JOURNAL OF FINANCIAL AND QUANTITATIVE ANALYSIS Vol. 50, No. 6, Dec. 2015, pp. 1389–1414
COPYRIGHT 2016, MICHAEL G. FOSTER SCHOOL OF BUSINESS, UNIVERSITY OF WASHINGTON, SEATTLE, WA 98195
doi:10.1017/S0022109015000642

https://doi.org/10.1017/S0022109015000642 Published online by Cambridge University Press

Heterogeneity in Beliefs and Volatility Tail Behavior

Gurdip Bakshi, Dilip Madan, and George Panayotov*

Abstract

We propose a model of volatility tail behavior in which investors display aversion to both low-volatility and high-volatility states, and hence, the derived pricing kernel exhibits an increasing and decreasing region in the volatility dimension. The model features investors who have heterogeneity in beliefs about volatility outcomes and maximize their utility by choosing volatility-contingent cash flows. Our empirical examination suggests that the model is better suited to reproduce data features in the left tail of the volatility distribution, both qualitatively and quantitatively.

I. Introduction

The purpose of this article is to propose a new framework for characterizing volatility tail behavior. Our central insight is that by allowing for heterogeneity in beliefs about volatility outcomes, one can obtain pricing kernels that are better suited to reconcile features, germane to traded market volatility, and to mimic the data counterparts, both qualitatively and quantitatively.

To motivate our modeling approach, let $v \in (0, \infty)$ be the market return volatility, where, for example, v^2 could be an estimate of the quadratic variation in the log of the market index. Suppose that the physical density of v, denoted by $\Phi[v]$, and the pricing (or risk-neutral) density, denoted by $q[v]$, are both lognormal. Specifically, $\Phi[v] = 1/(v\,\sigma_p\sqrt{2\pi})\exp(-1/(2\sigma_p^2)\ (\ln(v) - \mu_p)^2)$, $q[v] = 1/(v\,\sigma_q\sqrt{2\pi})\exp\left(-1/(2\sigma_q^2)(\ln(v) - \mu_q)^2\right)$, and hence, the change of measure (or pricing kernel) is $d\mathbb{Q}/d\mathbb{P} = \psi_0\ \exp\left(\psi_1\ (\ln(v))^2 + \psi_2\ \ln(v)\right)$, where ψ_0 is

*Bakshi (corresponding author), gbakshi@rhsmith.umd.edu, Madan, dbm@rhsmith.umd.edu, University of Maryland, Smith School of Business, College Park, MD 20742; and Panayotov, panayotov@ust.hk, Hong Kong University of Science and Technology, School of Business, Clearwater Bay, Hong Kong. The efforts of Hendrik Bessembinder (the editor) and Dacheng Xiu (the referee) have greatly improved the paper. The authors acknowledge helpful feedback from Charles Cao, Peter Carr, Fousseni Chabi-Yo, Xiaohui Gao, Steve Heston, Nikunj Kapadia, Tao Li, Huston McCulloch, Georgios Skoulakis, Réne Stulz, Ingrid Werner, Liuren Wu, and Jin Zhang. An earlier version of the paper was presented at the Chinese University of Hong Kong, Hong Kong University, Hong Kong University of Science and Technology, City University of Hong Kong, University of Maryland, and Ohio State University. We welcome comments, including references to related papers we have inadvertently overlooked. Any remaining errors are our responsibility alone.

a constant, $\psi_1 = 1/(2\sigma_p^2) - 1/(2\sigma_q^2)$, and $\psi_2 = \mu_q/\sigma_q^2 - \mu_p/\sigma_p^2$. It is seen now that if $\sigma_q = \sigma_p$, as in many standard models, then $d\mathbb{Q}/d\mathbb{P}$ could be monotonically increasing in the volatility dimension, which would imply geometrically that $q[v]$ crosses $\Phi[v]$ only once, and would entail an asset-pricing model where deeper out-of-the-money (OTM) volatility puts have higher expected returns. In contrast, if, for example, $\sigma_q > \sigma_p$, then $d\mathbb{Q}/d\mathbb{P}$ exhibits a decreasing region for low volatility levels, and increasing region for high volatility levels, implying that $q[v]$ crosses $\Phi[v]$ twice. OTM volatility puts and calls in this setting manifest low expected returns. Besides, expected returns could be decreasing for deeper OTM volatility puts.

Our interest in a model where investors display aversion to both low- and high-volatility states, giving rise to a nonmonotonic change of measure for volatility, is guided, intuitively, by certain specifics of the volatility market. In particular, i) the underlying volatility is not traded directly, but only via volatility-contingent claims, and ii) unlike equities, these claims are in 0 net supply and offer exposure to volatility movements on both the upside and the downside, for different investors (see also Chicago Board Options Exchange (CBOE) (2009)). The model prediction that both puts and calls on volatility will have low expected returns can be tested using newly available data from the VIX option market.

In our model of the volatility market, investors display heterogeneity in beliefs about volatility outcomes. Posited within the model are two types of investors who anticipate volatility to increase or to decline, respectively. In our setup, investors maximize their utility by choosing volatility-contingent cash flows, given their beliefs about the distribution of volatility, and we derive the pricing kernel in closed form.

We also contrast our model to a conventional approach, where the pricing kernel is monotonically declining in market returns and increasing in market return volatility, which reflects aversion to volatility. Under certain conditions, we show that a model of this type, where high-volatility states are disliked, predicts that the expected returns of puts (calls) on volatility are decreasing in strike and are positive (negative) for sufficiently low (high) strikes. Comparing these predictions with those from our model can be important, as they can be traced to ideas that are often at the core of finance theory, including mean-variance theory.

Elaborating, we explore whether there is empirical support for our model with heterogeneity in beliefs versus the conventional approach to modeling volatility tail behavior. We show that the two models share the same qualitative prediction about the pattern of expected returns of volatility calls and, hence, focus on the hypotheses that the expected returns of volatility puts either i) decrease in strike or ii) increase in strike. The results support, more generally, the model with heterogeneity in beliefs in our sample, which appears to be better suited for capturing the left-tail behavior of traded volatility.

Our article provides insights similar to Song and Xiu (2013), who estimate a U-shaped volatility pricing kernel, confirming empirically that investors display aversion to both low- and high-volatility states. Amengual and Xiu (2013) investigate the origin and differences in the risk pricing specifically of volatility jumps in the two tails, and Aït-Sahalia, Karaman, and Mancini (2013) study the risk premium from the term structure of variance swaps. Furthermore, Jackwerth and

https://doi.org/10.1017/S0022109015000642 Published online by Cambridge University Press

Vilkov (2013) examine the joint behavior of the risk-neutral distributions of the market index and its expected volatility, and show how to extract nonparametrically the expected risk-neutral correlation between them. We differ from many extant studies by demonstrating that a possible driver of the risk pricing for low- and high-volatility states is heterogeneity in beliefs about volatility outcomes, which could generate low average returns for OTM volatility calls and puts, which are also decreasing in moneyness.[1] Xiong (2013) provides an overview of the literature on the importance of modeling heterogeneity in beliefs in many empirically relevant contexts.

The modeling and empirical issues that are broached in our article are likely to gain traction, given the advent of active trading in volatility. Besides, our discussions could complement the insights from studies that rely on realized volatility, index-option inferred volatility, and variance swaps (e.g., Rubinstein (1994), Andersen, Bollerslev, Diebold, and Labys (2003), Carr and Wu (2009), Todorov (2010), Bollerslev and Todorov (2011), and Christoffersen, Heston, and Jacobs (2013)). Our contribution aims to fill some gaps from the perspective of the left tail of the volatility distribution.

II. A Model with Traded Volatility and Heterogeneity in Beliefs

This section first states our model assumptions, then describes the choice problem of the investors. Next, we present the solution for the pricing kernel and study its properties.

A. A Model of the Volatility Market

Assumption 1. The model is two period, with three dates (denoted 0, T, T').

In analogy with the VIX, and to provide a link to our data on returns of volatility options, volatility $v_T \in (0, \infty)$ reflects uncertainty over $[T, T']$, and without loss of generality, we set the current level of volatility to $v_0 = 1$. There is a continuum of realizations of volatility.

It is worth noting that although our model is cast in a 2-period setting, we also considered an analogue that treats volatility exposure in continuous time. Under the assumption that volatility follows a diffusion process, our analysis in Appendix B shows that the derived pricing kernel is monotonic in volatility. In contrast, we recognize that under richer volatility dynamics in continuous time, which allow for two-sided jumps, a nonmonotonic kernel could be obtained (e.g., Amengual and Xiu (2013), Song and Xiu (2013)).

[1] In the spirit of the index options literature, there is interest in understanding how alternative models of volatility dynamics fare in fitting the prices of VIX options and other volatility derivatives, and in producing accurate forecasts of quadratic variation. Among others, this promising line of research includes Andersen, Bollerslev, and Meddahi (2005), Buehler (2006), Zhu and Zhang (2007), Carr and Lee (2008), Gatheral (2008), Sepp (2008), Aït-Sahalia and Mancini (2008), Cont and Kokholm (2013), Lin (2009), and Egloff, Leippold, and Wu (2010).

https://doi.org/10.1017/S0022109015000642 Published online by Cambridge University Press

To examine the implications of nonmonotonic pricing kernels in the setting of a volatility market with heterogeneity in beliefs about volatility outcomes, we pursue a discrete-time model that is suited to our objectives and that could offer the flexibility of using general change of measure densities (see, e.g., Neftci (2007)).

Assumption 2. Options on volatility of all strikes are traded in the volatility market.

Any function of volatility v_T can be synthesized with a portfolio of options on volatility. The market for volatility is complete.

Assumption 3. The contingent claims on volatility are in 0 net supply in the aggregate.

This assumption ensures that the contingent cash flows, purchased by some investors, are sold by others.

Assumption 4. Investors can be divided into two types, denoted F and G, which are differentiated by their beliefs about volatility. They have personalized densities about volatility outcomes, denoted $\Phi_f[v_T]$ and $\Phi_g[v_T]$, respectively. In the model, the two investor types are present with mass $0 < \phi < 1$ and $1 - \phi$, respectively.

Differences in beliefs, and possibly risk aversion, can give rise to trading in contingent claims on volatility. We suppose the existence of a physical density $\Phi[v_T]$ and a pricing density $q[v_T]$.

For simplicity, the interest rate is set equal to 0, and hence, we omit discounting. The pricing density $q[v_T]$ is determined endogenously so as to clear the volatility market. Our approach generalizes to multiple types of investors.

Importantly, one should not construe our framework to imply that the future level of volatility is treated by investors as a variable independent from the past. On the contrary, we describe the investors' beliefs about future volatility via time 0 conditional densities, which implicitly incorporate all past information. Therefore, we adhere to the standard approach for representing the role of the past.

When the probability measures corresponding to $\Phi_f[v_T]$ and $\Phi_g[v_T]$ are absolutely continuous with respect to $\Phi[v_T]$, the change of measure densities for the two types of investors can be defined as follows:

$$(1) \qquad \mathcal{A}[v_T] = \frac{\Phi_f[v_T]}{\Phi[v_T]}, \qquad \mathcal{B}[v_T] = \frac{\Phi_g[v_T]}{\Phi[v_T]}.$$

It is assumed that the densities $\Phi_f[v_T]$, $\Phi_g[v_T]$, and $\Phi[v_T]$ belong to a finite parameter class, are twice continuously differentiable, and have finite moments up to order 4.

Assumption 5. Investors F and G hold volatility-contingent cash flows, denoted $f[v_T]$ and $g[v_T]$, respectively, that are self-financed, that is,

$$(2) \qquad 0 = \int f[v_T]\, q[v_T]\, dv_T \qquad \text{and} \qquad 0 = \int g[v_T]\, q[v_T]\, dv_T,$$

with $f \in \mathcal{C}$ and $g \in \mathcal{C}$, where $\mathcal{C} = \{c | c : \Re^+ \to \Re\}$.

https://doi.org/10.1017/S0022109015000642 Published online by Cambridge University Press

Equation (2) reflects the cost of synthesizing the volatility-contingent cash flows $f[v_T]$ and $g[v_T]$.

