Working Paper No. RHS-06-028

January 2006

A Theory of Volatility Spreads

Gurdip Bakshi

Dilip Madan

This Paper can be downloaded without charge from the
Social Science Research Network Electronic Paper Collection
http://ssrn.com/abstract=828244

A Theory of Volatility Spreads

Gurdip Bakshi and Dilip Madan*

January 30, 2006

Abstract

This study formalizes the departure between risk-neutral and physical index return volatilities, termed volatility spreads. Theoretically, the departure between risk neutral and physical index volatility is connected to the higher-order physical return moments and the parameters of the pricing kernel process. This theory predicts positive volatility spreads when investors are risk averse, and when the physical index distribution is negatively skewed and leptokurtic. Our empirical evidence is supportive of the theoretical implications of risk aversion, exposure to tail events, and fatter left-tails of the physical index distribution in markets where volatility is traded.

*Bakshi and Madan are both at Department of Finance, Robert H. Smith School of Business, University of Maryland, College Park, MD 20742. Bakshi can be reached at Tel: 301-405-2261, Email: gbakshi@rhsmith.umd.edu, and Website: www.rhsmith.umd.edu/faculty/gbakshi/; Madan at Tel: 301-405-2127, Email: dbm@rhsmith.umd.edu, and Website: www.rhsmith.umd.edu/faculty/dmadan/. The review process has substantially improved the paper, and we are grateful for the feedback of the referee, the anonymous associate editor, and David Hsieh (the department editor). The paper has benefited from discussions with Doron Avramov, Charles Cao, Steve Heston, Nengjiu Ju, Nikunj Kapadia, Inanc Kirgiz, Mark Loewenstein, Carl Ullrich, Liuren Wu, and Frank Zhang. Any remaining errors are ours. We welcome comments, including references to related papers we have inadvertently overlooked.

1 Introduction

Earlier empirical work from Canina and Figlewski (1993), Lamoureux and Lastrapes (1993), Bakshi, Cao, and Chen (2000), and Christoffersen, Heston, and Jacobs (2005) based on implied volatility, and now Bollerslev, Gibson, and Zhou (2005), Britten-Jones and Neuberger (2000), Carr and Wu (2004) and Jiang and Tian (2005) based on formal measures of risk-neutral volatility, suggests the finding that risk-neutral index volatility generally exceeds physical return volatility. Despite the advances in modeling volatility, there appears to be little theoretical work that connects the two entities within the same underlying economic equilibrium. The purpose of this article is to formalize the departure between risk-neutral and physical index return volatilities, termed volatility spreads. What are the theoretical determinants of volatility spreads? What role does investor behavior play in explaining dynamic movements in volatility spreads? Our contribution lies in identifying the sufficient conditions on physical and risk-neutral densities that give rise to positive volatility spreads.

The compelling force behind the theory of volatility spreads is the notion that risk-neutral probabilities are physical probabilities revised by investor's risk preference as determined by the pricing kernel (Harrison and Kreps, 1979). Rational risk-averse investors are sensitive to extreme loss states and are willing to counteract these exposures by buying protection. The desire to cover these losses typically drives up the risk-neutral probability relative to the actual probability of occurrence, and endogenously shifts probability mass to risk-neutral tails. This mechanism essentially argues that risk aversion introduces heterogeneity in the risk-neutral distributions and causes non-inverted volatility spreads.

To finesse the above elementary intuition, we first present a four-state example in a two-date setting. Through a judicious design of physical probabilities and returns, this model distinctly parameterizes physical skewness and physical kurtosis and accommodates exposures to both up and down extremes. Using a naive pricing kernel representation that amplifies both the up and down risk-neutral tail probabilities, we analytically derive the parametric restrictions on physical skewness and physical kurtosis that guarantee the existence of positive volatility spreads. This analysis verifies the core intuition that volatility spreads get bigger when the investor dislike for unpleasant tail events gets progressively more severe. Theoretically it is shown that the existence of positive volatility spreads warrant exposure to tail events that generate fat-tailed physical return distributions. Furthermore the introduction of negative physical skewness, at the margin, tends to magnify the effect of physical kurtosis on volatility spreads.

The general modeling framework richens the state-space example and provides a deeper understanding of the issues by (i) relaxing the assumption that the parametric form of the physical density is known, and by (ii) focusing on a family of theoretically appealing pricing kernels. Thus, it must be noted that our results hold for a wide class of physical return densities and are derived without imposing any structure on the dynamics of physical volatility and without assuming the functional form of the volatility risk premium. The main analytical characterization reveals an equation where volatility spreads are related to the pricing kernel and to the higher-order moments of the physical return distribution. For a broad set of pricing kernels that display risk aversion, we obtain approximate theoretical results that predict positive volatility spreads when the physical index distribution is negatively skewed and leptokurtic. We also develop extensions to more general marginal utilities and specify the conditions under which our results on volatility spreads are completely exact.

Based on the theoretical predictions, we empirically examine the determinants of volatility spreads using S&P 100 index returns and options on the S&P 100 index. Consistent with the modeling paradigm, the empirical work shows that volatility spreads are substantially positive with risk-neutral volatility exceeding physical volatility for the majority of the months (see also the contributions in Bollerslev, Gibson, and Zhou, 2005, Carr and Wu, 2004, and Jiang and Tian, 2005). Estimation of the volatility spread equation shows that the estimated risk aversion parameter is positive and statistically significant and comparable to Aït-Sahalia and Lo (2000) and Bliss and Panigirtzoglou (2004). This estimation provides a perspective in evaluating risk aversion in markets where volatility is traded. Such an exercise also confirms our theoretical prediction that positive volatility spreads are primarily due to the existence of risk aversion, exposure to tail events, and fatter left-tails of the physical index distribution.

In what follows, Section 2 outlines the framework for analyzing volatility spreads. The simple state-space example presented in Section 3 is aimed towards providing the intuition for the general theory of volatility spreads. Section 4 presents theoretical results that hold for arbitrary physical return distributions. Section 5 describes the empirical characteristics of volatility spreads. Section 6 presents the generalized method of moments estimation results from fitting theoretical volatility spreads and the estimates of risk aversion exploiting the volatility time-series. Conclusions are offered in Section 7.

2 A Framework for Analyzing Volatility Spreads

To fix the notation for the theoretical results, denote the τ-period return on the market index as $R(t,\tau) \equiv \ln\left[S\left(t+\tau\right)/S(t)\right]$, where $S(t)$ is the time-t level of the market index. For brevity of equation presentation, write $R(t,\tau)$ as R and denote the physical density of the market return by $p[R]$. Assume the existence of a risk-neutral (pricing) density, $q[R]$.

Ruling out extreme counterexamples make the assumption that the physical and risk-neutral return densities have finite moments up to order four:

$$\left| \int_{\Re} R^n \, p[R] \, dR \right| \; < \; \infty, \tag{1}$$

$$\left| \int_{\Re} R^n \, q[R] \, dR \right| \; < \; \infty, \tag{2}$$

for $n = 1, \ldots, 4$, and $\Re$ is the real line.

To streamline notation, define the higher-order return central moments under the physical measure as:

$$\sigma_p^2(t,\tau) \;\; \equiv \;\; \int_{\Re} (R - \mu_p)^2 \, p[R] \, dR, \tag{3}$$

$$\theta_p(t,\tau) \;\; \equiv \;\; \frac{\int_{\Re} (R - \mu_p)^3 \, p[R] \, dR}{\sigma_p^2(t,\tau)^{3/2}}, \tag{4}$$

$$\kappa_p(t,\tau) \;\; \equiv \;\; \frac{\int_{\Re} (R - \mu_p)^4 \, p[R] \, dR}{\sigma_p^2(t,\tau)^2}, \tag{5}$$

where the mean return under the physical probability measure is

$$\mu_p \;\; \equiv \;\; \int_{\Re} R \, p[R] \, dR. \tag{6}$$

Following a standard practice, define the risk-neutral market volatility as:

$$\sigma_{rn}^2(t,\tau) \equiv \int_{\Re} (R - \mu_{rn})^2 \, q[R] \, dR, \tag{7}$$

where μ_{rn} represents the mean of the risk-neutral return distribution.

The focus of this article is on characterizing *volatility spreads*, defined as:

$$\frac{\sigma_{rn}^2(t,\tau) - \sigma_p^2(t,\tau)}{\sigma_p^2(t,\tau)}. \tag{8}$$

We ask two questions. First, what are the theoretical determinants of volatility spreads? Second, what are the relative roles of risk preferences of the investors and return distributional properties in explaining movements in risk-neutral volatility and volatility spreads? In our theoretical exposition, we confine attention to a frugal description of $p[R]$ and $q[R]$ and then identify sufficient conditions that support the existence of positive volatility spreads.

