This is the accepted manuscript (post-print version) of the article. Contentwise, the post-print version is identical to the final published version, but there may be differences in typography and layout.

How to cite this publication
Please cite the final published version:

doi: [https://doi.org/10.1016/j.jbankfn.2018.11.012](https://doi.org/10.1016/j.jbankfn.2018.11.012)

Publication metadata

Title: A Non-Structural Investigation of VIX Risk Neutral Density
Author(s): Andrea Barletta, Paolo Santucci de Magistris, Francesco Violante
Journal: *Journal of Banking and Finance*
DOI/Link: [10.1016/j.jbankfn.2018.11.012](https://doi.org/10.1016/j.jbankfn.2018.11.012)
Document version: Accepted manuscript (post-print)
A Non-Structural Investigation of VIX Risk Neutral Density

Andrea Barletta, Paolo Santucci de Magistris, Francesco Violante

PII: S0378-4266(18)30252-8
DOI: https://doi.org/10.1016/j.jbankfin.2018.11.012
Reference: JBF 5460

To appear in: Journal of Banking and Finance

Received date: 21 January 2018
Revised date: 4 October 2018
Accepted date: 14 November 2018

Please cite this article as: Andrea Barletta, Paolo Santucci de Magistris, Francesco Violante, A Non-Structural Investigation of VIX Risk Neutral Density, Journal of Banking and Finance (2018), doi: https://doi.org/10.1016/j.jbankfin.2018.11.012

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.
A Non-Structural Investigation of VIX Risk Neutral Density *

Andrea Barle/t_ta† Paolo Santucci de Magistris ‡ Francesco Violante §

November 14, 2018

Abstract

We propose a non-structural method to retrieve the risk-neutral density (RND) implied by options on the CBOE Volatility Index (VIX). The methodology is based on orthogonal polynomial expansions around a kernel density and yields the RND of the underlying asset without the need for a parametric specification. The classic family of Laguerre expansions is extended to include the GIG and the generalized Weibull kernels. We show that orthogonal polynomial expansions yield accurate approximations of the RND of VIX and, in some cases, they outperform commonly used non-parametric methods. Based on a panel of observed VIX options, we retrieve the variance swap term structure, the time series of VVIX, the VIX risk-neutral moments and the Volatility-at-Risk, which reveal a number of stylized facts on the RND of VIX.

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

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

---

*We are grateful to Celso Brunetti, Fulvio Corsi, Friedrich Hubalek, Elisa Nicolato, Roberto Renò and Viktor Todorov for useful comments and suggestions that improved the quality of this work. The authors would like to thank also the participants of the CFE 2015 conference in London, the Empirical Finance 2016 workshop in Paris, the VCMF 2016 conference in Vienna, the HAX 2017 conference in Aarhus, the NEM 2017 conference in Tartu, the IAAE 2017 conference in Sapporo and the seminar participants at LUISS and at Aarhus University for helpful comments. Andrea Barletta acknowledges the research support of the European Union Seventh Framework Programme (FP7/2007-2013) under grant No. 289032 (HPCFinance) and of the Danish Council for Independent Research, Social Sciences. Paolo Santucci de Magistris and Francesco Violante acknowledge the research support of CREATES, funded by the Danish National Research Foundation (DNRF78). Francesco Violante also acknowledges the research support of the Danish Council for Independent Research, Social Sciences, Individual Research grant and Sapere Aude Research Talent grant.

†Department of Economics and Business Economics, Aarhus University, Denmark. abarletta@econ.au.dk
‡LUISS University, Department of Economics and Finance, Rome, Italy and CREATES. psantucci@econ.au.dk
§ENSAE - École Nationale de la Statistique et de l’Administration Économique, Paris, France and CREATES, Aarhus University, Denmark. francesco.violante@ensae.fr
1 Introduction

The Volatility Index (VIX), introduced in 1993 by the Chicago Board Options Exchange (CBOE), measures the market expected volatility based on call and put options on the S&P 500 index (SPX). The VIX is traded through derivative contracts, namely, futures and options, which have been introduced in 2006. Under standard assumptions of an arbitrage-free and complete market, the main goal of this paper is to provide a robust methodology to estimate the risk-neutral density (RND) of the VIX from option prices, as well as new insights on its empirical properties. The problem of the estimation of the RND is inextricably linked to that of derivative pricing: any pricing model fitted to the observed options entails an implicit estimate of the RND, and vice versa. For instance, an important strand of financial literature has addressed VIX derivative pricing under stochastic volatility models, mostly within the affine class, see 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). Bardgett et al. (2014) further generalize 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 Bayer et al. (2013). Modeling frameworks based on infinite-dimensional specifications of the variance swap term structure are proposed by Buehler (2006), Bergomi (2008), and Cont and Kokholm (2013). Finally, Lin (2013) and Huskaj and Nossman (2013) address the term structure of VIX futures, while ensuring absence of arbitrage opportunities across the SPX and the VIX markets.

A common feature of these contributions is that the RND is assumed to be fully described by stochastic dynamic equations of the state-variables, which are functions of the underlying model parameters. As discussed in Cont (2006), fully parametric specifications of the dynamics of price and volatility come at the cost of an intrinsic risk of misspecification affecting the implied RND estimates. The issue of misspecification remains when the focus is on the density of the VIX. For instance, although Christoffersen et al. (2010) and Wang and Daigler (2011) find some evidence in favor of models that assume lognormal dynamics for the instantaneous
variance, none of these models achieve sufficiently small pricing errors over the entire range of strike prices. Furthermore, the econometric analysis carried out by Mencia and Sentana (2013) reveals that the risk of model misspecification is particularly high during financial crises. Reducing the risk of misspecification of the RND provided by classic parametric models is possible, but it typically comes at the cost of analytic and numerical tractability.

Non-structural methods for estimating the RND directly from VIX options are a viable alternative to stochastic modeling. In general, the term "non-structural" refers to any option pricing method that does not postulate a fully parametric expression for the RND. This entails considerable reduction of the risk associated with misspecification. The idea that vanilla option prices can be linked to the RND through an explicit non-structural relation was introduced by Breeden and Litzenberger (1978). In the context of VIX option pricing, non-structural techniques based on robust computation of second-order derivatives have recently been employed by Song and Xiu (2016) and Völkert (2015). Specifically, using SPX and VIX options, Song and Xiu (2016) focus on the study of pricing kernels and their dependence on multiple volatility factors. Völkert (2015) carries out a similar analysis based only on the informational content provided by VIX options, with main focus on studying the predictive power of the RND for financial crises and the investors’ attitude towards risk.

In this paper we propose a non-structural approach to recover the RND by means of an orthogonal polynomial expansion around a kernel density, see for instance Szegő (1939) and Chihara (2011). Classic examples of orthogonal expansions are the Hermite, which are obtained when the kernel is a Gaussian density, and the Laguerre, which are obtained when the kernel is an exponential density. The key feature of orthogonal expansions is that they impose mild conditions on the form of the RND, proving to be particularly robust to misspecification. There is extensive literature on the use of orthogonal expansions in financial applications. Seminal examples are Jarrow and Rudd (1982), Corrado and Su (1996), Madan and Milne (1994), Coutant et al. (2001), and Jondeau and Rockinger (2001), while more recent contributions are Rompolis and Tzavalis (2008), Zhang et al. (2011), Ñíguez and Perote (2012), and Xiu (2014). In all these cases, the expansions are provided in terms of Hermite polynomials. Our methodology can be thought of as an alternative of Hermite expansions to the case
of kernels with positive support, which suits the case of volatility. We argue that adapting the expansion kernel to the data represents a better and more natural alternative to the adapting the data to the kernel (for instance by log-change or standardization).

The contribution of this paper is twofold. First, on the methodological side, we provide general convergence conditions of orthogonal expansions to the true RND. Based on these results, we extend the classic Laguerre expansions, used recently by Filipovic et al. (2013) and Mencia and Sentana (2017), by introducing a family of kernels that encompasses well known distributions such as the exponential, the Gamma, the Weibull and the Generalized Inverse Gaussian (GIG), among others. We show that the extended Laguerre kernels and the related expansions are particularly suited to approximate the RND of the VIX. The latter is characterized by peculiar features (e.g., upward sloping implied volatilities and a long right tail decaying slower than exponentially) which can only be captured through a flexible choice of the kernel density. In the spirit of Aït-Sahalia and Lo (1998), Jondeau and Rockinger (2001), and Bondarenko (2003), we estimate the parameters of the polynomial expansion using a minimum distance criterion based on the observed option prices. Our numerical experiments confirm that the methodology based on extended Laguerre polynomials is very effective and it proves more convenient than estimating second derivatives in the context of VIX options.

Second, on the empirical side, by estimating the RND at several points in time we recover time series of a number of relevant quantities embedded in the RND of the VIX, such as the VVIX, the variance swap (VS) term structure, the VIX risk-neutral higher moments and RND quantiles. Such quantities are used to carry out an extensive empirical investigation of the VIX under risk neutrality. Our analysis provides a novel contribution to the current literature as it unveils a number of stylized facts on the VIX under the risk-neutral measure, and it does so in a model-free fashion. Several studies have investigated the empirical features of the VIX but under the physical measure instead, see e.g. Becker et al. (2009) and Fernandes et al. (2014). Our findings are as follows. By investigating the VS term structure implicit in the VIX second moments, we find that it is coherent with the one directly computed from SPX options, suggesting that there are essentially no cross market arbitrage opportunities. Notably, this result entails that the second moment of VIX can be traded through a combination of long
and short positions on SPX options. The time series of the first four risk-neutral moments of VIX display interesting clues about volatility expectations. We find a common factor structure across moments and times-to-maturity, which is line with a multiplicative error model (MEM). In particular, fitting a MEM specification with jumps on the VIX risk-neutral moments reveals strong empirical evidence of a non-negligible jump term priced in the VIX options. Finally, by defining the conditional Volatility-at-Risk implied by the RND (CVolaR), we assess the expected excess VIX at given probability levels, which is of interest to volatility traders.

The paper is organized as follows. Section 2 recaps the definition of the VIX and defines its RND, while Section 3 summarizes some properties of orthogonal polynomial expansions. Section 3.4 discusses the estimation procedure. Section 4 provides numerical illustrations based on generated data to assess the accuracy of the estimation technique. Section 5 presents the empirical applications based on real VIX option data and Section 7 concludes. Appendix A contains the proof of the theorem presented in Section 3. Appendix B reports further information on the empirical application. The Supplementary document contains additional details on the fitting procedure and further empirical evidence.

