DEPARTMENT OF ECONOMICS
AND BUSINESS ECONOMICS
AARHUS UNIVERSITY

CREATES
Center for Research in Econometric Analysis of Time Series

A Non-Structural Investigation of VIX Risk Neutral Density

Andrea Barletta, Paolo Santucci de Magistris and Francesco Violante

CREATES Research Paper 2017-15

Department of Economics and Business Economics
Aarhus University
Fuglesangs Allé 4
DK-8210 Aarhus V
Denmark

Email: oekonomi@au.dk
Tel: +45 8716 5515

A Non-Structural Investigation of VIX Risk Neutral Density *

Andrea Barletta† Paolo Santucci de Magistris ‡ Francesco Violante §

March 31, 2017

Abstract

We propose a non-structural pricing method to derive the risk-neutral density (RND) implied by options on the CBOE Volatility Index (VIX). The methodology is based on orthogonal polynomial expansions around a kernel density and yields the RND of the underlying asset without the need for a parametric specification. The classic family of Laguerre expansions is extended to include the GIG and the generalized Weibull kernels, thus relaxing the conditions required on the tail decay rate of the RND to ensure convergence. We show that the proposed methodology yields an accurate approximation of the RND in a large variety of cases, also when the no-arbitrage and efficient option prices are contaminated by measurement errors. Our empirical investigation, based on a panel of traded VIX options, reveals some stylized facts on the RND of VIX. We find that a common stochastic factor drives the dynamic behavior of the risk neutral moments, the probabilities of volatility tail-events are priced in the options as jumps under the risk-neutral measure, and the variance swap term structure depends on two factors, one accounting for the slope and one for the mean-reverting behavior of the VIX.

Keywords: VIX options, orthogonal expansions, risk-neutral moments, volatility jumps, variance swaps

JEL Classification: C01, C02, C58, G12, G13 .

*We thank Fulvio Corsi, Friedrich Hubalek, Elisa Nicolato and Viktor Todorov for useful comments and suggestions. We also thank the participants of CREATES seminar series, the CFE 2015 conference in London, the Empirical Finance workshop in Paris, the VCMF 2016 conference in Vienna and the HAX 2017 conference in Aarhus. The research leading to these results received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement No. 289032 (HPCFinance). Paolo Santucci de Magistris and Francesco Violante acknowledge the research support of CREATES, funded by the Danish National Research Foundation (DNRF78). Francesco Violante also acknowledges the research support of the Danish Council for Independent Research, Social Sciences, Individual Research grant and Sapere Aude Research Talent grant.

†Department of Economics and Business Economics, Aarhus University, Denmark. abarletta@econ.au.dk

‡CREATES, Department of Economics and Business Economics, Aarhus University, Denmark. psantucci@econ.au.dk

§CREATES, Department of Economics and Business Economics, Aarhus University, Denmark. fviolante@econ.au.dk

1 Introduction

The Volatility Index (VIX) was introduced in 1993 by the 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 the growing interest in volatility trading, in 2004 the CBOE introduced the VIX futures alongside a revised formulation of the VIX which was based on the replication of variance swap contracts written on the broader S&P 500 (SPX) index. Specifically, in its current formulation, see CBOE (2015), the VIX is computed as the present value of a portfolio of SPX call and put options constructed as a static replication of a 30-days variance swap. In 2006, options written on the VIX also started trading. Since then, several authors have studied the pricing of VIX options. The main strand of literature addresses VIX derivative pricing under stochastic volatility models, mostly within the affine class. This branch is pioneered by Zhang and Zhu (2006) and Zhu and Zhang (2007), who derive dynamics for the VIX starting from a square-root model for the spot variance. The works of Sepp (2008a,b) extend this approach by introducing jumps in the spot variance within the affine jump-diffusion (AJD) framework of Duffie et al. (2000). The recent paper by Bardgett et al. (2014) further generalizes the framework of Sepp (2008a,b) by allowing for a stochastic long-run mean of variance. Non-affine pure-diffusion extensions of the square-root model for the spot variance are in Gatheral (2008) and Bayer et al. (2013). Finally, modeling frameworks based on infinite-dimensional specifications of the variance swap term-structure are proposed by Buehler (2006), Bergomi (2008), and Cont and Kokholm (2013).

A common feature of these contributions is that the risk-neutral density (RND) is assumed to be fully described by stochastic dynamic equations of the state-variables, which are functions of the underlying model parameters. Unfortunately, fully parametric specifications of the dynamics of price and volatility come at the cost of an intrinsic risk of misspecification, see for example Cont (2006) for a discussion. The problem of correct model specification in VIX option pricing is particularly troublesome since the linkage between VIX and SPX is not fully explicit, and they both depend on the variance, which is an unobservable quantity. Even when modeling is solely addressed to the marginal density of the VIX, a comparative analysis of the performance of simple stochastic volatility models in pricing of VIX options tends to confirm the potential issue of misspecification. 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 none of these models achieve small pricing errors over the entire range of strike prices. Moreover, the econometric analysis carried out by Mencia and Sentana (2013) reveals that the risk of model misspecification in structural pricing of VIX options is particularly high during financial crises. This reflects the general disagreement in the literature on the "nature" and the roughness of the instantaneous volatility. In particular, although the instantaneous volatility is most commonly modeled as a jump-diffusion process, Todorov and Tauchen (2011) find that it is best described by a pure jump process, with clear consequences on the VIX index

and its related options. Reducing the model risk concerned with VIX option pricing is possible but often comes at the cost of analytical tractability and availability of closed-form solutions. Thus, modeling frameworks conceived for capturing stylized facts of the VIX are rarely suited for estimation purposes.

In this view, non-structural methods for estimating the RND directly from VIX options represent a viable alternative to stochastic modeling. The term "non-structural" is referred, in general, to any option pricing method that does not rely on the postulation of a specific parametric expression for the RND. This entails considerable reduction of the risk associated with misspecification. The idea that vanilla option prices can be linked to the RND through an explicit non-structural relation was pioneered by Breeden and Litzenberger (1978). In the context of VIX option pricing, a non-structural technique has recently been employed by Song and Xiu (2016) with the purpose of estimating the volatility pricing kernel, which is the ratio between the physical and the risk-neutral density of the VIX. More precisely, as in many other works addressing non-structural option pricing, the authors propose to retrieve the RND by inferring the second-order derivatives of option prices with respect to strikes directly from the market. The non-structural approach proposed in this paper does not directly consider the second-order derivatives of option prices, but it recovers the RND by means of an orthogonal polynomial expansion around a kernel density, see for instance Szegö (1939). Classic examples of orthogonal expansions are the Hermite, which are obtained when the kernel is a Gaussian density, and the Laguerre, which are obtained when the kernel is an exponential density. The key feature of orthogonal expansions is that they yield an explicit functional form of the RND without the need of specifying stochastic dynamics of the state-variables. Instead, this method imposes mild integrability conditions on the form of the RND, proving to be particularly robust to misspecification. There is extensive literature on the use of orthogonal expansions in financial applications. Seminal examples are Jarrow and Rudd (1982), Corrado and Su (1996b), Madan and Milne (1994), Coutant et al. (2001), and Jondeau and Rockinger (2001), while more recent contributions are Rompolis and Tzavalis (2008), Zhang et al. (2011), Ñíguez and Perote (2012), and Xiu (2014). In all these cases, the expansions are provided in terms of Hermite polynomials. Our methodology can be thought of as an alternative of Hermite expansions to the case of kernels with positive support. Indeed, adapting the expansion kernel to the data (for instance by choosing a kernel with support on the positive axis only) may be a better alternative to the inverse approach of adapting the data to the kernel (for instance by log-change or standardization). In particular, we extend the Laguerre expansions, used recently by Filipovic et al. (2013) and Mencia and Sentana (2016), by introducing a family of kernels that encompasses well known distributions such as the exponential, the Gamma, the Weibull and the GIG, among others. We show that the introduction of the extended Laguerre (eLaguerre) kernels increases the adaptability of the approach by reducing the number of restrictions to be imposed on the form of the RND.

We contribute to the VIX option pricing literature on several aspects. First, we provide general convergence conditions of orthogonal expansions to the true RND. These conditions

relate to the rate of tail decay of the expansion kernel. We show that the log-normal density, due to its slow tail decay rate, does not represent a suitable candidate for the expansion kernel as this would generally lead to inaccurate approximations. Instead, our extended Laguerre kernels are better suited to approximate the RND associated with the VIX options, due to the very flexible decay rate on both tails. Indeed, owing to the irregular nature of the instantaneous volatility, see Todorov and Tauchen (2011) and Todorov et al. (2014), the tails of the RND of the VIX are expected to display features that can only be captured if a flexible choice of the kernel density is adopted. Second, in the spirit of Aït-Sahalia and Lo (1998), Jondeau and Rockinger (2001) and Aït-Sahalia and Duarte (2003), we propose a robust methodology to estimate the parameters of the polynomial expansion by a minimum distance criterion based on the observed option prices. We prove that the proposed methodology yields a very accurate approximation of the RND also when the no-arbitrage and efficient option prices are contaminated by measurement errors. Although this paper focuses on VIX options, our methodology is outlined in full generality, and hence it can be applied to any option to recover the implied RND.

The analysis of the RND is carried out on a panel of VIX options collected at monthly frequency for the period January 2010 - April 2016. The results highlight the reliability of our methodology to recover RND up to negligible rounding errors and to mimic relevant quantities embedded in the VIX options, such as the volatility of volatility index, i.e. VVIX. The time series of the first four risk-neutral moments of VIX display some interesting clues about volatility expectations. First, we discover the presence of a positive correlation between mean and variance of VIX (*reversed* leverage effect) as well as a common factor structure across moments and times-to-maturity. Second, by fitting a multiplicative error model (MEM) on the time series of the VIX risk-neutral moments, we find strong empirical evidence in favour of a non-negligible volatility jump probability under $\mathbb{Q}$ that is priced in the options. Third, the variance swap term structure is empirically studied. The variance swap term-structure implicit in the VIX second moments is coherent with the one directly computed from SPX options, proving that the two markets are consistent with each other. This result entails that the second moment of VIX can be traded through a combination of long and short positions on SPX options. Finally, we show that it is crucial to account for the mean-reverting behavior of the realized variance to describe the term-structure with a constant slope term.

The paper is organized as follows. Section 2 recaps the VIX formula and defines its RND, while Section 3 summarizes some properties of orthogonal polynomial expansions. Section 4 discusses the estimation procedure based on principal components regression of the expansion coefficients, under additional consistency constraints. Section 5 addresses whether and how the estimated RND is affected by option prices contaminated by measurement errors associated with no-arbitrage violations. Finally, Section 6 presents the empirical applications with real data. Appendix A contains the proofs of the theorems related to the orthogonal polynomials, while Appendix B contains some supplementary material with details on the fitting procedure, the numerical experiments and the robustness to no-arbitrage violations.

2 The VIX index

Introduced in 1993 by the Chicago Board Options Exchange (CBOE), the VIX is a risk-neutral forward measure of market volatility calculated as the fair value of a 30-days variance swap rate. Futures and options on the VIX are the first derivatives on volatility to be traded on a regulated security exchange, see Carr and Wu (2006) and Carr and Lee (2009) for a detailed historical review of the VIX. At the core of the VIX construction, which is detailed in CBOE (2015), lies the model-free methodology of Carr and Madan (1998), Britten-Jones and Neuberger (2000), and Jiang and Tian (2005), to price a variance swap via static replication. According to this methodology, under standard risk-neutrality and no-arbitrage assumptions, the s-variance swap rate at time t, $\mathrm{VS}_{t,s}$, can be computed as

$$\mathrm{VS}_{t,s} = -\frac{2}{s} E_t^{\mathbb{Q}} \left[\log \left(\frac{\mathrm{SPX}_{t+s}}{x_F} \right) \right], \tag{1}$$

where $\mathbb{Q}$ denotes the risk-neutral pricing measure and x_F is the SPX forward price observed at time t for maturity $t+s$. The right-hand term in (1) can be replicated by the following portfolio of SPX calls and puts

$$-\frac{2}{s} E_t \left[\log \left(\frac{\mathrm{SPX}_{t+s}}{x_F} \right) \right] = \frac{2e^{-r \cdot s}}{s} \left(\int_0^{x_F} \frac{1}{x^2} P^{\mathrm{SPX}}(x) dx + \int_{x_F}^{\infty} \frac{1}{x^2} C^{\mathrm{SPX}}(x) dx \right), \tag{2}$$

where r is the risk-free rate and $C^{\mathrm{SPX}}(x)$ and $P^{\mathrm{SPX}}(x)$ denote prices of SPX call and put options observed at time t, expressed as functions of the strike x. The square of VIX is defined by fixing s to 30 days and by discretizing the infinite strip of out-of-the-money options in (2) over the finite set of available strikes

$$\frac{\mathrm{VIX}_t^2}{100} = \frac{2e^{-r \cdot s}}{s} \left(\sum_{x_i \leq x_0} \frac{\Delta x_i}{x_i^2} P^{\mathrm{SPX}}(x_i) + \sum_{x_i \geq x_0} \frac{\Delta x_i}{x_i^2} C^{\mathrm{SPX}}(x_i) \right) + \frac{1}{s} \left[\frac{x_F}{x_0} - 1 \right]^2, \tag{3}$$

where Δx_i is half the distance between x_{i+1} and x_{i-1}, x_0 is the greatest available strike below x_F. In practice, when options with maturity exactly equal to s are not available, the VIX is calculated by interpolating the values of (3) calculated by using SPX options with the largest available maturity below 30 days (near term) and the smallest available maturity above 30 days (next term). We should also remark that, in principle, formula (1) is not exact when the trajectories of the SPX are subject to jumps. However, as pointed out by Carr and Wu (2009), the contribution of price jumps in the computation of (1) can be neglected in practice.

The VIX index typically exhibits high negative correlation to the stock market, thus proving an effective tool to hedge (or leverage) volatility separately from directional price moves. As a matter of fact, many investors deem the VIX to be a leading indicator of market sentiment - the index is often referred to as the market *fear gauge*. This is reflected by high trading volumes of VIX options, which stand at approximately 37% of the average daily volume of SPX options, as

noticed by Mencia and Sentana (2013). Under risk-neutrality and no-arbitrage assumptions, for fixed observation time t, time-to-maturity τ, and strike K, the price of a VIX call option $C_K(t,\tau)$ and the price of a VIX put option $\mathcal{P}_K(t,\tau)$ are given by

$$C_K(t,\tau) = e^{-r\tau}\int_0^\infty f_{\mathbb{Q}}(t,\tau;x)(x-K)^+dx, \quad \mathcal{P}_K(t,\tau) = e^{-r\tau}\int_0^\infty f_{\mathbb{Q}}(t,\tau;x)(K-x)^+dx\,, \quad (4)$$

where $f_{\mathbb{Q}}(t,\tau;\cdot)$ is the conditional RND of $\text{VIX}_{t+\tau}$, given all the past information up to time t. When not misleading, we will omit the dependence of the RND on t,τ and simply denote it by $f_{\mathbb{Q}}$. The RND of the VIX is the object of interest in this paper. In the next section, we discuss a robust methodology to retrieve $f_{\mathbb{Q}}$ directly from the VIX option prices.

3 Orthogonal polynomials

The methodology that we adopt builds upon the following expansion of the RND

$$f_{\mathbb{Q}} \approx f_{\mathbb{Q}}^{(n)}(x) := \phi(x)\left(1+\sum_{k=1}^{n} e_k(x)\right), \qquad n \geq 1\,, \quad (5)$$

where ϕ is a chosen probability density function (kernel) and e_k are corrective factors with $e_0 = 1$. The kernel function ϕ can be seen as the 0-order term in (5)

$$f_{\mathbb{Q}}^{(0)}(x) := \phi(x)\,,$$

and it can be interpreted as an initial proxy for the RND. In this work, we consider expansions where the corrective factors $e_1,\ldots,e_n$ take the following form

$$e_k(x) = c_k h_k^{\phi}(x),$$

where, for every $k = 1,\ldots,n$, c_k is a real constant and h_k^{ϕ} is a polynomial of degree k in the state variable x. The polynomials $h_1^{\phi},\ldots,h_n^{\phi}$ only depend on the kernel ϕ, and therefore the coefficients $c_1,\ldots,c_n$ embed all the information on the RND.

The consistency of this approach relies on two results of functional analysis discussed in Sections 3.1 and 3.2. The first result ensures that if $h_1^{\phi},\ldots,h_n^{\phi}$ are orthogonal to each other, in a sense that will be clarified below, then by construction the resulting approximate density $f^{(n)}$ enjoys unitary mass and other desirable properties regarding its moments. The second result reveals that the relation expressed in (5) is non-structural, meaning that no specific form of the RND needs to be postulated in order to ensure that (5) is an admissible expansion, i.e., that $f_{\mathbb{Q}}^{(n)} \to f_{\mathbb{Q}}$ as $n \to \infty$. The results in Sections 3.1 and 3.2 are obtained under the maintained assumption that ϕ is a probability density function with support $\mathcal{D} \subseteq \mathbb{R}$ and possessing finite

polynomial moments, that is

$$\int_{\mathcal{D}} |x|^k \phi(x) dx < +\infty, \quad \forall k \in \mathbb{N}.$$

3.1 Properties of orthogonal polynomials

The elements of the family $(h_k^\phi)_{k\in\mathbb{N}}$ are said to be *orthogonal polynomials* with respect to ϕ if, for all $k \in \mathbb{N}$ and all $j \in \mathbb{N}$ such that $j \neq k$,

$$\deg\left(h_k^\phi\right) = k \quad \text{and} \quad \int_{\mathcal{D}} h_k^\phi(x) h_j^\phi(x) \phi(x) dx = 0. \tag{6}$$

The existence of a family $(h_k)_{k\in\mathbb{N}}$ of orthogonal polynomials with respect to ϕ can be shown constructively, e.g., by applying the Gram-Schmidt orthogonalization to the basis $1, x, \ldots, x^n, \ldots$. For fixed $n \in \mathbb{N}$, $h_1^\phi, \ldots, h_n^\phi$ are uniquely determined, up to a sign, if they also obey to the following normality condition

$$\int_{\mathcal{D}} h_k^\phi(x)^2 \phi(x) dx = 1\,. \tag{7}$$

Henceforth, for a given kernel ϕ possessing finite moments, we denote by $(h_k^\phi)_{k\in\mathbb{N}}$ the unique (up to a sign) family of related orthogonal polynomials satisfying condition (7) for every $k \in \mathbb{N}$. Furthermore, for a given $n \in \mathbb{N}$, we denote by $\mathbb{W} := (w_{i,j})$ the $(n+1) \times (n+1)$ lower triangular matrix containing the coefficients of $h_0^\phi, \ldots, h_n^\phi$. Specifically, for $i = 1, \ldots, n$ we have

$$h_i^\phi(x) = w_{i,0} + w_{i,1} x + \ldots + w_{i,i} x^i, \tag{8}$$

and $w_{i,j} = 0$ for $j > i$.

For the numerical computation of the basis $h_1^\phi, \ldots, h_n^\phi$, the following recurrence relation can be used as a more efficient alternative to the Gram-Schmidt procedure. Define

$$h_0^\phi(x) = 1, \quad h_1^\phi(x) = \frac{x - \mathcal{M}^\phi}{\sqrt{\mathcal{V}^\phi}}\,,$$

where $\mathcal{M}^\phi$ and $\mathcal{V}^\phi$ are the mean and the variance of ϕ, respectively. The remaining terms, $h_2^\phi, \ldots, h_n^\phi$, can be computed as

$$h_k^\phi(x) = \frac{1}{C_k}\left[(x - a_k) h_{k-1}^\phi(x) - b_k h_{k-2}^\phi(x)\right], \tag{9}$$

where

$$\begin{aligned} a_k &= \int_{\mathcal{D}} x h_{k-1}^\phi(x)^2 \phi(x)^2 dx\,, \quad b_k = \int_{\mathcal{D}} x h_{k-1}^\phi(x) h_{k-2}^\phi(x) \phi(x)^2 dx\,, \\ C_k &= \left(\int_{\mathcal{D}} \left[(x - a_k) h_{k-1}^\phi(x) - b_k h_{k-2}^\phi(x)\right]^2 \phi(x)^2 dx\right)^{\frac{1}{2}}. \end{aligned}$$

An important property of orthogonal polynomials is that they fulfill a mass conservation principle that extends to all moments. More specifically, for fixed $p \geq 0$

$$\int_{\mathcal{D}} x^p f_{\mathbb{Q}}^{(n)}(x)dx = \int_{\mathcal{D}} x^p f_{\mathbb{Q}}^{(p)}(x)dx, \quad \forall n \geq p. \tag{10}$$

A major consequence of (10) is that, irrespective of the order n, the approximated RND obtained by (5) always integrates to 1, irrespective of the order n, i.e.,

$$\int_{\mathcal{D}} f_{\mathbb{Q}}^{(n)}(x)dx = 1 \quad \forall n \in \mathbb{N}. \tag{11}$$

Furthermore, in view of (10), we can interpret $c_1, \ldots, c_p$ as corrective factors of the first p moments of ϕ to match the corresponding moments of $f_{\mathbb{Q}}$. In general, $f_{\mathbb{Q}}^{(n)}$ is not guaranteed to be a positive function over its support, even under the assumption that ϕ is a positive function. This property is recovered when the n-th coefficient c_n fulfills some constraints. More precisely, $f_{\mathbb{Q}}^{(n)} \geq 0$ if and only if

$$c_n^{\inf} \leq c_n \leq c_n^{\sup}, \tag{12}$$

where

$$c_n^{\inf} = -\sup_{x\,:\,h_n^{\phi}(x)>0} \frac{f_{\mathbb{Q}}^{(n-1)}(x)}{h_n^{\phi}(x)}, \qquad c_n^{\sup} = \inf_{x\,:\,h_n^{\phi}(x)<0} \frac{f_{\mathbb{Q}}^{(n-1)}(x)}{h_n^{\phi}(x)}.$$

The result above follows after noticing that

$$f_{\mathbb{Q}}^{(n)} = f_{\mathbb{Q}}^{(n-1)} + c_n h_n^{\phi}.$$

3.2 Admissibility of orthogonal expansions

This paragraph is devoted to discussing the admissibility of expansion (5) for the RND. The following theorem represents a prerequisite for setting up a mathematically well-posed procedure to estimate the expansion coefficients $c_1, \ldots, c_n$ based on observed option prices.

