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

Andrea Barletta, Paolo Santucci de Magistris and Francesco Violante

CREATES Research Paper 2016-20

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

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

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

Andrea Barletta[†] Paolo Santucci de Magistris [‡] Francesco Violante [§]

June 28, 2016

Abstract

We propose a non-structural pricing method to retrieve the risk-neutral density implied by options contracts on the CBOE VIX. The method is based on orthogonal polynomial expansions around a kernel density and yields the risk-neutral density of the underlying asset without the need for modeling its dynamics. The method imposes only mild regularity conditions on shape of the density. The approach can be thought of as an alternative to Hermite expansions where the kernel has positive support. The family of Laguerre kernels is extended to include the GIG and the generalized Weibull densities, which, due to their flexible rate of decay, are better suited at modeling the density of the VIX. Based on this technique, we propose a simple and robust way to estimate the expansion coefficients by means of a principal components analysis. We show that the proposed methodology yields an accurate approximation of the risk-neutral density also when the no-arbitrage and efficient option prices are contaminated by measurement errors. A number of numerical illustrations support the adequacy and the flexibility of the proposed expansions in a large variety of cases.

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

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 and the Empirical Finance workshop in Paris. Paolo Santucci de Magistris and Francesco Violante acknowledge the research support of CREATES, funded by the Danish National Research Foundation (DNRF78).

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

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

[§]CREATES, Department of Economics and Business, 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). Following different approaches, based on infinite-dimensional specifications, Buehler (2006), Bergomi (2008), and Cont and Kokholm (2013) provide modeling frameworks aimed at pricing variance swaps jointly with the SPX.

A common feature of these contributions is that the risk-neutral density (RND) is assumed to be fully described by stochastic dynamic equations of state-variables that 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 model misspecification, see for example Cont (2006). 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. A comparative analysis of the performance of simple stochastic volatility models in pricing of VIX options tend to confirm these problems. 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. This also reflects the general disagreement in the literature on the "‘nature" and the roughness of the instantaneous volatility. For example, 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. 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. 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. As a consequence, consistent modeling frameworks conceived for capturing stylized facts of the VIX are rarely suited to be employed for estimation purposes.

In this view, non-structural methods for estimating the RND directly from VIX options represent a viable alternative to stochastic modeling. In this context "non-structural" indicates that only mild regularity conditions on the form of the RND are imposed. This entails considerable reduction of the risk associated to model misspecification. Non-structural option pricing has been recently employed by Song and Xiu (2016) with the purpose of estimating the volatility pricing kernel, which is proportional to the ratio between the physical and the risk-neutral density of the VIX. In particular, building upon Breeden and Litzenberger (1978), Song and Xiu (2016) adopt a direct method to extract the marginal risk-neutral densities of SPX and VIX from their options. The approach of Breeden and Litzenberger (1978) strongly relies on the condition that there exist traded derivatives on the same asset for a continuum of its future states, which is not the case in practice. Consequently, the original data needs to be complemented with artificial points, e.g. by interpolation techniques, see also Monteiro et al. (2008). The main shortcoming of this approach lies in the fact that estimated RNDs are highly sensitive to how observed data is complemented.

The method that we propose belongs to the class of non-structural approaches but removes the restrictive condition required in Breeden and Litzenberger (1978). In particular, we recover the RND implied by VIX options by means of a finite orthogonal expansion around a kernel density, see for instance Szegö (1939). Examples of these 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 a description 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 model 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 to the Hermite expansion in that the kernel has a 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), 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 literature on pricing of VIX options 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 to the VIX options, due to the very flexible decay rate on both tails. Indeed, due 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 same spirit of Aït-Sahalia and Lo (1998), Jondeau and Rockinger (2001), and Aït-Sahalia and Duarte (2003), we propose an econometric methodology to estimate the parameters of the polynomial expansion a minimum distance approach based on the observed option prices. This econometric methodology is robust under several points of view, e.g. multicollinearity of the regressors, absence of the intercept in the model, positivity, and unit mass of the estimated RND. Additionally, we prove that the proposed methodology yields a very accurate approximation of the risk-neutral density also when the no-arbitrage and efficient option prices are contaminated by measurement errors. Finally, we test the robustness of the proposed method on both option prices generated from known RNDs and market data. The results highlight the reliability of our methodology to recover risk-neutral densities up to negligible rounding errors and minor adjustments of the observed option prices. Interestingly, 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 paper is organized as follows. Section 2 defines the VIX risk-neutral density, while Section 3 introduces and discusses the properties of orthogonal polynomial expansions. Section 4 discusses the estimation procedure based on principal components regression of the expansion coefficients, under additional consistency constraints. In Section 5 we evaluate the methodology by studying its accuracy and robustness based on option data generated according to different risk-neutral specifications for the VIX. Section 6 addresses whether and how the estimated RND is affected by option prices contaminated by measurement errors associated to no-arbitrage violations. Finally, Section 7 presents the empirical applications with real data.

2 VIX option pricing

Introduced in 1993 by the Chicago Board Options Exchange (CBOE), the VIX (volatility index), in its current formulation, measures the expected market volatility 30 days ahead, based on near-term SPX options. Futures and options on the VIX were the first derivatives on the market volatility to be listed on a regulated security exchange, see Carr and Wu (2006) and Carr and Lee (2009) for a historical review of the VIX and a description of the improvements that have been made on this index over the years. The definition of the VIX (see CBOE, 2015) builds on the methodology of Carr and Madan (1998), Britten-Jones and Neuberger (2000), and Jiang and Tian (2005) to price variance swap via static replication based on options on the SPX. Specifically, under fairly general assumptions, the fair strike of a 30-days variance swap contract can be computed at time t as

$$\frac{2e^{r\tau}}{\tau}\left(\int_0^{F(\tau)} \frac{1}{x^2}P^{SPX}(x)dx + \int_{F(\tau)}^{\infty} \frac{1}{x^2}C^{SPX}(x)dx\right), \tag{1}$$

where r is the risk-free rate, τ is the time to maturity (30 days), $F(\tau)$ is the SPX τ-forward index level at time t and identifies the at-the-money option cutoff. Finally, $C^{SPX}(x)$ and $P^{SPX}(x)$ are SPX call and put mid-quotes, respectively, observed at time t and expiring at time $t+\tau$, here expressed as functions of the strike x.

The VIX is computed by taking the square root (expressed in percentage points) of an approximate fair price of a 30-days variance swap obtained by discretizing the infinite strip of out-of-the-money options in (1) over a finite set of available strikes as

$$\frac{2e^{r\tau}}{\tau}\left(\sum_{x_i \le X_0} \frac{\Delta x_i}{x_i^2}P^{SPX}(x_i) + \sum_{x_i \ge x_0} \frac{\Delta x_i}{x_i^2}C^{SPX}(x_i)\right) + \frac{1}{\tau}\left[\frac{F(\tau)}{x_0} - 1\right]^2, \tag{2}$$

where Δx_i is half the distance between the next and the previous strikes, while x_0 is the first strike below the forward index level $F(\tau)$. The last term in (2) represents a correction for the fact that there might not be perfectly at-the-money options. The fair price of the 30-days variance swap is then obtained by interpolating the values of (2) 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), respectively.

The VIX index typically exhibits high negative correlation to the stock market, thus offering the opportunity to hedge the volatility of a broad market portfolio separately from directional price moves, as well as to leverage volatility and take advantage of upward or downward moves in the stock market, without the need of taking a direct position on the underlying asset (the SPX). Therefore, VIX options represent a powerful risk management tool to hedge against changes in the market volatility and, as a matter of fact, many investors consider the VIX to be a leading indicator of market sentiment - often the index is referred to as the investors *fear gauge*. The popularity of the VIX is confirmed by great trading volume enjoyed by VIX options, which stand at approximately 37% of the average daily volume of SPX options, see Mencia and Sentana (2013).

Under risk neutrality and no-arbitrage, the price of VIX call and put options at time t expiring at time T with strike K are given by

$$\mathcal{C}_K(t, T, \text{VIX}_t = y) = e^{-r(T-t)} \int_0^{\infty} f_{\mathbb{Q}}(\text{VIX}_T = x|\text{VIX}_t = y)(x-K)^+ dx \tag{3}$$

$$\mathcal{P}_K(t, T, \text{VIX}_t = y) = e^{-r(T-t)} \int_0^{\infty} f_{\mathbb{Q}}(\text{VIX}_T = x|\text{VIX}_t = y)(K-x)^+ dx \tag{4}$$

where $f_{\mathbb{Q}}(\text{VIX}_T = x|\text{VIX}_t = y)$ is the RND of VIX_T, conditional to $\text{VIX}_t = y$. The RND of the VIX yields important information linking the investors expectations on the future states of market volatility with their aversion towards risk, and is the object of interest in this paper. For sake of compactness we abbreviate $f_{\mathbb{Q}}(\mathrm{VIX}_T = x|\mathrm{VIX}_t = y)$, $C_K(t, T, \mathrm{VIX}_t = y)$, and $\mathcal{P}_K(t, T, \mathrm{VIX}_t = y)$ as $f_{\mathbb{Q}}(x)$, $C_K(t, T)$, and $\mathcal{P}_K(t, T)$, respectively. In the next section, we outline a mathematical method to approximate $f_{\mathbb{Q}}(x)$ on the basis of a weighted polynomial expansion and we provide a robust method to retrieve $f_{\mathbb{Q}}(x)$ directly from the option prices.

3 An orthogonal polynomial expansion for the RND

Under fairly general conditions that will be discussed below, the function $f_{\mathbb{Q}}$ can be approximated by the following polynomial expansion, $f_{\mathbb{Q}}^{(n)}$, as

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

where ϕ, the kernel, is a probability density function and e_k are corrective factors. The kernel function ϕ represents the 0-order term in (5)

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

and therefore it can be interpreted as an initial proxy for $f_{\mathbb{Q}}$. In the following, we assume that ϕ is a probability density function with support $\mathcal{D} \subseteq \mathbb{R}$ and possessing finite polynomial moments, that is

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

Furthermore, the corrective factors $e_1, \ldots, e_n$ in (5) are defined as

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

where, for every $k = 1, \ldots, n$, h_k^{ϕ} is a polynomial function in x of degree k and c_k is a real constant. Furthermore, the polynomials $h_1^{\phi}, \ldots, h_n^{\phi}$ only depend on ϕ and therefore the expansion coefficients $c_1, \ldots, c_n$ embed all the information on $f_{\mathbb{Q}}$. In other words, the sequence $(c_k)_{k\in\mathbb{N}}$ yields a full description of the $f_{\mathbb{Q}}$, through the relation expressed in (5). Importantly, this relation is non-structural, since no specific form of $f_{\mathbb{Q}}$ needs to be postulated in order to ensure the admissibility of (5). In the rest of this section we provide elements to determine the polynomials $(h_k^{\phi})_{k\in\mathbb{N}}$ and we address mathematical conditions on $f_{\mathbb{Q}}$, ϕ, and $(c_k)_{k\in\mathbb{N}}$ ensuring the admissibility and the consistency of (5).

3.1 Determination of $(h_k^{\phi})_{k\in\mathbb{N}}$

The function $f_{\mathbb{Q}}^{(n)}$ is defined in (5) as the product between a probability density function and a polynomial. Therefore, in general, $f_{\mathbb{Q}}^{(n)}$ may not be a consistent approximation for $f_{\mathbb{Q}}$, unless the polynomials $(h_k^{\phi})_{k\in\mathbb{N}}$ are chosen so that $f_{\mathbb{Q}}^{(n)}$ is a probability density function, for every $n \in \mathbb{N}$. A convenient way to ensure unitary mass for $f_{\mathbb{Q}}^{(n)}$ is to require that the polynomials $(h_k^{\phi})_{k\in\mathbb{N}}$ form an orthogonal system with respect to the measure generated by ϕ. The elements of $(h_k^{\phi})_{k\in\mathbb{N}}$ are shortly said to be *orthogonal* *polynomials* with respect to ϕ if, for all $k \in \mathbb{N}$ and all $j \in \mathbb{N}$ such that $j \neq k$,

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

Lemma 3.1. *Let ϕ be a function with finite k-th moment, for every $k \in \mathbb{N}$. Then there exists a family $(h_k)_{k\in\mathbb{N}}$ of orthogonal polynomials with respect to ϕ. The family $(h_k)_{k\in\mathbb{N}}$ is uniquely determined, up to a sign, if the following additional condition is satisfied*

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

Proof. Appendix A.1 provides a constructive proof of this lemma. □

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 the polynomials $h_0^\phi, \ldots, h_n^\phi$. Specifically, for $i = 1, \ldots, n$ we have

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

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

3.2 Properties of $f_{\mathbb{Q}}^{(n)}$

A notable property of orthogonal polynomial expansions is that adding corrective terms to $f_{\mathbb{Q}}^{(n)}$ entails a mass reshaping not affecting the first n moments. More precisely, the following result holds.