To complete the description of the economic environment, consider any generic pricing kernel represented by $m[R]$. When the physical density (or the physical probability) is suitably adjusted, then assets can be priced in a risk-neutral manner (Harrison and Kreps, 1979). For any risk-neutral density, $q\,[R]$, and the physical density of the market index, $p\,[R]$, the pricing kernel, $m[R]$, is classically defined as (for a formal treatment, see Constantinides, 2002, Halmos, 1970, Harrison and Kreps, 1979, and Bakshi, Kapadia, and Madan, 2003):

$$m\,[R] = \frac{q[R]}{p\,[R]} \tag{9}$$

and note, in particular, that $\int_{\Re} q\,[R]\ dR = 1$. Alternatively one may state that the unnormalized risk-neutral density is of the form $m[R]p[R]$ for an arbitrary positive function $m[R]$, whereby on normalization we get:

$$q\,[R] = \frac{m[R]\,p\,[R]}{\int_{\Re} m[R]\,p\,[R]\ dR}, \tag{10}$$

where the pricing kernel, $m[R]$ - that prices all future payoffs - is some function of the market return R.

According to (10), the risk-neutral probabilities are physical probabilities revised by investor's risk preference reflecting the price that an investor would pay for receiving one unit of payoff in a particular state of the world. Finance theory suggests that risk-averse investors pay more attention to *unpleasant* extreme states relative to their actual probability of occurrence, and thus exaggerate the risk-neutral probability of unpleasant states in comparison to the physical counterpart. Intuitively, risk aversion - that induces a risk premium - can introduce a wedge between the moments of the risk-neutral and physical return distributions. When $m[R] = \hat{m} > 0$, for constant $\hat{m}$, equation (10) results in $q[R] = p[R]$, and implies the absence of risk premium on all claims contingent on the market return.

When it comes to claims written on the market return for which we have potential data, it does not matter how general the pricing kernel is and whether $m[R]$ admits state dependencies. To see this, the price at time t of a claim paying $c[R]$ at $t+\tau$ is,

$$\begin{aligned}\Pi_c(t) &= B(t,\tau)\, E^P\left(\Lambda\, c[R]\right), &(11)\\ &= B(t,\tau)\, E^P\Big(E^P\left(\Lambda\, c[R] \mid R\right)\Big) = B(t,\tau)\, E^P\left(\Lambda^*[R]\, c[R]\right), &(12)\end{aligned}$$

where $B(t,\tau)$ is the discount bond price at t paying unit at $t+\tau$ and $E^P(.)$ is expectation under the physical probability measure P. Moreover, Λ represents the change of probability with the bond price as numeraire, and $\Lambda^*[R] \equiv E^P(\Lambda \mid R)$. Thus, we need only model $E^P(\Lambda \mid R)$ and need not concern ourselves with any state dependent specifications of the pricing kernel.

3 Volatility Spreads in a Simple Model

To provide the basic intuition for positive volatility spreads, consider a state-space example in a two-date world. Through this static model, we can more directly understand how exaggerating the risk-neutral tails enables positive volatility spreads by making the risk-neutral return distribution less homogeneous. Albeit specialized to maintain parsimony, this framework is sufficiently flexible to study the impact of physical skewness and physical kurtosis on volatility spreads.

Consider a four state-space economy indexed by $\omega = (\omega_1, \omega_2, \omega_3, \omega_4)$. The associated physical probabilities and market return in the four states are shown below:

$$p[\omega] = \begin{pmatrix} \frac{1-\frac{(1-\upsilon)(1+\beta)}{2}}{2} \\ \frac{(1-\upsilon)\beta}{2} \\ \frac{(1-\upsilon)}{2} \\ \frac{1-\frac{(1-\upsilon)(1+\beta)}{2}}{2} \end{pmatrix}, \qquad R[\omega] = \begin{pmatrix} -\rho\,\hat{R} \\ -\hat{R} \\ \beta\hat{R} \\ \rho\,\hat{R} \end{pmatrix}, \qquad (13)$$

where $0 \leq \upsilon \leq 1$ and $\hat{R}$ is a strictly positive scaling constant. Because results are independent of scaling we henceforth set $\hat{R} = 1$. It is clear that $\sum_\omega p[\omega_j] = 1$. The construction of probabilities and returns outlined in (13) can be motivated as follows. First, the parameter υ calibrates the exposure to the extremes (and equals the probability of extremes in the symmetric case of $\beta = 1$) and $\rho > 1$ controls the kurtosis of the physical return distribution. Second, as verified in (16), the parameter $\beta \leq 1$ captures the negative asymmetry of the physical return distribution. Third, ω_1 (ω_4) represents the most unpleasant extreme state from the perspective of long (short) market position with exogenously specified return $-\rho$ (ρ). With small υ, both the large downward movement and the large upward movement are rare and share the same event probability. Finally, when $\upsilon \to 0$ and $\beta \approx 1$, the probability mass in the extremes shrinks to zero and the modeling structure collapses to the two-state classical binomial model.

Based on the assumed probabilities and returns,

$$\begin{aligned}
\mu_p &= \sum_{\omega} p[\omega_j]\, R[\omega_j] = 0, & (14)\\
\sigma_p^2 &= \sum_{\omega} p[\omega_j]\, (R[\omega_j] - \mu_p)^2 = \rho^2 \left(1 - \frac{(1-\upsilon)(1+\beta)}{2}\right) + \frac{\beta(1+\beta)(1-\upsilon)}{2}, & (15)\\
\theta_p &= \frac{1}{(\sigma_p^2)^{3/2}} \sum_{\omega} p[\omega_j]\, (R[\omega_j] - \mu_p)^3 = -\frac{(1-\upsilon)\,\beta\,(1-\beta^2)}{2(\sigma_p^2)^{3/2}} < 0, & (16)\\
\kappa_p &= \frac{1}{(\sigma_p^2)^2} \sum_{\omega} p[\omega_j]\, (R[\omega_j] - \mu_p)^4 = \frac{\rho^4 \left(1 - \frac{(1-\upsilon)(1+\beta)}{2}\right) + \frac{\beta(1+\beta^3)(1-\upsilon)}{2}}{\sigma_p^4}. & (17)
\end{aligned}$$

Although the four-state framework is chosen to provide the intuition for the general theoretical results on volatility spreads, the model nonetheless illustrates a number of economically sensible properties. Equation (15) shows that the presence of extremes adds to the dispersion of the return distribution and this dependence dominates at large ρ. Consider the symmetric case of $\beta = 1$, that makes $\theta_p = 0$ in (16), which leads to kurtosis: $\kappa_p = \frac{1+(\rho^4-1)\upsilon}{1+(\rho^2-1)^2\upsilon^2+2(\rho^2-1)\upsilon}$. Under the further consideration of small exposure to the extremes (i.e., $\upsilon \approx 0$), the kurtosis is of the order $(\rho^4-1)/(\rho^2-1)$ and kurtosis exceeds 3 for $\rho > \sqrt{2}$. Otherwise this model produces rich combinations of negative skewness and kurtosis, as respectively controlled by β and ρ.

With the structure of probabilities and returns as in (13), consider the following specification for the pricing kernel $m[\omega]$:

$$m[\omega] = \begin{pmatrix} \alpha \\ 1 \\ 1 \\ \alpha \end{pmatrix}, \qquad (18)$$

with $\alpha > 1$. This particularly parameterized choice of $m[\omega]$ reflects the idea that investor long (short) the market index are willing to pay a higher charge to hedge the payoff in $\omega_1(\omega_4)$. The bowl-shaped choice of $m[\omega]$ in ω-$m[\omega]$ space approximates a heterogeneous agent economy populated by agents with long and short positions in the market index (see Case 1 to come). Specifically the same weighting of α in state ω_1 and ω_4 inflates the un-normalized risk-neutral tail probabilities leaving the neck of the distribution unchanged. This assumption simplifies computation by keeping the risk-neutral mean at zero and isolating attention on the volatility effect. More realistically one would differentially inflate the downside with respect to the upside, as in the power utility $m[\omega] = [e^{\gamma\rho}, e^{\gamma}, e^{-\gamma\beta}, e^{-\gamma\rho}]'$ with risk aversion γ. Such a pricing kernel complicates the variance calculations by introducing a shift in the risk-neutral mean without changing the qualitative results on volatility spreads.

