Journal of Econometrics 190 (2016) 176–196

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

A tale of two option markets: Pricing kernels and volatility risk☆

Zhaogang Song [a,*], Dacheng Xiu [b]

[a] Carey Business School, John Hopkins University, 100 International Drive, Baltimore, MD 21202, USA

[b] University of Chicago, Booth School of Business, 5807 South Woodlawn Avenue, Chicago, IL 60637, USA

ARTICLE INFO

Article history:
Received 11 February 2014
Received in revised form
15 April 2015
Accepted 2 June 2015
Available online 25 September 2015

JEL classification:
G12
G13

Keywords:
Multi-factor volatility model
Pricing kernel
Variance swaps
VIX options

ABSTRACT

Using both S&P 500 option and recently introduced VIX option prices, we study pricing kernels and their dependence on multiple volatility factors. We first propose nonparametric estimates of marginal pricing kernels, conditional on the VIX and the slope of the variance swap term structure. Our estimates highlight the state-dependence nature of the pricing kernels. In particular, conditioning on volatility factors, the pricing kernel of market returns exhibit a downward sloping shape up to the extreme end of the right tail. Moreover, the volatility pricing kernel features a striking U-shape, implying that investors have high marginal utility in both high and low volatility states. This finding on the volatility pricing kernel presents a new empirical challenge to both existing equilibrium and reduced-form asset pricing models of volatility risk. Finally, using a full-fledged parametric model, we recover the joint pricing kernel, which is not otherwise identifiable.

1. Introduction

In addition to market risk, volatility risk has been well documented as an essential component of time-varying investment opportunities. A priced volatility factor leads to a pricing kernel (or stochastic discount factor) that depends on both the market return and volatility. Nevertheless, existing estimates of pricing kernels either ignore volatility factors in a nonparametric analysis or impose strong parametric restrictions because volatility is neither directly tradable nor observable.

The lack of tradable and observable volatility has changed substantially since the introduction of the Volatility Index (VIX) in 1993 by the Chicago Board of Options Exchange (CBOE), and the introduction of VIX derivatives such as futures and options in 2004 and 2006, respectively.[1] The VIX, derived from S&P 500 options as the square root of the expected average variance over the next 30 calendar days, provides investors with a direct measure of volatility; and VIX derivatives offer investors convenient instruments for trading on the volatility of the S&P 500 index.[2] As a result, the VIX is constantly exposed in the media spotlight, and VIX options have achieved abundant liquidity and become the third-most-active contracts at the CBOE as of October 2011.

Taking advantage of newly available VIX options, combined with S&P 500 index options, we propose a nonparametric framework to study the marginal pricing kernel as well as its dependence on volatility, incorporating multiple volatility factors. We also provide a complementary parametric analysis for the joint pricing kernel which is not identifiable nonparametrically. Together, we

☆ We are grateful to the Editor Oliver Linton and two anonymous referees for many insightful comments. We also benefited from discussions with Yacine Aït-Sahalia, Andrea Buraschi, Bjorn Eraker, Peter Carr, Peter Christoffersen, Fousseni Chabi-Yo, George Constantinides, Jianqing Fan, René Garcia, Kris Jacobs, Jakub Jurek, Ilze Kalnina, Ralph Koijen, Nicholas Polson, Eric Renault, Jeffrey Russell, Neil Shephard, George Tauchen (discussant), Viktor Todorov, Grigory Vilkov (discussant), and Hao Zhou, as well as seminar and conference participants at the University of Chicago, Northwestern, Princeton, Toulouse School of Economics, Liverpool School of Management, the 2012 CICF, the 5th Annual SoFiE Conference, the Measuring Risk conference 2012, the 2012 Financial Engineering and Risk Management International Symposium, and the 2012 International Symposium on Risk Management and Derivatives. Xiu acknowledges research support by the Fama-Miller Center for Research in Finance at Chicago Booth. The views expressed herein do not reflect those of the Federal Reserve System.

* Corresponding author.
E-mail addresses: zsong8@jhu.edu (Z. Song), dacheng.xiu@chicagobooth.edu (D. Xiu).

http://dx.doi.org/10.1016/j.jeconom.2015.06.024
0304-4076/

[1] The VIX, from its inception, was calculated from S&P 500 index options by inverting the Black–Scholes formula. In 2003, the CBOE amended this approach and adopted a model-free method to calculate the VIX using a portfolio of S&P 500 option price quotes.

[2] In order to trade volatility, investors previously had to take positions in a delta-hedged portfolio, or option portfolios such as straddles or strangles.

document several important empirical facts on the asset pricing implications of volatility risk.

First, we estimate marginal pricing kernels of market returns nonparametrically using daily data of S&P 500 options from January 1996 through December 2012. Given the documented importance of multiple volatility factors in capturing the dynamics of option prices (Christoffersen et al. (2008), Egloff et al. (2010), Mencia and Sentana (2012), and Bates (2012)), we accommodate two volatility factors in our nonparametric estimation framework. In particular, we use both the VIX and the slope of the variance swap term structure (defined as the ratio of 12-month and 3-month variance swap rates minus 1) as conditioning variables. Our estimates show that the pricing kernel of market returns strongly depends on volatility factors. Specifically, the pricing kernel is higher conditional on a low slope of the variance term structure, which signals stressful times. This provides corroborating evidence for the recently proposed asset pricing models in which the pricing kernel of market returns depends on the priced volatility risk (Bansal and Yaron (2004), Drechsler and Yaron (2011), Bollerslev et al. (2009), Zhou and Zhu (2012), and Branger and Völkert (2012)). Moreover, conditioning on volatility factors delivers a pricing kernel that exhibits a downward sloping shape in the range between negative returns and reasonably positive returns, but an upward sloping shape at the extreme right tail with a very wide confidence band. Recently, Linn et al. (2014) cast some doubt on the U-shape of the pricing kernel documented by Aït-Sahalia and Lo (2000) and Jackwerth (2000), claiming that "the 'pricing kernel puzzle' is a byproduct of econometric technique rather than a behavioral or economic phenomenon". Their main critique lies in the common practice of using unconditional physical density estimator in the construction of pricing kernels. Our finding suggests that conditioning on volatility factors helps achieve the desired downward slopping pattern within a reasonable range of returns. However, we cannot draw any definitive conclusion on the extreme right tail of the return distribution due to an insufficient amount of data.

Second, we provide nonparametric estimates of the volatility pricing kernel using daily data on VIX options from July 2007 through December 2012. Our estimates show that the volatility pricing kernel is also strongly state dependent: it is significantly higher conditional on a high lagged VIX mainly in the left tail, and conditional on a low slope factor mainly in the right tail. That is, consistent with the recently proposed asset pricing models mentioned above, the stochastic discount factor regarding volatility risk is higher in stressful times. More importantly, we find that the volatility pricing kernel exhibits a pronounced U-shape. Such a U-shape – in particular the left tail – implies that the stochastic discount factor increases as volatility decreases towards a very low level. This presents a new empirical challenge to most asset pricing models of volatility risk which lead to a monotonically increasing volatility pricing kernel. That being said, our empirical result is in agreement with the recent theoretical model proposed by Bakshi et al. (forthcoming), where they suggest that the heterogeneity in investors' beliefs can generate such a U-shape.

Third, we estimate the joint pricing kernel of the S&P 500 index return and the VIX using a full-fledged parametric model.[3] In particular, we employ a two-factor stochastic volatility model that is more flexible than most of the existing option pricing models in the literature. The estimates show that the joint pricing kernel achieves high values in the region of negative market return and relatively high VIX levels. These observations agree with the economic intuition that high marginal utility is associated with bad economic states. However, we caution that the parametric joint pricing kernel is subject to potential model misspecification error, whereas our nonparametric analysis of the marginal pricing kernels is more robust.

Finally, we compare the nonparametric and parametric marginal pricing kernels of the market return and the VIX to shed light on the limits of parametric models. We find that the parametric pricing kernel estimate of the market return is very close to the nonparametric estimate, despite that the parametric model causes slight overestimation in the region of extremely high returns. However, the parametric pricing kernel of volatility shows a monotonically increasing shape, in sharp contrast to our nonparametric VIX pricing kernel estimates that exhibit a pronounced U-shape. We also provide further evidence of potential model misspecification, e.g., the binding positivity constraints, which may prevent the parametric model from fully matching the data. Overall, such comparison reveals that the existing parametric models may capture how the market risk is priced, but largely fail to capture the price of volatility risk.

Estimating pricing kernels from option prices is discussed in Aït-Sahalia and Lo (1998), Aït-Sahalia and Duarte (2003), Jackwerth (2000), and Rosenberg and Engle (2002), which ignore the time-varying volatility risk. These studies document a puzzling U-shape of the pricing kernel. Subsequently, many studies have proposed alternative explanations for the U-shaped pricing kernel, including models with missing state variables in Chabi-Yo et al. (2008), Chabi-Yo (2012), and Christoffersen et al. (2013), and models with heterogeneous agents in Bakshi and Madan (2008) and Ziegler (2007). Our econometric analysis and empirical study contribute to this literature by documenting the dependence of the pricing kernel on multiple volatility factors. Importantly, we find that conditioning on volatility factors delivers a pricing kernel of a downward sloping shape up to the right tail where no definitive conclusion can be made due to limited amount of data.

Our paper is also related to the large body of literature on models with priced volatility risk, including both reduced-form option pricing models, e.g., Bakshi et al. (1997), Bates (2000), Pan (2002), Eraker (2004), and Broadie et al. (2007), and equilibrium models, such as Bansal et al. (2014), Bollerslev et al. (2012), and Campbell et al. (2012).[4] Unlike these studies, our framework does not depend on any parametric restrictions on volatility dynamics that may lead to misspecified pricing kernels.

A closely related study is Bakshi et al. (forthcoming), who suggest a U-shape volatility pricing kernel while exploring the link between the monotonicity of the pricing kernel and returns on VIX option portfolios. They further provide a static stylized model with heterogeneity in beliefs to account for the U-shape, in which the volatility market is dominated by investors with zero market risk. In contrast, we provide direct estimates of the volatility pricing kernel by nonparametric analysis, and further find that the U-shape kernel is state dependent. This finding implies that the price of volatility risk is dynamic, i.e., it depends on the time-varying economic states, which presents new empirical regularities that need to be incorporated into models of the volatility risk.

Methodologically, our paper is also related to Boes et al. (2007) and Li and Zhao (2009), who estimate pricing kernels of stock

[3] A follow-up paper to our study, Jackwerth and Vilkov (2013), implemented a semiparametric exercise using the parametric Frank copula along with two nonparametric marginal distributions. Their choice of copula, however, is ad hoc and inconsistent with the implied joint distribution from any dynamic asset pricing models.

[4] Several recent studies have constructed model-free measures of risk-neutral volatility from S&P 500 options, e.g., Bakshi and Kapadia (2003), Bollerslev et al. (2009), Carr and Wu (2009), and Todorov (2010), and compared them with measures of realized volatility. They focus on the sign, time variation, and the return predictability of variance risk premia, which relates only to the conditional mean of variance distributions under different measures.

market returns and interest rates, respectively, conditional on an ex-post volatility proxy filtered from historical time series. Our strategy differs from their approach as we take advantage of the functional relationship between the observed VIX and variance swap rates and the unobservable volatility factors. Therefore, our method avoids estimation errors arising from the filtering stage, which would otherwise be difficult to take into account. In view of this, our method can be considered as semiparametric, since it is built upon a parametric prediction supported by a large class of models.

Section 2 discusses the setup of pricing kernels and option prices, whereas Section 3 presents the nonparametric estimation framework and Monte Carlo simulations. Empirical estimates of state price densities and pricing kernels are provided in Section 4. Section 5 provides the estimate of the joint pricing kernel based on a parametric model. Section 6 concludes. The Appendix provides further details.

2. Model setup

We start by introducing some notations. The current date is denoted by t, whereas T denotes the maturity date of options, and $\tau = T - t$. We use capital letters to denote a random variable or a stochastic process. For any stochastic process X, we use small x to denote its realization at t, and x' to denote its realization at T. For convenience, we use $p(x'|\cdot)$ to denote the physical conditional density of X, $p^*(x'|\cdot)$ for its risk-neutral conditional density, and $p(x', y'|\cdot)$ and $p^*(x', y'|\cdot)$ for the joint densities of X and Y under the physical and risk-neutral measures, respectively.

To illustrate the idea, we start with an economy with two state variables: the log price of the S&P 500 index S and its unobserved volatility V, which determine the payoffs of assets and the pricing kernel. We introduce S&P 500 options into this economy, which in turn leads to the VIX, denoted by Z, and the VIX option market.[5,6] As a result, the information in the derivative markets is fully driven by the joint evolution of S and V, and Z is derived endogenously. Since V is not observable, there exist no Arrow–Debreu securities written on it. Instead, there are contingent claims with S and Z as their underlying indices, so that we focus on the Arrow–Debreu securities or state price densities of S and Z.

2.1. Nonparametric identification of marginal state-price densities

In this section, we show that the marginal state price densities of S and Z together span the two option markets, and that their joint densities, nevertheless, cannot be identified from option prices nonparametrically, unless additional contracts with payoff depending on both S and Z are introduced.

To fix ideas, we write the time-t price of a S&P 500 call option with maturity T and strike x as:

$$C(\tau, f, v, x, r_f) = e^{-r_f\tau}\mathbb{E}^{\mathbb{Q}}\left[(e^{S_T} - x)^+ | F_{t,\tau} = f, V_t = v\right] = e^{-r_f\tau}\int_{\mathbb{R}} (e^{s'} - x)^+ p^*(s'|\tau, f, v)ds',$$

where $p^*(s'|\tau, f, v)$ is the marginal state price density of S, $F_{t,\tau}$ denotes the log forward price of the S&P 500 index, $\tau = T - t$ is the time-to-maturity, and r_f is some deterministic risk-free rate between t and T.[7] Similarly, the price of a VIX call option with strike y is given by:

$$H(\tau, f, v, y, r_f) = e^{-r_f\tau}\mathbb{E}^{\mathbb{Q}}\left[(Z_T - y)^+ | F_{t,\tau} = f, V_t = v\right] = e^{-r_f\tau}\int_{\mathbb{R}} (z' - y)^+ p^*(z'|\tau, f, v)dz',$$

where $p^*(z'|\tau, f, v)$ is the marginal state price density of Z.

On the one hand, the two marginal state price densities completely determine these option prices. On the other hand, building upon the insight of Breeden and Litzenberger (1978), these densities can be recovered from the second-order derivatives of option prices with respect to strikes. Formally, we can obtain from S&P 500 options

$$p^*(s'|\tau, f, v) = e^{r_f\tau + s'}\frac{\partial^2 C(\tau, f, v, x, r_f)}{\partial x^2}\Big|_{x=e^{s'}}, \tag{1}$$

and from VIX options,

$$p^*(z'|\tau, f, v) = e^{r_f\tau}\frac{\partial^2 H(\tau, f, v, y, r_f)}{\partial y^2}\Big|_{y=z'}. \tag{2}$$

Therefore, $p^*(s'|\tau, f, v)$ and $p^*(z'|\tau, f, v)$ provide a sufficient and necessary summary of S&P 500 and VIX option prices, and the joint density of S and Z cannot be identified from the option prices without additional assumptions.

This fact delivers a somewhat surprising message – which may be neglected by practitioners and researchers who use S&P 500 options alone to estimate stochastic volatility models – that the joint dynamics of S and V cannot be inferred or calibrated entirely from the S&P 500 option prices. In fact, these options contain only information about the marginal state price density of S. Therefore, the conditional copula of S and V and the marginal state price density of V are determined by their model assumptions. The estimates of those parameters that only appear in the $\mathbb{Q}$ dynamics mainly come from the matching between the marginal risk-neutral density of S and the option price.

Nevertheless, estimating these two densities $p^*(s'|\tau, f, v)$ and $p^*(z'|\tau, f, v)$ in practice is not straightforward since V_t is unobservable. To deal with this problem, we replace the state variable V by the observable Z. That is, we rewrite option prices as $C(\tau, f, z, x, r_f)$ and $H(\tau, f, z, y, r_f)$.[8] Taking second-order derivatives as above, we obtain

$$p^*(s'|\tau, f, z) = e^{r_f\tau + s'}\frac{\partial^2 C(\tau, f, z, x, r_f)}{\partial x^2}\Big|_{x=e^{s'}}$$

$$\text{and} \quad p^*(z'|\tau, f, z) = e^{r_f\tau}\frac{\partial^2 H(\tau, f, z, y, r_f)}{\partial y^2}\Big|_{y=z'}. \tag{3}$$

This change of variable amounts to assuming that the information set generated by V_t and $F_{t,\tau}$ is the same as that by Z_t and $F_{t,\tau}$. In general, the latter set is coarser than the former one. Equating these

[5] The CBOE constructs Z_t from a portfolio of options weighted by strikes according to the following formula:

$$(Z_t/100)^2 = \frac{2e^{r_f\tau}}{\tau}\left(\int_0^{e^{f_{t,\tau}}}\frac{P(\tau, x)}{x^2}dx + \int_{e^{f_{t,\tau}}}^{\infty}\frac{C(\tau, x)}{x^2}dx\right)$$

where $P(\tau, x)$ and $C(\tau, x)$ are put and call options with time-to-maturity τ and strike x, and $f_{t,\tau}$ is the log price of forward contracts, see, e.g., Britten-Jones and Neuberger (2000) and Carr and Wu (2009).

[6] The VIX settlement value is determined by a Special Opening Quotation (SOQ), based on the opening trades of S&P 500 options instead of end-of-day option quotes. However, the SOQ quotes are only available at the monthly frequency on VIX option expiration dates. For parsimony and simplicity, we maintain the use of the VIX as the underlying asset of VIX options following the literature, and attribute the difference to the pricing error of VIX options.

[7] The risk-free interest rate r_f is time-varying despite being deterministic. For convenience and without ambiguity, we omit its dependence on t and τ. The time-t information set $\mathcal{F}_t$ contains stock prices, instantaneous volatility, interest rates and dividends. In the setting with a deterministic interest rate, the information set can be summarized by the log forward price $F_{t,\tau}$ and the volatility V_t.

[8] Strictly speaking, the function $C(\cdot)$ here is a composite function, differing from the previous call option pricing function. We recycle it to simplify our notations.

two effectively assumes that V_t is an invertible function of $F_{t,\tau}$ and Z_t. In practice, this assumption is justifiable, since the VIX is often regarded as a measure of the market risk, or at least proportional to it, despite the difference due to risk premia. In fact, this assumption holds rigorously for most parametric models in the literature (see Appendix A.1 for examples). In view of this, our approach can be regarded as semiparametric.

