$
\begin{align*}
\frac{dX_t}{X_t} &= -\frac{1}{2} \sum_{i=1}^{p} Y_{t}^{i} dt + \sum_{i=1}^{p} \sqrt{Y_{t}^{i}} dW_{t}^{X,i}, \\
\frac{dY_t}{Y_t} &= (\alpha Y_t + \beta) dt + \eta(t, Y_t) dW_{t}^{Y},
\end{align*}
$$

(2.1)

where $W = \left(W^{X}, W^{Y}\right)$ is an $n$-dimensional Brownian motion with $n \geq p$, $\alpha$ is an $n \times n$ matrix with non-positive diagonal elements, $\beta$ is a non-negative vector of $\mathbb{R}^{n}$, and $\eta$ is an $n \times n$ diagonal matrix-valued function such that (2.1) admits a solution and in particular the first $p$ components of $Y$ are all non-negative. Finally, the correlation matrix $\rho$ of $W$ is assumed to have the following blocks form

$$
\rho = \begin{bmatrix}
I_{p} & \rho^{X,Y} \\
\rho^{X,Y} & \rho^{Y,Y}
\end{bmatrix},
$$

where $I_{p}$ denotes the $p \times p$ identity matrix, $\rho^{X,Y}$ is a $p \times n$ matrix, and $\rho^{Y,Y}$ is an $n \times n$ positive semi-definite matrix with unitary diagonal elements. This particular form of $\rho$ ensures that

$$
[X]_{t}^{t+\Delta} = \sum_{i=1}^{p} \int_{t}^{t+\Delta} Y_{s}^{i} ds.
$$

Proposition 2.1. If $(X, Y)$ solves (2.1), then the square of the VIX is an affine function of $Y$, more precisely

$$
VIX_{t}^{2} = \sum_{i=1}^{p} R_{i}(Y_{t}; \alpha, \beta, \Delta),
$$

(2.2)

where $R_{i}$ is the $i$-th component of

$$
R(Y_{t}; \alpha, \beta, \Delta) = \left( \int_{0}^{\Delta} e^{\alpha s} ds \right) Y_{t} + \left( \int_{0}^{\Delta} e^{\alpha s} (\Delta - s) ds \right) \beta,
$$

(2.3)

with $e^{\alpha s}$ denoting the exponential matrix of $s \alpha$.

The Heston (H) model is the simplest sub-case of (2.1). A known drawback of this model is that it cannot adequately reproduce the upward slope that is empirically observed in the VIX implied volatilities. This is primarily due to the poor flexibility of the variance term structure induced by the model. To provide consistent pricing models for VIX options, it is fundamental to achieve a rich modelling structure of the forward variances, since they can be interpreted as spot variance swaps and therefore are crucially linked to the VIX. Hence, to enrich the variance term structure while remaining in the pure diffusion framework, it is necessary to introduce additional stochastic factors in the Heston model. In the double Heston (DH) model proposed in Christoffersen et al. (2009), a second stochastic volatility factor is proposed, originally to improve the model capability of fitting the implied volatilities observed in the equity market. The authors claim that their model has great potential to also provide a richer structure to the forward variances. In fact, the double Heston model is adopted in Da Fonseca et al. (2015) to price derivatives that jointly involve the stock and its volatility. Following a different approach, the double mean-reverting (DMR) model is proposed in Gatheral (2008) and Bayer et al. (2013) to provide consistent joint pricing of the SPX and variance swaps. Here, the long-term level of the volatility is allowed to be stochastic, and elasticity coefficients are introduced in the dynamics of both the spot variance and its long-term level to have better control on the shape of their tails. Another generalization of the variance process in the Heston model can be obtained by randomizing the vol-of-vol level (SVV model). As discussed in Sepp (2008), this may represent a valid alternative to adding jumps in the spot variance dynamics in order to obtain the same effect of increasing the probability of tail events in the volatility. A model that is conceptually similar to the SVV model has been applied in Ren et al. (2007) to perform a joint calibration on the implied volatilities generated by the SPX and the VIX. An even more generalized extension of the Heston model that is still included in (2.1) is the Chen model, which is proposed in Chen (1996) for the term structures of interest rates. Similarly to other models
for interest rates, this model can be reinterpreted as a model for the volatility, where both the mean and the volatility of volatility are assumed to be stochastic. However the Chen model has never been adopted, to the best of our knowledge, in the context of VIX options, possibly due to its difficult numerical tractability.

In Figure 2 is shown a comparison between the VIX implied volatilities generated by H, DH, DMR, and SVV models, where parametrizations are consistently chosen so that every model shares the same initial values for volatility, speed of mean reversion, long-term level, and vol-of-vol. From that illustration it clearly emerges that all the three extensions of the Heston model have potential to overcome its drawbacks in the context of VIX option pricing. Unfortunately, none of these models admit closed or semi-closed forms for VIX options and related implied volatilities. Therefore, it is relevant to provide an efficient numerical technique for pricing of VIX options that can be applied to the general model specified in (2.1).

3 Asymptotics of VIX futures and implied volatilities

In this work we derive first order at-the-money asymptotics of the VIX implied volatilities within the theoretical setting discussed in the previous section. Here we illustrate and discuss results for some key sub-cases of (2.1). Strong attention is dedicated to how enhancing different features of the Heston model variance process affects the model capability of generating upward sloping implied skews from VIX option prices. Throughout the section we denote by \( a = a(\alpha, \beta) \in \mathbb{R}^n \) and \( b = b(\alpha, \beta) \in \mathbb{R} \) the coefficients such that

\[
\text{VIX}_t = \sqrt{a(Y_t)} + b. \tag{3.1}
\]

where \( a \) and \( b \) can be uniquely determined in light of Proposition 2.1. Furthermore, for any \( K > 0 \) and \( 0 \leq t < T \), we indicate by \( \sigma_{t,T,K}^{\text{imp}} \) the VIX implied volatility (VIX-IV) at time \( t \) of a call option \( \text{VIX}_{t,T}^{\text{imp},K} \) with maturity \( T \) and strike \( K \), defined as the unique positive solution (in \( \sigma \)) of

\[
c_{t,T,K}^{\text{VIX}} = u^{\text{BS}}(\sigma; \log F_t^I, \log K, T-t), \tag{3.2}
\]

where \( u^{\text{BS}} \) is the Black-Scholes call price function.

The starting point for our discussion is the Heston model, where the variance \( Y_t \) is modelled through a square-root process and features constant levels of mean reversion and vol-of-vol

\[
dX_t = r - \frac{1}{2} (Y_t)^2 dt + \sqrt{Y_t} dW_t^X \\
dY_t = k (Y_t - \theta) dt + \sigma \sqrt{Y_t} dW_t^Y. \tag{3.3}
\]

For the Heston model we obtain the following exact expression for the at-the-money slope of the VIX implied volatility holding as \( T \to t \)

\[
\lim_{K \to \text{VIX}_{t,T}^{\text{imp}}} \frac{\partial}{\partial K} \sigma_{t,T,K}^{\text{imp}} = \frac{\epsilon (\text{VIX}_t^2 - 2aY_t)}{4\text{VIX}_t^2 \sqrt{Y_t}}. \tag{3.4}
\]

Remarkably, since the quantity \( \text{VIX}_t^2 - 2aY_t \) may be expected to be negative for relevant values of \( k \) and \( \theta \), the at-the-money implied volatility slope of VIX options will in turn be negative and decreasing in the parameter \( \epsilon \), controlling the vol-of-vol size. However, from Proposition 2.1 it turns out that \( a \) and \( b \) are decreasing functions of the speed of mean reversion \( k \) and \( b \) is increasing in both \( k \) and the long-term level \( \theta \). Hence, in principle it is possible to recover a positive slope for high values of \( k \) and \( \theta \), but that would imply that VIX deviates by a large extent from \( \sqrt{Y_t} \), which in general is not consistent with reality. We now investigate the effects of different enhancements of the Heston model on the implied skew of VIX options.