To derive the risk-neutral probabilities we appeal to the relation $q[\omega] = \frac{m[\omega]\, p[\omega]}{\sum_\omega m[\omega]\, p[\omega]}$, which is the state-space counterpart of (10), and arrive at the expression below:

$$q[\omega] = \begin{pmatrix} \frac{\alpha\left(1-\frac{(1-v)(1+\beta)}{2}\right)}{2\Psi} \\ \frac{(1-v)\beta}{2\Psi} \\ \frac{(1-v)}{2\Psi} \\ \frac{\alpha\left(1-\frac{(1-v)(1+\beta)}{2}\right)}{2\Psi} \end{pmatrix}, \tag{19}$$

where the normalization factor $\Psi \equiv \sum_\omega m[\omega]\, p[\omega] = \alpha\left(1 - \frac{(1-v)(1+\beta)}{2}\right) + \frac{(1-v)(1+\beta)}{2} > 0$. Comparing (13) and (19), the risk-neutral probability in state ω_1 and ω_4 is higher than the physical counterpart provided $\alpha > \Psi$, which is satisfied when $\alpha > 1$. The price of Arrow-Debreu security paying unit in state ω_1 (or ω_4) is $\frac{\alpha}{\Psi}\left(\frac{1-\frac{(1-v)(1+\beta)}{2}}{2}\right)$. Hence for all $\alpha > 1$ the Arrow-Debreu price of the tail event is greater than its physical probability.

Due to our distinct parameterization of higher-order physical moments, the risk-neutral volatility, σ^2_{rn}, will be driven by the skewness parameter β, the kurtosis parameter ρ, and the risk-neutral tail inflation factor α. Write $\sigma^2_{rn}[\alpha;\beta,\rho]$ to reflect this theoretical dependence. We wish to demonstrate analytically that volatility spreads will widen when investors care about avoiding unpleasant tail events. Since σ^2_p is independent of α, this result amounts to showing that $\left.\frac{\partial\sigma^2_{rn}[\alpha;\beta,\rho]}{\partial\alpha}\right|_{\alpha=1} > 0$. The rationale behind this restriction is that when $\alpha = 1$, $m[\omega]$ is unity in every state making risk-neutral and physical probabilities identical. It can be shown that the risk-neutral volatility is,

$$\sigma^2_{rn}[\alpha;\beta,\rho] \;=\; \sum_{\omega} q[\omega_j]\ (R[\omega_j]-\mu_q)^2 = \frac{\alpha\,\rho^2\left(1-\frac{(1-\upsilon)(1+\beta)}{2}\right)+\frac{\beta(1+\beta)(1-\upsilon)}{2}}{\alpha\left(1-\frac{(1-\upsilon)(1+\beta)}{2}\right)+\frac{(1-\upsilon)(1+\beta)}{2}}. \qquad (20)$$

Observe from (15) and (20) that for skewed return distributions, the difference between risk-neutral and physical volatilities is,

$$\sigma^2_{rn}-\sigma^2_p = \left(1-\frac{(1-\upsilon)(1+\beta)}{2}\right)\left(\frac{(\alpha-1)(1-\upsilon)(1+\beta)/2}{\alpha\left(1-\frac{(1-\upsilon)(1+\beta)}{2}\right)+\frac{(1-\upsilon)(1+\beta)}{2}}\right)\left(\rho^2-\beta\right). \qquad (21)$$

For fat-tailed distributions with kurtosis in excess of the stated benchmark level of unity, the right hand side of (21) is positive, and convex in ρ.

The derivative of risk-neutral volatility with respect to α evaluated at $\alpha = 1$ is,

$$\left.\frac{\partial\sigma^2_{rn}[\alpha;\beta,\rho]}{\partial\alpha}\right|_{\alpha=1} = \left(1-\frac{(1-\upsilon)(1+\beta)}{2}\right)\left(\frac{(1-\upsilon)(1+\beta)}{2}\right)\left(\rho^2-\beta\right) > 0, \qquad (22)$$

since $\rho^2 > 1$ and $\beta < 1$. Ceteris paribus, a higher α magnifies the tilt in the tails and causes more pronounced volatility spreads. The impact of return asymmetry can be addressed by computing the derivative of $\frac{\partial(\sigma^2_{rn}[\alpha;\beta,\rho])}{\partial\alpha}$ with respect to β evaluated at $\beta = 1$. This calculation yields,

$$\left.\frac{\partial^2\sigma^2_{rn}[\alpha;\beta,\rho]}{\partial\alpha\,\partial\beta}\right|_{\alpha=1,\beta=1} \;=\; \left(-\upsilon(1-\upsilon)-(\rho^2-1)(1-\upsilon)\left(\frac{1}{2}-\upsilon\right)\right). \qquad (23)$$

Introducing negative skewness by decreasing β does increase risk-neutral volatility, but for leptokurtic cases this cross-partial derivative is negative (provided $\upsilon < \frac{1}{2}$) with the consequence that the introduction of negative skewness at the margin inflates the effect of kurtosis on risk neutral volatilities.

4 Structure of Volatility Spreads in Dynamic Economies

This section provides general theoretical results on volatility spreads by imposing a plausible economic environment on the pricing kernel $m[R]$. The generality is related to the parametric form of the physical return density, $p[R]$, being unspecified. Our objective is to connect volatility spreads to higher-order moments of the physical density and the risk preference of investors. Specifically we expand on the state-space example in Section 3 and provide interpretations for the main drivers of volatility spreads in a broader economic setting.

Theorem 1 *Suppose the aggregate investor behavior is modeled through a class of pricing kernels $m[R]$. For pricing kernels satisfying the Taylor series expansion of $m[R]$, around zero, $m[R] \approx 1 - \mathcal{A}_1 R + \frac{1}{2}\mathcal{A}_2 R^2 + O[R^3]$, where $m[0] = 1$, $\mathcal{A}_1 \equiv -\frac{\partial m}{\partial R}\mid_{R=0}$, and $\mathcal{A}_2 \equiv \frac{\partial^2 m}{\partial R^2}\mid_{R=0}$, the τ-period volatility spreads are theoretically determined as,*

$$\begin{aligned}\frac{\sigma^2_{rn}(t,\tau) - \sigma^2_p(t,\tau)}{\sigma^2_p(t,\tau)} \approx & -\mathcal{A}_1 \left(\sigma^2_p(t,\tau)\right)^{1/2} \times \theta_p(t,\tau) \\ & + \frac{\mathcal{A}_2}{2}\left(\sigma^2_p(t,\tau)\right) \times \left(\kappa_p(t,\tau) - 1 - \frac{2\,\mathcal{A}_1^2}{\mathcal{A}_2}\right), \qquad (24)\end{aligned}$$

where the terms related to risk aversion, $\mathcal{A}_1 \equiv -\frac{\partial m}{\partial R}\mid_{R=0}$ and $\mathcal{A}_2 \equiv \frac{\partial^2 m}{\partial R^2}\mid_{R=0}$, are constants. The decreasing risk aversion assumption implies $\frac{\partial m}{\partial R} < 0$ and $\frac{\partial^2 m}{\partial R^2} > 0$. The skewness of the physical distribution, $\theta_p(t,\tau)$, and the kurtosis of the physical distribution, $\kappa_p(t,\tau)$, are respectively defined in (4) and (5). Furthermore, (24) is exact for quadratic pricing kernels.

Under the power utility class with $m[R] = e^{-\gamma R}$ for coefficient of relative risk aversion γ, up to a second-order of γ, we have $\mathcal{A}_1 = \gamma$, $\mathcal{A}_2 = \gamma^2$ and $\frac{\mathcal{A}_1^2}{\mathcal{A}_2} = 1$, and thus the approximate volatility spreads are governed by the relationship,

$$\frac{\sigma^2_{rn}(t,\tau) - \sigma^2_p(t,\tau)}{\sigma^2_p(t,\tau)} \approx -\gamma\left(\sigma^2_p(t,\tau)\right)^{1/2} \times \theta_p(t,\tau) + \frac{\gamma^2}{2}\left(\sigma^2_p(t,\tau)\right) \times \left(\kappa_p(t,\tau) - 3\right). \qquad (25)$$

The divergence between risk-neutral volatility and physical volatility can be attributed to (i) exposure to tail events, (ii) fatter left-tails of the physical distribution, and (iii) the risk averse behavior of the investors.

Proof: See the Appendix.

At a basic level Theorem 1 argues that the departure of risk-neutral index volatility from physical index volatility is related to non-zero risk aversion. Pricing kernels displaying $m[R] = 1$ (i.e., $\mathcal{A}_1 = 0$ and $\mathcal{A}_2 = 0$) will have $\sigma^2_{rn}(t,\tau) = \sigma^2_p(t,\tau)$ regardless of the distributional properties of the market index. The fact that option volatilities and realized volatilities are at odds is indicative of economies that display aversion to downside risk.

