# Capturing Option Anomalies with a Variance-Dependent Pricing Kernel

**Peter Christoffersen**
University of Toronto, Copenhagen Business School, and CREATES

**Steven Heston**
University of Maryland

**Kris Jacobs**
University of Houston and Tilburg University

We develop a GARCH option model with a new pricing kernel allowing for a variance premium. While the pricing kernel is monotonic in the stock return and in variance, its projection onto the stock return is nonmonotonic. A negative variance premium makes it U shaped. We present new semiparametric evidence to confirm this U-shaped relationship between the risk-neutral and physical probability densities. The new pricing kernel substantially improves our ability to reconcile the time-series properties of stock returns with the cross-section of option prices. It provides a unified explanation for the implied volatility puzzle, the overreaction of long-term options to changes in short-term variance, and the fat tails of the risk-neutral return distribution relative to the physical distribution. (*JEL* G12)

The literature on index option valuation has provided some significant improvements to the classical Black-Scholes setup. Most importantly, modeling stochastic volatility and incorporating a leverage effect reduce pricing error (see, for example, Bakshi, Cao, and Chen 1997). However, significant challenges remain. Bates (1996a) observed that "the central empirical issue in option research is whether the distributions implicit in option prices are consistent with the time-series properties of the underlying asset prices." While subsequent studies have addressed this issue, it has proved difficult to reconcile the empirical distributions of spot returns with the risk-neutral

Christoffersen and Jacobs thank IFM2 and SSHRC for financial support. We are grateful to our AFA discussant David Bates, as well as Gurdip Bakshi, Garland Durham, Richard Harris, Chris Jones, Stylianos Perrakis, Pietro Veronesi, an anonymous referee, and participants in seminars at Exeter, HKUST, University of Houston, UC Irvine, Concordia University, Luxembourg School of Finance, University of Southern California, University of Colorado, Boston College, Rice University, UC San Diego, University of Pittsburgh, the HEC Montreal Applied Financial Time Series Workshop, the University of Maryland Conference on Financial Economics and Accounting, and the FGV Conference on Financial Economics in Rio for helpful comments. Kadir Babaoglu, Chayawat Ornthanalai, Mehdi Karoui, and Nick Pan provided expert research assistance. Send correspondence to Steven Heston, R.H. Smith School of Business, University of Maryland, 4447 Van Munching Hall, College Park, MD 20742, USA; telephone (301) 405-9686. E-mail: sheston@rhsmith.umd.edu.


doi:10.1093/rfs/hht033 Advance Access publication June 28, 2013

distributions underlying option prices. This is a particularly acute puzzle, because the fundamental theorem of asset pricing states that absence of arbitrage guarantees the existence of a nonnegative pricing kernel that relates risk-neutral probabilities to true probabilities.

Moreover, several robust stylized facts and empirical puzzles have emerged from the index options literature. It is well known that volatilities implied by option prices tend to exceed realized volatility, which can be explained by a negative price of variance risk. Bakshi and Kapadia (2003) show that average returns on variance-sensitive option portfolios are indeed negative. The expectations puzzle highlights that implied variances do not provide an unbiased forecast of subsequent variance.[1] Stein (1989) and Poteshman (2001) show that long-term implied variance overreacts to changes in short-term variance. This puzzle involves movements in the term structure of implied volatility and is related to the expectations puzzle. Taken together, these anomalies indicate misspecification in the dynamic relationship between option values and the time series of spot returns. They are usually not discussed in the context of a parametric framework, and therefore the literature has typically not explicitly linked them to Bates's statement, but they are intimately related.

In addition to these longitudinal expectations puzzles, available models have difficulty explaining the cross-section of option prices, particularly the prices of out-of-the-money options. Several studies have recognized that this evinces a "pricing kernel puzzle" in the sense that available pricing kernels may not be general enough to explain option data. See, for instance, Brown and Jackwerth (2012), Bates (2008), Chabi-Yo, Garcia, and Renault (2008), and Bakshi, Madan, and Panayotov (2010). We further explore this puzzle by presenting new semiparametric evidence on the conditional pricing kernel. The natural logarithm of the conditional pricing kernel appears to be a U-shaped function of returns, in contrast with the monotonic pattern assumed in existing theory (Rubinstein 1976; Brennan 1979). The U shape is relatively stable over time.