1. Adding a constant of elasticity. Adding an elasticity constant in the variance process of the Heston model yields a CEV-type model with mean-reverting drift

\[
dX_t = r - \frac{1}{2} (Y_t)^2 dt + \sqrt{Y_t} dW_t^X \\
dY_t = k (Y_t - \theta) dt + \gamma \sqrt{Y_t} dW_t^Y. \tag{3.5}
\]

The introduction of an elasticity coefficient in the variance process of the Heston model allows for a better control on the tail of \( \sqrt{Y_t} \) and consequently of \( \text{VIX}_t \). In particular, the VIX process \( \text{VIX}_t \) is mesokurtic when \( \delta = \frac{1}{2}, \) platykurtic when \( \delta < \frac{1}{2}, \) and leptokurtic when \( \delta > \frac{1}{2}. \) The volatility and therefore the VIX are known to be leptokurtic, which suggests that the upward sloping of the VIX option implied skew can be matched only when \( \delta > \frac{1}{2}. \)

We obtain the following asymptotic result for \( \sigma_{t,T,K}^{\text{imp}}(t,T,K) \) under the dynamics specified in (3.5)

\[
\lim_{K \to \text{VIX}_{t,T}^{\text{imp}}} \frac{\partial}{\partial K} \sigma_{t,T,K}^{\text{imp}}(t,T,K) = \frac{\gamma Y_t^{\frac{1}{2}} (2\delta \text{VIX}_t^2 - 2aY_t)}{4\text{VIX}_t^2 \sqrt{Y_t}}. \tag{3.6}
\]
Equation (3.6) confirms that $\delta$ drastically affects the slope of $\sigma^\text{imp}$ for short maturities and when the strike is close to at-the-money. In particular, $b \geq 0$ implies $\text{VDX}_t^2 - a_t \geq b > 0$, meaning that there exists a certain value $\delta^* < 1$ such that the slope of $\sigma^\text{imp}(t,T,K)$ is positive for all $\delta > \delta^*$ and negative for all $\delta < \delta^*$. Again, the size of vol-of-vol only has an impact on the magnitude of the slope, but not on its sign. This result is consistent with the empirical analysis that has been carried out in Christoffersen et al. (2010), where the Heston model has been shown to be misspecified and outperformed by the VAR model, obtained from (3.5) for $\delta = 1$.

(2) Randomizing the size of vol-of-vol. We now consider the case where the level of vol-of-vol in the Heston model is allowed to be stochastic, and therefore the volatility process is driven by the following dynamics:

$$
\begin{align*}
\dot{X}_t &= r - \frac{1}{2} (Y_t^2)^2 dt + \sqrt{Y_t^2} dW_t^X \\
\dot{Y}_1^t &= k_1 (Y_1^t - \theta_1) dt + Y_1^t \sqrt{Y_t^2} dW_t^{Y_1} \\
\dot{Y}_2^t &= k_2 (Y_2^t - \theta_2) dt + \zeta(Y_2^t) dW_t^{Y_2} \\
d\langle W^{Y,1}, W^{Y,2} \rangle_t &= \sigma_t dt
\end{align*}
$$

(3.7)

where $\zeta$ is a non-negative function on $\mathbb{R}^+$ such that (3.7) has a solution. When $\zeta(y) = \sqrt{y}$, for every $y \geq 0$, we obtain the SV model. Here we obtain the following asymptotics:

$$
\lim_{K \to \text{VDX}_t, T \to t} \frac{\partial}{\partial K} \sigma^\text{imp}(t,T,K) = \frac{Y_1^t \left( \text{VDX}_t^2 - 2a_1 Y_1^t \right)}{4 \sqrt{Y_1^t \text{VDX}_t^2}} + \frac{\zeta(Y_1^t)}{2Y_1^t}.
$$

(3.8)

From (3.8) we observe that since $a_2 = 0$ (see Proposition 2.1) only the diffusion component of $Y_2^t$ has an impact on the at-the-money slope of short maturity VIX implied skew generated by the model defined in (3.7). In particular, randomizing the level of vol-of-vol results in an additive factor in the slope generated by the simpler Heston model. A first fact to be mentioned is that the sign of the additive factor is determined by the sign of $\zeta$. This is consistent with the discussion provided in Sepp (2008) that the correlation between the volatility and the vol-of-vol must be positive in order to generate an upward sloping implied skew. A second important observation is that the additive factor also determines the sign of the slope for large values of the level of vol-of-vol $Y_2^t$ at time $t$. Specifically, when $\lim_{y \to +\infty} \frac{\zeta(y)}{y} = +\infty$, then the slope of $\sigma^\text{imp}(t,T,K)$ will certainly be positive for short maturities and near at-the-money value for large values of $Y_2^t$. Conversely, when $\lim_{y \to +\infty} \frac{\zeta(y)}{y} = 0$ and for relevant choices of the model parameters, large values of vol-of-vol at time $t$ will imply a negative slope of $\sigma^\text{imp}(t,T,K)$.

(3) Adding another stochastic volatility factor When adding a second stochastic volatility factor in the Heston model, we obtain the DH model introduced in Christoffersen et al. (2009)

$$
\begin{align*}
\dot{X}_t &= r - \frac{1}{2} (Y_1^t)^2 dt + \sqrt{Y_1^t} dW_t^X \\
\dot{Y}_1^t &= k_1 (Y_1^t - \theta_1) dt + \epsilon_1 \sqrt{Y_1^t} dW_t^{Y_1} \\
\dot{Y}_2^t &= k_2 (Y_2^t - \theta_2) dt + \epsilon_2 \sqrt{Y_1^t} dW_t^{Y_2} \\
d\langle W^{Y,1}, W^{Y,2} \rangle_t &= 0
\end{align*}
$$

(3.9)

For this model, the asymptotic expression of the slope of $\sigma^\text{imp}$ is

$$
\lim_{K \to \text{VDX}_t, T \to t} \frac{\partial}{\partial K} \sigma^\text{imp}(t,T,K) = \frac{a_1^2 \text{VIX}_t^2 Y_1^t - 2a_1^2 (Y_1^t)^2}{4C} + \epsilon_2^2 \left( \frac{a_2^2 \text{VIX}_t^2 Y_2^t - 2a_2^2 (Y_2^t)^2}{4C} \right) - \epsilon_1^2 \epsilon_2 \frac{a_1^2 a_2^2 Y_1^t Y_2^t}{C}.
$$

(3.10)

where

$$
C = \text{VIX}_t^2 \left( a_1^2 \epsilon_1 + a_2^2 \epsilon_2 \right)^{3/2}.
$$

(3.11)

It is clear from (3.10) how adding a second volatility factor in the Heston model improves the model capability to match many different shapes of the volatility skew implied by VIX options, even without assuming correlation between the two factors. The flexibility of the model is essentially given by the fact that there exist several parametrizations all yielding the same initial value, long-term level, and the vol-of-vol of the variance process given by $Y_1^t + Y_2^t$. For instance, equation (3.10) suggests that upward sloping implied skews may be obtained by setting $\theta_1$, $k_1$ sufficiently large, $Y_2^t$ sufficiently small, and then adjusting the remaining parameters to get consistent structural properties of the variance process. This can be acknowledged by observing that as $Y_1^t \to 0$ the sign of (3.10) is determined by the following factor

$$
\text{VIX}_t^2 - 2a_1 Y_1^t = b - a_2 Y_2^t.
$$

(3.12)

Differently from the case of the Heston model, this quantity now depends on four parameters ($k_1, \theta_1, k_2, \theta_2$) and therefore can be controlled with less restrictions on the structural properties of the variance process.