The analytical expressions developed in Theorem 1 rely on approximate results that are based on ignoring all expansion terms of order higher than $O\left[\gamma^3\right]$. These expansion terms arise as we have restricted the shape of $m[R]$ but not $p[R]$. To clarify this feature of (24) and (25) suppose the market return is distributed Gaussian with $p[R] = \frac{1}{\sqrt{2\upsilon\sigma_p^2}} \exp\left(-\frac{(R-\mu_p)^2}{2\sigma_p^2}\right)$, and the pricing kernel is $m[R] = e^{-\gamma R}$. The risk-neutral density can be accordingly derived in *exact* form as:

$$q[R] = \frac{\exp(-\gamma R) \times \exp\left(-\frac{(R-\mu_p)^2}{2\sigma_p^2}\right)}{\int_{-\infty}^{+\infty} \exp(-\gamma R) \times \exp\left(-\frac{(R-\mu_p)^2}{2\sigma_p^2}\right) dR} = \frac{\hat{c}_0}{\sqrt{2\upsilon\sigma_p^2}} \exp\left(-\frac{(R-(\mu_p - \gamma\,\sigma_p^2))^2}{2\,\sigma_p^2}\right), \quad (26)$$

which, for some normalization constant $\hat{c}_0$, implies that the risk-neutral return distribution is mean-shifted version of the Gaussian physical distribution with the same volatility. The lack of exaggeration of risk-neutral volatility with Gaussian physical distribution agrees with the approximate results in Theorem 1 as $\theta_p = 0$ and $\kappa_p = 3$. Thus, Gaussian physical distributions have counterfactual implications for volatility spreads with power utility pricing kernels. An economic interpretation is that unpleasant extremes in a Gaussian world are sufficiently rare to counteract exponential risk aversion in investors.

Equations (24) and (25) both show that $\frac{\sigma_{rn}^2(t,\tau)-\sigma_p^2(t,\tau)}{\sigma_p^2(t,\tau)}$ can be decomposed into parts related to tail asymmetry and tail size of the physical index distribution. Specifically the volatility spread $\frac{\sigma_{rn}^2(t,\tau)-\sigma_p^2(t,\tau)}{\sigma_p^2(t,\tau)}$ is greatest when $p[R]$ is negatively-skewed and fat-tailed. More precisely according to our characterizations under power marginal utility, the physical skewness receives a weight equal to -$\gamma\ \left(\sigma_p^2(t,\tau)\right)^{1/2}$ while the physical excess kurtosis receives the weight $\frac{\gamma^2}{2}\left(\sigma_p^2(t,\tau)\right)$. These results are qualitatively consistent with the elementary but exact state-space model of Section 3. Since the U.S. post-war $p[R]$ are mildly left-skewed and substantially fat-tailed, risk aversion and excess kurtosis are the likely driving force behind the variations in volatility spreads in power utility economies. For example, Engle (2004) reports a value of -1.44 (-0.10) over the daily sample of 1963-2003 (1990-2003) for the S&P 500 index, and concludes that the sample skewness is small (page 411). The estimates of statistical skewness summarized in Table I, from Andersen, Bollerslev, Diebold, and Ebens (2001), Bollerslev and Zhou (2005), Brandt and Kang (2004), and Brock, Lakonishok, and Lebaron (1992), all point to physical distributions that are essentially symmetric.

The following cases consider some richer parameterizations of investor behavior and show the robustness of Theorem 1. Overall, the analytical results are geared towards understanding the combinations of physical skewness, physical kurtosis, and risk aversion that give rise to non-inverted volatility spreads. A particular observation to be made is that when risk aversion is balanced with respect to both extremes, the effects of physical skewness are mitigated and physical kurtosis remains the predominant force behind volatility spreads. The decomposition of volatility spreads into contributions of physical skewness and kurtosis remains an empirical issue, a topic to be addressed in the ensuing discussions.

CASE 1 *Consider $m[R] = e^{-\gamma R} + e^{\gamma R}$, which can be thought as a hyperbolic cosine pricing kernel representation. This type of $m[R] = \sum_j w_j U_j'[R]$ captures the idea that the marginal utility of the representative agent consists of equally weighted marginal utilities of the investor long and short the market index (e.g., Wang, 1996). Both class of investors share the same risk aversion γ. Application of Theorem 1 implies $\mathcal{A}_1 \equiv -\frac{\partial m}{\partial R} \mid_{R=0} = 0$ and $\mathcal{A}_2 \equiv \frac{\partial^2 m}{\partial R^2} \mid_{R=0} = 2\gamma^2$. Inserting these conditions into equation (24) will lead to physical kurtosis being the only determinant of volatility spreads, as in:*

$$\frac{\sigma_{rn}^2(t,\tau) - \sigma_p^2(t,\tau)}{\sigma_p^2(t,\tau)} \approx \gamma^2 \left(\sigma_p^2(t,\tau)\right) \times \left(\kappa_p(t,\tau) - 1\right), \tag{27}$$

since the weight on physical skewness becomes zero. Equation (27) offers the interpretation that volatility spreads are endogenously determined from physical kurtosis with the effect of skewness vanishing in the aggregate. We can make two further observations:

- *For this class of $m[R]$, the divergence between risk-neutral and physical volatility is positive even under Gaussian physical density. In this case the result may be computed exactly, as $q[R]$ is a mixture of two Gaussian and is consistent with (27).*

- *If long and short positions in the market index are weighted asymmetrically or if their risk aversions are unequal as in $m[R] = w_l \exp(-\gamma_l R) + w_s \exp(\gamma_s R)$, then the asymmetry of the physical distribution will impact the evolution of volatility spreads. The results of this case mirror the predictions of the model in Section 3.*

CASE 2 *Focus now on the implications of translog pricing kernels, that lead to the follow-*

ing relationship between the risk-neutral and physical densities:

$$q[R] = \mho_0 \left(\Delta_a + \Delta_b R + \Delta_c R^2\right) p[R], \tag{28}$$

$$\Delta_c > 0, \quad \textit{and} \quad 4\,\Delta_a \Delta_c > \Delta_b^2, \tag{29}$$

where the inequalities on the coefficients ensure the positivity of the likelihood ratio, and $\mho_0$ *is a normalizing constant. Economically the U-shaped measure change argues for a positive weighting to be given to the short side of the market. One may view* Δ_b *as a compensation for risk as seen as variance, while* Δ_c *is the compensation demanded for negative skewness when it is present. Suppose we mean-shift and set the risk-neutral mean to zero.*

$$\sigma_{rn}^2(t,\tau) = \mho_0 \Delta_a \sigma_p^2(t,\tau) + \mho_0 \Delta_b \left(\sigma_p^2(t,\tau)\right)^{3/2} \theta_p(t,\tau) + \mho_0\,\Delta_c \left(\sigma_p^2(t,\tau)\right)^2 \kappa_p(t,\tau), \tag{30}$$

or, the ratio of the standard deviation can be written as,

$$\frac{\sigma_{rn}(t,\tau)}{\sigma_p(t,\tau)} = \sqrt{\mho_0\,\Delta_a + \mho_0\,\Delta_b\,\sigma_p(t,\tau)\,\theta_p(t,\tau) + \mho_0\,\Delta_c\,\sigma_p^2(t,\tau)\,\kappa_p(t,\tau)}. \tag{31}$$

When $\Delta_b = 0$ *and the statistical process is not skewed then* $\mho_0 \equiv 1/(\Delta_a + \Delta_c\,\sigma_p^2(t,\tau))$ *and hence* $\frac{\sigma_{rn}(t,\tau)}{\sigma_p(t,\tau)} = \sqrt{\frac{\Delta_a}{\Delta_a + \Delta_c\,\sigma_p^2(t,\tau)} + \frac{\Delta_c\,\sigma_p^2(t,\tau)}{\Delta_a + \Delta_c\,\sigma_p^2(t,\tau)}\,\kappa_p(t,\tau)}$. *The volatility spreads are consistently positive in our generalization involving U-shaped measure changes.*

CASE 3 *Alter the pricing kernel,* $m[R]$, *to* $\int_0^\infty e^{-z\,R}\,\nu[dz]$ *for some measure* ν *on* $\Re^+$. *The pricing kernel of this case encompasses HARA marginal utility class and loss aversion utilities (Kahneman, and Tversky, 1979). For example, using the Gamma density with* $\nu[dz] = \frac{1}{\Gamma[\gamma] b^\gamma}\, z^{\gamma-1} e^{-z/b} dz$ *and its Fourier transform, we get* $m[R] = (1 + b\,R)^{-\gamma}$ *which represents HARA marginal utility. Define a* ϕ*-approximation as* $\int_0^\infty e^{-\phi z\,R}\,\nu[dz]$ *which implies that the risk aversion is proportional to* $\phi \int_0^\infty z\,\nu[dz]$. *For a general p[R], we arrive at the expression for volatility spreads below:*

$$\frac{\sigma_{rn}^2(t,\tau) - \sigma_p^2(t,\tau)}{\sigma_p^2(t,\tau)} \approx -\left(\phi \int_0^\infty z\,\nu[dz]\right) \left(\sigma_p^2(t,\tau)\right)^{1/2} \times \theta_p(t,\tau) + \frac{1}{2}\left(\phi^2 \int_0^\infty z^2\,\nu[dz]\right) \left(\sigma_p^2(t,\tau)\right) \times \left(\kappa_p(t,\tau) - 3\right). \tag{32}$$

Except for the weighting on physical skewness and kurtosis the relation (32) inherits a *structure similar to that in Theorem 1. The spread between* $\sigma_{rn}^2(t,\tau)$ *and* $\sigma_p^2(t,\tau)$ *is again a consequence of negative physical skewness and fatter physical tails.*

If the physical moments are stationary through time, then variations in the volatility spread can only be explained by variations in risk aversion. Cases 1 through 3 all imply that negative skewness and excess kurtosis translate into higher risk-neutral volatility relative to physical volatility in the presence of time-varying risk aversion. In Case 1 the absolute risk aversion evaluated at mean return is $\gamma\left(\frac{e^{-\gamma\,\mu_p}-e^{\gamma\,\mu_p}}{e^{-\gamma\,\mu_p}+e^{\gamma\,\mu_p}}\right)$ and in the HARA version of Case 3 it is $\gamma/(1+b\,\mu_p)$.