Theorem 3.1. *Assume that* $supp(f_{\mathbb{Q}}) \subseteq \mathcal{D} \subseteq \mathbb{R}^+$ *and that* $\phi^{-1} f_{\mathbb{Q}}^2$ *is integrable over its support. Moreover, assume that* $\lim_{x\to+\infty} \phi(x) e^{\varsigma x^{\frac{1}{2}}} = 0$ *for some* $\varsigma > 0$ *and that* $p\phi$ *is bounded for some polynomial* p*. Then:*

(a) there exists a sequence $(c_k)_{k\in\mathbb{N}}$ *such that, for a proper subsequence of indexes*[1] $(k_n)_{n\in\mathbb{N}}$*,*

$$f_{\mathbb{Q}}(x) = \lim_{n\to+\infty} \phi(x) \left(1 + \sum_{k=1}^{k_n} c_k h_k^{\phi}(x)\right) \quad \text{for a.e. } x \in \mathcal{D}\,;$$

[1] There is no need for a subsequence if the RND fulfills additional regularity assumptions, e.g., smoothness.

(b) the following holds in the limit

$$\lim_{n\to+\infty}\int_0^{+\infty}\Pi(x)f_{\mathbb{Q}}^{(n)}(x)dx=\int_0^{+\infty}\Pi(x)f_{\mathbb{Q}}(x)dx\,, \tag{13}$$

for any function Π *such that* $\Pi\phi^{\frac{1}{2}}\in L^2(\mathcal{D})$.

Proof. See Appendix A.1. □

Point (a) in Theorem 3.1 provides sufficient conditions to ensure that the RND admits the representation (5) or, in other words, that $f_{\mathbb{Q}}^{(n)}$ converges to $f_{\mathbb{Q}}$ for some $c_1,\dots,c_n,\dots$. Point (b) ensures that we can move the limit within the integral pricing formulas in (4) and obtain convergent expansions for prices of call and put options. Looking at the hypotheses of Theorem 3.1, we note that they are essentially sharp conditions, meaning that they cannot be further relaxed and, in fact, they sharpen the assumption of Filipovic et al. (2013), which allows for Gamma-type decay in the kernel. Indeed, if $\phi^{-1}f_{\mathbb{Q}}^2$ is not integrable, then the expansion (5) is not well-defined since it may diverge as $n\to\infty$. This means that the kernel should not decay faster than the RND. On the other hand, if $\lim_{x\to+\infty}\phi(x)e^{\varsigma x^{\frac{1}{2}-\gamma}}>0$ for some $\gamma,\varsigma>0$, then Theorem 3.1 is still well-defined but will not necessarily converge to $f_{\mathbb{Q}}$, as pointed out by Theorem A.3-(ii), reported in the Appendix A.1. This suggests that the expansion (5) may not be sufficiently informative on the RND when the kernel decays too fast, i.e., faster than $e^{-\sqrt{x}}$. However, in view of Theorem 3.1, the kernel ϕ mostly serves as an initializing state of the expansion (5). Therefore, as long as the choice of ϕ complies with the hypotheses underlying Theorem 3.1, its impact on the form of $f_{\mathbb{Q}}^{(n)}$ will be of marginal importance, provided that n is sufficiently large. The validity of this statement is supported by numerical illustrations provided in Section B.2 of the Supplementary material. Finally, the following remark highlights the existence of a linear mapping between the coefficients $(c_k)_{k\in\mathbb{N}}$ and the moments of $f_{\mathbb{Q}}$.

Remark 3.2. *If the hypotheses underlying Theorem 3.1 are satisfied, then one can show that for every* $k\in\mathbb{N}$

$$c_k=\int_{supp(\phi)}h_k^{\phi}(x)f_{\mathbb{Q}}(x)dx=\sum_{i=0}^{k}w_{k,i}\int_{-\infty}^{+\infty}x^i f_{\mathbb{Q}}(x)dx \tag{14}$$

where $w_{k,i}$ *is the* i*-th coefficient of* h_k^{ϕ}.

3.3 The extended Laguerre and the log-Hermite expansions

In this sub-section, we study the properties of a class of kernel densities that can form the basis for the polynomial expansions used to retrieve the RND of VIX. Based on restrictions imposed by Theorem 3.1, we propose the following family of kernels with support on $\mathcal{D}=[0,+\infty[$,

$$\phi(x)\propto x^{\alpha-1}e^{-\left(\beta x^p+\xi x^{-1}\right)}\mathbf{1}_{\mathcal{D}}(x),\quad \alpha,\beta,\xi,p\in\Theta, \tag{15}$$

where

$$\Theta = \{\alpha, \beta, \xi, p \in \mathbb{R} \mid \beta > 0, 0 < p \leq 1, (\alpha > 0, \xi = 0) \vee (\alpha \in \mathbb{R}, \xi > 0)\}.$$

The specification (15) 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 $\xi = 0$) kernel. Therefore, the orthogonal expansions arising from (15) extend the classic Laguerre expansions that are associated with a Gamma kernel. For this reason, we refer to these expansions as *extended Laguerre*. This family is flexible enough to capture different tail behaviors and it proves effective in reproducing the peculiar tail behavior of the RND implied by VIX options, as later shown in Section B.2. Indeed, the tail behavior of the GIG and GW kernels characterizes their ability to meet the condition $\phi^{-\frac{1}{2}} f_{\mathbb{Q}} \in L^2(\mathcal{D})$, required by Theorem 3.1. Looking at the left tail, the GW kernel clearly has the slowest decay, and thus it is the most flexible in terms of behavior around the origin. However, Theorem 3.1 also requires that $\lim_{x \to +\infty} \phi(x) e^{\varsigma x^{\frac{1}{2}}} = 0$ for some $\varsigma > 0$. This condition is always met by the GIG kernel, and it is met by the GW kernel when p is restricted to fall between $[\frac{1}{2}, 1]$.

As alternative to the extended Laguerre kernel family, one could consider the log-normal (LN) kernel, that is

$$\phi(x) \propto \frac{1}{x} e^{-\frac{1}{2\sigma^2}(\log(x)-\mu)^2}, \quad \mu \in \mathbb{R},\ \sigma > 0. \tag{16}$$

Expanding the underlying RND based on the LN kernel is conceptually similar to applying a Hermite expansion to the logarithm of the underlying. Furthermore, there is documented empirical evidence that the volatility, a quantity comparable to the VIX, is roughly log-normally distributed, see e.g. Christoffersen et al. (2010), Wang and Daigler (2011) and Bayer et al. (2013). This makes the LN kernel an interesting competitor for either the GIG or the GW kernel. Concerning the condition $\phi^{-\frac{1}{2}} f_{\mathbb{Q}} \in L^2(\mathcal{D})$, the LN kernel ensures the greatest flexibility in terms of requirements on the right tail of the RND. However, the condition $\lim_{x \to +\infty} \phi(x) e^{\varsigma x^{\frac{1}{2}}} = 0$ for some $\varsigma > 0$ is never met by the LN kernel. Thus, the LN kernel does not guarantee that the RND is fully recovered by the expansion (15). Therefore, using the LN kernel inherently entails further restrictions on the form of $f_{\mathbb{Q}}$, and it may reduce the "non-structurality" of the approach.

4 Retrieving the RND from the option prices

In this section, we outline a procedure to estimate the coefficients $c_1, \ldots, c_n$ of the expansion (5) by minimizing the distance between the RND implied and the observed option prices. The key feature of the proposed procedure is to estimate the option pricing formulas in (4) based on the observed market prices by choosing a sufficiently large n in the expansion (5), where the coefficients $c_1, \ldots, c_n$ are the unknown terms that convey the information about $f_{\mathbb{Q}}$. The consistency of this procedure relies on Theorem 3.1.

For fixed $K \geq 0$, $n \in \mathbb{N}$, and $c = [c_1, \ldots, c_n]^\top \in \mathbb{R}^n$, the prices associated with the expansion

of order n are defined as

$$C_K^{(n)}(c) := \int_K^{+\infty} \phi(x)\left(1 + \sum_{k=1}^{n} c_k h_k^{(\phi)}\right)(x-K)\,dx\,,$$
$$P_K^{(n)}(c) := \int_0^{K} \phi(x)\left(1 + \sum_{k=1}^{n} c_k h_k^{(\phi)}\right)(K-x)\,dx.$$

The expressions above can be rewritten in the following compact form

$$C_K^{(n)}(c) = A_0^{(K)} + A^{(K)}c\,, \quad P_K^{(n)}(c) = B_0^{(K)} + B^{(K)}c \tag{17}$$

where

$$A_0^{(K)} = \int_K^{+\infty} \phi(x)\,(x-K)\,dx, \quad B_0(K) = \int_0^{K} \phi(x)\,(K-x)\,dx\,,$$

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

$$A_i^{(K)} = \sum_{j=0}^{i} w_{i,j} \int_K^{+\infty} \left(x^{j+1} - Kx^j\right)\phi(x)dx\,, \quad B_i^{(K)} = \sum_{j=0}^{i} w_{i,j} \int_0^{K} \left(Kx^j - x^{j+1}\right)\phi(x)dx\,. \tag{18}$$

Therefore, for chosen ϕ and n, one may estimate $c_1, \ldots, c_n$ by collecting a cross-section of undiscounted market prices, $C_{K_m}^{Obs}(t,\tau)$ and $P_{K_m}^{Obs}(t,\tau)$, for $m = 1, \ldots, M$, and by finding the solution $\hat{c} = [\hat{c}_1, \ldots, \hat{c}_n]^\top$ of the following optimization problem

$$\hat{c} = \underset{c \in \mathbb{R}^n}{\operatorname{argmin}}\, Q(t,\tau;c)\,, \tag{19}$$

where $Q(t,\tau;\cdot)$ defines an objective function to be minimized. Given that the expressions in (17) are linear in the coefficients c, natural choice for $Q(t,\tau;\cdot)$ is the criterion function of the minimum least squares problem for the following linear model

$$\mathbf{Y} = \mathbf{X}_0 + \mathbf{X}c + \varepsilon, \tag{20}$$

where the $2M \times n$ matrix $\mathbf{X}$ is

$$\mathbf{X} = \begin{bmatrix} A_1^{(K_1)} & \ldots & A_n^{(K_1)} \\ \vdots & \ddots & \vdots \\ A_1^{(K_M)} & \ldots & A_n^{(K_M)} \\ B_1^{(K_1)} & \ldots & B_n^{(K_1)} \\ \vdots & \ddots & \vdots \\ B_1^{(K_M)} & \ldots & B_n^{(K_M)} \end{bmatrix},$$

$\mathbf{Y} = [C_{K_1}^{Obs}(t,\tau), \ldots, C_{K_M}^{Obs}(t,\tau), P_{K_1}^{Obs}(t,\tau), \ldots, P_{K_M}^{Obs}(t,\tau)]'$, and $\mathbf{X}_0 = [A_0^{K_1}, \ldots, A_0^{K_M}, B_0^{K_1}, \ldots, B_0^{K_M}]'$. The $2M \times 1$ vector ε represents the error term, whose properties are discussed more in detail in

Section 5. In this way, the objective function takes the following quadratic form

$$Q(t,\tau;c) = (\mathbf{Y}^* - \mathbf{X}c)'(\mathbf{Y}^* - \mathbf{X}c), \tag{21}$$

where $\mathbf{Y}^* = \mathbf{Y} - \mathbf{X}_0$ is a $2M \times 1$ vector and $\mathbf{X}_0$ is the option price vector generated by the kernel.

Unfortunately, the columns of $\mathbf{X}$ are functions of the first non-standardized n moments. As a consequence, they display an increasing degree of multicollinearity as the expansion order n grows. Employing the standard LS minimization to solve (21) is therefore not suitable when n is large. This is a well-known problem in the literature on orthogonal polynomial expansions. For example, Jarrow and Rudd (1982) and Corrado and Su (1996a,b) consider expansions only up to the fourth order, i.e., $n = 4$, and they calibrate the standardized skewness and kurtosis to the options on the SPX. Similarly, Jondeau and Rockinger (2001) estimate the RND of the Franc-Mark exchange rate by matching only the first four moments, which implies once again that n does not exceed 4. In this regard, it is important to stress that the RND of VIX is expected to be characterized by a long right tail, meaning that the moments higher than the fourth may provide significant information on the shape of the RND. We solve the problem of multicollinearity by orthogonalization of the regressors in (20). The latter, accomplished by means of PCA, also allows to achieve a dimensionality reduction of the problem in (19) without discarding a priory potentially relevant information. The Supplementary material, in Section B.1, provides further details on the implementation of the PCA analysis in this context. The vector of coefficients that minimizes the quadratic objective quadratic function under the PCA constraints is denoted by $\tilde{c}$. Given the vector $\tilde{c}$, the estimated RND function $\tilde{f}_{\mathbb{Q}}^{(n)}$ is determined as

$$\tilde{f}_{\mathbb{Q}}^{(n)}(x) = \phi(x;\theta)\left(1 + \sum_{k=1}^{n} \tilde{c}_k h_k^{\phi}(x)\right). \tag{22}$$

Note that the kernel ϕ in (22) is now expressed as function of an additional term, to highlight its dependence on a set of parameters $\theta \in \Theta$. For example, $\theta = [\alpha, \beta, \xi]'$ for the GIG kernel, $\theta = [\alpha, \beta, p]'$ for the GW kernel, and $\theta = [\mu, \sigma^2]$ for the LN kernel. The choice of θ does not play a crucial role in this context as long as $\phi(\cdot;\theta)$ fulfills the assumptions of Theorem 3.1. In practice, we set θ to the value that minimizes the residuals variance for the expansion of order 0. The minimization is performed under the restriction of zero-mean residuals, which implies absence of systematic pricing errors. A number of numerical examples, reported in Section B.2 of the Supplementary material, display the robustness and the flexibility of the proposed methodology.

5 No-arbitrage violations

Market prices are typically subject to a number of frictions that are typically functions of the market liquidity. Obviously, possible arbitrage opportunities can hardly be exploited due to the

presence of transaction costs in the form of bid and ask spread. However, from a purely mathematical perspective, the fact that the mid-quote is adopted to approximate the latent arbitrage-free option price can be seen as a violation of the no-arbitrage assumption. This means that even in absence of discretization/truncation errors it is not possible to achieve exact matching of all the observed option prices by minimizing (19), since the option prices obtained through a RND are free of (static) arbitrages by construction. Therefore, in defining the linear model (20) on which we build the estimation procedure based on orthogonal polynomials, we can assume that the error term ε subsumes all the uncertainty associated with the fact that the polynomial expansion is truncated to a finite n, that the number of available strikes M is finite and that the market prices may be subject to no-arbitrage violations. In other words, the error term in (20) can be split as $\varepsilon = \delta + \epsilon$, where δ is a vector of non-stochastic terms coming from the fact that both n and M are finite, while ϵ is a random term that approximates all the deviations from the latent arbitrage-free option prices resulting from the trading activity. In particular, we assume that ϵ is a vector of random variables with zero mean and that the vector $\mathcal{Y} = \mathbf{Y} - \epsilon$ is arbitrage-free. An example of specific distributional form for ϵ is given in Section B.6 in the Supplementary material.

In the following, an *infill* asymptotic analysis is carried out to show that, as the observed prices are sampled increasingly over a fixed interval of strikes and as $n \to \infty$, then $\delta \to 0$ and the only remaining error term is the noise associated with the no-arbitrage violations. Assuming that the observed option strikes fall in a fixed finite interval $I = [K_1, K_M]$, we define the "infill version" of (21) as

$$\mathbf{Q}^{(n)}(t, \tau; c = [c_1, \dots, c_n]^\top) := \frac{1}{K_M - K_1} \int_I \left(C_K^{Obs}(t, \tau) - C_K^{(n)}(c)\right)^2 + \left(P_K^{Obs}(t, \tau) - P_K^{(n)}(c)\right)^2 dK. \tag{23}$$

Note that there are other ways to define an "infill counterpart" of Q consistently. The definition that we adopt, justified by mathematical convenience, builds on the fact that the integral in (23) is the limit of $\frac{1}{M}Q$, under continuity assumptions for the integrand. Since multiplying Q by any constant does not affect the solution $\hat{c}$ of (19), then (23) can be interpreted as a valid infill version of (21). The observed prices $C^{Obs}(t, \tau)$, $P^{Obs}(t, \tau)$ appearing in (19) are assumed to take the form

$$C_K^{Obs}(t, \tau) = C_K(t, \tau) + \epsilon_K^C, \quad P_K^{Obs}(t, \tau) = \mathcal{P}_K(t, \tau) + \epsilon_K^P,$$

where $C_\cdot(t, \tau), C_\cdot(t, \tau) \in C^2(I)$ are defined as in (4), while $\epsilon^C = (\epsilon_K^C)_{K \in I}$ and $\epsilon^P = (\epsilon_K^P)_{K \in I}$ are zero mean processes on a probability space $(\Omega^\epsilon, \mathcal{F}^\epsilon, P^\epsilon)$, belonging to $L^2(\Omega^\epsilon \times I)$. Under these assumptions, the infill target function $\mathbf{Q}^{(n)}$ is well-defined and has finite expected value for every t, τ, n.

Proposition 5.1 (Infill asymptotics). *Assume that the hypotheses of Theorem 3.1 are satisfied and*

denote by $c^\star(n)$ *the minimum of* $\mathbf{Q}^{(n)}(t,\tau;\cdot)$*, for every* $n \in \mathbb{N}$*. The following inequality holds*

$$\lim_{n\to+\infty} E\left[\mathbf{Q}^{(n)}(t,\tau;c^\star(n))\right] \leq \frac{1}{K_M - K_1} E\left[\int_I \left(\epsilon_K^C\right)^2 + \left(\epsilon_K^P\right)^2 dK\right]. \tag{24}$$

The inequality in (24) *becomes an equality under the following additional hypotheses:*

(i) There exists $\bar{n} \in \mathbb{N}$ *such that, for all* $n \geq \bar{n}$*,* $c^\star(n)$ *is obtained by constraining* (19) *to the space of coefficients* $c_1, \ldots, c_n$ *such that*

$$\phi\left(1 + \sum_{k=1}^{k_n} c_k h_k^\phi\right) \geq 0 \text{ on } \mathcal{D}.$$

(ii) Let $C_\cdot^\star(t,\tau) \in C^2(I)$ *and* $\mathcal{P}_\cdot^\star(t,\tau) \in C^2(I)$ *be arbitrage-free call and put curves, respectively. Then, almost surely*

$$\int_I \left(C_K^{Obs}(t,\tau) - C_K^\star(t,\tau)\right)^2 + \left(P_K^{Obs}(t,\tau) - \mathcal{P}_K^\star(t,\tau)\right)^2 dK \geq \int_I \left(\epsilon_K^C\right)^2 + \left(\epsilon_K^P\right)^2 dK.$$

Proof. See Appendix A.2. □

The inequality (24) defines an upper-bound on the expected value of the target function, $\mathbf{Q}^{(n)}(t,\tau;c^\star(n))$. In particular, the expected value of the target function evaluated in $c^\star(n)$ is lower than the variance of the non-arbitrage residuals. Under the additional assumptions (i)-(ii), Proposition 5.1 states that our estimation method provides the arbitrage-free prices closest to the observed ones. In particular, assumption (i) requires that the estimation always returns a probability density function, while assumption (ii) can be interpreted as a uniqueness requirement on the target RND. This establishes an interesting linkage with the work of Aït-Sahalia and Lo (1998). Furthermore, under no-arbitrage (i.e. $\epsilon^C = 0$ and $\epsilon^P = 0$), Proposition 5.1 ensures that the sum of the squared residuals goes to zero as $n \to \infty$, so that the estimated and the observed prices coincide.

Summing up, Proposition 5.1 provides conditions ensuring that the estimation procedure based on orthogonal polynomials is robust to the presence of measurement errors in the option prices. This is a remarkable feature that is often lacked by non-structural techniques based on the straightforward computation of second-order derivatives, whose estimation is typically very sensitive to data inconsistencies. Below, we derive a theoretical lower bound for the estimation residuals that is inferred directly from the put-call parity violations affecting the observed option prices. In Section B.6 in the Supplementary material we show that this lower bound can be consistently used as a proxy for (24). Moreover, we show that the validity of Proposition 5.1 is empirically confirmed by a number of numerical tests.

5.1 An observable lower bound for the estimation residuals

Following Proposition 5.1, it seems natural to set a tolerance on the variability of the residuals, which are defined as $\tilde{\varepsilon} = \mathbf{Y}^* - \mathbf{X}\tilde{c}$. This tolerance should quantify the presence of arbitrage opportunities. Given a fixed threshold $\Delta^Q > 0$, we define as admissible any RND implying option prices whose distance from the observed ones is below Δ^Q. Consistently, we say that a density $\tilde{f}_{\mathbb{Q}}^{(n)}$ is admissible if

$$\frac{1}{M}\sum_{m=1}^{M}\left(C_{K_m}^{Obs} - \tilde{C}_{K_m}^{(n)}\right)^2 + \frac{1}{M}\sum_{m=1}^{M}\left(P_{K_m}^{Obs} - \tilde{P}_{K_m}^{(n)}\right)^2 \leq \Delta^Q\,,$$

where $\tilde{C}_{K_m}^{(n)}$ and $\tilde{P}_{K_m}^{(n)}$ are call and put prices associated with $\tilde{f}_{\mathbb{Q}}^{(n)}$. Since the option data are always affected by some noise, in view of Proposition 5.1 the existence of admissible RNDs is not guaranteed when Δ^Q is chosen to be too small. A lower bound for the set of all possible values of Δ^Q can be inferred from put-call parity violations. Given $\tilde{f}_{\mathbb{Q}}^{(n)}$, denote by $\text{Mean}^{\mathbb{Q}}$ the risk-neutral futures price computed as

$$\text{Mean}^{\mathbb{Q}} = \int_0^{+\infty} x\tilde{f}_{\mathbb{Q}}^{(n)}(x)dx\,,$$

and by Δ^{pcp} the variance of the put-call parity violations

$$\Delta^{\text{pcp}} = \frac{1}{M}\sum_{m=1}^{M}\left(C_{K_m}^{Obs}(t,\tau) - P_{K_m}^{Obs}(t,\tau) + K_m - \text{Mean}^{Obs}\right)^2\,, \tag{25}$$

where $\text{Mean}^{Obs} = \frac{1}{M}\sum_{m=1}^{M}\left(C_{K_m}^{Obs}(t,\tau) - P_{K_m}^{Obs}(t,\tau) + K_m\right)$. It follows that

$$\begin{aligned}\Delta^{\text{pcp}} &= \frac{1}{M}\sum_{m=1}^{M}\left[C_{K_m}^{Obs}(t,\tau) - P_{K_m}^{Obs}(t,\tau) - C_{K_m}^{(n)} + P_{K_m}^{(n)} + \left(F^{\mathbb{Q}} - F^{Obs}\right)\right]^2 \\ &\leq \frac{1}{M}\sum_{m=1}^{M}\left(C_{K_m}^{Obs}(t,\tau) - \tilde{C}_{K_m}^{(n)}\right)^2 + \frac{1}{M}\sum_{m=1}^{M}\left(P_{K_m}^{Obs}(t,\tau) - \tilde{P}_{K_m}^{(n)}\right)^2 + \left(\text{Mean}^{\mathbb{Q}} - \text{Mean}^{Obs}\right)^2\,,\end{aligned}$$

which yields the following inequality

$$\Delta^Q \geq \Delta^{\text{pcp}} - \left(\text{Mean}^{\mathbb{Q}} - \text{Mean}^{Obs}\right)^2\,. \tag{26}$$

From (26) we obtain a lower bound for the tolerance level that must be allowed on the estimation residual. Moreover, it proves that admissible solutions of (19) with tolerance level lower than $\Delta^{\text{pcp}} - \left(\text{Mean}_1^{\mathbb{Q}} - \text{Mean}^{Obs}\right)^2$ do not exist. Therefore, (26) suggests that setting the tolerance level as $\Delta^Q = \Delta^{\text{pcp}}$ is a convenient choice since Δ^{pcp} is an observable quantity and is expected to be only slightly greater than the lower bound.