2.2. Augmenting the information set with multiple volatility factors

In the previous section, we have discussed the setting in which the dynamics of two state variables S and V (or Z) determines both option markets. Nevertheless, it has been well documented, e.g., Egloff et al. (2010), Christoffersen et al. (2009), Bates (2012), and Amengual and Xiu (2014), that at least one additional volatility factor is essential in capturing the time variation of the volatility term structure. We now extend the previous analysis to deal with this case.

We augment the information set with one additional observable volatility factor. In principle, the choice of this factor depends on the assumption and data availability. Since we have adopted a short-term (one-month) volatility-level factor, the VIX, it is natural to include another factor that captures the long-term volatility level. For instance, the CBOE has created the VXV, the three-month volatility index, using a method that is similar to that used for constructing the VIX, but with options of longer maturities. The over-the-counter variance swap contract can also be used, since it is fairly similar to the squared VIX, but has a variety of maturities of up to at least two years. We choose to use the slope of the variance swap term structure, i.e., the relative difference between two variance swaps with two maturities, instead of a long-term volatility level, so as to mitigate the correlation between the two selected factors. Appendix A.2 contains examples of multi-factor volatility models, under which our choice of information set can be justified rigorously.

2.3. From state price densities to pricing kernels

Since the joint state price density is not identifiable, we thereby study the projections of the joint pricing kernel on S_T and Z_T, respectively, denoted as $\pi(s'|\tau,f,z,l)$ and $\pi(z'|\tau,f,z,l)$, where l denotes the realization of the slope factor L_t.[9] We call them the pricing kernel of the market return and the pricing kernel of the VIX. It turns out that they are also the ones that matter for option pricing, since we can write the price of an S&P 500 call option as

$$\begin{aligned} &C(\tau,f,z,l,x,r_f)\\ &\quad= e^{-r_f\tau}\mathbb{E}\left[\pi\cdot(e^{S_T}-x)^+|F_{t,\tau}=f,Z_t=z,L_t=l\right]\\ &\quad= e^{-r_f\tau}\int_{\mathbb{R}}\pi\left(s'|\tau,f,z,l\right)(e^{s'}-x)^+p(s'|\tau,f,z,l)ds', \end{aligned} \tag{4}$$

and the price of a VIX call option as

$$\begin{aligned} &H(\tau,f,z,l,y,r_f)\\ &\quad= e^{-r_f\tau}\mathbb{E}\left[\pi\cdot(z'-y)^+|F_{t,\tau}=f,Z_t=z,L_t=l\right]\\ &\quad= e^{-r_f\tau}\int_{\mathbb{R}}\pi\left(z'|\tau,f,z,l\right)(z'-y)^+p(z'|\tau,f,z,l)dz', \end{aligned} \tag{5}$$

where $p(s'|\tau,f,z,l)$ and $p(z'|\tau,f,z,l)$ are conditional densities of S_T and Z_T under the physical measure. Note that the law of iterated expectation is used to derive the second equalities in (4) and (5).

[9] The projection of π on S_T is defined as $\mathbb{E}^{\mathbb{P}}(\pi|S_T=s',F_{t,\tau}=f,Z_t=z,L_t=l)$.

Moreover, we can rewrite the pricing kernels, by combining the risk-neutral and physical densities.[10] Like (3), Eqs. (4) and (5) imply that the second-order derivatives of the S&P 500 and VIX call prices with respect to their strikes are also equal to $\pi\left(s'|\tau,f,z,l\right)p(s'|\tau,f,z,l)$ and $\pi\left(z'|\tau,f,z,l\right)p(z'|\tau,f,z,l)$, respectively. This fact, combined with (3), further implies that

$$\pi\left(s'|\tau,f,z,l\right)=\frac{p^*(s'|\tau,f,z,l)}{p(s'|\tau,f,z,l)},$$

$$\text{and}\quad \pi\left(z'|\tau,f,z,l\right)=\frac{p^*(z'|\tau,f,z,l)}{p(z'|\tau,f,z,l)}.$$

Therefore, by combining the risk-neutral and physical marginal densities of S and Z, we obtain the projections of π onto S_T and Z_T, respectively. These two pricing kernels bear all the information about how risks are priced in these option markets. In the equilibrium setup of Aït-Sahalia and Lo (2000) with a representative agent, the pricing kernel represents – up to a scale factor – the intertemporal marginal rate of substitution. While Aït-Sahalia and Lo (2000) and Jackwerth (2000) estimate the pricing kernels of the market return, our $\pi\left(s'|\tau,f,z,l\right)$ includes volatility factors in the conditional information set so that volatility becomes relevant to the price of risk regarding the expected returns. In addition, we identify the VIX pricing kernel $\pi\left(z'|\tau,f,z,l\right)$, which reflects how volatility risk is priced in volatility derivative markets.

3. Estimation strategy

3.1. Dimension reduction

In this section, we introduce nonparametric estimation strategies for state price densities. One of the major issues of nonparametric estimation is the curse of dimensionality, i.e., the rate of convergence decreases rapidly as the dimension of state variables increases. As discussed previously, our option pricing functions $C(\cdot)$ and $H(\cdot)$ depend on the time-to-maturity, the S&P 500 forward price, two volatility factors, the strike price, and the interest rate. To facilitate nonparametric estimation, we make some assumptions to reduce the number of state variables.

First, we assume that the option price, once multiplied by $e^{r_f\tau}$, depends on the interest rate and dividend only through the forward price. The same assumption is adopted by Aït-Sahalia and Lo (1998). Furthermore, following many existing studies such as Aït-Sahalia and Lo (1998) and Li and Zhao (2009), we assume that the S&P 500 option price is homogeneous of degree one in the forward price level: $C(\tau,f,z,l,x)=e^fC(\tau,0,z,l,x/e^f)$. Therefore we can define a new function $\overline{C}(\cdot)$, which depends on four state variables:

$$\overline{C}(\tau,z,l,m)=e^{r_f\tau-f}C(\tau,f,z,l,x,r_f), \tag{6}$$

where $m=x/e^f$ represents the moneyness of an option. As a result, we can obtain the nonparametric estimate of $C(\tau,f,z,l,x,r_f)$ by scaling the estimate of $\overline{C}(\tau,z,l,m)$. Moreover, we can write the marginal state price density of S_T using

$$p^*(s'|\tau,f,z,l)=e^{s'-f}\frac{\partial^2\overline{C}(\tau,z,l,m)}{\partial m^2}\Big|_{m=e^{s'-f}}. \tag{7}$$

Furthermore, we can rewrite the density in terms of the excess return $R_{t,\tau}=S_T-F_{t,\tau}$:

$$p^*(r'|\tau,z,l)=e^{r'}\frac{\partial^2\overline{C}(\tau,z,l,m)}{\partial m^2}\Big|_{m=e^{r'}},$$

[10] In the following analysis, our information set is generated from a combination of three state variables, Z_t, $F_{t,\tau}$, and L_t. Since L_T is not involved in the payoffs of either option, our joint pricing kernel does not depend on L_T.

where r' is the realization of $R_{t,\tau}$, since s' and f appear on the right-hand side of (7) only through m. This is in fact also due to the homogeneity assumption, which is satisfied by all parametric models discussed in Appendix A. See Joshi (2007) for additional discussions of the homogeneity.

As for VIX options, we assume that the information about $Z_{t'}$ in $F_{t,\tau}$ is fully incorporated into Z_t and L_t. In other words, conditional on Z_t and L_t, $Z_{t'}$ is independent of $F_{t,\tau}$, for any $t' > t$. This assumption further implies that the marginal state price density of Z_T, obtained from VIX option prices, depends on $F_{t,\tau}$ only through Z_t and L_t, i.e., $p^*(z'|\tau, f, z, l) = p^*(z'|\tau, z, l)$, and that the VIX option price does not depend on $F_{t,\tau}$. Thus, we name a new function $\overline{H}(\cdot)$, which also depends on four state variables:

$$\overline{H}(\tau, z, l, y) = e^{r_f \tau} H(\tau, f, z, l, y, r_f). \tag{8}$$

3.2. Local linear estimation of state price densities

We now construct nonparametric estimators of $\overline{C}$ and $\overline{H}$, and take derivatives to estimate the marginal densities. We adopt this nonparametric regression framework:

$$\overline{C}(\tau, z, l, m) = \mathbb{E}\left(\widetilde{C} \middle| \mathcal{T} = \tau, Z = z, L = l, M = m\right),$$
$$\overline{H}(\tau, z, l, y) = \mathbb{E}\left(\widetilde{H} \middle| \mathcal{T} = \tau, Z = z, L = l, Y = y\right),$$

where $\widetilde{C}$ and $\widetilde{H}$ are observed prices which are contaminated by some pricing errors. This is a cross-sectional regression, as option prices with different strikes and maturities across the sample are stacked together.

As opposed to the multivariate kernel regression approach adopted by Aït-Sahalia and Lo (1998), we prefer the local linear estimator (see, e.g. Fan and Gijbels (1996)) mainly for two reasons. First, the bias and variance of local polynomial estimators are of the same order of magnitude in the interior or near the boundary. This is remarkably different from the Nadaraya–Watson kernel estimator, which suffers from a larger order of bias on the boundary. This boundary effect could be substantial, especially when dealing with multivariate cases. As our empirical studies focus on the tail of pricing kernels, it is advantageous to adopt the local polynomial estimator. Second, local polynomial regression provides estimates of derivatives along with option prices, which makes it more convenient for our purpose.

It is efficient and common to use a local cubic estimator for second-order derivatives, as discussed in Fan and Gijbels (1996). However, among all state variables we are interested only in the second-order partial derivative with respect to the strike, and hence it is cumbersome to have all cross-terms of cubic polynomials included as in the standard local cubic regression. We therefore propose an alternative strategy which applies the local linear estimator to obtain first-order derivatives, and then takes another differentiation of the estimator with respect to the strike so as to estimate the density.

We take the nonparametric density estimation for S&P 500 options as an example. For any fixed $\mathbf{u} = (\tau, z, l, m)^\intercal$ at which we evaluate the densities and option prices, we consider the following minimization problem:

$$\min_{\alpha, \boldsymbol{\beta}} \sum_{i=1}^{n} \left\{\overline{C}_i - \alpha - \boldsymbol{\beta}^\intercal (\mathbf{u}_i - \mathbf{u})\right\}^2 K_h (\mathbf{u}_i - \mathbf{u}) \tag{9}$$

where $\mathbf{u}_i = (\tau_i, z_i, l_i, m_i)^\intercal$ is the characteristic of the ith S&P 500 option in sample, and $\overline{C}_i$ is the scaled option price. They are the data we observe. K_h is a kernel function multiplied by a bandwidth vector $\mathbf{h} = (h_\tau, h_z, h_l, h_m)^\intercal$:

$$K_h(\mathbf{u}_i - \mathbf{u}) = \frac{1}{h_\tau} k\left(\frac{\tau_i - \tau}{h_\tau}\right) \frac{1}{h_z} k\left(\frac{z_i - z}{h_z}\right) \frac{1}{h_l} k \times \left(\frac{l_i - l}{h_l}\right) \frac{1}{h_m} k\left(\frac{m_i - m}{h_m}\right), \tag{10}$$

where $k(\cdot)$ can be chosen, for example, as the density of the standard normal distribution.

The local linear regression approach is very intuitive. By minimizing the objective function, $\alpha + \boldsymbol{\beta}^\intercal(\mathbf{u}_i - \mathbf{u})$ is forced to provide a linear approximation to the function $\overline{C}_i$, in a neighborhood of $\mathbf{u}$, whose size is controlled by the bandwidth $\mathbf{h}$. In view of the Taylor expansion, α is an estimate of $\bar{C}(\mathbf{u})$, and $\boldsymbol{\beta}$ approaches $\partial \bar{C}/\partial \mathbf{u}(\mathbf{u})$. It is worth mentioning that the complexity of the implicit nonparametric function depends on the choice of the bandwidth, see e.g. Fan and Gijbels (1996). A larger bandwidth leads to a smoother estimate (and a simpler function), and a potentially larger bias yet smaller variance.

The solution to (9) has a closed-form representation:

$$\begin{bmatrix} \widehat{\alpha} \\ \widehat{\boldsymbol{\beta}} \end{bmatrix}_{(1+4)\times 1} = (\boldsymbol{\Omega}^\intercal \mathbf{K} \boldsymbol{\Omega})^{-1} \boldsymbol{\Omega}^\intercal \mathbf{K} \overline{\mathbf{C}}, \tag{11}$$

where

$$\boldsymbol{\Omega} = \begin{bmatrix} 1 (\mathbf{u}_1 - \mathbf{u})^\intercal \\ \vdots \quad \vdots \\ 1 (\mathbf{u}_n - \mathbf{u})^\intercal \end{bmatrix}, \qquad \overline{\mathbf{C}} = \begin{bmatrix} \overline{C}_1 \\ \vdots \\ \overline{C}_n \end{bmatrix},$$
$$\mathbf{K} = \begin{bmatrix} K_h (\mathbf{u}_1 - \mathbf{u}) & & \\ & \ddots & \\ & & K_h (\mathbf{u}_n - \mathbf{u}) \end{bmatrix}.$$

Thus, our nonparametric local linear estimator for function $\overline{C}(\mathbf{u})$ is

$$\widehat{\overline{C}}(\tau, z, l, m) = \widehat{\alpha} = \mathbf{e}_1^\intercal (\boldsymbol{\Omega}^\intercal \mathbf{K} \boldsymbol{\Omega})^{-1} \boldsymbol{\Omega}^\intercal \mathbf{K} \overline{\mathbf{C}}, \tag{12}$$

where $\mathbf{e}_1 = (1, 0, 0, 0, 0)^\intercal$. Also, since $\boldsymbol{\beta}$ provides an estimate of all the first-order derivatives, the estimator for the density $p^*(s'|\tau, f, z, l)$ can be derived from another differentiation of the fourth entry of $\boldsymbol{\beta}$ that corresponds to the first-order derivative of $C(\mathbf{u})$ with respect to m:

$$\widehat{p}^*(s'|\tau, f, z, l) = e^{s'-f} \frac{\partial \widehat{\boldsymbol{\beta}}_4}{\partial m}\Big|_{m=e^{s'-f}} = e^{s'-f} \frac{\partial \left(\mathbf{e}_5^\intercal (\boldsymbol{\Omega}^\intercal \mathbf{K} \boldsymbol{\Omega})^{-1} \boldsymbol{\Omega}^\intercal \mathbf{K} \overline{\mathbf{C}}\right)}{\partial m}\Big|_{m=e^{s'-f}}, \tag{13}$$

where $\mathbf{e}_5 = (0, 0, 0, 0, 1)'$. As for the density with respect to the excess return $r' = s' - f$, we have

$$\widehat{p}^*(r'|\tau, z, l) = e^{r'} \frac{\partial \left(\mathbf{e}_5^\intercal (\boldsymbol{\Omega}^\intercal \mathbf{K} \boldsymbol{\Omega})^{-1} \boldsymbol{\Omega}^\intercal \mathbf{K} \overline{\mathbf{C}}\right)}{\partial m}\Big|_{m=e^{r'}}. \tag{14}$$

Note that both $\boldsymbol{\Omega}$ and $\mathbf{K}$ are explicit functions of m, hence the derivative of $\widehat{\beta}_4$ with respect to m has a closed form using simple matrix calculus, see e.g. Magnus and Neudecker (1999). The nonparametric estimator $\widehat{\overline{H}}(\cdot)$ and $\widehat{p}^*(z'|\tau, z, l)$ can be constructed similarly.

It may be worth pointing out that the other entries of $\widehat{\boldsymbol{\beta}}$ in our local linear regression (11) are related to option Greeks, which could be useful for other applications, such as delta-hedging.

3.3. Estimation of pricing kernels

Since we have estimated the state price density from option prices, in order to obtain pricing kernels, we need to recover the physical density $p(r'|\tau, z, l)$ from the historical time series. We adopt the local linear method again. A similar strategy has been used by Aït-Sahalia et al. (2009). We collect time series of $(r_{t_i,\tau}, z_{t_i}, l_{t_i})$, $i = 1, 2, \ldots, k$, where $r_{t_i,\tau} = s_{t_i+\tau} - f_{t_i,\tau}$ with τ fixed. To estimate $p(r'|\tau, z, l)$, we minimize:

$$\min_{\gamma,\eta_z,\eta_l} \sum_{i=1}^{k} \left\{ K_{b_r}(r_{t_i,\tau} - r') - \gamma - \eta_z (z_{t_i} - z) - \eta_l (l_{t_i} - l) \right\}^2 \times K_{b_z}(z_{t_i} - z) K_{b_l}(l_{t_i} - l),$$

where b_r, b_z and b_l are bandwidths, and $K_b(\cdot) = 1/b \cdot k(\cdot/b)$ with $k(\cdot)$ being a kernel function.

To illustrate the intuition behind this estimator, we translate the conditional density estimation problem into another nonparametric regression problem because:

$$\mathbb{E}\left(K_{b_r}(R_{t,\tau} - r')|Z_t = z, L_t = l\right) \approx p(r'|\tau, z, l), \quad \text{as } b_r \to 0. \tag{15}$$

Therefore, following the same Taylor expansion idea, the density estimator is given by,

$$\widehat{p}(r'|\tau, z, l) = \widehat{\gamma}. \tag{16}$$

Consequently, our pricing kernel estimator can be constructed as

$$\widehat{\pi}(r'|\tau, z, l) = \frac{\widehat{p}^*(r'|\tau, z, l)}{\widehat{p}(r'|\tau, z, l)}.$$

Similarly, we can estimate the physical density of the VIX and hence its pricing kernel by constructing a time series sample of $(z_{t_i+\tau}, z_{t_i}, l_{t_i})$, for $i = 1, 2, \ldots, k$.