Lemma 3.2. *Let $p \geq 0$, then*

$$\int_{-\infty}^{+\infty} x^p f_{\mathbb{Q}}^{(n)}(x)dx = \int_{-\infty}^{+\infty} x^p f_{\mathbb{Q}}^{(p)}(x)dx, \quad \forall n \geq p. \tag{8}$$

Proof. See Appendix A.2. □

As a major consequence of Lemma 3.2, the total mass underlined by the n-th order expansion always equals 1, regardless of the order n. In other words

$$\int_{-\infty}^{+\infty} f_{\mathbb{Q}}^{(n)}(x)dx = 1 \quad \forall n \in \mathbb{N}. \tag{9}$$

Furthermore, in view of Lemma 3.2, we can interpret the coefficients $c_1, \ldots, c_p$ as corrective factors of the first p moments of ϕ to match the corresponding moments of $f_{\mathbb{Q}}$. For example, by applying (8) with $p = 1$ and $p = 2$ we get, respectively,

$$\int_{-\infty}^{+\infty} x f_{\mathbb{Q}}^{(n)}(x)dx = \mu_1 + c_1\sqrt{\mu_2 - \mu_1^2}, \tag{10}$$

and

$$\int_{-\infty}^{+\infty} x^2 f_{\mathbb{Q}}^{(n)}(x)dx = \mu_2 + c_1 \frac{\mu_3 - \mu_1\mu_2}{\sqrt{\mu_2 - \mu_1^2}} + c_2 \frac{\sqrt{-\mu_1^2\mu_4 + \mu_2(2\mu_1\mu_3 + \mu_4) - \mu_2^3 - \mu_3^2}}{\sqrt{\mu_2 - \mu_1^2}}. \tag{11}$$

where μ_i, $i = 1, \ldots, 4$, are the first four moments of ϕ. In particular, the quantity under square root in (11) is positive as resulting from integrating a positive function in the process of finding the normalizing constant C_2 for h_2^ϕ (see A.1).

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. However, 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

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

So far we have discussed under which conditions $f_{\mathbb{Q}}^{(n)}$ is a density function, i.e. integrates to one and is non-negative. The following lemma and the following remark are based on general properties of Hilbert spaces, and ensure that $f_{\mathbb{Q}}^{(n)}$ is an admissible expansion for $f_{\mathbb{Q}}$. For a proof the reader may refer e.g. to Rudin (1987), Theorem 4.14.

Lemma 3.3. *Assume that $\phi^{-\frac{1}{2}} f_{\mathbb{Q}} \in L^2(\mathcal{D})$ and $supp(f_{\mathbb{Q}}) \subseteq \mathcal{D}$. Define*

$$\mathcal{H}_\phi = \left\{\psi,\ \phi^{-\frac{1}{2}}\psi \in L^2(\mathcal{D})\right\}, \quad \mathcal{H}_\phi^* = Cl\left(span\left\{\phi h_k^\phi, k \in \mathbb{N}\right\}\right) \subseteq \mathcal{H}_\phi,$$
$$\mathrm{d}_\phi(\psi_1, \psi_2) = \left(\int_{\mathcal{D}} |\psi_1(x) - \psi_2(x)|^2 \frac{1}{\phi(x)} dx\right)^{\frac{1}{2}}, \quad \forall \psi_1, \psi_2 \in \mathcal{H}_\phi.$$

Then, there exist a sequence $(c_k)_{k\in\mathbb{N}}$ and a function

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

that solves the minimum distance problem

$$f_{\mathbb{Q}}^{(\infty)} = \underset{\psi\in\mathcal{H}_\phi^*}{argmin}\ \mathrm{d}_\phi(\psi, f_{\mathbb{Q}}), \quad f_{\mathbb{Q}} \in \mathcal{H}_\phi, \quad f_{\mathbb{Q}}^{(n)} \in \mathcal{H}_\phi^* \quad \forall n \in \mathbb{N}. \tag{13}$$

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

Remark 3.4. *The set $(h_k^{\phi})_{k\in\mathbb{N}}$ is an orthonormal set in the space $\mathcal{H}_\phi$. As a consequence, for every $k \in \mathbb{N}$ we have*

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

where $w_{k,i}$ is the i-th coefficient of h_k^{ϕ}. In particular there exists a linear mapping between the coefficients $(c_k)_{k\mathbb{N}}$ and the moments of the $f_{\mathbb{Q}}$.

Lemma 3.3 provides a set of theoretical conditions ensuring that $f_{\mathbb{Q}}$ admits the expansion (5), while allowing a certain freedom in the choice of the kernel ϕ. In view of this result, the kernel ϕ mostly serves as an initializing state of the expansion (5), and its choice should not affect the form of $f_{\mathbb{Q}}^{(n)}$, provided that n is sufficiently large. The validity of these results is thoroughly examined in Section 5 via numerical illustrations. In principle every function ϕ possessing all moments, up to normalization constants, can be used as kernel density in the expansion (5). To ensure convergent expansions, Lemma 3.3 requires that $\text{supp}(f_{\mathbb{Q}}) \subseteq \text{supp}(\phi)$ and, more importantly, that $\phi^{-\frac{1}{2}} f_{\mathbb{Q}} \in L^2(\mathcal{D})$. In other words, the tails of ϕ must decay at a slower rate than the tails of $f_{\mathbb{Q}}$. However, the general convergence conditions provided by Lemma 3.3 ensures that the limit of (5) equals a function $f_{\mathbb{Q}}^*$ that in general may differ from $f_{\mathbb{Q}}$ if $\mathcal{H}_\phi^* \neq \mathcal{H}_\phi$. In this paragraph, we prove that if the tails of ϕ decay at a certain exponential rate, the minimum distance problem (13) admits a solution for which the equality $f_{\mathbb{Q}}^* = f_{\mathbb{Q}}$ holds, whenever the assumptions of Lemma 3.3 are fulfilled. Specifically, this condition on the decay of ϕ is shown to guarantee the property of *closure* for the set $(h_k^{\phi})_{k\in\mathbb{N}}$ in the space $L^2_{\phi(x)dx}(\mathcal{D})$.

In the following, given a kernel function ϕ and an associated set of orthogonal polynomials $(h_k^{\phi})_{k\in\mathbb{N}}$, the notations $\mathcal{H}_\phi$ and $\mathcal{H}_\phi^*$ refer to the Hilbert spaces already defined in Lemma 3.3, where we recall that the underlying scalar product is defined as

$$\langle \psi_1, \psi_2 \rangle := \int_{\mathcal{D}} \psi_1(x)\psi_2(x) \frac{1}{\phi(x)} dx,$$

with $\text{supp}(f_{\mathbb{Q}}) \subseteq \mathcal{D} \subseteq \text{supp}(\phi)$.

Definition 3.5 (Closed polynomial set in $\mathcal{H}_\phi$)**.** The kernel ϕ is said to generate closed polynomial sets if

$$\text{Cl}\left(\text{span}\left\{x^k, k \in \mathbb{N}\right\}\right) = L^2_{\phi(x)dx}(\mathcal{D}). \tag{15}$$

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.

If $(h_k^{\phi})_{k\in\mathbb{N}}$ is closed with respect to ϕ, then $\mathcal{H}_\phi = \mathcal{H}_\phi^*$ and from Lemma 3.3 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 $a_0 + a_1 x + a_2 x^2 + \ldots$ in $L^2_{\phi(x)dx}(\mathcal{D})$ or equivalently that $f_{\mathbb{Q}}$ can be approximated by a certain series $c_0\phi h_0^{\phi} + c_1\phi h_1^{\phi} + c_2\phi h_2^{\phi} \ldots$ in $\mathcal{H}_\phi$.

Theorem 3.6 (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.

Proof. See Appendix A.3. □

The first statement of the theorem above remarks that any kernel density whose right-tail decay is faster than $e^{\varsigma\sqrt{x}}$ ensures $f_{\mathbb{Q}}^{(\infty)} = f_{\mathbb{Q}}$, whenever the conditions of Lemma 3.3 are satisfied. Notably, the classic result of closure of Laguerre polynomials is also included in statement (i), when $\phi(x) = x^{\alpha-1}e^{-\beta x}$ for some $\alpha, \beta > 0$. On the other hand, the second statement of Theorem 3.6 intuitively suggests that the "size" of the space $\mathcal{H}_{\phi}^{*}$, generated by all possible expansions with kernel ϕ, reduces as the right tail of ϕ thickens. Therefore, while heavy-tailed kernels tend to make the condition $\phi^{-\frac{1}{2}}f_{\mathbb{Q}} \in L^2(\mathcal{D})$ of Lemma 3.3 less restrictive, ϕ is required to possess a sufficiently rapid decay rate to ensure $f_{\mathbb{Q}}^{(\infty)} = f_{\mathbb{Q}}$. This determines a trade-off, since both features are required to guarantee full consistency of the expansion in (5).

3.3 The eLaguerre and the log-Hermite expansions

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

$$\phi(x) \propto x^{\alpha-1}e^{-\left(\beta x^{p}+\xi x^{-1}\right)}\mathbf{1}_{\mathcal{D}}(x), \quad \alpha, \beta, \xi, p \in \Theta, \tag{16}$$

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 general specification (16) 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 based on the kernel defined in (16) extends the classical Laguerre expansions. For this reason, we refer to this expansions as *extended Laguerre* (eLaguerre). To point out the great flexibility of this family of kernels in capturing different tail behaviours, which could not be achieved by a simple gamma kernel, we will focus on the GIG and the GW kernels. Here, we discuss their theoretical properties in relation with the consistency results provided in Lemma 3.3 and Theorem 3.6. Additionally, we consider the following log-normal (LN) kernel in our comparison

$$\phi(x) \propto \frac{1}{x}e^{-\frac{1}{2\sigma^2}(\log(x)-\mu)^2}, \quad \mu \in \mathbb{R}, \ \sigma > 0. \tag{17}$$

An expansion on the underlying risk-neutral density based on the LN kernel is conceptually similar to an Hermite expansion on the logarithm of the underlying, which makes the LN kernel an interesting competitor of the GIG and the GW kernels, due to the great interest that Hermite expansions have received by many authors in the literature concerning option pricing. The appeal of the LN kernel could be further motivated by documented empirical evidence that the volatility, which is 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).

The behavior on both left and right tails of the GIG, the GW and the LN kernels is therefore informative on their ability to meet the condition $\phi^{-\frac{1}{2}} f_{\mathbb{Q}} \in L^2(\mathcal{D})$, which is necessary for the convergence of $f_{\mathbb{Q}}^{(n)}$ to $f_{\mathbb{Q}}^{(\infty)}$, in view of Lemma 3.3. In this regard, it is worth to remark the following limit relations. Compared to the GIG and the GW kernels, Table 1 highlights that the LN kernel entails a less restrictive assumption on right tail behaviour of $f_{\mathbb{Q}}$ to satisfy $\phi^{-\frac{1}{2}} f_{\mathbb{Q}} \in L^2(\mathcal{D})$. On the other hand, the GW kernel, whose left decay rate is polynomial, imposes the smallest restriction on the left tail of $f_{\mathbb{Q}}$.