Assumption 6. Both types of investors have a utility function, defined over their respective volatility-contingent cash flow, and in the constant absolute risk aversion class (with $\gamma_f > 0$ and $\gamma_g > 0$):

$$(3) \quad U(f[v_T]) = \left(\frac{-1}{\gamma_f}\right) e^{-\gamma_f f[v_T]}, \qquad U(g[v_T]) = \left(\frac{-1}{\gamma_g}\right) e^{-\gamma_g g[v_T]}.$$

In specification (3), the exponential function guarantees real valued utility, even when the volatility-contingent cash flow is negative. The self-financing condition in equation (2) implies that both $f[v_T]$ and $g[v_T]$ exhibit a range of negative values; hence, for example, adopting an isoelastic utility does not result in a real-valued utility.

B. Choice Problem of Investors

It may be possible to characterize $f[v_T]$ and $g[v_T]$ endogenously and in analytical closed form. The choice problem of investors F and G can be stated as,

Problem 1. Choose $f \in \mathcal{C}$ and $g \in \mathcal{C}$ to maximize expected utility under the personalized densities, subject to the self-financing constraint (under 0 interest rate):

$$(4) \quad f^* = \underset{f \in \mathcal{C}}{\text{argmax}} \int \left(\frac{-1}{\gamma_f}\right) e^{-\gamma_f f[v_T]} \mathcal{A}[v_T] \Phi[v_T] \, dv_T$$
$$\text{subject to} \int f[v_T] \, q[v_T] \, dv_T = 0,$$

$$(5) \quad g^* = \underset{g \in \mathcal{C}}{\text{argmax}} \int \left(\frac{-1}{\gamma_g}\right) e^{-\gamma_g g[v_T]} \mathcal{B}[v_T] \Phi[v_T] \, dv_T$$
$$\text{subject to} \int g[v_T] \, q[v_T] \, dv_T = 0,$$

and the market-clearing condition for volatility-contingent claims,

$$(6) \quad \phi f^*[v_T] + (1 - \phi) \, g^*[v_T] = 0.$$

When the change of measure densities are parameterized in a way such that $\mathcal{A}[v_T]$ ($\mathcal{B}[v_T]$) is increasing (decreasing), then investors F (G) anticipate volatility to rise (fall).

In equation (4), observe that maximizing $\int (-1/\gamma_f) e^{-\gamma_f f[v_T]} \mathcal{A}[v_T] \Phi[v_T] dv_T$ is equivalent to minimizing $\int (1/\gamma_f) e^{-\gamma_f f[v_T]} \mathcal{A}[v_T] \Phi[v_T] dv_T$, which is i) a convex function of $f[v_T]$ and ii) bounded below by 0. At the same time, a strategy of 0 cash flow satisfies the self-financing constraint, with an expected utility of $-1/\gamma_f$. It follows that the optimal expected utility lies between $-1/\gamma_f$ and 0.

Using the optimal volatility-contingent cash flows $f[v_T]$ and $g[v_T]$ that solve the first-order conditions in equation (A-1) in Appendix A and satisfy equation (6)

https://doi.org/10.1017/S0022109015000642 Published online by Cambridge University Press

(and suppressing the $*$ superscript on $f^*[v_T]$ and $g^*[v_T]$ from now on), our goal is to derive the pricing kernel

$$m[v_T] \equiv \frac{q[v_T]}{\Phi[v_T]} \tag{7}$$

under suitable assumptions about the change of measure densities $\mathcal{A}[v_T]$ and $\mathcal{B}[v_T]$, and then to investigate the model properties, especially the tail behavior of the volatility pricing distribution.

C. Nonmonotonic Pricing Kernel in the Volatility Market

The following theorem presents the solution for the pricing kernel from Problem 1, together with conditions under which the pricing kernel is nonmonotonic in volatility. Section II.D builds on our results and develops examples that explicitly parameterize the physical and personalized densities of the investors.

Theorem 1. The following statements hold:

Part I. The pricing kernel is related to the change of measure densities as,

$$m[v_T] = m_0 \left(\mathcal{A}[v_T]\right)^{\eta} \left(\mathcal{B}[v_T]\right)^{1-\eta}, \qquad \eta \equiv \frac{\phi\,\gamma_g}{(1-\phi)\gamma_f + \phi\gamma_g}, \tag{8}$$

where m_0 is a constant of integration, chosen to ensure that $\int m[v_T]p[v_T]dv_T = 1$. The optimal volatility-contingent cash flows are,

$$\begin{aligned} f[v_T] &= f_0 + \left(\frac{(1-\phi)}{(1-\phi)\gamma_f + \phi\gamma_g}\right) \ln\left(\frac{\mathcal{A}[v_T]}{\mathcal{B}[v_T]}\right) \quad \text{and} \\ g[v_T] &= -\left(\frac{\phi}{1-\phi}\right) f[v_T], \end{aligned} \tag{9}$$

where f_0 in equation (9) is chosen to satisfy the self-financing constraint $\int f[v_T]\, q[v_T]\, dv_T = 0$.

Part II. Suppose that over some interval $v_T \in [\underline{v}, \overline{v}]$,

$$\text{(a) } \mathcal{A}'[v_T] \equiv \frac{d\mathcal{A}[v_T]}{dv_T} > 0 \quad \text{and} \tag{10}$$

$$\mathcal{B}'[v_T] \equiv \frac{d\mathcal{B}[v_T]}{dv_T} < 0,$$

$$\text{(b) } \left(\ln(\mathcal{A}[v_T])\right)'\Big|_{v=\underline{v}} \approx 0 \quad \text{and} \tag{11}$$

$$\left(\ln(\mathcal{B}[v_T])\right)'\Big|_{v=\overline{v}} \approx 0, \quad \text{and}$$

$$\text{(c) } \mathcal{A}[v_T] \text{ and } \mathcal{B}[v_T] \text{ are log-convex.} \tag{12}$$

Condition (c) is sufficient for the pricing kernel $m[v_T]$ to be log-convex. Additionally, conditions (a) and (b) ensure that $m[v_T]$ will have a minimum in the interior of the interval $v_T \in [\underline{v}, \overline{v}]$ and will admit a negatively (positively) sloped region at low (high) volatility levels.

Proof. See Appendix A. □

https://doi.org/10.1017/S0022109015000642 Published online by Cambridge University Press

Part I of Theorem 1 provides a characterization of the pricing kernel, which depends on the assumed change of measure densities. Importantly, the pricing kernel in equation (8) is determined by the beliefs of both types of investors in the volatility market, as the exponents on $\mathcal{A}[v_T]$ and $\mathcal{B}[v_T]$ are both positive. Furthermore, $m[v_T]$ is guaranteed to be positive.

With the pricing kernel determined, the pricing density $q[v_T]$ can be recovered accordingly, given the physical density $\Phi[v_T]$. Even if a relevant question remains about the magnitude of the exponent η, the proportion of each type of investors is assumed to be nonzero.

Although, in general, the functional equations from the optimization problem in equations (4)–(6) can be difficult to solve, the solution for the optimal $f[v_T]$ offered in equation (9) is made tractable under two features, namely: i) exponential utility and ii) 0 net supply for contingent claims for volatility. We will demonstrate that the solution for $m[v_T]$ in equation (8) provides flexibility and could possibly reconcile certain features of the observed pattern of VIX option returns under suitable parameterizations of $\mathcal{A}[v_T]$ and $\mathcal{B}[v_T]$.

The setup of Theorem 1 (and Problem 1) accommodates heterogeneity in beliefs, through $\mathcal{A}[v_T]$ and $\mathcal{B}[v_T]$, yet differs from the setup in, for example, Leland (1980), Detemple and Murthy (1994), Basak (2000), Calvet, Grandmont, and Lemaire (2004), Buraschi and Jiltsov (2006), Jouini and Napp (2007), Bakshi, Madan, and Panayotov (2010), and Xiong and Yan (2010). Our setup also differs from the corresponding setup for futures on the equity market index (or other traded assets), where the futures are also in 0 net supply, but both the spot-futures arbitrage condition and the martingale condition for the spot-price process must be taken into account when solving for equilibrium (e.g., Grossman and Stiglitz (1980), Briys, Crouhy, and Schlesinger (1990), and Hong (2000)).

What can be said about the concavity or the convexity of $g[v_T]$? Although not much can be inferred analytically without explicitly parameterizing $\mathcal{A}[v_T]$ and $\mathcal{B}[v_T]$, we nonetheless note, based on equation (9), that a sufficient condition for the convexity of $g[v_T]$ is that $\mathcal{A}[v]/\mathcal{B}[v]$ be concave, as log preserves concavity.

When $\mathcal{A}'[v_T] > 0$ ($\mathcal{B}'[v_T] < 0$), as postulated in condition (a) of Theorem 1 (Part II), it is provable that the mean of the personalized density $\Phi_f[v_T]$ ($\Phi_g[v_T]$) is higher (lower) than the mean of the physical density $\Phi[v_T]$. Therefore, the pricing kernel in Part II of Theorem 1 is designed to reflect the impact of investors with opposite beliefs about volatility outcomes. Moreover, the belief distributions of the investors are restricted to maintain a pricing (risk-neutral) density that dominates the physical density in both tails of the volatility distribution.

What specific change of measure densities $\mathcal{A}[v_T]$ and $\mathcal{B}[v_T]$ may satisfy the assumptions in Part II of Theorem 1? One example would be $\mathcal{A}[v_T] = \exp(\mu_a\, v_T^2)$ and $\mathcal{B}[v_T] = \exp(\mu_b\, v_T^{-1})$, for $\mu_a > 0$ and $\mu_b > 0$. In this case, $m[v_T] = m_0 \exp\left(\eta\mu_a v_T^2 + (1-\eta)\mu_b v_T^{-1}\right)$; hence, the pricing kernel is log-convex and has a decreasing (increasing) region at low (high) volatility levels. Alternatively, if $\mathcal{A}[v_T] = \exp(e^{\mu_a (v_T - 1)})$ and $\mathcal{B}[v_T] = \exp(e^{-\mu_b (v_T - 1)})$, then $\ln(m[v_T])$ is a generalized hyperbolic cosine function and has a minimum for v_T close to unity when $\eta\mu_a \approx (1-\eta)\mu_b$.

Under the assumption on the change of measure densities in equation (10), the optimal $g[v_T]$ is decreasing, being the logarithm of the ratio of a decreasing

https://doi.org/10.1017/S0022109015000642 Published online by Cambridge University Press

change of measure ($\mathcal{B}'[v_T] < 0$) and an increasing change of measure ($\mathcal{A}'[v_T] > 0$), and by analogy, the optimal $f[v_T]$ is increasing. Consequently, investors G have short volatility exposure and investors F have long volatility exposure.

One additional clarification relates to the question: How is the shape of $m[v_T]$ related to the properties of the cash flows $f[v_T]$ and $g[v_T]$? At the least, it can be shown that a necessary condition for the pricing kernel to be nonmonotonic is that $f[v_T]$ and $g[v_T]$ are nonlinear in v_T.

To establish this result, we return to equations (8) and (9) of Theorem 1 and consider change of measure densities of the form: $\mathcal{A}[v_T] = e^{\mu_a v_T}$ and $\mathcal{B}[v_T] = e^{-\mu_b v_T}$, with $\mu_a > 0$ and $\mu_b > 0$. In this case,

$$
\begin{aligned}
(13) \qquad f[v_T] &= f_0 + \left(\frac{(1-\phi)(\mu_a + \mu_b)}{(1-\phi)\gamma_f + \phi\gamma_g} \right) v_T, \\
g[v_T] &= -\frac{\phi}{(1-\phi)} f[v_T], \qquad m[v_T] = m_0 \, e^{\xi v_T},
\end{aligned}
$$

where $\xi \equiv \eta\,\mu_a - (1-\eta)\,\mu_b$ and $m_0 = (\int e^{\xi v_T} \Phi[v_T] dv_T)^{-1}$. Equation (13) shows that the linearity of $f[v_T]$ and $g[v_T]$ is accompanied by a monotonic pricing kernel, which has two further ramifications. First, contingent claims with payout linear in v_T may not allow one to fully discern the tail behavior of traded volatility. Second, to obtain nonmonotonicity of the pricing kernel in our framework with exponential utility, the change of measure densities, $\mathcal{A}[v_T]$ and $\mathcal{B}[v_T]$, must be kept outside the exponential class.