3.4. Asymptotic theory

To provide theoretical guidance for our approach, we derive the asymptotic distribution of the option price $\overline{C}(\tau, z, l, m)$ and density $p^*(s'|\tau, f, z, l)$ for S&P 500 options as an example. Suppose the sample size of the S&P 500 options is n. Using the equivalent kernels introduced in Fan and Gijbels (1996) and following the derivation in Ruppert and Wand (1994) and Aït-Sahalia and Lo (1998), we obtain:

$$n^{1/2}(h_\tau h_z h_l h_m)^{1/2}\left(\widehat{\overline{C}}(\tau, z, l, m) - \overline{C}(\tau, z, l, m)\right) \tag{17}$$
$$\xrightarrow{d} N\left(0, \left[\int k^2(c)\,dc\right]^3 s^2(\tau, z, l, m)/\pi(\tau, z, l, m)\right),$$
$$\text{as } nh_\tau h_z h_l h_m \longrightarrow \infty;$$

$$n^{1/2}h_m^2(h_\tau h_z h_l h_m)^{1/2}\left(\widehat{p}^*(r'|\tau, z, l) - p^*(r'|\tau, z, l)\right) \tag{18}$$
$$\xrightarrow{d} N\left(0, m^2\left[\int k^2(c)\,dc\right]^3 \times \left[\int \left(c\dot{k}(c) + k(c)\right)^2 dc\right] \Big/ \left[\int k(c)\,c^2 dc\right]^2 \times s^2(\tau, z, l, m')/\pi(\tau, z, l, m')\right),$$

as $nh_\tau h_z h_l h_m^5 \longrightarrow \infty$, where $m = e^{x-f}$, and $m' = e^{s'-f}$,

where $s^2(\tau, z, l, m)$ is the conditional variance for the local linear regression of $\overline{C}$ on the state variables, and $\pi(\tau, z, l, m)$ is the joint density of these variables. The estimator for $s^2(\cdot)$ can be constructed using similar nonparametric regressions of squared fitting errors on these state variables. The same asymptotic distributions apply to estimators for VIX option prices and their densities.

In addition, the asymptotic theory of the physical conditional density estimator (16) is given by:

$$k^{1/2}(b_r b_z b_l)^{1/2}\left(\widehat{p}(r'|\tau, z, l) - p(r'|\tau, z, l)\right) \xrightarrow{d} N\left(0, \left[\int k^2(c)dc\right]^3 p(r'|\tau, z, l)/\pi(z, l)\right),$$

as $kb_r b_z b_l \to \infty$ (see, e.g., Fan et al. (1996)). Certain regularity conditions are needed to establish this central limit result, such as stationarity and mixing properties of the process $(r_{t,\tau}, z_t, l_t)$. See Conditions 1–6 in the appendix of Aït-Sahalia et al. (2009) for references.

To obtain a valid asymptotic variance for the marginal pricing kernels, which is our main target in the empirical section, we use the joint independent normality between the time-series regression for physical density and the cross-sectional regression for the risk-neutral density, and apply the "delta" method to find the standard error for their ratio. More specifically, the asymptotic variance of the $\widehat{\pi}(r'|\tau, z, l)$ is given by

$$\widehat{\text{Avar}}(\widehat{\pi}(r'|\tau, z, l)) = \frac{1}{(\widehat{p}(r'|\tau, z, l))^2}\widehat{\text{Avar}}(\widehat{p}^*(r'|\tau, z, l)) + \frac{(\widehat{p}^*(r'|\tau, z, l))^2}{(\widehat{p}(r'|\tau, z, l))^4}\widehat{\text{Avar}}(\widehat{p}(r'|\tau, z, l)).$$

This approach is more conservative, as opposed to the approach by Aït-Sahalia and Lo (1998), in which they suggest taking into account only the estimation error from the risk-neutral density, due to a higher convergence rate of the physical density estimator.

3.5. Bandwidth selection

Bandwidth selection is important especially for multivariate nonparametric regressions. In theory, the optimal rate of bandwidth for estimating the pricing function is $n^{-1/(4+d)}$, whereas to estimate densities, we need to adopt a bandwidth with rate $n^{-1/(6+d)}$ due to the curse of differentiation. These bandwidth choices ensure that the nonparametric pricing function achieves the optimal rate of convergence in the mean-squared sense. Empirically, we can choose a bandwidth h_j ($j = \tau, z, l$, and m for S&P 500 options, or y for VIX options) as $h_j = c_j \sigma_j n^{-1/(4+d+2\nu)}$, where σ_j is the unconditional standard deviation of the regressor j, d is the number of regressors, and $\nu = 0$ and 1 for the pricing function or the density, respectively. The constant c_j is chosen by minimizing the mean-squared error of option prices via cross-validation. The cross-validation objective function for regression (12) is given by the weighted mean-squared errors:

$$\min_h \frac{1}{n}\sum_{i=1}^{n}\left(\overline{C}_i - \widehat{\overline{C}}_{h,-i}(\tau_i, z_i, l_i, m_i)\right)^2 \omega(\tau_i, z_i, l_i, m_i),$$

where $-i$ means leaving the ith observation out, and $\omega(\cdot)$ is the weighting function. To further accelerate the cross-validation, we adopt the popular K-fold cross-validation, which is faster than this leave-one-out method (Hastie et al., 2001).

The bandwidths of our nonparametric conditional density estimator (16) under the physical measure are chosen by the cross-validation following Fan and Yim (2004):

$$\min_b \frac{1}{k}\sum_{i=1}^{k}\omega(r_{t_i,\tau}, z_{t_i}, l_{t_i})\int(\widehat{p}_b(r'|\tau, z_{t_i}, l_{t_i}))^2 dr' - \frac{2}{k}\sum_{i=1}^{k}\widehat{p}_{b,-i}(r_{t_i,\tau}|\tau, z_{t_i}, l_{t_i})\omega(r_{t_i,\tau}, z_{t_i}, l_{t_i}),$$

where the first integral can be calculated in a closed form from (16). Alternative choices of bandwidths have been discussed in Yao and Tong (1998) and Ruppert et al. (1995).

3.6. Monte Carlo simulations

Here we provide simulation studies of our local linear estimators. The Monte Carlo experiments are designed to match our empirical study. First, we select a sample path generated from the multi-factor stochastic volatility model with both jumps in volatility and prices as discussed in Section 5.1. The parameters are taken from our estimates reported in Table 3. We then calculate the VIX, the variance swap rates, and S&P 500 and VIX option prices, according to the closed-form formulae given in Appendices D and F. The simulated cross-section of options has the same characteristics as those traded on the CBOE. The number of observations for each contract matches exactly the volume of the real data we have, see Section 4.1.

To estimate the risk-neutral densities, we use the cross-section of option prices, which are polluted by multiplicative measurement errors that follow log-normal distribution with a 5% standard deviation. We conduct the simulation 1000 times, by repeatedly generating the cross-section of pricing errors with the sample path fixed. As to the physical densities, for each sample path simulated from the parametric model, we build a time series of observations 42 days apart, so as to estimate the conditional densities. The procedure is repeated for 1000 sample paths generated from the parametric model.

We compare in Fig. 1 the nonparametric densities averaged over 1000 replications, with their true values, calculated using the parametric model. We also compare the pricing kernel estimates and pricing errors. We report these estimates conditional on $Z_t = 18.60$ and $L_t = 0.17$, which correspond to $V_t = 0.03$ and $\xi_t = 0.02$ given our parametric model. The time to maturity is $\tau = T - t = 42$ days, which matches our empirical study. The results for other conditional values and different times to maturity are the same, and hence are omitted here.

We observe from the bottom two panels that the average of the estimated implied volatilities for S&P 500 options and VIX option prices are almost identical with their true values. The top left and middle left panels show that our nonparametric density and kernel estimators also perform fairly well. The true pricing kernel is a bit bumpy towards the very right tail so we truncate it off, because, as can be seen from the top left panel of density estimates, both risk-neutral and physical densities are close to zero in that region, so that the pricing kernel as a ratio of the two exhibits some idiosyncrasies. From the top right and middle right panels, we observe that the VIX density and kernel estimators are also close to the true, though the errors are larger than their counterparts for S&P 500 options. This is because there are more observations of S&P 500 options. Moreover, the changes in the consecutive strikes of VIX options account for a much larger percentage of the VIX, compared to the consecutive strikes of S&P 500 options, which also explains the inferior performance of their risk-neutral density estimators.

4. Empirical results

In this section, we estimate nonparametric pricing kernels using both S&P 500 and VIX options, and present our empirical findings. Before delving into the details, we introduce the data.

4.1. Data

We obtain daily close bid and offer prices of S&P 500 and VIX options from the OptionMetrics. Our sample period is January 4, 1996 through December 31, 2012 for S&P 500 index options, and from July 1, 2007 through December 31, 2012 for VIX options. While VIX options are available from February 2006, we discard earlier ones due to liquidity concerns. We also obtain the VIX from the CBOE and 3-month and 12-month variance swap rates from an anonymous bank from January 4, 1996 through December 31, 2012.[11]

We follow the data-cleaning routine commonly used in the literature; see, e.g., Aït-Sahalia and Lo (1998). First, observations with bid or ask prices smaller than \$0.025 are eliminated to mitigate the effect of decimalization. We use the mid-quote as the observed option price. Moreover, we eliminate any options with zero open interests or trading volumes. In addition, we only consider out-of-the-money (OTM) S&P 500 options with maturity between 7 and 252 days. In estimation, we use the put–call parity to convert OTM puts into in-the-money (ITM) calls as it is well known that in-the-money S&P 500 options are less liquid than out-of-the-money options. For VIX options, we restrict the maturity to be between 7 and 126 days. VIX calls have a higher trading volumes than the puts, so we consider only VIX call options in the estimation. The last step is to eliminate option contracts that violate no-arbitrage conditions. After cleaning, there are in total 715,549 S&P 500 out-of-the-money options over 4,263 days, and 122,139 VIX call options over 1,376 days left for our nonparametric analysis.

The top panels of Fig. 2 plot the time series of the S&P 500 index and the VIX over 1996–2012, while the bottom panels plot time series of the 3-month and 12-month variance swap rates and the slope of volatility term structure defined as the ratio of the 12-month and 3-month variance swap rates minus 1. We observe a prevailing negative relationship between the S&P 500 index returns and changes of the VIX from Fig. 2, consistent with the literature on leverage effect. We also observe that the slope of the volatility term structure varies significantly over time, although the two variance swap rates track each other closely, and they have a pattern similar to the VIX.

Table 1 provides summary statistics of these series. We observe that the VIX ranges between 9.89 and 80.86, which is large enough to have both relatively low and high volatility levels. Moreover, the slope of the volatility term structure reaches both negative and positive values in the sample period of 1996–2012. As documented in, e.g., Amengual and Xiu (2012), negative values of the slope signal stressful times.

Finally, Table 2 presents summary statistics of options, including the total volume and open interest in millions of contracts, and the mean, median, minimum, and maximum of the moneyness, time-to-maturity, and implied volatility (or option prices) for the cleaned S&P 500 index and VIX options.

The volume and open interest of VIX call options are roughly twice those of VIX puts, suggesting that VIX calls are more liquid than VIX puts. This is not surprising as VIX call options can be used to hedge against the upside volatility movement and the downside market risk. We thereby use VIX call options for the empirical study.

For both S&P 500 and VIX options, the range of moneyness is wide, covering both deep out-of-the-money options, which is important for recovering the tails of the risk-neutral densities. The median time-to-maturity is about 40 days, and we hence mainly use 42 days (six calendar weeks) in our empirical study. In addition, the implied volatilities of S&P 500 options and prices of VIX options have a wide range, because of the recent financial crisis covered in our sample.

[11] We take the square-root of the variance swap rates and multiply them by 10, so that they are of the same scale as the VIX. Without ambiguity, hereafter our variance swap rates are scaled.

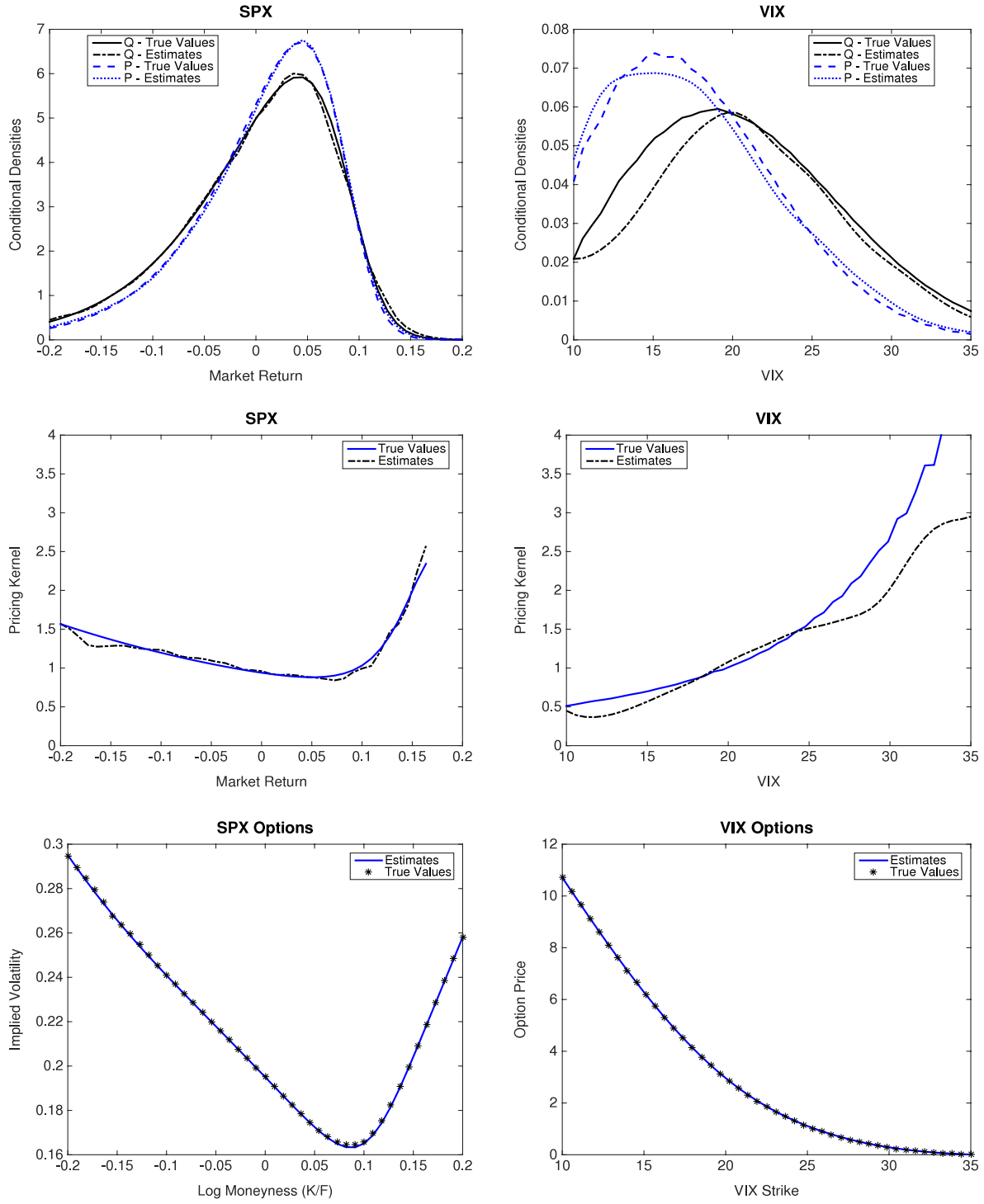

Fig. 1. Monte Carlo simulations. Note: This figure plots the average of nonparametric estimates of the (risk-neutral and physical) densities (top), pricing kernels (middle), and option prices (bottom) along with their true values in Monte Carlo simulations. The left panels are for S&P 500 index options and the right panels are for VIX options. The number of Monte Carlo samples is 1000.

Table 1
Summary statistics of the S&P 500 index and VIX.

	Mean	Std	Skew	Kurt	Min	25%	75%	Max
Index	1154.18	222.27	−0.50	2.54	598.48	1007.30	1326.80	1565.15
Return	0.020	1.30	−0.27	10.18	−9.47	−0.58	0.64	10.42
VIX	22.00	8.47	1.93	9.62	9.89	16.41	25.24	80.86
Slope	0.03	0.08	0.04	3.54	−0.31	−0.02	0.09	0.38

Note: This table reports summary statistics of the daily S&P 500 index, the S&P 500 index return, the VIX, and the slope of the volatility term structure defined as the ratio of 12-month and 3-month variance swap rates minus 1. The kurtosis is the total (rather than excess) kurtosis. The sample period is from January 4, 1996 through December 31, 2012.

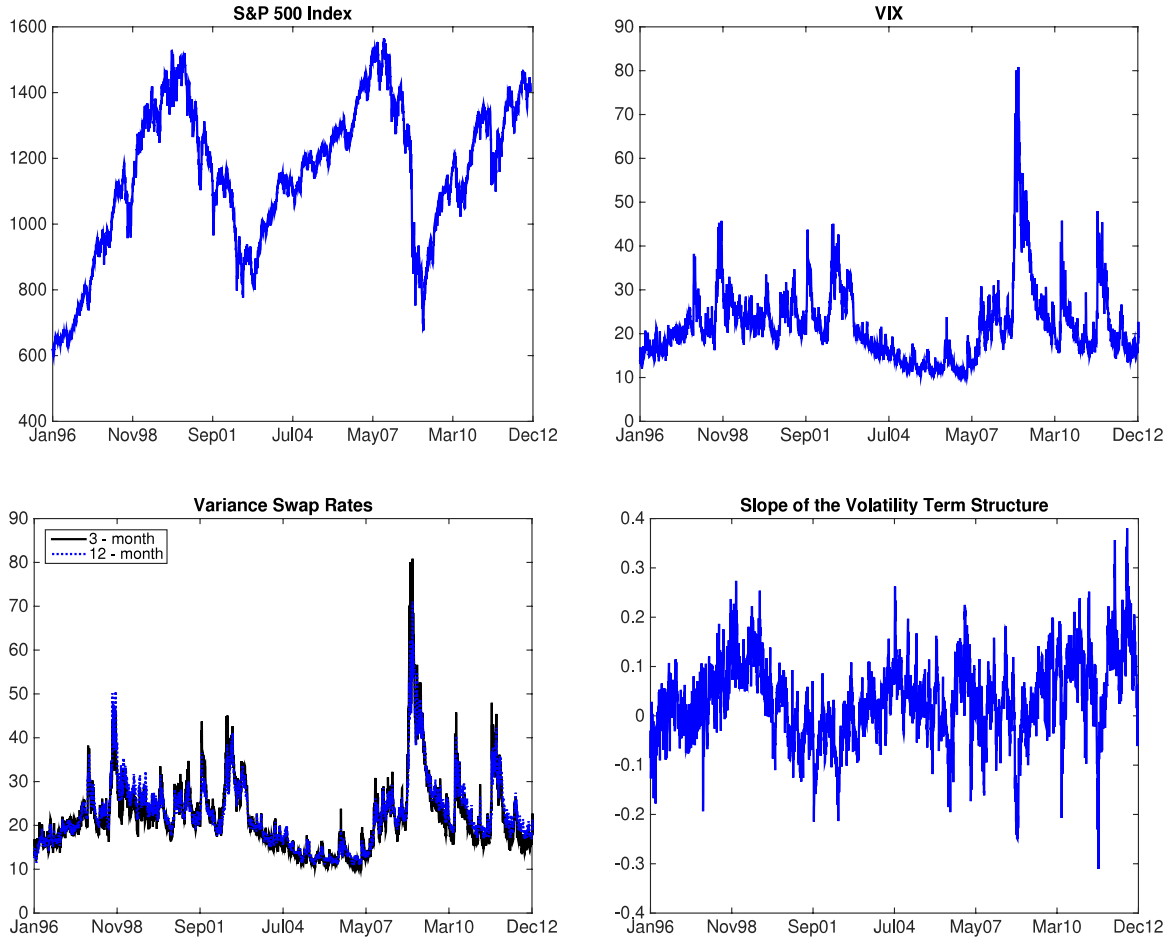

Fig. 2. Time series of the S&P 500 Index, VIX, and variance swap rates. Note: The top panels plot time series of the S&P 500 index and VIX, while the bottom panels plot time series of the 3-month and 12-month variance swap rates, transformed to the same scale of the VIX, and the slope defined as the ratio of the transformed 12-month and 3-month variance swap rates minus 1. The sample period is from January 4, 1996 through December 31, 2012.