2 The VIX index

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

$$V_{t,s} = -\frac{2}{s} E_t^Q \log \left( \frac{SPX_{t+s}}{XF} \right), \quad (1)$$
where $Q$ denotes the unique risk-neutral pricing measure, $\text{SPX}_{t+s}$ is the price of the S&P 500 index at time $t + s$ and $x_F$ is the associated forward price observed at time $t$. The relation in (1) is exact only under the assumption the SPX process follows continuous trajectories, while it is approximate in presence of jumps. However, as empirically observed in Carr and Wu (2009), the approximation error is negligible in practice. The right-hand term in (1) can be replicated by the following portfolio of SPX calls and puts

$$-\frac{2}{s} E_t \log \left( \frac{\text{SPX}_{t+s}}{x_F} \right) = \frac{2e^{r-s}}{s} \left( \int_0^{x_F} \frac{1}{x^2} p^{\text{SPX}}(x) dx + \int_{x_F}^{\infty} \frac{1}{x^2} C^{\text{SPX}}(x) dx \right),$$

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

$$\frac{\text{VIX}_t^2}{100} = \frac{2e^{r-s}}{s} \sum_{x_i \leq x_0} \frac{\Delta x_i}{x_i^2} p^{\text{SPX}}(x_i) + \sum_{x_i \geq x_0} \frac{\Delta x_i}{x_i^2} C^{\text{SPX}}(x_i) + \frac{1}{s} \frac{x_F}{x_0} - 1^2,$$

where $\Delta x_i$ is half the distance between $x_{i+1}$ and $x_{i-1}$, $x_0$ is the greatest available strike below $x_F$. In practice, when options with maturity exactly equal to $s$ are not available, the VIX is calculated by interpolating the values of (3) calculated by using SPX options with the largest available maturity below 30 days (near term) and the smallest available maturity above 30 days (next term). Under risk-neutrality and no-arbitrage assumptions, for fixed observation time $t$, time-to-maturity $\tau$, and strike $K$, the price of a VIX call option $C_K(t, \tau)$ and the price of a VIX put option $P_K(t, \tau)$ are given by

$$C_K(t, \tau) = e^{-r\tau} \int_0^{\infty} f_Q(t; x; x - K)^+ dx, \quad P_K(t, \tau) = e^{-r\tau} \int_0^{\infty} f_Q(t; x; K - x)^+ dx,$$

where $f_Q(t; x; \cdot)$ is the conditional RND of $\text{VIX}_{t+\tau}$, given all the past information up to time $t$. When not misleading, we will omit the dependence of the RND on $t, \tau$ and simply denote it by $f_Q$. In the next section, we discuss a robust methodology to retrieve $f_Q$ from the VIX options.
3 Orthogonal polynomials

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

\[ f_Q(x) \approx f_Q^{(n)}(x) := \phi(x) + \sum_{k=1}^{n} e_k(x), \quad n \geq 1, \]

(5)

where \( \phi \) is a chosen probability density function (kernel) and \( e_k \) are corrective factors with \( e_0 = 1 \). The kernel function \( \phi \) can be seen as the 0-order term in (5), i.e. \( f_Q^{(0)}(x) := \phi(x) \), and it can be interpreted as an initial proxy for the RND. In this work, we consider expansions where the corrective factors \( e_1, \ldots, e_n \) take the following form \( e_k(x) = c_k h_k^\phi(x) \), where, for every \( k = 1, \ldots, n \), \( c_k \) is a real constant and \( h_k^\phi \) is a polynomial of degree \( k \) in the state variable \( x \) depending only on the kernel \( \phi \). Hence, the coefficients \( c_1, \ldots, c_n \) contain all the information on the RND. The consistency of this approach, i.e., that \( f_Q^{(n)} \rightarrow f_Q \) as \( n \rightarrow \infty \) does not require postulating an explicit parametric form for the RND, see the discussion in Section 3.2. Thus the relation expressed in (5) is non-structural. The results presented in this section are obtained under the maintained assumption that \( \phi \) is a probability density function with support \( \mathcal{D} \subseteq \mathbb{R} \) and possessing finite polynomial moments, that is \( \int_{\mathcal{D}} |x|^k \phi(x) dx < +\infty, \forall k \in \mathbb{N} \).

3.1 Definition and properties of orthogonal polynomials

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

\[ \deg (h_k^\phi) = k \quad \text{and} \quad \int_{\mathcal{D}} h_k^\phi(x) h_j^\phi(x) \phi(x) dx = 0. \]

(6)

The existence of a family \( (h_k)_{k \in \mathbb{N}} \) of orthogonal polynomials with respect to \( \phi \) follows from the general properties of the theory of Hilbert spaces, see e.g. Rudin (1987). It can also be shown constructively, e.g., by applying the Gram-Schmidt orthogonalization to the basis \( 1, x, \ldots, x^n, \ldots. \). For fixed \( n \in \mathbb{N} \), \( h_1^\phi, \ldots, h_n^\phi \) are uniquely determined, up to a sign, if they
obey the normality condition
\[ \int_{D} h_k^\phi(x)^2 \phi(x) dx = 1. \] (7)

Henceforth, for a given kernel \( \phi \) with finite moments, we denote by \((h_k^\phi)_{k \in \mathbb{N}}\) the unique (up to a sign) family of orthogonal polynomials satisfying condition (7) for every \( k \in \mathbb{N} \). For the numerical computation of the basis \( h_1^\phi, \ldots, h_n^\phi \), the following well-known recurrence relation, see e.g. Chihara (2011), can be used as a more efficient alternative to the Gram-Schmidt procedure. Define
\[ h_0^\phi(x) = 1, \quad h_1^\phi(x) = \frac{x - M^\phi}{\sqrt{V^\phi}}, \]
where \( M^\phi \) and \( V^\phi \) are the mean and the variance of \( \phi \), respectively. The remaining terms, \( h_2^\phi, \ldots, h_n^\phi \), can be computed as
\[ h_k^\phi(x) = \frac{1}{C_k} \left( x - a_k \right) h_{k-1}^\phi(x) - b_k h_{k-2}^\phi(x), \]
where
\[ a_k = \int_{D} x h_{k-1}^\phi(x)^2 \phi(x)^2 dx, \quad b_k = \int_{D} x h_{k-1}^\phi(x) h_{k-2}^\phi(x) \phi(x)^2 dx, \]
\[ C_k = \left( \int_{D} (x - a_k) h_{k-1}^\phi(x) - b_k h_{k-2}^\phi(x) \phi(x)^2 dx \right)^{\frac{1}{2}}. \]

As shown in Filipovic et al. (2013), orthogonal polynomials fulfill a mass conservation principle that extends to all moments. More specifically, for fixed \( p \geq 0 \)
\[ \int_{D} x^p f_Q^{(p)}(x) dx = \int_{D} x^p f_Q^{(p)}(x) dx, \quad \forall n \geq p. \] (8)

Consequence of (8) is that, irrespective of the order \( n \), the approximated RND obtained by (5) always integrates to 1. However, in general, \( f_Q^{(n)} \) is not guaranteed to be a positive function over its support, even under the assumption that \( \phi \) is a positive function.

3.2 Admissibility of orthogonal expansions

In this section, we discuss the admissibility of expansion (5) for the RND. The following theorem is central for setting up a mathematically well-posed procedure to estimate the expansion coefficients \( c_1, \ldots, c_n \) based on observed option prices.

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

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

$$f_Q(x) = \lim_{n \to +\infty} \phi(x) \left( 1 + \sum_{k=1}^{k_n} c_k h^k(x) \right) \text{ for a.e. } x \in D;$$

(b) the following limit

$$\lim_{n \to +\infty} \int_0^{+\infty} \Pi(x) f_Q^{(n)}(x) dx = \int_0^{+\infty} \Pi(x) f_Q(x) dx,$$

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

Proof. See Appendix A.1. □

Point (a) in Theorem 3.1 provides sufficient conditions to ensure that the RND admits the representation (5) or, in other words, that $f_Q^{(n)}$ converges to $f_Q$ for some $c_1, \ldots, c_n, \ldots$. It is worth noting that, in some cases, we can recover true almost everywhere convergence, thus not only along a certain subsequence. For instance, under the hypotheses of Theorem 3.1 and in the special case $\phi(x) = \exp(-x)x^\alpha$, sharp conditions on $f_Q$ can be given in order for the orthogonal expansion (5) to converge almost everywhere, see Corollary 1 in Muckenhoupt (1970). Point (b) ensures that we can move the limit within the integral pricing formulas in (4) and obtain convergent expansions for prices of call and put options.

The hypotheses of Theorem 3.1 represent sharp conditions as they cannot be further relaxed. In particular, the condition on the decay rate of the kernel density are milder than those adopted in Filipovic et al. (2013), which only allows for Gamma-type decay. If $\phi^{-1}f_Q^2$ is not integrable (e.g. heavy-tailed RND), then the expansion (5) is not well-defined and it may diverge as $n \to \infty$. If $\phi^{-1}f_Q^2$ is integrable but $\lim_{x \to +\infty} \phi(x)e^{\frac{1}{2}\gamma x^2+\zeta} > 0$ for some $\gamma, \zeta > 0$, then $f_Q^{(n)}$ is well-defined in the sense that it will converge in the limit, but this limit is not necessarily $f_Q$ - see Appendix A.1, Theorem A.3-(ii). On the one hand, this means that the kernel should not decay faster than the RND to ensure convergent expansions. On the other
hand, the expansion (5) may not be flexible enough if the kernel density decays too slowly (e.g. lognormal-type decay). As a matter of fact, the vast majority of parametric models used for VIX option pricing are associated with RNDs fulfilling the hypotheses of Theorem 3.1. For example, a wide class of affine models for volatility satisfies the theoretical conditions ensuring convergence of Laguerre-like expansions to the true RND, see Filipovic et al. (2013). By adding flexibility to the decay rate on right tail, Theorem 3.1 allows for an even wider class of densities to be represented by orthogonal polynomial expansions. Examples of extended Laguerre expansions are discussed in the next section.

3.3 The extended Laguerre expansions

We now illustrate the properties of a class of kernel densities that can form the basis for the polynomial expansions used to retrieve the RND of VIX. Based on restrictions imposed by Theorem 3.1, we propose the following family of kernels with support on \( D = [0, +\infty[ \),

\[ \phi(x) \propto x^{\alpha-1}e^{-(\beta x^p+\xi x^{-1})}1_D(x), \quad \alpha, \beta, \xi, p \in \Theta, \]

where \( \Theta = \{\alpha, \beta, \xi, p \in \mathbb{R} \mid \beta > 0, 0 < p \leq 1, (\alpha > 0, \xi = 0) \lor (\alpha \in \mathbb{R}, \xi > 0)\} \). The specification (10) 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. Indeed, the tail behavior of the GIG and GW kernels characterizes their ability to meet the condition \( \phi^{-\frac{1}{2}}f \in L^2(D) \), required by Theorem 3.1. Looking at the left tail, the GW kernel clearly has the slowest decay, and thus it is the most flexible in terms of behavior around the origin. However, Theorem 3.1 also requires that \( \lim_{x \to +\infty} \phi(x)e^{\xi x^{-\frac{1}{2}}} = 0 \) for some \( \zeta > 0 \). This condition is met by the GW kernel when \( p \) is restricted between \([\frac{1}{2}, 1]\) and it is always met by the GIG kernel.

The orthogonal expansions arising from (10) extend the classic Laguerre expansions that are associated with a Gamma kernel. For this reason, we refer to these expansions as extended Laguerre. This family proves effective in reproducing the peculiar tail behavior of the RND implied by VIX options. As alternative to the extended Laguerre kernel family, one could
consider a log-normal (LN) kernel. This choice is endorsed by documented empirical evidence that the volatility, a quantity comparable to the VIX, is roughly log-normally distributed, see e.g. Christoffersen et al. (2010), Wang and Daigler (2011) and Bayer et al. (2013). However, the condition \( \lim_{x \to +\infty} \phi(x) e^{cx^2} = 0 \) for some \( \zeta > 0 \) is never met by the LN kernel. Thus, the LN kernel does not guarantee that the RND is fully recovered by the expansion (10). This means that using the LN kernel inherently entails further restrictions on the form of \( f_Q \), and it may reduce the "non-structural" nature of the approach.