	GIG	GW	LN
Density	$x^{\alpha-1} e^{-\frac{1}{2}(\beta x + \xi x^{-1})}$	$x^{\alpha-1} e^{-\beta x^p}$	$\frac{1}{x} e^{-\frac{1}{2\sigma^2}(\log(x)-\mu)^2}$
Parameters	$\alpha \in \mathbb{R},\ \beta > 0,\ \xi \geq 0$	$\alpha > 0,\ \beta > 0,\ 0 < p \leq 1$	$\mu \in \mathbb{R},\ \sigma > 0$
Right tail of $\phi(x)^{-1}$	$O\left(e^{\varsigma x}\right), \forall \varsigma > \beta$	$O\left(e^{\varsigma x^p}\right), \forall \varsigma > \beta$	$O\left(e^{\varsigma \log(x)^2}\right),\ \forall \varsigma \geq \frac{1}{2\sigma^2}$
Left tail of $\phi(x)^{-1}$	$O\left(e^{\zeta x^{-1}}\right), \forall \zeta > \xi$	$O\left(x^{\zeta}\right), \forall \zeta \leq 1 - \alpha$	$O\left(e^{\zeta \log(x)^2}\right), \forall \zeta < \frac{1}{2\sigma^2}$
Closure of (h_k^{ϕ})	always	$\frac{1}{2} \leq p \leq 1$	never

Table 1: The table reports, for each kernel, the behavior on the left and right tails of ϕ^{-1} and the parametric conditions to guarantee that $f_{\mathbb{Q}}^{(\infty)} = f_{\mathbb{Q}}$.

Moreover, Theorem 3.6 imposes a restriction on the tail behavior of the kernel to guarantee the closure of related orthogonal polynomials. Notably, this property is always satisfied by the GIG kernel. Differently, for the orthogonal polynomials related to a GW kernel, closure is achieved only when the parameter p falls in the interval $[\frac{1}{2}, 1]$. Finally, the log-normal kernel never meets the closure condition for the related orthogonal polynomials, independently of the parameters choice. Thus, the log-normal kernel does not guarantee that the expansion $f_{\mathbb{Q}}^{(\infty)}$ converges to $f_{\mathbb{Q}}$ in all cases. This means that the LN kernel is less capable than GIG and GW to recover $f_{\mathbb{Q}}$, as intuitively the set of polynomial expansions in (5) does not cover a sufficiently large space to incorporate $f_{\mathbb{Q}}$. Therefore, using the LN density implicitly translates in assigning some restrictions to the expansion, drastically reducing the flexibility of the non-structural approach. In Section 5 we illustrate the pitfalls of choosing a LN kernel to estimate the RND of the VIX.

4 Retrieving $f_{\mathbb{Q}}$ from option prices

In this section, we outline a procedure to estimate the coefficients $(c_k)_{k\in\mathbb{N}}$ of the expansion (5) by minimizing the distance with the market option prices. The key feature of the proposed procedure is to estimate the option pricing formulas in (3) and (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 the procedure fully relies on the following theorem, which combines the results of Lemma 3.3 and Theorem 3.6.

Theorem 4.1. *Assume that ϕ is a probability density function with support on $[0, +\infty[$ and satisfying $\lim_{x\to+\infty} \phi(x)e^{\varsigma x^{\frac{1}{2}}} = 0$ for some $\varsigma > 0$. Moreover assume that $\phi^{-\frac{1}{2}} f \in L^2(\mathcal{D})$. Then:*

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

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

(b) the following holds in the limit

$$\lim_{n\to+\infty} \int_0^{+\infty} \Pi(x) f_{\mathbb{Q}}^{(n)}(x)dx = \int_0^{+\infty} \Pi(x) f_{\mathbb{Q}}(x)dx, \tag{18}$$

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

Proof. The proof of this theorem is obtained by a direct application of classic results on Hilbert's spaces. For completeness, a short proof is reported in Appendix A.4. □

Theorem 4.1 represents a prerequisite to define a mathematically well-posed procedure to estimate the coefficients $(c_k)_{k\in\mathbb{N}}$ based on observed option prices. In particular, point (a) shows that $f_{\mathbb{Q}}$ admits the representation (5) for $n \to \infty$, or, in other words, that $f_{\mathbb{Q}}^{(n)}$ converges to $f_{\mathbb{Q}}$ for some $(c_k)_{k\in\mathbb{N}}$. Note that if $f_{\mathbb{Q}}$ is sufficiently regular then k_n can be replaced by n for all $n \in \mathbb{N}$. Point (b) ensures that we can move the limit within the integral pricing formulas (3)-(4) and thus obtain convergent expansions also for the option prices related to $f_{\mathbb{Q}}$.

Remark 4.2 (Pricing formulas for vanilla options)**.** *Let ϕ be a fixed probability density function whose support is contained in $[0, +\infty[$. Let us define, for every $K \geq 0$, $n \in \mathbb{N}$ and $c_1, \dots, c_n \in \mathbb{R}$*

$$C_K^{(n)}(c_1, \dots, c_n) := \int_K^{+\infty} \phi(x)\left(1 + \sum_{k=1}^{n} c_k h_k^{(\phi)}\right)(x - K)\, dx$$

$$P_K^{(n)}(c_1, \dots, c_n) := \int_0^{K} \phi(x)\left(1 + \sum_{k=1}^{n} c_k h_k^{(\phi)}\right)(K - x)\, dx$$

Then, for every $n \in \mathbb{N}$ and every real vector $c = [c_1, \dots, c_n]'$, the expressions for $C_K^{(n)}$ and $P_K^{(n)}$ can be rewritten in the compact form

$$C_K^{(n)}(c_1, \dots, c_n) = A_0^{(K)} + A^{(K)}c, \tag{19}$$

$$P_K^{(n)}(c_1, \dots, c_n) = B_0^{(K)} + B^{(K)}c, \tag{20}$$

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

$$A_i^{(K)} = \sum_{j=0}^{i} w_{i,j} \int_K^{+\infty} \left(x^{j+1} - Kx^j\right) \phi(x)dx, \quad B_i^{(K)} = \sum_{j=0}^{i} w_{i,j} \int_0^{K} \left(Kx^j - x^{j+1}\right) \phi(x)dx. \tag{21}$$

Notably, the vectors $A^{(K)}$ and $B^{(K)}$ are only functions of the strike price, of the kernel and the related polynomials through the coefficients $w_{k,i}$. A method to retrieve values of $(c_k)_{k\in\mathbb{N}}$ from observed option prices given an expansion of order n is discussed in the next paragraph.

4.1 Estimation of $(c_k)_{k\in\mathbb{N}}$

Given a fixed kernel ϕ and a finite order expansion n, to obtain an estimate of the coefficients $c_1, \ldots, c_n$ one can collect a cross-section of undiscounted market prices, $C_{K_m}^{Obs}(t,T)$ and $P_{K_m}^{Obs}(t,T)$, for $m = 1, \ldots, M$, then find the parameters $\hat{c}_1, \ldots, \hat{c}_n$ that solve the problem

$$[\hat{c}_1, \ldots, \hat{c}_n] = \underset{c_1,\ldots,c_n \in \mathbb{R}^n}{\operatorname{argmin}} \; Q(t,T; c_1, \ldots, c_n), \tag{22}$$

where $Q(t,T;c_1,\ldots,c_n)$ defines a particular objective function to be minimized. Given that (19)-(20) are linear in the coefficients c, the objective function $Q(t,T;c_1,\ldots,c_n)$ can be naturally characterized as criterion function of a problem of minimum least squares for the following linear model

$$\mathbf{Y} = \mathbf{X}_0 c_0 + \mathbf{X}c + \varepsilon, \tag{23}$$

where $c = [c_1, \ldots, c_n]'$,

$$\mathbf{Y} = [C_{K_1}^{Obs}(t,T), \ldots, C_{K_M}^{Obs}(t,T), P_{K_1}^{Obs}(t,T), \ldots, P_{K_M}^{Obs}(t,T)]',$$

$$\mathbf{X}_0 = [A_0^{K_1}, \ldots, A_0^{K_M}, B_0^{K_1}, \ldots, B_0^{K_M}]',$$

and

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

is a $2M \times n$ matrix. The $2M \times 1$ vector ε represents the error term, whose properties are discussed more in detail in Section 6. For what concerns the regressors, since the constraint $c_0 = 1$ is necessary to ensure unitary mass by construction, the $2M \times 1$ vector of dependent variables could be defined as $\mathbf{Y}^* = \mathbf{Y} - \mathbf{X}_0$, where $\mathbf{X}_0$ is the option price vector generated by the kernel. Hence, the objective function could take the following quadratic form

$$Q(t, T; c_1, \dots, c_n) = (\mathrm{Y}^* - \mathbf{X}c)'(\mathrm{Y}^* - \mathbf{X}c), \tag{24}$$

which, in turns, allows for a closed-form solution for the vector of coefficients given by

$$\hat{c} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathrm{Y} \tag{25}$$

if $\mathbf{X}$ has full column rank. Unfortunately, the columns of $\mathbf{X}$ are functions of the first non-standardized n moments. Therefore, $\mathbf{X}$ tends to display increasing degree of multicollinearity as n grows. As a consequence, employing the standard OLS estimator to solve (24) is not suitable when n is large. This problem is well known in the literature concerned with orthogonal polynomial expansions. For example, Jarrow and Rudd (1982) and Corrado and Su (1996a,b) consider expansions only up to the fourth order and 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 again implies $n = 4$. In this regard, it is important to stress that the RND of VIX is expected to be characterized by a fat right tail, especially during turmoil periods, meaning that the information carried by higher moments may provide significant correction to the expansion kernel.

4.1.1 Orthogonal regressors

We propose to solve the problem of multicollinearity outlined above 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

$$\mathbf{Z}_i = \frac{\mathbf{X}_i - \frac{1}{2M}\sum_{j=1}^{2M}\mathbf{X}_{ji}}{\sqrt{\frac{1}{2M-1}\sum_{j=1}^{2M}\left(\mathbf{X}_{ji} - \frac{1}{2M}\sum_{j=1}^{2M}\mathbf{X}_{ji}\right)^2}}, \quad i = 1, \dots, n \tag{26}$$

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,

$$\mathrm{Y}^* = \tilde{\mathbf{V}}\gamma + u, \tag{27}$$

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

$$\tilde{c} = (\mathbf{O}\hat{\gamma}) \circ \left[\sqrt{\frac{2M-1}{\sum_{j=1}^{2M}(\mathbf{X}_{j1} - \frac{1}{2M}\sum_{j=1}^{2M}\mathbf{X}_{j1})^2}}, \dots, \sqrt{\frac{2M-1}{\sum_{j=1}^{2M}(\mathbf{X}_{jn} - \frac{1}{2M}\sum_{j=1}^{2M}\mathbf{X}_{jn})^2}} \right]', \tag{28}$$

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

Given the vector $\tilde{c}$, the estimated RND function $\tilde{f}_{\mathbb{Q}}^{(n)}$ is determined as

$$\tilde{f}_{\mathbb{Q}}^{(n)}(x) = \phi(x;\theta)\left(1 + \sum_{k=1}^{n} \tilde{c}_k h_k^{\phi}(x)\right). \tag{29}$$

Note that the kernel ϕ in (29) is now expressed as function of an additional term, to highlight its dependence on a the 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. It is important to highlight that for the polynomial expansion defined in (5) the results in Lemma 3.3 and Theorems 3.6-4.1 are obtained for any ϕ satisfying certain regularity conditions on the decay rate on their tails. Therefore, any $\theta \in \Theta$ such that the kernel satisfies these conditions can be chosen to initialize the expansion and Theorem 4.1 guarantees a convergent approximation $f_{\mathbb{Q}}^{(\infty)}$ to $f_{\mathbb{Q}}$. In other words, the initialization of the kernel has an impact on the approximation of $f_{\mathbb{Q}}$ that quickly decreases as n increases. This is clearly illustrated in the numerical examples in Section 5.3. However, in order to minimize the impact of the initialization also for finite n (that is the relevant case in practice), we determine θ as the set of parameters 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, as discussed in the next paragraph.