There are two features of the volatility market that distinguish it from other futures markets, for example, those on the Standard & Poor's (S&P) 500 equity index. Specifically, the mean reversion and nontradability of volatility allow the delinking of pricing kernels at different dates, whereas tradability enforces restrictions on the pricing kernels at different dates. However, these issues are not explored in this article.

In a setting different from ours, Song and Xiu (2013) estimate nonparametrically a U-shaped pricing kernel of volatility using VIX options, and their finding serves as direct evidence in support of the theoretical findings of our article. In addition, Amengual and Xiu (2013) use variance swaps to study downward jumps in volatility. The models in both these studies are cast in continuous time with jumps, and point to the possibility of obtaining a U-shaped pricing kernel in a different framework.

D. Model Implications in Example Economies

The implications in Part II of Theorem 1, that is, an asset-pricing model where investors exhibit aversion to both low- and high-volatility states, are illustrated on examples that parameterize the belief distribution of investors.

1. Associated Pricing Kernels

Example 1. Suppose volatility v_T is distributed lognormally under the physical measure:

$$
(14) \qquad \Phi[v_T; \mu_p, \sigma_p] = \frac{1}{v_T\,\sigma_p\,\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma_p^2} \{\ln(v_T) - \mu_p\}^2 \right).
$$

https://doi.org/10.1017/S0022109015000642 Published online by Cambridge University Press

The personalized densities of investors F and G are also lognormal, with parameters μ_f and σ_f, and μ_g and σ_g, respectively.

From equation (1), the log change of measure densities are quadratic in $\ln(v_T)$:

$$\ln(\mathcal{A}[v_T]) = \ln\left(\frac{\sigma_p}{\sigma_f}\right) - \frac{(\mu_f - \ln(v_T))^2}{2\sigma_f^2} + \frac{(\mu_p - \ln(v_T))^2}{2\sigma_p^2}, \tag{15}$$

$$\ln(\mathcal{B}[v_T]) = \ln\left(\frac{\sigma_p}{\sigma_g}\right) - \frac{(\mu_g - \ln(v_T))^2}{2\sigma_g^2} + \frac{(\mu_p - \ln(v_T))^2}{2\sigma_p^2}. \tag{16}$$

We specify that

$$\mu_f > \mu_p > \mu_g, \qquad \sigma_f > \sigma_p, \qquad \text{and} \qquad \sigma_g > \sigma_p, \tag{17}$$

differentiating the beliefs about volatility outcomes of the two types of investors. □

From equations (15) and (16), the change of measures for investors F and G satisfy

$$\begin{aligned} \frac{\mathcal{A}'[v_T]}{\mathcal{A}[v_T]} &= \frac{\left(\sigma_f^2 - \sigma_p^2\right)\ln(v_T) - \mu_p\sigma_f^2 + \mu_f\sigma_p^2}{v_T\,\sigma_p^2\sigma_f^2}, \\ \frac{\mathcal{B}'[v_T]}{\mathcal{B}[v_T]} &= \frac{\left(\sigma_g^2 - \sigma_p^2\right)\ln(v_T) - \mu_p\sigma_g^2 + \mu_g\sigma_p^2}{v_T\,\sigma_p^2\sigma_g^2}. \end{aligned} \tag{18}$$

It follows that $\mathcal{A}[v_T]$ is increasing for $v_T > \underline{v}$, and $\mathcal{B}[v_T]$ is decreasing for $v_T < \overline{v}$, where

$$\begin{aligned} \underline{v} &= \exp\left(-\left(\frac{\sigma_p^2\sigma_f^2}{\sigma_f^2 - \sigma_p^2}\right)\left(\frac{\mu_f}{\sigma_f^2} - \frac{\mu_p}{\sigma_p^2}\right)\right), \\ \overline{v} &= \exp\left(\left(\frac{\sigma_p^2\sigma_g^2}{\sigma_g^2 - \sigma_p^2}\right)\left(\frac{\mu_p}{\sigma_p^2} - \frac{\mu_g}{\sigma_g^2}\right)\right). \end{aligned} \tag{19}$$

Our focus is on change of measure parameterizations, for which $[\underline{v}, \overline{v}]$ is a suitably wide interval over which $\mathcal{A}[v_T]$ and $\mathcal{B}[v_T]$ in equations (15) and (16) are bounded.

Invoking equation (8) of Theorem 1,

$$\ln(m[v_T]) = \frac{\alpha}{2}(\ln(v_T))^2 + \beta\,\ln(v_T) + (\varsigma + \ln(m_0)), \tag{20}$$

where m_0 is a constant of integration, ensuring that $\int m[v_T]\Phi[v_T]dv_T = 1$, and

$$\alpha = \eta\left(\frac{1}{\sigma_p^2} - \frac{1}{\sigma_f^2}\right) + (1-\eta)\left(\frac{1}{\sigma_p^2} - \frac{1}{\sigma_g^2}\right), \tag{21}$$

$$\beta = \eta\left(\frac{\mu_f}{\sigma_f^2} - \frac{\mu_p}{\sigma_p^2}\right) + (1-\eta)\left(\frac{\mu_g}{\sigma_g^2} - \frac{\mu_p}{\sigma_p^2}\right), \tag{22}$$

https://doi.org/10.1017/S0022109015000642 Published online by Cambridge University Press

$$\varsigma = \eta\left(\ln\left(\frac{\sigma_p^2}{\sigma_f^2}\right)+\frac{\mu_p^2}{2\sigma_p^2}-\frac{\mu_f^2}{2\sigma_f^2}\right) + (1-\eta)\left(\ln\left(\frac{\sigma_p^2}{\sigma_g^2}\right)+\frac{\mu_p^2}{2\sigma_p^2}-\frac{\mu_g^2}{2\sigma_g^2}\right). \tag{23}$$

Importantly, the belief distributions implied by equation (17) ensure that $\alpha > 0$, and consequently the quadratic function in equation (20), decreases for low values of volatility and increases for high values of volatility. The log pricing kernel is convex in $\ln(v_T)$ and achieves a minimum at $\exp(-\beta/\alpha)$.

A greater dispersion in the personalized volatility distributions, relative to the physical distribution, as in equation (17), may be needed to support a broader class of pricing kernels in the volatility market, since if $\sigma_f = \sigma_g = \sigma_p$, then the pricing kernel is monotonic in v_T (as $\alpha = 0$).

If $\sigma_f < \sigma_p$ and $\sigma_g < \sigma_p$, then a hump-shaped pricing kernel is obtained, which implies that the physical density dominates the pricing density in both tails of the volatility distribution. Such an implication is also counterfactual.

From equation (9), the optimal volatility-contingent cash flow for investors F is:

$$f[v_T] = f_0 + \frac{(1-\phi)}{(1-\phi)\gamma_f + \phi\gamma_g}\left\{\ln\left(\frac{\sigma_g}{\sigma_f}\right)+\frac{1}{2\sigma_g^2}\left(\mu_g-\ln(v_T)\right)^2 - \frac{1}{2\sigma_f^2}\left(\mu_f-\ln(v_T)\right)^2\right\}, \tag{24}$$

and the optimal volatility-contingent cash flow for investors G is $g[v_T] = -(\phi/(1-\phi)f[v_T]$. Furthermore, over $[\underline{v}, \overline{v}]$, where $\mathcal{A}'[v_T] > 0$ and $\mathcal{B}'[v_T] < 0$, the optimal cash flow $f[v_T]$ is increasing and $g[v_T]$ is decreasing.

It may be observed that $f[v_T]$ can be concave or convex over various intervals,

$$f''[v_T] = \frac{1}{v_T^2}\left(\frac{(1-\phi)}{(1-\phi)\gamma_f+\phi\gamma_g}\right) \times \left(\left(\frac{1}{\sigma_f^2}-\frac{1}{\sigma_g^2}\right)\ln(v_T) - \frac{1}{\sigma_f^2}(\mu_f+1) + \frac{1}{\sigma_g^2}(\mu_g+1)\right). \tag{25}$$

For instance, when $\sigma_f < \sigma_g$, one could get $f''[v_T] < 0$ for low values of v_T and $f''[v_T] > 0$ for high values of v_T. The converse obtains when $\sigma_f > \sigma_g$. When $\sigma_f = \sigma_g$, then $f[v_T]$ is concave over $[\underline{v}, \overline{v}]$.

The crucial point to observe is that either investors F or investors G are holding convex functions of volatility in the aggregate over some intervals.

To highlight some additional aspects of our example economy, Figure 1 plots the pricing kernel and the optimal cash flows $f[v_T]$ and $g[v_T]$, based on the following parameters:

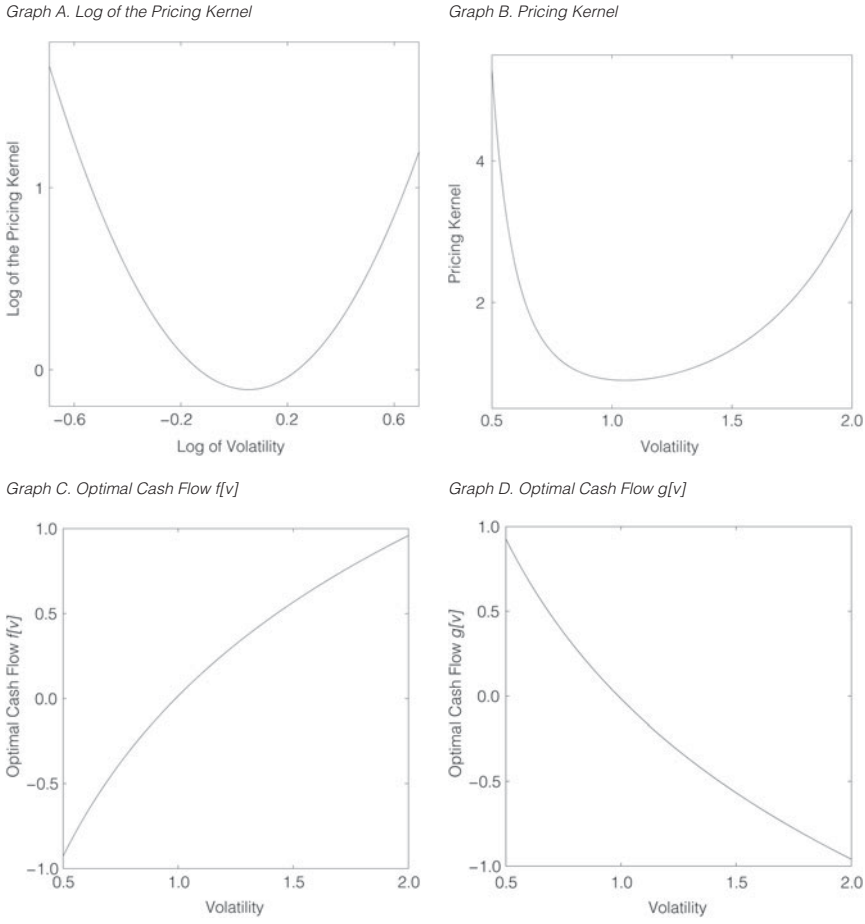
$$\mu_p = 0.000, \quad \mu_f = 0.150, \quad \mu_g = -0.150, \quad \sigma_p = 0.171, \quad \sigma_f = 0.184, \quad \sigma_g = 0.184, \tag{26}$$

https://doi.org/10.1017/S0022109015000642 Published online by Cambridge University Press

FIGURE 1

Pricing Kernel and Optimal Volatility-Contingent Cash Flows from Example 1

Figure 1 plots i) the pricing kernel $m[v_T]$ and ii) the optimal volatility-contingent cash flows, $f[v_T]$ and $g[v_T]$, for the two types of investors. The physical density (as shown in equation (14)) and the personalized densities are assumed to be lognormal, and the personalized change of measure densities are as shown in equations (15) and (16). We use $\mu_p = 0.000$, $\mu_f = 0.150$, $\mu_g = -0.150$, $\sigma_p = 0.171$, $\sigma_f = 0.184$, and $\sigma_g = 0.184$. The coefficients of absolute risk aversion are set to $\gamma_f = 3.5$, $\gamma_g = 3.0$, and $\phi = 0.50$. Graphs A and B are based on equation (20), and Graphs C and D, which plot the optimal volatility-contingent cash flows, are based on equation (24).



with $\gamma_f = 3.5$, $\gamma_g = 3.0$, and $\phi = 0.50$.[2] Under the assumed parameterization, $v_T \in (0.50, 2.00)$.