3.4 Implementation

In this section, we outline a procedure to estimate the coefficients \( c_1, \ldots, c_n \) of the expansion (5) by minimizing the distance between the option prices implied by the RND and the observed ones. The fitting procedure detailed below can be easily implemented with the MATLAB App *rndfittool*, see Barletta and Santucci de Magistris (2018) for a tutorial. The key feature of the proposed procedure is to estimate the option pricing formulas in (4) based on the observed market prices by choosing a sufficiently large \( n \) in the expansion (5), where the coefficients \( c_1, \ldots, c_n \) are the unknown terms that convey the information about \( f_Q \). The consistency of this procedure relies on Theorem 3.1. For a fixed strike \( K \geq 0 \), \( n \in \mathbb{N} \), and \( c = [c_1, \ldots, c_n]^T \in \mathbb{R}^n \), the prices associated with the expansion of order \( n \) are defined as

\[
C_K^{(n)}(c) := \int_K^{+\infty} \phi(x) (x - K) dx, \quad P_K^{(n)}(c) := \int_0^K \phi(x) (K - x) dx.
\]

The expressions above can be rewritten in the following compact form

\[
C_K^{(n)}(c) = A^{(K)}_0 + A^{(K)} c, \quad P_K^{(n)}(c) = B^{(K)}_0 + B^{(K)} c \tag{11}
\]

where \( A^{(K)}_0 = \int_K^{+\infty} \phi(x) (x - K) dx \), \( B^{(K)}_0 = \int_0^K \phi(x) (K - x) dx \), and \( A^{(K)} \) and \( B^{(K)} \) are \( 1 \times n \) vectors, whose \( i \)-th element is given by

\[
A^{(K)}_i = \sum_{j=0}^{i} w_{i,j} \int_K^{+\infty} (x^{j+1} - Kx^j) \phi(x) dx, \quad B^{(K)}_i = \sum_{j=0}^{i} w_{i,j} \int_0^K (Kx^j - x^{j+1}) \phi(x) dx. \tag{12}
\]
In general, even when $\phi$ is given, the integrals in $A_0^{(K)}, B_0^{(K)}, A^{(K)}$ and $B^{(K)}$ cannot be reduced to any explicit form and thus require numerical approximations. However, they are semi-explicitly computable in a few particular cases. For example, when $\phi(x) \propto x^{a-1}e^{-\beta x^p}$, they can be expressed in terms of incomplete Gamma functions. For given $\phi$ and $n$, one can estimate $c_1, \ldots, c_n$ by collecting a cross-section of market prices, $C_{K_m}^{\text{Obs}}(t, \tau)$ and $P_{K_m}^{\text{Obs}}(t, \tau)$, for $m = 1, \ldots, M$, and by finding the solution $\hat{c} = [\hat{c}_1, \ldots, \hat{c}_n]^\top$ of the following optimization problem

$$
\hat{c} = \hat{c}(t, \tau) = \arg\min_{c \in \mathbb{R}^n} Q(t, \tau; c),
$$

(13)

where $Q(t, \tau; \cdot)$ defines an objective function to be minimized. Given that the expressions in (11) are linear in the coefficients $c$, a natural choice for $Q(t, \tau; \cdot)$ is the criterion function of the minimum least squares problem for the following linear model

$$
Y = X_0 + Xc + \varepsilon,
$$

(14)

where $Y = [C_{K_1}^{\text{Obs}}(t, \tau), \ldots, C_{K_M}^{\text{Obs}}(t, \tau), P_{K_1}^{\text{Obs}}(t, \tau), \ldots, P_{K_M}^{\text{Obs}}(t, \tau)]'$ is the vector with the observed option prices, $X_0 = [A_0^{(K_1)}, \ldots, A_0^{(K_M)}, B_0^{(K_1)}, \ldots, B_0^{(K_M)}]'$ is the price vector of option prices generated by the kernel. The $2M \times n$ matrix of the regressors, $X$, is such that its $j$-th column, denoted as $X_j$, is $X_j = [A_j^{(K_1)}, \ldots, A_j^{(K_M)}, B_j^{(K_1)}, \ldots, B_j^{(K_M)}]'$. The $2M \times 1$ vector $\varepsilon$ represents the error term. The objective function therefore takes the following quadratic form

$$
Q(t, \tau; c) = (Y^* - Xc)'(Y^* - Xc),
$$

(15)

where $Y^* = Y - X_0$. Denoting by $\tilde{c} = \tilde{c}(t, \tau)$ the vector of coefficients minimizing the objective function, the estimated RND function $\tilde{f}_Q^{(n)}$ is determined as

$$
\tilde{f}_Q^{(n)}(x) = \phi(x; \theta) + \sum_{k=1}^{n} \tilde{c}_k h_k^{(q)}(x).
$$

(16)

Note that the kernel $\phi$ in (16) is now expressed as a function of an additional term, to highlight its dependence on a set of parameters $\theta \in \Theta$. For example, $\theta = [\alpha, \beta, \xi]'$ for the GIG kernel,
\( \theta = [\alpha, \beta, p]' \) for the GW kernel, and \( \theta = [\mu, \sigma^2] \) for the LN kernel. As shown in Section 4, the choice of \( \theta \) does not play a crucial role in practice, provided that \( n \) is sufficiently large and the hypothesis on the decay rate of the kernel is not violated as prescribed by Theorem 3.1. A heuristic criterion to select \( n \) is discussed in Section 6. In our empirical applications, we set \( \theta \) to the value that minimizes the residuals variance for the expansion of order 0. The minimization is performed under the restriction of zero-mean residuals, which implies absence of systematic pricing errors. Furthermore, we deal with possible multicollinearity in the columns of \( X \) by principal components analysis (PCA) and we introduce additional constraints in the objective function to ensure the positivity of \( \tilde{f}_Q^{(n)} \). These and other technical details about the determination of \( \tilde{c} \) are provided in Section 1 of the Supplementary document.

4 Numerical validation

We now assess the accuracy of the methodology presented above and compare it to other existing methods. This is addressed by means of numerical experiments. In particular, we determine finite sets of call and put prices from known parametric densities playing the role of the true RND, \( f_Q \), and then estimate \( \tilde{f}_Q^{(n)} \) as described in Section 3.4. We consider multiple realistic RND parameter combinations, each of them obtained by calibrating \( f_Q \) to the sample of 1-month VIX options and strikes used in Section 5. We compare the accuracy of orthogonal polynomial expansions with the spanning formula of Carr and Madan (1998) and Bakshi et al. (2003) in estimating the mean, variance, skewness and kurtosis of the true RND. We also make a direct comparison of the RND estimates provided by orthogonal polynomials with those obtained through numerical evaluation of the second-order derivatives of the interpolated call prices as in Breeden and Litzenberger (1978).

Beyond the baseline implementation of the Carr-Madan formula by midpoint quadrature scheme based on available options, we consider an improved version obtained by refining the integration domain through interpolation. In particular, following the approach popularized by Shimko (1993), we apply cubic spline interpolation to the available implied volatilities and then obtain interpolated call prices via the Black-Scholes formula. Typically when dealing
with noisy option data, spline interpolation is coupled with smoothing. In the illustrations below smoothing is unnecessary, since option prices enjoy smoothness by construction as they are exactly generated out of known densities. The effects and the benefits of combining smoothing with spline interpolation are investigated using real data in Section 6. We employ the endpoints spline coefficients obtained under "not-a-knot" conditions to extrapolate information on the RND beyond the range of available strikes. To minimize the sensitivity of the results to the choice of the range of extrapolated strikes, the true RND and the associated estimates based on orthogonal expansions are truncated to zero outside this interval.

The choice of the kernel is a practical issue with orthogonal expansions. Theorem 3.1 ensures that, in the limit of "infinite expansion order", the functional form of the kernel and its specific parametrization will not affect the validity of expansion (5) – provided that the kernel is consistent with the hypotheses of the theorem. However, for finite expansion orders, the choice of the kernel might have an impact on the accuracy of estimated RND, $\hat{f}_Q^{(n)}$. To assess the robustness to the choice of the kernel, we consider different functional forms of the kernel, namely the GIG and GW. Furthermore, in all cases we choose random values for the kernel parameters, so that the moments implied by the kernel exhibit large deviations from the true ones. In our first example we assume that the true unknown RND is a Cox et al. (1985) density (CIR), that is

$$f_Q(x) = Ce^{-C(u-x)} \left(\frac{x}{u}\right)^{\frac{q}{2}} I_q(2C\sqrt{ux}) , \quad C = \frac{2\kappa}{\bar{e}^2(1-e^{-\kappa})} , \quad u = \bar{e}^{-\kappa} , \quad q = \frac{2\kappa\bar{e}}{\bar{e}^2 - 1} ,$$

where $I_q(\cdot)$ denotes the modified Bessel function of the first kind of order $q$. The CIR density features exponential tail behavior and therefore admits orthogonal polynomial expansions associated with both the GW and the GIG kernel, see Theorem 3.1.

Table 1 reports the mean absolute percentage errors (MAPEs) on the CIR-implied moments. The MAPE is computed as

$$\text{MAPE} = \frac{100}{H} \sum_{i=1}^{H} \left| \frac{M_i}{\hat{M}_i} - 1 \right| ,$$

where $M_i$ is the true moment (mean, variance, skewness and kurtosis) and $\hat{M}_i$ is the associated
estimate. For each moment, the average is computed over $H = 76$ values obtained for different CIR densities calibrated to the sample of 1-month VIX options collected in the period 2011-2016. The extended extrapolation range used in the spline interpolation is set to $[0, 300]$ in all cases. Table 1 indicates that a straightforward implementation of the Carr-Madan formula is enough to obtain reliable – albeit not perfect – estimates of the mean, variance, skewness and kurtosis of the CIR process. Estimates based on orthogonal expansions are extremely accurate, irrespective of the functional form of the kernel density and its parameter specification. In this case, the Carr-Madan formula improved by spline interpolation performs even better, yielding nearly error-less estimates of all moments. The robustness of orthogonal expansions to the choice of the kernel is highlighted in the left panel of Figure 1, reporting extremely high absolute percentage errors (APEs) associated with the expansion of order 0. For both kernels

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Variance</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carr-Madan (baseline)</td>
<td>0.00%</td>
<td>0.49%</td>
<td>3.03%</td>
<td>4.33%</td>
</tr>
<tr>
<td>Carr-Madan (spline)</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.04%</td>
<td>0.02%</td>
</tr>
<tr>
<td>GIG-expansion</td>
<td>0.00%</td>
<td>0.04%</td>
<td>0.14%</td>
<td>0.28%</td>
</tr>
<tr>
<td>GW-expansion</td>
<td>0.00%</td>
<td>0.13%</td>
<td>0.50%</td>
<td>1.06%</td>
</tr>
</tbody>
</table>

Table 1: Mean absolute percentage error (MAPE) on the CIR-implied moments.

Figure 1: Absolute percentage errors on CIR option-implied RNDs recovered by different techniques. We compare the true and the estimated RNDs under the $L^2$ norm. The absolute error is determined as the square root of the integral of the square difference between true and the estimated RND. The APE is obtained by normalizing the absolute error by the square root of the integral of $f_Q$. The lefth panel reports the errors associated with the GIG and the GW kernels ($n = 0$). The right plot reports errors associated with the GIG-expansion, the GW-expansion and the cubic spline interpolation method.
the correction provided by orthogonal polynomials leads to large reduction in the magnitude of the errors, making it always comparable with that associated with spline interpolation.