4.1.2 Regression through the origin

In regression (27), 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 (27) 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 (27) onto the solution of (23), a constant term equal to $\sum_{i=1}^{n} d_i \mathbf{R}_i$ appears, where $\mathbf{R} = \mathbf{O}\hat{\gamma}$ and $d_i = \frac{\sqrt{\frac{1}{2M-1}}\sum_{j=1}^{2M}\mathbf{X}_{ji}}{\sqrt{\sum_{j=1}^{2M}(\mathbf{X}_{ji} - \frac{1}{2M}\sum_{j=1}^{2M}\mathbf{X}_{ji})^2}}$, $i = 1, \dots, n$. Therefore, in order to guarantee that the relation in equation (23) holds for any $n \geq 1$, the following constrained optimization is performed

$$
\begin{aligned}
[\gamma_1, \ldots, \gamma_s] \quad &= \quad \underset{\gamma_1,\ldots,\gamma_s}{\operatorname{argmin}}\ \tilde{Q}(t, T; \gamma_1, \ldots, \gamma_s), \\
&\quad \text{s.t.} \quad \sum_{i=1}^{n} d_i \mathbf{R}_i = 0
\end{aligned}
\tag{30}
$$

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.

4.1.3 Positivity and unitary mass

As pointed out in Section 3.1 a potential issue of orthogonal expansions is that they generate approximate RNDs that might not be positive functions, even if the convergence of $f_{\mathbb{Q}}^{(n)}$ to $f_{\mathbb{Q}}$ ensures that possible negative mass becomes negligible as $n \to +\infty$. However, 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 (30). In particular, the optimal parameters $\gamma_1, \ldots, \gamma_s$ are found by solving the following constrained minimum distance problem

$$
\begin{aligned}
[\gamma_1, \ldots, \gamma_s] \quad &= \quad \underset{\gamma_1,\ldots,\gamma_s}{\operatorname{argmin}}\ \tilde{Q}(t, T; \gamma_1, \ldots, \gamma_s | \theta), \\
&\quad \text{s.t.} \quad \sum_{i=1}^{n} d_i \mathbf{R}_i = 0 \\
&\quad \text{s.t.} \quad 1 - \Delta^{\text{pos}} < \int_0^{\infty} \left| \phi(x; \theta) \left(1 + \sum_{k=1}^{n} c_k(\gamma) h_k^{\phi} \right) \right| \mathrm{d}x < 1 + \Delta^{\text{pos}}
\end{aligned}
\tag{31}
$$

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 deterministic functions of the parameters $\gamma_1, \ldots, \gamma_s$, determined as in (28).

4.1.4 Kernel displacement

Consistency conditions ensured by Theorem 4.1 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 (31), associated with $c_1, \ldots, c_n = 0$, does not satisfy the constraint of unit mass in $\operatorname{supp}(f)$. Even if in general we may assume that $\operatorname{supp}(f) = \mathbb{R}^+$, when the left tail of the true RND is particularly short around a point $K_{\min} > 0$, 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

$$\int_0^{K_{min}} \phi(x, \theta)\mathrm{d}x = 0. \tag{32}$$

A simple way to guarantee (32) 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 $K^* = [K_1 - K_{min}, \dots, K_M - K_{min}]$, 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 (31) based on K^*, then the estimated RND is determined as follows

$$\tilde{f}_{\mathbb{Q}}^{(n)}(x) = \phi(x - K_{min}; \theta)\left(1 + \sum_{k=1}^{n} \tilde{c}_k h_k^{\phi}(x - K_{min})\right). \tag{33}$$

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 7.

5 Numerical illustrations under no-arbitrage

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 4.1 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 3.3

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 4.1).

5.1 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 (3) and (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 4.1 is met for any choice of the kernel among the families of GIG, GW and LN densities. Figure 1 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 1 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 3.6), and may serve as an interpretive example of the importance of the hypotheses required by Theorem 4.1.

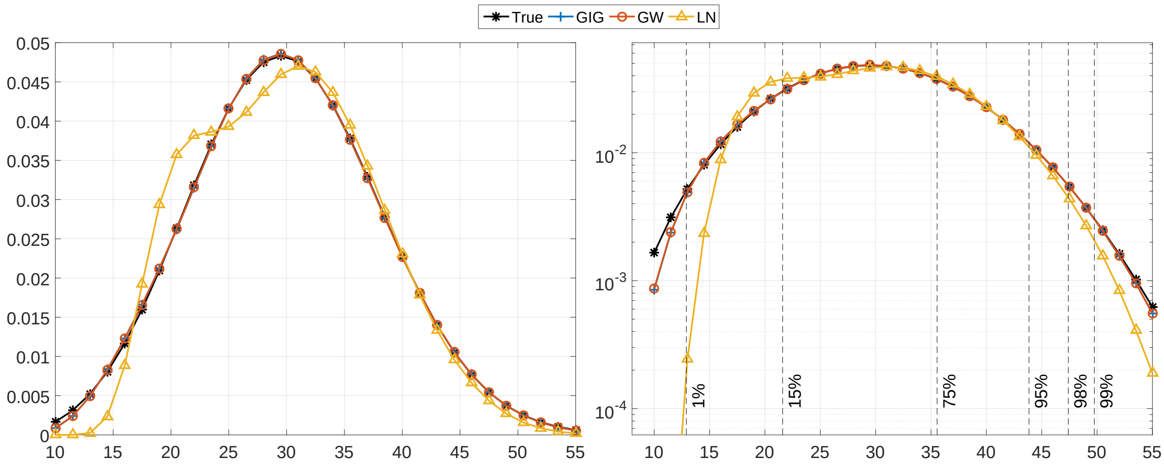


Figure 1: Probability density functions in standard scale (left) and semi-logarithmic scale (right). Comparison between the true density of VIX implied by the Heston model and the estimated RNDs of order 20. The parameters for the Heston model are: $k = 1.71$, $\bar{v} = 0.097$, $\eta = 0.577$, $v(0) = \bar{v}$ and $T - t = 30/365$. The dashed vertical lines on the right panel identify several relevant probability levels and the corresponding quantiles.

5.2 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 2 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.

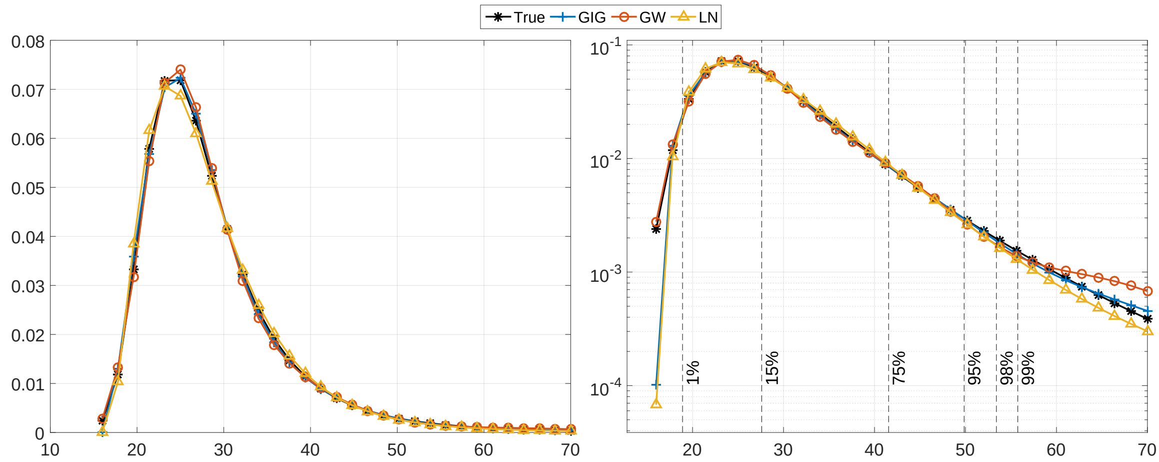


Figure 2: Probability density functions in standard scale (left) and semi-logarithmic scale (right). Comparison between the true density of VIX implied by the NIG density and the estimated RNDs of order 20. The parameters for the NIG density are chosen as follows: $\alpha = 14.36$, $\beta = 9.8$, $\mu = 2.97$, $\gamma = 0.38$. The dashed vertical lines on the right panel identify relevant probability levels and the corresponding quantiles.

5.3 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 2 we report the first four moments of the GIG and

	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 2: 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.

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 4.1 are always satisfied. Figure 3 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.

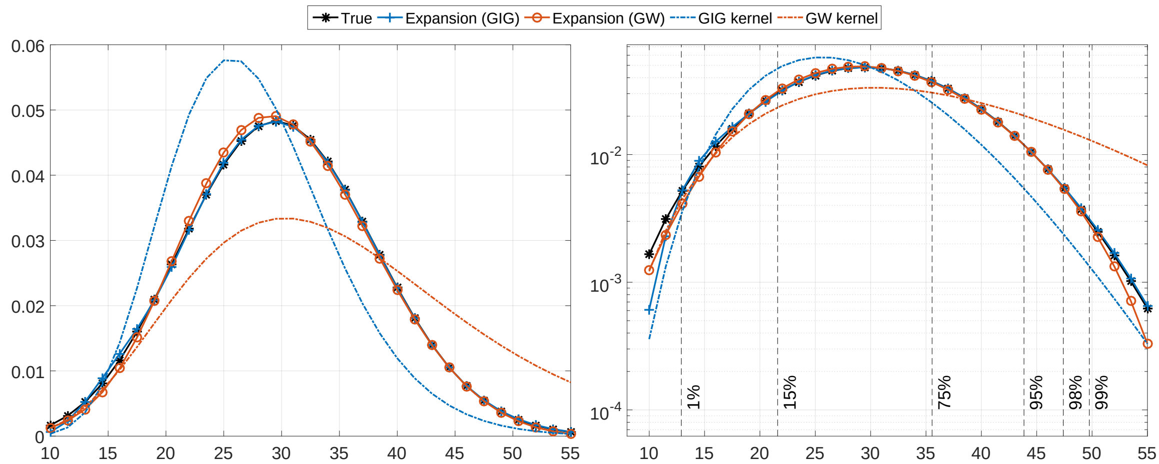
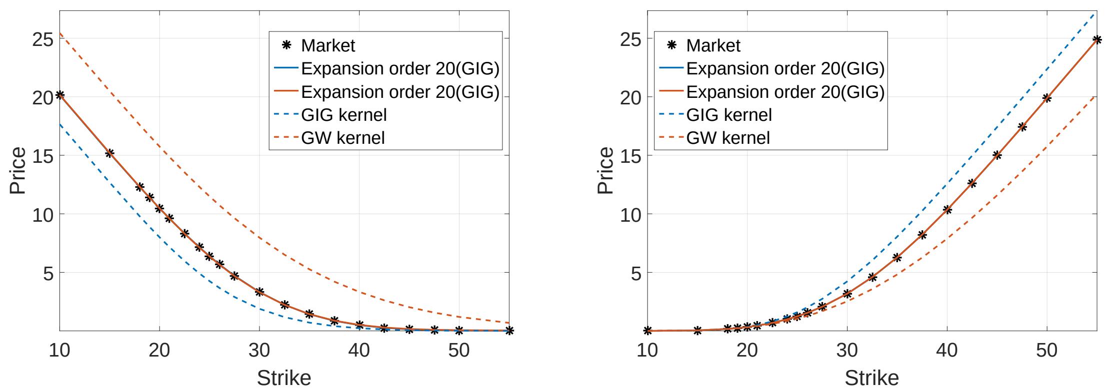


Figure 3: Probability density functions in standard scale (left) and semi-logarithmic scale (right). Comparison between the true density implied by the Heston model, the "mismatching" kernels, and the related estimated expansions. The dashed vertical lines locate some relevant mass levels and the corresponding 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 4: Call and put option prices implied by the Heston model, the mismatching GIG and GW kernels, and related expansions of order 20.

Figure 4, depicting the option prices generated by the densities reported in Figure 3, 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.

6 No-arbitrage violations

In defining the linear model (23) on which we build the estimation procedure based on orthogonal polynomials, we can assume that the error term ε subsumes all the uncertainty associated to 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 (23) can be defined 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 related to the no-arbitrage violations. Indeed, the observed option prices are subject to a number of frictions that are typically function of the market liquidity. While displaying non-negligible trading volume from most strikes, the VIX option market is still not as liquid as the equity market and, as a matter of fact, the noise produced by no-arbitrage violations may be significantly reflected in the estimation residuals. Although the arbitrage opportunities can hardly be exploited in reality due to the presence of transaction costs in the form of bid and ask spread, from a mathematical perspective the fact that the mid-quote is adopted to approximate the latent arbitrage-free option price can be seen a violation of the no-arbitrage assumption. In particular, the noise term ϵ approximates all the deviations from the efficient option price that are the result of the trading activity on the option markets and it is assumed to be the only component of ε that is subject to randomness. The presence of no-arbitrage violations makes it impossible to achieve absolutely perfect matching of all the observed option prices by minimizing (22), even if we rule out the error associated to the discretization/truncation ($\delta = 0$), since the option prices obtained by a RND are free of (static) arbitrages by construction.