6 Empirical Analysis

In this section we estimate the RND of the VIX on a panel of option prices observed in the period from January 2010 to April 2016, sampled at monthly frequencies, and with time-to-maturity, τ, ranging from 1 to 5 months.[2] The data is obtained from the *OptionMetrics* database. For each month in the sample, we collect option prices observed on the first Tuesday following the third Friday of the month. This ensures that observations will not overlap in the base monthly frequency, meaning that 1-month options observed at time t always expire prior to the subsequent observation occurring at time $t+1$. We operate minimum pre-filtering of the data. More specifically, we exclude all OTM puts (calls) with mid-quote below 0.025\$ together with the ITM calls (puts) with the same strikes. These contracts turn out to be highly illiquid (if traded at all) and therefore they are likely subject to mispricing. With this filtering criterion, we have an average total number of 56 available contracts for each date and time-to-maturity. Under normal market conditions, the strike values that are taken into consideration typically fall between 10\$ and 45\$, with this range remaining quite stable over time and maturities due to the mean-reverting behavior of the VIX. However, the interval of available strikes enlarges during turmoil periods, with maximum values reaching peaks of 100\$. For further details see Table 7 in Section B.7 of the Supplementary material. In the following, we focus on a specific date and maturity in the sample to illustrate the effectiveness of the orthogonal polynomials in correcting ϕ and to confirm the robustness of the methodology to the choice of the kernel.

6.1 November 16, 2011

On November 16, 2011, the cross-section of VIX options expiring on December 21, 2011 ($\tau = 1$) quoted by the CBOE consists of 64 contracts. The choice of this date is not coincidental. Indeed, the end of the year 2011 is characterized by high levels of market volatility, registered in connection with the European sovereign debt crisis and the US sovereign debt downgrading. As a consequence, on November 16, 2011, the VIX index reached a value of 33.51%. At the same time, the trading of VIX options spanned strike values in the range between 15\$ and 90\$. After the pre-filtering, we end up with 52 contracts (26 calls and 26 puts) with strikes ranging between 21\$ and 80\$. Hence, on the chosen date, trading of deep out-of-the-money options is sufficiently high to ensure informative market prices in a wide range of strikes and, in particular, suggests uncommonly long right-tail in the RND. To highlight the robustness of our results to the choice of the kernel, we consider both the GIG and the GW kernels. The parameters of the two kernels are chosen to minimize the variance of Y^*. As discussed in Section 4, we can increase the expansion order n to high values, thanks to the orthogonalization of regressors and the dimension reduction thereby operated by means of PCA. Therefore, we set the order to a relatively high value $n = 18$. We find that 5 and 6 principal components explain the 99%

[2] Weekly VIX options and options with time-to-maturity $\tau = 6$ are also traded, but unfortunately they are not available for all dates in the sample.

of the total variance of $\mathbb{X}$ for the GW and the GIG kernel, respectively. The estimated RNDs obtained using expansions of order $n = 18$ are reported in Figure 1 (solid lines), together with the corresponding kernels (dashed lines). From the visual inspection of Figure 1, it emerges that the

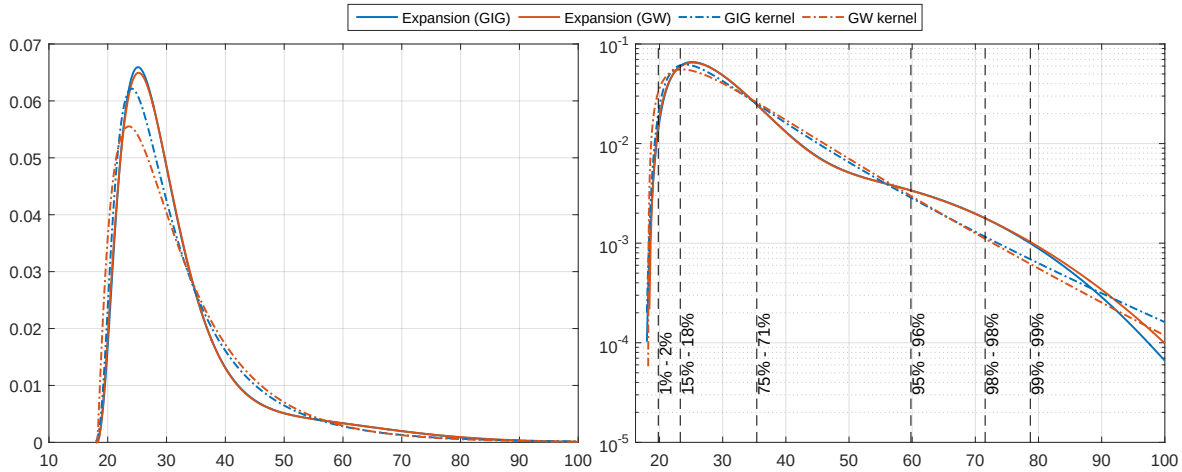
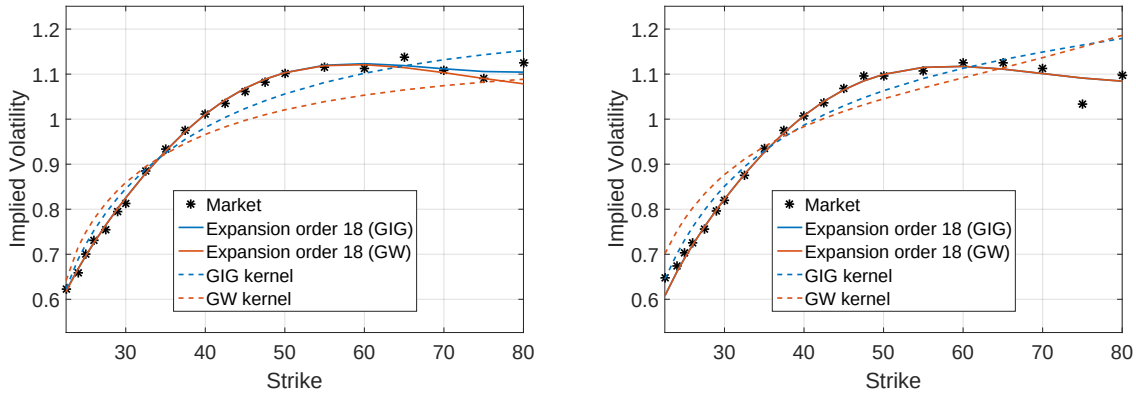


Figure 1: Estimated RNDs. The left panel depicts the graphs of the two kernels considered in this section together with the corresponding estimated RNDs obtained for $n = 18$. The right panel reports the same contents of the left panel in semi-log scale, to highlight tail features that are difficult to observe in linear scale. Each couple of percentage values denotes, from top to bottom, the average mass levels of the two kernels and the two corresponding estimated RNDs, related to the quantiles identified by the dashed vertical lines.

RND of VIX should have short left tails. The densities and the associated kernels display relevant differences on the right tails. The differences mostly occur in a part of the domain above the 98% quantile, thus signaling the relevance of a proper characterization of the right tail in pricing OTM calls and ITM puts of VIX. The importance of the correction provided by the orthogonal polynomials is better understood by comparing the implied volatility curves generated by the two kernels and their corresponding expansions, reported in Figure 2. The implied volatilities generated by both kernels are considerably different from those generated by market prices, while the expansions are able to produce implied volatilities that closely replicate the observed ones. All these observations point in the same direction: neither of the two baseline kernels is able to reproduce the tail-features of the RND of VIX. In particular, they both display positive excess mass between the 75% and 95% quantiles while, on the other hand, they display negative excess mass in the area covered by the last 5 percentiles. Coherently with Theorem 3.1, the expansions obtained using the two different kernels converge to the same density, as the two functions diverge from each other only starting from roughly the 99.5% quantile.

6.2 Stylized facts on VIX risk-neutral moments

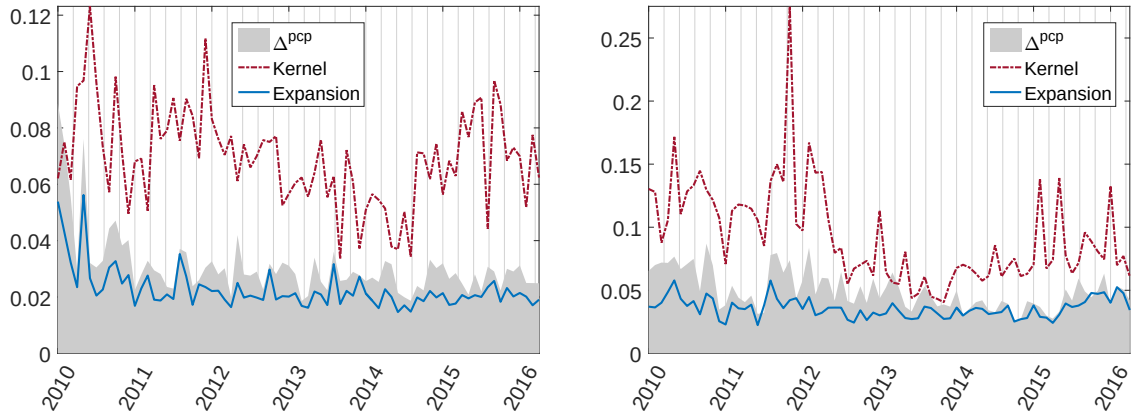
The analysis carried out for November 11, 2016 is replicated for all dates and maturities in the sample. From Figure 3, it emerges that the orthogonal expansion (blue line) outperforms the GW kernel (red line) and, in all cases, it generates a root mean square error that lies around the



(a) Call Imp. Vol. (b) Put Imp. Vol.

Figure 2: Black and Scholes implied volatility curves obtained from market prices, GIG kernel, GW kernel, and the resulting approximated RNDs with expansion order equal to $n = 18$.

market implied threshold $\sqrt{\Delta^{\text{pcp}}}$.[3] This signals the goodness of fit for all the estimated RNDs in



(a) $\tau = 1$ (b) $\tau = 5$

Figure 3: Root mean square error between market VIX option prices and approximate prices implied by both the GW kernel and its related expansion. Option prices have been collected based on monthly observations taken from January 2010 to April 2016 with $\tau = 1$ and $\tau = 5$ months. The grey area is the market implied threshold, defined as $\sqrt{\Delta^{\text{pcp}}}$, see equation (25).

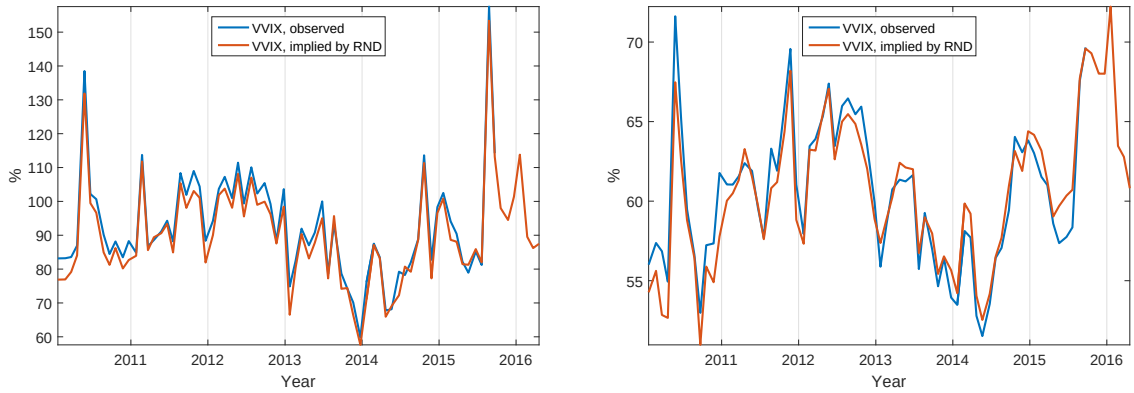
the sample and the accuracy gain achieved by correcting the GW kernel through the orthogonal expansion. To further assess the consistency of the estimated RND with market data, we can look at the VVIX, i.e., the volatility-of-volatility index inferred from VIX options through the same algorithm used for the VIX itself. The VVIX can be linked to the RND of the VIX through

[3]Similar results are obtained with other maturities, $\tau = 2, 3, 4$ and are reported in the Supplementary material together with a table reporting the summary statistics associated with the fit.

the following formula, which is analogous to (1)

$$\frac{\mathrm{VVIX}^2_{t,\tau}}{100} = -\frac{2}{\tau}\int_0^{+\infty} \log\left(\frac{x}{\mathrm{Mean}^{\mathbb{Q}}_{t,\tau}}\right) f_{\mathbb{Q}}(t,\tau;x)dx\,, \tag{27}$$

where $\mathrm{Mean}^{\mathbb{Q}}_{t,\tau} := E^{\mathbb{Q}}_t[\mathrm{VIX}_{t+\tau}]$ denotes the forward price of VIX at time t for maturity $t+\tau$. Figure 4, which reports the observed time series of the VVIX and that computed by (27), confirms that the estimated RND implies generally consistent estimates of the VVIX. Discrepancies are small in size and can be attributed to discretization errors affecting the CBOE formula for the computation of the VVIX.

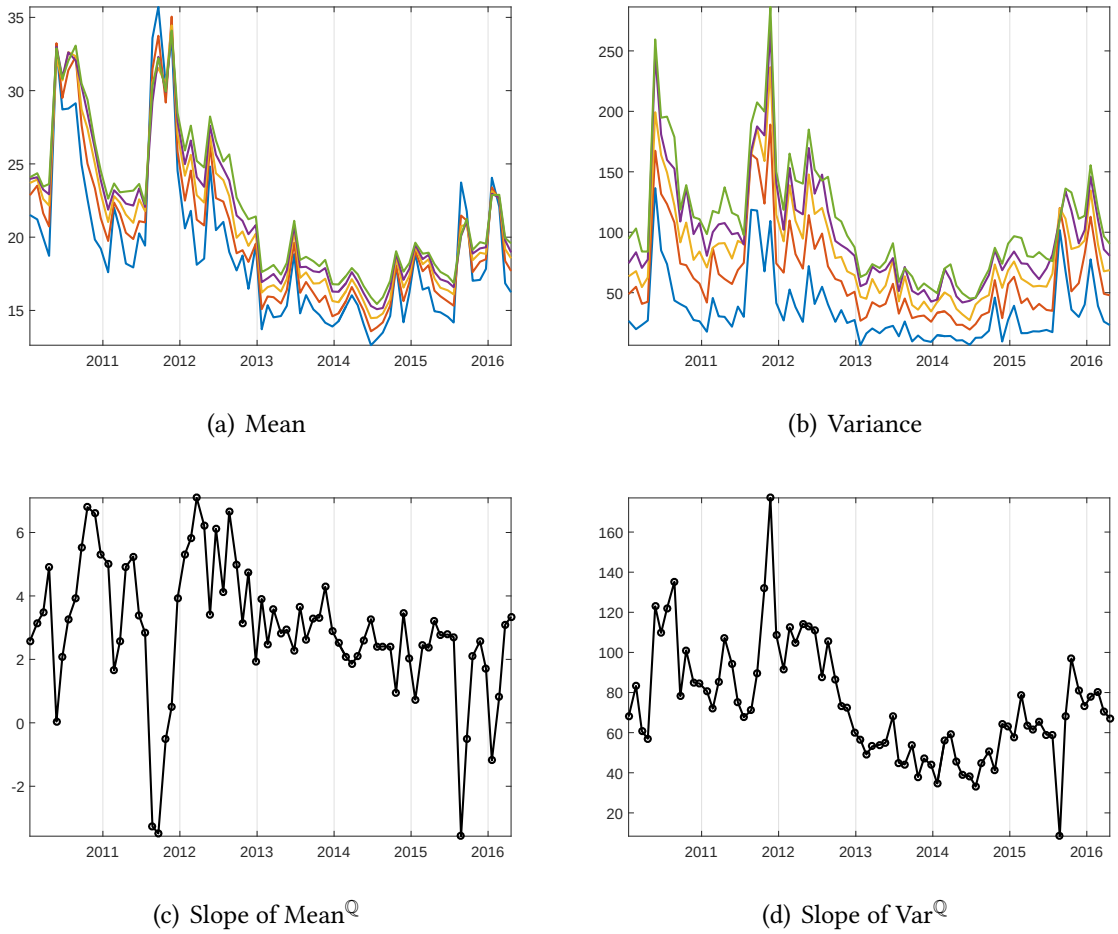


(a) $\tau = 1$ (b) $\tau = 5$

Figure 4: VVIX time series. Comparison between observed values of the VVIX and those obtained by formula (27), for $\tau = 1$ and $\tau = 5$ months.

By direct inversion of the linear relation in (14), it is possible to retrieve the first k risk-neutral moments of the VIX for each value of t and τ. An empirical analysis of the risk-neutral moments of the VIX reveals a set of stylized facts related to the market expectation of volatility (and its powers). Figure 5 reports the time series of mean ($\mathrm{Mean}^{\mathbb{Q}}_{t,\tau}$) and variance ($\mathrm{Var}^{\mathbb{Q}}_{t,\tau}$), for all times-to-maturity. From Panels (a) and (b), we notice that the first two expected risk-neutral moments exhibit large variations associated with the escalation of the European sovereign debt crisis in the second half of 2011. From 2013 to the end of the sample, both moments display lower levels and more stable patterns reflecting agents' confidence in more stable market conditions. For what concerns the first moment, the spread between maturities has remained largely uniform along the sample, i.e., an average spread between $\tau = 5$ and $\tau = 1$ of 3.094 with a standard deviation of 2.135. Additionally, from Panel (c) we can see that changes of sign in the slope of the term structure are sporadic and tend to occur concurrently with extreme market conditions, in analogy with the well-known case of inverted yield curve. The periods with negative slope of the mean term structure are associated with the first Greek debt crisis (April-May 2010), the escalation of the European sovereign debt crisis (October-November 2011), the second Greek debt crisis (July 2015) and the slowdown of Chinese production (February 2016), which all reflect

expectations of increasing market volatility in the short term.



(a) Mean

(b) Variance

(c) Slope of $\text{Mean}^{\mathbb{Q}}$

(d) Slope of $\text{Var}^{\mathbb{Q}}$

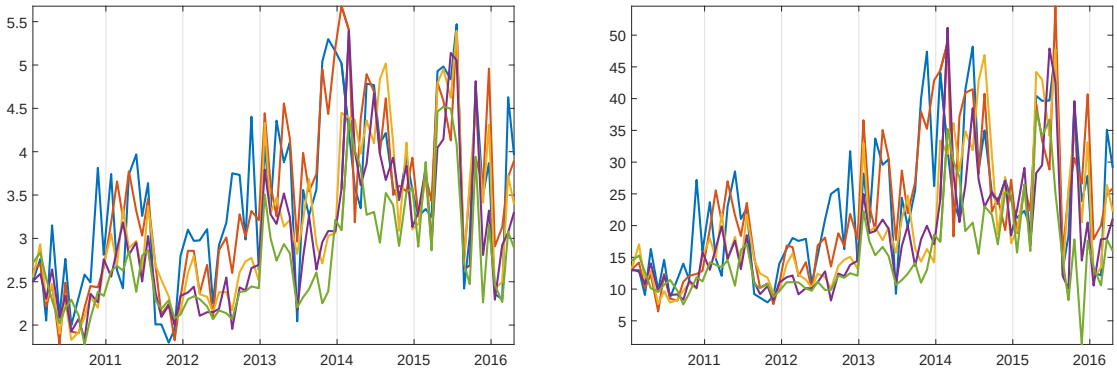
Figure 5: Time series of the VIX risk-neutral mean and variance. Panel a) and b) report the time series of $\text{Mean}^{\mathbb{Q}}_{t,\tau}$) and $\text{Var}^{\mathbb{Q}}_{t,\tau}$ for different times-to-maturity, $\tau = 1$ (blue), $\tau = 2$ (red), $\tau = 3$ (yellow), $\tau = 4$ (purple) and $\tau = 5$ (green). Panels c) and d) also report the time series of the slope of $\text{Mean}^{\mathbb{Q}}$ and $\text{Var}^{\mathbb{Q}}$. The slope is defined as $\text{Mean}^{\mathbb{Q}}_{t,1} - \text{Mean}^{\mathbb{Q}}_{t,1}$ and $\text{Var}^{\mathbb{Q}}_{t,1} - \text{Var}^{\mathbb{Q}}_{t,1}$ $\forall t = 1, \ldots, T$ for mean and variance, respectively.

A generally positive slope in the term structure is also observed for $\text{Var}^{\mathbb{Q}}_{t,\tau}$, whose dynamics closely resemble those followed by $\text{Mean}^{\mathbb{Q}}_{t,\tau}$, see Panel (b). The slope of the variance term structure, measured by the spread between values of $\text{Var}^{\mathbb{Q}}_{t,\tau}$ observed for $\tau = 5$ and $\tau = 1$, peaks during periods of market turmoil up to 177.022, against an average value of 73.751 over the whole period. This behavior reflects agents' lack of confidence about the level of market stability in the long term. Only in one case, during the second Greek debt crisis in July 2015, does the slope not increase to abnormal levels in response to the market turmoil, although it remains positive.

We further investigate the link between $\text{Mean}^{\mathbb{Q}}_{t,\tau}$ and $\text{Var}^{\mathbb{Q}}_{t,\tau}$ to measure the leverage effect. The leverage effect generally relates to the negative correlation between the change of an asset price and its volatility. We perform regressions of $\Delta\text{Mean}^{\mathbb{Q}}_{t,\tau}$ on a constant and $\Delta\text{Var}^{\mathbb{Q}}_{t,\tau}$ for each maturity τ. The coefficients associated with the change in $\text{Var}^{\mathbb{Q}}_{t,\tau}$ are all positive and increasing

in τ, ranging from 0.241 ($\tau = 1$) to 0.315 ($\tau = 5$), and the regressions R-squared are around 75%. Such evidence suggests the presence of a leverage effect with positive (*reversed*) sign. As the uncertainty about future levels of market volatility increases, it is arguable that this has direct implications on the expected level of market volatility itself, which increases proportionally.