Table 2
Summary statistics of S&P 500 and VIX options.

		SPO				VXO	
		OTM Call	OTM Put			Call	Put
Volume		317.11	629.75	Volume		192.66	99.66
Open Interest		4010.86	8586.76	Open Interest		2164.82	1176.18
Log Moneyness	Mean	0.08	−0.17	Log Moneyness	Mean	1.45	1.12
	Median	0.06	−0.11		Median	1.29	1.04
	Min	0	−3.18		Min	0.12	0.12
	Max	0.95	0		Max	7.01	7.01
Time-to-Maturity	Mean	59	61	Time-to-Maturity	Mean	53	48
	Median	38	39		Median	48	41
	Min	7	7		Min	7	7
	Max	252	252		Max	126	126
Implied Volatility	Mean	0.19	0.3	Option Price	Mean	3.52	4.92
	Median	0.17	0.26		Median	2.15	2.3
	Min	0.06	0.08		Min	0.05	0.05
	Max	1.6	3.24		Max	58.5	84.6

Note: This table reports summary statistics for selected S&P 500 options from January 4, 1996 to December 31, 2012 and VIX options from July 1, 2007 to December 31, 2012, including the total trading volume and open interest in millions of contracts and the mean, median, minimum, and maximum of the log moneyness log(K/F), time-to-maturity in days, and implied volatility (prices) for S&P 500 (VIX) options. We include only out-of-the-money S&P 500 options, because of their better liquidity than in-the-money ones. The put options are later converted to calls using the put–call parity. While we provide both puts and calls for VIX options, only the calls are used in the empirical study.

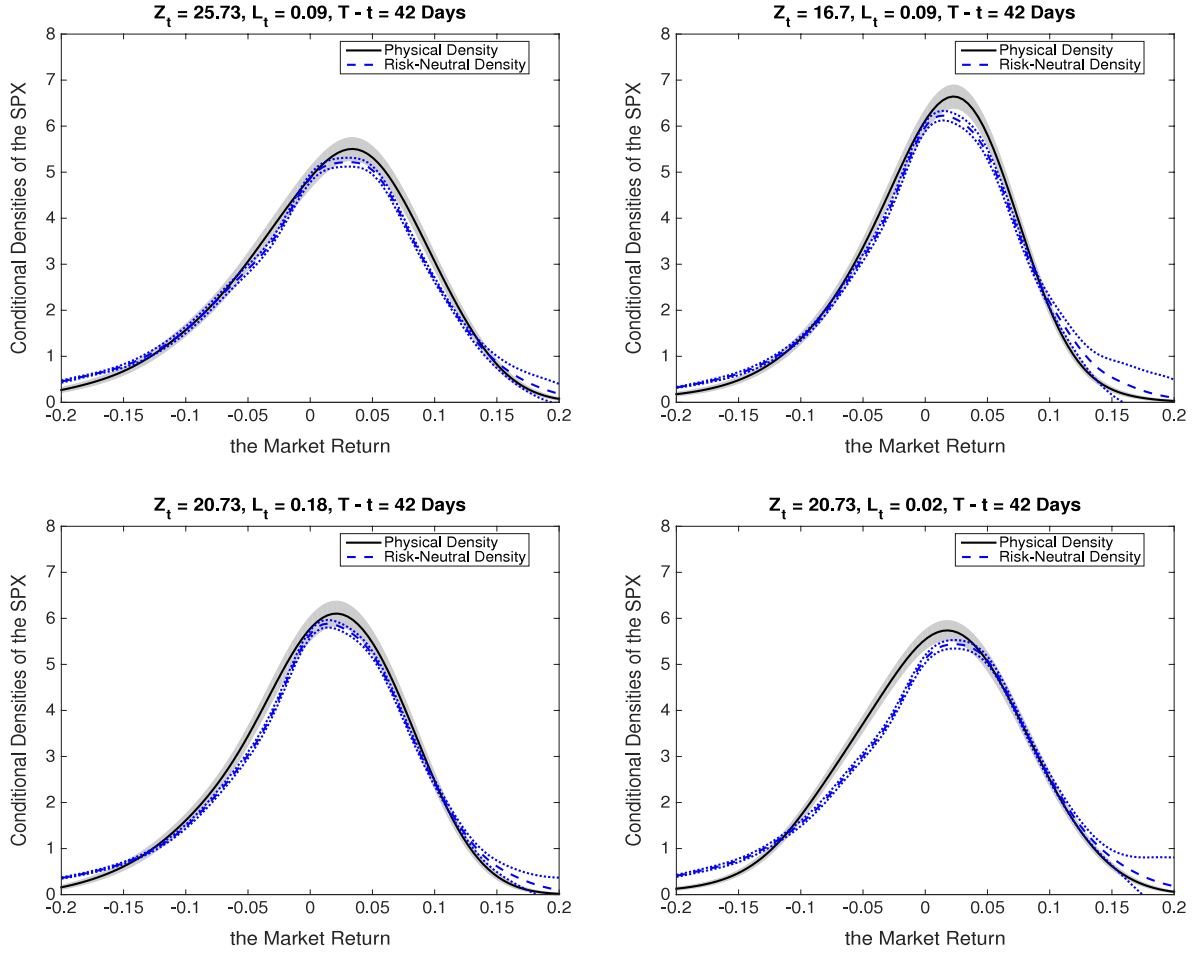

Fig. 3. Conditional marginal densities of the market return. Note: The top panels provide nonparametric estimates of marginal risk-neutral and physical densities of the market return with time-to-maturity at 42 days, L_t at the median level of its time series, and Z_t at the 75% (top left panel) and 25% (top right panel) quantiles. The bottom panels provide the nonparametric estimates of marginal densities of the market return at maturity of 42 days, with Z_t set at the median level and L_t at the 75% (bottom left panel) and 25% (bottom right panel) quantiles. The sample period ranges from January 4, 1996 through December 31, 2012.

4.2. Pricing kernels of the market return

The upper panels of Fig. 3 provide nonparametric risk-neutral and physical density estimates of the market return, conditional on a median level of the slope, and both low and high levels of VIX at the 25% and 75% quantiles of the VIX column in the S&P 500 option sample, respectively. The lower panels present the densities conditional on a median level of VIX and both low and high levels of the slope L_t.[12] We find that the risk-neutral and physical densities depend on both Z_t and L_t. In particular, the densities are more dispersed conditional on a higher Z_t or a lower L_t, both of which signal stressful times. Moreover, the risk-neutral density exhibits a heavier left tail than the physical density across all figures, and the gap seems to be larger conditional on a higher Z_t or a lower L_t. On the right sides of the tails, the differences between the risk-neutral and physical densities are not as clear.

Fig. 4 provides corresponding nonparametric estimates of the pricing kernel of the market return conditional on the VIX and the slope of the volatility term structure. For comparison, we also provide unconditional pricing kernel estimates following Aït-Sahalia and Lo (1998). These estimates reveal several important properties of the pricing kernel and volatility risk.

First, our estimates of conditional pricing kernels differ significantly from the unconditional ones, implying that volatilities are important factors driving pricing kernels. Missing volatility factors seems to cause overestimation of the tails of the pricing kernel of market returns. Moreover, when L_t is low, the pricing kernel is clearly higher than when L_t is high, which agrees with the fact that lower L_t signals stressful time periods. When L_t is fixed at its median, there is not a clear-cut significant difference.

Second, our conditional pricing kernel estimates exhibit a downward sloping shape, in contrast to the U-shape of the unconditional pricing kernels for a wide range of positive market returns up to 10%–15% (depending on the volatility level). This suggests that conditioning on volatility factors helps explain the "pricing kernel puzzle." As first documented in Aït-Sahalia and Lo (1998) and Jackwerth (2000), a U-shape of the pricing kernel estimate – the right tail in particular – implies that the (stochastic) discount factor is increasing when the market return is positive and increasing, inconsistent with standard asset pricing theories. Our estimates of conditional pricing kernels show that conditioning on volatility factors delivers a pricing kernel with a more sharply downward sloping shape within certain threshold where we have sufficient statistical power.

Finally, our conditional pricing kernel estimates still exhibit a bit of a U-shape on the extreme end of the right tail, but the pattern is much less significant than that of the AL unconditional pricing kernel estimates. More importantly, the higher values on the right end are associated with extremely wide confidence bands, since

[12] It appears that the standard errors of the physical density estimates are small. This is because we use 16 years of *daily* SPX returns with more than 4000 observations. The standard errors of the VIX density estimates are larger, because we use 6 years of the VIX series.

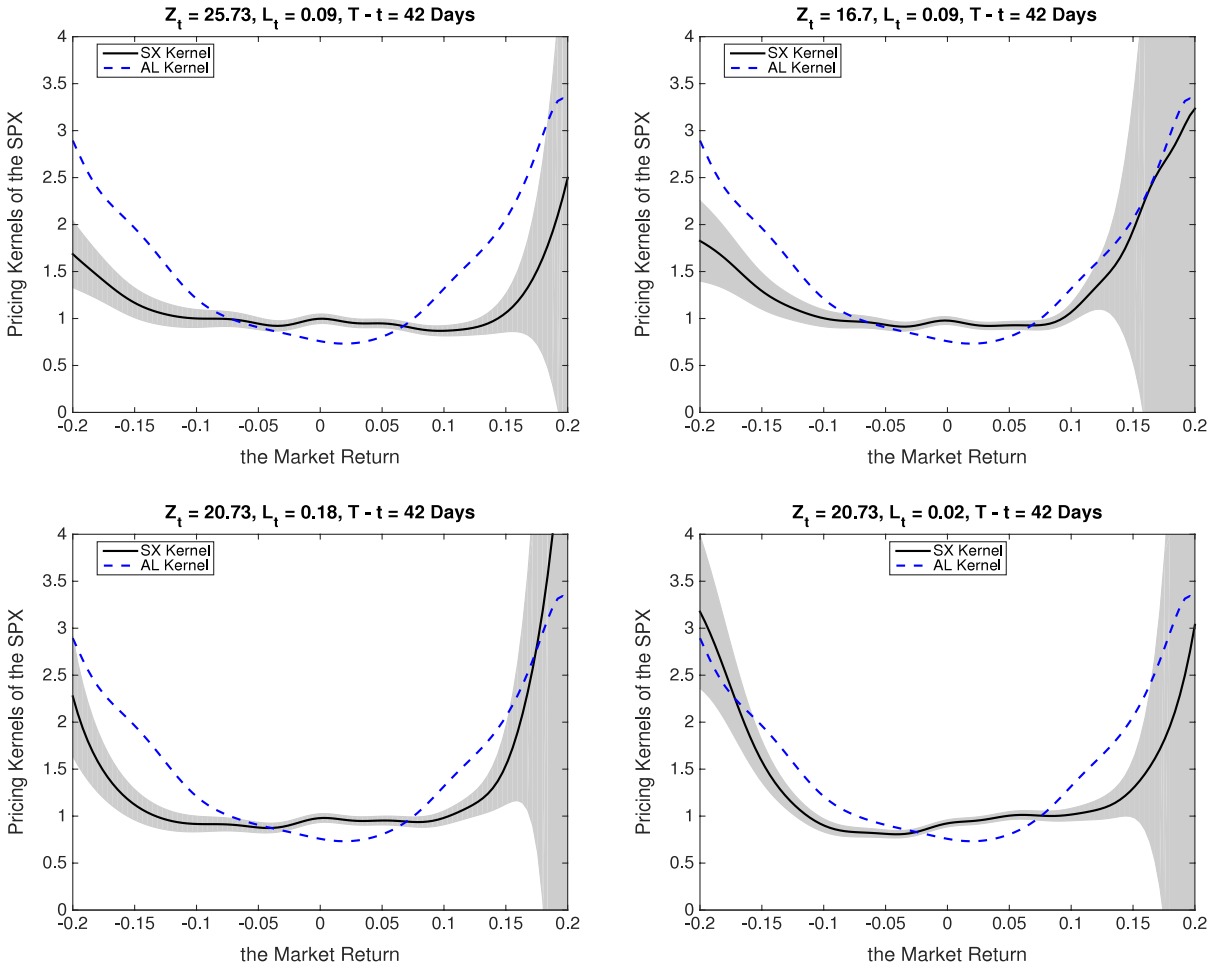

Fig. 4. Pricing kernels of the market return. Note: The top panels provide our nonparametric estimates (SX) of the marginal pricing kernel of the S&P 500 index return at maturity of 42 days, with L_t set at the median level while Z_t at the 75% (top left panel) and 25% (top right panel) quantiles. The bottom panels provide the nonparametric estimates (SX) of the marginal pricing kernel of the S&P 500 index return at maturity of 42 days, with Z_t set at the median level while L_t at the 75% (bottom left panel) and 25% (bottom right panel) quantiles. The unconditional nonparametric pricing kernel estimate following Aït-Sahalia and Lo (1998) (AL) is also reported for comparison. The gray areas are the 95% confidence intervals constructed by the asymptotic distribution theory in (18). The sample period is January 4, 1996–December 31, 2012.

there are very few extremely positive returns realized in particular when volatility is low, and we do not have strong evidence to reach any conclusion. We hence argue that the U-shape puzzle around extremely positive returns is a statistical artifact due to limited data.

4.3. Pricing kernels of the VIX

We study the pricing kernels of volatility in this section. The top panels of Fig. 5 provide nonparametric risk-neutral and physical density estimates of the VIX conditional on a median level of L_t for both low and high levels of Z_t at its 25% and 75% quantiles, respectively. The lower panels present the same estimates conditional on high and low values of L_t and a median value of Z_t. We observe that both the risk-neutral and physical densities of the VIX are positively skewed, suggesting a large risk of volatility spikes. Moreover, both the left and right tails of risk-neutral densities are rather heavier than those of the physical densities, which is different from the S&P 500 case where the gap exists mainly on the left tail.

Fig. 6 provides the corresponding nonparametric estimates of the volatility pricing kernel conditional on the VIX and the slope. We make several observations here. First, our estimates show that the volatility pricing kernel is strongly state dependent. In particular, from the top two panels, the volatility pricing kernel is significantly higher conditional on a high Z_t than a low Z_t on the left tail. From the bottom two panels, the volatility pricing kernel is significantly higher conditional on a low L_t than a high L_t mainly on the right tail. These two observations are consistent with the recently proposed asset pricing models with volatility risk such as Bansal and Yaron (2004), Drechsler and Yaron (2011), Bollerslev et al. (2009), Zhou and Zhu (2012), and Branger and Völkert (2012): the stochastic discount factor regarding volatility risk is higher in stressful times that usually experience a high Z_t and low L_t. Without imposing any parametric restrictions, we provide model-free evidence on how the volatility pricing kernel is state dependent.

Second, and more importantly, we observe that the volatility pricing kernel exhibits a pronounced U-shape, in particular on the left end. In other words, investors attach high marginal utility to payoffs received when the future volatility is low, which usually means a good state of the economy. It presents empirical challenges to the asset pricing models of Bansal and Yaron (2004), Drechsler and Yaron (2011), Bollerslev et al. (2009), Zhou and Zhu (2012), and Branger and Völkert (2012), in which the volatility kernel is monotonically increasing. Bakshi et al. (forthcoming) on the other hand, provide some evidence of a U-shape volatility pricing kernel, by exploring the link between the monotonicity of the pricing kernel and returns of VIX option portfolios. They also provide a model with heterogeneity in beliefs to account for the U-shape, in which the volatility market is dominated by investors with zero market risk. Our nonparametric estimates conform with their theoretical predictions.