Deferring to Section 6.2 a specific distributional form for the error term, we now assume that ϵ is a vector of independent random variables, with zero mean, and such that the vector defined by $\mathcal{Y} = \mathbf{Y} - \epsilon$ is arbitrage-free. In the following, an *infill* asymptotic analysis is adopted 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 interval $I = [K_1, K_M]$, we define the "infill version" of (22) as [1]

$$[\hat{c}_1, \dots, \hat{c}_n] = \underset{c_1,\dots,c_n \in \mathbb{R}^n}{\text{argmin}} \; \mathbf{Q}^{(n)}(t, T; c_1, \dots, c_n), \tag{34}$$

where

$$\mathbf{Q}^{(n)}(t, T; c) = \frac{1}{K_M - K_1} \int_I \left(C_K^{Obs}(t, T) - C_K^{(n)}(c) \right)^2 + \left(P_K^{Obs}(t, T) - P_K^{(n)}(c) \right)^2 dK.$$

[1]There is no unique way to define an "infill counterpart" of Q. The definition that we adopt, justified by mathematical convenience, builds on the fact that the integral in (34) is the limit of $\frac{1}{M}Q$, under continuity assumptions for the integrand. On the other hand, multiplying Q by any constant does not affect the solution $\hat{c}$ of (22) and therefore (34) can be legitimately interpreted as a continuous version of (22).

The observed prices $C^{Obs}(t,T)$, $P^{Obs}(t,T)$ appearing in (34) are assumed to take the form

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

where $C.(t,T), P.(t,T) \in C^2(I)$ are arbitrage-free price curves as in (3)-(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 $\mathrm{Q}^{(n)}(t,T;c_1,\dots,c_n)$ is well-defined and has finite expected value for every n. Then, the following result holds

Proposition 6.1 (Infill asymptotics). *Under the hypotheses of Theorem 4.1 and denoting by $\hat{c}(n)$ the solution of* (22)*, for every $n \in \mathbb{N}$, we have that*

$$\lim_{n\to+\infty} E\left[\mathrm{Q}^{(n)}(t,T;\hat{c}(n))\right] \leq \frac{1}{K_M - K_1} E\left[\int_I \left(\epsilon_K^C\right)^2 + \left(\epsilon_K^P\right)^2 dK\right]. \tag{35}$$

The inequality in (35) *becomes an equality under the following additional hypotheses:*

(i) There exists $\bar{n} \in \mathbb{N}$ such that, for all $n \geq \bar{n}$, $\hat{c}(n)$ is obtained by constraining (22) *to the space of coefficients $c_1,\dots,c_n$ such that*

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

(ii) If $\bar{C}.(t,T), \bar{\mathcal{P}}.(t,T) \in C^2(I)$ are arbitrage-free curves, then almost surely

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

Proof. See Appendix A.5. □

The inequality (35) defines an upper-bound on the expected value of the target function, $\mathrm{Q}^{(n)}(t,T;\tilde{c}(n))$. In particular, the expected value of the target function evaluated in $\hat{c}(n)$ is lower than the variance of the non-arbitrage residuals when $C.(t,T)$ and $\mathcal{P}.(t,T)$ are calculated on the basis of any probability density function that satisfies the hypotheses of Theorem 4.1. Under the additional assumptions (i)-(ii), Proposition 6.1 states that our estimation method provides the closest arbitrage-free prices 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 6.1 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 6.1 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 as compared to the approach of Breeden and Litzenberger (1978), which, instead, is extremely sensitive to data inconsistency. However, it may be argued that the theoretical information provided by Proposition 6.1 could be not very meaningful in practice, as it builds on a condition that might be unfeasible to be verified. In particular, condition (ii) imposes the possibly strong requirement that the noise affecting the observed prices does not cause indeterminacy of the target RND.

The aim of the following section is to show that the validity of Proposition 6.1 is empirically confirmed. To this purpose, we first derive a theoretical lower bound for the estimation residuals, to be inferred directly from the put-call parity violations based on the observed option prices. This bound is subsequently and consistently used as a proxy for (35) in a number of numerical tests, to establish to what extent the estimated RND is affected by the noise produced by no-arbitrage violations.

6.1 An observable lower bound for the estimation residuals

Following the discussion above and the result of Proposition 6.1, it is natural to tolerate a minimum degree of variability on estimation residuals $\tilde{\varepsilon}$, below a certain threshold that should quantify the presence of potential arbitrage opportunities. Given a fixed threshold $\Delta^{Q} > 0$, we define "admissible RND" any probability density function that implies option prices whose distance from the observed ones is below Δ^{Q}. Consistently, we say that a solution $\tilde{c}_1, \ldots, \tilde{c}_n$ of (30) is admissible if

$$\frac{1}{M}\sum_{m=1}^{M}\left(C_{K_m}^{Obs} - C_{K_m}^{(n)}(\tilde{c}_1, \ldots, \tilde{c}_n)\right)^2 + \frac{1}{M}\sum_{m=1}^{M}\left(P_{K_m}^{Obs} - P_{K_m}^{(n)}(\tilde{c}_1, \ldots, \tilde{c}_n)\right)^2 \leq \Delta^{Q}.$$

Since option data always contain some noise, in view of Proposition 6.1 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 expressed in terms of the residual Δ^{pcp} generated by the put-call parity violations.

Denote by $\mu_1^{\mathbb{Q}}$ the price of the future calculated under $\mathbb{Q}$, given $\tilde{c}_1, \ldots, \tilde{c}_n$, as

$$\mu_1^{\mathbb{Q}} = \int_0^{+\infty} x f_{\mathbb{Q}}^{(n)}(x; \hat{\theta}, \tilde{c}_1, \ldots, \tilde{c}_n)dx,$$

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

$$\Delta^{\mathrm{pcp}} = \frac{1}{M}\sum_{m=1}^{M}\left(C_{K_m}^{Obs}(t,T) - P_{K_m}^{Obs}(t,T) + K_m - \mu^{Obs}\right)^2, \tag{36}$$

where $\mu^{Obs} = \frac{1}{M}\sum_{m=1}^{M}\left(C_{K_m}^{Obs}(t,T) - P_{K_m}^{Obs}(t,T) + K_m\right)$.

We observe that

$$\begin{aligned}\Delta^{\mathrm{pcp}} &= \frac{1}{M}\sum_{m=1}^{M}\left[C_{K_m}^{Obs}(t,T) - P_{K_m}^{Obs}(t,T) - C_{K_m}^{(n)}(\tilde{c}_1, \ldots, \tilde{c}_n) + P_{K_m}^{(n)}(\tilde{c}_1, \ldots, \tilde{c}_n) + \left(\mu_1^{\mathbb{Q}} - \mu_1^{Obs}\right)\right]^2 \\ &\leq \frac{1}{M}\sum_{m=1}^{M}\left(C_{K_m}^{Obs}(t,T) - \tilde{C}_{K_m}^{(n)}(\tilde{c}_1, \ldots, \tilde{c}_n)\right)^2 + \frac{1}{M}\sum_{m=1}^{M}\left(P_{K_m}^{Obs}(t,T) - \tilde{P}_{K_m}^{(n)}(\tilde{c}_1, \ldots, \tilde{c}_n)\right)^2 + \left(\mu_1^{\mathbb{Q}} - \mu_1^{Obs}\right)^2,\end{aligned}$$

which yields the following inequality

$$\Delta^{Q} \geq \Delta^{\text{pcp}} - d_\mu^2. \tag{37}$$

where $d_\mu = \left(\mu_1^{\mathbb{Q}} - \mu_1^{Obs}\right)$. From (37) we obtain a lower bound for the tolerance level that must be allowed on the estimation residual. Moreover, it proves that admissible solutions of (28) with tolerance level lower than $\Delta^{\text{pcp}} - \left(\mu_1^{\mathbb{Q}} - \mu_1^{Obs}\right)^2$ do not exist. [2] Therefore, (37) 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.

6.2 Robustness to no-arbitrage violations

We now assess the practical validity of Proposition 6.1 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 (22). 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 are given by

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

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

$$\sigma_{C,i}^2 = \text{Var}[\epsilon_i^C], \quad \sigma_{P,i}^2 = \text{Var}[\epsilon_i^P], \quad i = 1, \ldots, M.$$

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, $\sigma_C^2 = [\sigma_{C,1}^2, \ldots, \sigma_{C,M}^2]$ and $\sigma_P^2 = [\sigma_{P,1}^2, \ldots, \sigma_{P,M}^2]$ are assumed to be an increasing and a decreasing vector, respectively.

To identify σ_C^2 and σ_P^2 we further assume that the arbitrage error ϵ^F induced on the vector $\mathbf{F}$ of future prices implied by put-call parity consists of i.i.d. components, that is

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

where

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

is the unique arbitrage-free future price. Hence $\epsilon^F = \epsilon^C - \epsilon^P$ and $E[\epsilon^F] = 0$. Assuming $\text{Var}[\epsilon^F] :=$

[2] Note that this level is certainly positive if the underlying future price is fitted sufficiently well by the estimated RND.

$\sigma_F^2 < \infty$, the identification of $\mathrm{Var}[\epsilon^C]$ and $\mathrm{Var}[\epsilon^P]$ can therefore be achieved by

$$\sigma_{C,i}^2 + \sigma_{P,i}^2 = \sigma_F^2, \quad \frac{\sigma_{C,i}^2}{\sigma_{P,i}^2} = \frac{\mathbf{C}_i}{\mathbf{P}_i}, \quad i = 1, \dots, M,$$

which gives

$$\sigma_{P,i}^2 = \frac{\sigma_F^2}{1 + \frac{\mathbf{C}_i^2}{\mathbf{P}_i^2}}, \quad \sigma_{C,i}^2 = \sigma_F^2 - \sigma_{P,i}^2, \quad i = 1, \dots, M. \tag{38}$$

Note that observable quantity Δ^{pcp} defined in (36) is an 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 (35), Δ^{pcp} can consistently approximate the right hand of (35). We therefore carry out a number $N = 1000$ of Monte Carlo simulations with the purpose of investigating the robustness of the proposed method to the no-arbitrage violations and the usefulness of the threshold Δ^{pcp} to provide an useful 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 $T - t = 30/365$. 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 (38), 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 $\sigma_F = 0.01, 0.03, 0.05$.[3] The results of these Monte Carlo simulations are summarized in Table (3). The so called *divergence rate*, which is associated to the cases in which the RSME exceeds the threshold $2\sqrt{\Delta^{\mathrm{pcp}}}$, is intended to give of proxy on how frequently the conditions of Proposition 6.1 are not met. On the other hand, the second column of Table 3 endorses the validity of (35), since the RSME is below $\sqrt{\Delta^{\mathrm{pcp}}}$ in a large percentage of cases. Furthermore, by looking at the Monte Carlo average RMSE it emerges that variance of the error associated to the expansion of order 10 decreases with σ_F and 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 3 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 on data, namely σ_F, decreases. This is consistent with Proposition 6.1 since the hypotheses 6.1.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

[3]Typical values of $\sqrt{\Delta^{\mathrm{pcp}}}$ 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.

- 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 3 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 σ_F^2 decreases, as shown in the fifth column. Finally, the last column of Table 3 confirms that the estimation based on expansions of order 10 outperforms the related kernel, in all cases that we consider.

					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 3: The table summarizes the results of the 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 (37), 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).