The CIR density provides a theoretical framework where both the GIG and the GW kernels can be compared in light of Theorem 3.1. Unfortunately, the CIR density is not a realistic choice for the RND of the VIX because it delivers downward sloping implied volatilities, which are inconsistent with the evidence provided by the market. We therefore carry out an additional experiment where a more realistic functional form for the RND is considered. Specifically, we

![Figure 2](image_url)

**Figure 2**: Comparison of Kappa risk-neutral moments estimated with different techniques. The plots report the mean, variance, skewness and kurtosis computed with: the baseline implementation of the Carr-Madan formula (blue downward triangles), the Carr-Madan formula improved by spline interpolation (green upward triangles), the GW kernel with misspecified parameters (green filled circles) and the related orthogonal expansion (red empty circles). The plots also report the true moments of the Kappa densities (solid black line). The x-axis reports the dates available in the original sample of VIX options.
assume that the true unknown RND is now a Kappa density, that is
\[ f_Q(x) \sim x^{\beta-1} K_{\nu-L} 2 \sqrt{\frac{L \nu}{\mu}} x^{-\frac{\nu-L}{2}}, \]

where \( K_{\nu-L} \) denotes the modified Bessel function of the second kind of order \( \nu-L \). The right tail of the Kappa density decays slower than exponentially and, as shown in Caporin et al. (2017), it better represents the relatively high probability of extreme events in volatility. The effect of this slow tail decay is reflected by upward sloping implied volatilities. The GW kernel has the same tail behavior of the Kappa density and therefore fulfills the hypotheses of Theorem 3.1. Conversely, the GIG kernel is not consistent with the theory and it is ruled out in this case.

Figure 2 reports the first four risk-neutral moments (mean, variance, skewness and kurtosis) retrieved by the Carr-Madan formulas and with the GW expansion together with the true ones based on the Kappa density. The visual inspection of the figure highlights that using a discrete quadrature scheme over finite sets of available strikes leads to fairly accurate approximations of mean and variance, but not of the higher moments. Refining and extending the grid of available points through cubic spline interpolation/extrapolation leads to a neat improvement over the baseline scheme but, especially in estimating the kurtosis, this direct approach is still outperformed by the GW-expansion. Notably, here the choice of the extended strike range used for spline interpolation/extrapolation is more troublesome than in case of the CIR and the extrapolated values are now less accurate. To ensure numerical stability, the extrapolation range is set proportional to the available strikes as \([0.75K_1, 1.5K_M]\). These conclusions are supported by the MAPEs in Table 2.

<table>
<thead>
<tr>
<th>Mean Variance Skewness Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carr-Madan (baseline) 0.00% 1.16% 10.72% 30.22%</td>
</tr>
<tr>
<td>Carr-Madan (spline) 0.00% 0.33% 1.10% 6.21%</td>
</tr>
<tr>
<td>GW Kernel 5.04% 15.44% 58.16% 59.55%</td>
</tr>
<tr>
<td>GW-Expansion 0.00% 0.09% 1.07% 2.07%</td>
</tr>
</tbody>
</table>

**Table 2:** Mean absolute percentage error (MAPE) on the Kappa-implied moments. The true Kappa densities are calibrated to the sample of 1-month VIX options collected in the period 2011-2016.

Finally, Figure 3 reports the APEs associated with the GW kernels and related expansion...
together with the RND estimates obtained by spline interpolation. The latter is systematically outperformed by the GW orthogonal polynomial expansion. Moreover, the orthogonal polynomial expansion proves particularly robust to the choice of the kernel parameters. Indeed, the APEs in the estimation of the RND are reduced on average by nearly ten times.

![Figure 3](image_url)

**Figure 3:** Absolute percentage errors (APEs) on option-implied RNDs recovered by different techniques. We compare the true and the estimated RNDs under the $L^2$ norm. The absolute error is determined as the square root of the integral of the square difference between true and the estimated RND. The APE is obtained by normalizing the absolute error by the square root of the integral of $f_Q$. The figure reports the errors associated with the GW kernel, the GW-expansion and the cubic spline interpolation method. The true RND is the Kappa density for different parameters calibrated to the 1-month VIX options in the period 2011-2016.

The results of the numerical experiments carried out in this section can be summarized as follows. First, they confirm the validity of Theorem 3.1 and show its importance in the empirical implementation of the method. Second, methods based on spline interpolation work very well on options generated by relatively thin-tailed RNDs, such as the CIR density, while its level of accuracy significantly deteriorates when the underlying RND features a thicker right tail, such as the Kappa. The conclusion is that, in the context of VIX options, the expansion based on the GW kernel is expected to be the most flexible in terms of accommodating different tail behaviors among all the techniques considered in this section. For these reasons, the GW will be the chosen kernel in the analysis reported in Section 5.

## 5 Recovering the RND from VIX options

In this section we discuss the goodness-of-fit of the RND on a panel of VIX options observed in the period from January 2010 to April 2016, sampled at monthly frequencies, and with
time-to-maturity, \( \tau \), ranging from 1 to 5 months. The data is obtained from the OptionMetrics database. Weekly VIX options and options with time-to-maturity \( \tau = 6 \) are also traded, but unfortunately they are not available for all dates in the sample. Following the CBOE calendar, for each month in the sample, we collect option prices observed on the first Tuesday following the third Friday of the month, the latter being the option expiration date. This ensures that observations will not overlap in the base monthly frequency, meaning that 1-month options observed at time \( t \) always expire prior to the subsequent observation occurring at time \( t + 1 \).

We operate minimum pre-filtering of the data. More specifically, we exclude all options with zero bid quote, for which the Black and Scholes implied volatility can not be computed and those that violate upper and lower no-arbitrage bounds. We also exclude OTM puts (calls) with mid-quote below $0.025 together with the corresponding ITM calls (puts) with the same strikes. These contracts turn out to be highly illiquid (if traded at all) and therefore they are likely subject to mispricing. With this filtering criterion, we have an average total number of 56 available contracts for each date and time-to-maturity. Under normal market conditions, the strike values that are taken into consideration typically fall between $10 and $45, with this range remaining quite stable over time and maturities due to the mean-reverting behavior of the VIX. However, the interval of available strikes enlarges during turmoil periods, with maximum values reaching $100. In the following, we focus on a specific date and maturity to illustrate the effectiveness of the orthogonal polynomials in correcting \( \phi \) and the robustness of the methodology to the choice of the kernel.

To assess the precision of orthogonal polynomial expansions, we analyze the error on the option prices associated with the fitted RND. This is the error term \( \varepsilon \) appearing in the linear model (14), where we inherently assume that \( \varepsilon \) subsumes all the uncertainty associated with the fact that the polynomial expansion is truncated to a finite \( n \), that a strikes are available in a finite number \( M \) and that market prices are subject to no-arbitrage violations, which are function of the market liquidity. Particular attention should be paid to the latter source of error as it implies that, even in absence of discretization/truncation errors, it is not possible to achieve exact matching of all the observed option prices by solving (13). This is a direct consequence of the well-known fact that option prices obtained through any probability density function
are free of (static) arbitrages by construction. In light of this, it is natural to set a tolerance on
the variability of the fitting residuals $\tilde{e} = Y^* - X\hat{c}$, which quantifies the magnitude of arbitrage
opportunities. Given a fixed threshold $\Delta^Q > 0$, we define “admissible” any RND implying
option prices whose distance from the observed ones is below $\Delta^Q$. The error norm that we
adopt is the root mean square error (RMSE), which is defined as

$$\text{RMSE}_{t, \tau} = \left\{ \frac{1}{M} \sum_{m=1}^{M} C_{K_m}^{\text{obs}}(t, \tau) - C_{K_m}^{(n)}(\tilde{c}(t, \tau)) \right\}^2 + \frac{1}{M} \sum_{m=1}^{M} P_{K_m}^{\text{obs}}(t, \tau) - P_{K_m}^{(n)}(\tilde{c}(t, \tau)) \right\}^2 \right\}^{\frac{1}{2}} .$$

(17)

In particular, we regard the estimated RND, $f_Q^{(n)}$, as admissible if $\text{RMSE}_{t, \tau} \leq \sqrt{\Delta^Q}$. A lower
bound for the set of all possible values of $\Delta^Q$ can be directly inferred from the violations of
the put-call parity. Specifically, the following inequality must hold

$$\text{RMSE}_{t, \tau}^2 \geq \Delta_{\text{pcp}} - \left( \text{Mean}^{(n)}(\tilde{c}(t, \tau)) - \text{Mean}^{\text{obs}}(t, \tau) \right)^2 ,$$

where $\text{Mean}^{\text{obs}}(t, \tau) = \frac{1}{M} \sum_{m=1}^{M} \left( C_{K_m}^{\text{obs}}(t, \tau) - P_{K_m}^{\text{obs}}(t, \tau) + K_m \right)$, $\text{Mean}^{(n)}(\tilde{c}(t, \tau))$ denotes the
mean of $f_Q^{(n)}$, and

$$\Delta_{\text{pcp}} = \frac{1}{M} \sum_{m=1}^{M} \left( C_{K_m}^{\text{obs}}(t, \tau) - P_{K_m}^{\text{obs}}(t, \tau) + K_m - \text{Mean}^{\text{obs}}(t, \tau) \right)^2 .$$

(18)

is the variance of the put-call parity violations. In other words, $\Delta_{\text{pcp}}$ approximately repre-
sents the maximum precision that can be achieved by an estimate of the RND guaranteeing
arbitrage-free option prices. As a consequence, a RMSE below this threshold might be associ-
ated with overfitting. Thus, we choose the expansion order $n$ as the lowest integer such that
$f_Q^{(n)}$ is an admissible RND estimate given the tolerance $\Delta_{\text{pcp}}$, which is an observable quantity.

We start our analysis by illustrating the pricing performance of orthogonal polynomial
expansions on a fixed day, i.e., 16 November 2011 and for a fixed maturity, i.e., 21 December
2011 ($\tau = 1$). The goal is to provide an example of the goodness-of-fit typically obtained with
orthogonal polynomial expansions. The cross-section of VIX options quoted by the CBOE
on 16 November 2011 consists of 64 contracts. Notably, this period is characterized by high
levels of market volatility (the VIX reached a value of 33.51%) leading VIX options to trade
at unusually high strike values (up to $90). This suggests an uncommonly long right-tail in the RND. We consider both the GIG and the GW kernels and we set the expansion order to a relatively high value, i.e., $n = 18$. We find that 5 and 6 principal components explain the

Figure 4: Estimated RNDs. The left panel depicts the graphs of GW and the GIG kernels, together with the corresponding orthogonal polynomial expansion. The right panel reports the same contents of the left panel in semi-log scale, to highlight tail features. Each pair of percentage values denotes, from top to bottom, the average mass levels of the two kernels and their corresponding orthogonal polynomial expansion, associated with the quantile identified by the dashed vertical lines.

99% of the total variance of $X$ for the GW and the GIG kernel, respectively. The plots of the estimated RNDs are reported in Figure 4 (solid lines), together with the corresponding kernels (dashed lines). From a visual inspection of Figure 4, it emerges that the RND estimates and the associated kernels display remarkable differences on the right tails, more specifically in a part of the domain above the 98% quantile. The importance of the correction provided by the orthogonal polynomials is better understood by comparing the implied volatility curves generated by the two kernels and their corresponding expansions, reported in Figure 5. The implied volatilities generated by both kernels are considerably different from those generated by market prices, while the expansions are able to produce implied volatilities that closely replicate the observed ones. Notably, the fitting of the implied volatilities is equally good for any moneyness region, which suggests that no weighting scheme is required in the least squares optimization. All these observations point in the same direction: neither of the two baseline kernels is able to reproduce the tail-features of the RND of VIX. But, coherently with Theorem 3.1, they lead to expansions converging to the same density.
Figure 5: Black and Scholes implied volatility curves obtained from market prices, GIG kernel, GW kernel and the resulting orthogonal polynomial expansion.