(4) Randomizing the level of mean-reversion The last model considered in this section is the generalization of the Heston model obtained by allowing the variance long-term level to be stochastic, and in turn to be a square root process:

$$
\begin{align*}
\dot{X}_t &= r - \frac{1}{2} (Y_1^t)^2 dt + \sqrt{Y_1^t} dW_t^X \\
\dot{Y}_1^t &= k_1 (Y_1^t - \theta_1) dt + \epsilon_1 \sqrt{Y_1^t} dW_t^{Y_1} \\
\dot{Y}_2^t &= k_2 (Y_2^t - \theta_2) dt + \epsilon_2 \sqrt{Y_1^t} dW_t^{Y_2} \\
d\langle W^{Y,1}, W^{Y,2} \rangle_t &= 0
\end{align*}
$$

(3.13)
The model defined in (3.13) is a particular case of the DMR model of Gatheral (2008), obtained when $\delta_1 = \delta_2 = \frac{1}{3}$ and $g = 0$. Here, the asymptotic expression for the ATM slope of $\sigma^{\text{imp}}$ is identical to (3.10), which is given for the model defined in (3.9). However, we notice that the coefficients $a_1, a_2, b$ appearing throughout the formula are different functions of parameters of the two models, therefore Equation (3.10) leads to different results when $a_1, a_2$, and $b$ are written explicitly. An intuitive motivation of this result is given by the fact that the two equations for $Y$ given in (3.9) and (3.13) only differ in the drift component and, as time-to-maturity tends to zero, the drift component of $Y$ has no direct impact on the implied volatility skew generated by VIX options. However, as compared to the model defined in (3.9), the fact that the process $Y^2$ in (3.13) represents the long-term level of $Y^1$ may restrict the flexibility in the choice of parameters, and therefore, for relevant parametrizations, it may be unlikely to yield upward sloping implied skews. As discussed in Bayer et al. (2013) a great improvement is obtained by adding an elasticity coefficient, which is confirmed by comparing Equation (3.4) and Equation (3.5).

4 Numerical illustrations

As already mentioned, in this work we obtain expansions of the VIX implied volatilities that converge asymptotically as time-to-maturity and log-moneyness approach the origin. In this section we illustrate the accuracy of our technique in fitting the short-maturity implied volatilities generated by VIX call options, based on numerical tests made under the DH model (Figure 1), the DMR model (Figure 2), and the SVV model (Figure 3). We refer the reader to the previous section for the definition of these models. All plots read as follows: Monte Carlo simulation (blue solid) and expansions of order 2 (red asterisks), order 3 (yellow circles) and order 4 (violet crosses). Our approximating technique relies on asymptotic results holding for short time-to-maturity and strike near ATM. Hence, to assess the overall performance of the technique we must consider two different aspects: the range of maturities and the range of strikes for which it provides sufficiently good approximations of the true solution. In all of our tests approximations of the implied volatilities generated by VIX options are benchmarked against results of Monte Carlo simulations. In order to obtain a reliable benchmark, for every considered example we simulate $10^6$ paths over a discrete mesh of $\sqrt{\frac{n}{12}} \times 10^3$ points for every time unit, where $n$ denotes the number of Brownian motions involved in the simulation. From our tests it emerges that our technique performs significantly better when applied to the SVV model, both over time and moneyness. Under the DMR model it seems more difficult to achieve accuracy for large maturities, while under the DH model the approximation seems to be more stable over time but more unstable as regards the strikes range.

5 List of figures

Figure 1. Implied volatility smiles of VIX call options under the DH model, as functions of the strike moneyness.

Figure 2. Implied volatility smiles of VIX call options under the DMR model, as functions of the strike moneyness.

Figure 3. Implied volatility smiles of VIX call options under the SVV model, as functions of the strike moneyness.
References


SUBPROJECT 3
Retrieving Risk-Neutral Densities Embedded in VIX Options: a Non-Structural Approach

Andrea Barletta¹*, Paolo Santucci de Magistris², Francesco Violante³

Abstract
We propose a non-structural pricing method to retrieve the VIX risk-neutral density directly from the related option chains. This method is based on finite orthogonal polynomial expansions around a kernel density. The polynomial expansion yields a description of the risk-neutral density without the need of assuming risk-neutral dynamics for the VIX. The methodology can be thought of as an alternative to Hermite expansions where the underlying asset is a volatility and therefore is proxied by a kernel supported on the positive real axis. Importantly, the method only imposes mild regularity conditions on the shape of the risk-neutral density. Based on this technique, we derive a simple and robust way to estimate the expansion coefficients through OLS regression. To tackle the problem of multicollinearity that arises as the expansion order increases we rely on principal component analysis. Examples based on either model-generated or real data are provided to support the proposed technique in a large variety of cases.

Keywords
VIX options — orthogonal expansions — non-structural — principal components

1 Introduction
The Volatility Index (VIX) was introduced in 1993 by Chicago Board Options Exchange (CBOE) to measure the market expected volatility. In its first formulation, the VIX was defined as an average of S&P 100 call and put implied volatilities. In response to growing interest in trading of volatility, in 2004 CBOE introduced VIX futures alongside a revised formulation of the VIX, based on replication of variance swap contracts written on the broader S&P 500 (SPX) index. Specifically, in its current formulation reported by CBOE (2009), the VIX is computed as present value of a portfolio of SPX call and put options constructed as a static replication attempt of a 30-days variance swap. Following up VIX futures, in 2006 VIX options were introduced. Since then, several authors have investigated pricing of VIX options. For example, Sepp (2008a,b) prices VIX derivatives on the basis of Fourier inversion techniques applied to stochastic volatility models that are nested in the AJD framework of Duffie et al. (1999). The recent paper by Bardgett et al. (2014) extends the method of Sepp (2008a,b) for the dynamic estimation of the volatility risk premium based on the joint calibration on both the SPX and VIX options. Following two different approaches, Bergomi (2008), Gatheral (2008), and Cont and Kokholm (2013) provide modeling frameworks aimed at pricing variance swaps jointly with the SPX.

A central feature of any such approach is that option prices are obtained based on parametric assumptions on the risk-neutral density (RND) of the VIX. In this way, the RND is fully described as a function of the underlying model parameters, which entails an intrinsic model risk (see e.g. Cont (2006)) mainly concerned with the possibility of model incorrect specification. Handling model risk in VIX option pricing is particularly troublesome since its linkage with the SPX is not explicit. On the other hand, modeling the instantaneous variance is equally subject to misspecification risks, as the quantity being modeled in this case is not observable. Comparative analyses of the performance of simple stochastic volatility models in pricing of VIX options tend to confirm this fact. For example, Christoffersen et al. (2010) and Wang and Daigler (2011) find some evidence in favor of models that assume log-normal dynamics for the instantaneous variance, although remarking that no model has small pricing errors over the entire range of strike prices. The econometric analysis carried out by Mencia and Sentana (2013) reveals especially high misspecification risk in structural pricing of VIX options during financial crises. Reducing the model risk concerned with VIX option pricing is possible but often comes at the cost of mathematical tractability and availability of closed-form solutions. As a consequence, consistent modeling frameworks conceived for capturing stylized facts of the VIX are rarely suited to be employed for estimation purposes.

In this view, non-structural methods for estimating the
RND directly from VIX options represent a valid alternative to stochastic modeling. Non-structural means that only mild regularity conditions on the form of the RND are imposed, which entails considerable reduction of misspecification risks. Non-structural option pricing has been recently employed by Song and Xiu (2015) with the purpose of estimating the volatility pricing kernel, which is proportional to the ratio between the physical and the risk-neutral density of the VIX. In particular, Song and Xiu (2015) adopt a direct method to extract the marginal risk-neutral densities of SPX and VIX from the related options, building upon the insight of Breeden and Litzenberger (1978). Unfortunately, this approach strongly relies on the condition that there exist traded derivatives on the same asset for a continuum of its future states, which is not the case in practice. Consequently, the original data needs to be complemented with artificial points, e.g. by interpolation techniques (see Monteiro et al. (2008) and Song and Xiu (2015)). The main shortcoming of this approach lies in the fact that estimated RNDs are highly sensitive to how observed data is complemented.