7 Empirical Analysis

In this section we apply the methodology for the estimation of the RND of VIX based on observed option prices. The sample consists of 64 contracts quoted by the CBOE on November 16, 2011 and expiring on December 21, 2011. The data is obtained from the *OptionMetrics* database. The choice of this date is not coincidental. Indeed, the end of 2011 is characterized by extremely high levels of market volatility, registered in connection with the turmoil caused by the European sovereign debt crisis and the US sovreign debt downgrading . As a consequence, on November 16, 2011, the VIX reached the high value of 33.51%. Consistently, actively traded VIX options spanned a strike range between 15$ and 90$, while under normal market conditions VIX option strikes would typically fall between 10$ and 45$, with the range being expected to remain stable over time, due to mean-reversion of the VIX. Hence, on the chosen date, deep out-of-the-money options were associated to a sufficiently high trading volume to ensure informative market prices in a wide range of strikes and, in particular, suggest an uncommonly long right-tail of the RND.

Before proceeding to the estimation of the RND, we operate minimum pre-filtering of the data. More specifically we exclude all OTM puts (calls) with mid-quote below 0.05\$ as well as the corresponding ITM calls (puts) with the same strike. These contracts turn out to be highly illiquid (if traded at all) and likely to be mis-priced, thus they are not informative on the RND. Obviously, this solution has little impact when, as in the instance considered here, contracts are traded over a sufficiently large set of strikes. After filtering, we end up with 52 contracts (26 calls and 26 puts) with strikes ranging between 21\$ and 80\$.

We consider two different kernels: the GIG and the GW. The kernel parameters - $\theta = [\alpha, \beta, \xi]'$ for the GIG kernel and $\theta = [\alpha, \beta, p]'$ for the GW kernel - are estimated using the procedure detailed in Section 4.1.2, which ensures a correct specification of the option prices representation given in (23). We also operate a displacement of the kernel to ensure that $\int_0^{K_{min}} \tilde{f}_{\mathbb{Q}}^{(n)} \mathrm{d}x = 0$. Following the arguments detailed in Section 4.1.4, K_{min} is set to 16.5\$ for the GIG kernel and 18.2\$ for the GW kernel. Table 4 reports the parameters estimated for the two kernels. As discussed in Section 4.1.1, the main advantage of the

	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 4: The table reports the kernel parameters for the GIG and GW kernels respectively, estimated using the procedure detailed in Section 4.1.2.

procedure that we propose in this paper lies in the fact that we can arbitrarily increase the expansion order n thanks to the orthogonalization of the regressors in (23) and the dimension reduction operated by applying the principal component approach. Therefore, we set a relatively high order, that is $n = 18$, and we estimate the RND by using the solution $c_1(\beta), \ldots, c_{18}(\beta)$ of (30). Given a threshold of 99% on the explained total variance, this corresponds to using 6 (GIG kernel) and 5 (GW kernel) principal components. The estimated RNDs obtained using expansions of order $n = 18$ are reported in Figure 5 (solid lines), together with the corresponding kernels (dashed lines). In the following, we discuss the results in terms of estimation residuals and features related to the estimated RNDs.

7.1 Analysis of the estimation residuals

In order to evaluate the accuracy of the estimated RND obtained by using the expansion (29) for $n = 18$, we plot in Figure 6 the residuals from (27). As a reference, we also report the residuals for the intermediate case of an expansion of order $n = 9$, i.e. $\tilde{f}_{\mathbb{Q}}^{(9)}$. These expansions use 5 principal components for both GIG and GW kernels. As the order of expansion increases, and with virtually no cost in terms of estimation of extra parameters, we note a dramatic drop in the cross-sectional dependence (across

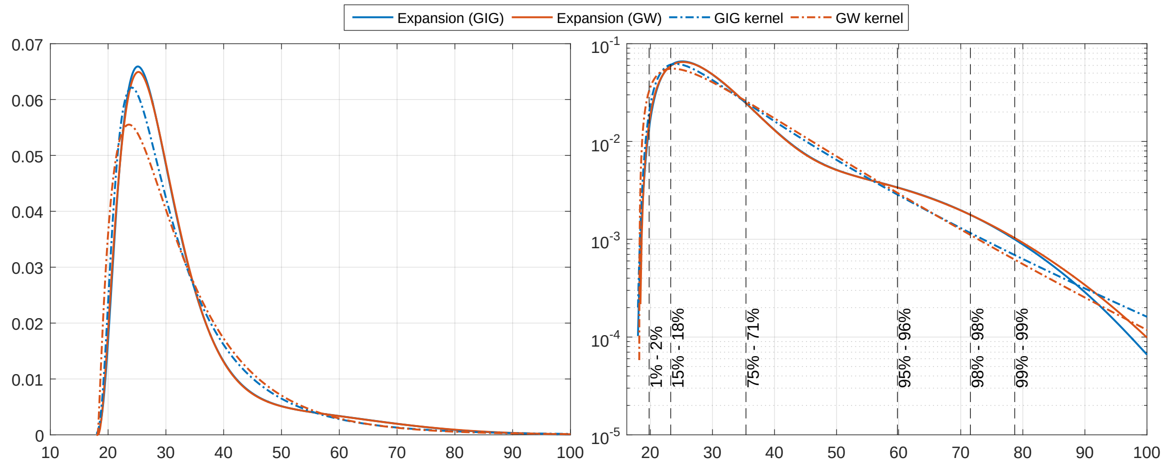
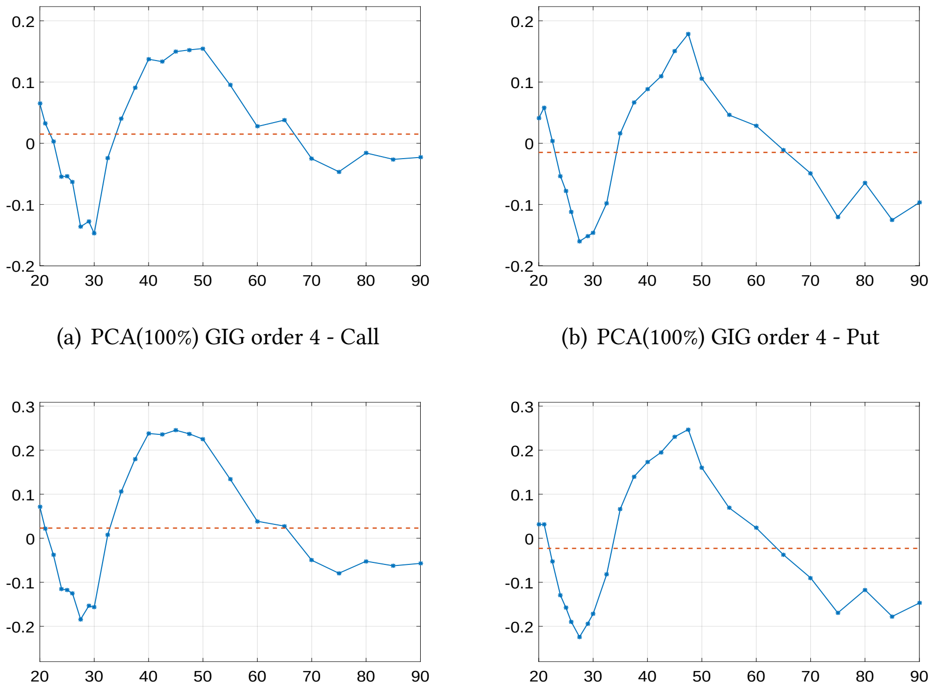


Figure 5: Estimated RNDs. The left panel depicts the graphs of the two kernels considered in this section as well as the corresponding estimated RNDs obtained for $n = 18$. The right panel reports the same contents of the left panel in semi-log scale, to highlight tail features that are difficult to observe in linear scale. The percentage values denote average mass levels of the two kernels and the two corresponding estimated RNDs, respectively, related to the quantiles identified by the dashed vertical lines.

strikes) and a reduction in the mispricing for blocks of neighboring strikes. Similarly, we observe a steady increase in the accuracy in the fitting. Indeed, the root-mean squared error (RMSE) of the residuals drops from 0.1 for the GIG kernel (0.16 for the GW kernel) to 0.04 when the expansion is of order 9, and it lands on 0.03 when $n = 18$. In both cases the RMSE with $n = 18$ is very close to the value $\sqrt{\Delta^{\text{pcp}}} = 0.031$, computed as described in (36), meaning that we have reached convergence. Following the discussion provided in Section 6, we can conclude that the error ε mostly consists of a component due to no-arbitrage violations. In other words, we achieved the best possible fitting to option data, see Proposition 6.1. For sake of comparison, we also report the fitting results obtained by using the an expansion of order 4 with 100% of explained variance by the PCA, which is comparable to what has been done in Corrado and Su (1996b) and Jondeau and Rockinger (2001), among others. Since all the principal components are used, the PCA is adopted in this case only to orthogonalize the regressors without reducing the dimensionality of the problem. This is analogous in spirit to the standardization adopted in Corrado and Su (1996b) to the third and fourth cumulants to overcome the multicollinearity problem discussed above.

As discussed in Section 4.1, the multicollinearity of the regressors in (23) implicitly constraints the order of expansion to a limited number of terms (four in this case), which proves to be unreasonably restrictive in practice for the VIX. Indeed, the RND of the VIX exhibits a characteristic tail behavior, meaning that moments higher than the fourth may provide significant information on its shape. Figure 7 shows that such a low order expansion proves inadequate and does not provide any sizable improvement to the corresponding zero-order expansions, i.e., the baseline GIG and GW kernels. The same result is confirmed by examining the residuals of the regression, reported in Figure 8. For both kernel expansions, the residuals exhibit a high degree of cross sectional dependence, systematic over (under) pricing over subset of neighboring strikes and virtually the same variability of the ones computed using the corresponding simple kernels (i.e., the RMSE of the residuals is 0.09 and 0.14 for the GIG and the GW expansions, respectively).



(a) PCA(100%) GIG order 4 - Call

(b) PCA(100%) GIG order 4 - Put

(c) PCA(100%) GW order 4 - Call

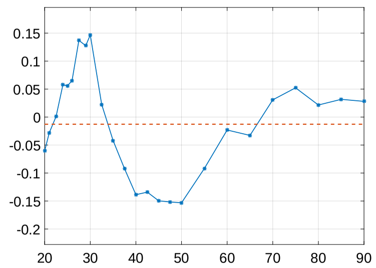
(d) PCA(100%) GW order 4 - Put

Figure 8: Call and put price residuals obtained by estimation based on the GIG and GW kernel. Parameters estimated with $n = 4$ and PCA and 100% of explained variance.

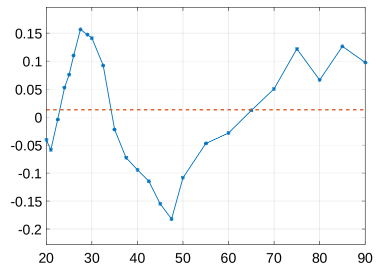
7.2 Features of the estimated RND

Although showing seemingly comparable values in terms of absolute difference, see first and second column of Table 5, the moments of the two alternative kernels exhibit notable relative differences, see third column of Table 5, in particular regarding the kurtosis. This is reflected into tangibly different right tails of the two densities, in a part of the domain that identifies roughly 98% of their average mass, as it emerges from Figure 5. The left tails are instead rather similar and both suggest that a RND consistent with VIX options should decay very quickly on the left tail, which is located very far from the origin. This supports the use of the kernel displacement discussed above.