Drawing from the research of Bollerslev, Gibson, and Zhou (2005), Carr and Wu (2004), and Jiang and Tian (2005), we can similarly shed light on the empirical determinants of volatility spreads without imposing parametric assumptions on the dynamics of market volatility and the volatility risk premium. The representation (24) indicates that when physical index distributions are left-skewed and fat-tailed, the volatility spread is positive provided the risk aversion coefficient is non-zero. In fact, the volatility characterization (25) relying on power utility assumption can be thought as an orthogonality condition which is potentially testable. The novelty of this estimation strategy is that it permits statistical inference on risk aversion solely based on the time-series of risk-neutral volatility, physical volatility, and physical skewness and kurtosis.

5 Empirical Characteristics of Volatility Spreads

For the remainder of this article, S&P 100 index options and returns are employed in our empirical study of volatility spreads. S&P 100 index option (ticker OEX) data is taken from the *Wall Street Journal* over the period January 1984 through December 1999. The S&P 100 index prices are available daily and obtained from Dow Jones Interactive.

To maintain the scope of the investigation, a monthly sampling procedure is followed to construct volatility spreads. Specifically, we select sampling days for OEX by moving backward 28 calendar days from each expiration date. This procedure provides calls and puts with maturity of roughly 28 days. Monthly sampled option observations result in nonoverlapping time periods for successive expiration cycles and mitigates autocorrelation problem. The monthly sample facilitates comparison with the related findings in Bollerslev and Zhou (2005), Bollerslev, Gibson, and Zhou (2005), Canina and Figlewski (1993), Carr and Wu (2004), Lamoureux and Lastrapes (1993), and Jiang and Tian (2005).

Recent literature has shown that the risk-neutral volatility, $\sigma^2_{rn}(t,\tau)$, can be inferred from the prices of out-of-money calls and puts observed in options market. Tapping the model-free approach formalized in Bakshi and Madan (2000), Britten-Jones and Neuberger (2000), Carr and Madan (2001), Carr and Wu (2004), and Jiang and Tian (2005), we write:

$$\sigma^2_{rn}(t,\tau) \;=\; \int_{\Re} R^2\, q[R]\, dR - \left(\int_{\Re} R\, q[R]\, dR\right)^2, \tag{33}$$

where

$$\int_{\Re} R^2 q[R] dR = e^{r\tau} \int_{S(t)}^{\infty} \frac{2\left(1-\ln\left(\frac{K}{S(t)}\right)\right)}{K^2} C[K]\, dK + e^{r\tau}\int_0^{S(t)} \frac{2\left(1+\ln\left(\frac{S(t)}{K}\right)\right)}{K^2} P[K]\, dK, \tag{34}$$

and

$$\int_{\Re} R\, q[R]\, dR = e^{r\tau} - 1 - e^{r\tau}\left(\int_0^{S(t)} \frac{1}{K^2} P\,[K]\, dK + \int_{S(t)}^{\infty} \frac{1}{K^2} C\,[K]\, dK\right). \tag{35}$$

In (34) and (35), $C[K]$ and $P[K]$ respectively represent the price of the call option and the put option with strike price K and τ-periods to expiration, and r is the interest rate. Equation (34) is a consequence of spanning and pricing the payoff $(\ln(S\,(t+\tau)\,/S(t)))^2$.

The computation of risk-neutral volatility requires options with constant maturity. Fixing τ as 28 days, consider the Riemann integral approximation of $\int_{\Re} R^2\, q[R]\, dR$ in (34). First, discretize the integral for the long position in calls as:

$$\sum_{j=1}^{\mathcal{J}-1} \left(N\,[j-1] + N\,[j]\right) \frac{\Delta K}{2}, \tag{36}$$

where $N[j] \equiv z[K_{\max} - j\Delta K] \times C[K_{\max} - j\Delta K]$, $K_{\max}$ is the maximum level of the strike price, $\mathcal{J}$ is the number of call/put options, and $z[K] \equiv \frac{2}{K^2}\left(1-\ln\left(\frac{K}{S(t)}\right)\right)$. Second, the integral for the long position in puts can be discretized as:

$$\sum_{j=1}^{\mathcal{J}-1} \left(M\,[j-1] + M\,[j]\right) \frac{\Delta K}{2}, \tag{37}$$

where $M[j] \equiv z[K_{\min} + j\Delta K] \times P[K_{\min} + j\Delta K]$ and $K_{\min}$ represents the minimum level of the strike price. According to Table 1 in Jiang and Tian (2005), the implementation with finite number of options works reasonably well in practice with small approximation errors. The risk-neutral volatility calculation uses an input for r that is proxied by the three-month treasury bill rate. The resulting $\sigma_{rn}^2(t,\tau)$ is annualized by scaling the 28-day volatility estimate by a factor of 360/28.

Relying on Merton (1980) and Andersen, Bollerslev, Diebold, and Labys (2003), the realized physical (annualized) volatility is also computed in a model-free manner, using daily OEX index returns, as:

$$\sigma_p(t) \equiv \sqrt{\frac{360}{28}\sum_{\ell=1}^{28}\left(R(t+\ell)-\overline{R}\right)^2} \tag{38}$$

where $R(t+\ell)$ is the daily return and $\overline{R}$ represents the rolling mean during this period.

Table 2 presents four statistics: (i) the risk-neutral volatility, (ii) the realized physical volatility, (iii) the percentage volatility spread as $\frac{\sigma_{rn}(t)-\sigma_p(t)}{\sigma_p(t)}$, and (iv) the fraction of months in which risk-neutral volatility exceeds realized volatility, labeled $\mathcal{Z}_T$.

The results support our fundamental claim that volatility spreads are systematically and substantially positive. When averaged over the entire sample, the risk-neutral index volatility is 16.29% and the realized physical index volatility is 13.98%. The average relative volatility spread is 23.45%, with risk-neutral volatility higher than physical volatility in 74% of the months. The volatility spread is statistically significant with a t-statistic of 9.41. The 84:01-91:12 and 92:01-99:12 sub-samples reveals a similar picture of positive volatility spreads. The reported volatility spread estimates are roughly in line with the corresponding estimates in Bollerslev, Gibson, and Zhou (2005), Carr and Wu (2004), and Jiang and Tian (2005). For instance, over the sample of January 1990 to May 2004, Bollerslev, Gibson, and Zhou (2005) state that the average risk-neutral volatility for S&P 500 index is 20.08% and the realized volatility is 12.68%, implying large volatility spreads. Jiang and Tian (2005) present evidence on positive volatility spreads for maturities up to 180 days.

Analyzing Table 2 further, we can see that positive volatility spreads are mostly a post-crash phenomenon. During the pre-crash period of 84:01-87:09, $\frac{\sigma_{rn}(t)-\sigma_p(t)}{\sigma_p(t)}$ is 0.47% with a t-statistic of 0.13, while over the post-crash period the corresponding average volatility spread is 30.49% with a t-statistic of 10.85. Our theory of volatility spreads provides an explanation for the findings in Jackwerth and Rubinstein (1996) who argue that the shape of the risk-neutral index distribution has changed after the stock market crash of 1987. In terms of the example of Section 3, the change in the risk-neutral index density noted in

Jackwerth and Rubinstein (1996) is likely related to higher exposure to tail events (i.e., υ and ρ), risk aversion (i.e., α), and fatter left-tails of the physical index distribution (i.e., β).