By means of the normalized versions of the third and fourth moments reported in Figure 6, namely the skewness ($\text{Sk}^{\mathbb{Q}}_{t,\tau}$), and kurtosis ($\text{Kurt}^{\mathbb{Q}}_{t,\tau}$), we can study how the shape of the RND changes over time. Both $\text{Sk}^{\mathbb{Q}}_{t,\tau}$ and $\text{Kurt}^{\mathbb{Q}}_{t,\tau}$ appear highly volatile and share similar dynamics, as also suggested by their sample correlations reaching levels close to 97%. Contrary to what we have observed for mean and variance, $\text{Sk}^{\mathbb{Q}}_{t,\tau}$ and $\text{Kurt}^{\mathbb{Q}}_{t,\tau}$ plummet during periods of market turmoil. Thus, the third and the fourth moment increase at a slower rate with respect to the variance. As the conditional mean and variance of VIX increase, moving away from the zero lower bound, the distribution tends to become more symmetric and thinner-tailed. Notably,



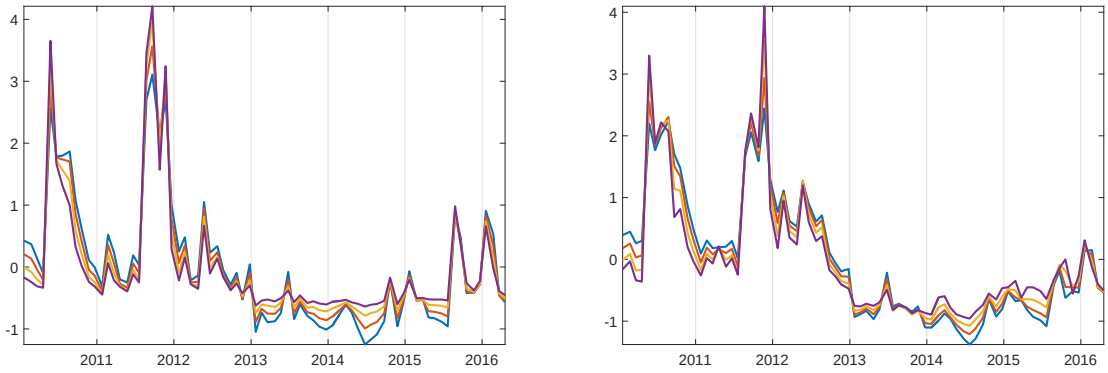
(a) Skewness

(b) Kurtosis

Figure 6: Time series of the VIX risk-neutral skewness and kurtosis. Panel a) and b) report the time series of $\text{Sk}^{\mathbb{Q}}_{t,\tau}$) and $\text{Kurt}^{\mathbb{Q}}_{t,\tau}$ for different times-to-maturity, $\tau = 1$ (blue), $\tau = 2$ (red), $\tau = 3$ (yellow), $\tau = 4$ (purple) and $\tau = 5$ (green).

both $\text{Sk}^{\mathbb{Q}}_{t,\tau}$ and $\text{Kurt}^{\mathbb{Q}}_{t,\tau}$ exhibit a term structure that is systematically downward sloped. Sample averages decrease from 3.43 ($\tau = 1$) to 2.78 ($\tau = 5$) for the $\text{Sk}^{\mathbb{Q}}_{t,\tau}$ and from 22.90 ($\tau = 1$) to 15.55 ($\tau = 5$) for $\text{Kurt}^{\mathbb{Q}}_{t,\tau}$. As the prediction horizon increases, the distribution implied by the VIX options reflects higher expectations and volatility for the level of VIX, together with a more symmetric and platykurtic distribution.

The similarity in the dynamic behavior of the first four moments becomes striking after simple rescaling. As an example, Figure 7 plots the standardized time series of the first four moments for $\tau = 1$ and $\tau = 5$, with lines tracking one another. The sample correlations are well above 90% for both $\tau = 1$ and $\tau = 5$, supporting the existence of a strong link between (standardized) moments. To further explore the behavior of the implied moments both over the time dimension and the term-structure, we test for the existence of a common factor structure across maturities and moment orders. This is conceptually similar to the analysis carried out by Andersen et al. (2015b) on the SPX implied volatility surfaces. The PCA, carried out on the



(a) $\tau = 1$ (b) $\tau = 5$

Figure 7: Standardized moments for short ($\tau = 1$) and long $\tau = 5$ time-to-maturity. The standardization is performed as $Z_i = \frac{M_i - \bar{M}_i}{S(M_i)}$ for $i = 1, \ldots, 4$ where $\bar{M}$ and $S(M_i)$ are the time-series sample mean and standard deviation of the i-th moment.

cross-section of τ, shows that the first principal component of $\text{Mean}^{\mathbb{Q}}_{t,\tau}$ explains 98% of its total covariance. Similarly, the first principal component of $\text{Var}^{\mathbb{Q}}_{t,\tau}$ explains almost 96% and, for the third and fourth moments, the first principal component explains 89% and 82% of the related cross sectional (across maturities) variability.

When regressing $\text{Mean}^{\mathbb{Q}}_{t,\tau}$ on a constant and the first principal component, we find regression intercepts that increase with the maturity, ranging from 19.26 ($\tau = 1$) to 22.17 ($\tau = 5$), but slopes that are fairly invariant. This suggests that VIX options prices incorporate premia for the uncertainty about future market conditions as the horizon increases, but also that these premia are constant over time, and thus deterministic in nature. Similar results are obtained when repeating the same exercise for the variance, although in this case the slope coefficients, increasing with τ from 12.1 to 21.8, reflect a (less than proportional) increase in the variability of $\text{Var}^{\mathbb{Q}}_{t,\tau}$ with the time-to-maturity. The PCA for the third and fourth moments shows qualitatively the same results, up to higher orders of magnitude. We also test whether and to what extent the high degree of correlation between the four moments, as illustrated in Figure 7, is due to the presence of a common factor. We find that the first principal component explains nearly 93% of the total covariation between the 20 variables. This result provides strong evidence of a main driver for the RND of VIX over time.

6.3 VIX jumps under $\mathbb{Q}$

The peculiar proportional structure of the moments of the VIX under risk neutrality suggests that there might exist a multiplicative error model, MEM, driving the dynamics of future states of the VIX over time under the $\mathbb{Q}$-measure. The MEM model is defined as

$$X_t = \mu_t \eta_t, \quad t = 1, \ldots, T, \tag{28}$$

where, for a fixed maturity τ, $X_t = \text{VIX}_{t+\tau}$, μ_t is the expectation of $\text{VIX}_{t+\tau}$ computed at the time t, η_t is an i.i.d. stochastic term, independent of μ_t, with positive support and with mean equal to 1, and T is the sample size. MEM type dynamics for the VIX index are also studied in Mencia and Sentana (2016), who adopt a Laguerre expansion for the innovation term η_t under the physical measure $\mathbb{P}$. They also analyze the process under $\mathbb{Q}$, using the VIX futures and assuming an exponentially affine stochastic discount factor to model the risk premium. In this way, they retrieve information on the risk-neutral higher moments of VIX by a structural assumption to link $\mathbb{P}$ and $\mathbb{Q}$. We instead perform a direct analysis of the distributional assumptions on the VIX under $\mathbb{Q}$, exploiting higher risk-neutral moments obtained from the VIX options. The idea of using transformations of option prices to back out estimates of a parametric model is in line with the methodology adopted in Pastorello et al. (2000) and Pan (2002), among others.

By setting $\eta_t \sim i.i.d.\Gamma(1,\nu)$ (mean-shape form) in (28), which corresponds to a zero-order Laguerre expansion we obtain

$$E_t^{\mathbb{Q}}[\text{VIX}_{t+\tau}] = \mu_t\,, \qquad E_t^{\mathbb{Q}}[\text{VIX}_{t+\tau}^2] = \mu_t^2\left(\frac{1}{\nu}+1\right),$$
$$E_t^{\mathbb{Q}}[\text{VIX}_{t+\tau}^3] = \mu_t^3\left(\frac{2}{\nu^2}+\frac{3}{\nu}+1\right), \qquad E_t^{\mathbb{Q}}[\text{VIX}_{t+\tau}^4] = \mu_t^4\left(\frac{6}{\nu^3}+\frac{11}{\nu^2}+\frac{6}{\nu}+1\right).$$

Under the MEM-Gamma the proportionality between second, third and fourth moment is controlled by a single parameter, ν. The MEM-Gamma can be augmented with jumps, thus obtaining the MEM-J of Caporin et al. (2017), defined as

$$X_t = \mu_t Z_t \eta_t, \quad t = 1,\ldots,T, \tag{29}$$

where Z_t is the VIX jump component, and the innovation $\eta_t \sim i.i.d.\Gamma(1,\nu)$. The jump term, Z_t is defined as

$$Z_t = \begin{cases} d_\lambda & N_t = 0 \\ \sum_{j=1}^{N_t} Y_{j,t} & N_t > 0 \end{cases} \tag{30}$$

where N_t is a Poisson random variable with intensity $\lambda > 0$ and $d_\lambda = \left(e^{-\lambda}+\lambda\right)^{-1}$ is a scalar positive function of λ, denoting the baseline value of Z_t in absence of jumps. When $N_t > 0$, the process Z_t is a compound Poisson process. Following Caporin et al. (2017), we assume that the jump sizes are driven by a Gamma distribution $\Gamma(d_\lambda,\varsigma)$ (in mean-shape form), from which it follows that the innovation term $\xi_t = Z_t\eta_t$ is driven by a countably infinite mixture of Gamma

	MEM-Gamma		MEM-J		
	$\hat{\nu}$	$\chi^2(2)$	$\hat{\nu}$	$\hat{\lambda}$	$\chi^2(1)$
$\tau = 1$	26.38^a	0.000	55.79^a	0.351^a	0.053
$\tau = 2$	12.87^a	0.000	47.33^a	0.513^a	0.200
$\tau = 3$	12.65^a	0.000	41.04^a	0.591^a	0.161
$\tau = 4$	9.433^a	0.000	26.65^a	0.585^a	0.117
$\tau = 5$	7.411^a	0.000	19.66^a	0.563^a	0.054

Table 1: GMM estimates of the parameters of the MEM-Gamma and the MEM-J. The estimation is performed by matching the parametric expressions of $E_t^{\mathbb{Q}}[\text{VIX}_{t+\tau}^2]$, $E_t^{\mathbb{Q}}[\text{VIX}_{t+\tau}^3]$ and $E_t^{\mathbb{Q}}[\text{VIX}_{t+\tau}^4]$ outlined above with their empirical counterpart obtained from the RNDs retrieved from the option prices. The first moment is exactly matched, i.e. $\text{Mean}_{t,\tau}^{\mathbb{Q}} := E_t^{\mathbb{Q}}[\text{VIX}_{t+\tau}] = \mu_t$, by construction, and it is used as a driver for the dynamics of the higher order moments. The subscripts a, b and c stand for significance at 1%, 5% and 10%, respectively. $\chi^2(2)$ and $\chi^2(1)$ are the p-values of the J-test for over-identifying restrictions whose distribution is $\chi^2(m-r)$, where $m = 3$ is the number of moment conditions adopted in the estimation and r is the number of free parameters.

and *Kappa* distributions with closed-form moments given by

$$
\begin{aligned}
E_t^{\mathbb{Q}}[\text{VIX}_{t,\tau}] &= \mu_t\,, \\
E_t^{\mathbb{Q}}[\text{VIX}_{t,\tau}^2] &= \mu_t^2\left[\frac{\lambda}{\varsigma} + e^{-\lambda} + \left(\lambda+\lambda^2\right)\right] d_\lambda^2\left(1+\nu^{-1}\right)\,, \\
E_t^{\mathbb{Q}}[\text{VIX}_{t,\tau}^3] &= \mu_t^3\left[e^{-\lambda}\left(\frac{2d_\lambda^3}{\nu^2} + 3\frac{d_\lambda^3}{\nu} + d_\lambda^3\right) + (\lambda^3+3\lambda^2+\lambda)d_\lambda^3\varsigma^2 C_{\nu,\varsigma} + 3(\lambda^2+\lambda)d_\lambda^3\varsigma C_{\nu,\varsigma} + 2\lambda d_\lambda^3 C_{\nu,\varsigma}\right]\,, \\
E_t^{\mathbb{Q}}[\text{VIX}_{t,\tau}^4] &= \mu_t^4\left[e^{-\lambda}\left(6\frac{d_\lambda^4}{\nu^3} + 11\frac{d_\lambda^4}{\nu^2} + 6\frac{d_\lambda^4}{\nu} + d_\lambda^4\right) + (\lambda^4+6\lambda^3+7\lambda^2+\lambda)d_\lambda^4\varsigma^3 D_{\nu,\varsigma} + 6(\lambda^3+3\lambda^2+\lambda)d_\lambda^4\varsigma^2 D_{\nu,\varsigma}\right] \\
&\quad + \mu_t^4\left(11(\lambda^2+\lambda)d_\lambda^4\varsigma D_{\nu,\varsigma} + 6\lambda d_\lambda^4 D_{\nu,\varsigma}\right)\,,
\end{aligned}
\tag{31}
$$

where

$$
C_{\nu,\varsigma} = \frac{\nu^2+3\nu+2}{\nu^2\varsigma^2}\,, \qquad D_{\nu,\varsigma} = \frac{\nu^3+6\nu^2+11\nu+6}{\nu^3\varsigma^3}\,.
$$

The MEM-J generates a mixture of distributions conceptually similar to the mixture generated by the Laguerre expansion used in Mencia and Sentana (2016), but it has a further interpretation for the tail-events as generated by jumps.

For every $\tau = 1,\ldots,5$, based on the parametric expressions (31), we estimate the MEM-Gamma and the MEM-J models by GMM exploiting the time-series of the first four risk-neutral moments of the VIX. The GMM estimates of the parameters are reported in Table 1. The estimates of the parameter ν obtained under the Gamma distribution are always significant and they decrease with maturity. Since ν is the inverse of the variance of η_t in the Gamma specification, this evidence correctly signals the increase in the uncertainty around future values of the VIX at longer horizons. However, the MEM with Gamma distributed innovations does a poor job matching the structure of the higher moments of the VIX under the $\mathbb{Q}$-measure. Indeed, the

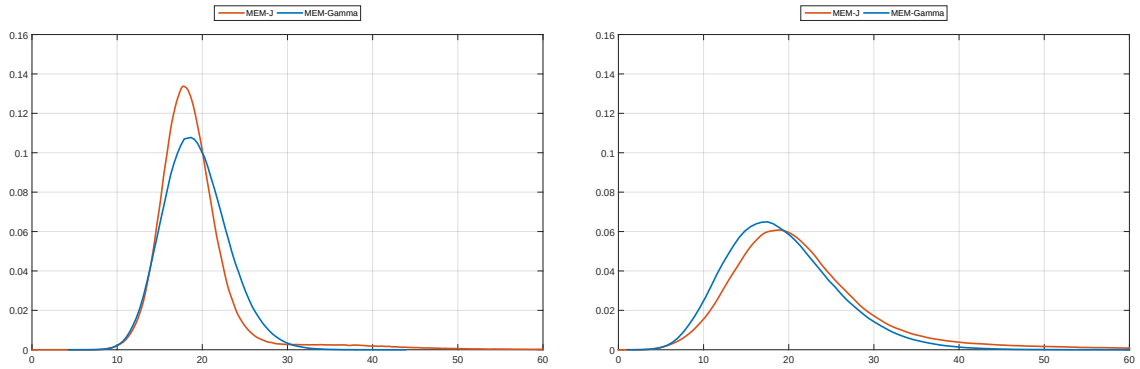


Figure 8: Steady-state risk neutral densities of VIX_τ implied by the MEM-Gamma model and by the MEM-J model of Caporin et al. (2017) for $\tau = 1$ (upper panel) and $\tau = 5$ (lower panel).

J-test for over-identifying restrictions, which is distributed as a χ^2 with 2 degrees of freedom, strongly rejects the null hypothesis that the higher moments are those implied by a Gamma distribution. Instead, adding flexibility to the higher moments of VIX by adding the jump term, as in the MEM-J specification, provides a good description of the linkages between the VIX risk-neutral moments, since the p-value of the J-test is larger than 5% for every $\tau = 1, \ldots, 5$. Looking at the parameter estimates, we note that $\hat{\nu}$ decreases with maturity for both MEM specifications, thus signaling an increase of the variance at longer horizons. In the MEM-J, the mixing parameter λ, which governs the expected number of jumps in each month, is rather stable across τ and close to 0.5. This means an average of 1 jump every second month. Notably, the parameter λ is always significant, thus indicating that tail events, namely those generated by jumps, are priced in the VIX options. A graphical illustration of contribution of jumps on the probability of tail events is provided in Figure 8. The figure reports the model-implied RND of the VIX under $\mathbb{Q}$ in "steady state" that is obtained by setting μ_t to its long-run value represented by its sample mean, $\bar{\mu} = \frac{1}{T}\sum_{t=1}^{T} \text{Mean}_{t,\tau}^{\mathbb{Q}}$. In steady state, the option-implied probability of observing VIX larger than 31% is negligible under the MEM-Gamma for $\tau = 1$, while it is approximately 4.5% if jumps are included in the model. Analogous evidence emerges for $\tau = 5$.

6.4 Variance swap term-structure

In this paragraph, we analyze the term structure of the (annualized percentage) realized variance (RV) in terms of variance swaps (VS). As discussed in Section 2, the square of VIX can be seen as the swap rate of a VS maturing in one month, that is $\text{VS}_{t,t+1} = E_t^{\mathbb{Q}}[\text{RV}_{t,t+1}] = \text{VIX}_t^2$, and we assume zero interest rate, which is indeed negligible in the sample under investigation. This relationship can be generalized to link 1-month forward VS prices to the second moment of the VIX, as shown below

$$E_t^{\mathbb{Q}}[\text{RV}_{t+\tau,t+\tau+1}] = E_t^{\mathbb{Q}}[\text{VIX}_{t+\tau}^2]. \tag{32}$$

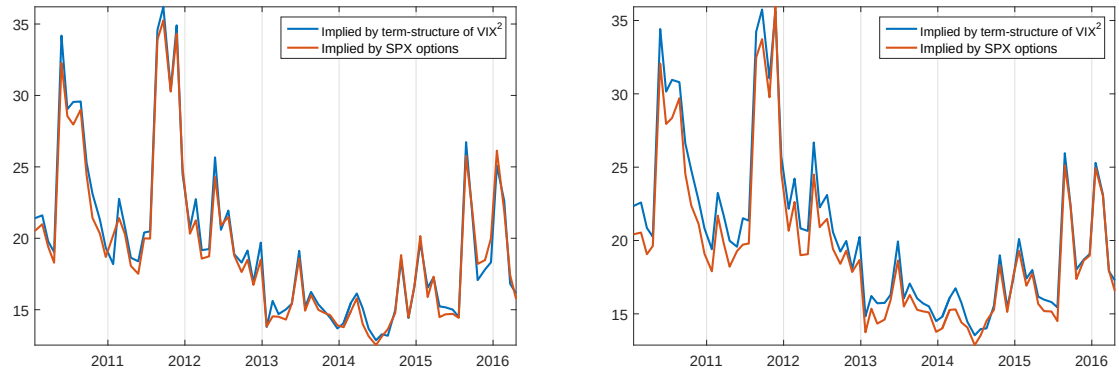
Equation (32) clearly generalizes the definition of VIX, which is recovered for $\tau = 0$. By aggregating the terms on the RHS of (32) over τ one can obtain variance swap prices for maturities longer than one month, that is

$$\mathrm{VS}_{t,t+n} = E_t^{\mathbb{Q}}[\mathrm{RV}_{t,t+n}] = \frac{1}{n}\sum_{\tau=1}^{n} E_t^{\mathbb{Q}}[\mathrm{RV}_{t+\tau-1,t+\tau}] = \frac{1}{n}\sum_{\tau=1}^{n} E_t^{\mathbb{Q}}[\mathrm{VIX}^2_{t+\tau-1}]. \tag{33}$$

The relation in (33) holds under the inherent assumption that the joint market of VIX and RV is priced consistently under the unique risk neutral measure $\mathbb{Q}$, so that

$$E_t^{\mathbb{Q}}[\mathrm{VIX}^2_{t+\tau}] = E_t^{\mathbb{Q}}\left[E_{t+\tau}^{\mathbb{Q}}[\mathrm{RV}_{t+\tau,t+\tau+1}]\right] = E_t^{\mathbb{Q}}[\mathrm{RV}_{t+\tau,t+\tau+1}]\,.$$

In principle, by generalizing the VIX formula (3), one can always replicate $\mathrm{VS}_{t,\tau}$ through a portfolio of SPX options expiring in n months. This provides a tool to assess whether the SPX and the VIX markets are consistent with each other. In Figure 9 we report the time series of $\mathrm{VS}_{t,t+\tau}$ for $\tau = 3$ months (left panel) and $\tau = 4$ months (right panel). Each plot displays the VS computed from SPX options by extending formula (3) (red line) and the VS implied by the RND of VIX through (33) (blue line). Figure 9 suggests that there are no profitable arbitrage opportunities based on trading VS across the SPX and the VIX markets. In particular, this means that the second risk-neutral moment of the VIX can be regarded as an equity derivative, as it can be replicated by combining short and long positions on SPX options expiring in $\tau + n - 1$ and $\tau + n$ months, respectively.



(a) $\tau = 3$ (b) $\tau = 4$

Figure 9: Time series of 3-months and 4-months VS. The figures display VS implied by SPX market (red line) and the VS implied by the VIX market (blue line).

Based on the time series of $E_t^{\mathbb{Q}}[\mathrm{VIX}^2_{t+\tau}]$, we characterize the driving factors of the VS term-structure. As starting point we consider the simple model below

$$y_{t,\tau} = e^{\kappa(\tau+1)} x_t \tag{34}$$

	$\tau=0$	$\tau=1$	$\tau=2$	$\tau=3$	$\tau=4$	$\tau=5$
$\hat{\kappa}$	0.3130	0.2795	0.2447	0.2110	0.1835	0.1627
$S.E.(\kappa)$	0.0676	0.0342	0.0247	0.0194	0.0162	0.0139
t-test	4.6303^a	8.1650^a	9.9131^a	10.8702^a	11.3064^a	11.7156^a

Table 2: Estimate of the time-to-maturity compounding factor. The standard errors are computed with the Newey-West robust estimator. *a*, *b* and *c* stand for significance at 1%, 5%, and 10%, respectively. For $\tau=0$, $E_t^{\mathbb{Q}}[\text{VIX}_{t+\tau}^2]$ is VIX_t^2.

where $y_{t,\tau}=E_t^{\mathbb{Q}}[\text{RV}_{t+\tau,t+\tau+1}]$, $x_t=\text{RV}_{t-1,t}$. This model represents a spot-future parity, where κ is a time-to-maturity compounding factor, which determines the slope of the term-structure of VS.