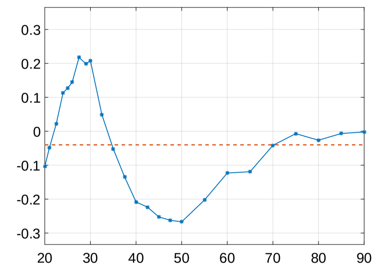
Irrespective of the initial choice of the kernel, the expansion yields a substantial correction to the estimated kernel. As compared to the simple kernels, the relative spread between the moments of the estimated densities is reduced by a factor of 4 to 5 times, see last column of Table 5, and both the estimated RNDs assign a higher probability to tail-events than the corresponding kernels. The right tail of the RND plays a determinant role in the pricing of OTM calls and ITM puts. In this sense, the importance of the right tail correction provided by the orthogonal polynomials is better understood by comparing the implied volatility curves obtained from OTM call and ITM put options generated by the



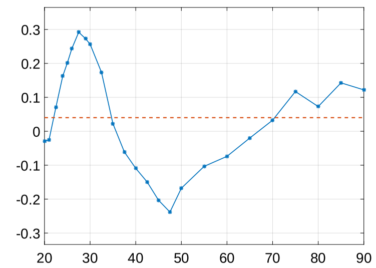
(a) GIG Kernel - Call



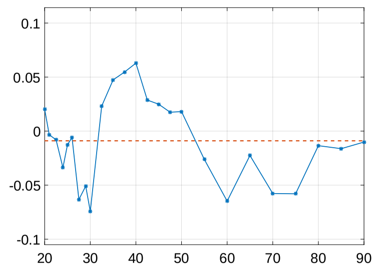
(b) GIG Kernel - Put



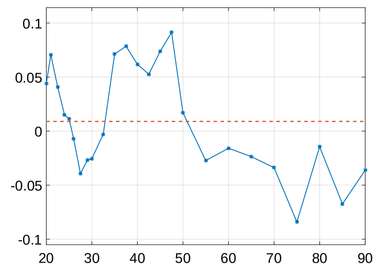
(c) GW Kernel - Call



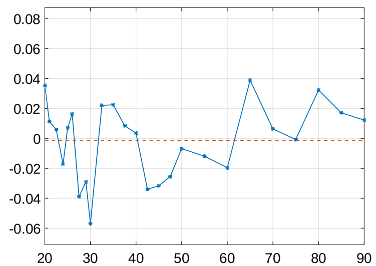
(d) GW Kernel - Put



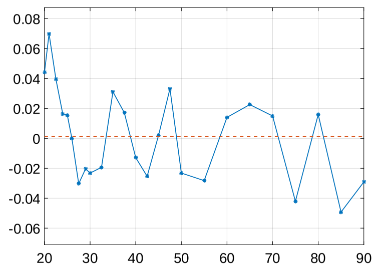
(e) PCA(99%) GIG order 9 - Call



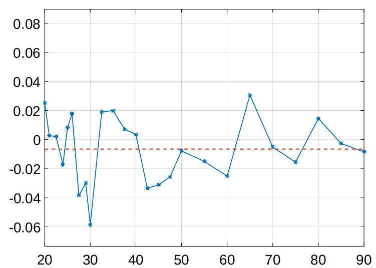
(f) PCA(99%) GIG order 9 - Call



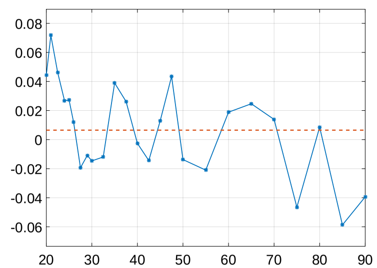
(g) PCA(99%) GW order 9 - Call



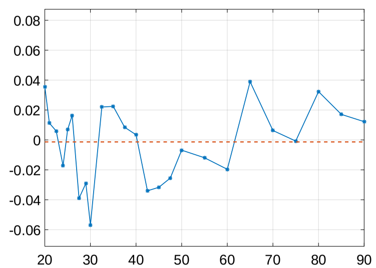
(h) PCA(99%) GW order 9 - Call



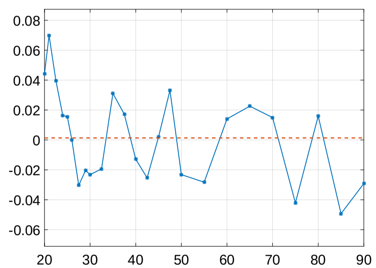
(i) PCA(99%) GIG order 18 - Call



(j) PCA(99%) GIG order 18 - Call



(k) PCA(99%) GW order 18 - Call



(l) PCA(99%) GW order 18 - Call

Figure 6: Call and put price residuals obtained by estimations based on the GIG and GW kernels of increasing order.

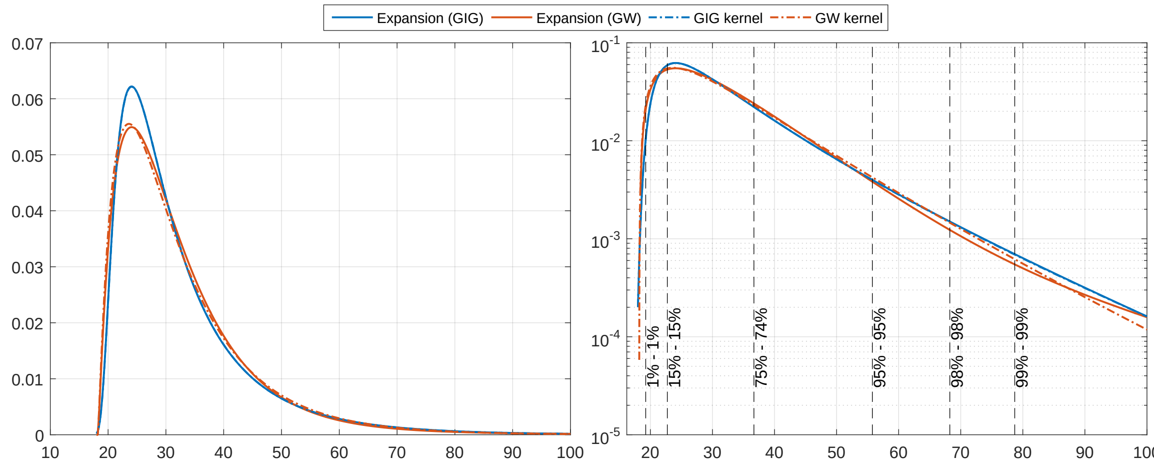


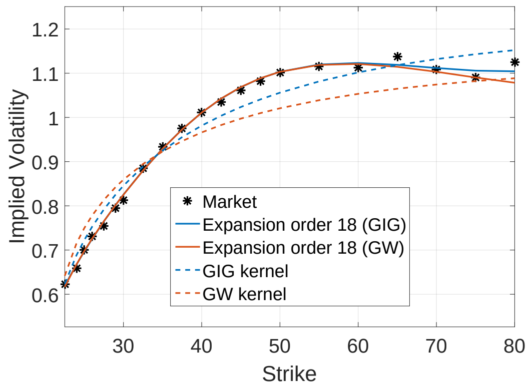
Figure 7: Estimated RNDs. The left panel depicts the graphs of the two kernels considered in this section as well as the corresponding estimated RNDs obtained for $n = 4$ with PCA and 100% of explained variance. The right panel reports the same contents of the left panel in semi-log scale, to highlight tail features that are difficult to observe in linear scale. The percentage values denote average mass levels of the two kernels and the two corresponding estimated RNDs, respectively, related to the quantiles identified by the dashed vertical lines.

	GIG kernel	GW kernel	Perc. Diff.	Expansion (GIG)	Expansion (GW)	Perc. Diff.
Mean	32.63	32.58	0.15%	32.67	32.66	0.03%
Variance	151.13	144.24	4.56%	153.60	151.93	1.08%
Skewness	21.41	20.57	3.92%	20.31	20.17	0.69%
Kurtosis	12.44	11.29	9.24%	11.15	10.91	2.15%

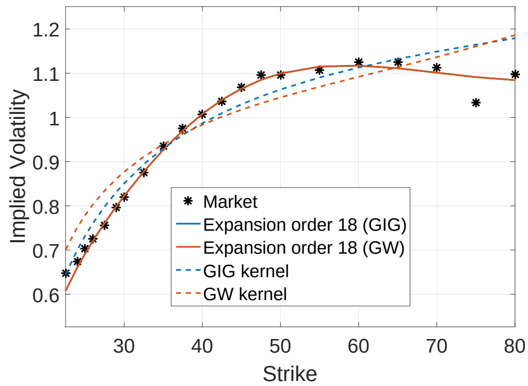
Table 5: The table reports mean, variance, standardized skewness and kurtosis of the kernel densities (GIG kernel and GW kernel) calibrated to VIX option prices and of the related orthogonal expansions of order 18. The third column reports the relative difference between the first and second column, while the sixth column reports the relative difference between the fourth and the fifth column.

two kernels and their corresponding expansions, reported in Figure 9. First, we observe that the implied volatility curves, associated with ITM put and generated by the two kernels (right panel, dashed lines), are more similar than the corresponding implied volatility curves of OTM call (left panel, dashed lines). Second, we observe that the implied volatilities generated by both kernels are considerably different from the implied volatilites generated by market prices (solid lines), with the differences tending to enlarge as the options go OTM. Third, the correction provided by the orthogonal expansion is substantial and, as a matter of fact, the expansion proves 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 VIX risk-neutral density, and 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. The degree of flexibility that the expansion offers emerges clearly, in that it adequately approximates densities with complicated shapes. Also, ac-



(a) Call Imp. Vol.



(b) Put Imp. Vol.

Figure 9: Black and Scholes implied volatility curves obtained from market prices, GIG kernel, GW kernel, and the resulting approximated RNDs with expansion order equal to $n = 18$.

cordingly to Lemma 3.3, the expansions obtained using the two different kernels converge to the same density, with small differences between the GIG and the GW expansions that become noticeable only deep into the right tail - the two functions diverge from each other starting from roughly the 99.5% quantile. In terms of implied volatilities, the two expansions generate indistinguishable curves in the range of strikes between 20$ and 70$, while small differences are observed for strikes above 70$. In particular, the expansion based on the GIG kernel seems to overprice call options with strike above 70$, as compared to the expansion based on the GW kernel. This might be related to a slight overestimate of VIX variance and kurtosis provided by the GIG-based expansion, see third and fourth column in Table 5.

8 Conclusion and future research

In this paper, we proposed a methodology based on a finite orthogonal expansion to infer the RND of the underlying asset from option prices. The methodology closely relates to Hermite expansions and generalizes the Laguerre expansions to suit cases where the density is supported over the positive real axis. We emphasize the importance of a proper choice of the kernel, which turns out to be a crucial aspect not only from a theoretical but also from an empirical point of view. 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. Our approach drastically reduces the intrinsic risk of misspecification entailed when working with transformed data. Particular attention is dedicated to the statistical aspects and practical implementation of the method. Adopting PCA proves a robust and effective solution to overcome the problem of multicollinearity and to reduce the number of coefficients to be estimated. We use the developed technique to recover the RND underlying VIX options, although the same principle can be applied to different classes of financial derivatives sharing the same characteristics, e.g. the VVIX or interest rates. The empirical application to VIX options shows that this technique is very successful in retrieving the underlying RND. Thus, our results may be particularly useful when the estimation of time-varying risk-premia embedded in option prices is of interest, see Bollerslev and Todorov (2011) and Andersen et al. (2015), or to compute a VaR on volatility (*VolaR*, see Caporin et al., 2014), adjusted for risk aversion as in Aït-Sahalia and Lo (2000). The investigation of the dynamic properties of the RND or the extension to the multivariate case are further natural extensions of this work. However, we leave such extensions to future research.

A Appendix

A.1 Proof of Theorem 3.1

We start by introducing the following notation for the i-th moment of ϕ

$$\mu_i = \int_{-\infty}^{+\infty} x^i \phi(x) dx, \tag{39}$$

and by defining

$$h_0^\phi(x) = 1, \quad h_1^\phi(x) = \frac{x - \mu_1}{\sqrt{\mu_2 - \mu_1^2}}.$$

It is straightforward to verify that

$$\int_{-\infty}^{+\infty} h_1^\phi(x) h_0^\phi(x) \phi(x) dx = 0, \quad \int_{-\infty}^{+\infty} h_1^\phi(x)^2 \phi(x) dx = 1.$$

To determine h_k^ϕ, for $k \geq 2$, we impose the following recursive relation

$$h_k^\phi(x) = \frac{1}{C_k}\left[(x - a_k) h_{k-1}^\phi - b_k h_{k-2}^\phi\right], \tag{40}$$

and seek for a solution a_k, b_k, and $C_k \neq 0$. We notice that, by construction, for every $a_k, b_k \in \mathbb{R}$ and $C_k \neq 0$, we have

$$\int_{-\infty}^{+\infty} h_k^\phi(x) h_j^\phi(x) \phi(x) dx = 0, \qquad \forall 0 \leq j < k-2.$$

Then, we only need to plug (40) into the system

$$\int_{-\infty}^{+\infty} h_k^\phi(x) h_j^\phi(x) \phi(x) dx = \delta_{ik}, \qquad j = k-2, k-1, k, \tag{41}$$

Solving (41) with respect to a_k,b_k, and C_k yields

$$a_n = \int_{-\infty}^{+\infty} h_{k-1}^\phi(x)^2 x \phi(x) dx, \quad b_n = \int_{-\infty}^{+\infty} h_{k-1}^\phi(x) h_{k-2}^\phi(x) x \phi(x) dx,$$
$$C_k^2 = \int_{-\infty}^{+\infty} \left[(x - a_k) h_{k-1}^\phi - b_k h_{k-2}^\phi\right]^2 \phi(x) dx.$$

To show that $\left(h_k^\phi\right)_{k\in\mathbb{N}}$ is uniquely determined up to a sign let us consider a polynomial h of degree k such that

$$\int_{-\infty}^{+\infty} h(x) h_j^\phi(x)\phi(x)dx = 0, \qquad j = 0, \ldots, k-1 \tag{42}$$

then the $k+1$ coefficients of h must solve a linear system of k equations, therefore are determined in a one-dimensional space and uniquely up to a sign if (7) is also required.