6 Reconciling the Time-Series of Volatility Spreads

Suggestive from Bollerslev, Gibson, and Zhou (2005), and Jiang and Tian (2005), we can examine the theoretical restrictions on volatility spreads emerging from (25) of Theorem 1. This equation provides a different perspective in evaluating risk aversion in financial markets. Provisionally treating (25) as exact, define, the error term, $\epsilon(t+1)$, from Theorem 1, as:

$$\begin{aligned}\epsilon(t+1) \;&\equiv\; \frac{\sigma_{rn}^2(t+1) - \sigma_p^2(t+1)}{\sigma_p^2(t+1)} + \gamma \left(\sigma_p^2(t+1)\right)^{1/2} \times \theta_p(t+1) \\ &\quad - \frac{\gamma^2}{2} \left(\sigma_p^2(t+1)\right) \times \left(\kappa_p(t+1) - 3\right), \qquad (39)\end{aligned}$$

where $\theta_p(t+1)$ and $\kappa_p(t+1)$ are respectively the t+1-conditional skewness and kurtosis of the physical index distribution. This characterization suggests that volatility spreads are due to the existence of risk aversion and fatter left-tails of the physical index distribution. The restriction that this theory imposes on the evolution of the risk-neutral volatility and risk aversion γ is testable using generalized method of moments (Hansen (1982)).

Relying on the orthogonality of $\epsilon(t+1)$ with respect to time-t information variables $Z(t)$, we have $E\{\epsilon(t+1)|Z(t)\} = 0$, where $E\{.\}$ denotes the unconditional expectation operator. That is, by a standard argument, $E\{\epsilon(t+1) \otimes Z(t)\} = 0$. If the theoretical linkages between physical and risk-neutral densities are correct, then

$$g_T[\gamma] \equiv \frac{1}{T} \sum_{t=1}^{T} \epsilon(t+1) \otimes Z(t) \qquad (40)$$

should be sufficiently close to zero as the sample size, T, increases. The GMM estimator of γ is based on minimizing the criterion function:

$$J_T \equiv \arg \min_{\gamma} \; g_T' \, W_T \, g_T \qquad (41)$$

where W_T is a symmetric, weighting matrix. Hansen has shown that $T \times J_T$ is asymptotically

χ^2-distributed, with degrees of freedom equal to the difference between the number of orthogonality conditions, L, and the number of parameters to be estimated.

A high value of J_T implies that the risk-neutral volatility representation (25) may be misspecified and, in particular, may contain a constant term representing a summary of other effects. The source of misspecification can be alternative class of pricing kernels or the relevance of physical moments beyond four for volatility spreads. A related approach in Bliss and Panigirtzoglou (2004) relies on constructing the risk-neutral density from the cross-section of options and then selecting the risk aversion coefficient to maximize the forecasting ability of the resulting physical density, which can be computationally intensive. The virtue of the GMM procedure, in the present context of estimating γ, is that it requires simple inputs of risk-neutral volatility, physical volatility, skewness, and kurtosis.

Realize that although physical volatility can be estimated with a reasonable degree of confidence, the estimate of physical skewness and kurtosis warrants a reasonably long return-series. If a relatively short window (i.e., 28 days) is selected to estimate higher moments, the skewness and kurtosis are likely to be underestimated (Jackwerth and Rubinstein, 1996). Failing to properly account for the length of the window can result in the rejection of (39) even when Theorem 1 provides a reasonable characterization of volatility spreads. To minimize the biases associated with the underestimation of skewness and kurtosis in the GMM estimation, the physical skewness and kurtosis are estimated using OEX returns trailed by 90 days and 200 days, respectively. The skewness and kurtosis are updated every 90 days and matched with the corresponding estimate of risk-neutral and physical volatility.

Three sets of instrumental variables are adopted for robustness: SET 1 contains a constant and $\sigma^2_{rn}(t)$; SET 2 contains a constant, $\sigma^2_{rn}(t)$, and $\sigma^2_{rn}(t-1)$; Finally, SET 3 contains a constant, $\sigma^2_{rn}(t)$, $\sigma^2_{rn}(t-1)$, and $\sigma^2_{rn}(t-2)$. Each choice of instrumental variables ensures that the number of orthogonality conditions do not get too large relative to the sample size.

Parameter estimates for γ reported in Table 3 are based on a covariance matrix that accommodates conditional heteroskedasticity and autocorrelation. The lag in the moving average is set to 4, which is roughly the cube root of the number of observations.

Not at odds with the theoretical predictions, the overidentifying restrictions afforded by the modeling structure are not rejected. In addition, as theory suggests, the estimate of γ are positive and statistically significant and stable across instrumental variables. The minimum t-statistic on γ is 4.12. Taken all together, the estimates of risk aversion are comparable to that in Aït-Sahalia and Lo (2000) who find that the weighted average of relative risk aversion is 12.70 using S&P 500 index options. With S&P 500 futures options and power utility, Bliss and Panigirtzoglou (2004) report γ of 9.52 (5.38) for one (two) week forecasting horizon. Finally, the Euler equation approach in Aït-Sahalia, Parker, and Yogo (2004) yields a point estimate of $\gamma = 7$ when luxury retail series is adopted to surrogate aggregate consumption.

When the window used in the estimation of skewness and kurtosis is altered from 90 days to 200 days, both the goodness-of-fit measure J_T and the estimate of risk aversion fall. More exactly, the estimate of γ decreases from about 17 to about 12.8. This is to be expected since the use of longer time-series will make the estimate of higher moments closer to their theoretical counterparts. For instance, with 90 days, the (annualized) sample θ_p is -0.0142 and κ_p is 3.0065. On the other hand, with 200 days history, the θ_p and κ_p are respectively -0.0267 and 3.0144. Our results are consistent with the view that volatility spreads are a consequence of risk aversion and fatter left-tails.

Specialized GMM estimations can be performed to study the relative importance of skewness and kurtosis in fitting the evolution of volatility spreads. In the first estimation, we omit a role for fat-tailed physical distributions by setting $\kappa_p(t+1) = 3$ in (39). When physical skewness is the only source of volatility spreads, we obtain:

$$\epsilon_\theta(t+1) \equiv \frac{\sigma_{rn}^2(t+1) - \sigma_p^2(t+1)}{\sigma_p^2(t+1)} + \gamma \left(\sigma_p^2(t+1)\right)^{1/2} \times \theta_p(t+1). \tag{42}$$

The second estimation considers the impact of symmetric physical distributions on volatility spreads by setting $\theta_p(t+1) = 0$ in (39). When excess kurtosis is the only source of volatility spreads, the volatility equation specializes to:

$$\epsilon_\kappa(t+1) \equiv \frac{\sigma_{rn}^2(t+1) - \sigma_p^2(t+1)}{\sigma_p^2(t+1)} - \frac{\gamma^2}{2} \left(\sigma_p^2(t+1)\right) \times \left(\kappa_p(t+1) - 3\right). \tag{43}$$

Two points can be made based on the evidence presented in Table 3. First observe that the overidentifying restrictions imposed by the theory are not rejected in either of the specialized estimations. This evidence can be interpreted as stating that the presence of excess kurtosis does not swamp out the effect of skewness and vice versa. Therefore, both physical skewness and kurtosis can be regarded as equally important sources of volatility spreads. The second point concerns the estimated level of risk aversion: the estimate of γ when physical skewness is constrained to zero is 50% of the corresponding value when kurtosis is constrained to 3. A possible economic interpretation is that the zero excess kurtosis restriction is too stringent, meaning that implausible levels of γ are needed to reconcile volatility spread dynamics when only physical skewness is allowed to impact risk-neutral volatility. Viewed from this angle, the fat-tailed feature may be instrumental to producing more realistic volatility spread levels.

7 Conclusions

This paper theoretically links risk-neutral index volatility to the higher-order moments of the physical index distribution and to the parameters of the pricing kernel process. Specifically we establish that the spread between risk-neutral and physical volatility is more pronounced when the physical index distribution is leptokurtic and left-skewed. Our framework is sufficiently general and encompasses a broad class of pricing kernels and market return dynamics.

Consistent with our theory, the empirical investigation finds that the risk-neutral distribution embedded in S&P 100 index options is more volatile than its physical counterpart. Overall, the estimation approach supports the view that the observed divergence between the risk-neutral index volatility and the physical volatility is systematically related to the higher-order moments of the physical index density, and to risk aversion.

Among the various possible extensions, the theoretical and empirical framework can refined to pin down the sources of the volatility risk premium (e.g., Bakshi and Kapadia, 2003, Bollerslev, Gibson, and Zhou, 2005, and Carr and Wu, 2004) and for devising methods to hedge volatility risks (e.g., Brenner, Ou, and Zhang, 2005) in not only equity markets, but other markets.