For each given τ, the parameter κ can be estimated by a simple regression in logs as

$$\zeta_{t,\tau}=\kappa(\tau+1)+u_{t,\tau},\quad t=1,\dots T \tag{35}$$

where $\zeta_{t,\tau}=\log(E_t^{\mathbb{Q}}[\text{VIX}_{t+\tau}^2])-\log(\text{RV}_{t-1,t})$, with observations on $\text{RV}_{t-1,t}$ constructed by cumulating daily RV over the 30 days period between $t-1$ and t. The daily RV is constructed from 5 minute returns on SPX. The estimation results are reported in Table 2. The value of κ estimated for $\tau=0$ is twice as large as the value of κ estimated for $\tau=5$, thus signaling a rather steep downward sloped term structure of VS. The decreasing behavior of the estimates of k signals that model (34) does not provide an accurate description of the term structure of VS, as it prescribes a constant positive drift in RV under $\mathbb{Q}$.

We therefore suggest a generalization of (34) by including a term accounting for the mean-reversion of RV, that is

$$y_{t,\tau}=e^{\kappa(\tau+1)}x_t\cdot mr_t \tag{36}$$

where mr_t is a multiplicative term that adjusts for the mean-reversion. We assume that mr_t is a function of the ratio between $\text{RV}_{t,t-1}$ and VIX_t, and it determines the level of the term-structure of VS relative to the current level of the underlying. The econometric specification becomes

$$\zeta_{t,\tau}=\kappa(\tau+1)+\alpha\xi_t+u_t,\quad t=1,\dots T \tag{37}$$

where $\xi_t=\log(\text{RV}_{t-1,t})-\log(\text{VIX}_t^2)$ can be interpreted as the residuals from the long-run equilibrium between RV and VIX^2. This term is mean-reverting due to the cointegration relationship between the two quantities, see Bollerslev et al. (2013) among others. Table 3 shows that the correction based on the mean-reversion makes the parameter κ much more stable across τ, and approximately equal to 13.8%. We can interpret this number as the monthly growth rate, adjusted by mean-reversion, of the VS. Similarly, Johnson (2012) finds that the term structure implicit in VIX futures cannot be well described by an unique slope factor, but another state-dependent term is needed to provide a good approximation of the surface.

	$\tau = 1$	$\tau = 2$	$\tau = 3$	$\tau = 4$	$\tau = 5$
$\hat{\kappa}$	0.1388	0.1547	0.1461	0.1324	0.1204
$S.E.(\kappa)$	0.0105	0.0118	0.0107	0.0095	0.0086
t-test(κ)	13.2102	13.0934	13.6426	13.9938	14.0555

Table 3: Estimate of the time-to-maturity compounding factor. The standard errors are computed with the Newey-West robust estimator. *a*, *b* and *c* stand for significance at 1%, 5% and 10%, respectively.

7 Conclusion and directions for future research

In this paper, we proposed a methodology based on a finite orthogonal expansion to infer the RND underlying the VIX option prices. The method generalizes the Laguerre expansions and suits cases where the density is supported over the positive real axis. The approach is non-structural since it does not require restrictive parametric assumptions on the underlying asset dynamics, reducing the number of restrictions to be imposed on the form of the RND. Therefore, we drastically reduce the intrinsic risk of misspecification entailed in parametric models. While in our applications we addressed the RND underlying VIX options, the same technique can be applied to different classes of financial derivatives sharing the same characteristics, e.g., to interest rates and inflation. Our empirical study on VIX options highlights the usefulness of this technique to directly study several features of the VIX through its RND and associated moments.

The empirical analysis suggests that the proposed methodology may be particularly useful when the study of risk-premia embedded in option prices is of interest, see Bollerslev and Todorov (2011), Andersen et al. (2015b) and Schneider (2015), or to compute a VaR on volatility (*VolaR*, see Caporin et al., 2017), adjusted for risk aversion as in Aït-Sahalia and Lo (2000). A multivariate functional dynamic model for the RND would be a further natural extension of this work, see Grith et al. (2013). This would provide an alternative to the parametric methodology of Andersen et al. (2015a) to carry out inference on the underlying processes based on panels of options. Along these lines, the study of the risk premia embedded in VIX options could be carried out by studying the shape and the variability over time of the pricing kernel in a bivariate setting that also includes the SPX. This would further extend the work of Song and Xiu (2016) to potentially assess the risk aversion parameter of investors using the VIX for hedging purposes as opposed to those taking speculative positions on it.

References

Abramowitz, M. and Stegun, I. A. (1964). *Handbook of mathematical functions with formulas, graphs, and mathematical tables*. Dover, New York.

Aït-Sahalia, Y. and Duarte, J. (2003). Nonparametric option pricing under shape restrictions. *Journal of Econometrics*, 116(1-2):9–47.

Aït-Sahalia, Y. and Lo, A. W. (1998). Nonparametric estimation of state-price densities implicit in financial asset prices. *The Journal of Finance*, 53(2):499–547.

Aït-Sahalia, Y. and Lo, A. W. (2000). Nonparametric risk management and implied risk aversion. *Journal of Econometrics*, 94(1fi?!2):9 – 51.

Andersen, T. G., Fusari, N., and Todorov, V. (2015a). Parametric inference and dynamic state recovery from option panels. *Econometrica*, 83(3):1081–1145.

Andersen, T. G., Fusari, N., and Todorov, V. (2015b). The risk premia embedded in index options. *Journal of Financial Economics*, 117(3):558 – 584.

Bardgett, C., Gourier, E., and Leippold, M. (2014). Inferring volatility dynamics and risk premia from the S&P 500 and VIX markets. Technical report, Swiss Finance Institute Research Paper No. 13-40.

Bayer, C., Gatheral, J., and Karlsmark, M. (2013). Fast Ninomiya–Victoir calibration of the double-mean-reverting model. *Quantitative Finance*, 13(11):1813–1829.

Bergomi, L. (2008). Smile dynamics III. *Risk*, October:90–96.

Bollerslev, T., Osterrieder, D., Sizova, N., and Tauchen, G. (2013). Risk and return: Long-run relations, fractional cointegration, and return predictability. *Journal of Financial Economics*, 108(2):409 – 424.

Bollerslev, T. and Todorov, V. (2011). Tails, fears and risk premia. *The Journal of Finance*, 66(6):2165–2211.

Breeden, D. T. and Litzenberger, R. H. (1978). Prices of state-contingent claims implicit in option prices. *The Journal of Business*, 51(4):621–51.

Brigo, D. and Mercurio, F. (2002). Lognormal-mixture dynamics and calibration to market volatility smiles. *International Journal of Theoretical and Applied Finance*, 5(4):427–446.

Britten-Jones, M. and Neuberger, A. (2000). Option prices, implied price processes, and stochastic volatility. *The Journal of Finance*, 55(2):839–866.

Buehler, H. (2006). Consistent variance curve models. *Finance and Stochastics*, 10(2):178–203.

Caporin, M., Rossi, E., and Santucci de Magistris, P. (2017). Chasing volatility - a persistent multiplicative error model with jumps. Technical report, Forthcoming on the Journal of Econometrics.

Carr, P. and Lee, R. (2007). Realized volatility and variance: Options via swaps. *Risk*, 20(5):76–83.

Carr, P. and Lee, R. (2009). Volatility derivatives. *Annual Review of Financial Economics*, pages 14.1–14.21.

Carr, P. and Madan, D. (1998). Towards a theory of volatility trading. In *Volatility: New estimation techniques for pricing derivatives, ed. R. Jarrow, chap. 29*, pages 417–427. Risk Publications.

Carr, P. and Wu, L. (2006). A tale of two indices. *The Journal of Derivatives*, 13(3):13–29.

Carr, P. and Wu, L. (2009). Variance risk premiums. *Review of Financial Studies*, 22(3):1311–1341.

CBOE (2015). The CBOE volatility index VIX. White Paper. Available at .

Christoffersen, P., Jacobs, K., and Mimouni, K. (2010). Volatility dynamics for the S&P500: evidence from realized volatility, daily returns, and option prices. *Review of Financial Studies*, 23(8):3141–3189.

Cont, R. (2006). Model uncertainty and its impact on the pricing of derivative instruments. *Mathematical Finance*, 16(3):519–547.

Cont, R. and Kokholm, T. (2013). A consistent pricing model for index options and volatility derivatives. *Mathematical Finance*, 23(2):248–274.

Corrado, C. J. and Su, T. (1996a). Skewness and kurtosis in S&P 500 index returns implied by option prices. *Journal of Financial Research*, 19(2):175–192.

Corrado, C. J. and Su, T. (1996b). S&P 500 index option tests of Jarrow and Rudd's approximate option valuation formula. *Journal of Futures Markets*, 16(6):611–629.

Coutant, S., Jondeau, E., and Rockinger, M. (2001). Reading PIBOR futures options smiles: The 1997 snap election. *Journal of Banking & Finance*, 25(11):1957–1987.

Duffie, D., Pan, J., and Singleton, K. (2000). Transform analysis and asset pricing for affine jump-diffusions. *Econometrica*, 68(6):1343–1376.

Filipovic, D., Mayerhofer, E., and Schneider, P. (2013). Density approximations for multivariate affine jump-diffusion processes. *Journal of Econometrics*, 176(2):93 – 111.

Gatheral, J. (2008). Consistent modeling of SPX and VIX options. In *Presentation at the Fifth World Congress of the Bachelier Finance*.

Grith, M., Hardle, W., and Park, J. (2013). Shape invariant modeling of pricing kernels and risk aversion. *Journal of Financial Econometrics*, 11(2):370.

Hazewinkel, M. (1988). *Encyclopaedia of mathematics: C An updated and annotated translation of the Soviet 'Mathematical Encyclopaedia'*. Springer.

Hewitt, E. (1954). Remark on orthonormal sets in L2(a, b). *The American Mathematical Monthly*, 61(4):249–250.

Huskaj, B. and Nossman, M. (2013). A term structure model for VIX futures. *Journal of Futures Markets*, 33(5):421–442.

Jarrow, R. and Rudd, A. (1982). Approximate option valuation for arbitrary stochastic processes. *Journal of Financial Economics*, 10(3):347–369.

Jiang, G. J. and Tian, Y. S. (2005). The model-free implied volatility and its information content. *Review of Financial Studies*, 18(4):1305–1342.

Johnson, T. L. (2012). Equity risk premia and the VIX term structure. Technical report, Forthcoming on the Journal of Financial and Quantitative Analysis.

Jondeau, E. and Rockinger, M. (2001). Gram-Charlier densities. *Journal of Economic Dynamics and Control*, 25(10):1457–1483.

Lee, R. and Wang, D. (2009). Displaced lognormal volatility skews: analysis and applications to stochastic volatility simulations. *Annals of Finance*, 8(2):159–181.

Madan, D. B. and Milne, F. (1994). Contingent claims valued and hedged by pricing and investing in a basis. *Mathematical Finance*, 4(3):223–245.

Mencia, J. and Sentana, E. (2013). Valuation of VIX derivatives. *Journal of Financial Economics*, 108(2):367 – 391.

Mencia, J. and Sentana, E. (2016). Volatility-related exchange traded assets: an econometric investigation. Technical report, Forthcoming on the Journal of Business & Economic Statistics.

Ñíguez, T.-M. and Perote, J. (2012). Forecasting heavy-tailed densities with positive Edgeworth and Gram-Charlier expansions. *Oxford Bulletin of Economics and Statistics*, 74(4):600–627.

Pan, J. (2002). The jump-risk premia implicit in options: evidence from an integrated time-series study. *Journal of Financial Economics*, 63(1):3 – 50.

Pastorello, S., Renault, E., and Touzi, N. (2000). Statistical inference for random-variance option pricing. *Journal of Business & Economic Statistics*, 18(3):358–367.

Rompolis, L. S. and Tzavalis, E. (2008). Recovering risk neutral densities from option prices: a new approach. *Journal of Financial and Quantitative Analysis*, 43(04):1037–1053.

Rudin, W. (1987). *Real and complex analysis*. Tata McGraw-Hill Education.

Schneider, P. (2015). Generalized risk premia. *Journal of Financial Economics*, 116(3):487 – 504.

Sepp, A. (2008a). Pricing options on realized variance in the Heston model with jumps in returns and volatility. *Journal of Computational Finance*, 11(4):33–70.

Sepp, A. (2008b). VIX options pricing in a jump-diffusion model. *Risk*, pages 84–89.

Shohat, J. (1942). Note on closure for orthogonal polynomials. *Bulletin of the American Mathematical Society*, 48(6):488–490.

Song, Z. and Xiu, D. (2016). A tale of two option markets: pricing kernels and volatility risk. *Journal of Econometrics*, 90:176–196.

Szegö, G. (1939). *Orthogonal polynomials*. American Mathematical Society.

Todorov, V. and Tauchen, G. (2011). Volatility jumps. *Journal of Business & Economic Statistics*, 29(3):356–371.

Todorov, V., Tauchen, G., and Grynkiv, I. (2014). Volatility activity: Specification and estimation. *Journal of Econometrics*, 178, Part 1:180 – 193.

Wang, Z. and Daigler, R. T. (2011). The performance of VIX option pricing models: Empirical evidence beyond simulation. *Journal of Futures Markets*, 31(3):251–281.

Xiu, D. (2014). Hermite polynomial based expansion of European option prices. *Journal of Econometrics*, 179(2):158–177.

Zhang, J. E. and Zhu, Y. (2006). VIX futures. *Journal of Futures Markets*, 26(6):521–531.

Zhang, L., Mykland, P. A., and Aït-Sahalia, Y. (2011). Edgeworth expansions for realized volatility and related estimators. *Journal of Econometrics*, 160(1):190–203.

Zhu, Y. and Zhang, J. E. (2007). Variance term structure and VIX futures pricing. *International Journal of Theoretical and Applied Finance*, 10(01):111–127.

A Appendix

A.1 Proof of Theorem 3.1

Some preliminary are stated before the main proof. First, we recall a standard result of functional analysis. For a proof the reader may refer, e.g., to Rudin (1987)-Theorem 4.14.

Lemma A.1. *Assume that* $\phi^{-\frac{1}{2}} f_{\mathbb{Q}} \in L^2(\mathcal{D})$ *and* $supp(f_{\mathbb{Q}}) \subseteq \mathcal{D}$. *Consider the Hilbert space* $(\mathcal{H}_\phi, \langle \cdot\,, \cdot \rangle)$ *defined by*

$$\mathcal{H}_\phi = \left\{ \psi,\ \phi^{-\frac{1}{2}}\psi \in L^2(\mathcal{D}) \right\}, \quad \langle \psi_1, \psi_2 \rangle := \int_{\mathcal{D}} \psi_1(x)\psi_2(x)\frac{1}{\phi(x)}dx\,, \quad \forall \psi_1, \psi_2 \in \mathcal{H}_\phi\,,$$

and the subspace

$$\mathcal{H}_\phi^* = Cl\left(span\left\{ \phi h_k^\phi, k \in \mathbb{N} \right\} \right) \subseteq \mathcal{H}_\phi\,.$$

Then, there exists a sequence $(c_k)_{k\in\mathbb{N}}$ *such that function*

$$f_{\mathbb{Q}}^{(\infty)} := \lim_{n\to+\infty} \phi\left(1 + \sum_{k=1}^{k_n} c_k h_k^\phi \right) \quad in\ \mathcal{H}_\phi$$

solves the minimum distance problem

$$f_{\mathbb{Q}}^{(\infty)} = \underset{\psi\in\mathcal{H}_\phi^*}{argmin} \left\langle \psi - f_{\mathbb{Q}}, \psi - f_{\mathbb{Q}} \right\rangle^{\frac{1}{2}}. \tag{38}$$

In particular, if $\mathcal{H}_\phi^* = \mathcal{H}_\phi$ *we have* $f_{\mathbb{Q}}^{(\infty)} = f_{\mathbb{Q}}$ *almost everywhere.*

Definition A.2 (Closed polynomial set in $\mathcal{H}_\phi$)**.** The kernel ϕ is said to generate closed polynomial sets if

$$\mathrm{Cl}\left(\mathrm{span}\left\{ x^k, k \in \mathbb{N} \right\} \right) = L^2_{\phi(x)dx}(\mathcal{D}). \tag{39}$$

In this case, we say that either $(x^k)_{k\in\mathbb{N}}$ or $(h_k^\phi)_{k\in\mathbb{N}}$ is closed with respect to ϕ.

The following result provides necessary and sufficient conditions to determine whether ϕ generates closed polynomial sets. The results, whose proof is deferred to the end of the section, extends the classic result of closure of Laguerre polynomials.

Theorem A.3 (Conditions to the closure of $(h_k^\phi)_{k\in\mathbb{N}}$)**.** *Let* ϕ *be a positive integrable function and* $\mathcal{D} = [0, +\infty[$.

(i) If $\lim_{x\to+\infty} \phi(x)e^{\varsigma x^{\frac{1}{2}}} = 0$ *for some* $\varsigma > 0$ *and there exists a polynomial* p *such that* $p\phi$ *is bounded, then* ϕ *generates closed polynomial sets.*

(ii) If $\lim_{x\to+\infty} \phi(x)e^{\varsigma x^{\frac{1}{2}-\gamma}} > 0$ *for some* $\gamma, \varsigma > 0$*, then* ϕ *does not generate closed polynomial sets.*

We can now prove Theorem 3.1. In the following, $\mathcal{H}_\phi$ and $\mathcal{H}_\phi^*$ refer to the Hilbert spaces defined in Lemma A.1.

If $(h_k^\phi)_{k\in\mathbb{N}}$ is closed with respect to ϕ, then $\mathcal{H}_\phi = \mathcal{H}_\phi^*$ and from Lemma A.1 it follows that $f_{\mathbb{Q}}^* = f_{\mathbb{Q}}$ whenever $\phi^{-\frac{1}{2}} f_{\mathbb{Q}} \in L^2(\mathcal{D})$. The first implication can be readily shown by noticing that $f_{\mathbb{Q}} \in \mathcal{H}_\phi$ implies $\phi^{-1} f_{\mathbb{Q}} \in L^2_{\phi(x)dx}(\mathcal{D})$. Then, the closure of $(h_k^\phi)_{k\in\mathbb{N}}$ implies that $\phi^{-1} f_{\mathbb{Q}}$ can be approximated by a certain polynomial series $a_0 + a_1 x + a_2 x^2 + \dots$ in $L^2_{\phi(x)dx}(\mathcal{D})$ or equivalently that $f_{\mathbb{Q}}$ can be approximated by a certain series $c_0\phi h_0^\phi + c_1\phi h_1^\phi + c_2\phi h_2^\phi \dots$ in $\mathcal{H}_\phi$.

Then, (a) follows immediately from Lemma A.1 by noticing that the assumptions on ϕ imply $\mathcal{H}_\phi = \mathcal{H}_\phi^*$, in view of Theorem A.3-(i). To prove (13), we observe that for every $\phi \in \mathcal{H}_\phi^*$ and every $n \in \mathbb{N}$

$$\begin{aligned}\left|\int_0^{+\infty} \Pi(x) f_{\mathbb{Q}}^{(n)}(x)dx - \int_0^{+\infty} \Pi(x) f_{\mathbb{Q}}^{(\infty)}(x)dx\right| &\le \int_0^{+\infty} \Pi(x)\left|f_{\mathbb{Q}}^{(n)}(x) - f_{\mathbb{Q}}^{(\infty)}(x)\right|dx \\ = \int_{\mathcal{D}} \phi^{\frac{1}{2}}(x)\Pi(x)\left|\phi^{-\frac{1}{2}}(x) f_{\mathbb{Q}}^{(n)}(x) - \phi^{-\frac{1}{2}}(x) f_{\mathbb{Q}}^{(\infty)}(x)\right|dx &\le b \cdot \left\langle f_{\mathbb{Q}}^{(n)} - f_{\mathbb{Q}}^{(\infty)}, f_{\mathbb{Q}}^{(n)} - f_{\mathbb{Q}}^{(\infty)}\right\rangle^{\frac{1}{2}},\end{aligned}$$

where $b = \left\langle \phi^{\frac{1}{2}}\Pi, \phi^{\frac{1}{2}}\Pi\right\rangle^{\frac{1}{2}}$ is finite by hypothesis. Then, (13) follows from Lemma A.1.

Proof of Theorem A.3

The following lemma is needed to prove one part of the theorem.

Lemma A.4. *Suppose that* ϕ^* *generates closed polynomial sets and* $\phi = h \cdot \phi^*$*, where* h *is bounded and positive a.e. on* $\mathcal{D}$*. Then* ϕ *generates closed polynomial sets.*

Proof. By the Riesz-Fischer characterization it suffices to prove that if there exists $f \in L^2_{\phi(x)dx}(\mathcal{D})$ such that

$$\int_{\mathcal{D}} f(x)x^k\phi(x)dx = 0 \quad \forall k \in \mathbb{N},$$

then it must hold that $f(x) = 0$ a.e. on $\mathcal{D}$. Define $g(x) = h(x)f(x)$, then

$$\int_{\mathcal{D}} g^2(x)\phi^*(x)dx \le \max_{x\in\mathcal{D}} h(x) \cdot \int_{\mathcal{D}} f^2(x)\phi(x)dx < +\infty$$

which proves $g \in L^2_{\phi^*(x)dx}(\mathcal{D})$. Furthermore

$$\int_{\mathcal{D}} g(x)x^k\phi^*(x)dx = \int_{\mathcal{D}} f(x)x^k\phi(x)dx = 0$$

for every $k \in \mathbb{N}$, which implies in view of hypothesis that $g(x) = 0$ a.e. on $\mathcal{D}$ and therefore $f(x) = 0$ a.e. on $\mathcal{D}$ due to positivity assumptions on $h(x)$. □

To prove statement (i) we start by recalling a classic result due to Hewitt (1954) showing that every bounded function ψ supported on the entire real line and such that

$$\lim_{|x|\to+\infty} \psi(x)e^{\varsigma|x|} = 0 \tag{40}$$

generates closed polynomial sets. Based on this result, statement (i) can be proven under the additional hypothesis that ϕ is bounded. Indeed, under this assumption, the function $\psi(x) = |x|\phi(|x|^{\frac{1}{2}})$ is bounded on $\mathbb{R}$ and satisfies (40), and therefore it generates closed polynomial sets. Statement (i) is then a straightforward consequence of the main theorem reported in Shohat (1942). To prove statement (i) with no additional requirements on ϕ, we remark that by hypothesis there exist a polynomial p and $\varsigma^* > 0$ such that the function ϕ^* defined by

$$\phi^*(x) := p(x)e^{\varsigma^*\sqrt{x}}\phi(x)$$

is bounded on $\mathcal{D}$. Since ϕ^* clearly preserves the same integrability and asymptotic properties of ϕ, then it generates closed polynomial sets. Now, consider f such that $f^2\phi$ is integrable and

$$\int_{\mathcal{D}} f(x)x^k\phi(x)dx = 0, \quad \forall k \in \mathbb{N}.$$

Moreover, define g as

$$g(x) = e^{-\varsigma^*\sqrt{x}}f(x), \quad x \in \mathcal{D}.$$

We have

$$\int_{\mathcal{D}} g(x)^2\phi^*(x)dx \le \sup_{x\in\mathcal{D}} \left|p(x)e^{-\varsigma^*\sqrt{x}}\right| \int_{\mathcal{D}} f^2(x)\phi(x)dx < +\infty.$$

On the other hand, for every $k \in \mathbb{N}$

$$\int_{\mathcal{D}} g(x)x^k\phi^*(x)dx = \int_{\mathcal{D}} f(x)x^k p(x)\phi(x)dx = 0,$$

which proves $g(x) = 0$ and therefore $f(x) = 0$ a.e. on $\mathcal{D}$. Then, statement (i) is proved.