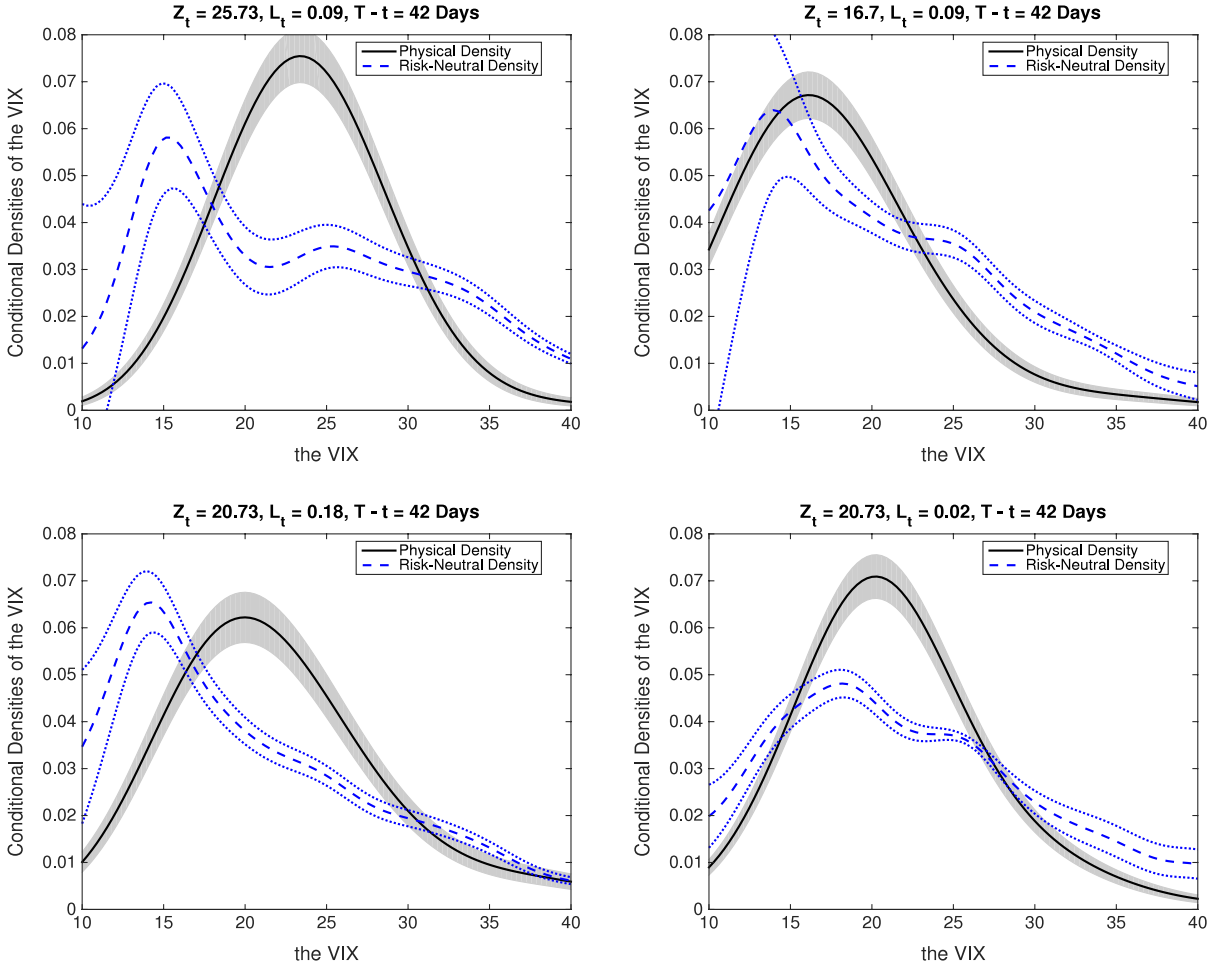

Fig. 5. Conditional marginal densities of the VIX. Note: The top panels provide our nonparametric estimates of marginal risk-neutral and physical densities of VIX at maturity of 42 days, with L_t set at the median level while Z_t at the 75% (top left panel) and 25% (top right panel) quantiles. The bottom panels provide the nonparametric estimates of marginal densities of VIX at maturity of 42 days, with Z_t set at the median level while L_t at the 75% (bottom left panel) and 25% (bottom right panel) quantiles. The sample period is July 1, 2007–December 31, 2012.

5. The joint pricing kernel: a parametric analysis

5.1. The parametric model

To study the joint pricing kernel derived from the S&P 500 and VIX option markets, more parametric assumptions are essential, since the joint state price densities are not identified nonparametrically. We consider a full-fledged parametric model with two volatility factors.[13] Specifically, the risk-neutral dynamics follows:

$$dS_t = \left(r_f - d - \frac{1}{2}V_t\right)dt + \sqrt{V_t}dW_t^{\mathbb{Q}} + J_S^{\mathbb{Q}}dN_t - \mu\lambda_t dt,$$
$$dV_t = (\eta + \kappa(\xi_t - V_t))dt + \sigma\sqrt{V_t}dB_t^{\mathbb{Q}} + J_V^{\mathbb{Q}}dN_t,$$
$$d\xi_t = \alpha(\theta - \xi_t)dt + \gamma\sqrt{\xi_t}dM_t^{\mathbb{Q}},$$

where $W_t^{\mathbb{Q}}$ and $B_t^{\mathbb{Q}}$ are standard Brownian motions with a correlation ρ, $M_t^{\mathbb{Q}}$ is another independent Brownian motion, $J_S^{\mathbb{Q}}$ and $J_V^{\mathbb{Q}}$ are random jump sizes, dN_t is a pure-jump process with intensity $\lambda_t = \lambda_0 + \lambda_1 V_t$, and $\mu = \mathbb{E}^{\mathbb{Q}}(e^{J_S^{\mathbb{Q}}} - 1)$. The jump sizes follow:

$$J_V^{\mathbb{Q}} \sim \exp(\beta_V), \qquad J_S^{\mathbb{Q}} \sim \begin{cases} \exp(\beta_+) & \text{with probability } q \\ - \exp(\beta_-) & \text{with probability } 1-q. \end{cases}$$

The pricing formulae for S&P 500 and VIX options are given in the Appendix.

To study the pricing kernel, we also specify the dynamics under the historical measure $\mathbb{P}$:

$$dS_t = (\mu_0 + \mu_1 V_t)dt + \sqrt{V_t}dW_t^{\mathbb{P}} + J_S^{\mathbb{P}}dN_t,$$
$$dV_t = (\eta + \kappa\xi_t - \kappa^{\mathbb{P}}V_t)dt + \sigma\sqrt{V_t}dB_t^{\mathbb{P}} + J_V^{\mathbb{P}}dN_t,$$
$$d\xi_t = (\alpha\theta - \alpha^{\mathbb{P}}\xi_t)dt + \gamma\sqrt{\xi_t}dM_t^{\mathbb{P}},$$

where the jump sizes follow:

$$J_V^{\mathbb{P}} \sim \exp(\beta_V^{\mathbb{P}}), \qquad J_S^{\mathbb{P}} \sim \begin{cases} \exp(\beta_+^{\mathbb{P}}) & \text{with probability } q \\ - \exp(\beta_-^{\mathbb{P}}) & \text{with probability } 1-q. \end{cases}$$

This model nests many existing parametric models in the literature, e.g., Bakshi et al. (1997), Bates (2000), Pan (2002), Chernov and Ghysels (2000), Eraker (2004), and Broadie et al. (2007). Moreover, two volatility factors have been shown by, e.g., Egloff et al. (2010) and Mencia and Sentana (2012), to capture the volatility term structure effectively. While multi-factor stochastic models have been shown to match S&P 500 options well, their performance of pricing VIX derivatives is rarely investigated in the empirical derivative pricing literature. There are a couple of exceptions. Papanicolaou and Sircar (2014) find that a regime-switching Heston model bridges the shortcomings of the basic Heston model,

[13] We thank an anonymous referee for suggesting the study of the joint pricing kernel.

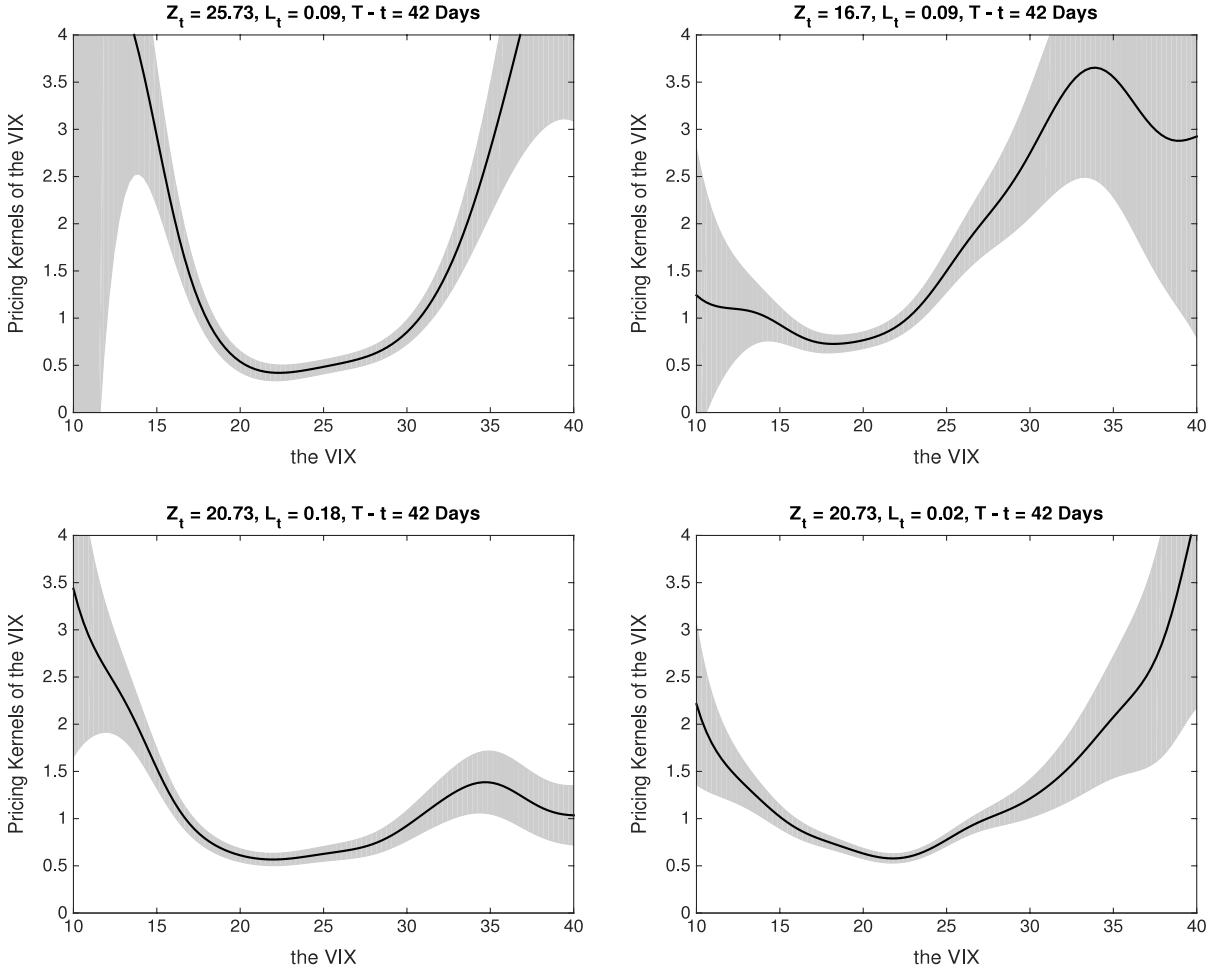

Fig. 6. Pricing kernels of the VIX. Note: The top panels provide our nonparametric estimates (SX) of the marginal pricing kernel of the VIX at maturity of 42 days, with L_t set at the median level of its time series while Z_t at the 75% (top left panel) and 25% (top right panel) quantiles. The bottom panels provide the nonparametric estimates (SX) of the marginal pricing kernel of the VIX at maturity of 42 days, with Z_t set at the median level while L_t at the 75% (bottom left panel) and 25% (bottom right panel) quantiles. The gray areas are the 95% confidence intervals constructed by the asymptotic distribution theory in (18). The sample period is July 1, 2007–December 31, 2012.

which cannot capture the implied volatilities of VIX options. Using a similar model to ours, Bardgett et al. (2014) find that the S&P 500 and VIX derivatives markets contain conflicting information on the variance, especially during market distress. See Gatheral (2006) for further reference. Differently, our focus here is on pricing kernels.

We derive, in Appendix E, closed-form formulae of marginal and joint conditional densities under both the risk-neutral and physical measures, respectively. We also develop closed-form pricing formulae for S&P 500 and VIX options in Appendix F, which are used in our estimation procedure detailed in Section 5.2.

5.2. Estimation of the parametric model

The parametric model is estimated using particle filters. One of the first and most used particle-filtering algorithm is known as the sampling importance resampling (SIR) method. The use of this method has been suggested by Berzuini et al. (1997), Gordon et al. (1993), and Kitagawa (1996). Despite its sample impoverishment problem, which could be fixed using the auxiliary particle filtering of Pitt and Shephard (1999), the SIR algorithm is very convenient when incorporating option prices. Moreover, it performs well in most settings, as discussed in Johannes and Polson (2009) and Johannes et al. (2009). We thereby adopt the SIR filtering algorithm combined with the maximum likelihood estimation, similar to Christoffersen et al. (2010) and Doucet et al. (2014).

The procedure relies on the following Euler approximation of the dynamics under the physical measure. The discretization error can be ignored when sampling frequency is small, as discussed in Johannes and Polson (2009). Let $y_{t+1} = (S_{t+1} - S_t, C_{t+1}, H_{t+1}, Z_{t+1})$, which contains the available information we observe. Suppose ϵ_t^S, ϵ_t^V, and ϵ_t^ξ are standard normal variables where $corr(\epsilon_t^S, \epsilon_t^V) = \rho$. Also, N_{t+1} follows a binomial distribution with the probability of taking 1 equal to $\lambda_t \Delta$. We have

$$S_{t+1} - S_t = (\mu_0 + \mu_1 V_t)\Delta + \sqrt{V_t \Delta}\epsilon_{t+1}^S + J_S^{\mathbb{P}} N_{t+1}, \tag{19}$$

$$V_{t+1} = V_t + (\eta + \kappa\xi_t - \kappa^{\mathbb{P}} V_t)\Delta + \sigma\sqrt{V_t \Delta}\epsilon_{t+1}^V + J_V^{\mathbb{P}} N_{t+1}, \tag{20}$$

$$\xi_{t+1} = \xi_t + (\alpha\theta - \alpha^{\mathbb{P}}\xi_t)\Delta + \gamma\sqrt{\xi_t \Delta}\epsilon_{t+1}^\xi. \tag{21}$$

For any given parameter value, we evaluate the likelihood function by following the SIR algorithm, which we summarize in Box 1. We use the Covariance Matrix Adaptation Evolution Strategy (CMAES) for likelihood optimization. We also impose Feller's conditions, i.e., $2\eta > \sigma^2$ and $2\alpha\theta > \gamma^2$, to ensure the positivity of V_t and ξ_t. Otherwise, the conditional densities are not well-defined.

5.3. Estimates of the joint pricing kernel

Prior to discussing the joint pricing kernel, we provide some comment on the fitting of our parametric model discussed in

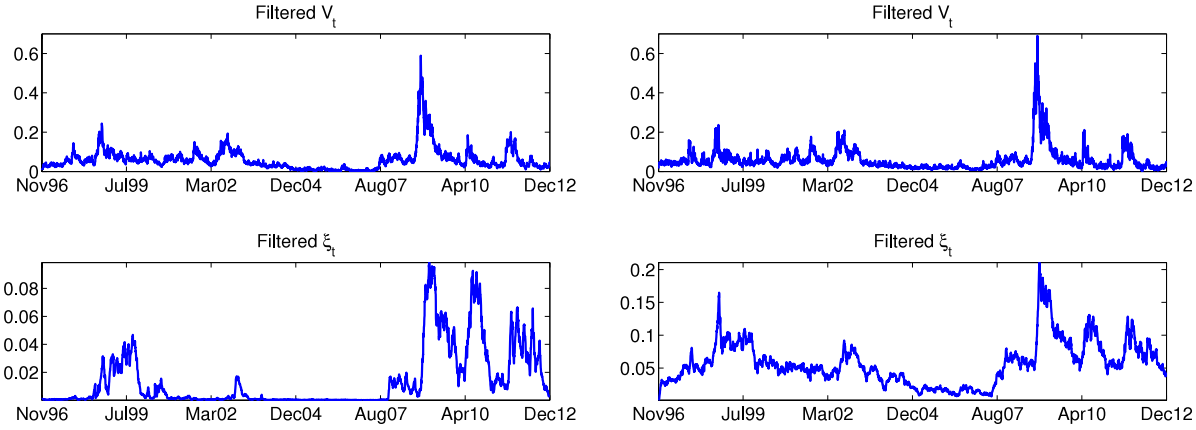

Fig. 7. Filtered state variables. Note: In this figure, we compare the filtered V_t and ξ_t from the same parametric model. The left panels impose Feller's constraints, whereas the right panels do not. The results are based on the same subsample of S&P 500 and VIX options.

1 Given current parameter value and current particles $\{(V_t^{(i)}, \xi_t^{(i)})\}_{i=1}^I$, simulate shocks from their distributions $(\epsilon_{t+1}^S, \epsilon_{t+1}^V, \epsilon_{t+1}^{\xi}, (J_S^{\mathbb{P}})_{t+1}, (J_V^{\mathbb{P}})_{t+1}, N_{t+1})^{(i)}$ for each $i = 1, 2, \ldots, I$. Calculate new states $(V_{t+1}^{(i)}, \xi_{t+1}^{(i)})$, using (20) and (21).

2 Evaluate $p(y_{t+1}|V_{t+1}^{(i)}, \xi_{t+1}^{(i)}, (J_S^{\mathbb{P}})_{t+1}^{(i)}, (J_V^{\mathbb{P}})_{t+1}^{(i)}, N_{t+1}^{(i)})$ for each $i = 1, 2, \ldots, I$. The density follows a multivariate normal. Compute the resampling probabilities

$$\omega_{t+1}^{(i)} = \frac{p(y_{t+1}|V_{t+1}^{(i)}, \xi_{t+1}^{(i)}, (J_S^{\mathbb{P}})_{t+1}^{(i)}, (J_V^{\mathbb{P}})_{t+1}^{(i)}, N_{t+1}^{(i)})}{\sum_{i=1}^I p(y_{t+1}|V_{t+1}^{(i)}, \xi_{t+1}^{(i)}, (J_S^{\mathbb{P}})_{t+1}^{(i)}, (J_V^{\mathbb{P}})_{t+1}^{(i)}, N_{t+1}^{(i)})}.$$

Draw with replacement using the multinomial distribution with probabilities ω_{t+1}:

$$z(i) \sim Mult(I; \omega_{t+1}^{(1)}, \ldots, \omega_{t+1}^{(I)}).$$

Set $(V_{t+1}^{(i)}, \xi_{t+1}^{(i)}, (J_S^{\mathbb{P}})_{t+1}^{(i)}, (J_V^{\mathbb{P}})_{t+1}^{(i)}, N_{t+1}^{(i)}) = (V_{t+1}^{z(i)}, \xi_{t+1}^{z(i)}, (J_S^{\mathbb{P}})_{t+1}^{z(i)}, (J_V^{\mathbb{P}})_{t+1}^{z(i)}, N_{t+1}^{z(i)})$, for each $i = 1, 2, \ldots, I$.

3 Repeat step 1 and step 2 to generate a sequence of observations $(y_t, V_t^{(i)}, \xi_t^{(i)})_{t=1,i=1}^{T,I}$, based on which we can evaluate the log-likelihood of the data $\sum_{t=1}^T \log \frac{1}{I} \sum_{i=1}^I p(y_t|V_t^{(i)}, \xi_t^{(i)})$.