Appendix

Proof of Theorem 1: Suppress time-dependence and set the physical first-moment $\mu_p = 0$. Consider the moment generating function of the physical return distribution:

$$\overline{\mathcal{C}}[\lambda] \equiv \int_{-\infty}^{\infty} e^{\lambda R}\, p[R]\, dR \;\;=\;\; 1 + \frac{\lambda^2}{2}\,\sigma_p^2 + \frac{\lambda^3}{6}\,\theta_p\,(\sigma_p^2)^{3/2} + \frac{\lambda^4}{24}\,\kappa_p\,(\sigma_p^2)^2 + o[\lambda^4]. \tag{44}$$

Based on (10) and assumed $m[R] = e^{-\gamma R}$, the moment generating function of the risk-neutral distribution is,

$$\mathcal{C}[\lambda] \;\;\equiv\;\; \int_{-\infty}^{\infty} e^{\lambda R}\, q[R]\, dR = \frac{\int_{-\infty}^{\infty} e^{\lambda R}\, e^{-\gamma\, R}\, p[R]\, dR}{\int_{-\infty}^{\infty} e^{-\gamma\, R}\, p[R]\, dR} = \frac{\overline{\mathcal{C}}[\lambda-\gamma]}{\overline{\mathcal{C}}[-\gamma]}, \tag{45}$$

where,

$$\overline{\mathcal{C}}[\lambda-\gamma] \;\;=\;\; 1 + \frac{(\lambda-\gamma)^2}{2}\,\sigma_p^2 + \frac{(\lambda-\gamma)^3}{6}\,\theta_p\,(\sigma_p^2)^{3/2} + \frac{(\lambda-\gamma)^4}{24}\,\kappa_p\,(\sigma_p^2)^2 + o[\lambda^4], \tag{46}$$

$$\overline{\mathcal{C}}[-\gamma] \;\;=\;\; 1 + \frac{(-\gamma)^2}{2}\,\sigma_p^2 + \frac{(-\gamma)^3}{6}\,\theta_p\,(\sigma_p^2)^{3/2} + \frac{(-\gamma)^4}{24}\,\kappa_p\,(\sigma_p^2)^2 + o[\lambda^4]. \tag{47}$$

Therefore, the first risk-neutral moment is,

$$\int_{-\infty}^{\infty} R\, q[R]\, dR \;\;=\;\; \frac{\overline{\mathcal{C}}'[\lambda-\gamma]\,|_{\lambda=0}}{\overline{\mathcal{C}}[-\gamma]}, \tag{48}$$

$$=\;\; \left(1 + \frac{\gamma^2}{2}\sigma_p^2 + \ldots\right)^{-1} \left(-\gamma\,\sigma_p^2 + \frac{\gamma^2}{2}\,\theta_p\,(\sigma_p^2)^{3/2} - \frac{\gamma^3}{6}\,\kappa_p\,(\sigma_p^2)^2 + \ldots\right), \tag{49}$$

$$=\;\; -\gamma\,\sigma_p^2 + \frac{\gamma^2}{2}\,\theta_p\,(\sigma_p^2)^{3/2} + O\left[\gamma^3\right], \tag{50}$$

and the second uncentered risk-neutral moment satisfies,

$$\int_{-\infty}^{\infty} R^2\, q[R]\, dR \;\;=\;\; \frac{\overline{\mathcal{C}}''[\lambda-\gamma]\,|_{\lambda=0}}{\overline{\mathcal{C}}[-\gamma]}, \tag{51}$$

$$=\;\; \left(1 + \frac{\gamma^2}{2}\sigma_p^2 + \ldots\right)^{-1} \left(\sigma_p^2 - \gamma\,\theta_p\,(\sigma_p^2)^{3/2} + \frac{\gamma^2}{2}\,\kappa_p\,(\sigma_p^2)^2 + \ldots\right), \tag{52}$$

$$=\;\; \sigma_p^2 - \gamma\,\theta_p\,(\sigma_p^2)^{3/2} + \frac{\gamma^2}{2}(\sigma_p^2)^2\,\left(\kappa_p - 1\right) + O\left[\gamma^3\right]. \tag{53}$$

Thus, we arrive at the second-order approximation in γ for risk-neutral volatility as,

$$\begin{aligned}\sigma_{rn}^2 &= \int_{-\infty}^{\infty} R^2\, q[R]\, dR - \left(\int_{-\infty}^{\infty} R\, q[R]\, dR\right)^2 \\ &\approx \sigma_p^2 - \gamma\, \theta_p\, (\sigma_p^2)^{3/2} + \frac{\gamma^2}{2}(\sigma_p^2)^2\ (\kappa_p - 1) - \gamma^2\, (\sigma_p^2)^2. \end{aligned} \tag{54}$$

Finally, we may write (25) of Theorem 1 as below:

$$\frac{\sigma_{rn}^2 - \sigma_p^2}{\sigma_p^2} \approx -\gamma\, \theta_p\, (\sigma_p^2)^{1/2} + \frac{\gamma^2}{2}\, \sigma_p^2\ (\kappa_p - 3)\,. \tag{55}$$

The first part of Theorem 1 is a modification of the above steps. Taking a Taylor series expansion of $m[R]$ around zero, we get: $m[R] \approx 1 - \mathcal{A}_1\, R + \frac{1}{2}\, \mathcal{A}_2\, R^2 + \ldots$, where $m[0] = 1$, $\mathcal{A}_1 \equiv -\frac{\partial m}{\partial R}\, |_{R=0}$, $\mathcal{A}_2 \equiv \frac{\partial^2 m}{\partial R^2}\, |_{R=0}$, and so on. Therefore, the risk-neutral density is,

$$q[R] = \frac{(1 - \mathcal{A}_1\, R + \frac{1}{2}\, \mathcal{A}_2\, R^2 + \ldots)\, p[R]}{1 + \frac{1}{2}\, \mathcal{A}_2\, \sigma_p^2 + \ldots}, \tag{56}$$

since $\int_{\Re} m[R]\, p[R]\, dR = 1 + \frac{1}{2}\, \mathcal{A}_2\, \sigma_p^2 + \ldots$, and thus,

$$\int_{-\infty}^{\infty} R\, q[R]\, dR \approx \left(1 - \frac{1}{2}\, \mathcal{A}_2\, \sigma_p^2\right) \left(-\mathcal{A}_1\, \sigma_p^2 + \frac{1}{2}\, \mathcal{A}_2\, \theta_p (\sigma_p^2)^{3/2}\right), \tag{57}$$

$$\int_{-\infty}^{\infty} R^2\, q[R]\, dR \approx \left(1 - \frac{1}{2}\, \mathcal{A}_2\, \sigma_p^2\right) \left(\sigma_p^2 - \mathcal{A}_1\, \theta_p (\sigma_p^2)^{3/2} + \frac{1}{2}\, \mathcal{A}_2\, \kappa_p (\sigma_p^2)^2\right). \tag{58}$$

Ignoring expansion terms higher than $\mathcal{A}_2$ in our calculation, we obtain the volatility characterization in (24). $\square$

References

Aït-Sahalia, Y., A. Lo, 2000. Nonparametric risk management and implied risk aversion. *Journal of Econometrics* 94, 9-51.

Aït-Sahalia, Y., J. Parker, M. Yogo, 2004. Luxury goods and the equity premium. *Journal of Finance* 59 (6), 2959-3004.

Andersen, T., T. Bollerslev, F. Diebold, H. Ebens, 2001. The distribution of realized stock return volatility. *Journal of Financial Economics* 61, 43-76.

Andersen, T., T. Bollerslev, F. Diebold, P. Labys, 2003. Modeling and forecasting realized volatility. *Econometrica* 71, No. 2, 579-625.

Bakshi, G., C. Cao, Z. Chen, 2000. Pricing and hedging long-term options. *Journal of Econometrics* 94, 277-318.

Bakshi, G., N. Kapadia, D. Madan, 2003. Stock return characteristics, skew laws, and the differential pricing of individual equity options. *Review of Financial Studies* 16, 101-143.

Bakshi, G., N. Kapadia, 2003. Delta hedged gains and the negative market volatility risk premium. *Review of Financial Studies* 16, 527-566.

Bakshi, G., D. Madan, 2000. Spanning and derivative security valuation. *Journal of Financial Economics* 55, 205-238.

Bliss, R., N. Panigirtzoglou, 2004. Option-implied risk aversion estimates. *Journal of Finance* 59 (1), 407-446.

Bollerslev, T., H. Zhou, 2005. Volatility puzzles: a simple framework for gauging return-volatility regressions. *Journal of Econometrics* (forthcoming).

Bollerslev, T., M. Gibson, H. Zhou, 2005. Dynamic estimation of volatility risk premia and investor risk aversion from option-implied and realized volatilities. *Mimeo*, Duke University and Federal Reserve Board.