The proof of statement (ii) is based on a known counterexample in the theory of orthogonal polynomials (cf. entry "Closed system of elements" in Hazewinkel (1988)), showing that every function ψ of the form

$$\psi(x) = e^{-|x|^{\frac{2m}{2m+1}}}, \quad x \in \mathbb{R},\ m \in \mathbb{N},$$

does not generate closed polynomial sets. By combining this counterexample with the results of Shohat (1942), we can prove that the function ψ supported on $[0, +\infty[$ and defined by

$$\psi(x) = e^{-x^{\frac{m}{2m+1}}}, \quad x \ge 0,\ m \in \mathbb{N}.$$

does not generate closed polynomial sets. By a change of variable and through the Riesz-Fischer characterization, one can extend the latter result to the case where ψ is supported on $[x_0, +\infty[$ and is of the form

$$\psi(x) = e^{-\varsigma(x-x_0)^{\frac{m}{2m+1}}}, \quad x \geq x_0,\ m \in \mathbb{N}.$$

for some $\varsigma > 0$ and $x_0 \geq 0$. To prove statement (ii), then, we proceed by contradiction and suppose that there exists an integrable function ϕ, supported on $[0, +\infty[$ and such that $\lim_{x\to+\infty} \phi(x)e^{\varsigma x^{\frac{1}{2}-\gamma}} > 0$ for some $\gamma, \varsigma > 0$, which generates closed polynomial sets. To this aim, we observe that by the hypothesis made on the right-tail of ϕ, there exists $x_0 \geq 0$ such that $\phi(x) > 0$ for all $x \geq x_0$. The closure property of polynomial sets with respect to ϕ holds in particular when the support is restricted, by truncation, to $[x_0, +\infty[$. Furthermore, the function h defined by

$$h(x) = e^{-\varsigma(x-x_0)^{\frac{m}{2m+1}}}\phi(x)^{-1},$$

is bounded on $[x_0, +\infty[$, for m sufficiently large. Then, as a consequence of Lemma A.4, the function $e^{-\varsigma(x-x_0)^{\frac{m}{2m+1}}}$ generates closed polynomial sets on $[x_0, +\infty[$, which is a contradiction. The proof is thereby concluded.

□

A.2 Proof of Proposition 5.1

For the notational simplicity, throughout the proof we omit the dependence on t, τ of $\mathbf{Q}^{(n)}$, C^{Obs}, P^{Obs}, C, and $\mathcal{P}$. We also set $r = 0$. Moreover, we denote by $(c_k)_{k\in\mathbb{N}}$ and $f_{\mathbb{Q}}^{(n)}$ the quantities defined in Theorem 3.1-(a). For every $n \in \mathbb{N}$ we have

$$\begin{aligned}
&(K_M - K_1) \cdot E\left[\mathbf{Q}^{(n)}(c^\star(n))\right] = E\left[\int_I \left(C_K^{Obs} - C_K^{(n)}(c^\star(n))\right)^2 + \left(P_K^{Obs} - P_K^{(n)}(c^\star(n))\right)^2 dK\right] \\
&\leq E\left[\int_I \left(C_K + \epsilon_K^C - C_K^{(n)}(c)\right)^2 + \left(\mathcal{P}_K + \epsilon_K^P - P_K^{(n)}(c)\right)^2 dK\right] \\
&= E\left[\int_I \left(C_K - C_K^{(n)}(c)\right)^2 + \left(\epsilon_K^C\right)^2 + 2\epsilon_K^C\left(C_K - C_K^{(n)}(c)\right) dK\right] \\
&+ E\left[\int_I \left(\mathcal{P}_K - P_K^{(n)}(c)\right)^2 + \left(\epsilon_K^P\right)^2 + 2\epsilon_K^P\left(\mathcal{P}_K - P_K^{(n)}(c)\right) dK\right] \\
&= \int_I \left(C_K - C_K^{(n)}(c)\right)^2 + \left(\mathcal{P}_K - P_K^{(n)}(c)\right)^2 dK + E\left[\int_I \left(\epsilon_K^C\right)^2 + \left(\epsilon_K^P\right)^2 dK\right].
\end{aligned}$$

Then, by (4), we get

$$\begin{aligned}
&\int_I \left(C_K - C_K^{(n)}(c)\right)^2 + \left(\mathcal{P}_K - P_K^{(n)}(c)\right)^2 dK \\
&= \int_I \left[\int_{\mathcal{D}} \left(f_{\mathbb{Q}}(x) - f_{\mathbb{Q}}^{(n)}(x)\right)(K-x)^+ dx\right]^2 + \left[\int_{\mathcal{D}} \left(f_{\mathbb{Q}}(x) - f_{\mathbb{Q}}^{(n)}(x)\right)(x-K)^+ dx\right]^2 dK \\
&\leq \int_I \left[\int_{\mathcal{D}} \left|f_{\mathbb{Q}}(x) - f_{\mathbb{Q}}^{(n)}(x)\right| |K-x| dx\right]^2 dK \leq b \cdot \int_{\mathcal{D}} \left|f_{\mathbb{Q}}(x) - f_{\mathbb{Q}}^{(n)}(x)\right|^2 \frac{1}{\phi(x)} dx\,,
\end{aligned}$$

where $b = \int_I \int_0^{+\infty} (K-x)^2 \phi(x) dx dK$ is finite. Then, in view of Theorem 3.1

$$\lim_{n\to+\infty} \int_I \left(C_K - C_K^{(n)}(c)\right)^2 + \left(\mathcal{P}_K - P_K^{(n)}(c)\right)^2 dK = 0\,,$$

which proves (24). The proof is concluded by noticing that, under the additional hypotheses (i) and (ii)

$$E\left[\mathbf{Q}^{(n)}(c^\star(n))\right] \geq \frac{1}{K_M - K_1} E\left[\int_I \left(\epsilon_K^C\right)^2 + \left(\epsilon_K^P\right)^2 dK\right], \quad \forall n \geq \bar{n}\,.$$

B Supplementary material

B.1 A Robust Technique

B.1.1 Orthogonal regressors

We propose to solve the problem of multicollinearity outlined in Section 4 by means of a principal component analysis (PCA), which allows to estimate the coefficients of an expansion of any arbitrarily large order n. The PCA analysis is implemented as follows: first, to avoid scale effects, we standardize each column of $\mathbf{X}$, as

$$\mathbf{Z}_i = \frac{\mathbf{X}_i - \frac{1}{2M}\sum_{j=1}^{2M}\mathbf{X}_{ji}}{\sqrt{\frac{1}{2M-1}\sum_{j=1}^{2M}\left(\mathbf{X}_{ji} - \frac{1}{2M}\sum_{j=1}^{2M}\mathbf{X}_{ji}\right)^2}}, \quad i = 1, \dots, n \tag{41}$$

where $\mathbf{X}_i$ and $\mathbf{X}_{ji}$ denote the i-th column and the j,i-th element of $\mathbf{X}$, respectively. Then, we determine the $2M \times n$ matrix of principal components as $\mathbf{V} = \mathbf{PZ}$ and the $n \times n$ orthonormal matrix of weights $\mathbf{P}$ from the spectral decomposition $\mathbf{ZP} = \mathbf{P}\Lambda$. Lastly, we extract the sub-matrix $\tilde{\mathbf{V}} = \mathbf{V}_{\cdot,1:s}$ of the first s principal components, associated with a given threshold on the explained total variance (e.g. 99%), to be used as regressor. For example, when $n > 10$, the first 4-5 principal components typically explain at least the 99% of the total variance.

Once we have obtained $\mathbf{V}$, we estimate the coefficients $\hat{\gamma} = (\hat{\gamma}_1, \dots, \hat{\gamma}_s)$ of the following regression,

$$\mathbf{Y}^* = \tilde{\mathbf{V}}\gamma + u, \tag{42}$$

where γ represents the loading on the first s principal components. The estimated coefficients $\tilde{c}$ are finally retrieved by reverting the orthogonalization as follows

$$\tilde{c} = (\mathbf{O}\hat{\gamma}) \circ \left[\sqrt{\frac{2M-1}{\sum_{j=1}^{2M}(\mathbf{X}_{j1} - \frac{1}{2M}\sum_{j=1}^{2M}\mathbf{X}_{j1})^2}}, \dots, \sqrt{\frac{2M-1}{\sum_{j=1}^{2M}(\mathbf{X}_{jn} - \frac{1}{2M}\sum_{j=1}^{2M}\mathbf{X}_{jn})^2}}\right]', \tag{43}$$

where $\mathbf{O}$ is the $n \times s$ matrix obtained from the first s columns of $\mathbf{P}$ and $\circ$ denotes the Hadamard product.

B.1.2 Regression through the origin

In regression (42), there is no intercept and the columns of $\tilde{\mathbf{V}}$ have zero-mean by construction, while $\mathbf{Y}^*$ has zero mean if and only if the sample mean of $\mathbf{Y}$ and $\mathbf{X}_0$ coincide. To enforce that $E(\mathbf{Y}^*) = E(\mathbf{Y} - \mathbf{X}_0) = 0$, the initial estimation of the kernel parameters, θ, must be constrained such that $E(\mathbf{X}_0) = E(\mathbf{Y})$. This represents a very mild constraint but it has several practical 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 (42) have zero mean for any order $n \geq 1$ by construction.

Since the principal components are constructed from the standardized regressors $\mathbf{Z}$, when remapping the solution of (42) onto the solution of (20), a constant term equal to $\sum_{i=1}^{n} d_i \mathbf{R}_i$ appears, where $\mathbf{R} = \mathbf{O}\hat{\gamma}$ and $d_i = \frac{\sqrt{\frac{1}{2M-1}} \sum_{j=1}^{2M} \mathbf{X}_{ji}}{\sqrt{\sum_{j=1}^{2M} (\mathbf{X}_{ji} - \frac{1}{2M} \sum_{j=1}^{2M} \mathbf{X}_{ji})^2}}$, $i = 1, \ldots, n$. Therefore, in order to guarantee that the relation in equation (20) holds for any $n \geq 1$, the following constrained optimization is performed

$$
\begin{aligned}
[\gamma_1, \ldots, \gamma_s] \quad = \quad & \underset{\gamma_1, \ldots, \gamma_s}{\operatorname{argmin}} \, \tilde{Q}(t, T; \gamma_1, \ldots, \gamma_s), \\
& \text{s.t.} \quad \sum_{i=1}^{n} d_i \mathbf{R}_i = 0
\end{aligned}
\tag{44}
$$

where $\tilde{Q}(t, T; \gamma_1, \ldots, \gamma_s) = (\mathbf{Y} - \tilde{\mathbf{V}}\gamma)'(\mathbf{Y} - \tilde{\mathbf{V}}\gamma)$. This restriction also guarantees that $\tilde{c}$ is such that there is no systematic pricing error or, in other words, that $\tilde{\varepsilon} = \mathbf{Y} - \mathbf{X}_0 - \mathbf{X}\tilde{c}$ are centered around zero for any n.

B.1.3 Positivity and unitary mass

Since the estimation of $\tilde{c}_1, \ldots, \tilde{c}_n$ is performed on a finite set of option prices, the estimated RND $\tilde{f}_{\mathbb{Q}}^{(n)}$ could display significant negative mass even for large values of n. Therefore, we add an extra implicit constraint to the optimization problem (44). In particular, the optimal parameters $\gamma_1, \ldots, \gamma_s$ are found by solving the following constrained minimum distance problem

$$
\begin{aligned}
[\gamma_1, \ldots, \gamma_s] \quad = \quad & \underset{\gamma_1, \ldots, \gamma_s}{\operatorname{argmin}} \, \tilde{Q}(t, T; \gamma_1, \ldots, \gamma_s | \theta), \\
& \text{s.t.} \quad \sum_{i=1}^{n} d_i \mathbf{R}_i = 0 \\
& \text{s.t.} \quad 1 - \Delta^{\text{pos}} < \int_0^{\infty} \left| \phi(x; \theta) \left(1 + \sum_{k=1}^{n} c_k(\gamma) h_k^{\phi} \right) \right| \mathrm{d}x < 1 + \Delta^{\text{pos}}
\end{aligned}
\tag{45}
$$

where Δ^{pos} = is the tolerance on the unity mass constraint, e.g. $\Delta^{\text{pos}} = 0.000001$, while the coefficients $c_1(\gamma), \ldots, c_n(\gamma)$ are functions of the parameters $\gamma_1, \ldots, \gamma_s$, determined as in (43).

B.1.4 Kernel displacement

Consistency conditions ensured by Theorem 3.1 are rather flexible with respect to the support of the kernel. In principle, it is sufficient that the support of the RND is contained in the support of the kernel. However, if the support of ϕ is too large with respect to the support of $f_{\mathbb{Q}}$, then the expansion (5) is "forced" to converge to zero for all points that are outside the support of

$f_{\mathbb{Q}}$. This has clear disadvantages from an empirical perspective, since the kernel, which is the starting point of the optimization in (45), associated with $c_1, \ldots, c_n = 0$, does not satisfy the constraint of unit mass in $\text{supp}(f_{\mathbb{Q}})$. Even if in general we may assume that $\text{supp}(f_{\mathbb{Q}}) \subseteq \mathbb{R}^+$, when the left tail of the true RND is particularly short around a point $K_{\text{min}} > 0$, this implies that that nearly the whole probability mass is concentrated away from the origin. Since the put price curve of VIX contract normally becomes quickly linear as the strike price approaches the deep OTM region, the RND is expected to display a strong negative skewness associated with a very short left tail. Hence, when $f_{\mathbb{Q}}$ displays such a behavior on the left tail, it may be convenient to choose the kernel ϕ so that the following condition is satisfied

$$\int_0^{K_{min}} \phi(x, \theta) \mathrm{d}x = 0. \tag{46}$$

A simple way to guarantee (46) is to displace the domain of a kernel by K_{min}. The idea of using displaced densities is not new to finance, see e.g. Brigo and Mercurio (2002), and has attracted particular interest in the context of volatility derivatives, see e.g. Carr and Lee (2007) and Lee and Wang (2009). The kernel displacement is done by considering a set K^* of shifted strikes defined as $K^* = [K_1 - K_{min}, \ldots, K_M - K_{min}]$, and defining the matrix of regressors $\mathbf{X}$ with respect to K^*. Once the optimal $\tilde{c}$ are obtained as solution of the of the problem (45) based on K^*, then the estimated RND is determined as follows

$$\tilde{f}_{\mathbb{Q}}^{(n)}(x) = \phi(x - K_{min}; \theta) \left(1 + \sum_{k=1}^{n} \tilde{c}_k h_k^{\phi}(x - K_{min}) \right). \tag{47}$$

which guarantees that $\int_0^{K_{min}} \tilde{f}_{\mathbb{Q}}^{(n)}(x) \mathrm{d}x = 0$. The choice of K_{min} is based on the analysis of the convexity of deeply OTM put prices, see the discussion in Section 6.

B.2 Numerical illustrations

In this section, we test the accuracy of the proposed approach by means of two numerical examples under no-arbitrage. The purpose here is to show that the orthogonal polynomials are able to approximate RNDs, belonging to different families, with a high degree of accuracy. Therefore, we perform the estimation on option prices generated by structural models for which the RND is known in closed-form. The option prices that are thereby considered are arbitrage-free by construction. In this section we illustrate the practical relevance of the asymptotic conditions on the RND required in Theorem 3.1 to ensure a convergent estimation. To obtain the target RND and the related option prices, we consider two simple but popular models. In the first case, the VIX is determined as a function of the instantaneous variance process of the Heston model, as explained in Zhang and Zhu (2006). In the second case, the RND of the log-VIX is assumed to be normal inverse Gaussian (NIG), an approach that is adopted in Huskaj and Nossman (2013).

In both cases, the estimation of $c_1, \dots, c_n$ is performed according to the methodology outlined in Section 4 on a set of $M = 42$ option prices relative to strikes in the interval $[K_1, K_M] = [10, 55]$. The option prices to be matched are generated through direct integration of the RND 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. Notably, the true risk-neutral densities associated with the two models display different decay rates on the tails, thus offering an interesting evaluation on how violating the conditions of Lemma A.1 may possibly generate divergent expansions. Moreover, this numerical exercise provides valuable information on the robustness of the estimates to the initial choice of the expansion kernel. In particular, although the asymptotic properties of ϕ have a tangible effect on the accuracy of the approximation, the choice of its parameters has only a marginal impact, provided that it guarantees that convergence and closure conditions are respected (see Theorem 3.1).

B.3 Heston model

Under Heston dynamics, the undiscounted SPX price $(S_t)_{t\geq 0}$ and its variance $(v_t)_{t\geq 0}$ are generated according to the following SDE

$$\begin{aligned} \mathrm{d}\log S_t &= -\frac{1}{2} v_t \mathrm{d}t + \sqrt{v_t}\mathrm{d}W_t, \\ \mathrm{d}v_t &= k(\bar{v} - v_t)\mathrm{d}t + \eta\sqrt{v_t}\mathrm{d}W_t^*, \end{aligned}$$

where $\mathrm{d}W_t$ and $\mathrm{d}W_t^*$ are correlated Brownian motions with constant correlation ρ. The parameters k and $\bar{v}$ govern the speed of mean reversion and long-run value of v_t respectively, while η is the volatility-of-volatility parameter. Following the approach of Zhang and Zhu (2006), under the Heston model the square of the VIX at time T can be expressed as the following linear function of v_T

$$\mathrm{VIX}_T = 100 \cdot (a_1 \cdot v_T + a_2)^{\frac{1}{2}}, \quad a_1 = \frac{1 - e^{k\tau}}{k\tau}, \quad a_2 = \bar{v}(1 - a_1), \quad \tau = \frac{30}{365}.$$

Moreover, the density of v_T given $v_t = z$ has the following closed-from expression

$$(v_T \mid v_t = z) \sim g, \quad g(s) = C_1 s^{\frac{k\bar{v}}{\eta^2} - \frac{1}{2}} e^{-\frac{2ks}{\eta^2(1-e^{-kT})}} I_{\frac{2k\bar{v}}{\eta^2}-1}(C_2\sqrt{s}),$$

where $C_1 = \frac{2k}{\eta^2(1-e^{-k(T-t)})}$ and $C_2 = 2C_1\sqrt{e^{-k(T-t)}z}$ do not depend on the state variable s and I_ν denotes the modified Bessel function of first kind of order ν. Hence, the RND of VIX_t is also

known in closed-form

$$f_{\mathbb{Q}}(x) = \frac{2}{a} x \cdot g\left(\frac{x^2 - b}{a}\right)$$

and vanilla options prices can be generated through the integral formulas in (4). The support of $f_{\mathbb{Q}}$ is $[\sqrt{a_2}, +\infty[$ and, by the asymptotic properties of I_ν (see for example Abramowitz and Stegun, 1964), it can be shown that $f_{\mathbb{Q}}(x) \sim x^{\alpha^*} e^{-\beta^* x^2 + \gamma^* x}$ as $x \to +\infty$, where the leading term is clearly $e^{-\beta^* x^2}$. Moreover, whenever the support of ϕ strictly contains $[\sqrt{a_2}, +\infty[$, the left-tail decay of $f_{\mathbb{Q}}$ does not influence the integrability of $f_{\mathbb{Q}}^2 \phi^{-1}$. Therefore, the condition $\phi^{\frac{1}{2}} f \in L^2(\mathcal{D})$ of Theorem 3.1 is met for any choice of the kernel among the families of GIG, GW and LN densities. Figure 10 portrays the true RND of the Heston model and the related orthogonal polynomial expansions based on different choices of the kernel. The approximated densities reported in Figure 10 highlight the ability of the expansions based on the GIG and the GW kernels to well recover the original density $f_{\mathbb{Q}}$. On the contrary, the LN kernel fails in approximating $f_{\mathbb{Q}}$, although several corrective terms are considered in the expansion and the convergence condition $f_{\mathbb{Q}} \phi^{-1/2} \in L^2(\mathcal{D})$ is satisfied. The expansion based on the LN kernel proves particularly ineffectual on both tails of $f_{\mathbb{Q}}$. This is a practical consequence of the fact that the LN density does not generate closed polynomial sets (see Theorem A.3), and may serve as an interpretive example of the importance of the hypotheses required by Theorem 3.1.

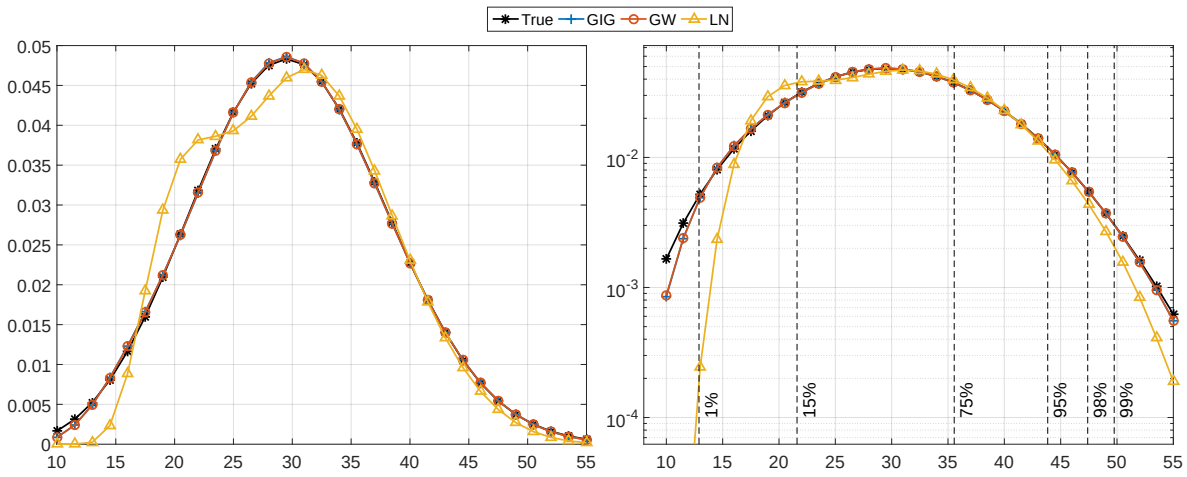


Figure 10: Probability density functions in standard scale (left) and semi-logarithmic scale (right). Comparison between the true density of VIX implied by the Heston model and the estimated RNDs of order 20. The parameters for the Heston model are: $k = 1.71$, $\bar{v} = 0.097$, $\eta = 0.577$, $v(0) = \bar{v}$ and $T - t = 30/365$. The dashed vertical lines on the right panel identify several relevant probability levels and the corresponding quantiles.