The analysis carried out for 11 November 2011 is replicated for all dates and maturities in the sample, adopting the GW kernel. This choice is endorsed by Theorem 3.1 and by the numerical illustrations reported in Section 4, which prove that the GW can accommodate a wider range of right tail behaviors of the true RND than the GIG kernel. The time series of the RMSE obtained for $\tau = 1$ and $\tau = 5$ are reported in Figure 6. As expected, the orthogonal expansion (blue line) outperforms the GW kernel (red line) by a large extent. Furthermore, in all cases, the RMSE associated with the orthogonal polynomial expansion lies around the market implied threshold $\sqrt{\Delta_{\text{pcp}}}$, which is a function of the magnitude of the noise. This evidence signals that the GW polynomial expansion is never associated with overfitting of the noisy option prices, while systematically leading to a near-to-maximum level of precision.\footnote{Similar results are obtained with the other maturities, see Appendix B for details on the dataset of VIX options and for further evidence on the goodness-of-fit.}

Since the results presented in the Section 6 involve the analysis of the risk-neutral moments, jumps and quantiles of VIX, it is important that values of the RND lying on the right-tail of the RND are robustly extrapolated. Therefore, to further assess the robustness of orthogonal polynomial expansions to well approximate options outside the range of available strikes, we perform a comparison with an interpolation-based technique widely adopted in the literature. In contrast to orthogonal polynomial expansions, non-structural techniques based on the raw interpolation of the observed data do not typically enjoy robustness to no-arbitrage
violations. For example, applying cubic spline interpolation to noisy implied volatilities will harshly affect the estimate of the RND through second-order derivatives of the interpolated price curve. Therefore, we combine cubic spline interpolation with some form of "smoothing" to make the RND estimate robust to noise, as proposed by Malz (1997), Bliss and Panigirtzoglou (2002) and, more recently, in Völker (2015) in the context of VIX options. This class of methods is typically referred to as cubic smooth spline interpolation (SSI). The adopted SSI method is the same as in Bliss and Panigirtzoglou (2002) and Völker (2015), the only difference being that option prices are not transformed into the so-called "delta domain". This technique relies on the choice of a smoothness parameter, which we operate by resorting to the algorithm of Reinsch (1967). Reinsch’s method provides an elegant way to achieve direct control of the goodness-of-fit through the smoothing parameter. The smoothing parameter is chosen to guarantee the smoothest interpolation of the observed implied volatilities for a pre-determined level of goodness-of-fit, the latter measured by means of the RMSE with respect to the observed option prices. This allows us to control the trade-off between pricing accuracy and smoothness of the RND. For a fair comparison we match the RMSE of SSI with that of orthogonal polynomials, where the latter is set to the market implied threshold $\sqrt{\Delta^{\text{pcp}}}$.

Once the level of pricing accuracy is fixed, we can compare the SSI and the orthogonal

---

2See Figure 13 in Appendix B.
polynomial expansions in terms of degree of smoothness of the extracted RND. This exercise complements the analysis carried out in Section 4 based on model-implied and arbitrage-free prices. The degree of smoothness of the estimated RND is measured by its total curvature which is a standard measure in numerical analysis, see Ramsay (2005). The total curvature $TC_\alpha(f)$ over $[\alpha, +\infty]$ of a second-order differentiable function $f$ is defined as $TC_\alpha(f) = \int_\alpha^{+\infty} |f''(z)|^2 dz$. For ease of interpretation, we used a normalized version of the total curvature. Specifically, given a probability density function with mean $\mu$, we define its "roughness" over $[\mu q, +\infty]$ as

$$R_q(f) = 1 + \frac{1}{2q} \log 2\mu^5 \cdot TC(q\mu) .$$

(19)

It is straightforward to verify that $R_q(f) = 0$ for all $q \geq 0$ when $f$ is an exponential density with mean $\mu$. Thus, the functional defined in (19) provides a relative measure of roughness using the exponential density as benchmark. In the present context, the parameter $\mu$ is the VIX futures while $q$ is the moneyness of the underlying VIX options.

We focus on the right tail behavior of the RND, which turns out to be of crucial importance when estimating high order moments and investigating volatility extreme quantiles as done in Section 6. Hence, we set four thresholds of moneyness starting from the at-the-money, i.e. $q = 1, 1.15, 1.25, 1.5$ with the purpose of moving progressively deeper into the right tail. For completeness, we consider both a short horizon (i.e. one month, $\tau = 1$) and a long horizon (i.e. five months, $\tau = 5$). Table 3 reports sample averages of $R_q(f_i)$ and the corresponding $t$-statistic under the null hypothesis of equal degree of smoothness. When the roughness measure is computed at $q = 1$ and $\tau = 1$, we do not find significant systematic differences between the two methodologies. The hypothesis that both the orthogonal polynomial expansion and SSI share on average equal degree of smoothness cannot be rejected at 5% significance level. This is not surprising because the at-the-money threshold $q = 1$ gives inherently a large weight on the central portion of the distribution whose estimation does not represent a challenge for either approach. As we move deeper into the right tail, i.e. we increase the threshold $q$, the differences between the two methodologies emerge and become more pronounced. When $q = 1.15$ the average roughness differential becomes three times as large as for $q = 1$, with the right tail of the RNDs estimated by means of orthogonal polynomial expansion resulting
Table 3: Sample averages of $R_q(f_i) = \text{OPE, SSI}$ for for the RND estimated by means of orthogonal polynomial expansion and SSI, denoted $R_q(f_i) = \text{OPE, SSI}$ respectively, average roughness differential $d = R_q(f_{\text{OPE}}) - R_q(f_{\text{SSI}})$ and corresponding $t$-statistics, where $t = \sqrt{T}d/\sqrt{\text{Var}(d)} - (T - 1)$, under the null hypothesis of equal degree of smoothness. Results are reported for different moneyness thresholds $q$ and horizons $\tau$. Newey-West standard errors are reported in parentheses.

<table>
<thead>
<tr>
<th>$q$</th>
<th>$R_q(f_{\text{OPE}})$</th>
<th>$R_q(f_{\text{SSI}})$</th>
<th>$\Delta$</th>
<th>$t$-stat.</th>
<th>$R_q(f_{\text{OPE}})$</th>
<th>$R_q(f_{\text{SSI}})$</th>
<th>$\Delta$</th>
<th>$t$-stat.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4.016 (0.063)</td>
<td>4.065 (0.056)</td>
<td>-0.049</td>
<td>-1.925</td>
<td>2.580 (0.025)</td>
<td>2.861 (0.041)</td>
<td>-0.281</td>
<td>-8.405</td>
</tr>
<tr>
<td>1.15</td>
<td>3.082 (0.036)</td>
<td>3.203 (0.034)</td>
<td>-0.120</td>
<td>-5.205</td>
<td>2.062 (0.020)</td>
<td>2.365 (0.029)</td>
<td>-0.303</td>
<td>-11.065</td>
</tr>
<tr>
<td>1.25</td>
<td>2.428 (0.026)</td>
<td>2.671 (0.039)</td>
<td>-0.243</td>
<td>-6.781</td>
<td>1.729 (0.017)</td>
<td>2.012 (0.027)</td>
<td>-0.283</td>
<td>-10.660</td>
</tr>
<tr>
<td>1.5</td>
<td>0.942 (0.045)</td>
<td>1.779 (0.066)</td>
<td>-0.837</td>
<td>-11.975</td>
<td>1.014 (0.013)</td>
<td>1.334 (0.035)</td>
<td>-0.320</td>
<td>-8.935</td>
</tr>
</tbody>
</table>

significantly smoother than their SSI counterpart. Moving even further on the right tail, i.e. setting $q = 1.25$ and $1.5$, the average degree of smoothness of the SSI relative to the orthogonal polynomial expansion substantially worsens. We argue that this result is mostly due to the SSI struggling in those areas of the domain of the RND where strikes are sparse and the underlying option prices are more noisy. On the contrary, the RNDs estimated by means of the orthogonal polynomial expansion proves less sensitive to these issues while assuring the same level of pricing accuracy thanks to the structure on the tail decay implicit in GW kernel consistently with Theorem 3.1. Similar considerations still hold and become even more striking as we move along the horizon dimension by looking at $\tau = 5$. Summing up the SSI proves more sensitive than orthogonal polynomials to sparse and less informative prices, and this is amplified as the time-to-maturity increases.

6 The RND of VIX

Supported by the empirical results outlined in the previous section, we now look at the characteristics of the RND of VIX retrieved by means of the orthogonal polynomial expansions. In particular, we study the stylized facts of the RND of VIX along several dimensions: the coherence with the options on the SPX in terms of VVIX and variance swap term structure,
the dynamics of the VIX risk-neutral moments, the presence of significant jumps and the tail risk as measured by the volatility-at-risk.

6.1 Coherence across markets

To further assess the consistency of the estimated RND with market data, we can look at the VVIX, defined by CBOE as the volatility-of-volatility index inferred from VIX options through the same algorithm used for the VIX itself. Therefore, the VVIX can be linked to the RND of the VIX through the following formula, which is analogous to (1)

\[
\frac{\text{VVIX}_{t, \tau}^2}{100} = -\frac{2}{\tau} \int_0^{+\infty} \log \left( \frac{x}{\text{Mean}_{t, \tau}^Q} \right) f_{Q,t}(t, \tau; x) \, dx,
\]

(20)

where \(\text{Mean}_{t, \tau}^Q := \mathbb{E}_{t}^Q[\text{VIX}_{t+\tau}]\) denotes the VIX futures at time \(t\) for maturity \(t + \tau\). By construction, the RHS of (20) is a non-negative quantity since the logarithm of the VIX futures is a super-martingale. Figure 7 reports the observed time series of the VVIX and the synthetic version computed by (20). The plots confirm that the estimated RND is generally associated with consistent estimates of the VVIX. Discrepancies are small in size and can be attributed to discretization errors affecting the CBOE formula for the computation of the VVIX. We also

![Figure 7: VVIX time series. Comparison between observed values of the VVIX and those obtained by formula (20), for \(\tau = 1\) and \(\tau = 5\) months.](image)

analyze the term structure of the (annualized percentage) realized variance (RV) in terms of variance swaps (VS). The RV is a non-parametric estimator of the (annualized percentage)
quadratic variation of the SPX log-returns, see Barndorff-Nielsen and Shephard (2002). As discussed in Section 2, the square of VIX can be seen as the swap rate of a VS maturing in one month, that is \( VST_{t,t+1} = VIX_t^2 \approx E_t^Q[RV_{t,t+1}] \). For ease of notation, here we assume zero interest rate, which is indeed negligible in the sample under investigation. This relationship can be generalized to link 1-month forward VS rates to the second moment of the VIX, as shown below

\[
E_t^Q[RV_{t+\tau,t+\tau+1}] \approx E_t^Q[VIX_{t+\tau}^2]. \tag{21}
\]

Equation (21) clearly generalizes the definition of VIX, which is recovered for \( \tau = 0 \). By aggregating the terms on the RHS of (21) over \( \tau \) one can obtain variance swap rates for maturities longer than one month, that is

\[
VST_{t,t+n} \approx E_t^Q[RV_{t,t+n}] = \frac{1}{n} \sum_{\tau=1}^{n} E_t^Q[RV_{t+\tau-1,t+\tau}] \approx \frac{1}{n} \sum_{\tau=1}^{n} E_t^Q[VIX_{t+\tau-1}^2]. \tag{22}
\]