Algorithm 1: Sampling-Importance Resampling (SIR)

Section 5.2. We select a subsample of S&P 500 and VIX options in order to accelerate the estimation. The nonparametric methods are computationally efficient whereas the parametric approach is much slower. We characterize all options into five moneyness categories, and three maturity categories, and select from each category per day one option that has the highest trading volume. While some deep in/out of the money options are missing over certain days, we accumulate a large number of records covering the same sampling period of our nonparametric analysis—58,054 S&P 500 options for 4263 days and 12,762 VIX options for 1376 days.

Table 3 reports the parameter estimates. We find that both estimates of $\kappa^{\mathbb{P}}$ and $\alpha^{\mathbb{P}}$ are larger than their counterparts under $\mathbb{Q}$ indicating that the variance risk premia are negative—the expected volatility under $\mathbb{Q}$ is higher than it is under $\mathbb{P}$. This agrees with most parametric studies in the literature. Also, the fact that $\alpha^{\mathbb{P}}$ is much smaller than $\kappa^{\mathbb{P}}$ indicates that ξ_t is more persistent than V_t, so that ξ_t captures some long-run volatility variation. The pricing errors are fairly small, suggesting a decent fitting under the risk-neutral measure.

To investigate the quality of the time-series fitting, we plot the filtered estimates in Fig. 7. We compare two scenarios here. On the left panels, we impose Feller's conditions throughout the estimation procedure, whereas on the right, we ignore such constraints. On both the top left and right panels, the level of V_t matches the ups and downs of the market from 1996 through 2012. In the early part of the sample, we observe volatility spikes in part due to the Asian, Russian and LTCM crises. After the Dot-Com bubble, the market becomes extremely calm between 2004 and 2007 until the recent financial meltdown, the European sovereign debt crisis, and the U.S. debt-ceiling confrontation. The dynamics of ξ_t on the left bottom panel seems unsatisfactory, since it touches the zero lower bound fairly often. This is due to Feller's constraint. In order to avoid hitting zero, 2η has to be larger than σ^2, which pushes ξ_t towards zero. On the right bottom panel of Fig. 7 where Feller's constraints are relaxed, the dynamics of ξ_t appears more satisfactory. Needless to say, the pricing errors in the unconstrained case are also smaller. Nevertheless, we have to stick to the constrained fitting, in this case that warrants well-defined conditional densities.[14] Alternatively, log-type volatility models are superior to square-root models since they are free from Feller's restrictions and they allow more persistent time series, as discussed in Amengual and Xiu (2014). Unfortunately, they do not allow closed-form option prices hence they cannot be used in our parametric study.

We now discuss estimates of the joint pricing kernel based on the model. The top panels of Fig. 8 provide the joint risk-neutral and physical densities of the market return and the VIX, conditional on median levels of Z_t and L_t, whereas the bottom panels present the corresponding contour plots. We find that both plots display a clear negative correlation between the marginal distributions of the market return and the VIX, and as a result, for a wide range of scenarios, both joint densities are fairly close to 0. Moreover, it is clear from the contour plot that the tail of the risk-neutral density with a low market return and a high VIX level is heavier than that of the physical density, which agrees with our findings from the nonparametric marginal density plots.

Fig. 9 then provides the contour plot of the joint pricing kernel. The values of the joint pricing kernel that correspond to the high market return and high VIX level, and low market return and low VIX level are undetermined, because these scenarios are so rare that the risk-neutral and physical densities are rather close to 0,

[14] We also implement an alternative specification of two-factor square-root volatility models introduced in Appendix A.2, for which the empirical fitting is also affected by binding Feller's constraints.

Table 3
Summary of Parameter Estimates.

	Estimate	Std Err		Estimate	Std Err		Estimate	Std Err
κ	2.8332	(0.1293)	μ_0	0.0386	(0.0018)	ϵ_1^{SPO}	0.0732	(0.0033)
σ	0.5111	(0.0214)	κ^P	6.2466	(0.2621)	ϵ_2^{SPO}	0.0252	(0.0011)
ρ	−0.8407	(0.0330)	β_+^P	0.0254	(0.0010)	ϵ_3^{SPO}	0.0247	(0.0010)
β_+	0.0081	(0.0003)	β_-^P	0.0146	(0.0007)	ϵ_1^{VXO}	1.3709	(0.0570)
β_-	0.0196	(0.0008)	β_V^P	0.0063	(0.0002)	ϵ_2^{VXO}	1.0704	(0.0438)
q	0.0853	(0.0030)	α^P	1.5089	(0.0823)	ϵ_3^{VXO}	1.3913	(0.0614)
β_V	0.0094	(0.0004)	μ_1	−0.1197	(0.0050)	ϵ^{VIX}	5.2425	(0.2166)
λ_1	8.1313	(0.3591)						
λ_0	0.3023	(0.0123)						
α	0.6432	(0.0302)						
γ	0.1714	(0.0059)						
θ	0.0236	(0.0010)						
η	0.1306	(0.0056)						

Note: This table reports the maximum likelihood estimates of the parameters introduced in Section 5.1. ε_1^{SPO}, ε_2^{SPO}, and ε_3^{SPO} denote the standard errors of the differences between the observed and fitted implied volatilities for short (<3 months), medium (3–6 months), and long maturities (6–12 months), respectively. Similarly, ε_1^{VXO}, ε_2^{VXO}, and ε_3^{VXO} are for VIX options with maturity less than 2 months, 2–4 months, and 4–6 months. ε_1^{VIX} denotes the error for the VIX.

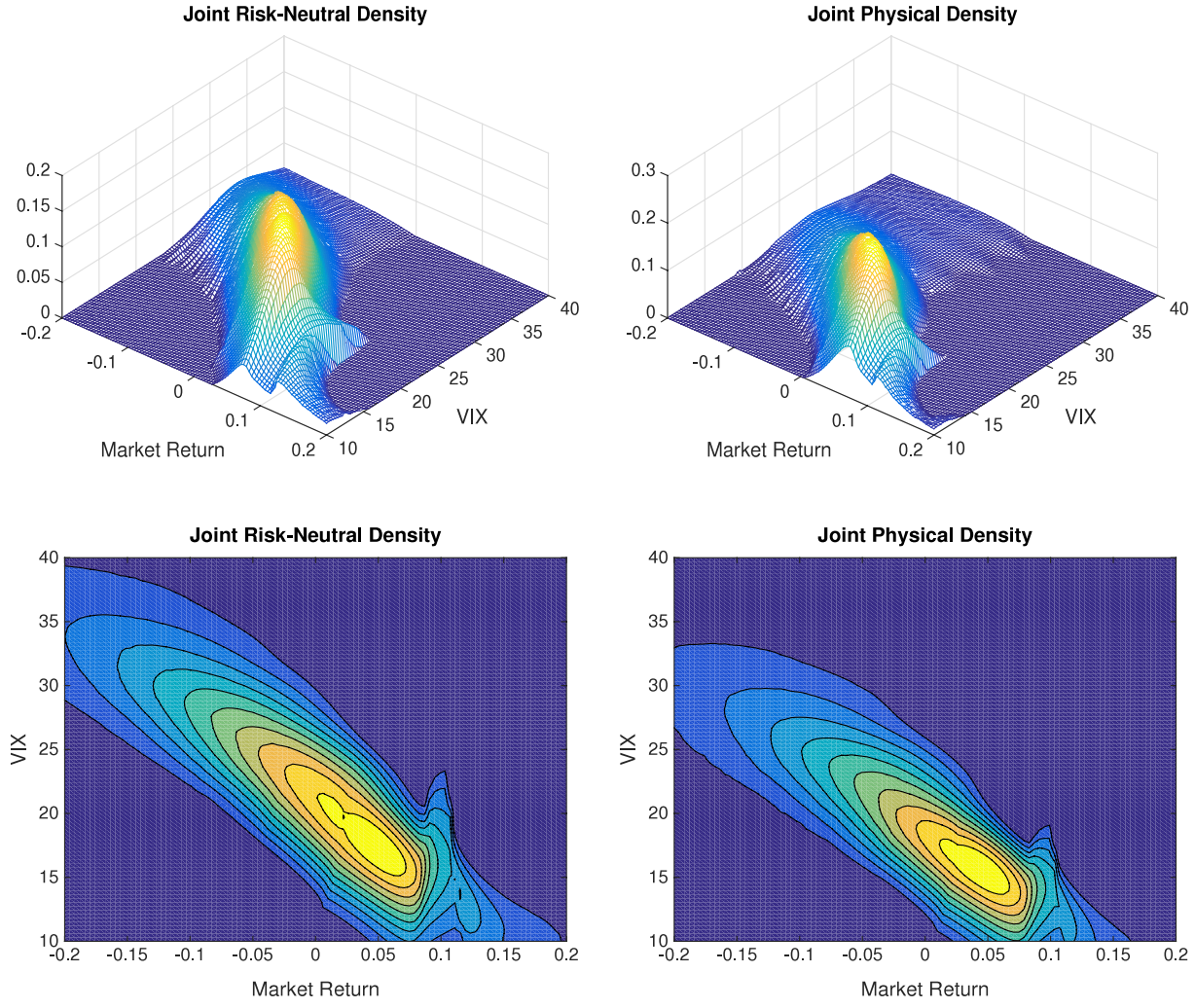

Fig. 8. Estimates of the joint risk-neutral and physical densities. Note: The top panels provide 3D plots of our estimates of the joint risk-neutral and physical densities of the S&P 500 index return and VIX at the maturity of 42 days, with L_t and Z_t set at their median levels. The bottom panels present the corresponding contour plots. The sample period is July 1, 2007–December 31, 2012. Lighter colors in the contour plot indicate larger values. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

as can be seen from Fig. 8. Yet, our estimates show that the joint pricing kernel achieves higher values in the region of negative market return and relatively large VIX level, which is in agreement with the intuition that high pricing kernels are associated with unpleasant economic states.

5.4. Comparison between nonparametric and parametric estimates

Finally, we compare the nonparametric and parametric estimates. Such a comparison sheds light on the limitation of the parametric approach and points to the direction of improvement.

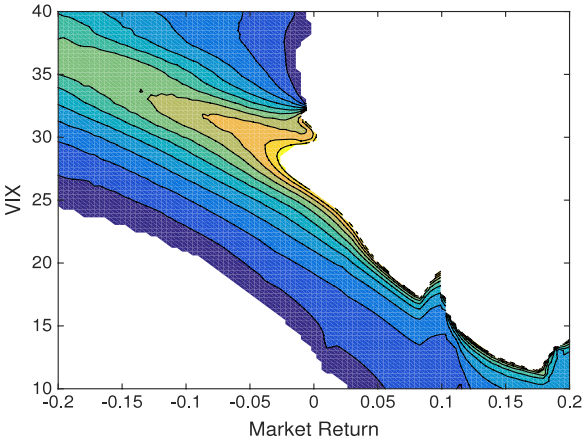

Fig. 9. Estimates of the joint pricing kernel. Note: This figure provides the contour plot of the joint pricing kernel estimates of the market return and the VIX at maturity of 42 days, with L_t and Z_t fixed at their median levels. Lighter colors in the contour plot indicate larger values. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The left panels of Fig. 10 compare the estimates of the risk-neutral density, physical density, and pricing kernel of the market return, whereas the right panels provide the corresponding estimates for the VIX. The parametric densities and pricing kernel estimates of the market return agree with the nonparametric estimates. This is not surprising given that the parametric estimates of the volatility level are very satisfactory.

Regarding the right panels, we find that the parametric density and pricing kernel of the VIX fail to match our nonparametric estimates by a wide margin. In particular, the parametric pricing kernel shows a monotonically increasing shape, consistent with models in Bansal and Yaron (2004), Drechsler and Yaron (2011), Bollerslev et al. (2009), Zhou and Zhu (2012), and Branger and Völkert (2012).[15] This is in sharp contrast to our nonparametric VIX pricing kernel estimates that exhibit a pronounced U-shape. The parametric models often impose restrictions on the data for tractability or computational feasibility, which may not be satisfied by the real data. While square-root affine models provide closed-form pricing for options, there remains a substantial gap between the specified model and the real data, as is evident from binding constraints, whereas a rigorous nonparametric approach offers a better alternative.

Overall, the comparisons of the parametric and nonparametric estimates of the marginal pricing kernels reveal that the existing models capture how the market risk is priced reasonably well, but fail to correctly reflect the price of the volatility risk, potentially due to the model misspecification error.

6. Conclusion

Volatility has been well documented as a priced risk factor, and hence an essential component of pricing kernels. Taking advantage of the rapidly developed volatility derivative markets, we provide nonparametric estimates of risk-neutral and physical densities as well as pricing kernels. We show that all these estimates strongly depend on multiple volatility factors. In particular, we document a U-shape pricing kernel of the VIX, which cannot be captured by the state-of-the-art parametric model, and which is also in conflict with the predictions of most equilibrium asset-pricing models. Extensions of these models are therefore necessary to reconcile them with our empirical findings.

[15] A stylized equilibrium model is provided in Appendix B to illustrate how the pricing kernel depend on volatility factors.

Appendix A. Nested reduced-form option-pricing models

A.1. One-factor volatility models

We justify the strategy of replacing V_t by Z_t using reduced-form option pricing models widely used in the literature. We first consider the class of option pricing models that induce an affine relationship between the unobservable variance and the squared VIX. This class of models has the following risk-neutral dynamics:

$$\begin{aligned} dS_t &= \left(r_f - d - \frac{1}{2}V_t\right)dt + \sqrt{V_t}dW_t^{\mathbb{Q}} + dL_t^S, \\ dV_t &= \kappa(\xi - V_t)dt + \sigma(V_t)dB_t^{\mathbb{Q}} + dL_t^V, \end{aligned} \tag{22}$$

where dL_t^S and dL_t^V can be finite activity compound Poisson processes with correlated jump sizes J_t^S and J_t^V independent of V_t, and with intensities being affine in V_t. Such models include those discussed in Bakshi et al. (1997), Bates (2000), Pan (2002), Chernov and Ghysels (2000), Eraker (2004), Carr et al. (2003), Eraker et al. (2003),and Broadie et al. (2007). Jumps can be driven by Lévy processes such as the CGMY process in Carr et al. (2003) and Bates (2012). Note that this class also includes non-Gaussian OU processes, as introduced in Barndorff-Nielsen and Shephard (2001); see Shephard (2005) for a collection of similar models. For models of this class, we have

$$Z_t^2 = aV_t + b,$$ [16]

where a and b are functions of model parameters. See Carr and Wu (2009) and Li and Xiu (2012) for details. Apparently, prices of the risk are incorporated into parameters, and Z_t^2 is a linear function of V_t, hence Z_t and V_t generate the same information set.

The second class of models introduces a non-affine structure between the squared VIX and the unobservable variance, such as the exponential-OU-L models in Shephard (2005). More specifically, under the risk-neutral measure, such models specify the volatility process as

$$\log V_t = \alpha + \beta X_t, \quad dX_t = \kappa X_t dt + dL_t^V,$$

where L_t^V is a Lévy process. The squared VIX is, as calculated by Tauchen and Todorov (2011) and Li and Xiu (2012),

$$Z_t^2 = \frac{1}{\tau}\int_0^\tau \gamma + (\eta+1)\exp\Big(\alpha + e^{\kappa u}(\log V_t - \alpha) + C(u)\Big)du, \tag{23}$$

where $C(u)$ is determined by the characteristic exponent of the Lévy process L_t^V, and γ and η are constants determined by the quadratic variation of L_t^S. It is straightforward to show that this function $V_t \mapsto Z_t$ is invertible, so that the information sets generated by V_t and by Z_t are equivalent.

A.2. Multi-factor volatility models

Our framework also extends to many multi-factor volatility models. For example, we consider two common specifications of option-pricing models with two volatility factors. For exam-

[16] The squared VIX is given by $100^2 \times \frac{1}{\tau}\mathbb{E}^{\mathbb{Q}}(\int_t^{t+\tau} V_s ds + \sum_{t\le s\le t+\tau}(\exp(J_s^S) - 1 - J_s^S)|\mathcal{F}_t)$, where $\tau = 21$ days.