B.4 NIG distribution

We assume that the log-VIX at maturity T follows a NIG distribution, that is

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

where $C = \frac{\alpha\delta}{\pi} e^{\delta\gamma}$ is the normalization constant and K_ν denotes the modified Bessel function of the second kind (cf. Abramowitz and Stegun (1964)). Therefore, by the change of variable $s = \log(x)$ we obtain the RND of the VIX

$$f_{\mathbb{Q}}(x) = \frac{1}{x} \cdot g\left(\log(x)\right).$$

The asymptotic properties of K_ν determine polynomial decay of $f_{\mathbb{Q}}$ both on the right and the left tail. It follows that none of the kernels considered here meets the condition $f_{\mathbb{Q}}\phi^{-\frac{1}{2}} \in L^2(\mathbb{R}^+)$. Figure 11 reports the true RND implied by the log-NIG density, and related expansions based on different choices of the kernel. As expected, in all cases the main convergence issues involve the tails. In particular, the expansion based on the GIG kernel is defective on both tails, which is consistent with the fact that a GIG kernel decays more rapidly than the true RND, at both sides. Due to the polynomial decay of the GW kernel on left tail, which accommodates the slow decay of the true RND, the GW-based expansion proves inexact only on the right tail. 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 is a consequence of the fact that the LN is nested within the log-NIG family, and therefore here $f_{\mathbb{Q}}$ is intuitively "closer" to a log-normal than in the previous case.

B.5 Robustness to kernel specification

We now test the robustness of our estimation to the initialization of the kernel parameters, θ. So far, the parameters of the kernels were optimally determined by minimizing the residuals variance for the expansion of order 0. However, it is interesting to empirically assess how the initial choice of the parameters $\theta \in \Theta$ affects the accuracy of (5). To answer this question, we perturb the parameters of optimally calibrated kernels, so that the moments and the option prices implied by the kernels heavily mismatch those generated by the true $f_{\mathbb{Q}}$. In Table 4 we report the first four moments of the GIG and the GW kernels, where mean and variance are drastically perturbed as compared to the values implied by the true density of the Heston model. The last two columns of the table highlight the capability of the polynomial expansions to yield precise fitting of the moments of the RND, even when the kernel largely deviates from the true density. It is inherently assumed, however, that the assumptions of Theorem 3.1 are

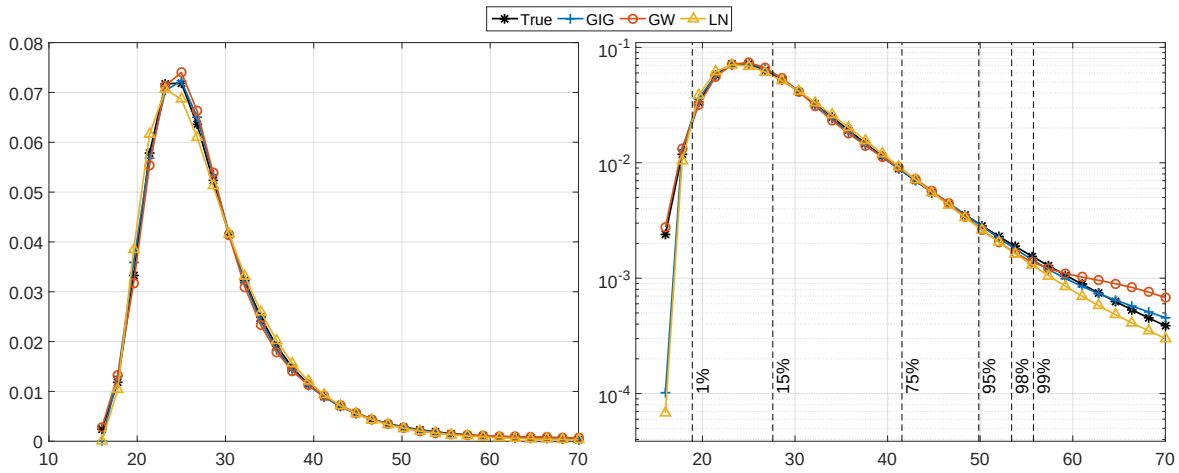


Figure 11: Probability density functions in standard scale (left) and semi-logarithmic scale (right). Comparison between the true density of VIX implied by the NIG density and the estimated RNDs of order 20. The parameters for the NIG density are chosen as follows: $\alpha = 14.36$, $\beta = 9.8$, $\mu = 2.97$, $\gamma = 0.38$. The dashed vertical lines on the right panel identify relevant probability levels and the corresponding quantiles.

	True	GIG kernel	GW kernel	GIG (order 20)	GW (order 20)
Mean	30.13	27.65	35.44	30.14	30.17
Variance	65.36	50.79	165.78	65.27	65.81
Skewness	50.26	21.07	56.86	50.16	50.40
Kurtosis	2.86	0.80	6.07	2.82	2.87

Table 4: 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 20.

always satisfied. Figure 12 portrays the true density implied by the Heston model, the perturbed kernels, and the RNDs obtained by estimating the coefficients of the corresponding orthogonal expansions.

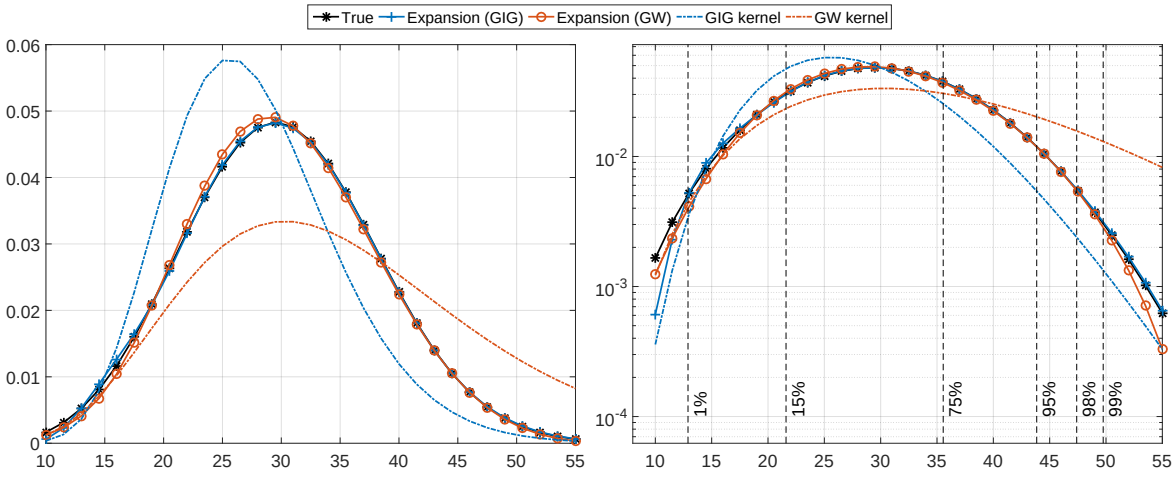
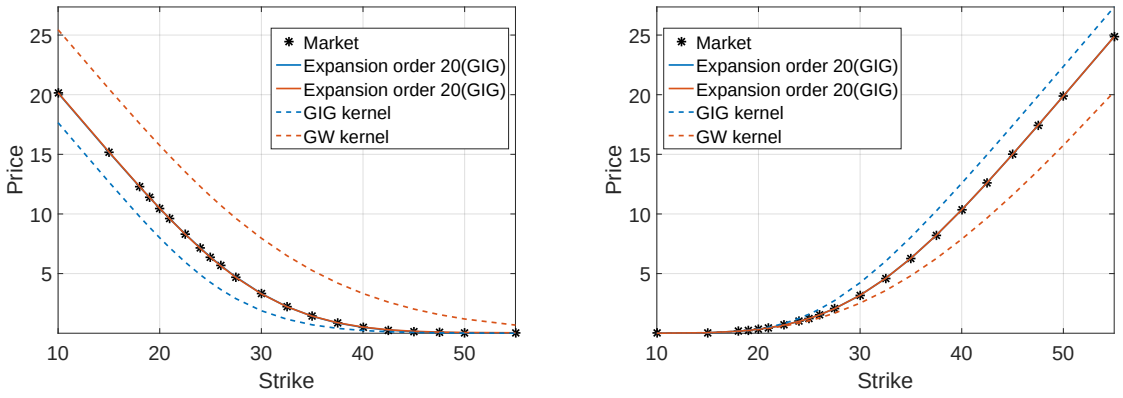


Figure 12: Probability density functions in standard scale (left) and semi-logarithmic scale (right). Comparison between the true density implied by the Heston model, the "mismatching" kernels, and the related estimated expansions. The dashed vertical lines locate some relevant mass levels and the corresponding quantiles.

The non-calibrated kernels clearly mismatch the true RND and totally deviate from each other, but almost perfect approximations of the RND are attained in both cases through expansions of order 20. Thus, the estimation based on the orthogonal expansions proves to be very robust with respect to the initialization of ϕ.



(a) Call options (b) Put options

Figure 13: Call and put option prices implied by the Heston model, the mismatching GIG and GW kernels, and related expansions of order 20.

Figure 13, depicting the option prices generated by the densities reported in Figure 12, confirms that the accuracy of the estimated RND is affected by the choice of the kernel only to a minor extent. Indeed, the kernel has almost no impact on the estimation, provided that the conditions of convergence have been guaranteed and that the expansion order can be set sufficiently large - which is our case.

B.6 Robustness to no-arbitrage violations

We assess here the practical validity of Proposition 5.1 by means of Monte Carlo simulations. To this end, we evaluate how adding a random noise to a discrete set of arbitrage-free prices obtained from a known RND affects the estimates of RND obtained by solving the problem in (19). Consistently with notations of Section 4, we assume the following form for the vector $\mathbf{Y}$ of observed prices

$$\mathbf{Y} = \mathcal{Y} + \epsilon,$$

where $\mathcal{Y}$ are arbitrage free option prices and ϵ is a vector of random shocks embedding all the violations from the no-arbitrage assumption. Specifically, we assume that the vector of observed call and put prices are given by

$$\mathbf{C} = C + \epsilon^C, \quad \mathbf{P} = \mathcal{P} + \epsilon^P$$

where $C = [C_{K_1}(t,\tau), \dots, C_{K_M}(t,\tau)]'$, $\mathcal{P} = [\mathcal{P}_{K_1}(t,\tau), \dots, \mathcal{P}_{K_M}(t,\tau)]'$, ϵ^C and ϵ^P are independent vectors of independent centered Gaussian variables with non-constant variance

$$\sigma^2_{C,i} = \mathrm{Var}[\epsilon^C_i], \quad \sigma^2_{P,i} = \mathrm{Var}[\epsilon^P_i], \quad i = 1, \dots, M.$$

Choosing a non-constant variance is owed to the fact that the magnitude of no-arbitrage violations must be consistent with the magnitude of option prices, which are monotonic quantities. Therefore, $\sigma^2_C = [\sigma^2_{C,1}, \dots, \sigma^2_{C,M}]$ and $\sigma^2_P = [\sigma^2_{P,1}, \dots, \sigma^2_{P,M}]$ are assumed to be an increasing and a decreasing vector, respectively. To identify σ^2_C and σ^2_P we further assume that the arbitrage error ϵ^F induced on the vector $\mathbf{F}$ of future prices implied by the put-call parity consists of i.i.d. components, that is

$$\mathbf{F} = \mathbf{C} - \mathbf{P} + K = \mathcal{F} + \epsilon^F,$$

where

$$\mathcal{F} = C_i - \mathcal{P}_i + K_i \quad \forall i = 1, \dots, M$$

is the unique arbitrage-free future price. Hence $\epsilon^F = \epsilon^C - \epsilon^P$ and $E[\epsilon^F] = 0$. Assuming $\sigma^2_F := \mathrm{Var}[\epsilon^F] < \infty$, identification of $\mathrm{Var}[\epsilon^C]$ and $\mathrm{Var}[\epsilon^P]$ can therefore be achieved by

$$\sigma^2_{C,i} + \sigma^2_{P,i} = \sigma^2_F, \quad \frac{\sigma^2_{C,i}}{\sigma^2_{P,i}} = \frac{\mathbf{C}_i}{\mathbf{P}_i}, \quad i = 1, \dots, M,$$

which gives

$$\sigma^2_{P,i} = \frac{\sigma^2_F}{1 + \frac{\mathbf{C}^2_i}{\mathbf{P}^2_i}}, \quad \sigma^2_{C,i} = \sigma^2_F - \sigma^2_{P,i}, \quad i = 1, \dots, M. \tag{48}$$

Note that observable quantity Δ^{pcp} defined in (25) is a sample counterpart of σ_F^2. Therefore, since $\sigma_F^2 = \sigma_{C,i}^2 + \sigma_{P,i}^2$ by construction, up to switching the integration order in (24), Δ^{pcp} can consistently approximate the right hand of (24). We therefore carry out Monte Carlo simulations with the purpose of investigating the robustness of the orthogonal polynomial expansion to the no-arbitrage violations and the usefulness of the threshold Δ^{pcp} to provide an indication for the lower bound on the variance of residuals. Each Monte Carlo simulation consists of a set of perturbed option prices $\mathbf{Y}$ over a fixed number $M = 25$ of strikes. The vector of arbitrage-free call and put prices, $\mathcal{Y}$, is generated only once, by direct integration of the VIX-RND implied by the Heston model, with parameters $k = 1.71$, $\bar{v} = 0.097$, $\eta = 0.577$, $v(0) = \bar{v}$ and $\tau = 30/365$. The arbitrage components ϵ^C and ϵ^P for each Monte Carlo simulation are obtained as

$$\begin{bmatrix} \epsilon^C \\ \epsilon^P \end{bmatrix} = \begin{bmatrix} \sigma_C \\ \sigma_P \end{bmatrix} \circ R,$$

where $\circ$ denotes the Hadamard product, σ_C, σ_P are determined as in (48), and R is a $2M \times 1$ vector of i.i.d. standard Gaussian realizations, symmetrically truncated to ensure $\mathbf{Y} \geq 0$. We repeat the procedure based on either the GIG or the GW kernel, and for $\sigma_F = 0.01, 0.03, 0.05$.[4] The results of these Monte Carlo simulations are summarized in Table (5). The so called *divergence rate*, which is associated to the cases in which the RSME exceeds the threshold $2\sqrt{\Delta^{\text{pcp}}}$, is intended to approximate the frequency of violations of the conditions of Proposition 5.1. On the other hand, the second column of Table 5 endorses the validity of (24), since the RSME is below $\sqrt{\Delta^{\text{pcp}}}$ in a large percentage of cases. Furthermore, by looking at the Monte Carlo average of the RMSE, it emerges that the variance of the error associated to the expansion of order 10 decreases with σ_F and it is of the same order of σ_F in most cases. Differently, the two kernels on average are not associated to a residual variance that is comparable to Δ^{pcp} and the RMSE remains very high also when $\sigma_F = 0.01$. The third column of Table 5 reports the *filtering rate* as a measure of how often, among the convergent cases, the noise produced on data does not affect the estimated RND. As expected, the filtering rate increases as the level of noise, namely σ_F, decreases. This is consistent with Proposition 5.1 since the hypotheses 5.1.i)-ii) are expected to be less restrictive as σ_F decreases. These additional hypotheses require that the estimated RND is constrained to be positive - which is the case here - and that the observed prices do not embed multiple arbitrage-free curves. Intuitively, under these hypotheses, the estimated RND is not affected by the arbitrage noise existing in the observed prices. The figures reported in Table 5 provide a solid confirmation of this intuition, since the percentage of cases where the noise does not affect the estimated RND grows as σ_F^2 decreases, which in turn implies reducing the uncertainty on the RND. Consistently, the L^2 distance between the estimated and the true RND decreases as

[4] Typical values of $\sqrt{\Delta^{\text{pcp}}}$ determined on real data fall in the interval $[0.01, 0.05]$, which roughly correspond to an uncertainty between 1 and 5 cents of dollar on the futures prices implied by the put-call parity.

σ_F^2 decreases, as shown in the fifth column. Finally, the last column of Table 5 confirms that the estimation based on expansions of order 10 outperforms the related kernel in all cases.

					Order 10		Kernel	
		Div. rate	Fitt. rate	Filt. rate	RMSE	L2	RMSE	L2
$\sigma_F = 0.01$	GIG	14.4 %	71.4 %	100 %	0.0123	0.0069	0.0969	0.0218
	GW	15.6 %	71.3 %	100 %	0.0125	0.0070	0.0962	0.0217
$\sigma_F = 0.03$	GIG	5.6 %	74 %	82.57 %	0.0362	0.0135	0.0988	0.0217
	GW	4.7 %	76.3 %	80.47 %	0.0356	0.0131	0.0980	0.0217
$\sigma_F = 0.05$	GIG	6.8 %	81.5 %	56.81 %	0.0713	0.0186	0.1024	0.0218
	GW	4.7 %	83.2 %	57.81 %	0.0536	0.0174	0.1017	0.0217

Table 5: The table summarizes the results of $N = 1000$ Monte Carlo tests described above, corresponding to different kernels and different values of σ_F^2. The first column reports the divergence rate of the estimation, determined as the percentage of tests such that the residual root-mean squared error (RMSE) is greater than $2\sqrt{\Delta^{\text{pcp}}}$. The second column reports the rate of optimal fitting according to Equation (26), that is the percentage of tests yielding a RMSE lower or equal to σ_F. The third column reports the percentage of tests for which the arbitrage component is successfully "filtered". The arbitrage component is considered to be filtered when the RND estimated on the perturbed data and the RND estimated on arbitrage-free data achieve the same level of accuracy, in terms of magnitude ($\sim 10^{-3}$) of their distance from the true RND, measured as L^2 norm (L2). Only convergent tests are considered in this computation. The last four columns report the Monte Carlo average of RMSE and L2 relative to the expansion of order 10 and kernel (order 0).

B.7 Details on the Empirical Analysis

	GIG		GW
Par.	Estimate	Par.	Estimate
α	-0.899	α	2.467
β	0.090	β	0.874
ξ	33.99	p	0.605

Table 6: The table reports the kernel parameters for the GIG and GW kernels respectively, estimated using the procedure detailed in Section B.1.2.