The relation in (22) holds under the inherent assumption that the joint market of VIX and RV is priced consistently under the unique risk neutral measure \( Q \), so that

\[
E_t^Q[VIX_{t+\tau}^2] \approx E_t^Q E_{t+\tau}^Q[RV_{t+\tau,t+\tau+1}] = E_t^Q[RV_{t+\tau,t+\tau+1}].
\]

In principle, by generalizing the VIX formula (3), one can always replicate \( VST_{t,t+\tau} \) through a portfolio of SPX options expiring in \( n \) months. This provides a tool to assess whether the SPX and the VIX markets are consistent with each other. In Figure 8 we report the time series of \( VST_{t,t+\tau} \) for \( \tau = 3 \) months (left panel) and \( \tau = 4 \) months (right panel). Each plot displays the VS computed from SPX options by extending formula (3) (red line) and the VS implied by the RND of VIX through (22) (blue line). Figure 8 shows that the two series follow the same dynamic pattern, with the SPX-implied VS being slightly above the one implied by VIX throughout most of the sample, especially at \( \tau = 4 \) months. As a matter of fact, the approximating formula (3) truncates the non-negative integrand in (2), based on the range of available strikes on SPX options, generating a negative bias. Accordingly, the difference between the two time series is much less noticeable starting from the second half of 2014, as the set of available
strikes from the OptionMetrics database has become larger and thicker, thus reducing the
discretization/truncation error in (3). Based on this evidence, we conclude that the small
differences between the SPX-implied VS and the VIX-implied VS are mostly due to numerical
issues and not by presence of profitable cross-market arbitrage opportunities. This means that
the second risk-neutral moment of the VIX can be regarded as an equity derivative, as it can
be replicated by combining short and long positions on SPX options expiring in $\tau + n - 1$ and
$\tau + n$ months, respectively.

6.2 Risk-neutral moments

By integration of the estimated RND, it is possible to retrieve the first $k$ risk-neutral moments
of the VIX for each value of $t$ and $\tau$. An empirical analysis of the risk-neutral moments of
the VIX reveals a set of stylized facts related to the market expectation of volatility (and its
powers). Some descriptive statistics on the estimated values of mean, variance, skewness and
kurtosis are presented in Table 4.

<table>
<thead>
<tr>
<th>( \tau )</th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>19.26</td>
<td>5.30</td>
<td>12.63</td>
<td>35.71</td>
<td>35.19</td>
<td>27.92</td>
<td>6.57</td>
<td>136.54</td>
<td>3.43</td>
<td>1.80</td>
<td>5.47</td>
<td>22.90</td>
<td>10.53</td>
<td>7.91</td>
<td>48.19</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>20.36</td>
<td>5.17</td>
<td>13.58</td>
<td>35.05</td>
<td>64.31</td>
<td>37.31</td>
<td>19.51</td>
<td>189.05</td>
<td>3.41</td>
<td>1.78</td>
<td>5.68</td>
<td>22.86</td>
<td>10.84</td>
<td>6.47</td>
<td>54.56</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>21.13</td>
<td>5.00</td>
<td>14.47</td>
<td>34.45</td>
<td>83.71</td>
<td>41.61</td>
<td>27.24</td>
<td>236.48</td>
<td>3.14</td>
<td>1.83</td>
<td>5.39</td>
<td>19.91</td>
<td>10.13</td>
<td>7.65</td>
<td>47.78</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>21.67</td>
<td>4.92</td>
<td>15.08</td>
<td>33.99</td>
<td>97.85</td>
<td>45.27</td>
<td>41.87</td>
<td>268.13</td>
<td>2.99</td>
<td>1.83</td>
<td>5.39</td>
<td>18.05</td>
<td>9.09</td>
<td>8.21</td>
<td>51.16</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>22.17</td>
<td>4.88</td>
<td>15.44</td>
<td>34.10</td>
<td>109.50</td>
<td>48.82</td>
<td>45.11</td>
<td>286.35</td>
<td>2.78</td>
<td>1.78</td>
<td>4.52</td>
<td>15.55</td>
<td>7.07</td>
<td>1.28</td>
<td>38.13</td>
<td></td>
</tr>
</tbody>
</table>

Table 4: Descriptive statistics on the estimated risk-neutral moments of VIX.
kurtosis are reported in Table 4. By means of the skewness ($\text{Sk}_{Q_{t,\tau}}$) and the kurtosis ($\text{Kurt}_{Q_{t,\tau}}$), reported in Panels (c) and (d) of Figure 9, we study how the shape of the RND changes over time. Both $\text{Sk}_{Q_{t,\tau}}$ and $\text{Kurt}_{Q_{t,\tau}}$ appear highly volatile and share similar dynamics, as suggested by their sample correlations reaching levels close to 97%. In contrast to what we observe for the mean and the variance in Panels (a) and (b), the skewness and the kurtosis plummet during periods of market turmoil, meaning that the third and the fourth moment increase at a slower rate compared to the variance. As the conditional mean and variance of VIX increase, moving away from the zero lower bound, the distribution tends to become more symmetric and thinner-tailed. Notably, both $\text{Sk}_{Q_{t,\tau}}$ and $\text{Kurt}_{Q_{t,\tau}}$ exhibit a term structure that is systematically downward sloped. Sample averages decrease from 3.43 ($\tau = 1$) to 2.78 ($\tau = 5$) for the $\text{Sk}_{Q_{t,\tau}}$ and from 22.90 ($\tau = 1$) to 15.55 ($\tau = 5$) for $\text{Kurt}_{Q_{t,\tau}}$. As the prediction horizon increases,

![Figure 9: Time series of the VIX risk-neutral mean, variance, skewness and kurtosis for different times-to-maturity, $\tau = 1$ (blue), $\tau = 2$ (red), $\tau = 3$ (yellow), $\tau = 4$ (purple) and $\tau = 5$ (green).](image-url)
the distribution implied by the VIX options reflects higher expectations and volatility for the level of VIX, together with a more symmetric and platykurtic distribution. The similarity in the dynamic behavior of the first four moments becomes striking after simple rescaling. As an example, Figure 10 plots the standardized time series of the first four moments for $\tau = 1$ and $\tau = 5$, with lines tracking one another. The sample correlations are well above 90% for both $\tau = 1$ and $\tau = 5$, supporting the existence of a strong link between (standardized) moments.

To further explore the behavior of the implied moments both over the time dimension and the term structure, we test for the existence of a common factor structure across maturities and moment orders. This is conceptually similar to the analysis carried out by Christoffersen et al. (2009) and Andersen et al. (2015b) on the SPX implied volatility surface. The PCA, carried out on the cross-section of $\tau$, shows that the first principal component of $\text{Mean}_{t,\tau}^{Q}$ explains 98% of its total covariance. Similarly, the first principal component of $\text{Var}_{t,\tau}^{Q}$ explains almost 96% and, for the third and fourth moments, the first principal component explains 89% and 82% of the related cross sectional (across maturities) variability. When regressing $\text{Mean}_{t,\tau}^{Q}$ on a constant and the first principal component, we find regression intercepts that increase with the maturity, ranging from 19.26 ($\tau = 1$) to 22.17 ($\tau = 5$), but slopes that are fairly invariant.

This suggests that VIX options prices incorporate premia for the uncertainty about future market conditions as the horizon increases, but also that these premia are constant over time,
and thus deterministic in nature. The PCA for the higher order moments shows qualitatively the same results. We also test whether and to what extent the high degree of correlation between the four moments, as illustrated in Figure 10, is due to the presence of a common factor. We find that the first principal component explains nearly 93% of the total covariation between the 20 variables. This result provides strong evidence of a single main driver for the RND of VIX over time.

6.3 Jumps under \( Q \)

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

\[
X_t = \mu_t \eta_t, \quad t = 1, \ldots, T,
\]

where, for a fixed maturity \( \tau \), \( X_t = \text{VIX}_{t+\tau} \), \( \mu_t \) is the expectation of \( \text{VIX}_{t+\tau} \) computed at the time \( t \), \( \eta_t \) is an i.i.d. stochastic term, independent of \( \mu_t \), with positive support and with mean equal to 1, and \( T \) is the sample size. MEM type dynamics for the VIX index are also studied in Mencía and Sentana (2017), who adopt a Laguerre expansion for the innovation term \( \eta_t \) under the physical measure \( \mathbb{P} \). They also analyze the process under \( Q \), using the VIX futures and assuming an exponentially affine stochastic discount factor to model the risk premium. In this way, they retrieve information on the risk-neutral higher moments of VIX by a structural assumption to link \( \mathbb{P} \) and \( Q \). We instead perform a direct analysis of the distributional assumptions on the VIX under \( Q \), exploiting higher risk-neutral moments obtained from the VIX options. The idea of using transformations of option prices to back out estimates of a parametric model is in line with the methodology adopted in Pastorello et al. (2000) and Pan (2002), among others. By setting \( \eta_t \sim i.i.d. \Gamma(1, \nu) \) (mean-shape form), which corresponds to a zero-order Laguerre expansion, we obtain

\[
E_t^Q [\text{VIX}_{t+\tau}] = \mu_t, \\
E_t^Q [\text{VIX}^2_{t+\tau}] = \mu_t^2 \left( \frac{1}{\nu} + 1 \right), \\
E_t^Q [\text{VIX}^3_{t+\tau}] = \mu_t^3 \left( \frac{2}{\nu^2} + \frac{3}{\nu} + 1 \right), \\
E_t^Q [\text{VIX}^4_{t+\tau}] = \mu_t^4 \left( \frac{6}{\nu^3} + \frac{11}{\nu^2} + \frac{6}{\nu} + 1 \right).
\]
Under the MEM-Gamma the proportionality between second, third and fourth moment is controlled by a single parameter, $\nu$. The MEM-Gamma can be augmented with jumps (MEM-J),

$$X_t = \mu_t Z_t \eta_t, \quad t = 1, \ldots, T,$$

(23)

where $Z_t$ is the VIX jump component, and the innovation $\eta_t \sim i.i.d. \Gamma(1, \nu)$, see Caporin et al. (2017). The jump term, $Z_t$ is defined as $Z_t = d_\lambda \cdot 1_{N_t = 0} + \sum_{j=1}^{N_t} Y_{j,t} \cdot 1_{N_t > 0}$, where $1$ is the indicator function, $N_t$ is a Poisson random variable with intensity $\lambda > 0$ and $d_\lambda = \left( e^{-\lambda} + \lambda \right)^{-1}$ denotes the baseline value of $Z_t$ in absence of jumps. When $N_t > 0$, the process $Z_t$ is a compound Poisson process. Following Caporin et al. (2017), we assume that the jump sizes are driven by a Gamma distribution $\Gamma(d_\lambda, \varsigma)$ (in mean-shape form), from which it follows that the innovation term $\xi_t = Z_t \eta_t$ is driven by a countably infinite mixture of Gamma and Kappa distributions with closed-form moments given in Caporin et al. (2017). The MEM-J generates a mixture of distributions conceptually similar to the mixture generated by the Laguerre expansion used in Mencia and Sentana (2017), with further interpretation for tail-events as generated by jumps.