[2]In our illustration, the assumed values of μ_p and σ_p roughly correspond to the mean and standard deviation of monthly changes in the log of VIX over Jan. 1990–Dec. 2013. We also note that the monthly log changes in the VIX are uncorrelated according to the Ljung–Box statistic at lag 12 (p-value of 0.229). The maximum-likelihood estimate of the first-order autoregressive coefficient is -0.125 (p-value of 0.058). Furthermore, the frequency of observations in certain intervals of monthly log changes in the VIX are tabulated below:

Log change in the VIX (%)	<-0.2	$[-0.2, -0.1)$	$[-0.1, 0)$	$(0, 0.1]$	$(0.1, 0.2]$	>0.2
Frequency (%)	0.10	0.18	0.25	0.21	0.13	0.12

https://doi.org/10.1017/S0022109015000642 Published online by Cambridge University Press

The derived pricing kernel is declining at low levels of volatility and then increasing. It is this property of the pricing kernel that could impart negative returns of volatility puts. Incorporating the behavior of the two types of investors with $\mu_f > \mu_p$ and $\mu_g < \mu_p$ could give rise to a pricing kernel that is consistent with negative returns of both OTM volatility puts and calls.

In general, a relative increase in γ_f steepens the slope of the pricing kernel in the region of low volatility levels and shifts it to the right, a trait that can broadly induce more negative put returns. When $\phi\gamma_g = (1-\phi)\gamma_f$, then $\eta = 0.50$ (from equation (8)), and the response of $m[v_T]$ to v_T is symmetric.

Recapitulating the implications of Example 1, the optimal cash flows can be concave or convex over various intervals, and the pricing kernel can be asymmetric, displaying both a decreasing region for low levels of volatility and an increasing region for high levels of volatility.

Example 2. Keep the physical density, $\Phi[v_T; \mu_p, \sigma_p]$, lognormal, as in equation (14), but assume an alternative functional form for the change of measures for the two types of investors:

$$\mathcal{A}[v_T] = \varpi_a + (1-\varpi_a)\left(\frac{1}{\sigma_a\sqrt{2\pi}}\int_{-\infty}^{\ln(v_T)} \exp\left(-\frac{\{x-\mu_a\}^2}{2\sigma_a^2}\right) dx\right), \tag{27}$$

$$\mathcal{B}[v_T] = \varpi_b + (1-\varpi_b)\left(\frac{1}{\sigma_b\sqrt{2\pi}}\int_{-\infty}^{-\ln(v_T)} \exp\left(-\frac{\{x-\mu_b\}^2}{2\sigma_b^2}\right) dx\right), \tag{28}$$

where $\varpi_a \in (0,1)$ and $\varpi_b \in (0,1)$. The parameters μ_i and σ_i, for $i = \{a, b\}$, control the mean and standard deviation of the cumulative Normal distribution in equations (27) and (28), which is chosen as a device to bound the change of measure densities between ϖ and 1. □

Such a choice of change of measure densities also preserves an increasing $\mathcal{A}[v_T]$ and a decreasing $\mathcal{B}[v_T]$ over the entire domain $v_T \in (0,\infty)$:

$$\begin{aligned} \mathcal{A}'[v_T] &= \frac{(1-\varpi_a)\exp\left(-\frac{\{\ln(v_T)-\mu_a\}^2}{2\sigma_a^2}\right)}{v_T\sigma_a\sqrt{2\pi}} > 0, \\ \mathcal{B}'[v_T] &= -\frac{(1-\varpi_b)\exp\left(-\frac{\{-\ln(v_T)-\mu_b\}^2}{2\sigma_b^2}\right)}{v_T\sigma_b\sqrt{2\pi}} < 0, \end{aligned} \tag{29}$$

bringing certain deviation from the setup of Example 1.

One common feature across different parameter combinations is that the pricing kernel $m[v_T]$ associated with this example is flat in the tails, reflecting the boundedness of the change of measures. Moreover, when μ_a and μ_b are both positive, and not too small, one generally obtains $m[v_T]$, which displays a

indicating that a sizable fraction of VIX moves exceed 20% in both the left and right tails. The analysis of Amengual and Xiu (2013) indicates that a source of downward movements in volatility is resolution of policy uncertainty.

https://doi.org/10.1017/S0022109015000642 Published online by Cambridge University Press

decreasing region at some low volatility levels and then an increasing region for high volatility levels.[3]

Figure 2 plots the pricing kernel and the optimal $f[v_T]$ and $g[v_T]$ for the following parameters:

$$\text{(30)} \qquad \begin{aligned} &\mu_p = 0.000, \quad \mu_a = 0.150, \quad \mu_b = 0.200, \quad \sigma_p = 0.171, \\ &\sigma_a = 0.100, \quad \sigma_b = 0.100, \end{aligned}$$

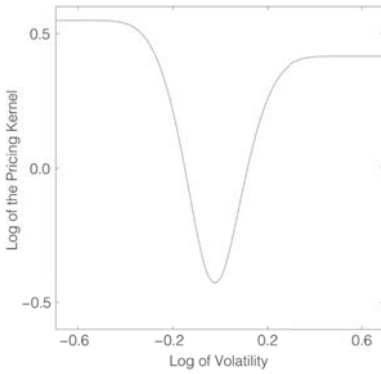
with $\gamma_f = 3.5$, $\gamma_g = 3.0$, $\phi = 0.50$, $\varpi_a = 0.060$, and $\varpi_b = 0.070$.

FIGURE 2

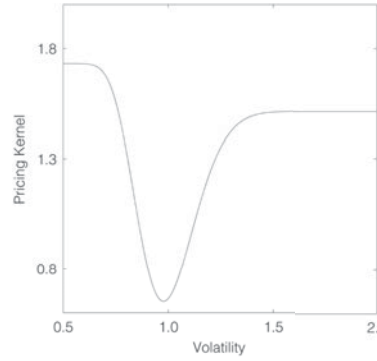
Pricing Kernel and Optimal Volatility-Contingent Cash Flows from Example 2

Figure 2 plots i) the pricing kernel $m[v_T]$ and ii) the optimal volatility-contingent cash flows, $f[v_T]$ and $g[v_T]$, for the two types of investors. The physical density (as shown in equation (14)) is assumed to be lognormal, the personalized change of measures are as displayed in equations (27) and (28), and the personalized densities are $\Phi_f[v_T] = \mathcal{A}[v_T]\Phi[v_T]/\int \mathcal{A}[v_T]\Phi[v_T]dv_T$ and $\Phi_g[v_T] = \mathcal{B}[v_T]\Phi[v_T]/\int \mathcal{B}[v_T]\Phi[v_T]dv_T$. We use $\mu_p = 0.000$, $\mu_a = 0.150$, $\mu_b = 0.200$, $\sigma_p = 0.171$, $\sigma_a = 0.100$, and $\sigma_b = 0.100$. The coefficients of absolute risk aversion are set to $\gamma_f = 3.5$, $\gamma_g = 3.0$, $\phi = 0.50$, $\varpi_a = 0.060$, and $\varpi_b = 0.070$. Our parameterization implies means of 15% and −15% for the personalized densities of investors F and G, respectively.

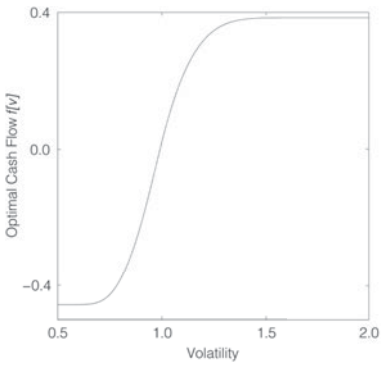
Graph A. Log of the Pricing Kernel



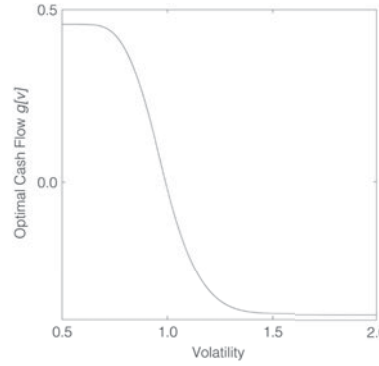
Graph B. Pricing Kernel



Graph C. Optimal Cash Flow f[v]



Graph D. Optimal Cash Flow g[v]



[3]We note that $f[v_T]$ is increasing and $g[v_T]$ is decreasing, and explicit expressions for $f[v_T]$ and $g[v_T]$, and their derivatives, can be obtained analytically in terms of the cumulative Normal distribution function.

https://doi.org/10.1017/S0022109015000642 Published online by Cambridge University Press

This part of the analysis establishes two model traits. First, the derived $m[v_T]$ in Figure 2 exhibits bending in the tails, mirroring the boundedness of the change of measure densities. In between, $m[v_T]$ is first declining and then increasing. For our parameter set, $f[v_T]$ ($g[v_T]$) is convex for low (high) levels of volatility. These regions correspond to states where the respective investors are selling, rather than buying, volatility-contingent cash flows. Examples 1 and 2 thus differ in some economic dimensions but still rest on heterogeneity in investors' beliefs about the distribution of volatility outcomes and their risk attitudes. The shape of the pricing kernel implies that some investors dislike both low- and high-volatility states.

2. Expected Option Returns

To provide a link to the empirical evidence presented in Section III, we next discuss the implications of our parametric examples for the pricing density $q[v_T]$, and for the pattern of expected returns of options on volatility.

Based on equations (7) and (8),

$$q[v_T] = \frac{(\mathcal{A}[v_T])^{\eta}\,(\mathcal{B}[v_T])^{1-\eta}\,\Phi[v_T]}{\int (\mathcal{A}[v_T])^{\eta}\,(\mathcal{B}[v_T])^{1-\eta}\,\Phi[v_T]\,dv_T}. \tag{31}$$

Although in Example 1, equation (31) reduces to $q[v_T] = (\Phi_f[v_T])^{\eta}\,(\Phi_g[v_T])^{1-\eta} / \int (\Phi_f[v_T])^{\eta}\,(\Phi_g[v_T])^{1-\eta}\,dv_T$, which is multiplicative in the belief densities, Example 2 does not offer a simple analytical expression for the pricing density.

In columns 1–6 of Table 1, we report the expected returns of options on volatility obtained from Examples 1 and 2 under parameterizations of investors' volatility belief distributions and risk aversions. In this illustrative exercise, columns 1 and 5 correspond to Figures 1 and 2, respectively. Both examples exhibit asymmetries in the returns of volatility options, and most notably, expected put returns increase in strike, whereby deep OTM put returns can be the most negative and expected call returns decrease in strike.[4] Our parameterizations thus illustrate a pricing measure that dominates the physical measure in both tails of the volatility distribution.

Recognize that our model incorporates the opposing views about volatility of only two types of investors, which leads to expected returns in Table 1 that are somewhat structured. In contrast, observed option returns likely reflect other opinions; for example, some investors may expect no change in volatility. We anticipate that, for instance, a model adding investors with centrist beliefs about volatility outcomes could deliver a different pattern of near-the-money option returns. Such an extension could impart a pricing kernel that places greater weight on outcomes in the neck of the volatility distribution, as opposed to our model that reinforces movements in the tails.