The method that we propose still belongs to the class of non-structural approaches but removes the restrictive condition required in Breeden and Litzenberger (1978). In particular, we recover the RND underlying VIX options by means of a finite orthogonal expansion (see e.g. Szegö (1939)) around a kernel density. As notable representatives of this class we mention Hermite expansions, which are obtained when the kernel is a Gaussian density, and Laguerre expansions, which are obtained when kernel is an exponential density. The key feature of orthogonal expansions is that they yield a description of the RND without the need of specifying stochastic dynamics. This method imposes mild integrability conditions on the form of the RND, proving to be particularly robust as estimator. There is exhaustive literature related to the use of orthogonal expansions for financial applications. Seminal examples are Jarrow and Rudd (1982), Corrado and Su (1996), Madan and Milne (1994), Coutant et al. (2001), and Jondeau and Rockinger (2001), while more recent contributions are Zhang et al. (2011), Níguez and Perote (2012), and Xiu (2014). Noticeably, in all cases above the considered expansions are in terms of Hermite polynomials. Our methodology can be thought of as an alternative to the Hermite expansion where the kernel is supported over the positive real axis. In particular, we extend the classic Laguerre expansion proposed by Filipovic et al. (2013) by allowing full flexibility in the choice of the kernel.

We make the following contributions to the present literature on pricing of VIX options. First, we provide rather general convergence conditions of orthogonal expansions to the true RND and show that they are concerned with the tail behavior of the expansion kernel. Second, consistently with theoretical results, we show that the log-normal density is not a suitable choice for the expansion kernel, due to its slow decay rate that implies stable but generally inaccurate approximations. Instead, we base our expansions on a wide family of kernels encompassing the exponential, the Gamma, the Weibull, and the GIG distributions. Since this family is characterized by a flexible decay rate on both tails, the related orthogonal polynomials prove to be particularly suited for modeling the RND associated to the VIX options. Third, in the same spirit of Ait-Sahalia and Duarte (2003) and Jondeau and Rockinger (2001), we propose an econometric methodology to estimate the parameters of the polynomial expansion by minimizing the distance between the observed option prices and those implied by the model. Finally, we test the robustness of the proposed method on both option prices generated from known RNDs and market data. The results highlight the reliability of the suggested methodology to recover risk-neutral densities up to negligible rounding errors and minor adjustments of the observed prices. Interestingly, although this paper focuses on VIX options, our methodology is outlined in full generality and it can be applied to any type of options to recover the underlying RND.

2 Methodology

Our methodology is developed under the general risk-neutral assumptions, meaning that for a fixed maturity $T$ there exists a function $f_Q$ such that the price $V^\Pi$ of a European option with payoff $\Pi$ on a generic asset $X$ can be written as

$$V^\Pi = \int_0^{+\infty} \Pi(x) f_Q(x) \, dx. \quad (2.1)$$

The function $f_Q$ appearing in (2.1) is referred to as risk-neutral density (RND). In order to obtain a closed-form approximation $f_Q^{(n)}$ of the density $f_Q$ associated with the risk-neutral measure $Q$, we consider the following expansion

$$f_Q^{(n)}(x) = \phi(x) \left( 1 + \sum_{k=1}^{n} \epsilon_k(x) \right), \quad n \geq 1 \quad (2.2)$$

where $\phi$ is a kernel density and $\epsilon_k$ are corrective factors.

The kernel density is such that, after being properly normalized, it represents the 0-order term of the expansion

$$f_Q^{(0)}(x) = \phi(x).$$

The corrective factors $\epsilon_1, \ldots, \epsilon_n$ are defined as

$$\epsilon_k := c_k h_k,$$

where $c_k$ are constant coefficients embedding all the information on $f_Q$, while $h_k$ is a polynomial function of degree $k$, that only depends on the choice of $\phi$. Given $\phi$ and $h_k$, we can approximate $V^\Pi$ by replacing (2.2) in (2.1)

$$V^\Pi \approx \int_0^{+\infty} \Pi(x) f_Q^{(n)}(x) \, dx. \quad (2.3)$$

The following results provide the method with consistency:
Definition 2.1 (Orthogonal Polynomials). The elements of a family \((h_k)_{k \in \mathbb{N}}\) are said to be orthogonal polynomials with respect to \(\phi\), if \(h_k\) is a polynomial for every \(k \in \mathbb{N}\) and there exists a function \(\phi\), possessing all moments, such that
\[
\int_{-\infty}^{\infty} h_i(x)h_j(x)\phi(x)dx = 0, \quad \forall i, j \in \mathbb{N}. \tag{2.4}
\]

Henceforth, we denote such a family of orthogonal polynomials by \((h_k^\phi)_{k \in \mathbb{N}}\) to emphasize its dependence on \(\phi\). Furthermore, we denote by \(\mathcal{W} := (w_{ij})\) the \((n+1) \times (n+1)\) lower triangular matrix containing the coefficients of the polynomials \(h_0^\phi, \ldots, h_n^\phi\), for a given \(n \in \mathbb{N}\). Specifically,
\[
h_i^\phi(x) = w_{i,0} + w_{i,1}x + \ldots + w_{i,n}x^i, \quad i = 1, \ldots, n
\]
and \(w_{ij} = 0\) for \(j > i\). Uniqueness and existence results for \((h_k^\phi)_{k \in \mathbb{N}}\) are provided in the following theorem.

Theorem 2.2. If \(\phi\) is a function possessing all moments, then there always exists a family \((h_k^\phi)_{k \in \mathbb{N}}\) of polynomials orthogonal with respect to \(\phi\). Furthermore, \(h_k^\phi\) is uniquely defined up to a multiplicative constant and uniquely defined up to a sign if the following additional condition is required
\[
\int_{-\infty}^{\infty} h_k^\phi(x)^2\phi(x)dx = 1. \tag{2.5}
\]
□

Remark 2.3. In general, \(f_Q^{(n)}\) is not guaranteed to be a positive function over its support, even under the assumption that \(\phi\) is a positive function. However, this property is recovered when the \(n\)-th coefficient \(c_n\) fulfills some constraints. More precisely, \(f_Q^{(n)}\) is only if and only if
\[
c_n \leq c_n \leq c_n^{\sup}, \tag{2.6}
\]
where
\[
c_n^{\inf} = \inf_{x : h_k^\phi(x) > 0} \frac{f_Q^{(n-1)}(x)}{h_k^\phi(x)}, \quad c_n^{\sup} = \inf_{x : h_k^\phi(x) < 0} \frac{f_Q^{(n-1)}(x)}{|h_k^\phi(x)|}.
\]
□

Theorem 2.4. Let \(\phi\) be a chosen kernel for (2.2) and denote by \(\mathcal{D}\) its support. Define
\[
\mathcal{H}_\phi = \left\{ \psi : \psi \cdot \phi \in L^2(\mathcal{D}) \right\},
\]
\[
\mathcal{H}_\phi^* = \text{Cl}(\text{span}\{\phi h_k^\phi, k \in \mathbb{N}\}) \subseteq \mathcal{H}_\phi,
\]
\[
d_\phi(\psi_1, \psi_2) = \left( \int_{\mathcal{D}} |\psi_1(x) - \psi_2(x)|^2 \frac{1}{\phi(x)}dx \right)^{1/2}, \quad \psi_1, \psi_2 \in \mathcal{H}_\phi.
\]
Then:

(a) If \((c_k)_{k \in \mathbb{N}}\) is a real sequence satisfying
\[
\exists \lim_{n \to \infty} \sum_{k=0}^{n} c_k^2 < \infty,
\]
there exists a unique function \(f_Q^{(0)}\) such that
\[
\lim_{n \to \infty} d(f_Q^{(0)}, f_Q^{(n)}) = 0.
\]
(b) For every function \(\Pi\) such that \(\Pi^2\phi\) is integrable the function \(f_Q^{(0)}\) defined in (a) satisfies
\[
\lim_{n \to \infty} \int_{\mathcal{D}} \Pi(x)f_Q^{(0)}(x)dx = \int_{\mathcal{D}} \Pi(x)f_Q^{(0)}(x)dx.
\]
(2.7)
(c) If \(f_Q\) satisfies \(\phi^{-1/2} f_Q \in L^2(\mathcal{D})\) and \(\text{supp}(f_Q) \subseteq \mathcal{D}\), there exists a sequence \((c_k^*)_{k \in \mathbb{N}}\) such that the function
\[
f_Q^* = \lim_{n \to \infty} \left(1 + \sum_{k=1}^{k} c_k^* h_k^\phi \right) \text{in } \mathcal{H}_\phi
\]
solves
\[
f_Q^* = \arg\min_{\psi \in \mathcal{H}_\phi} d_\phi(\psi, f_Q). \tag{2.8}
\]
Specifically, for every \(k \in \mathbb{N}\)
\[
c_k^* = \int_{\mathcal{D}} h_k^\phi(x)f_Q^*(x)dx = \frac{1}{k} \sum_{i=1}^{k} w_{k,i} \int_{-\infty}^{\infty} h_k^\phi(x)x^i f_Q^*(x)dx
\]
(2.9)
is determined by a linear combination of the first \(k\) moments of \(f_Q\), where \(w_{k,i}\) is the \(i\)-th coefficient of \(h_k^\phi\).
□

In the following, given a kernel function \(\phi\) and an associated set of orthogonal polynomials \((h_k^\phi)_{k \in \mathbb{N}}\), the notations \(\mathcal{H}_\phi^*\) and \(\mathcal{H}_\phi^\dagger\) refer to the Hilbert spaces already defined in Theorem 2.4, where we recall that the underlying scalar product is defined as
\[
\langle \psi_1, \psi_2 \rangle := \int_{\mathcal{D}} \psi_1(x)\psi_2(x) \frac{1}{\phi(x)}dx,
\]
with \(\mathcal{D} = \text{supp}(\phi)\).

Definition 2.5 (Closed polynomials system in \(\mathcal{H}_\phi\)). The orthogonal set of polynomials \((h_k^\phi)_{k \in \mathbb{N}}\) is said to be closed with respect to \(\phi\) if
\[
\text{Cl}(\text{span}\{x^k, k \in \mathbb{N}\}) = L^2_{\phi(x)dx}(\mathcal{D}). \tag{2.10}
\]
Theorem 2.6 (Conditions to the closure of \((h^k_k)_{k\in\mathbb{N}}\)). Let \(\phi\) be a positive integrable function and \(\mathcal{D} = [0, +\infty[\). 

(i) If \(\lim_{x \to +\infty} \phi(x) e^{\pm \xi^2} = 0\) for some \(\xi > 0\), then \((h^k_k)_{k\in\mathbb{N}}\) is closed with respect to \(\phi\).

(ii) If \(\lim_{x \to +\infty} \phi(x) e^{\pm \xi^2} > 0\) for some \(\xi > 0\), then \((h^k_k)_{k\in\mathbb{N}}\) is not closed with respect to \(\phi\).

\[\square\]

2.1 Generalized Laguerre and log-Hermite expansions

In this paragraph we introduce and discuss the properties of the following family of kernel functions with support \(\mathcal{D} = [0, +\infty[\)

\[\phi(x) = x^{\alpha-1} e^{-(\beta x^\rho + \xi x^{-\sigma})}, \quad \alpha, \beta, \xi > 0, \ p \in [0, 1],\]

which, up to normalization, embeds a number of notable sub-cases such as the Gamma (for \(p = 1, \xi = 0\), the generalized inverse Gaussian (GIG, for \(p = 1\)), and the generalized Weibull (GW, for \(\xi = 0\)) density functions. Therefore, the orthogonal expansions based on the kernel defined above generalize the classical Laguerre expansions. To point out the great flexibility of this family of kernels in capturing tail behaviours that could not be achieved by a simple Gamma kernel, we will focus on the GIG and the GW kernels. Here, we discuss their theoretical properties in relation with consistency results provided in Theorem 2.4 and Theorem 2.6. In Section 4 we provide some numerical illustrations involving expansions based on the GIG and the GW kernels. Additionally, we consider the following log-normal (LN) kernel in our comparison

\[\phi(x) = \frac{1}{x} e^{-(\log(x) - \mu)^2}, \quad \mu \in \mathbb{R}, \sigma > 0.\]

An expansion on the underlying risk-neutral density based on the LN kernel (log-Hermite) is conceptually similar to an Hermite expansion on the logarithm of the underlying, which makes the LN kernel an interesting competitor of the GIG and the GW kernels, due to the great interest that Hermite expansions have received by many authors in the financial literature. The appeal of the LN kernel is enhanced by documented empirical evidence that the volatility, which is comparable to the VIX, is roughly log-normally distributed.

The behaviour on both left and right tails of the GIG, the GW and the LN kernels is informative on their ability to meet the condition \(\phi^{-1} f_q \in L^2(\mathcal{D})\), which is necessary for the convergence of \(f_q^{(n)}\) to \(f_q^{(\infty)}\), in view of Theorem 2.4. We observe that the LN kernel is certainly least demanding in terms of assumptions on the right tail behaviour of \(f_q\), while the GW kernel, whose decay rate is polynomial, imposes the smallest restriction on the left tail behaviour. Theorem 2.6 allows to assess the different impact that the tails of the three kernels have on the closure of related orthogonal polynomials: in particular, this property is always satisfied by the orthogonal polynomials related to the GIG kernel. On the contrary, for the orthogonal polynomials related to a GW kernel, closure is achieved only when the parameter \(p\) falls in the interval \([\frac{1}{2}, 1]\). Finally, the orthogonal polynomials related to the log-normal kernel never meet the closure condition, independently of the parameters choice. This means that the LN kernel is less capable than GIG and GW to recover \(f_q\), as intuitively the set of polynomial expansions in (2.2) does not cover a sufficiently large space to incorporate \(f_q\). Therefore, using the LN density could implicitly assign too much structure to the initial condition of the expansion, drastically reducing the flexibility of the approach. In Section 4 we illustrate the pitfalls of choosing a LN density as kernel function in the pricing of VIX options.

3 Estimation on market prices

In view of Theorem 2.4, the unique representation (2.2) of \(f_q\) in terms of the sequence \((c_k)_{k\in\mathbb{N}}\) can be seen as the explicit solution of a certain moment problem, where moments are the unknowns to be inferred from market data. In this section we outline a procedure to estimate the coefficients \(c_k\) of the expansion (2.2) by minimizing the distance between approximated and market price curves.