For every $\tau = 1, \ldots, 5$, we estimate the MEM-Gamma and the MEM-J models by GMM exploiting the time-series of the first four risk-neutral moments of the VIX. The GMM estimates of the parameters are reported in Table 5. The estimates of the parameter $\nu$ obtained under the Gamma distribution are always significant and they decrease with maturity. Since $\nu$ is the inverse of the variance of $\eta_t$ in the Gamma specification, this evidence correctly signals the increase in the uncertainty around future values of the VIX at longer horizons. However, the MEM with Gamma distributed innovations does a poor job matching the structure of the higher moments of the VIX under the $Q$-measure. Indeed, the J-test for over-identifying restrictions, which is distributed as a $\chi^2$ with 2 degrees of freedom, strongly rejects the null hypothesis that the higher moments are those implied by a Gamma distribution. Instead, adding flexibility to the higher moments of VIX by allowing for a jump term, provides a good description of the linkages between the VIX risk-neutral moments, since the $p$-value of the J-test is larger than 5% for every $\tau = 1, \ldots, 5$. Looking at the parameter estimates of the MEM-J, we note that $\hat{\nu}$ decreases with maturity, thus signaling an increase of the variance at longer horizons. The mixing parameter $\lambda$ is rather stable across $\tau$ and close to 0.5. This means an
average of 1 jump every second month. Notably, the parameter $\lambda$ is always significant, thus indicating that tail events, like the jumps, are priced in the VIX options. This evidence is in line with the recent results of Zang et al. (2017). Finally, Table 5 also reports the estimates and the fitting of the MEM model when pooling all maturities and imposing $\nu_\tau = e^{-\kappa \tau} \nu$ to account for the term structure. The results suggest that the fitting of the MEM-J models to the first four moment and across maturities is very good with a $p$-value of the J-test equal to 76.5%.

In particular, the estimate of $\kappa$ is 0.180, which means that the variance of VIX increases by a factor of approximately 20% when extending the maturity by one month.

A graphical illustration of the jump contribution to the probability of tail events is provided in Figure 11. The figure reports the model-implied RND of the VIX under $\mathbb{Q}$ in “steady state”, obtained by setting $\mu_t$ to its long-run value, $\bar{\mu} = \frac{1}{T} \sum_{t=1}^{T} \text{Mean}_t^\mathbb{Q}$. In steady state, the option-implied probability of observing VIX larger than 31% is negligible under the MEM-Gamma for $\tau = 1$, while it is approximately 4.5% if jumps are included in the model. Analogous evidence emerges for $\tau = 5$.

<table>
<thead>
<tr>
<th>$\tau$</th>
<th>MEM-Gamma</th>
<th>MEM-J</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\hat{\nu}$</td>
<td>$\hat{\kappa}$</td>
</tr>
<tr>
<td>1</td>
<td>26.38$^a$</td>
<td>0.000</td>
</tr>
<tr>
<td>2</td>
<td>12.87$^a$</td>
<td>0.000</td>
</tr>
<tr>
<td>3</td>
<td>12.66$^a$</td>
<td>0.000</td>
</tr>
<tr>
<td>4</td>
<td>9.433$^a$</td>
<td>0.000</td>
</tr>
<tr>
<td>5</td>
<td>7.411$^a$</td>
<td>0.000</td>
</tr>
<tr>
<td>ALL</td>
<td>8.515$^a$</td>
<td>0.155$^a$</td>
</tr>
</tbody>
</table>

Table 5: GMM estimates of the parameters of the MEM-Gamma and the MEM-J. The estimation is performed by matching the parametric expressions of $E_t^Q[VIX_{t+\tau}^2]$, $E_t^Q[VIX_{t+\tau}^3]$ and $E_t^Q[VIX_{t+\tau}^4]$ outlined above with their empirical counterpart obtained from the RNDs retrieved from the option prices for different maturities ($\tau$) or pooling all maturities (ALL). The first moment is exactly matched, i.e. $\text{Mean}_t^\mathbb{Q}$ is $E_t^Q[VIX_{t+\tau}] = \mu_t$, by construction, and it is used as a driver for the dynamics of the higher order moments. The subscripts $a$, $b$ and $c$ stand for significance at 1%, 5% and 10%, respectively. $p(J)$ is the $p$-value of the J-test for over-identifying restrictions whose distribution is $\chi^2(m-r)$, where $m$ is the number of moment conditions adopted in the estimation and $r$ is the number of free parameters.
Figure 11: Steady-state risk neutral densities of VIX\(_t\) implied by the MEM-Gamma model and by the MEM-J model of Caporin et al. (2017) for \(\tau = 1\) (upper panel) and \(\tau = 5\) (lower panel).

6.4 Tail risk under \(Q\)

In this section, we exploit the RND of VIX to compute the probability of volatility tail events implied in VIX options. Similarly to Aït-Sahalia and Lo (2000) and by analogy to the Value-at-Risk introduced for quantifying the risk of extremely negative returns, we can define the "economic" Volatility-at-Risk (E-VolaR), i.e. the risk of extreme high volatility under \(Q\). In particular, for given \(\alpha\), E-VolaR\((t, \tau; \alpha)\) is the value satisfying \(Q_t(\text{VIX}_{t+\tau} > \text{E-VolaR}(t, \tau; \alpha)) = \alpha\), where \(Q_t\) denotes the risk-neutral probability at time \(t\). Therefore, E-VolaR\((\alpha)\) is the VIX level that may be exceeded with probability \(\alpha\).

Differently from the so-called "statistical" VolaR (S-VolaR), derived under the real-world probability, the E-VolaR relies on the information contained in Arrow-Debreu securities or, equivalently, in vanilla options. As such, the E-VolaR is not a forecast of extreme losses based on their likelihood, but it provides an economic valuation of risk that "incorporates many other aspects of market risk that are central to the practice of risk management" (Aït-Sahalia and Lo, 2000). By contrasting the E-VolaR and the S-VolaR one can obtain a measure of the risk aversion for extreme volatility of market participants. Thus, the knowledge of E-VolaR allows for designing volatility trading strategies where the amount of a volatility hedge under extreme conditions is determined by also accounting for the market price of risk. Based on
the notion of E-VolaR, we can also define the economic conditional VolaR as

$$\text{E-CVolaR}_Q(t, \tau; \alpha) = E_Q^{t}[\text{VIX}_{t+\tau} | \text{VIX}_{t+\tau} > \text{E-VolaR}(\alpha)] = \frac{1}{\alpha} \int_0^{\infty} f_Q(t, \tau; x) (x - K)^+ dx,$$

which is an extension of the concept of risk-neutral expected shortfall to the volatility context.

A number of clues emerge from the empirical properties of the E-CVolaR, whose time series are reported in Figure 12. First, the E-CVolaR strongly exhibit a term structure pattern, with levels of E-CVolaR associated to longer times-to-maturity being consistently above those of shorter maturities. At 1-month maturity, the E-CVolaR is on average around 35% and 45% at 5% and 1% probability levels, respectively. Instead, at 5-months maturity, the E-CVolaR is on average around 55% and 80% at 5% and 1% probability levels, respectively. This means, that the risk-neutral expected excess VIX over the E-VolaR threshold is approximately 20% larger when extending the time horizon from 1 to 5 months. Second, the E-CVolaR is time-varying and it increases during turmoil periods. For instance, during the European debt crisis in the fall of 2011, the E-CVolaR reaches values well above 100% for the 5-months horizons at 1% probability level. Finally, the dynamics of E-CVolaR for different maturities are driven by the same stochastic trend, which again supports the existence of a common factor structure across maturities, in line with the multiplicative error structure across time and maturities adopted in Section 6.3.

**Figure 12:** Economic conditional risk-neutral VaR (E-CVolaR) implied by VIX options. Each panel reports time series of the E-CVolaR associated with a different probability for several times-to-maturity. In particular, $\tau = 1$ (blue), $\tau = 2$ (red), $\tau = 3$ (yellow), $\tau = 4$ (purple) and $\tau = 5$ (green).
7 Conclusions and directions for future research

In this paper, we proposed a methodology based on a finite orthogonal expansion to infer the RND underlying the VIX option prices. The method generalizes the Laguerre expansions and suits cases where the density is supported over the positive real axis. The approach is non-structural since it does not require restrictive parametric assumptions on the underlying asset dynamics, reducing the number of restrictions to be imposed on the form of the RND. Therefore, we drastically reduce the intrinsic risk of misspecification entailed in parametric models. A number of numerical experiments confirms that the extended Laguerre polynomials are often superior to methods based on the estimation of the second derivatives in a wide range of cases. Furthermore, our empirical study on VIX options highlights the robustness of this technique when the goal is to study features of the RND of VIX involving its right tail behavior. In particular, we shed some light on the main stylized facts associated with the the RND of VIX and compute a number of relevant quantities such as the VVIX, the variance swap (VS) term structure, the VIX risk-neutral higher moments and RND quantiles. These quantities play a crucial role in assessing the presence of cross-market arbitrage opportunities with SPX options, in testing for the presence of VIX risk-neutral jumps and to measure the volatility tail risk under $Q$.

The proposed methodology may be particularly useful to study the risk-premia embedded in volatility or inflation options, see Bollerslev and Todorov (2011), Kitsul and Wright (2013) and Schneider (2015). This would provide an alternative to the parametric methodology of Andersen et al. (2015a) to carry out inference on the underlying processes based on panels of options. A multivariate functional dynamic model for the RND would be a further natural extension of this work. Along these lines, the analysis of the risk premia embedded in VIX options could be carried out by studying the shape and the variability over time of the pricing kernel (see Grith et al., 2013) in a bivariate setting that also includes the SPX. Indeed, given that the VIX is computed through a portfolio of SPX options, the volatility index has obviously a close connection with the SPX, which means that the two quantities cannot be modeled independently. However, since the functional linkage between the SPX and the related options
is, in general, not explicitly known, the problem of correct specification of the bivariate RND is particularly troublesome in the context of joint SPX-VIX option pricing. This would further extend the recent work of Song and Xiu (2016).

References


### A Proofs

#### A.1 Proof of Theorem 3.1

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

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

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

and the subspace

$$
\mathcal{H}_\phi^* = \text{Cl}(\text{span} \ \phi h^\phi_k, \ k \in \mathbb{N}) \subseteq \mathcal{H}_\phi .
$$

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

$$
f_Q^{(\infty)} := \lim_{n \to +\infty} \phi^{1/2} \left( 1 + \sum_{k=1}^{n} c_k h^\phi_k \right) \text{ in } \mathcal{H}_\phi
$$
solves the minimum distance problem
\[ f_Q^{(\infty)} = \arg\min_{\psi \in \mathcal{H}_\phi^*} \left\{ \psi - f_Q, \psi - f_Q \right\}^{\frac{1}{2}}. \] (25)

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

**Definition A.2** (Closed polynomial set in \( \mathcal{H}_\phi \)). The kernel \( \phi \) is said to generate closed polynomial sets if
\[ \text{Cl} \left( \text{span } x^k, k \in \mathbb{N} \right) = L^2_{\phi(x)dx}(\mathcal{D}). \] (26)

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 \( \phi \).

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

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

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

(ii) If \( \lim_{x \to +\infty} \phi(x)e^{\gamma x^{-\gamma}} > 0 \) for some \( \gamma, \xi > 0 \), then \( \phi \) does not generate closed polynomial sets.