[4]We have ascertained that it is possible to preserve the core features of the pricing kernel, as in Examples 1 and 2, under Gamma-distributed physical and personalized densities. This analysis is omitted to maintain parsimony in the number of examples.

https://doi.org/10.1017/S0022109015000642 Published online by Cambridge University Press

TABLE 1

Implications of the Model with Heterogeneity in Beliefs for Expected VIX Option Returns

The pricing density for Example 1 and Example 2 can be computed as,

$$q[v_T] = \begin{cases} \dfrac{(\Phi_f[v_T])^{\eta}\ (\Phi_g[v_T])^{1-\eta}}{\int (\Phi_f[v_T])^{\eta}\ (\Phi_g[v_T])^{1-\eta}\, dv_T} & \text{(Example 1)}, \\ \dfrac{(\mathcal{A}[v_T])^{\eta}\,(\mathcal{B}[v_T])^{1-\eta}\,\Phi[v_T]}{\int (\mathcal{A}[v_T])^{\eta}\,(\mathcal{B}[v_T])^{1-\eta}\,\Phi[v_T]\, dv_T} & \text{(Example 2)}, \end{cases}$$

where $\Phi[v_T] = \Phi[v_T; \mu_p, \sigma_p]$ is the lognormal physical density of volatility, as displayed in equation (14), and $\Phi_f[v_T] = \Phi[v_T; \mu_f, \sigma_f]$ and $\Phi_g[v_T] = \Phi[v_T; \mu_g, \sigma_g]$ are the personalized densities of the two types of investors. The change of measure densities $\mathcal{A}[v_T]$ and $\mathcal{B}[v_T]$ are presented in equations (27) and (28). The expected returns of the respective options are calculated using the ratios of their expected payoffs under the physical and pricing densities, that is, for puts, $(\int (K - v_T)^+ \Phi[v_T] dv_T / \int (K - v_T)^+ q[v_T] dv_T) - 1$. OTM stands for out of the money. The parameters employed in columns 1 and 5 of Table 1 correspond to those used in Figure 1 and Figure 2.

	Example 1					Example 2	
	1	2	3	4		5	6
Panel A. Model Parameters							
μ_p	0.000	0.000	0.000	0.000	μ_p	0.000	0.000
σ_p	0.171	0.171	0.171	0.171	σ_p	0.171	0.171
γ_f	3.500	3.500	3.000	3.000	γ_f	3.500	3.500
γ_g	3.000	3.000	3.500	3.500	γ_g	3.000	3.000
μ_f	0.150	0.100	0.150	0.100	μ_a	0.150	0.150
μ_g	−0.150	−0.100	−0.150	−0.100	μ_b	0.200	0.200
σ_f	0.184	0.179	0.184	0.179	σ_a	0.100	0.100
σ_g	0.184	0.179	0.184	0.179	σ_b	0.100	0.100
					ϖ_a	0.060	0.160
					ϖ_b	0.070	0.150
Panel B. Means of Investors' Volatility Belief Distributions							
Investor type F	15%	10%	15%	10%		15%	10%
Investor type G	−15%	−10%	−15%	−10%		−15%	−10%
Panel C. Expected VIX Option Returns							
25% OTM put	−46%	−35%	−28%	−20%	25% OTM put	−41%	−38%
15% OTM put	−31%	−23%	−13%	−9%	15% OTM put	−35%	−32%
5% OTM put	−19%	−14%	−4%	−3%	5% OTM put	−23%	−22%
5% OTM call	−5%	−4%	−21%	−15%	5% OTM call	−24%	−18%
15% OTM call	−13%	−9%	−30%	−22%	15% OTM call	−30%	−24%
25% OTM call	−21%	−15%	−39%	−29%	25% OTM call	−33%	−27%

III. Empirical Evaluation

This section formulates the testable hypothesis regarding our model and compares its implications to a framework where volatility is disliked. Next, we outline the essential features of VIX options, given that we measure volatility by the VIX index. Then we discuss the evidence from our empirical tests.

A. Testable Implications

The salient feature of our model with heterogeneity in beliefs is the nonmonotonic kernel, which exhibits a decreasing region at low volatility levels and an increasing region at high volatility levels. The specific prediction is that expected returns of calls on volatility decrease in strike and expected returns of puts on volatility increase in strike (Theorem 1, and as demonstrated in the context of our example economies in Table 1).

How does this prediction compare to a conventional modeling approach, where volatility is disliked? Specifically, in Theorem 2 of Appendix C, we show

https://doi.org/10.1017/S0022109015000642 Published online by Cambridge University Press

that a model of this type leads to a pricing kernel that is monotonically increasing in the volatility dimension and, hence, predicts that the expected returns of puts (calls) on volatility are decreasing in strike and are positive (negative) for sufficiently low (high) strikes.

We note that both models imply a pricing kernel with an increasing region for high levels of volatility and, hence, yield identical predictions with respect to the right tail of the volatility distribution or the pattern of expected returns of OTM volatility calls.

However, the models differ dramatically in their predictions about the left tail of the volatility distribution. Therefore, it is the pattern of expected returns of OTM volatility puts that can help discriminate between the two models, and hence, Theorem 2 offers a useful counterpoint to Theorem 1 that brings expected returns of volatility puts into the focus of our examination.

B. VIX Option Data

Proceeding to the empirical analysis, we employ VIX options with 28 days to expiration and calculate the return of a VIX call and put as,[5]

$$1 + r_{t,T}^{c,\text{vix}}[y] = \frac{(\text{VIX}_{t+T} - y\,\text{FVIX}_{t,T})^+}{C_{t,T}^{\text{vix}}[y]} \quad \text{and} \tag{32}$$

$$1 + r_{t,T}^{p,\text{vix}}[y] = \frac{(y\,\text{FVIX}_{t,T} - \text{VIX}_{t+T})^+}{P_{t,T}^{\text{vix}}[y]},$$

where $\text{FVIX}_{t,T}$ is the front-month VIX futures price, and $C_{t,T}^{\text{vix}}[y]$ ($P_{t,T}^{\text{vix}}[y]$) is the price of the VIX call (put) with moneyness $y \equiv K/\text{FVIX}_{t,T}$. Using ask quotes on VIX options from the CBOE, we obtain one return observation per month, corresponding to a fixed moneyness, and avoid overlapping returns.

For the calculation in equation (32), we identify VIX calls and puts, which are closest to 5%, 15%, and 25% OTM, respectively. The specific moneyness levels are chosen, taking into consideration the minimum strike intervals for VIX options at the CBOE. We compute returns to expiration, using the settlement value of VIX to determine the option payoffs. Our data set thus consists of six series of OTM VIX option returns, both calls and puts, with 28 days to expiration.

Important to our themes, trading in volatility has seen rapid growth, with the VIX options pit the second most active, after the put for S&P 500 index options. Because exchange trading in VIX options started in Mar. 2006, our sample spans the 93-month period from Mar. 2006 to Dec. 2013.

[5]While VIX option attributes are standard, we nonetheless note that VIX at time t, denoted VIX_t, is the square root of the expected quadratic variation of the logarithm of the market index over the next 30 days, under the pricing measure $\mathbb{Q}$. In practice, VIX_t is approximated from the price of a traded index option portfolio (e.g., Carr and Wu (2009)). The price at time t of the VIX futures with time to maturity T, denoted $\text{FVIX}_{t,T}$, is $\mathbb{E}_t^{\mathbb{Q}}(\text{VIX}_{t+T})$, where $\mathbb{E}_t^{\mathbb{Q}}$ is the time t conditional expectation under the pricing measure $\mathbb{Q}$. Similarly, the time t prices of a call and a put on the VIX with expiration date $t+T$ and strike K is $C_{t,T}^{\text{vix}}[K] = e^{-r_f}\,\mathbb{E}_t^{\mathbb{Q}}\left((\text{VIX}_{t+T} - K)^+\right)$ and $P_{t,T}^{\text{vix}}[K] = e^{-r_f}\,\mathbb{E}_t^{\mathbb{Q}}\left((K - \text{VIX}_{t+T})^+\right)$, where r_f is the T-period risk-free return.

https://doi.org/10.1017/S0022109015000642 Published online by Cambridge University Press

https://doi.org/10.1017/S0022109015000642 Published online by Cambridge University Press

Pertinent to our empirical approach, two other points deserve further comment. First, to uncover a pricing kernel with increasing and decreasing regions, one needs to rely on instruments that can separate up and down volatility movements. Therefore, variance swaps, which do not offer Arrow-Debreu-like payoffs, may be uninformative and may not be suitable for this purpose. Lacking other suitable instruments that offer tail exposures, we employ VIX option data in our model assessments, despite the nascency of the volatility option market. Second, if conditional expected returns are lower for volatility puts deeper OTM, then average returns of volatility puts, which reflect unconditional expectations, should exhibit the same property (e.g., Coval and Shumway (2001), Bakshi et al. (2010)).

C. Tests Based on the Moneyness Pattern of VIX Put Returns

Providing an essential link to the theoretical models, our exercise now focuses on the VIX option returns, whose pattern across option types and moneyness is revealing about the pricing density and underlying pricing kernel that are supported in the data. Panels A–C of Table 2 yield potential insights into tail behavior in the volatility market.

Do returns of VIX options get more negative deeper OTM? While addressing the small-sample concern, in Panel A of Table 2 we test for differences in average VIX option returns across strikes by drawing 50,000 pairwise bootstrap samples of returns of VIX puts (calls) that are 25% and 15%, 25% and 5%, and 15% and 5% OTM, respectively. Then we calculate the difference between the average returns for each pair of bootstrap samples. Reported are the *p*-values for the differences between the average returns of the respective VIX options, calculated as the proportion of bootstrap samples where the first option in a pair has higher average return than the second option.

Most importantly, we find that the average return of 25% OTM VIX puts is lower than the return of 15% (5%) OTM VIX puts at a 0.8% (1.7%) significance level. At the same time, these *p*-values indicate that the differences across moneyness are statistically insignificant for VIX calls. Therefore, the pairwise differences in average VIX put returns across the three moneyness levels provide statistically significant evidence, contradicting Theorem 2 but possibly consistent with our model of the volatility market. As highlighted before, it is the returns of volatility puts that may help discriminate between the competing hypotheses.

Complementing the above, we follow Patton and Timmermann (2010) and apply the MR, UP, and DOWN bootstrap-based procedures they advocate to test for a monotonic relation between expected returns of options on volatility and their strikes. The null hypothesis for the MR test is that average VIX option returns are flat with respect to moneyness, versus the alternative of an increasing pattern. Furthermore, the UP and DOWN tests, which combine information on both the number and magnitude of departures from a flat pattern, are designed to detect positive or negative segments in the relation between average option returns and moneyness.

Panel B of Table 2 reports for VIX puts *p*-values of 0.049, 0.005, and 0.723, for the MR, UP, and DOWN tests, respectively. The MR (UP) test thus supports a statistically significant increasing pattern (positive segment) in the relation

TABLE 2

Assessing Differences in Returns of VIX Puts and Calls across Strikes

Returns to expiration of 28-day VIX options are constructed as,

$$1 + r^{c,\text{vix}}_{t,T}[y] = \frac{(\text{VIX}_{t+T} - y\,\text{FVIX}_{t,T})^+}{C^{\text{vix}}_{t,T}[y]} \quad \text{and} \quad 1 + r^{p,\text{vix}}_{t,T}[y] = \frac{(y\,\text{FVIX}_{t,T} - \text{VIX}_{t+T})^+}{P^{\text{vix}}_{t,T}[y]}$$

for each month from Mar. 2006 to Dec. 2013 (93 months), using *ask quotes*. The moneyness $y \equiv K/\text{FVIX}_{t,T}$ of a VIX option at time t, which expires at time $t + T$, for $T = 28$ days, is determined relative to the time t price $\text{FVIX}_{t,T}$ of the VIX futures contract with settlement date $t + 28$ days. In Panel A of Table 2, we test for differences in average VIX option returns across strikes and option types. We consider differences between puts (calls) that are 25% and 15%, 25% and 5%, and 15% and 5% OTM, respectively. For each such pair of moneyness levels, we draw 50,000 pairwise bootstrap samples of the respective option returns, and for each pair of bootstrap samples we calculate the difference between their average returns. A negative difference indicates that in the respective bootstrap sample, the first option in a pair has higher average return than the second option. The reported *p*-values in Panel A are the proportion of negative differences in the 50,000 bootstrap samples for each pair of option moneyness levels. Panel B presents the results from the bootstrap-based MR, UP, and DOWN monotonicity tests of Patton and Timmermann (2010). The null hypothesis for the MR test is that the pattern of VIX option returns with respect to moneyness is flat, versus the alternative that the pattern is monotonic. The UP and DOWN tests detect positive or negative segments in the pattern. The *p*-values lower than 5% imply rejection of the null. Panel C shows i) the average option returns (28 days, in %), ii) bootstrapped 90% confidence intervals (in square brackets) for average returns, iii) open interest (average daily number of contracts), and iv) traded volume (in curly brackets, average daily number of contracts).