	τ = 1 month			τ = 2 months			τ = 3 months			τ = 4 months			τ = 5 months		
	M	Min	Max	M	Min	Max	M	Min	Max	M	Min	Max	M	Min	Max
20 Jan 2010	21	17	60	22	15	75	20	18	75	19	15	70	20	15	75
24 Feb 2010	19	17	50	23	17	75	18	18	70	21	10	75	20	15	80
24 Mar 2010	18	15	42.5	21	15	55	21	15	80	20	15	80	20	15	80
21 Apr 2010	20	15	47.5	21	15	60	22	15	70	24	15	80	23	15	75
26 May 2010	25	19	100	28	10	90	28	15	100	26	15	90	31	15	100
23 Jun 2010	21	20	80	23	18	90	28	15	100	30	16	100	20	15	80
21 Jul 2010	21	17	75	23	17	85	31	15	100	19	20	80	21	10	80
25 Aug 2010	19	22.5	75	27	18	95	22	10	80	21	10	80	21	10	80
22 Sep 2010	22	17	65	27	15	80	22	20	80	20	20	75	22	15	80
20 Oct 2010	23	15	55	23	18	75	22	18	80	24	10	80	21	15	80
24 Nov 2010	21	15	47.5	24	15	70	23	15	75	22	15	80	22	15	80
22 Dec 2010	21	15	50	26	15	75	23	15	75	21	15	75	20	15	75
26 Jan 2011	18	14	40	23	13	55	27	13	75	27	13	75	26	13	70
23 Feb 2011	22	15	60	25	15	75	25	14	70	26	14	75	28	13	80
23 Mar 2011	20	15	47.5	25	15	70	25	15	75	27	14	80	26	13	70
20 Apr 2011	22	13	47.5	27	14	75	27	14	75	27	14	80	25	15	75
25 May 2011	19	14	42.5	25	13	60	25	13	65	25	15	75	24	15	70
22 Jun 2011	21	15	50	24	14	65	24	15	70	24	15	70	26	15	80
20 Jul 2011	22	14	55	24	15	70	25	15	75	25	15	75	28	15	80
24 Aug 2011	22	21	90	26	19	100	25	19	95	29	16	90	31	17	95
21 Sep 2011	24	22.5	95	26	20	95	30	16	95	31	18	100	33	16	100
26 Oct 2011	23	19	75	28	17	90	32	17	100	33	16	100	34	15	100
23 Nov 2011	24	21	90	30	19	100	34	15	100	35	10	100	34	15	100
21 Dec 2011	26	17	70	29	16	80	31	15	85	33	15	95	31	16	90
25 Jan 2012	25	15	55	27	16	70	32	16	95	31	15	85	32	15	90
22 Feb 2012	26	15	60	30	16	85	29	15	75	31	15	85	33	15	95
21 Mar 2012	28	12	55	31	12	70	31	12	70	34	10	75	35	11	85
25 Apr 2012	24	13	45	30	12	65	33	12	80	32	12	75	34	10	95
23 May 2012	27	16	70	31	14	80	33	14	90	31	15	85	30	15	80
20 Jun 2012	31	14	80	36	13	100	30	15	80	29	15	75	30	15	80
25 Jul 2012	28	14	65	31	15	85	30	15	80	32	15	90	30	15	80
22 Aug 2012	31	13	75	31	13	75	31	13	75	32	13	80	31	13	75
26 Sep 2012	26	13	50	32	11	70	30	12	65	31	12	70	32	12	75
24 Oct 2012	27	13	55	31	12	70	31	12	70	31	12	70	32	13	80
21 Nov 2012	26	13	50	33	12	80	31	12	70	30	13	70	30	13	70
26 Dec 2012	25	14	50	32	12	75	29	13	65	31	12	70	32	12	75
23 Jan 2013	20	12	32.5	28	11	50	34	11	80	31	10	60	31	10	60
20 Feb 2013	22	12	37.5	30	11	60	33	10	70	35	10	80	30	11	60
20 Mar 2013	25	11	42.5	30	11	60	35	9	75	30	11	60	34	11	80
24 Apr 2013	23	11	37.5	31	11	65	33	11	75	36	9	80	31	11	65
22 May 2013	25	12	45	31	11	65	32	11	70	31	10	60	31	11	65
26 Jun 2013	22	13	40	28	13	60	33	11	75	31	12	70	30	12	65
24 Jul 2013	20	12	32.5	28	11	50	29	11	55	31	11	65	28	12	55
21 Aug 2013	23	12	40	30	11	60	29	11	55	32	10	65	30	11	60
25 Sep 2013	19	12	30	25	12	45	29	12	60	31	11	65	30	11	60
23 Oct 2013	22	12	37.5	28	12	55	29	11	55	31	11	65	27	12	50
20 Nov 2013	22	12	37.5	27	12	50	31	11	65	31	11	65	29	11	55
24 Dec 2013	21	13	37.5	27	12	50	28	11	50	31	11	65	30	11	60
22 Jan 2014	25	12	45	29	11	55	30	11	60	30	11	60	30	11	60
26 Feb 2014	21	12	35	27	12	50	31	11	65	30	12	65	31	11	65
26 Mar 2014	20	13	35	24	13	45	30	12	65	30	11	60	30	11	60
23 Apr 2014	19	12	30	24	12	42.5	29	12	60	32	11	70	29	11	55
21 May 2014	19	12	30	23	12	40	30	11	60	30	11	60	33	11	75
25 Jun 2014	23	10.5	28	24	10	37.5	29	10	50	30	11	60	30	10	55
23 Jul 2014	23	10.5	28	30	10	40	29	10	50	31	10	60	31	10	60
20 Aug 2014	26	10.5	32.5	29	10	50	31	10	60	31	10	60	33	10	70
24 Sep 2014	24	11.5	32.5	30	11	45	30	10	55	33	10	70	32	11	70
22 Oct 2014	31	12.5	60	31	11	65	30	12	65	32	11	70	31	11	65
26 Nov 2014	25	11.5	35	32	12	60	31	11	65	31	11	65	30	12	65
24 Dec 2014	28	12	45	34	11.5	65	30	11	60	30	12	65	32	11	70
21 Jan 2015	28	13.5	55	30	12	65	29	12	60	30	12	65	31	12	70
25 Feb 2015	25	12.5	40	31	12.5	60	31	12	70	31	12	70	33	11	75
25 Mar 2015	23	13	37.5	31	12	55	29	12	60	30	12	65	32	11	70
22 Apr 2015	26	12	40	31	12	55	30	12	65	32	11	70	31	12	70
20 May 2015	25	12	37.5	31	11	65	31	12	70	31	11	65	32	11	70
24 Jun 2015	27	11.5	40	32	11.5	55	31	11	65	31	12	70	31	12	70
22 Jul 2015	26	11.5	37.5	29	11	55	32	11	70	30	12	65	31	12	70
26 Aug 2015	31	14	75	35	12.5	80	31	12	70	30	12	65	32	11	70
22 Sep 2015	29	14	65	32	13	70	30	12	65	31	12	70	31	12	70
20 Oct 2015	30	12.5	55	29	12	60	31	11	65	32	11	70	31	12	70
24 Nov 2015	25	12	37.5	32	12.5	65	30	12	65	31	12	70	31	12	70
22 Dec 2015	29	13	55	33	12	65	31	11	65	33	12	80	31	12	70
19 Jan 2016	30	13.5	65	29	14	70	32	13	80	31	12	70	32	11	70
23 Feb 2016	25	14.5	50	29	14.5	70	28	14	65	30	13	70	32	13	80
22 Mar 2016	28	13	50	32	12.5	65	29	12	60	31	12	70	30	12	65
19 Apr 2016	27	12	42.5	29	12	60	30	12	65	31	12	70	31	12	70

Table 7: Summary of the panel of VIX options.

	τ = 1 month			τ = 2 months			τ = 3 months			τ = 4 months			τ = 5 months		
	$\sqrt{\Delta^{\text{pcp}}}$	Kernel	Expans.	$\sqrt{\Delta^{\text{pcp}}}$	Kernel	Expans.	$\sqrt{\Delta^{\text{pcp}}}$	Kernel	Expans.	$\sqrt{\Delta^{\text{pcp}}}$	Kernel	Expans.	$\sqrt{\Delta^{\text{pcp}}}$	Kernel	Expans.
20 Jan 2010	0.0886	0.0621	0.0538	0.0556	0.103	0.0362	0.0664	0.0799	0.0355	0.0807	0.116	0.0438	0.0654	0.13	0.037
24 Feb 2010	0.0744	0.075	0.0431	0.032	0.0943	0.0239	0.0458	0.103	0.0333	0.0532	0.106	0.0514	0.07	0.128	0.0364
24 Mar 2010	0.0554	0.0616	0.0322	0.0416	0.0782	0.0232	0.0344	0.0689	0.022	0.0518	0.103	0.0297	0.072	0.0877	0.0402
21 Apr 2010	0.0296	0.0948	0.0235	0.0326	0.0775	0.0206	0.0337	0.0935	0.0239	0.0416	0.0948	0.03	0.0717	0.106	0.0486
26 May 2010	0.0761	0.0968	0.0562	0.0798	0.253	0.0603	0.0776	0.17	0.0459	0.0869	0.151	0.0511	0.0767	0.172	0.0578
23 Jun 2010	0.0321	0.123	0.0267	0.0459	0.121	0.0393	0.0465	0.177	0.0379	0.0608	0.0992	0.0419	0.0668	0.11	0.0433
21 Jul 2010	0.0305	0.096	0.0205	0.0403	0.0812	0.0257	0.0713	0.125	0.0465	0.0685	0.0906	0.0383	0.0708	0.128	0.0376
25 Aug 2010	0.0328	0.0733	0.0226	0.0463	0.075	0.0301	0.0469	0.129	0.0272	0.085	0.143	0.0448	0.0749	0.133	0.0414
22 Sep 2010	0.0443	0.0572	0.0305	0.0369	0.063	0.0239	0.0389	0.0935	0.0287	0.0367	0.0783	0.0209	0.0493	0.145	0.031
20 Oct 2010	0.0471	0.0982	0.0327	0.0706	0.118	0.0503	0.0405	0.124	0.0248	0.0813	0.142	0.042	0.087	0.13	0.0474
24 Nov 2010	0.0383	0.0713	0.0248	0.0367	0.0738	0.0281	0.0757	0.145	0.0482	0.0555	0.172	0.0427	0.0698	0.122	0.0436
22 Dec 2010	0.0402	0.0496	0.0278	0.0558	0.123	0.0336	0.0333	0.109	0.0308	0.0585	0.105	0.0346	0.0347	0.107	0.0254
26 Jan 2011	0.0224	0.0681	0.0169	0.023	0.0706	0.0176	0.0351	0.0994	0.0227	0.0334	0.0812	0.0228	0.0383	0.0708	0.0229
23 Feb 2011	0.028	0.0691	0.023	0.0519	0.0971	0.0327	0.0491	0.0757	0.0352	0.0547	0.0828	0.037	0.0536	0.113	0.0402
23 Mar 2011	0.0329	0.0506	0.0276	0.0303	0.0935	0.0262	0.0349	0.0805	0.0246	0.0339	0.097	0.021	0.0442	0.118	0.0357
20 Apr 2011	0.0295	0.0952	0.019	0.03	0.0852	0.0227	0.039	0.114	0.0359	0.0526	0.107	0.0534	0.0406	0.117	0.0351
25 May 2011	0.0282	0.0761	0.0187	0.0267	0.102	0.0217	0.0227	0.0927	0.0202	0.0322	0.0871	0.0231	0.0458	0.115	0.0386
22 Jun 2011	0.0236	0.079	0.0209	0.027	0.077	0.0215	0.0296	0.0789	0.0257	0.0374	0.127	0.036	0.0312	0.106	0.0224
20 Jul 2011	0.023	0.0906	0.0193	0.0278	0.0873	0.024	0.0299	0.143	0.0226	0.0331	0.0917	0.0211	0.0323	0.0851	0.0382
24 Aug 2011	0.0371	0.0754	0.0352	0.0445	0.121	0.0362	0.0385	0.143	0.0336	0.0411	0.183	0.0417	0.0779	0.137	0.0577
21 Sep 2011	0.0358	0.0903	0.0274	0.0273	0.0985	0.022	0.0385	0.187	0.0307	0.0583	0.118	0.059	0.0795	0.15	0.0432
26 Oct 2011	0.0238	0.0843	0.0172	0.0284	0.0808	0.0252	0.0272	0.0714	0.0203	0.0333	0.0999	0.0324	0.0641	0.136	0.0359
23 Nov 2011	0.026	0.0692	0.0245	0.0423	0.0925	0.033	0.0398	0.115	0.0274	0.0571	0.261	0.0329	0.0738	0.275	0.0421
21 Dec 2011	0.0304	0.112	0.0235	0.0343	0.0889	0.0289	0.0442	0.153	0.0286	0.0462	0.156	0.0265	0.0437	0.103	0.0437
25 Jan 2012	0.0326	0.0831	0.0221	0.0337	0.0569	0.0284	0.0438	0.0543	0.032	0.0447	0.0853	0.0337	0.058	0.0974	0.0352
22 Feb 2012	0.028	0.0762	0.0222	0.0305	0.0684	0.0291	0.0289	0.175	0.0286	0.0553	0.161	0.0322	0.0837	0.167	0.0447
21 Mar 2012	0.0302	0.0704	0.0191	0.0334	0.0546	0.0218	0.04	0.0614	0.0319	0.0442	0.157	0.0305	0.0468	0.143	0.0302
25 Apr 2012	0.02	0.0771	0.0164	0.0423	0.107	0.0268	0.0394	0.0654	0.0232	0.0437	0.074	0.0271	0.0599	0.144	0.0322
23 May 2012	0.042	0.0611	0.025	0.0461	0.0676	0.0295	0.0475	0.061	0.0314	0.035	0.153	0.0261	0.0594	0.106	0.0362
20 Jun 2012	0.028	0.0742	0.0198	0.0426	0.0661	0.0271	0.0308	0.0687	0.0236	0.0317	0.0823	0.031	0.0418	0.0797	0.0363
25 Jul 2012	0.0276	0.066	0.0205	0.0386	0.083	0.0254	0.0453	0.0804	0.0293	0.0536	0.159	0.0305	0.0645	0.0837	0.0363
22 Aug 2012	0.029	0.07	0.0198	0.042	0.0744	0.0281	0.0303	0.0598	0.0232	0.0439	0.0523	0.0334	0.0415	0.0548	0.0266
26 Sep 2012	0.0227	0.0757	0.019	0.0279	0.118	0.0215	0.0263	0.0737	0.0197	0.0294	0.0545	0.0205	0.0394	0.0672	0.0244
24 Oct 2012	0.0318	0.075	0.0297	0.034	0.0794	0.0278	0.0331	0.0857	0.0259	0.0375	0.0662	0.0338	0.0532	0.0702	0.0342
21 Nov 2012	0.0281	0.0771	0.0191	0.0417	0.0331	0.0255	0.0302	0.0704	0.0296	0.0404	0.0721	0.0267	0.04	0.0734	0.0264
26 Dec 2012	0.0322	0.0525	0.0203	0.0291	0.0958	0.024	0.0408	0.0572	0.0292	0.0561	0.0945	0.0324	0.0593	0.0617	0.0323

Table 8: Root mean square error (RMSE): errors between the observed VIX option prices and the approximate prices implied by GW kernel and related expansion.

	τ = 1 month			τ = 2 months			τ = 3 months			τ = 4 months			τ = 5 months		
	$\sqrt{\Delta^{\text{pcp}}}$	Kernel	Expans.	$\sqrt{\Delta^{\text{pcp}}}$	Kernel	Expans.	$\sqrt{\Delta^{\text{pcp}}}$	Kernel	Expans.	$\sqrt{\Delta^{\text{pcp}}}$	Kernel	Expans.	$\sqrt{\Delta^{\text{pcp}}}$	Kernel	Expans.
23 Jan 2013	0.0311	0.0566	0.0201	0.0235	0.053	0.0192	0.0495	0.0672	0.0303	0.0485	0.106	0.0313	0.0434	0.113	0.0302
20 Feb 2013	0.0281	0.0605	0.0214	0.036	0.0576	0.0233	0.0385	0.0862	0.0284	0.0597	0.108	0.0376	0.0523	0.0642	0.0316
20 Mar 2013	0.0186	0.0624	0.0168	0.0402	0.0674	0.0309	0.0364	0.123	0.0311	0.0487	0.0676	0.0348	0.0641	0.0561	0.0398
24 Apr 2013	0.0203	0.0556	0.0162	0.0293	0.0621	0.0238	0.045	0.0602	0.0347	0.0539	0.117	0.0384	0.0523	0.0551	0.0337
22 May 2013	0.034	0.0641	0.0219	0.0295	0.0651	0.0213	0.0325	0.0483	0.0264	0.038	0.105	0.033	0.0369	0.0805	0.028
26 Jun 2013	0.0254	0.0758	0.0209	0.0235	0.0591	0.0228	0.037	0.102	0.0304	0.0366	0.0417	0.0289	0.0332	0.0438	0.0272
24 Jul 2013	0.0224	0.0554	0.0171	0.0271	0.0612	0.0199	0.0334	0.0544	0.0227	0.0442	0.0363	0.0252	0.0445	0.0473	0.0278
21 Aug 2013	0.0361	0.0629	0.0317	0.049	0.0986	0.0371	0.0249	0.0913	0.0235	0.0573	0.103	0.0408	0.0531	0.0609	0.0371
25 Sep 2013	0.0257	0.0335	0.0175	0.0246	0.0588	0.0203	0.0447	0.0646	0.0367	0.0489	0.0588	0.0373	0.045	0.045	0.0359
23 Oct 2013	0.0261	0.0721	0.0222	0.0286	0.0712	0.0214	0.0305	0.051	0.0254	0.0439	0.0461	0.0265	0.0373	0.0432	0.0317
20 Nov 2013	0.0289	0.0608	0.0205	0.0272	0.0663	0.0195	0.0347	0.0918	0.0208	0.0353	0.0405	0.0244	0.0388	0.0406	0.0273
24 Dec 2013	0.0277	0.0372	0.0272	0.0305	0.0884	0.029	0.0655	0.0897	0.0624	0.0386	0.0668	0.0299	0.037	0.0561	0.0277
22 Jan 2014	0.0256	0.0509	0.0212	0.0236	0.128	0.0233	0.0369	0.0933	0.0341	0.041	0.0546	0.0273	0.05	0.0677	0.0363
26 Feb 2014	0.0269	0.0565	0.0187	0.0189	0.0841	0.0179	0.0253	0.108	0.0224	0.0303	0.0815	0.0249	0.0306	0.0703	0.03
26 Mar 2014	0.0253	0.0545	0.016	0.027	0.04	0.0263	0.0249	0.056	0.022	0.0363	0.107	0.0367	0.0354	0.0681	0.0338
23 Apr 2014	0.0328	0.0514	0.0228	0.0242	0.0195	0.0182	0.0362	0.0652	0.0337	0.0467	0.0985	0.0421	0.0405	0.0633	0.036
21 May 2014	0.0316	0.0379	0.02	0.0255	0.0652	0.0208	0.0237	0.037	0.0193	0.0306	0.0635	0.0267	0.0421	0.0576	0.0353
25 Jun 2014	0.0215	0.0369	0.0147	0.021	0.0492	0.0194	0.0304	0.0303	0.0196	0.0241	0.0631	0.0184	0.0341	0.0614	0.0312
23 Jul 2014	0.0199	0.0503	0.017	0.0234	0.091	0.02	0.0207	0.0628	0.0174	0.0337	0.0869	0.027	0.0326	0.0858	0.0319
20 Aug 2014	0.0187	0.0343	0.0148	0.0233	0.056	0.0179	0.0209	0.0623	0.0213	0.0425	0.106	0.039	0.0415	0.0609	0.0327
24 Sep 2014	0.0239	0.0713	0.0198	0.0237	0.0823	0.018	0.0309	0.0622	0.0257	0.0392	0.131	0.0361	0.0393	0.0687	0.0379
22 Oct 2014	0.0223	0.071	0.0185	0.0259	0.102	0.0237	0.0314	0.0834	0.0307	0.0284	0.0793	0.0222	0.0262	0.0752	0.0252
26 Nov 2014	0.0332	0.0618	0.0221	0.0261	0.129	0.0187	0.0221	0.127	0.0221	0.0295	0.0838	0.0298	0.027	0.0613	0.0271
24 Dec 2014	0.0303	0.0745	0.0198	0.0337	0.129	0.0213	0.0371	0.104	0.0353	0.0347	0.0778	0.0311	0.0413	0.063	0.028
21 Jan 2015	0.0319	0.0564	0.0214	0.0264	0.122	0.0265	0.0179	0.103	0.018	0.0279	0.0522	0.0224	0.0395	0.0704	0.038
25 Feb 2015	0.0265	0.0684	0.0172	0.0208	0.066	0.0209	0.0149	0.0722	0.0141	0.0261	0.0616	0.0251	0.0368	0.139	0.0289
25 Mar 2015	0.0285	0.0629	0.0177	0.0179	0.0928	0.0171	0.0263	0.0521	0.0262	0.0296	0.0693	0.0296	0.0299	0.0673	0.0283
22 Apr 2015	0.0248	0.0859	0.0207	0.0222	0.0833	0.0204	0.0233	0.0584	0.0204	0.0232	0.11	0.0236	0.0276	0.0741	0.0242
20 May 2015	0.0208	0.0767	0.0195	0.0287	0.106	0.0208	0.0275	0.0756	0.0239	0.0224	0.102	0.0184	0.033	0.139	0.0307
24 Jun 2015	0.0281	0.089	0.0207	0.0342	0.0696	0.0308	0.0343	0.124	0.032	0.0348	0.0744	0.0274	0.0507	0.0778	0.0395
22 Jul 2015	0.0219	0.0908	0.02	0.0255	0.0421	0.0219	0.031	0.0983	0.0279	0.0411	0.0894	0.0384	0.0418	0.0635	0.0368
26 Aug 2015	0.0307	0.0442	0.0235	0.046	0.0786	0.044	0.0565	0.0702	0.0418	0.0425	0.0974	0.041	0.0593	0.0715	0.0379
22 Sep 2015	0.0291	0.0969	0.0257	0.0286	0.0768	0.0236	0.0314	0.102	0.0309	0.0481	0.115	0.0489	0.0491	0.0959	0.0407
20 Oct 2015	0.0223	0.0883	0.0183	0.023	0.0408	0.021	0.0206	0.11	0.0208	0.0364	0.116	0.0264	0.0445	0.0888	0.0479
24 Nov 2015	0.03	0.0684	0.0232	0.0335	0.044	0.0261	0.0422	0.0701	0.0418	0.0396	0.0875	0.0379	0.0495	0.0809	0.0472
22 Dec 2015	0.0291	0.073	0.0202	0.027	0.114	0.0272	0.0369	0.101	0.0249	0.0571	0.0638	0.0427	0.0632	0.0746	0.0485
19 Jan 2016	0.0312	0.0699	0.0214	0.0249	0.059	0.0219	0.0448	0.0816	0.0438	0.0453	0.105	0.0455	0.0377	0.133	0.0401
23 Feb 2016	0.025	0.0518	0.0201	0.0282	0.0698	0.0215	0.036	0.0683	0.0338	0.0338	0.0708	0.03	0.0532	0.0699	0.0525
22 Mar 2016	0.0249	0.0775	0.0171	0.0324	0.0768	0.0254	0.0312	0.0772	0.0214	0.0472	0.0661	0.0436	0.0506	0.0771	0.0477
19 Apr 2016	0.0249	0.0623	0.0191	0.0233	0.098	0.022	0.0282	0.0524	0.0256	0.0362	0.0525	0.0306	0.042	0.0609	0.0343

Table 9: Root mean square error (RMSE): errors between the observed VIX option prices and the approximate prices implied by GW kernel and related expansion.

Research Papers 2016

2016-31: Carlos Vladimir Rodríguez-Caballero: Panel Data with Cross-Sectional Dependence Characterized by a Multi-Level Factor Structure

2016-32: Lasse Bork, Stig V. Møller and Thomas Q. Pedersen: A New Index of Housing Sentiment

2016-33: Joachim Lebovits and Mark Podolskij: Estimation of the global regularity of a multifractional Brownian motion

2017-01: Nektarios Aslanidis, Charlotte Christiansen and Andrea Cipollini: Predicting Bond Betas using Macro-Finance Variables

2017-02: Giuseppe Cavaliere, Morten Ørregaard Nielsen and Robert Taylor: Quasi-Maximum Likelihood Estimation and Bootstrap Inference in Fractional Time Series Models with Heteroskedasticity of Unknown Form

2017-03: Peter Exterkate and Oskar Knapik: A regime-switching stochastic volatility model for forecasting electricity prices

2017-04: Timo Teräsvirta: Sir Clive Granger's contributions to nonlinear time series and econometrics

2017-05: Matthew T. Holt and Timo Teräsvirta: Global Hemispheric Temperatures and Co–Shifting: A Vector Shifting–Mean Autoregressive Analysis

2017-06: Tobias Basse, Robinson Kruse and Christoph Wegener: The Walking Debt Crisis

2017-07: Oskar Knapik: Modeling and forecasting electricity price jumps in the Nord Pool power market

2017-08: Malene Kallestrup-Lamb and Carsten P.T. Rosenskjold: Insight into the Female Longevity Puzzle: Using Register Data to Analyse Mortality and Cause of Death Behaviour Across Socio-economic Groups

2017-09: Thomas Quistgaard Pedersen and Erik Christian Montes Schütte: Testing for Explosive Bubbles in the Presence of Autocorrelated Innovations

2017-10: Jeroen V.K. Rombouts, Lars Stentoft and Francesco Violante: Dynamics of Variance Risk Premia, Investors' Sentiment and Return Predictability

2017-11: Søren Johansen and Morten Nyboe Tabor: Cointegration between trends and their estimators in state space models and CVAR models

2017-12: Lukasz Gatarek and Søren Johansen: The role of cointegration for optimal hedging with heteroscedastic error term

2017-13: Niels S. Grønborg, Asger Lunde, Allan Timmermann and Russ Wermers: Picking Funds with Confidence

2017-14: Martin M. Andreasen and Anders Kronborg: The Extended Perturbation Method: New Insights on the New Keynesian Model

2017-15: Andrea Barletta, Paolo Santucci de Magistris and Francesco Violante: A Non-Structural Investigation of VIX Risk Neutral Density