Together, the expectations puzzles, the overreaction puzzle, the pricing kernel puzzle, and the problems in reconciling option prices and underlying returns pose a collective challenge to option valuation. This paper attempts to provide a unified explanation for these puzzles by studying a pricing kernel that is more general than the Rubinstein (1976) and Brennan (1979) pricing kernels, which are monotonic in market returns, and analyzing its implications. We incorporate a variance premium in the pricing kernel. The pricing kernel is monotonic in returns and is also monotonic in volatility. However, because volatility tends to be high when the stock return is large and positive or large and negative, the pricing kernel is nonmonotonic. A negative variance premium makes it a

[1] See Day and Lewis (1992), Canina and Figlewski (1993), Lamoureux and Lastrapes (1993), Jorion (1995), Fleming (1998), Blair, Poon, and Taylor (2001), and Chernov (2007) among others.

U-shaped function of the stock return after projecting the volatility effect onto returns.

We generalize the Heston-Nandi (2000) GARCH model to include this more general pricing kernel.[2] The volatility premium in the new model contains two components: one is related to the equity risk premium, while the other is an independent component that originates in volatility preferences. In this model, the logarithm of the pricing kernel is a quadratic function of the market return, qualitatively accounting for the stylized facts. The pricing kernel also explains the other anomalies in terms of a volatility premium. A negative premium for volatility can explain why implied volatilities are high and average option returns are low. It is important that the magnitude of this (negative) premium grows as volatility rises. This amplifies the movements of long-term option values in response to short-term fluctuations in volatility. The pricing kernel can therefore explain Stein's (1989) observation that long-maturity implied volatilities appear to overreact to temporary increases in volatility.

These theoretical and empirical results indicate that the suggested pricing kernel is able to qualitatively account for a number of important puzzles. In order to demonstrate that these implications are quantitatively important, we conduct an empirical analysis that uses an objective function with a return and an option component. This analysis also investigates whether the new kernel specification is flexible enough to resolve the dichotomy between option prices and the time series of underlying index returns pointed out by Bates (1996a). The discrete-time structure of the model greatly facilitates the filtering problem and the empirical implementation, and we use an option data set that is substantially larger than existing studies. We use an ad hoc benchmark that fits both components of the data separately. We then estimate the model with the new pricing kernel, as well as using the more restrictive traditional pricing kernel.

The empirical results are quite striking. Imposing the new pricing kernel with the variance premium dramatically improves model fit compared with a traditional pricing kernel with equity risk only. The fit of the new model comes much closer to that of the unrestricted ad hoc model. The new pricing kernel adequately captures the volatility premium between the physical and risk-neutral volatility, as well as the differences between the physical and risk-neutral volatility of variance. We subsequently repeat the empirical analyses using the model data. We confirm that the new model is able to quantitatively capture the stable U-shaped pricing kernel, Stein's overreaction hypothesis, and returns on straddles.

A number of existing studies on option valuation and general equilibrium modeling are related to our findings. Several studies have argued that modifications to standard preferences are needed to explain option data. See, for

[2] The GARCH model was developed in Engle (1982) and Bollerslev (1986).

instance, Bates (2008), Pan (2002), Benzoni, Collin-Dufresne, and Goldstein (2011), and Liu, Pan, and Wang (2004). Ait-Sahalia and Lo (2000) and Jackwerth (2000) have noted the surprising implications of option prices for risk aversion, and Shive and Shumway (2006) suggest using nonmonotonic pricing kernels. Rosenberg and Engle (2002) and Chernov (2003) document nonmonotonicities in pricing kernels using parametric assumptions on the underlying returns. Chabi-Yo (2012) uses Taylor series expansions of marginal utilities and documents nonmonotonicities after projecting on the market return. Brown and Jackwerth (2012) argue that in order to explain option prices, the pricing kernel needs a momentum factor. Bollerslev, Tauchen, and Zhou (2009) show that incorporating variance risk in the pricing kernel can explain why option volatilities predict market returns. Bakshi, Madan, and Panayotov (2010) show that the prices of S&P 500 calls are inconsistent with monotonically declining kernels, and motivate U-shaped pricing kernels using a heterogeneous agent economy. They also show that the mimicking portfolio for the pricing kernel is U shaped.

The remainder of the paper is organized as follows. Section 1 reviews the standard SV model and presents a new discrete-time GARCH model incorporating a quadratic pricing kernel. Section 2 discusses a number of existing stylized facts and also presents new evidence on the shape of the conditional pricing kernel. Section 3 estimates the new GARCH model with a quadratic pricing kernel jointly on returns and options. Section 4 estimates the model sequentially on returns and options and compares the properties of the estimated model with the model-free stylized facts found in Section 2. Section 5 concludes. The Appendix collects proofs of propositions.