Panel A. Bootstrap p-Values for Differences in Average VIX Option Returns

	Puts	Calls
25% OTM versus 15% OTM	0.008	0.359
25% OTM versus 5% OTM	0.017	0.272
15% OTM versus 5% OTM	0.246	0.242

Panel B. Bootstrap-Based Tests of Monotonic Relation (p-Values)

	MR	UP	DOWN
VIX puts	0.049	0.005	0.723
VIX calls	0.639	0.668	0.272

Panel C. Average VIX Option Returns (28 Days, in %), Open Interest, and Traded Volume

	Average Returns (%) [90% CI]		Open Interest {Volume}	
	Puts	Calls	Puts	Calls
25% OTM	−32.68 [−68, 8]	−34.84 [−83, 26]	34,621 {2,493}	70,684 {8,660}
15% OTM	4.53 [−24, 35]	−31.30 [−72, 20]	39,943 {6,560}	55,047 {6,716}
5% OTM	10.21 [−11, 32]	−27.36 [−63, 16]	33,657 {5,672}	49,199 {6,024}

between VIX put returns and moneyness, and provides statistical confirmation that average returns of VIX puts get more negative deeper OTM.

To clarify our result, Panel C of Table 2 shows the average returns per month, which are 10.21%, 4.53%, and −32.68% for VIX puts that are 5%, 15%, and 25% OTM, respectively. Therefore, VIX put returns increase in strike and are conspicuously low for the farthest OTM puts. In contrast, the average returns are −27.36%, −31.30%, and −34.84% for 5%, 15%, and 25% OTM VIX calls, respectively, and are thus consistent with both models. We note here that the average VIX option returns are not statistically different from 0, as indicated by the reported 90% confidence intervals obtained in 50,000 bootstrap draws, whereas the pairwise test is able to furnish statistical significance in support of the model that hinges on heterogeneity in beliefs.

In conclusion, the average VIX put returns in our sample appear inconsistent with a theory postulating a pricing kernel that is upward sloping in the volatility

https://doi.org/10.1017/S0022109015000642 Published online by Cambridge University Press

dimension, as per Theorem 2, and Table 2 indicates that the risk-neutral probability in both tails of the volatility distribution can potentially exceed the corresponding physical probability. In particular, the model of the volatility market that we suggest in Section II appears better suited to reflect the empirical features of observed VIX option returns, and it is more consistent with the left tail of the volatility distribution, thus pointing to a direction for further volatility modeling.

IV. Concluding Remarks

This article features a model that incorporates heterogeneity in beliefs about volatility and allows for a kernel that is nonmonotonic in the volatility dimension.

We provide examples demonstrating that within the model with heterogeneity in beliefs, expected returns of puts on volatility can be increasing in strike and negative for deep OTM strikes. Contrasting our model's implications for volatility tail behavior, a model where high-volatility states are disliked, predicts that expected returns of puts on volatility are decreasing in strike and are positive for sufficiently low strikes. Using the two models' opposite predictions about the behavior of volatility in the left tail, we conduct tests indicating that the model with heterogeneity in beliefs appears to better reflect features of the data, both qualitatively and quantitatively.

Although we have a discrete-time model, there is some progress in modeling heterogeneity in beliefs in continuous time, as in Dieckmann (2011) and Chen, Joslin, and Tran (2012). The recent works of Song and Xiu (2013) and Amengual and Xiu (2013) highlight the possible role of jumps in generating realistic pricing kernels of the volatility market.

It bears emphasizing that our model presumes a volatility market that is dominated by investors with 0 equity delta. In a more elaborate model setting, the investors would trade the equity index, as well as volatility-contingent claims, exploiting their beliefs about both equity returns and volatility outcomes. The relevant optimization problem of the investors would then involve i) a two-dimensional physical density, ii) change of measure densities that are functions of equity returns and volatility, and iii) long and short exposures in equity and volatility cash flows. However, this extension may require a richer market that trades product options of all strike pairs in equity and volatility, and is left to a follow-up study.

Taken together, our theoretical models and empirical results can pave the way for a better appreciation of the motives for trading volatility and hopefully provide impetus for improved models for pricing and hedging volatility. When VIX options are available to offer exposure to volatility tails, much can be learned in the realm of traded volatility.

Appendix A. Proof of Theorem 1

Proof of Part I of Theorem 1. The choice problem of the two investors in Problem 1 implies the optimality conditions,

$$\text{(A-1)} \qquad e^{-\gamma_f f[v_T]}\,\mathcal{A}[v_T] \;=\; \lambda^f\,\frac{q[v_T]}{\Phi[v_T]} \quad \text{and} \quad e^{-\gamma_g g[v_T]}\,\mathcal{B}[v_T] \;=\; \lambda^g\,\frac{q[v_T]}{\Phi[v_T]},$$

https://doi.org/10.1017/S0022109015000642 Published online by Cambridge University Press

where λ^f and λ^g are the Lagrange multipliers associated with the budget constraints of the investors.

Now define the pricing kernel as $m[v_T] \equiv q[v_T]/\Phi[v_T]$. From equation (A-1), take a logarithmic derivative to obtain the equations:

$$\text{(A-2)} \qquad \begin{aligned} f'[v_T] &= \frac{1}{\gamma_f}\left(\frac{\mathcal{A}'[v_T]}{\mathcal{A}[v_T]} - \frac{m'[v_T]}{m[v_T]}\right) \quad \text{and} \\ g'[v_T] &= \frac{1}{\gamma_g}\left(\frac{\mathcal{B}'[v_T]}{\mathcal{B}[v_T]} - \frac{m'[v_T]}{m[v_T]}\right). \end{aligned}$$

Because the claims are in 0 net supply, we have $\phi f'[v_T] + (1-\phi)\, g'[v_T] = 0$. This implies,

$$\text{(A-3)} \qquad \begin{aligned} \phi f'[v_T] + (1-\phi)\, g'[v_T] &= 0 = \frac{\phi}{\gamma_f}\left(\frac{\mathcal{A}'[v_T]}{\mathcal{A}[v_T]} - \frac{m'[v_T]}{m[v_T]}\right) \\ &\quad + \frac{1-\phi}{\gamma_g}\left(\frac{\mathcal{B}'[v_T]}{\mathcal{B}[v_T]} - \frac{m'[v_T]}{m[v_T]}\right). \end{aligned}$$

Rearranging equation (A-3), we get the differential equation,

$$\text{(A-4)} \qquad \frac{m'[v_T]}{m[v_T]} = \left(\frac{\phi\,\gamma_g}{(1-\phi)\gamma_f + \phi\gamma_g}\right)\frac{\mathcal{A}'[v_T]}{\mathcal{A}[v_T]} + \left(\frac{(1-\phi)\,\gamma_f}{(1-\phi)\gamma_f + \phi\gamma_g}\right)\frac{\mathcal{B}'[v_T]}{\mathcal{B}[v_T]}.$$

The solution to equation (A-4), when expressed in terms of the change of measure densities, and letting $\eta \equiv (\phi\,\gamma_g)/((1-\phi)\gamma_f + \phi\gamma_g)$, is,

$$\text{(A-5)} \qquad \ln(m[v_T]) = \ln(m_0) + \eta\, \ln(\mathcal{A}[v_T]) + (1-\eta)\, \ln(\mathcal{B}[v_T]),$$

where the constant m_0 satisfies $\int m[v_T]\Phi[v_T]dv_T = 1$. Hence, $m[v_T] = m_0\,(\mathcal{A}[v_T])^{\eta} \times (\mathcal{B}[v_T])^{1-\eta}$.

Based on equations (A-3) and (A-4), we arrive at the optimal volatility-contingent cash flows,

$$\text{(A-6)} \qquad f'[v_T] = \left(\frac{(1-\phi)}{(1-\phi)\gamma_f + \phi\gamma_g}\right)\left(\frac{\mathcal{A}'[v_T]}{\mathcal{A}[v_T]} - \frac{\mathcal{B}'[v_T]}{\mathcal{B}[v_T]}\right),$$

$$\text{(A-7)} \qquad g'[v_T] = -\left(\frac{\phi}{(1-\phi)\gamma_f + \phi\gamma_g}\right)\left(\frac{\mathcal{A}'[v_T]}{\mathcal{A}[v_T]} - \frac{\mathcal{B}'[v_T]}{\mathcal{B}[v_T]}\right).$$

Solving the differential equation (A-6) and imposing $\int f[v_T]q[v_T]\,dv_T = 0$, we obtain:

$$\text{(A-8)} \qquad \begin{aligned} f[v_T] &= -\left(\frac{(1-\phi)}{(1-\phi)\gamma_f + \phi\gamma_g}\right)\int \ln\left(\frac{\mathcal{A}[v_T]}{\mathcal{B}[v_T]}\right) q[v_T]\,dv_T \\ &\quad + \left(\frac{(1-\phi)}{(1-\phi)\gamma_f + \phi\gamma_g}\right)\ln\left(\frac{\mathcal{A}[v_T]}{\mathcal{B}[v_T]}\right). \end{aligned}$$

For $f[v_T]$ to be concave, we need $f''[v_T] < 0$. If $\mathcal{A}[v_T]/\mathcal{B}[v_T]$ is posited to be concave, then $\ln(\mathcal{A}[v_T]/\mathcal{B}[v_T])$ preserves concavity. □

Proof of Part II of Theorem 1. A function $\mathcal{L}[v]$ is log-convex on an interval if $\mathcal{L}[v]$ is positive and $\ln(\mathcal{L}[v])$ is convex (Bagnoli and Bergstrom (2005)). It follows from equation (A-5) and condition (c) that $\ln(m[v_T])$ is convex, being a convex combination of two convex functions. Therefore, $m''[v_T]/m[v_T] - (m'[v_T]/m[v_T])^2 > 0$, and hence, $m''[v_T] > 0$.

Together with conditions (a) and (b), this implies that $m[v_T]$ is log-convex, decreasing at $\underline{v}$ and increasing at $\overline{v}$, and achieves a minimum in the interior of $[\underline{v}, \overline{v}]$. □

https://doi.org/10.1017/S0022109015000642 Published online by Cambridge University Press

Appendix B. A Model of Volatility Exposure in Continuous Time

The setup of the continuous-time economy is similar in its information structure and investor beliefs to those adopted in Detemple and Murthy (1994), Basak (2000), (2005), Buraschi and Jiltsov (2006), and Kogan, Ross, Wang, and Westerfield (2006). Hence, we intend to be brief.

An important property of our setup is that instantaneous market volatility, which evolves stochastically over time, is not a traded asset (see, e.g., CBOE (2009)). Therefore, in our analysis we assume that trading opportunities are confined to a position in a futures contract that settles into the instantaneous volatility at maturity. We further assume that the volatility market is dynamically complete, and our aim is to show that the pricing kernel in this dynamic setting is monotonic in volatility.

Other features of the continuous-time model are as follows. First, the current time is 0, the futures contract matures at time T, and investors consume only at time T. Second, the economy is inhabited by two investor types, F and G, who are endowed with some initial cash, $Z_0^f > 0$ and $Z_0^g > 0$, respectively. Third, the two investor types are present with mass $0 < \phi < 1$ and $1 - \phi$, respectively.

Uncertainty about the evolution of volatility is described by a 1-dimensional standard Brownian motion $\omega_t^{\mathbb{P}}$, and it is assumed that volatility v_t follows a geometric Brownian motion, under the physical probability measure $\mathbb{P}$,

$$dv_t \quad = \quad \mu_p\, v_t\, dt \;+\; \sigma_p\, v_t\, d\omega_t^{\mathbb{P}}, \tag{B-1}$$

where μ_p and σ_p are constants.