Brandt, M., Q. Kang, 2004. On the relationship between the conditional mean and volatility of stock returns: a latent VAR approach. *Journal of Financial Economics* 72, 217-257.

Brenner, M., E. Ou, J. Zhang, 2005. Hedging volatility risk. *Journal of Banking and Finance* (forthcoming).

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

Brock, W., J. Lakonishok, B. LeBaron, 1992. Simple technical trading rules and the stochastic properties of stock returns. *Journal of Finance* 47 (5), 1731-1764.

Canina, L., S. Figlewski, 1993. The information content of implied volatilities. *Review of Financial Studies* 6, 659-681.

Carr, P., D. Madan, 2001. Optimal positioning in derivative securities. *Quantitative Finance* 1, 19-37.

Carr, P., L. Wu, 2004. Variance risk premia. *Mimeo*, Baruch College.

Constantinides, G., 2002. Rational asset prices. *Journal of Finance* 57 (4), 1567-1591.

Christoffersen, P., S. Heston, K. Jacobs, 2005. Option valuation with conditional skewness. *Journal of Econometrics* (forthcoming).

Engle, R., 2004. Risk and volatility: econometric models and financial practice. *American Economic Review* 94, No. 3, 405-420.

Halmos, P., 1974, Measure Theory. Springer Verlag. New York.

Hansen, L., 1982. Large sample properties of generalized method of moments estimators. *Econometrica* 50, 1029-1084.

Harrison, M., D. Kreps, 1979. Martingales and arbitrage in multiperiod securities markets. *Journal of Economic Theory* 20, 381-408.

Jackwerth, J., M. Rubinstein, 1996. Recovering probability distributions from option prices. *Journal of Finance* 51, 1611-1631.

Jiang, G., Y. Tian, 2005. The model-free implied volatility and its information content. *Review of Financial Studies* 18, 1305-1342.

Kahneman, D., and A. Tversky, 1979, Prospect theory: an analysis of decision under risk. *Econometrica* 47 (2), 263-292.

Lamoureux, C., W. Lastrapes, 1993. The information content of implied volatilities. *Review of Financial Studies* 6, 659-681.

Merton, R., 1980. On estimating the expected return on the market: an exploratory investigation. *Journal of Financial Economics* 8, 323-361.

Wang, J., 1996. The term structure of interest rates in a pure exchange economy with heterogeneous investors. *Journal of Financial Economics* 41, 75-110.

Table 1: Estimates of Physical Return Skewness from Previous Studies

This table summarizes sample return skewness estimates from previous studies. The DJIA refers to the Dow Jones Industrial Average and 30 DJIA refers to the average skewness across the 30 stocks in the DJIA. The CRSP VWI series represents value-weighted index returns in excess of the one-month Treasury bill rate. The daily skewness number reported in the row "Present Study" is obtained by first calculating the sample skewness based on all daily observations in each year, and then averaging them over 1984-1999. Under the hypothesis of i.i.d. market returns, the daily (monthly) skewness can be annualized by dividing them by a factor of $\sqrt{252}$ ($\sqrt{12}$).

Study	Sample Period	Frequency	Underlying Index	Skewness Level
Engle (2004)	1963-2003	Daily	S&P 500	-1.440
Engle (2004)	1990-2003	Daily	S&P 500	-0.100
Brock et. el. (1992)	1897-1986	Daily	DJIA	-0.104
Andersen et. el. (2001)	1993-1998	Daily	30 DJIA	0.172
Present Study	1984-1999	Daily	S&P 100	-0.558
Brandt and Kang (2004)	1946-1998	Monthly	CRSP VWI	-0.625
Bollerslev and Zhou (2005)	1990-2002	Monthly	S&P 500	-0.621

Table 2: Empirical Estimates of Volatility Spreads

Reported in this table are (i) the average risk-neutral standard deviations (σ_{rn}), (ii) the average physical standard deviations (σ_p), and (iii) the percentage volatility spreads as ($\sigma_{rn}/\sigma_p - 1$). The reported t-statistics are computed as the average estimate divided by its standard error. $\mathcal{Z}_T$ is an indicator function that is assigned a value of 1 if $\sigma_{rn} > \sigma_p$, and is zero otherwise. The risk-neutral volatility and physical realized volatility are as respectively computed in equations (33) and (38). To calculate the risk-neutral volatility, the option sampling days are selected by moving back 28 calendar days from the maturity date. This is to maintain a nonoverlapping risk-neutral volatility series. All numbers are annualized. The sample period is January 1984 through December 1999 (192 observations).

Year	σ_{rn} Avg. (%)	σ_p Avg. (%)	$\sigma_{rn}/\sigma_p - 1$ Avg. (%)	t-stat	$\mathcal{Z}_T$ (%)
1984	12.74	13.85	-5.61	-0.87	50
1985	10.10	10.98	-5.34	-0.84	25
1986	14.76	14.71	3.07	0.51	67
1987	26.03	22.24	17.74	1.69	58
1988	19.82	15.48	31.44	3.77	75
1989	14.23	13.73	11.67	1.33	75
1990	19.19	16.16	27.05	2.78	75
1991	15.15	14.13	13.57	1.69	58
1992	12.65	9.63	35.44	4.61	92
1993	11.25	8.56	39.45	3.47	92
1994	11.87	9.77	26.27	3.36	83
1995	11.19	8.17	44.98	4.35	92
1996	14.87	11.75	33.41	3.55	83
1997	21.30	17.77	31.18	2.94	83
1998	22.97	18.73	43.43	2.71	75
1999	22.59	17.97	27.52	5.48	100
84:01-99:12	16.29	13.98	23.45	9.41	74
84:01-91:12	16.50	15.16	11.70	3.80	60
92:01-99:12	16.09	12.79	35.21	9.92	88
84:01-87:09	13.44	13.88	0.47	0.13	49
87:10-99:12	17.17	14.01	30.49	10.85	82

Table 3: Estimates of Risk Aversion from Volatility Spreads

When $\theta_p(t+1)$ and $\kappa_p(t+1)$ are respectively the t+1-conditional skewness and kurtosis of the physical index distribution, equation (25) of Theorem 1 implies:

$$\epsilon(t+1) \equiv \frac{\sigma_{rn}^2(t+1) - \sigma_p^2(t+1)}{\sigma_p^2(t+1)} + \gamma \left(\sigma_p^2(t+1)\right)^{1/2} \times \theta_p(t+1) - \frac{\gamma^2}{2} \left(\sigma_p^2(t+1)\right) \times (\kappa_p(t+1) - 3)$$

with $E\{\epsilon(t+1) \otimes Z(t)\} = 0$, for a set of information variables $Z(t)$. The risk aversion parameter, γ, is estimated by generalized method of moments (GMM). The degrees of freedom, df, is the number of instruments minus one. The reported p-value is based on the minimized value (multiplied by T) of the GMM criterion function, J_T, and is χ^2-distributed with degrees of freedom, df. Three sets of instrumental variables are considered: SET 1 contains a constant plus $\sigma_{rn}^2(t)$; SET 2 contains a constant, $\sigma_{rn}^2(t)$, and $\sigma_{rn}^2(t-1)$; SET 3 contains a constant, $\sigma_{rn}^2(t)$, $\sigma_{rn}^2(t-1)$, and $\sigma_{rn}^2(t-2)$. In Panel A and Panel B the skewness and kurtosis are calculated from returns lagged by 90 days and 200 days, respectively. The inputs σ_{rn}^2, σ_p^2, θ_p and κ_p are annualized.

		Unrestricted Estimation				**Restricted $\kappa_p = 3$**				**Restricted $\theta_p = 0$**			
Z(t)	df	γ	t(γ)	J_T	p-value	γ	t(γ)	J_T	p-value	γ	t(γ)	J_T	p-value
Panel A: Skewness and Kurtosis Calculated from 90 Day Returns													
SET 1	1	17.22	4.12	1.22	0.27	46.93	2.98	1.08	0.30	21.71	4.89	1.24	0.27
SET 2	2	17.26	4.17	1.22	0.54	46.96	3.04	1.09	0.58	21.76	4.93	1.24	0.54
SET 3	3	17.33	4.36	1.39	0.71	46.20	3.34	1.36	0.72	21.95	5.08	1.38	0.71
Panel B: Skewness and Kurtosis Calculated from 200 Day Returns													
SET 1	1	12.71	4.63	1.16	0.28	32.95	3.72	0.99	0.32	16.27	5.19	1.21	0.27
SET 2	2	12.71	4.65	1.16	0.56	31.83	4.32	1.12	0.57	16.31	5.19	1.21	0.55
SET 3	3	12.94	4.71	1.43	0.70	29.04	5.48	2.30	0.51	16.64	5.07	1.23	0.75