A.2 Proof of Lemma 3.2

We first point that

$$\int_{-\infty}^{+\infty} x^p f_{\mathbb{Q}}^{(n)}(x)dx = \int_{-\infty}^{+\infty} x^p f_{\mathbb{Q}}^{(p)} dx + \sum_{k=p+1}^{n} c_k \int_{-\infty}^{+\infty} x^p h_k^\phi(x)\phi(x)dx.$$

Then, since for every $k \in \mathbb{N}$ we can write x^k as a linear combination of $h_0^\phi(x), \ldots, h_k^\phi(x)$, by (6) we get

$$\int_{-\infty}^{+\infty} x^p h_k^\phi(x)\phi(x)dx = 0,$$

which gives

$$\int_{-\infty}^{+\infty} x^p f_{\mathbb{Q}}^{(n)}(x)dx = \int_{-\infty}^{+\infty} x^p f_{\mathbb{Q}}^{(p)} dx.$$

A.3 Proof of Theorem 3.6

The proof of this theorem relies on standard results on orthogonal polynomials. In order to adapt these results to our context, we need the following lemma.

Lemma A.1. *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 $f \in L^2_{\phi(x)dx}(\mathcal{D})$ and

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

then $f(x) = 0$ 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 $h(x)$. $\square$

Proof of 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

$$\lim_{|x|\to+\infty} \psi(x)e^{\varsigma|x|} = 0 \tag{43}$$

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 (43), 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 $\mathcal{D}$. This concludes the proof of statement (i).

Proof of statement (ii) We base our proof on a counterexample by V.A. Steklov (cf. entry "Closed system of elements" in Hazewinkel (1988)), showing that every function ψ of the form

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

By combining this counterexample with the results of Shohat (1942), we can conclude that the function ψ supported on $[0, +\infty[$ and defined by

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

does not generate closed polynomial sets. By proceeding similarly to the proof of (i), this result can be proven also when ψ is of the form

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

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.1, 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.4 Proof of Theorem 4.1

We remark that $\mathcal{H}_\phi$ and $\mathcal{H}_\phi^*$ endowed with the scalar product

$$\langle \psi_1, \psi_2 \rangle := \int_{\mathcal{D}} \psi_1(x) \psi_2(x) \frac{1}{\phi(x)} dx$$

are separable Hilbert spaces and d_ϕ is the induced distance. Then, (a) follows immediately from Lemma 3.3 by noticing that the assumptions on ϕ imply $\mathcal{H}_\phi = \mathcal{H}_\phi^*$ in view of Theorem 3.6. To prove (18), we observe that for every $\phi \in \mathcal{H}_\phi^*$ and every $n \in \mathbb{N}$

$$\begin{aligned} \left| \int_0^{+\infty} \Pi(x) f_{\mathbb{Q}}^{(n)}(x) dx - \int_0^{+\infty} \Pi(x) f_{\mathbb{Q}}^{(\infty)}(x) dx \right| &\leq \int_0^{+\infty} \Pi(x) \left| f_{\mathbb{Q}}^{(n)}(x) - f_{\mathbb{Q}}^{(\infty)}(x) \right| dx \\ = \int_{\mathcal{D}} \phi^{\frac{1}{2}}(x) \Pi(x) \left| \phi^{-\frac{1}{2}}(x) f_{\mathbb{Q}}^{(n)}(x) - \phi^{-\frac{1}{2}}(x) f_{\mathbb{Q}}^{(\infty)}(x) \right| dx &\leq C \cdot \mathrm{d}_\phi \left(f_{\mathbb{Q}}^{(n)}, f_{\mathbb{Q}}^{(\infty)} \right), \end{aligned}$$

where $C = \| \phi^{\frac{1}{2}} \Pi \|_{L^2(\mathcal{D})}$ is finite by hypothesis. Then (18) follows from Lemma 3.3.

A.5 Proof of Proposition 6.1

For the sake of readability, throughout this proof we omit the dependence on t, T of $\mathbf{Q}^{(n)}, C^{Obs}, P^{Obs}, \mathcal{C}, \mathcal{P}$. Moreover, we denote by $(c_k)_{k\in\mathbb{N}}$ and $f_{\mathbb{Q}}^{(n)}$ the quantities defined in Theorem 4.1-(a). For every $n \in \mathbb{N}$ we have

$$
\begin{aligned}
&(K_M - K_1) \cdot E\left[\mathbf{Q}^{(n)}(\hat{c}(n))\right] = E\left[\int_I \left(C_K^{Obs} - C_K^{(n)}(\hat{c}(n))\right)^2 + \left(P_K^{Obs} - P_K^{(n)}(\hat{c}(n))\right)^2 dK\right] \\
&\leq E\left[\int_I \left(\mathcal{C}_K + \epsilon_K^C - C_K^{(n)}(c_1,\dots,c_n)\right)^2 + \left(\mathcal{P}_K + \epsilon_K^P - P_K^{(n)}(c_1,\dots,c_n)\right)^2 dK\right] \\
&= E\left[\int_I \left(\mathcal{C}_K - C_K^{(n)}(c_1,\dots,c_n)\right)^2 + \left(\epsilon_K^C\right)^2 + 2\epsilon_K^C\left(\mathcal{C}_K - C_K^{(n)}(c_1,\dots,c_n)\right) dK\right] \\
&+ E\left[\int_I \left(\mathcal{P}_K - P_K^{(n)}(c_1,\dots,c_n)\right)^2 + \left(\epsilon_K^P\right)^2 + 2\epsilon_K^P\left(\mathcal{P}_K - P_K^{(n)}(c_1,\dots,c_n)\right) dK\right] \\
&= \int_I \left(\mathcal{C}_K - C_K^{(n)}(c_1,\dots,c_n)\right)^2 + \left(\mathcal{P}_K - P_K^{(n)}(c_1,\dots,c_n)\right)^2 dK + E\left[\int_I \left(\epsilon_K^C\right)^2 + \left(\epsilon_K^P\right)^2 dK\right].
\end{aligned}
$$

Since C, P are arbitrage-free, the following identities hold

$$
\mathcal{C}_K = \int_K^{+\infty} f_{\mathbb{Q}}(x)(K-x)dx, \quad \mathcal{P}_K = \int_0^K f_{\mathbb{Q}}(x)(x-K)dx, \quad K \geq 0.
$$

Then

$$
\begin{aligned}
&\int_I \left(\mathcal{C}_K - C_K^{(n)}(c_1,\dots,c_n)\right)^2 + \left(\mathcal{P}_K - P_K^{(n)}(c_1,\dots,c_n)\right)^2 dK \\
&= \int_I \left[\int_K^{+\infty} \left(f_{\mathbb{Q}}(x) - f_{\mathbb{Q}}^{(n)}(x)\right)(K-x)dx\right]^2 + \left[\int_0^K \left(f_{\mathbb{Q}}(x) - f_{\mathbb{Q}}^{(n)}(x)\right)(x-K)dx\right]^2 dK \\
&\leq \int_I \left[\int_0^{+\infty} \left|f_{\mathbb{Q}}(x) - f_{\mathbb{Q}}^{(n)}(x)\right| |K-x| dx\right]^2 dK \leq \mathrm{d}_\phi\left(f_{\mathbb{Q}}, f_{\mathbb{Q}}^{(n)}\right)^2 \cdot \int_I \int_0^{+\infty} (K-x)^2 \phi(x) dx dK,
\end{aligned}
$$

where d_ϕ is the L^2 distance defined in Lemma 3.3. Since

$$
\int_I \left[\int_0^{+\infty} (K-x)^2 \phi(x) dx\right]^2 dK < +\infty,
$$

then, in view of Theorem 4.1

$$
\lim_{n\to+\infty} \int_I \left(\mathcal{C}_K - C_K^{(n)}(c_1,\dots,c_n)\right)^2 + \left(\mathcal{P}_K - P_K^{(n)}(c_1,\dots,c_n)\right)^2 dK = 0,
$$

which proves (35). The proof is concluded by noticing that, under the additional hypotheses (i) and (ii)

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

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. (2015). 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. 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. (2014). Chasing volatility - A persistent multiplicative error model with jumps. CREATES Research Papers 2014-29.

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.

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

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.*

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.

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.

Monteiro, A. M., Tütüncü, R. H., and Vicente, L. N. (2008). Recovering risk-neutral probability density functions from options prices using cubic splines and ensuring nonnegativity. *European Journal of Operational Research*, 187(2):525–542.

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.

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.

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.

Research Papers 2016

CREATES
Center for Research in Econometric Analysis of Time Series

2016-03: Bent Jesper Christensen and Rasmus T. Varneskov: Dynamic Global Currency Hedging

2016-04: Markku Lanne and Jani Luoto: Data-Driven Inference on Sign Restrictions in Bayesian Structural Vector Autoregression

2016-05: Yunus Emre Ergemen: Generalized Efficient Inference on Factor Models with Long-Range Dependence

2016-06: Girum D. Abate and Luc Anselin: House price fluctuations and the business cycle dynamics

2016-07: Gustavo Fruet Dias, Cristina M. Scherrer and Fotis Papailias: Volatility Discovery

2016-08 N. Haldrup, O. Knapik and T. Proietti: A generalized exponential time series regression model for electricity prices

2016-09: Ole E. Barndorff-Nielsen: Assessing Gamma kernels and BSS/LSS processes

2016-10: Tim Bollerslev, Andrew J. Patton and Rogier Quaedvlieg: Modeling and Forecasting (Un)Reliable Realized Covariances for More Reliable Financial Decisions

2016-11: Tom Engsted and Thomas Q. Pedersen: The predictive power of dividend yields for future inflation: Money illusion or rational causes?

2016-12: Federico A. Bugni, Mehmet Caner, Anders Bredahl Kock and Soumendra Lahiri: Inference in partially identified models with many moment inequalities using Lasso

2016-13: Mikko S. Pakkanen and Jani Lukkarinen: Arbitrage without borrowing or short selling?

2016-14: Andrew J.G. Cairns, Malene Kallestrup-Lamb, Carsten P.T. Rosenskjold, David Blake and Kevin Dowd: Modelling Socio-Economic Differences in the Mortality of Danish Males Using a New Affluence Index

2016-15: Mikkel Bennedsen, Ulrich Hounyo, Asger Lunde and Mikko S. Pakkanen: The Local Fractional Bootstrap

2016-16: Martin M. Andreasen and Kasper Jørgensen: Explaining Asset Prices with Low Risk Aversion and Low Intertemporal Substitution

2016-17: Robinson Kruse, Christian Leschinski and Michael Will: Comparing Predictive Accuracy under Long Memory - With an Application to Volatility Forecasting

2016-18: Søren Johansen and Bent Nielsen: Tightness of M-estimators for multiple linear regression in time series

2016-19: Tim Bollerslev, Jia Li and Yuan Xue: Volume, Volatility and Public News Announcements

2016-20: Andrea Barletta, Paolo Santucci de Magistris and Francesco Violante: Retrieving Risk-Neutral Densities Embedded in VIX Options: a Non-Structural Approach