## 1. From Continuous-Time Stochastic Volatility to Discrete-Time GARCH

In order to value options, we need both a statistical description of the physical process and a pricing kernel. This section reviews the Heston (1993) model and then builds a GARCH option model with the same features.[3]

### 1.1 Continuous-time stochastic volatility

The Heston (1993) model assumes the following dynamics for the spot price $S(t)$:

$$\begin{aligned} dS(t) &= (r+\mu v(t))S(t)dt+\sqrt{v(t)}S(t)dz_1(t), \\ dv(t) &= \kappa(\theta-v(t))dt+\sigma\sqrt{v(t)}\left(\rho dz_1(t)+\sqrt{1-\rho^2}dz_2(t)\right), \end{aligned} \tag{1}$$

where $r$ is the risk-free interest rate, the parameter $\mu$ governs the equity premium, and $z_1(t)$ and $z_2(t)$ are independent Wiener processes. The notation

[3] See Hull and White (1987), Melino and Turnbull (1990), and Wiggins (1987) for other examples of option valuation with stochastic volatility. See Fleming and Kirby (2003) for a comparison between GARCH and stochastic volatility.

in (1) emphasizes the separate sources of equity risk, $z_1(t)$, and independent volatility risk, $z_2(t)$. An important aspect of our analysis is the separate premia for these risks.

In addition to the physical dynamics (1), we assume the pricing kernel takes the exponential-affine form

$$M(t)=M(0)\left(\frac{S(t)}{S(0)}\right)^{\phi}\exp\left(\delta t+\eta\int_0^t v(s)ds+\xi(v(t)-v(0))\right), \tag{2}$$

where parameters $\delta$ and $\eta$ govern the time preference, while $\phi$ and $\xi$ govern the respective aversion to equity and variance risk. When variance is constant, (2) amounts to the familiar power utility from Rubinstein's (1976) preference-based derivation of the Black-Scholes model. But, with stochastic variance, it has distinctive implications for option valuation.[4] Appendix A shows that the risk-neutral process takes the form

$$\begin{aligned} dS(t)&=rS(t)dt+\sqrt{v(t)}S(t)dz_1^*(t),\\ dv(t)&=(\kappa(\theta-v(t))-\lambda v(t))dt+\sigma\sqrt{v(t)}\left(\rho dz_1^*(t)+\sqrt{1-\rho^2}dz_2^*(t)\right), \end{aligned} \tag{3}$$

where $z_1^*(t)$ and $z_2^*(t)$ are independent Wiener processes under the risk-neutral measure and the reduced-form parameter $\lambda$ governs the variance risk premium. The pricing kernel in (2) is the unique arbitrage-free specification consistent with both the physical (1) and risk-neutral (3) dynamics. We can express the equity premium $\mu$ and variance premium $\lambda$ parameters in terms of the underlying preference parameters $\phi$ and $\xi$,

$$\begin{aligned} \mu&=-\phi-\xi\sigma\rho,\\ \lambda&=-\rho\sigma\phi-\sigma^2\xi=\rho\sigma\mu-\left(1-\rho^2\right)\sigma^2\xi. \end{aligned} \tag{4}$$

This allows us to interpret both the equity risk premium $\mu$ and the variance risk premium $\lambda$ in terms of two distinct components originating in preferences. One component is related to the risk-aversion parameter $\phi$, and the other one is related to the variance preference parameter $\xi$. We can therefore use economic intuition to sign the equity premium and the variance premium. If the pricing kernel is decreasing in the spot price, we have $\phi<0$, because marginal utility is a decreasing function of stock index returns. If hedging needs increase in times of uncertainty, then we anticipate the pricing kernel to be increasing in volatility, $\xi>0$. Empirically, the correlation between stock market returns and variance $\rho$ is strongly negative. Therefore, from (4) the equity premium $\mu$ must

[4] Stochastic variance in the pricing kernel could result, for instance, if $v(t)$ governs the variance of aggregate production in a Cox-Ingersoll-Ross (1985) model with time-seperable and non-logarithmic utility. It could also result from the model of Benzoni, Collin-Dufresne, and Goldstein (2006) where uncertainty directly affects Epstein-Zin style preferences. Testing and distinguishing these models would require consumption data. See also Bakshi, Madan, and Panayotov (2010) who consider short-sale constraints.

be positive. The variance premium $\lambda$ has a component based on covariance with equity risk, and a separate component based on the variance preference $\xi$. With a negative correlation $\rho$, we see that $\lambda$ must be negative.