In choosing their trading policy in the futures contract on volatility, each investor uses a subjective probability measure $\mathbb{P}^i$, for $i \in \{f, g\}$. We follow, among others, Detemple and Murthy (1994) and Kogan et al. (2006), and define,

$$\omega_t^{\mathbb{P}_i} \quad = \quad \omega_t^{\mathbb{P}} \;+\; \vartheta_i\, t, \vartheta_i \quad \equiv \quad \frac{\mu_p - \mu_i}{\sigma_p}, \quad i \in \{f, g\}, \quad 0 \le t \le T, \tag{B-2}$$

which implies that investor i believes that volatility v_t evolves as,

$$v_t \quad = \quad v_0 \exp\left((\mu_i - \frac{1}{2}\,\sigma_p^2)t \;+\; \sigma_p\, \omega_t^{\mathbb{P}_i} \right), \qquad i \in \{f, g\}. \tag{B-3}$$

It follows from the Radon–Nikodym theorem (e.g., Karatzas and Shreve (1991)) that the change of measure density for investor i at time T, ξ_T^i, is,

$$\xi_T^i \quad = \quad \exp\left(-\frac{1}{2}\vartheta_i^2\, T - \vartheta_i\, \omega_T^{\mathbb{P}} \right). \tag{B-4}$$

To fix the remaining notation, let the futures price be y_t and r be the constant instantaneous interest rate. As investor i's position in the futures contract dynamically changes, her wealth evolves as (see also Briys et al. (1990), Hong (2000)),

$$dW_t^i \quad = \quad r\, W_t^i\, dt \;+\; N_t^i\, dy_t, \tag{B-5}$$

where N_t^i is the number of shares held (long or short) in the futures contract at time t, and hence, equation (B-5) reflects marking to market. The futures contract is in 0 net supply, and so $\phi N_t^f + (1 - \phi) N_t^g = 0$ for each t.

At maturity $y_T = v_T$. The futures price at time t is determined as $y_t = \mathbb{E}_t^{\mathbb{P}}\left((m_T/m_t)v_T\right)$, where m_T denotes the pricing kernel process at time T, and we normalize $\mathbb{E}_0^{\mathbb{P}}(m_T) = 1$.

Following Cox and Huang (1989), one can convert the investor's maximization problem into a static problem, as stated next.

https://doi.org/10.1017/S0022109015000642 Published online by Cambridge University Press

Problem 2. Choose wealth W_T^i to maximize $(-1/\gamma_i)\,\mathbb{E}_0^{\mathbb{P}}\left(\xi_T^i \exp\left(-\gamma_i W_T^i\right)\right)$, subject to $\mathbb{E}_0^{\mathbb{P}}\left(m_T W_T^i\right) = W_0^i$. The market-clearing condition is $\phi W_T^f + (1-\phi) W_T^g = \phi Z_0^f + (1-\phi) Z_0^g = \overline{Z}$.

We now state a theoretical result from our continuous-time economy.

Proposition 1. In the continuous-time economy, the pricing kernel m_T is related to the change of measure densities ξ_T^f and ξ_T^g as,

$$\text{(B-6)} \qquad m_T \;=\; m_* \left(\xi_T^f\right)^{\eta} \left(\xi_T^g\right)^{1-\eta}, \quad \eta \;\equiv\; \frac{\phi\,\gamma_g}{(1-\phi)\gamma_f + \phi\gamma_g},$$

where ξ_T^f and ξ_T^g are as shown in equation (B-4), and $m_* = 1/\mathbb{E}_0^{\mathbb{P}}\left(\left(\xi_T^f\right)^{\eta}\left(\xi_T^g\right)^{1-\eta}\right)$. Furthermore, the pricing kernel m_T is monotonic in volatility v_T.

Proof. The first-order optimality conditions for the two investors are,

$$\text{(B-7)} \qquad \xi_T^f \exp\left(-\gamma_f W_T^f\right) \;=\; \lambda^f m_T, \quad \text{and} \quad \xi_T^g \exp\left(-\gamma_g W_T^f\right) \;=\; \lambda^g m_T.$$

Imposing the market-clearing condition, and normalizing $\mathbb{E}_0^{\mathbb{P}}(m_T) = 1$ to eliminate the two Lagrangian multipliers λ^f and λ^g, yields equation (B-6).

Note that $\omega_T^{\mathbb{P}} = (1/\sigma_p)\log(v_T/v_0) + (1/\sigma_p)\left(\sigma_p^2/2 - \mu_p\right) T$, and the pricing kernel is multiplicative in ξ_T^f and ξ_T^g. As a result, ξ_T^f and ξ_T^g are exponential in $\log(v_T/v_0)$.

Using these features, along with the fact that μ_p, μ_f, μ_g, and σ_p enter as constants in the derived solution (B-6) implies that the pricing kernel is monotonic in $\log(v_T/v_0)$. Such a pricing kernel can only admit an upward- or downward-sloping region in v_T/v_0. □

Appendix C. Implications of a Model Where Volatility Is Disliked

The model presented here provides one set of predictions about the tails of the volatility distribution, which could be explored using VIX option data. It also allows us to place in perspective the model of the volatility market that we suggest in Section II.

Let $S_{t+T} \in (0, \infty)$ be the level of the market index at time $t + T$ and $v_{t+T} \in (0, \infty)$ be the market return volatility.

Claims, contingent on S_{t+T} and/or v_{t+T}, can be generically represented by their payoff $h[S_{t+T},\ v_{t+T}]$. When considering generic contingent claims dependent on the 2-dimensional vector $(S_{t+T},\ v_{t+T})$, for which there could be data from financial markets, it suffices to work with the projection of the pricing kernel onto the space, generated by the market index and return volatility, even if the pricing kernel may admit higher dimensional state dependencies.[6] Hence, we consider here a pricing kernel that is dependent on S_{t+T} and v_{t+T}, and is denoted as $m[S_{t+T}, v_{t+T}]$.

If we normalize $S_t = 1$, then $z_{t+T} \equiv \ln(S_{t+T}) \in (-\infty, \infty)$ represents the logarithmic return of the market. One can write the pricing kernel as $m[z_{t+T}, v_{t+T}]$ or, for brevity, as $m[z, v]$. Let $\Phi[z, v]$ represent the joint density function corresponding to the physical probability measure $\mathbb{P}$. For the discussion to come, we denote the conditional density of z given v as $\widehat{\Phi}_1[z|v] = \Phi[z, v] / \int_{-\infty}^{\infty} \Phi[z, v] dz$, and the marginal density of v as $\Phi_2[v] = \int_{-\infty}^{\infty} \Phi[z, v]\, dz$.

[6] The time t price S_t^h of the claim with payoff $h[S_{t+T}, v_{t+T}]$ at $t + T$ is $S_t^h = (1 + r_f)^{-1} \times \mathbb{E}^{\mathbb{P}}\left(h[S_{t+T}, v_{t+T}] \times \mathfrak{m}\right) = (1 + r_f)^{-1}\mathbb{E}^{\mathbb{P}}(h[S_{t+T}, v_{t+T}] \times \mathbb{E}^{\mathbb{P}}(\mathfrak{m}|S_{t+T}, v_{t+T})) = \mathbb{E}^{\mathbb{P}}(h[S_{t+T}, v_{t+T}] \times \widetilde{\mathfrak{m}}[S_{t+T}, v_{t+T}]/1 + r_f)$, where $\mathbb{E}^{\mathbb{P}}(.)$ is expectation under the physical probability measure, and $r_f \geq 0$ is the T-period risk-free return. Moreover, $\mathfrak{m}$ represents the probability change of measure with the bond price as numeraire, and $\widetilde{\mathfrak{m}}[S_{t+T}, v_{t+T}] \equiv \mathbb{E}^{\mathbb{P}}\left(\mathfrak{m}\middle|S_{t+T}, v_{t+T}\right)$. Accordingly, all that is required to be modeled is $\mathbb{E}^{\mathbb{P}}\left(\mathfrak{m}\middle|S_{t+T}, v_{t+T}\right)$.

https://doi.org/10.1017/S0022109015000642 Published online by Cambridge University Press

Define the expected return of the put on volatility with strike price K as,

$$1 + \mu_v^{\text{put}}[K] = \frac{\int_0^\infty \int_{-\infty}^\infty (K - v)^+ \Phi[z, v]\, dz\, dv}{\int_0^\infty \int_{-\infty}^\infty (K - v)^+ m[z, v]\, \Phi[z, v]\, dz\, dv}, \tag{C-1}$$

and analogously for the call on volatility, where $\ell^+ \equiv \max(\ell, 0)$.

Three assumptions underpin our analysis concerning expected returns of options on volatility.

Assumption C.1. The pricing kernel $m[z, v]$ is declining and convex in market index returns z for each volatility v of market index returns.

Assumption C.1 is broadly consistent with asset-pricing theory that often postulates a pricing kernel that is declining and convex in market index returns (e.g., Rubinstein (1976), Brennan (1979), Hansen and Jagannathan (1991), and Aït-Sahalia and Lo (2000)).

Assumption C.2. $m[z, v]$ is increasing in the volatility v of market index returns for each z.

Assumption C.2 reflects the economic intuition that high-volatility states are generally disliked and, hence, may be associated with a high level of the pricing kernel.

Assumption C.3. The conditional density of market index returns z given volatility v, $\widehat{\Phi}_1[z|v]$, with lower volatility dominates, in the sense of second-order stochastic dominance, those with higher volatility.

Whereas Assumptions C.1 and C.2 relate to the pricing kernel $m[z, v]$, Assumption C.3 is about the physical probability measure. Assumption C.3 is based on the argument that increases in volatility that are not compensated by an increase in the expected market return would be disliked by concave utilities, and hence the second-order stochastic dominance (e.g., Huang and Litzenberger (1988)).

Theorem 2. Under Assumptions C.1, C.2, and C.3, expected returns of puts (calls) on volatility decrease in strike and are positive (negative) for sufficiently low (high) strikes.

Proof. Follows from Lemma 1, Lemma 2, and Lemma 3 below.

Recall that in our notation $z \in (-\infty, \infty)$ is market index return, $v \in (0, \infty)$ is volatility of market index return, and $\Phi[z, v]$ is their joint probability density.

The expected return $\mu_v[K]$ of an option on volatility with strike K is,

$$1 + \mu_v[K] = \frac{\int_0^\infty \int_{-\infty}^\infty h[v\,;K]\, \Phi[z, v]\, dz\, dv}{\int_0^\infty \int_{-\infty}^\infty h[v\,;K]\, m[z, v]\, \Phi[z, v]\, dz\, dv}, \tag{C-2}$$

where $h[v\,;K] = (K - v)^+$ or $h[v\,;K] = (v - K)^+$, respectively, for volatility puts and calls.

Upon integrating the market index return z, we may write equation (C-2) as,

$$1 + \mu_v[K] = \frac{\int_0^\infty h[v\,;K]\, \Phi_2[v]\, dv}{\int_0^\infty h[v\,;K]\, \widehat{m}[v]\, \Phi_2[v]\, dv}, \tag{C-3}$$

where $\widehat{m}[v]$ and the marginal density $\Phi_2[v]$ of v are defined as,

$$\begin{aligned} \widehat{m}[v] &= \int_{-\infty}^\infty m[z, v]\, \widehat{\Phi}_1[z|v]\, dz \quad \text{and} \\ \Phi_2[v] &= \int_{-\infty}^\infty \Phi[z, v]\, dz \ \text{ for } v \in (0, \infty), \end{aligned} \tag{C-4}$$

https://doi.org/10.1017/S0022109015000642 Published online by Cambridge University Press

and $\widehat{\Phi}_1[z|v] = \Phi[z, v]/\Phi_2[v]$ is the conditional density of z, with $\int_{-\infty}^{\infty} \widehat{\Phi}_1[z|v]\, dz = 1$ (see, e.g., DeGroot and Schervish (2002)).