Remark 3.1 (Pricing formulas for vanilla options). Let \(\phi\) be a fixed probability density function whose support is contained in \([0, +\infty[\). Let us define, for every \(K \geq 0, n \in \mathbb{N} \) and \(c_1, \ldots, c_n \in \mathbb{R}\)

\[C^{(n)}_K(c_1, \ldots, c_n) := \int_K^{+\infty} \left( \sum_{k=1}^{n} c_k h^k_k \phi(x; \hat{\theta}) \right) (x - K) \, dx\]

\[P^{(n)}_K(c_1, \ldots, c_n) := \int_0^K \left( \sum_{k=1}^{n} c_k h^k_k \phi(x; \hat{\theta}) \right) (K - x) \, dx\]

where \(\hat{\theta}\) is the estimated vector of the kernel parameters. Then, for every \(n \in \mathbb{N}\) and every real vector \(c = (c_1, \ldots, c_n)^\top\), the expressions for the approximated prices \(C^{(n)}_K\) and \(P^{(n)}_K\) can be rewritten in the compact form

\[C^{(n)}_K(c) = A^{(K)}_0 + A^{(K)} c, \quad (3.1)\]

\[P^{(n)}_K(c) = B^{(K)}_0 + B^{(K)} c, \quad (3.2)\]

where

\[A^{(K)}_0 = \int_K^{+\infty} \phi(x) (x - K) \, dx, \quad B^{(K)}_0 = \int_0^K \phi(x) (K - x) \, dx,\]

and \(A^{(K)}\) and \(B^{(K)}\) are \(1 \times n\) vectors, whose \(i\)-th element is given by

\[A^{(K)}_i = \sum_{j=0}^{i} w_{i,j} \int_K^{+\infty} \left( x^{i+1} - K x^j \right) \phi(x) \, dx, \quad (3.3)\]

\[B^{(K)}_i = \sum_{j=0}^{i} w_{i,j} \int_0^K \left( K x^j - x^{i+1} \right) \phi(x) \, dx.\]
To obtain an estimate of the coefficients \(c_1, \ldots, c_n\), one can collect a cross-section of market prices, \(C_{m}^{\text{Obs}}(t, T)\) and \(P_{m}^{\text{Obs}}(t, T)\), for \(m = 1, \ldots, M\), then find the parameters \(\hat{c}_1, \ldots, \hat{c}_n\) that solve the problem
\[
[\hat{c}_1, \ldots, \hat{c}_n] = \arg\min_{c_1, \ldots, c_n} Q(t, T; c_1, \ldots, c_n). \tag{3.4}
\]

The objective function \(Q(t, T; c_1, \ldots, c_n)\) can be characterized as criterion function of a problem of minimum least squares for the following linear model
\[
Y_{\ast} = Xc + e. \tag{3.5}
\]

Since \(c_0 = 1\) by construction, the \(2M \times 1\) vector of dependent variables is defined as \(Y_{\ast} = Y - X_0\) with
\[
Y = [C_{K_1}^{\text{Obs}}(t, T), \ldots, C_{K_M}^{\text{Obs}}(t, T), P_{K_1}^{\text{Obs}}(t, T), \ldots, P_{K_M}^{\text{Obs}}(t, T)]',
\]
and
\[
X_0 = [A_0^{K_1}, \ldots, A_0^{K_M}, B_0^{K_1}, \ldots, B_0^{K_M}],
\]
while \(e\) is the \(2M \times 1\) vector of error terms and
\[
X = \begin{bmatrix}
A_1^{(K_1)} & \ldots & A_n^{(K_1)} \\
\vdots & \ddots & \vdots \\
A_1^{(K_M)} & \ldots & A_n^{(K_M)} \\
B_1^{(K_1)} & \ldots & B_n^{(K_1)} \\
\vdots & \ddots & \vdots \\
B_1^{(K_M)} & \ldots & B_n^{(K_M)}
\end{bmatrix}
\]
is a \(2M \times n\) matrix.

The objective function has the following quadratic form
\[
Q(t, T; c_1, \ldots, c_n) = (Y - Xc)'(Y - Xc). \tag{3.6}
\]
which, in turns, allows for a closed-form solution for the vector of coefficients given by
\[
\hat{c} = (X'X)^{-1}X'Y \tag{3.7}
\]
if \(X\) has full column rank. Unfortunately, as \(n\) increases, the columns of \(X\), which are functions of the first non-standardized \(n\) moments, become more and more collinear by construction, thus making \(X\) close to singular and the OLS estimation unfeasible. To tackle the problem of multicollinearity we make use of a procedure based on the orthogonalization of regressors, which allows to estimate the coefficients of an expansion of any arbitrary large \(n\) order.

### 3.1 Orthogonal regressors

We propose a method that allows to choose a large \(n\), e.g. 15, but to reduce the dimensionality of \(X\) by orthogonalizing the matrix of regressors by means of principal components. This solves the curse of multicollinearity which typically arises when \(n > 4\). To avoid scale effects, we first standardize each column of \(X\), as
\[
Z_i = \frac{X_i - \xi_i}{\sigma_i}, \quad i = 1, \ldots, n \tag{3.8}
\]
with \(\xi_i = \frac{1}{2M} \sum_{j=1}^{2M} X_{ji}\) and \(\sigma^2 = \frac{1}{2M-1} \sum_{j=1}^{2M} (X_{ji} - \xi_i)^2\), where \(X_{ji}\) is the \(j, i\)-th element of matrix \(X\). We then apply the PCA to \(Z\) and obtain the \(2M \times n\) matrix of orthogonal principal components \(V = PZ\). The \(n \times n\) matrix of weights, \(P\), is obtained from the eigenvectors associated to the eigenvalues of the covariance matrix of \(Z\). Given a threshold on the explained total variance (e.g. 99%), we extract the sub-matrix with the first \(s\) principal components, \(\tilde{V} = V_{1:s}\), to be used as regressors. Typically, the first 4-5 principal components explain at least the 99% of the total variance when \(n > 10\). We then estimate the coefficients of the following regression,
\[
\tilde{Y}_{\ast} = \tilde{V}\beta + u \tag{3.9}
\]
where \(\beta\) is an \(s \times 1\) with the loadings of the first \(s\) principal components, while the estimated coefficients \(\hat{c}\) are retrieved as
\[
\hat{c} = (\tilde{O}\hat{\beta}) \circ \sigma^{-1}
\]
where \(\tilde{O}\) is the \(n \times s\) matrix of the first \(s\) columns of the inverse of the PCA weights, \(\circ\) denotes the Hadamard product, and \(\sigma = [\sigma_1, \ldots, \sigma_n]'\) is an \(n \times 1\) vector.

### 3.2 Regression through the origin

In regression (3.9), there is no intercept and the columns of \(\tilde{V}\) have zero-mean by construction, while \(Y_{\ast}\) have zero mean if and only if the sample mean of \(Y\) and \(X_0\) coincide. To enforce that \(E(Y_{\ast}) = E(Y - X_0) = 0\), the initial estimation of the kernel parameters, \(\hat{\theta}\), must be constrained such that \(E(X_0) = E(Y)\). This has several advantages. First, it ensures that the approximation of order 0 does not produce systematic mispricing since the observed market prices are centered around the estimated price curve generated by the kernel. Second, the residuals of regression (3.9) have zero mean for any order \(n > 0\) by construction. Third, all the expansion terms for \(n \geq 1\) operate as a correction of the pricing expansion around \(X_0\). Fourth, it enforces identification of the kernel parameters.

Since the principal components are constructed from the standardized regressors, \(Z\), when remapping the parameters of equation (3.9) to the ones in equation (3.5), a constant term equal to \(\sum_{i=1}^{n} \frac{\xi_i}{\sigma_i} R_{i1}\), with \(R = \tilde{O}\beta\), appears. Therefore, in order to guarantee that the relation in equation (3.5) holds for any \(n \geq 1\), the following constrained optimization is performed...
\[ [\beta_1, \ldots, \beta_s] = \arg\min_{\beta_1, \ldots, \beta_s} \tilde{Q}(t, T; \beta_1, \ldots, \beta_s), \quad (3.10) \]

where \( \tilde{Q}(t, T; \beta_1, \ldots, \beta_s) = (Y - \tilde{\nu})'(Y - \tilde{\nu}) \). This restriction guarantees that also \( \hat{\epsilon} \) are such that there is no systematic pricing error, given that the residuals, \( \hat{\epsilon} = Y - X_0 - \lambda \hat{c} \), are centered in zero for any \( n \).