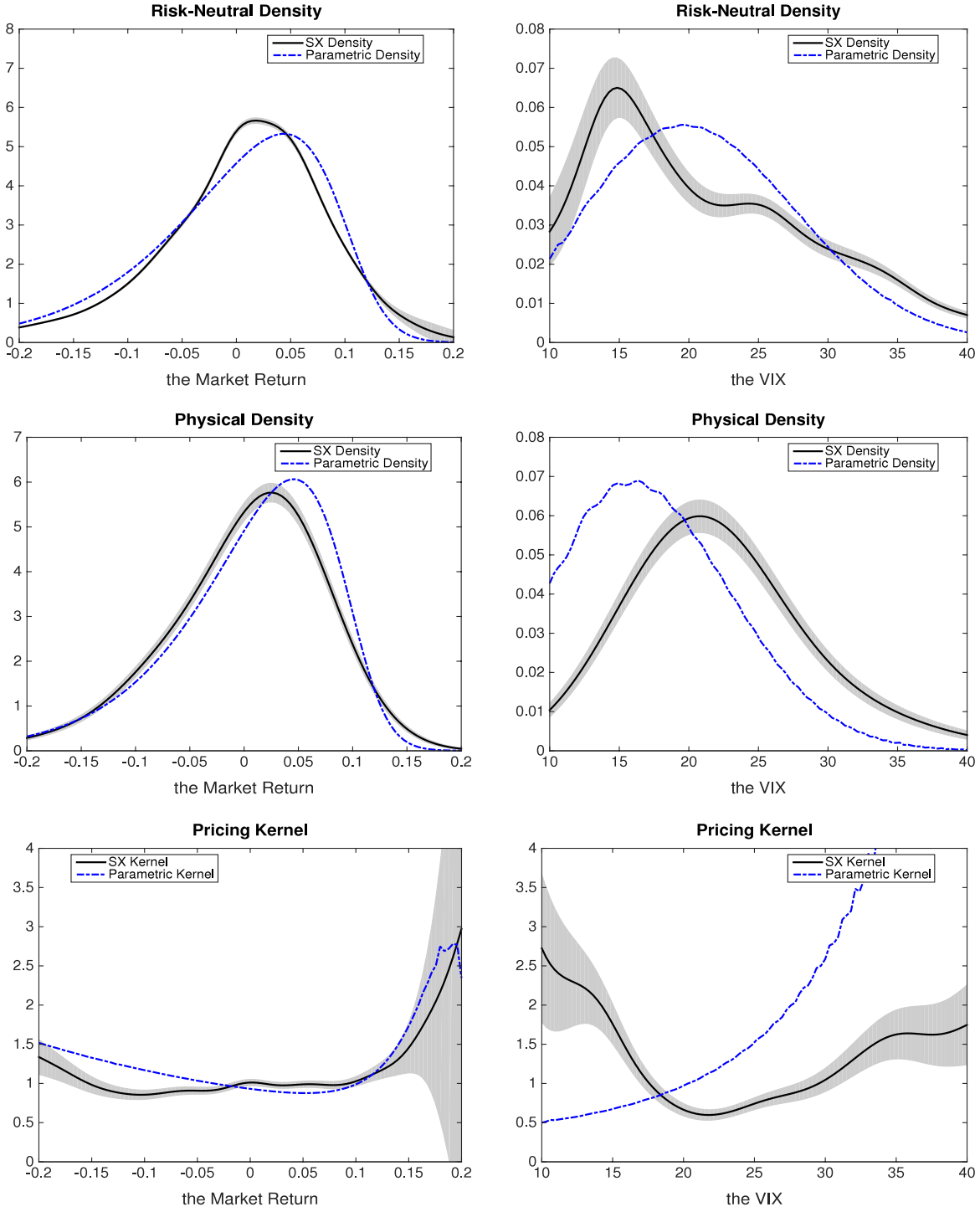

Fig. 10. Comparison between parametric and nonparametric estimates. Note: The left panels compare our nonparametric estimates (SX) with the parametric estimate of the marginal risk neutral density (top), physical density (middle), and pricing kernel (bottom) of the market return. The right panels provide the corresponding results for the VIX. We fix the time-to-maturity at 42 days, and L_t and Z_t at their median levels.

ple, Egloff et al. (2010) and Mencia and Sentana (2012) specify the volatility dynamics as:

$$\begin{aligned} dV_t &= (\eta + \kappa(\xi_t - V_t))dt + \sigma\sqrt{V_t}dB_t^{\mathbb{Q}} + dL_t^V, \\ d\xi_t &= \alpha(\theta - \xi_t)dt + \gamma\sqrt{\xi_t}dM_t^{\mathbb{Q}} + dL_t^{\xi}, \end{aligned} \tag{24}$$

where $B_t^{\mathbb{Q}}$ and $M_t^{\mathbb{Q}}$ are two potentially correlated Brownian motions, and the compensator for L_t^{ξ} is affine in ξ_t. This model leads to an affine pricing equation for both the squared VIX and variance swap rates with maturity τ:

$$Z_t^2 = aV_t + b\xi_t + c, \quad \text{and} \quad Z_{t,\tau}^2 = \bar{a}_\tau V_t + \bar{b}_\tau \xi_t + \bar{c}_\tau, \tag{25}$$

where $a, b, c, \bar{a}_\tau$, $\bar{b}_\tau$ and $\bar{c}_\tau$ are constants that depend nonlinearly on τ and model parameters.[17] The derivation is given by Appendix D.

[17] The variance swap rate with maturity τ is $100 \times \frac{1}{\tau}\mathbb{E}^{\mathbb{Q}}(\int_t^{t+\tau} V_s ds + \sum_{t\le s\le t+\tau}(J_s^S)^2|\mathcal{F}_t)$, where $QV_{t,\tau}$ denotes the quadratic variation of the log return process from t to $t+\tau$.

Apparently, V_t and ξ_t can be written as affine functions of $Z^2_{t,\tau}$ with at least two distinct maturities, Z^2_{t,τ_1} and Z^2_{t,τ_2}, and hence they generate the same information set. For the same reason, our proposal to use the VIX Z_t and the slope $L_t := Z_{t,\tau_2}/Z_{t,\tau_1} - 1$ can also be justified, where τ_1, and τ_2 are some fixed time-to-maturities, such as 3 months and 1 year.

Chernov et al. (2003) and Bates (2012) adopt an alternative two-factor additive volatility model, in which $V_t = X_{1t} + X_{2t}$, where for $i = 1, 2$,

$$dX_{it} = \kappa_i(\xi_i - X_{it})dt + \sigma_i\sqrt{X_{it}}dB^{\mathbb{Q}}_{it} + dL^{X_i}_t,$$

where $L^{X_i}_t$ is some pure jump process with a compensator being affine in X_{it}. This model also yields an affine relationship among Z_t, $Z^2_{t,\tau}$, X_{1t} and X_{2t}, hence our choices of Z_t and L_t remain valid.

Appendix B. A stylized equilibrium model

In this section, we sketch a stylized general equilibrium model, which is a simplified version of Branger and Völkert (2012) and Zhou and Zhu (2012) that extend the models of Bansal and Yaron (2004), Bollerslev et al. (2009), and Drechsler and Yaron (2011) to include multiple volatility factors. Such a model provides the economics motivation for the pricing kernel defined in Section 2 based on the no-arbitrage principal. It also predicts the shape of the pricing kernel, against which we shall interpret and compare our empirical estimates.

More specifically, we assume a representative agent economy with a recursive preference over consumption (Epstein and Zin (1989)). The log pricing kernel at time $t+1$ is given by

$$m_{t+1} = \theta \log\delta - \theta\psi^{-1}\Delta c_{t+1} + (\theta - 1)r_{c,t+1}, \tag{26}$$

where $\theta = (1-\gamma)/(1-\psi^{-1})$, $0 < \delta < 1$ is the subjective discount factor, γ is the risk-aversion coefficient, ψ is the intertemporal elasticity of substitution, Δc_{t+1} is the growth rate of log consumption, and r_{t+1} is the time t to $t+1$ return on the aggregate wealth claim.[18] The state vector of the economy follows

$$\begin{aligned}
\Delta c_{t+1} &= \mu_c + \sigma_t z_{c,t+1},\\
\sigma^2_{t+1} &= \xi^2_{t+1} + \rho_\sigma\sigma^2_t + \sigma_t z_{\sigma,t+1},\\
\xi^2_{t+1} &= \mu_\xi + \rho_\xi\xi^2_t + \xi_t z_{\xi,t+1},
\end{aligned} \tag{27}$$

where $\{z_{c,t}\}$, $\{z_{\sigma,t}\}$, and $\{z_{\xi,t}\}$ are independent i.i.d. N(0, 1) processes.

By the standard log-linearization approach following Campbell and Shiller (1988), we have

$$r_{c,t+1} = \kappa_0 + \kappa_1 w_{t+1} - w_t + \Delta c_{t+1}, \tag{28}$$

where the price–wealth ratio w_t is conjectured to be affine in the state vector:

$$w_t = A_0 + A_\sigma\sigma^2_t + A_\xi\xi^2_t, \tag{29}$$

with $A_0 > 0$ and $A_\sigma < 0$ as functions of the model parameters we suppress for notational brevity, see Branger and Völkert (2012) and Zhou and Zhu (2012) for details. The coefficients A_0, A_σ, A_ξ, as well as the linearizing constants κ_0 and κ_1 solve a system of (non-linear) equations.

Using (28) and (29), we obtain

$$\begin{aligned}
r_{c,t+1} = {} & \Delta c_{t+1} + \kappa_1 A_\sigma\sigma^2_{c,t+1} + \kappa_1 A_\xi\xi^2_{t+1} - A_\sigma\sigma^2_{c,t}\\
& - A_\xi\xi^2_t + \kappa_0 + \kappa_1 A_0 - A_0.
\end{aligned}$$

Therefore, both the long-run and short-run volatility factors ξ^2_{t+1} and σ^2_{t+1} show up in $r_{c,t+1}$, and hence in the pricing kernel m_{t+1} given in (26). In most scenarios where $\gamma > 1$ and $\psi > 1$, $A_\sigma < 0$ and $A_\xi < 0$ hold. The wealth–consumption ratio, and hence the pricing kernel m_{t+1}, is decreasing in σ^2_t and ξ^2_t, implying that an increase in the short-run or the long-run volatility worsens economic conditions and increases the marginal utility.

[18] The literature usually assumes that $\gamma > 1$ and $\psi > 1$, which implies $\theta < 0$. This assumption ensures that the representative agent has a preference for early resolution of uncertainty, which is the key for the price of volatility risk.

Appendix C. Proof of the asymptotic theory

We provide the proof of the asymptotic theory (17) as part of it is non-standard. Notice from (11) that

$$\begin{bmatrix}\widehat{\alpha}\\ \widehat{\boldsymbol{\beta}}\end{bmatrix}_{(1+4)\times 1} = (\boldsymbol{\Omega}^\intercal\mathbf{K}\boldsymbol{\Omega})^{-1}\boldsymbol{\Omega}^\intercal\mathbf{K}\overline{\mathbf{C}}.$$

Using the properties of Gaussian kernel, we have

$$\begin{aligned}
&\frac{1}{n}(\boldsymbol{\Omega}^\intercal\mathbf{K}\boldsymbol{\Omega})^{-1}\\
&= \begin{bmatrix} \frac{1}{n}\sum_{i=1}^n K_h(\mathbf{u}_i-\mathbf{u}) & \frac{1}{n}\sum_{i=1}^n K_h(\mathbf{u}_i-\mathbf{u})(\mathbf{u}_i-\mathbf{u})^\intercal\\ \frac{1}{n}\sum_{i=1}^n K_h(\mathbf{u}_i-\mathbf{u})(\mathbf{u}_i-\mathbf{u}) & \frac{1}{n}\sum_{i=1}^n K_h(\mathbf{u}_i-\mathbf{u})(\mathbf{u}_i-\mathbf{u})(\mathbf{u}_i-\mathbf{u})^\intercal\end{bmatrix}\\
&\xrightarrow{p} \begin{bmatrix} f(\mathbf{u}) & \mathbf{0}^\intercal\\ \mathbf{0} & f(\mathbf{u})\cdot\int c^2k^2(c)dc\cdot\mathrm{diag}(h^2_\tau,h^2_z,h^2_l,h^2_m)\end{bmatrix},
\end{aligned}$$

as $h\to 0, n\to 0$.

Therefore, we can write the estimators in their equivalent kernel forms:

$$\widehat{\alpha} \approx \frac{1}{nf(\mathbf{u})}\sum_{i=1}^n K_h(\mathbf{u_i}-\mathbf{u})\cdot\overline{C}_i,$$

$$\widehat{\beta}_4 \approx \frac{1}{nh^2_m f(\mathbf{u})\int c^2k^2(c)dc}\sum_{i=1}^n K_h(\mathbf{u_i}-\mathbf{u})(m_i-m)\cdot\overline{C}_i.$$

Using the standard kernel asymptotic results, we can obtain the desired asymptotic theory.

Appendix D. Closed-form derivation of the VIX and the variance swap rate

Given the model in Section 5.1, we solve for the VIX by noting that

$$\begin{aligned}
d\mathbb{E}^{\mathbb{Q}}_t V_s &= (\eta + \kappa(\mathbb{E}^{\mathbb{Q}}_t\xi_s - \mathbb{E}^{\mathbb{Q}}_t V_s))ds + \beta_V\mathbb{E}^{\mathbb{Q}}_t(\lambda_0 + \lambda_1 V_s)ds,\\
d\mathbb{E}^{\mathbb{Q}}_t\xi_s &= \alpha(\theta - \mathbb{E}^{\mathbb{Q}}_t\xi_s)ds.
\end{aligned}$$

So, we can obtain

$$\mathbb{E}^{\mathbb{Q}}_t\xi_s = \xi_t e^{-\alpha(s-t)} + \theta(1-e^{-\alpha(s-t)})$$

$$\begin{aligned}
\mathbb{E}^{\mathbb{Q}}_t V_s = {} & V_t e^{-(\kappa-\beta_V\lambda_1)(s-t)} - \frac{\xi_t\kappa}{\kappa-\alpha-\beta_V\lambda_1}\\
& \times\left(e^{-(s-t)(\kappa-\beta_V\lambda_1)} - e^{-\alpha(s-t)}\right)\\
& + \frac{\beta_V\lambda_0+\eta}{(\kappa-\beta_V\lambda_1)}\left(1-e^{-(s-t)(\kappa-\beta_V\lambda_1)}\right)\\
& + \left(\frac{\kappa}{\kappa-\beta_V\lambda_1} - \frac{\kappa}{-\alpha+\kappa-\beta_V\lambda_1}e^{-\alpha(s-t)}\right.\\
& \left. + \frac{\kappa\alpha}{(\kappa-\beta_V\lambda_1)(-\alpha+\kappa-\beta_V\lambda_1)}e^{-(s-t)(\kappa-\beta_V\lambda_1)}\right)\cdot\theta.
\end{aligned}$$

Hence, it implies that

$$\begin{aligned}\mathbb{E}_t^{\mathbb{Q}}\Big(\int_t^{t+\tau} V_s ds\Big) &= V_t \frac{1-e^{-(\kappa-\beta_V\lambda_1)\tau}}{\kappa-\beta_V\lambda_1} - \frac{\xi_t\kappa}{\kappa-\alpha-\beta_V\lambda_1}\\ &\times\Big(\frac{1-e^{-(\kappa-\beta_V\lambda_1)\tau}}{\kappa-\beta_V\lambda_1} - \frac{1-e^{-\alpha\tau}}{\alpha}\Big)\\ &+ \frac{\beta_V\lambda_0+\eta}{(\kappa-\beta_V\lambda_1)}\Big(\tau - \frac{1-e^{-(\kappa-\beta_V\lambda_1)\tau}}{\kappa-\beta_V\lambda_1}\Big)\\ &+ \Big(\frac{\kappa\tau}{\kappa-\beta_V\lambda_1} - \frac{\kappa}{-\alpha+\kappa-\beta_V\lambda_1}\frac{1-e^{-\alpha\tau}}{\alpha}\\ &+ \frac{\kappa\alpha}{(\kappa-\beta_V\lambda_1)(-\alpha+\kappa-\beta_V\lambda_1)}\frac{1-e^{-(\kappa-\beta_V\lambda_1)\tau}}{\kappa-\beta_V\lambda_1}\Big)\cdot\theta. \end{aligned} \tag{30}$$

Therefore, by definition we have

$$\begin{aligned}(\mathrm{VIX}_t/100)^2 &= \frac{1}{\tau}\mathbb{E}_t^{\mathbb{Q}}\Big(\int_t^{t+\tau} V_s ds\Big) + \frac{2}{\tau}\mathbb{E}_t^{\mathbb{Q}}\Big(\int_t^{t+\tau}\lambda_s ds\Big)\\ &\times\mathbb{E}^{\mathbb{Q}}\Big(e^{J_S^{\mathbb{Q}}} - 1 - J_S^{\mathbb{Q}}\Big) = aV_t + b\xi_t + c, \end{aligned} \tag{31}$$

$$\begin{aligned}\mathrm{VS}_t/100 &= \frac{1}{\tau}\mathbb{E}_t^{\mathbb{Q}}\Big(\int_t^{t+\tau} V_s ds\Big) + \frac{2}{\tau}\mathbb{E}_t^{\mathbb{Q}}\Big(\int_t^{t+\tau}\lambda_s ds\Big)\mathbb{E}^{\mathbb{Q}}(J_S^{\mathbb{Q}})^2\\ &= \bar{a}V_t + \bar{b}\xi_t + \bar{c}, \end{aligned} \tag{32}$$

where

$$a = \frac{1-e^{-(\kappa-\beta_V\lambda_1)\tau}}{(\kappa-\beta_V\lambda_1)\tau}(1+\lambda_1\chi),$$

$$b = -\frac{\kappa}{\kappa-\alpha-\beta_V\lambda_1}\Big(\frac{1-e^{-(\kappa-\beta_V\lambda_1)\tau}}{(\kappa-\beta_V\lambda_1)\tau} - \frac{1-e^{-\alpha\tau}}{\alpha\tau}\Big)(1+\lambda_1\chi),$$

$$\begin{aligned}c &= \Big(\frac{\beta_V\lambda_0+\eta}{(\kappa-\beta_V\lambda_1)}\Big(1-\frac{1-e^{-(\kappa-\beta_V\lambda_1)\tau}}{(\kappa-\beta_V\lambda_1)\tau}\Big)\\ &+\Big(\frac{\kappa}{\kappa-\beta_V\lambda_1} - \frac{\kappa}{-\alpha+\kappa-\beta_V\lambda_1}\frac{1-e^{-\alpha\tau}}{\alpha\tau}\\ &+\frac{\kappa\alpha}{(\kappa-\beta_V\lambda_1)(-\alpha+\kappa-\beta_V\lambda_1)}\frac{1-e^{-(\kappa-\beta_V\lambda_1)\tau}}{(\kappa-\beta_V\lambda_1)\tau}\Big)\cdot\theta\Big)\\ &\times(1+\lambda_1\chi)+\lambda_0\chi,\end{aligned}$$

$$\begin{aligned}\chi &= 2\Big(q(1-\beta_+)^{-1} + (1-q)(1+\beta_-)^{-1}\\ &-1-q\beta_+ + (1-q)\beta_-\Big),\end{aligned}$$

and $\bar{a}$, $\bar{b}$, and $\bar{c}$ are defined similarly to a, b, and c, by replacing χ by $\bar{\chi} = 2(q\beta_+^2 + (1-q)\beta_-^2)$. *VS* denotes the variance swap rates.