According to Lemma 1, the expectation of $m[z, v]$ conditional on volatility is increasing in volatility (i.e., $d\widehat{m}[v]/dv > 0$). Besides, $\int_0^{\infty} \widehat{m}[v]\, \Phi_2[v]\, dv = 1$. Therefore, the conditions of Lemma 3 are satisfied for $\widehat{m}[v]$ and $\Phi_2[v]$ in equation (C-3), and the statements in Theorem 2 follow. □

Lemma 1. For a function $m[z, v] > 0$, let $\partial m[z, v]/\partial z < 0$ (Assumption C.1), and $\partial m[z, v]/\partial v > 0$ (Assumption C.2). Then, by Assumption C.3 and Lemma 2, we have $d\widehat{m}[v]/dv > 0$, where $\widehat{m}[v]$ is as defined in equation (C-4).

Proof. Observe that

$$\text{(C-5)} \qquad \frac{d\widehat{m}[v]}{dv} = \int_{-\infty}^{\infty} \frac{\partial m[z, v]}{\partial v}\, \widehat{\Phi}_1[z|v]\, dz + \int_{-\infty}^{\infty} m[z, v]\, \frac{\partial \widehat{\Phi}_1[z|v]}{\partial v}\, dz.$$

The first term in equation (C-5) is positive, given our condition on $\partial m[z, v]/\partial v > 0$. The statement of the lemma follows, since the second term is also positive given equation (C-6) of Lemma 2, for all v. □

Lemma 2. Suppose $v^c \leq v^d$. Furthermore, suppose lower volatility conditional densities second-order stochastically dominate the higher volatility conditional densities with $\widehat{\Phi}_1[z|v^c] \overset{\text{ssd}}{\geq} \widehat{\Phi}_1[z|v^d]$ (our Assumption C.3). Then,

$$\text{(C-6)} \qquad \int \mathcal{M}[z]\, \widehat{\Phi}_1[z|v^c]\, dz \leq \int \mathcal{M}[z]\, \widehat{\Phi}_1[z|v^d]\, dz, \qquad \text{for } v^c \leq v^d,$$

for a decreasing convex function $\mathcal{M}[z]$, given some fixed volatility.

Proof. From Theorem 2.93 of Föllmer and Schied (2004), p. 103), we have

$$\text{(C-7)} \qquad \int (-\mathcal{M}[z])\, \widehat{\Phi}_1[z|v^c]\, dz \geq \int (-\mathcal{M}[z])\, \widehat{\Phi}_1[z|v^d]\, dz, \qquad \text{for } v^c \leq v^d,$$

for an increasing concave function $(-\mathcal{M}[z])$. Reversing the inequality gives equation (C-6). □

Lemma 3. Suppose $m[v]$ is a pricing kernel, $\int_0^{\infty} m[v]\, \Phi[v]\, dv = 1$, and $\Phi[v]$ is the density of v. When $dm[v]/dv > 0$, the following statements are true: i) expected put returns decrease in strike and are positive for sufficiently low strikes, and ii) expected call returns decrease with strike and are negative for sufficiently high strikes.

Proof. For the proof of the signs and strike dependence of expected option returns in i) and ii), see, for example, Coval and Shumway (2001), with the exception that our results pertain to a pricing kernel that is monotonically increasing in volatility (i.e., $dm[v]/dv > 0$). □

The proof of Theorem 2 hinges on the result that the expectation of $m[z, v]$, conditional on volatility, is increasing in volatility (see Lemma 1). Such a result can also be derived under the alternative stronger assumption that the conditional density of z given v, $\widehat{\Phi}_1[z|v]$, with lower volatility first-order stochastically dominates the higher volatility counterparts, in which case one can relax the assumption of convexity. First-order stochastic dominance could be obtained, for instance, when there is a strong leverage effect (e.g., Black (1976)) and high-volatility states map to reduced expected market returns.

One may view Theorem 2 as extending the theoretical results on expected returns of options written on the market index (Coval and Shumway (2001)).

Theorem 2 and, hence the pattern of expected returns of volatility options across moneyness, provides testable implications for the change of measures, or pricing kernels, that can be supported in the volatility market.

https://doi.org/10.1017/S0022109015000642 Published online by Cambridge University Press

https://doi.org/10.1017/S0022109015000642 Published online by Cambridge University Press

References

Aït-Sahalia, Y.; M. Karaman; and L. Mancini. "The Term Structure of Variance Swaps and Risk Premia." Working Paper, Princeton University (2013).

Aït-Sahalia, Y., and A. Lo. "Nonparametric Risk Management and Implied Risk Aversion." *Journal of Econometrics*, 94 (2000), 9–51.

Aït-Sahalia, Y., and L. Mancini. "Out of Sample Forecasts of Quadratic Variation." *Journal of Econometrics*, 147 (2008), 17–33.

Amengual, D., and D. Xiu. "Resolution of Policy Uncertainty and Sudden Declines in Volatility." Working Paper, University of Chicago (2013).

Andersen, T.; T. Bollerslev; F. Diebold; and P. Labys. "Modeling and Forecasting Realized Volatility." *Econometrica*, 71 (2003), 529–626.

Andersen, T.; T. Bollerslev; and N. Meddahi. "Correcting the Errors: Volatility Forecast Evaluation Using High-Frequency Data and Realized Volatilities." *Econometrica*, 73 (2005), 279–296.

Bagnoli, M., and T. Bergstrom. "Log-Concave Probability and its Applications." *Economic Theory*, 26 (2005), 445–469.

Bakshi, G.; D. Madan; and G. Panayotov. "Returns of Claims on the Upside and the Viability of U-Shaped Pricing Kernels." *Journal of Financial Economics*, 97 (2010), 130–154.

Basak, S. "A Model of Dynamic Equilibrium Asset Pricing with Heterogeneous Beliefs and Extraneous Risk." *Journal of Economic Dynamics and Control*, 24 (2000), 63–95.

Basak, S. "Asset Pricing with Heterogeneous Beliefs." *Journal of Banking and Finance*, 29 (2005), 2849–2881.

Black, F. "Studies of Stock Price Volatility Changes." In *Proceedings of the 1976 American Statistical Association, Business and Economical Statistics Section*, Alexandria, VA: American Statistical Association (1976), 177–181.

Bollerslev, T., and V. Todorov. "Tails, Fears, and Risk Premia." *Journal of Finance*, 66 (2011), 2165–2211.

Brennan, M. "The Pricing of Contingent Claims in Discrete Time Models." *Journal of Finance*, 34 (1979), 53–68.

Briys, E.; M. Crouhy; and H. Schlesinger. "Optimal Hedging under Intertemporally Dependent Preferences." *Journal of Finance*, 45 (1990), 1315–1324.

Buehler, H. "Consistent Variance Curve Models." *Finance and Stochastics*, 10 (2006), 178–203.

Buraschi, A., and A. Jiltsov. "Model Uncertainty and Option Markets with Heterogeneous Beliefs." *Journal of Finance*, 61 (2006), 2841–2897.

Calvet, L.; J.-M. Grandmont; and I. Lemaire. "Aggregation of Heterogeneous Beliefs, Asset Pricing and Risk Sharing in Complete Markets." Working Paper, HEC Paris, and CREST (2004).

Carr, P., and R. Lee. "Robust Replication of Volatility Derivatives." Working Paper, New York University and University of Chicago (2008).

Carr, P., and L. Wu. "Variance Risk Premia." *Review of Financial Studies*, 22 (2009), 1311–1341.

Chen, H.; S. Joslin; and N.-K. Tran. "Rare Disasters and Risk Sharing with Heterogeneous Beliefs." *Review of Financial Studies*, 25 (2012), 2189–2224.

Chicago Board Options Exchange. "VIX: Fact & Fiction." Working Paper, Chicago Board Options Exchange (2009).

Christoffersen, P.; S. Heston; and K. Jacobs. "Capturing Option Anomalies with a Variance-Dependent Pricing Kernel." *Review of Financial Studies*, 26 (2013), 1963–2006.

Cont, R., and T. Kokholm. "A Consistent Pricing Model for Index Options and Volatility Derivatives." *Mathematical Finance*, 23 (2013), 248–274.

Coval, J., and T. Shumway. "Expected Option Returns." *Journal of Finance*, 56 (2001), 983–1009.

Cox, J. C., and C. Huang. "Optimal Consumption and Portfolio Policies When Asset Prices Follow a Diffusion Process." *Journal of Economic Theory*, 49 (1989), 33–83.

DeGroot, M., and M. Schervish. *Probability and Statistics*. New York, NY: Addison Wesley (2002).

Detemple, J., and S. Murthy. "Intertemporal Asset Pricing with Heterogeneous Beliefs." *Journal of Economic Theory*, 62 (1994), 294–320.

Dieckmann, S. "Rare Event Risk and Heterogeneous Beliefs: The Case of Incomplete Markets." *Journal of Financial and Quantitative Analysis*, 46 (2011), 459–488.

Egloff, D.; M. Leippold; and L. Wu. "The Term Structure of Variance Swap Rates and Optimal Variance Swap Investments." *Journal of Financial and Quantitative Analysis*, 45 (2010), 1279–1310.

Föllmer, H., and A. Schied. *Stochastic Finance: An Introduction in Discrete Time*. Berlin, Germany: Walter de Gruyer (2004).

Gatheral, J. "Consistent Modeling of SPX and VIX Options." In *The Fifth World Congress of the Bachelier Finance Society*. London, UK: Wiley (2008).

Grossman, S., and J. Stiglitz. "On the Impossibility of Informationally Efficient Markets." *American Economic Review*, 70 (1980), 393–408.

Hansen, L., and R. Jagannathan. "Implications of Security Market Data for Dynamic Economies." *Journal of Political Economy*, 99 (1991), 225–261.
Hong, H. "A Model of Returns and Trading in Futures Markets." *Journal of Finance*, 55 (2000), 959–988.
Huang, C., and R. Litzenberger. *Foundations for Financial Economics*. Upper Saddle River, NJ: Prentice Hall (1988).
Jackwerth, J., and G. Vilkov. "Asymmetric Volatility Risk: Evidence from Option Markets." Working Paper, University of Konstanz (2013).
Jouini, E., and C. Napp. "Consensus Consumer and Intertemporal Asset Pricing with Heterogeneous Beliefs." *Review of Economic Studies*, 74 (2007), 1149–1174.
Karatzas, I., and S. Shreve. *Brownian Motion and Stochastic Calculus*. New York, NY: Springer-Verlag (1991).
Kogan, L.; S. Ross; J. Wang; and M. Westerfield. "The Price Impact and Survival of Irrational Traders." *Journal of Finance*, 61 (2006), 195–229.
Leland, H. "Who Should Buy Portfolio Insurance?" *Journal of Finance*, 35 (1980), 581–594.
Lin, Y.-N. "VIX Option Valuation." Working Paper, National Chung Hsing University (2009).
Neftci, S. *An Introduction to the Mathematics of Financial Derivatives*. New York, NY: Academic Press (2007).
Patton, A., and A. Timmermann. "Monotonicity in Asset Returns: New Tests with Applications to the Term Structure, the CAPM and Portfolio Sorts." *Journal of Financial Economics*, 98 (2010), 605–625.
Rubinstein, M. "The Valuation of Uncertain Income Streams and the Pricing of Options." *Bell Journal of Economics*, 7 (1976), 407–425.
Rubinstein, M. "Implied Binomial Trees." *Journal of Finance*, 49 (1994), 771–818.
Sepp, A. "VIX Option Pricing in a Jump-Diffusion Model." *Risk*, April (2008), 84–89.
Song, Z., and D. Xiu. "A Tail of Two Option Markets: State-Price Densities and Volatility Risk." Working Paper, Federal Reserve Board and University of Chicago (2013).
Todorov, V. "Variance Risk Premium Dynamics: The Role of Jumps." *Review of Financial Studies*, 23 (2010), 345–383.
Xiong, W. "Bubbles, Crises, and Heterogeneous Beliefs." Working Paper, Princeton University (2013).
Xiong, W., and H. Yan. "Heterogeneous Expectations and Bond Markets." *Review of Financial Studies*, 23 (2010), 1433–1466.
Zhu, Y., and J. Zhang. "Variance Term Structure and VIX Futures Pricing." *International Journal of Theoretical and Applied Finance*, 10 (2007), 111–127.

https://doi.org/10.1017/S0022109015000642 Published online by Cambridge University Press