We can now prove Theorem 3.1. In the following, \( \mathcal{H}_\phi \) and \( \mathcal{H}_\phi^* \) refer to the Hilbert spaces defined in Lemma A.1. If \( (h_k^\phi)_{k \in \mathbb{N}} \) is closed with respect to \( \phi \), then \( \mathcal{H}_\phi = \mathcal{H}_\phi^* \) and from Lemma A.1 it follows that \( f_Q^n = f_Q \) whenever \( \phi^{\frac{1}{2}}f_Q \in L^2(\mathcal{D}) \). The first implication can be readily shown by noticing that \( f_Q \in \mathcal{H}_\phi \) implies \( \phi^{-\frac{1}{2}}f_Q \in L^2_{\phi(x)dx}(\mathcal{D}) \). Then, the closure of \( (h_k^\phi)_{k \in \mathbb{N}} \) implies that \( \phi^{-\frac{1}{2}}f_Q \) can be approximated by a certain polynomial series \( a_0 + a_1x + a_2x^2 + \ldots \) in \( L^2_{\phi(x)dx}(\mathcal{D}) \) or equivalently that \( f_Q \) 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 \( \mathcal{H}_\phi \). Then, (a) follows immediately from Lemma A.1 by noticing that the assumptions on \( \phi \) imply \( \mathcal{H}_\phi = \mathcal{H}_\phi^* \), in view of Theorem A.3-(i). To prove (9), we observe that for every \( \phi \in \mathcal{H}_\phi^* \) and every \( n \in \mathbb{N} \)

\[
\left| \int_0^{+\infty} \Pi(x)f_Q^{(n)}(x)dx - \int_0^{+\infty} \Pi(x)f_Q^{(\infty)}(x)dx \right| \leq \int_0^{+\infty} \Pi(x) \left| f_Q^{(n)}(x) - f_Q^{(\infty)}(x) \right| dx \quad = \int_{\mathcal{D}} \phi^{\frac{1}{2}}(x)\Pi(x) \left| \phi^{-\frac{1}{2}}(x)f_Q^{(n)}(x) - \phi^{-\frac{1}{2}}(x)f_Q^{(\infty)}(x) \right| dx \leq b \cdot \left( f_Q^{(n)} - f_Q^{(\infty)}, f_Q^{(n)} - f_Q^{(\infty)} \right)^{\frac{1}{2}},
\]

where \( b = \left( \phi^{\frac{1}{2}}\Pi, \phi^{\frac{1}{2}}\Pi \right)^{\frac{1}{2}} \) is finite by hypothesis. Finally, (9) follows from Lemma A.1.

41
Proof of Theorem A.3

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

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

Proof. By the Riesz-Fischer characterization it suffices to prove that if there exists $f \in L^2_{\phi^*(x)dx}(\mathcal{D})$ such that
\[
\int_{\mathcal{D}} f(x)x^k\phi(x)dx = 0 \quad \forall k \in \mathbb{N},
\]
then it must hold that $f(x) = 0$ a.e. on $\mathcal{D}$. Define $g(x) = h(x)f(x)$, then
\[
\int_{\mathcal{D}} g^2(x)\phi^*(x)dx \leq \max_{x \in \mathcal{D}} h(x) \cdot \int_{\mathcal{D}} f^2(x)\phi(x)dx < +\infty,
\]
which proves $g \in L^2_{\phi^*(x)dx}(\mathcal{D})$. Furthermore, $\int_{\mathcal{D}} g(x)x^k\phi^*(x)dx = \int_{\mathcal{D}} f(x)x^k\phi(x)dx = 0$ for every $k \in \mathbb{N}$, which implies in view of hypothesis that $g(x) = 0$ a.e. on $\mathcal{D}$ and therefore $f(x) = 0$ a.e. on $\mathcal{D}$ due to positivity assumptions on $h(x)$. □

To prove statement (i) we start by recalling a classic result due to Hewitt (1954) showing that every bounded function $\psi$ supported on the entire real line and such that
\[
\lim_{|x| \rightarrow +\infty} \psi(x)e^{\varsigma|x|} = 0 \quad (27)
\]
generates closed polynomial sets. Based on this result, statement (i) can be proven under the additional hypothesis that $\phi$ is bounded. Indeed, under this assumption, the function $\psi(x) = |x|\phi(|x|^{\frac{1}{2}})$ is bounded on $\mathbb{R}$ and satisfies (27), 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 $\phi$, we remark that by hypothesis there exist a polynomial $p$ and $\varsigma^* > 0$ such that the function $\phi^*$ defined by
\[
\phi^*(x) := p(x)e^{\varsigma^*\sqrt{x}}\phi(x)
\]
is bounded on $\mathcal{D}$. Since $\phi^*$ clearly preserves the same integrability and asymptotic properties of $\phi$, then it generates closed polynomial sets. Now, consider $f$ such that $f^2\phi$ is integrable and
\[
\int_{\mathcal{D}} f(x)x^k\phi(x)dx = 0, \quad \forall k \in \mathbb{N}.
\]
Furthermore, define $g$ as $g(x) = e^{-\varsigma^*\sqrt{x}}f(x)$ for $x \in \mathcal{D}$. We have
\[
\int_{\mathcal{D}} g(x)^2\phi^*(x)dx \leq \sup_{x \in \mathcal{D}} |p(x)e^{-\varsigma^*\sqrt{x}}| \int_{\mathcal{D}} f^2(x)\phi(x)dx < +\infty.
\]
On the other hand, for every \( k \in \mathbb{N} \),
\[
\int_D g(x)x^k\phi^*(x)dx = \int_D f(x)x^k\phi(x)dx = 0,
\]
which proves \( g(x) = 0 \) and therefore \( f(x) = 0 \) a.e. on \( D \). Then, statement (i) is proved.

The proof of statement (ii) is based on a known counterexample in the theory of orthogonal polynomials (cf. entry "Closed system of elements" in Hazewinkel (1988)), showing that every function \( \psi \) of the form
\[
\psi(x) = e^{-|x|^2 m^2 + 1}, \quad x \in \mathbb{R}, \ m \in \mathbb{N},
\]
does not generate closed polynomial sets. By combining this counterexample with the results of Shohat (1942), we can prove that the function \( \psi \) supported on \([0, +\infty[\) and defined by
\[
\psi(x) = e^{-x m^2}, \quad x \geq 0, \ m \in \mathbb{N}.
\]
does not generate closed polynomial sets. By a change of variable and through the Riesz-Fischer characterization, one can extend the latter result to the case where \( \psi \) is supported on \([x_0, +\infty[\) and is of the form
\[
\psi(x) = e^{-\zeta(x-x_0) m^2}, \quad x \geq x_0, \ m \in \mathbb{N}.
\]
for some \( \zeta > 0 \) and \( x_0 \geq 0 \). To prove statement (ii), then, we proceed by contradiction and suppose that there exists an integrable function \( \phi \), supported on \([0, +\infty[\) and such that, for some \( \gamma > 0 \),
\[
\lim_{x \to +\infty} \phi(x)e^{\gamma x} > 0,
\]
which generates closed polynomial sets. To this aim, we observe that by the hypothesis made on the right-tail of \( \phi \), 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 \( \phi \) holds in particular when the support is restricted, by truncation, to \([x_0, +\infty[\). Furthermore, the function \( h \) defined by
\[
h(x) = e^{-\zeta(x-x_0) m^2} \phi(x)^{-1},
\]
is bounded on \([x_0, +\infty[\), for \( m \) sufficiently large. Then, as a consequence of Lemma A.4, the function \( e^{-\zeta(x-x_0) m^2} \phi(x)^{-1} \) generates closed polynomial sets on \([x_0, +\infty[\), which is a contradiction. This concludes the proof. \( \square \)

B Details on VIX option data and goodness-of-fit
Figure 13: Comparison of RMSEs. The figure reports the RMSE of pricing errors obtained by SSI (solid-red line) and OPE (dashed-blue line) for $\tau = 1$ and $\tau = 5$ for all dates in the sample. The t-test for the difference of the mean RMSE, $t = \sqrt{T}d/\text{Var}(d)$ with $d = \text{RMSE}_{\text{OPE}} - \text{RMSE}_{\text{SSI}}$, is -4.437 and -7.255 for $\tau = 1$ and $\tau = 5$ respectively.

Table 6: Summary of the panel of VIX options considered in Section 5. Each row reports the information on the option prices used on a specific date. Specifically, each row reports, for several values of the time-to-maturity $\tau$, the total number of strikes $M$, the minimum (Min) and the maximum (Max) values of the strike range, collected after cleaning the original data as described in Section 5.
<table>
<thead>
<tr>
<th>Dates</th>
<th>τ = 1 month</th>
<th>τ = 2 months</th>
<th>τ = 3 months</th>
<th>τ = 4 months</th>
<th>τ = 5 months</th>
</tr>
</thead>
<tbody>
<tr>
<td>20 Jan 2010</td>
<td>0.0886</td>
<td>0.0621</td>
<td>0.0556</td>
<td>0.0503</td>
<td>0.0416</td>
</tr>
<tr>
<td>24 Feb 2010</td>
<td>0.0741</td>
<td>0.0475</td>
<td>0.0451</td>
<td>0.0453</td>
<td>0.0448</td>
</tr>
<tr>
<td>24 Mar 2010</td>
<td>0.0554</td>
<td>0.0416</td>
<td>0.0402</td>
<td>0.0401</td>
<td>0.0396</td>
</tr>
<tr>
<td>21 Apr 2010</td>
<td>0.0296</td>
<td>0.0194</td>
<td>0.0185</td>
<td>0.0184</td>
<td>0.0182</td>
</tr>
<tr>
<td>26 May 2010</td>
<td>0.0173</td>
<td>0.0111</td>
<td>0.0105</td>
<td>0.0104</td>
<td>0.0101</td>
</tr>
<tr>
<td>21 Jun 2010</td>
<td>0.0131</td>
<td>0.0086</td>
<td>0.0081</td>
<td>0.0080</td>
<td>0.0078</td>
</tr>
<tr>
<td>25 Sep 2010</td>
<td>0.0087</td>
<td>0.0056</td>
<td>0.0051</td>
<td>0.0050</td>
<td>0.0049</td>
</tr>
<tr>
<td>24 Oct 2010</td>
<td>0.0065</td>
<td>0.0042</td>
<td>0.0040</td>
<td>0.0039</td>
<td>0.0038</td>
</tr>
<tr>
<td>24 Nov 2010</td>
<td>0.0049</td>
<td>0.0030</td>
<td>0.0029</td>
<td>0.0028</td>
<td>0.0027</td>
</tr>
<tr>
<td>22 Dec 2010</td>
<td>0.0038</td>
<td>0.0024</td>
<td>0.0023</td>
<td>0.0022</td>
<td>0.0021</td>
</tr>
<tr>
<td>20 Jan 2011</td>
<td>0.0027</td>
<td>0.0017</td>
<td>0.0016</td>
<td>0.0015</td>
<td>0.0014</td>
</tr>
<tr>
<td>24 Feb 2011</td>
<td>0.0019</td>
<td>0.0012</td>
<td>0.0012</td>
<td>0.0011</td>
<td>0.0010</td>
</tr>
<tr>
<td>24 Mar 2011</td>
<td>0.0013</td>
<td>0.0008</td>
<td>0.0008</td>
<td>0.0007</td>
<td>0.0007</td>
</tr>
<tr>
<td>21 Apr 2011</td>
<td>0.0005</td>
<td>0.0003</td>
<td>0.0003</td>
<td>0.0003</td>
<td>0.0003</td>
</tr>
<tr>
<td>26 May 2011</td>
<td>0.0002</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>21 Jun 2011</td>
<td>0.0001</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>25 Sep 2011</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>24 Oct 2011</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>24 Nov 2011</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>22 Dec 2011</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Table 7: Root mean square error (RMSE): errors between the observed VIX option prices and the approximate prices implied by GW kernel and related expansion.