Appendix E. Closed-form marginal and joint conditional densities

Let $Y = (S_t, V_t, \xi_t)$. Denote their realizations at t (or T) as y (or y'). Let $u = (u_s, u_v, u_\xi) \in \mathbb{R}^3$, $\gamma = (\gamma_s, \gamma_v, \gamma_\xi) \in \mathbb{R}^3$, and $\tau = T - t$. We consider the following transformation:

$$\Psi^Y(\gamma, u, \tau, y) = \mathbb{E}_t^{\mathbb{Q}}(e^{(\gamma+iu)\cdot Y_T}).$$

The corresponding partial differential equation for the generalized characteristic function is given by:

$$\begin{aligned}0 &= -\frac{\partial\Psi^Y}{\partial\tau} + \Big(r_f - d - \frac{1}{2}v - \mu(\lambda_0+\lambda_1 v)\Big)\frac{\partial\Psi^Y}{\partial s}\\ &+ (\eta+\kappa(\xi-v))\frac{\partial\Psi^Y}{\partial v} + \alpha(\theta-\xi)\frac{\partial\Psi^Y}{\partial\xi} + \frac{1}{2}\gamma^2\xi\frac{\partial^2\Psi^Y}{\partial\xi^2}\\ &+ \frac{1}{2}v\frac{\partial^2\Psi^Y}{\partial s^2} + \frac{1}{2}\sigma^2 v\frac{\partial^2\Psi^Y}{\partial v^2} + \rho\sigma v\frac{\partial^2\Psi^Y}{\partial s\partial v}\\ &+ \int_E\Big(\Psi^Y(\gamma,u,\tau,s,v,\xi) - \Psi^Y(\gamma,u,\tau,y)\Big)m(t,d\eta)\\ &+ (\lambda_0+\lambda_1 v)\iint_{\mathbb{R}^2}\Big(\Psi^Y(\gamma,u,\tau,s+j_s,v+j_v,\xi)\\ &- \Psi^Y(\gamma,u,\tau,y)\Big)\nu_s(j_s)\nu_v(j_v)dj_s dj_v.\end{aligned}$$

So we have:

$$\Psi^Y(\gamma,u,\tau,y) = e^{(\gamma_s+iu_s)s + A(\tau) + B_v(\tau)v + B_\xi(\tau)\xi},$$

where $A(\tau)$, $B_v(\tau)$, and $B_\xi(\tau)$ satisfy the system of ordinary differential equations below:

$$\begin{aligned}0 &= -\dot{B}_\xi - \alpha B_\xi + \frac{1}{2}\gamma^2 B_\xi^2 + \kappa B_v\\ 0 &= -\dot{B}_v + \frac{1}{2}\sigma^2 B_v^2 + ((\gamma_s+iu_s)\rho\sigma - \kappa)B_v\\ &\quad - \Big(\frac{1}{2} + \mu\lambda_1\Big)(\gamma_s+iu_s) + \frac{1}{2}(\gamma_s+iu_s)^2 + \lambda_1(l(B_v)-1)\\ 0 &= -\dot{A} + \alpha\theta B_\xi + \eta B_v + (r_f - d - \mu\lambda_0)(\gamma_s+iu_s)\\ &\quad + \lambda_0(l(B_v)-1)\end{aligned}$$

where $A(0) = 0$, $B_v(0) = \gamma_v + iu_v$, and $B_\xi(0) = \gamma_\xi + iu_\xi$.

To simplify our notation, we write $\Psi^Z(\gamma_z, u_z) = \Psi^Y(\gamma_z e_z, u_z e_z, \tau, y)$, where Z denotes S, V or ξ, and e_z is a unit vector with the zth component equal to 1.

The joint conditional transition density $p(\tau, y'|y)$ can be represented via inverse Fourier Transform as:

$$p^*(\tau,y'|y) = \frac{e^{-\gamma\cdot y'}}{(2\pi)^3}\iiint_{\mathbb{R}^3} e^{-iu\cdot y'}\Psi^Y(\gamma,u,\tau,y)du. \tag{33}$$

Moreover, the marginal densities can be derived by plugging 0 into the irrelevant components of γ and u. For instance, the marginal conditional density of S is given by a one-dimensional integral:

$$p^*(\tau,s'|y) = \frac{e^{-\gamma_s s'}}{2\pi}\int_{\mathbb{R}} e^{-iu_s s'}\Psi^S(\gamma_s,u_s,\tau,y)du.$$

The tuning parameter γ_y in general can be selected to ensure the absolute convergence of the integrals under consideration as well as the conciseness of the formulae. We fix them at 0 for densities.

To develop the joint density of s' and z', we first consider changes of variables: $G : (s', v', \xi') \to (s', z', \xi')$, where $z' = 100\times\sqrt{av' + b\xi' + c}$ is given by (31). Note that by (33), taking $\gamma = (0,0,0)$,

$$\begin{aligned}&p^*(\tau,s',z',\xi'|s,v,\xi)\\ &= p^*(\tau,s',v',\xi'|s,v,\xi)\cdot|\det(\partial(s',v',\xi')/\partial(s',z',\xi'))|\\ &= \frac{2z'}{100^2(2\pi)^3 a}\iiint_{\mathbb{R}^3} e^{-i(u_s s' + u_v\frac{(z'/100)^2 - b\xi' - c}{a} + u_\xi\xi')}\\ &\quad\times\Psi^Y((0,0,0),u,\tau,s,v,\xi)\,du.\end{aligned}$$

Therefore, we have

$$p^*(\tau, s', z'|s, v, \xi) = \int_{-\infty}^{\infty} p^*(\tau, s', z', \xi'|s, v, \xi)d\xi'$$
$$= \frac{z'}{2 \times 100^2\pi^2 a} \iint_{\mathbb{R}^2} e^{-i(u_s s' + u_v \frac{(z'/100)^2 - c}{a})}$$
$$\times \Psi^Y\left((0,0,0), \left(u_s, u_v, \frac{b}{a}u_v\right), \tau, s, v, \xi\right) du_s du_v.$$

Based on this, we can further derive the marginal density of z':

$$p^*(\tau, z'|s, v, \xi) = \int_{-\infty}^{\infty} p^*(\tau, s', z'|s, v, \xi)ds'$$
$$= \frac{z'}{100^2\pi a} \int_{\mathbb{R}} e^{-iu_v \frac{(z'/100)^2 - c}{a}}$$
$$\times \Psi^Y\left((0,0,0), \left(0, u_v, \frac{b}{a}u_v\right), \tau, s, v, \xi\right) du_v.$$

The above derivations utilize the Fourier transform of generalized functions (distributions). For further discussion see Kanwal (2004). The conditional density under $\mathbb{P}$ measure can be derived similarly.

Appendix F. Closed-form pricing of the S&P 500 and VIX options

The price of the S&P 500 European call option with maturity T and strike price K, can be written in terms of the log forward price $f_{t,\tau}$:

$$C_{t,T}^{\text{S\&P}} = e^{-r_{t,\tau}\tau}\mathbb{E}_t^{\mathbb{Q}}((e^{F_T} - K)^+)$$
$$= e^{f_{t,\tau} - r_{t,\tau}\tau} - \frac{\sqrt{K}e^{-r_{t,\tau}\tau}}{\pi} \int_0^{\infty} \text{Re}\left[\frac{\Psi^F(0, u_f - \frac{i}{2})e^{-iu_f \log K}}{(u_f^2 + \frac{1}{4})}\right] du_f.$$

The VIX option price with maturity T and strike price K is given by:

$$C_{t,T}^{\text{VIX}} = e^{-r_{f,\tau}\tau}\mathbb{E}_t^{\mathbb{Q}}\left((100\sqrt{aV_T + b\xi_T + d} - K)^+\right)$$
$$= 100e^{-r_{f,\tau}\tau} \iint_{\mathbb{R}^2} (\sqrt{av' + b\xi' + d} - K')^+ p^*(v', \xi'|s, v, \xi)dv'd\xi'$$
$$= \frac{25e^{-r_{f,\tau}\tau}}{\pi^2} \iint_{\mathbb{R}^2} \Psi^Y((0, \gamma_v, \gamma_\xi), (0, u_v, u_\xi), \tau, y)$$
$$\cdot \iint_{\mathbb{R}^2} e^{-(\gamma_v + iu_v)v' - (\gamma_\xi + iu_\xi)\xi'} (\sqrt{av' + b\xi' + c} - K')^+ dv'd\xi' du_v du_\xi$$
$$= 50e^{-r_{f,\tau}\tau}\sqrt{\frac{a}{\pi}} \int_0^{\infty} \text{Re}$$
$$\times \left[\frac{\Psi^Y((0, \gamma_v, \frac{b}{a}\gamma_v), (0, u_v, \frac{b}{a}u_v), \tau, y)e^{\frac{c}{a}(\gamma_v + iu_v)}\text{erfc}\left(K'(\frac{\gamma_v + iu_v}{a})^{\frac{1}{2}}\right)}{(\gamma_v + iu_v)^{\frac{3}{2}}}\right] du_v$$

where $\gamma_v > 0$, $K' = K/100$, and $\text{erfc}(x) = \frac{2}{\sqrt{\pi}}\int_x^{\infty} e^{-t^2}dt$ is the complementary error function.

References

Aït-Sahalia, Y., Duarte, J., 2003. Nonparametric option pricing under shape restrictions. J. Econometrics 116, 9–47.
Aït-Sahalia, Y., Fan, J., Peng, H., 2009. Nonparametric transition-based tests for jump-diffusions. J. Amer. Statist. Assoc. 104, 1102–1116.
Aït-Sahalia, Y., Lo, A., 1998. Nonparametric estimation of state-price-densities implicit in financial asset prices. J. Finance 53, 499–547.
Aït-Sahalia, Y., Lo, A., 2000. Nonparametric risk management and implied risk aversion. J. Econometrics 94, 9–51.
Amengual, D., Xiu, D., 2012. Delving into Risk Premia: Reconciling Evidence from the S&P 500 and VIX Derivatives, Tech. Report. CEMFI and University of Chicago Booth School of Business.
Amengual, D., Xiu, D., 2014. Resolution of Policy Uncertainty and Sudden Declines in Volatility, Tech. Report. University of Chicago.
Bakshi, G., Cao, C., Chen, Z., 1997. Empirical performance of alternative option pricing models. J. Finance 52, 2003–2049.
Bakshi, G., Kapadia, N., 2003. Delta-hedged gains and the negative market volatility risk premium. Rev. Financial Stud. 16, 527–566.
Bakshi, G., Madan, D., 2008. Investor Heterogeneity and the Non-Monotonicity of the Aggregate Marginal Rate of Substitution in the Market Index, Working paper. University of Maryland.
Bakshi, G., Madan, D., Panayotov, G., 2014. Heterogeneity in beliefs and volatility tail behavior. J. Financial Quantitative Anal. (forthcoming).
Bansal, R., Kiku, D., Shaliastovich, I., Yaron, A., 2014. Volatility, the macroeconomy, and asset prices. J. Finance 69.
Bansal, R., Yaron, A., 2004. Risks for the long run: a potential resolution of asset pricing puzzles. J. Finance 59.
Bardgett, C., Gourier, E., Leippold, M., 2014. Inferring Volatility Dynamics and Risk Premia from the S&P 500 ad VIX Markets, Tech. Report. University of Zurich and Swiss Finance Institute (SFI).
Barndorff-Nielsen, O.E., Shephard, N., 2001. Non-Gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics. J. Roy. Statist. Soc. Ser. B 63, 167–241.
Bates, D.S., 2000. Post-'87 crash fears in the S&P 500 futures option market. J. Econometrics 94, 181–238.
Bates, D.S., 2012. US stock market crash risk, 1926–2010. J. Financial Econom. 105, 229–259.
Berzuini, C., Best, N.B., Gilks, W.R., Larizza, C., 1997. Dynamic conditional independence models and Markov chain Monte Carlo methods. J. Amer. Statist. Assoc. 92, 1403–1412.
Boes, M., Drost, F., Werker, B.J., 2007. Nonparametric Risk-Neutral Return and Volatility Distributions, Tech. Report. Tilburg University.
Bollerslev, T., Sizova, N., Tauchen, G., 2012. Volatility in equilibrium: asymmetries and dynamic dependencies. Rev. Finance 16, 31–80.
Bollerslev, T., Tauchen, G.E., Zhou, H., 2009. Expected stock returns and variance risk premia. Rev. Financial Stud. 22, 4463–4492.
Branger, N., Völkert, C., 2012. What is the Equilibrium Price of Variance Risk? A Long-Run Risks Model with Two Volatility Factors, Working paper.
Breeden, D., Litzenberger, R.H., 1978. Prices of state-contingent claims implicit in option prices. J. Bus. 51, 621–651.
Britten-Jones, M., Neuberger, A., 2000. Option prices, implied price processes, and stochastic volatility. J. Finance 55, 839–866.
Broadie, M., Chernov, M., Johannes, M.S., 2007. Model specification and risk premia: evidence from futures options. J. Finance 62.
Campbell, J., Christopher, P., Turley, B., Giglio, S., 2012. An Intertemporal CAPM with Stochastic Volatility, Tech. Report. Harvard University.
Campbell, J.Y., Shiller, R.J., 1988. Stock prices, earnings, and expected dividends. J. Finance 43, 661–676.
Carr, P., Geman, H., Madan, D.B., Yor, M., 2003. Stochastic volatility for Lévy processes. Math. Finance 13, 342–345.
Carr, P., Wu, L., 2009. Variance risk premiums. Rev. Financial Stud. 22, 1311–1341.
Chabi-Yo, F., 2012. Pricing kernels with stochastic skewness and volatility risk. Manage. Sci. 58 (3), 624–640.
Chabi-Yo, F., Garcia, R., Renault, E., 2008. State dependence can explain the risk aversion puzzle. Rev. Financial Stud. 21, 973–1011.
Chernov, M., Gallant, A.R., Ghysels, E., Tauchen, G.T., 2003. Alternative models for stock price dynamics. J. Econometrics 116, 225–257.
Chernov, M., Ghysels, E., 2000. A study towards a unified approach to the joint estimation of objective and risk neutral measures for the purpose of options valuation. J. Financial Economics 57, 407–458.
Christoffersen, P., Heston, S., Jacobs, K., 2009. The shape and term structure of the index option smirk: why multifactor stochastic volatility models work so well. Manage. Sci. 55, 1914–1932.
Christoffersen, P., Heston, S., Jacobs, K., 2013. Option anomalies and the pricing kernel. Rev. Financial Stud. 1963–2006.
Christoffersen, P., Jacobs, K., Mimouni, K., 2010. Models for S&P 500 dynamics: evidence from realized volatility, daily returns, and option prices. Rev. Financial Stud. 23, 3141–3189.
Christoffersen, P., Jacobs, K., Ornthanalai, C., Wang, Y., 2008. Option valuation with long-run and short-run volatility components. J. Financial Economics 90, 272–297.
Doucet, A., Malik, S., Pitt, M., 2014. Simulated likelihood inference for stochastic volatility models using continuous particle filtering. J. Econometrics 179, 99–111.
Drechsler, I., Yaron, A., 2011. What's vol got to do with it? Rev. Financial Stud. 24, 1–45.
Egloff, D., Leippold, M., Wu, L., 2010. The term structure of variance swap rates and optimal variance swap investments. J. Financial Quantitative Anal. 45, 1279–1310.
Epstein, L., Zin, S., 1989. Substitution, risk aversion, and the temporal behavior of consumption and asset returns: a theoretical framework. Econometrica 57, 937–969.
Eraker, B., 2004. Do stock prices and volatility jump? Reconciling evidence from spot and option prices. J. Finance 59.
Eraker, B., Johannes, M.S., Polson, N., 2003. The impact of jumps in equity index volatility and returns. J. Finance 58, 1269–1300.
Fan, J., Gijbels, I., 1996. Local Polynomial Modelling and its Applications. Chapman & Hall, London, U.K.
Fan, J., Yao, Q., Tong, H., 1996. Estimation of conditional densities and sensitivity measures in nonlinear dynamical systems. Biometrika 83, 189–206.
Fan, J., Yim, T.H., 2004. A crossvalidation method for estimating conditional densities. Biometrika 91, 819–834.

Gatheral, J., 2006. The Volatility Surface: A Practitioner's Guide. Wiley.
Gordon, N.J., Salmond, D.J., Smith, A.F.M., 1993. A novel approach to non-linear and non-Gaussian Bayesian state estimation. IEEE Proceedings F 140.
Hastie, T., Tibshirani, R., Friedman, J., 2001. The Elements of Statistical Learning: Data Mining, Inference and Prediction. Springer-Verlag, New York.
Jackwerth, J., 2000. Recovering risk aversion from option prices and realized returns. Rev. Financial Stud. 13, 433–451.
Jackwerth, J., Vilkov, G., 2013. Asymmetric Volatility Risk: Evidence from Option Markets, Working paper.
Johannes, M.S., Polson, N., 2009. Particle filtering. In: Handbook of Financial Time Series. Springer-Verlag.
Johannes, M.S., Polson, N., Stroud, J.R., 2009. Optimal filtering of jump diffusions: extracting latent states from asset prices. Rev. Financial Stud. 22.
Joshi, M., 2007. Log-type models, Homogeneity of Option Prices and Convexity, Tech. Report. Melbourne University.
Kanwal, R.P., 2004. Generalized Functions: Theory and Applications, third ed. Birkhäuser.
Kitagawa, G., 1996. Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. J. Comput. Graph. Statist. 5, 1–25.
Li, H., Zhao, F., 2009. Nonparametric estimation of state-price-densities implicit in interest rate cap prices. Rev. Financial Stud. 22, 4335–4376.
Li, J., Xiu, D., 2012. Spot Variance Regression, Tech. Report. Duke University and University of Chicago Booth School of Business.
Linn, M., Shive, S., Shumway, T., 2014. Pricing Kernel Monotonicity and Conditional Information, Tech. Report. University of Michigan.
Magnus, J.R., Neudecker, H., 1999. Matrix Differential Calculus with Applications in Statistics and Economics. Wiley.
Mencia, J., Sentana, E., 2012. Valuation of VIX derivatives. J. Financial Economics (forthcoming).
Pan, J., 2002. The jump-risk premia implicit in options: evidence from an integrated time-series study. J. Financial Economics 63, 3–50.
Papanicolaou, A., Sircar, R., 2014. A regime-switching Heston model for VIX and S&P 500 implied volatilities. Quant. Finance 14 (10), 1811–1827.
Pitt, M.K., Shephard, N., 1999. Filtering via simulation: auxiliary particle filters. J. Amer. Statist. Assoc. 94, 590–599.
Rosenberg, J.V., Engle, R.F., 2002. Empirical pricing kernels. J. Financial Economics 64, 341–372.
Ruppert, D., Sheather, S., Wand, M.P., 1995. An effective bandwidth selector for local least squares kernel regression. J. Amer. Statist. Assoc. 90.
Ruppert, D., Wand, M., 1994. Multivariate locally weighted least squares regression. Ann. Statist. 22, 1346–1370.
Shephard, N., 2005. Stochastic Volatility. Oxford University Press.
Tauchen, G.E., Todorov, V., 2011. Volatility jumps. J. Bus. Econom. Statist. 29.
Todorov, V., 2010. Variance risk premium dynamics: the role of jumps. Rev. Financial Stud. 23, 345–383.
Yao, Q., Tong, H., 1998. Cross-validatory bandwidth selections for regression estimation based on dependent data. J. Statist. Plann. Inference 68, 387–415.
Zhou, G., Zhu, Y., 2012. Volatility trading: what is the role of the long-run volatility component? J. Financial Quantitative Anal. 47, 273–307.
Ziegler, A., 2007. Why does implied risk aversion smile? Rev. Financial Stud. 20, 859–904.