### 3.3 Kernel displacement

Consistency conditions ensured by Theorems 2.4-2.6 are rather flexible with respect to the kernel support. In principle, kernels whose support strictly contains the support of \( f_Q \) are also admitted. However, in the latter case, the expansion (2.2) is "forced" to converge to zero on all points that are outside the support of \( f_Q \). This necessary condition impacts on problem (3.10) as an additional constraint, meaning that it negatively affects the optimality of its solution. The argument holds true when the left tail of \( f_Q \) is particularly short and thick, implying that nearly the whole mass underlined by \( f_Q \) is concentrated away from the origin. When \( f_Q \) displays such a behavior, in practice we find it more convenient to displace the kernel domain, so that its support is bounded away from zero. The need of this displacement is somehow concerned with possible negative skewness enjoyed by the \( f_Q \), which reflects in how quickly the price curves become linear as strikes become OTM and OTM. For the VIX, we can observe that the convexity of OTM options tends to zero outside the support of \( f_Q \). This behavior suggests that the RND implied by the VIX is negative skewed, which is confirmed by our experience. To include a support displacement in our routine we proceed as follows:

1. We determine a value \( K_{\text{min}} \) such that the mass outside \([0, K_{\text{min}}]\) can be neglected.
2. We compute the matrix of regressors \( X \) as a function of a new set of shifted strikes \( K_1 - K_0, \ldots, K_M - K_0 \).
3. We find the optimal solution \( \tilde{\sigma} \) of the problem (3.10) with respect to \( K_1 - K_0, \ldots, K_M - K_0 \).
4. We determine \( f_Q^{(n)}(x) \) as follows

\[
f_Q^{(n)}(x) = \phi(x - K_{\text{min}}) \left( 1 + \sum_{k=1}^n \tilde{c}_k h_k^\phi(x - K_{\text{min}}) \right). \quad (3.11)
\]

### 4 Accuracy and robustness

In this section, we test the accuracy of the proposed approach by means of two numerical examples. The purpose is to estimate the risk-neutral density on option prices generated by structural models for which \( f_Q \) is known in closed-form. Specifically, we consider two models that are popular in the option pricing literature. In the first case, the VIX is determined by a function of the instantaneous variance process \( \nu(t) \) of the Heston model, as explained in Sepp (2008b). In the second case, the risk-neutral density of \( \log(VIX(t)) \) is assumed normal inverse Gaussian (NIG), as studied in Huskaj and Nossman (2013). In both cases, the estimation of \( c_1, \ldots, c_n \) is performed according to the methodology outlined in Section 3 on a set of 42 call and put prices relative to strike prices in the interval [10, 55].

The option prices to be matched are generated through direct integration of the RND \( f_Q \) implied by the two models. The expansions order is set to 20, which is sufficiently high to ensure that the fitting cannot be further improved by adding more terms to the expansion. Furthermore, choosing a high order for the expansion illustrates the convergence and stability properties of the approach. Interestingly, the risk-neutral densities implicit in the above models decay rates that are respectively exponential (convergent case) and polynomial (non-convergent case). This exercise provides valuable information on the robustness of the estimates to the initial choice of the expansion kernel \( \phi \). In particular, although the asymptotic properties of \( \phi \) have a tangible effect on the accuracy of the approximation, the initial calibration of the parameters of \( \phi \) has only a marginal impact, provided that it guarantees that convergence and closure conditions are respected (see Theorems 2.4-2.6).

### 4.1 Heston model

According to Sepp (2008b), the Heston model implies that, for a fixed maturity \( T \), \( VIX(T) \) can be expressed as a function of the instantaneous variance \( \nu(T) \):

\[
VIX(T) = 100 \cdot (a_1 \cdot \nu(T) + a_2)^{\frac{1}{2}}.
\]

Moreover, the density of \( \nu(T) \) given \( \nu(t) = z \) has the following closed-from expression

\[
(v_T | v_t = z) \sim g, \quad g(s) = C_1 s^{\frac{1}{2} - \frac{1}{2}} e^{-\frac{1}{2} \theta s} \sqrt{2 \pi I_{\nu}^{\frac{1}{2}}(C_2 \sqrt{s})},
\]

where \( C_1 = \frac{\theta}{\sqrt{2 \pi I_{\nu}^{\frac{1}{2}}(C_2 \sqrt{s})}} \) and \( C_2 = 2C_1 \sqrt{2 \pi I_{\nu}^{\frac{1}{2}}} \). The support of \( f_Q \) is \([\sqrt{\nu_0}, +\infty]\) and its right tail decays as \( e^{-x^2} \). Moreover, whenever the support of \( \phi \) strictly contains \([\sqrt{\nu_0}, +\infty]\), the decay of \( f_Q \) on the left tail does not influence the integrability of \( f_Q^2 \phi^{-1} \). Therefore, the convergence condition of Theorem 2.4 is met for any choice of the kernel density in the family of GIG, GW and LN.
Figure 4.1 portrays the true risk-neutral density of the Heston model and the related orthogonal polynomial expansions based on different choices of the kernel. The approximated densities reported in Figure 4.1 highlight the ability of the expansions based on the GIG and the GW kernels to well recover the original density $f_Q$. On the contrary, the LN kernel does not well approximate $f_Q$ although several corrective terms are considered in the expansion and the convergence condition $f_Q \phi^{-1/2} \in L^2(D)$ is satisfied. The expansion based on the LN kernel proves particularly unable of capturing the tails features of $f_Q$. This is a practical consequence of the fact that the set $(x^k)_{k \in \mathbb{N}}$ is not closed with respect to a LN density (see Theorem 2.6), i.e. it doesn’t span a space sufficiently large to incorporate functions that are “too dissimilar”. Hence, the poor fitting obtained by the expansion based on the LN kernel is an illustrative example on the importance that the conditions stated in both Theorem 2.4 and Theorem 2.6 must be satisfied by the kernels.

4.2 NIG distribution for $\log(\text{VIX}(t))$

We assume now that the logarithm of the VIX follows a NIG distribution. Specifically, we assume that

$$
\log(\text{VIX}(t)) \sim g, \quad g(s) = C \cdot \frac{K_1(\alpha \sqrt{\beta^2 + (s-\mu)^2})}{\sqrt{\beta^2 + (s-\mu)^2}} e^{\delta(s-\mu)},
$$

where $C = \frac{\alpha}{\pi} e^{\delta \gamma}$ is the normalization constant and $K_v$ denotes the modified Bessel function of the second kind (cfr Abramowitz and Stegun (1964)). Therefore, by the change of variable $s = \log(x)$ the risk-neutral density of VIX is

$$
f_Q(x) = \frac{1}{x} \cdot g(\log(x)).
$$

The asymptotic properties of $K_v$ determine a polynomial decay of $f_Q$ both on the right and the left tails. It follows that the condition $f_Q \phi^{-1/2} \in L^2(\mathbb{R})$ is never met by any of the kernel densities discussed so far.

Figure 4.2 reports the true risk-neutral density of the NIG model and the related orthogonal polynomial expansions based on different choices of the kernel. As expected, the main problems of convergence appear in the tails. In particular, the expansion based on the GIG kernel is not accurate on both left and right tails: this is consistent with the fact that a GIG kernel decays exponentially at both sides. The GW-based expansion, though, is inexact only on the right tail. Again, this is consistent with the theory, as the GW kernel decays exponentially towards $+\infty$ and as a polynomial in the origin. Finally, the LN kernel provides again the weakest performance, but it is worth noticing that here the approximation is more accurate than in the previous test. This can be explained by the fact that the normal inverse Gaussian includes and generalizes the Gaussian distribution, which means that here $f_Q$ is “closer” to a log-normal than in the previous case.

4.3 Robustness to initial conditions

We study here the robustness to the initial calibration of the parameters of the kernel. In particular, all kernels that intervene in Figure 4.1 and in Figure 4.2 were initially calibrated to match the mean and the variance of $f_Q$. However, it is interesting to empirically assess how the initial choice of the parameters of $\phi$ affects the accuracy of (2.2). To answer this question, we perturb the parameters of well calibrated kernels, so that the moments and the option prices implied by the kernels heavily mismatch those generated by the true $f_Q$. In Table 4.1 we report the first four moments of the GIG and GW kernels when the mean and variance parameters of the latter are drastically perturbed compared to the values implied by the true density of the Heston model. The

Figure 4.2. Probability density functions in standard scale (left) and semi-logarithmic scale (right). Comparison between the true density of VIX implied by the Heston model (solid line) and the polynomial expansions of order 20 calibrated on the option prices generated by the Heston model. The expansions are based on: GIG kernel (red-asterisks), GW kernel (yellow-stars) and LN kernel (green-crosses). Parameters for the Heston variance process are: $k = 1.71$, $\theta = 0.097$, $\eta = 0.577$, $\psi(0) = 1$ and $T - t = 30/365$. The dashed vertical lines on the right panel identify several relevant probability levels and the corresponding quantiles.
Figure 4.2. Probability density functions in standard scale (left) and semi-logarithmic scale (right). Comparison between the true density of VIX implied by the NIG density (solid line) and the polynomial expansions of order 20 calibrated on the option prices generated by the NIG model. The expansions are based on: GIG kernel (red-asterisks), GW kernel (yellow-stars) and LN kernel (green-crosses). Parameters for the NIG density are chosen as follows: $\alpha = 29.3$, $\beta = 26.4$, $\mu = 2.66$, $\gamma = 0.11$. The dashed vertical lines on the right panel identify several relevant probability levels and the corresponding quantiles.

<table>
<thead>
<tr>
<th></th>
<th>True</th>
<th>GIG ker.</th>
<th>GW ker.</th>
<th>GIG exp.</th>
<th>WB exp.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>30.13</td>
<td>27.65</td>
<td>35.44</td>
<td>30.14</td>
<td>30.17</td>
</tr>
<tr>
<td>Var</td>
<td>65.36</td>
<td>50.79</td>
<td>165.78</td>
<td>65.27</td>
<td>65.81</td>
</tr>
<tr>
<td>Skew</td>
<td>50.26</td>
<td>21.07</td>
<td>56.86</td>
<td>50.16</td>
<td>50.40</td>
</tr>
<tr>
<td>Kur</td>
<td>2.86</td>
<td>0.80</td>
<td>6.07</td>
<td>2.82</td>
<td>2.87</td>
</tr>
</tbody>
</table>

Table 4.1. The table reports mean, variance, standardized skewness, and kurtosis of the true density of the Heston model, of the calibrated kernel densities (GIG kernel and GW kernel) and of their related expansions of order larger than 5.

The last two columns of the table highlight the capability of the polynomial expansions to provide a precise fitting of the moments of the true density even when the initial conditions on the kernels are subject to large perturbations. It is inherently assumed, though, that convergence conditions stated in Theorems 2.4-2.6 are always preserved.

Figure 4.3 shows that the choice of $\phi$ influences the accuracy of the calibration of the polynomial expansion only up to a minimum extent. In other words, the choice of the kernel is fundamental only in order to correctly match the required asymptotic decay of the tails of $f_G$, while it doesn’t directly affect the fitting accuracy once the conditions of convergence have been guaranteed.

5 An example with real data

In this section we illustrate the reliability of our methodology by means of an estimation exercise on real VIX option prices. Our data sample consists of VIX option prices observed on November 16, 2011 and expiring on December 21, 2011. The option data is acquired from the OptionMetrics database. November 2011 is a period of particular interest since extremely high levels of market volatility were registered in connection with the sovereign debt crisis. As a consequence, on that day, available VIX option strikes also spiked and settled in a strip between 15 and 90 dollars. Under normal market conditions, VIX option strikes typically fall between 10 and 45. Moreover, due to mean-reversion of volatility, this range is normally expected to remain constant over time, resulting in poor liquidity of deeply OTM and ITM options, as discussed below. Contrariwise, on November 16, 2011, VIX options with maturity December 21, 2011 enjoyed trading volumes that were sufficiently high to provide reliable market prices for OTM and ITM contracts. This data can be used to obtain an estimate of the RND that is robust from an econometric point of view.

The estimate is obtained by finding the optimal solution $\hat{c}_1, \ldots, \hat{c}_s$ of (3.10) with respect to $\beta_1, \ldots, \beta_s$. The sample size for the estimation is $2M = 52$ out of a total number of 64 contracts, from which we symmetrically excluded all options priced below 0.05 dollars. Therefore, the options used in the estimation range between $K_0 = 21$ and $K_M = 80$. Moreover, since all put options with strikes below $K_{min} = 18$ share the same price of 0.03 dollars, which is below the bid-ask spread, we assume that the RND encloses negligible mass within $[0, K_{min}]$ by adopting the displacement procedure described in Section 3.3. We additionally find ex-post that $K_{max} = 140$ is a sufficiently large value to exclude only negligible mass from the interval $[K_{min}, K_{max}]$. Consistently with the notation of Section 3 we indicate by $K_1, \ldots, K_M$ the strikes contained in our sample and by $n$ the expansion order. The correctness of our procedure is confirmed by excellent fitting that we obtain on both option prices and implied volatilities for $n = 18$, as portrayed in Figure 6.2. Furthermore, the two volatility curves implied by observed call and put options, which are reported in Figure 6.2, display almost identical values, mean-
ing that no evident mispricing affects the data used in our estimation. Differently, the volatilities implied by market option prices that are not considered in the estimation exhibit anomalous behavior. In both cases, this is a consequence of low trading volumes and justifies our exclusion criteria.

From the price residuals displayed in Figures 6.3 and 6.4, it clearly emerges that the GIG kernel alone does not provide to be optimal to fit the prices as the residuals are still very high and display a large degree of correlation across strikes. Increasing the expansion to $n = 4$ and using OLS improves over the fitting of the kernel, but it is still not optimal, as the residuals still display some degree of persistence. Since we cannot rely on OLS based on an expansion order larger than $n$ due to the multicollinearity problem discussed above, we use the PCA method to reduce the problem dimensionality for all choices of $n > 4$. The residuals with $n = 9$ and $n = 18$ already display a large improvement in terms of variability of the residual compared to the kernel. This means that the moments higher than the fourth have some explanatory power in determining the risk neutral density and hence the market prices. Moreover, the correlation across consecutive strikes is largely reduced although not completely eliminated. This is the main reason for adopting a block bootstrap scheme in computing the $p$-values of the diagnostic tests.

Figure 6.1 reports the estimated density with $n = 18$ and the GIG kernel, whose parameters $\theta = [\alpha, \beta, \gamma]$ are found by minimizing the squared residuals at a first stage of the estimation routine outlined in Section 3. The shapes of the two densities are quite different, especially in the right tail. In particular, we observe that the RND based on the expansion of order 18 assigns a higher probability to tail-events than the simple kernel fitted to option prices. The left tails of the two densities are rather similar and both suggest that a RND that is consistent with VIX option data should decay very quickly and far from the origin. This confirms that expanding the RND on a displaced support is a good practice to approximate this behavior. The huge difference in fitting performances obtained by the kernel alone and the expansion of order 18 can be appreciated by looking at the implied volatility curves, that are reported in Figure 6.2. In general, fitting the VIX implied volatilities through structural pricing proves to be a difficult task. As compared to the performance of structural pricing techniques (see e.g. Bayer et al. (2013) and Kokholm and Stisen (2015)), we achieve much higher accuracy. However, differently from our technique, structural methods are not purely targeted at fitting a cross-sectional set of implied volatilities but are equally addressed to replicate the evolution over time of one or more assets. Therefore, we may conclude that when there is no interest in the transition between states over time, our technique is generally more suited to estimate the RND than structural models.
6 List of figures

Figure 6.1. GIG kernel and expansion-based estimated densities.

Figure 6.2. Implied volatility curves implied by market observations, GIG kernel and orthogonal expansion of order $n = 18$

Figure 6.3. Call residuals
References