It is important to note that the reduced-form risk-neutral dynamics of variance in (3) do not distinguish whether the variance risk premium $\lambda v(t)$ emanates exclusively from $\phi$ (and therefore indirectly from the equity premium $\mu$) or whether it has a separate component $\xi$. In other words, assuming $\xi = 0$ in (2) is consistent with a nonzero variance risk premium $\lambda$, as can be seen from (4). Therefore, when estimating option models with stochastic volatility using both return data and option data, it is important to explicitly write down the pricing kernel that provides the link between the physical dynamic (1) and the risk-neutral dynamic (3). It is not sufficient to simply state that (3) holds for arbitrary (negative) $\lambda$, because this assumption is consistent with the pricing kernel (2) but also with the special case with $\xi = 0$, and the economic implications of those sets of assumptions are very different. This paper explores the distinct implications of a variance premium $\xi \neq 0$ for option prices.

### 1.2 Discrete-time GARCH

The option pricing model in (1), (2), and (3) captures important stylized facts, and can be extended with additional volatility factors, stochastic jumps, and jump risk premia.[5] However, the resulting models are often cumbersome to estimate because of the complexity of the resulting filtering problem.[6] A discrete-time analog of the physical square-root return process (1) is the Heston-Nandi (2000) GARCH process

$$\begin{aligned} \ln(S(t)) &= \ln(S(t-1)) + r + \left(\mu - \frac{1}{2}\right) h(t) + \sqrt{h(t)} z(t), \\ h(t) &= \omega + \beta h(t-1) + \alpha \left(z(t-1) - \gamma \sqrt{h(t-1)}\right)^2, \end{aligned} \tag{5}$$

where $r$ is the daily continuously compounded interest rate and $z(t)$ has a standard normal distribution. We will implement this model using daily data, and we are therefore interested in its predictions for a fixed daily interval. In line with the properties of the diffusion model (1), the expected future variance is a linear function of current variance

$$E_{t-1}(h(t+1)) = \left(\beta + \alpha\gamma^2\right) h(t) + \left(1 - \beta - \alpha\gamma^2\right) E(h(t)), \tag{6}$$

---

[5] Andersen, Benzoni, and Lund (2002), Bakshi, Cao, and Chen (1997), Bates (1996b, 2000, 2006), Broadie, Chernov, and Johannes (2007), Chernov and Ghysels (2000), Eraker (2004), Eraker, Johannes, and Polson (2003), and Pan (2002) investigate jumps in returns. Broadie, Chernov, and Johannes (2007), Eraker (2004), and Eraker, Johannes, and Polson (2003) estimate models with additional jumps in volatility. Bates (2012), Carr and Wu (2004), and Huang and Wu (2004) investigate infinite-activity Levy processes. Bates (2000) and Christoffersen, Heston, and Jacobs (2009) investigate multifactor volatility models.

[6] See Lamoureux and Paseka (2009) on the role of discretization bias in implementations of the Heston (1993) model.

where $E(h(t))=(\omega+\alpha)/(1-\beta-\alpha\gamma^2)$. In words, the variance reverts to its long-run mean with daily autocorrelation of $\beta+\alpha\gamma^2$. The conditional variance of the $h(t)$ process is also linear in past variance.

$$Var_{t-1}(h(t+1))=2\alpha^2+4\alpha^2\gamma^2h(t). \tag{7}$$

The parameter $\gamma$ determines the correlation of the variance $h(t+1)$ with stock returns $R(t)=\ln(S(t)/S(t-1))$, via

$$Cov_{t-1}(R(t),h(t+1))=-2\alpha\gamma h(t). \tag{8}$$

The data robustly indicate sizable negative correlation, which means that $\gamma$ must be positive.

In existing GARCH models, Duan (1995) and Heston and Nandi (2000) use Rubinstein's (1976) power pricing kernel. In a lognormal context, this is equivalent to using the Black-Scholes formula for one-period options. Instead, we value securities using a discrete analog of the continuous-time pricing kernel in (2), namely,

$$M(t)=M(0)\left(\frac{S(t)}{S(0)}\right)^{\phi}\exp\left(\delta t+\eta\sum_{s=1}^{t}h(s)+\xi(h(t+1)-h(1))\right). \tag{9}$$

The discrete-time specification (9) is identical to the continuous pricing kernel (2) with the integral replaced by a summation. Recall that in the diffusion model, the variance process follows square-root dynamics with different parameters under the physical (1) and risk-neutral (3) measures. The following proposition shows an analogous result in the discrete model—the risk-neutral process remains in the same GARCH class.

**Proposition 1.** The risk-neutral stock price process corresponding to the physical Heston-Nandi GARCH process (5) and the pricing kernel (9) is the GARCH process

$$\begin{aligned}\ln(S(t))&=\ln(S(t-1))+r-\frac{1}{2}h^*(t)+\sqrt{h^*(t)}z^*(t),\\ h^*(t)&=\omega^*+\beta h^*(t-1)+\alpha^*\left(z^*(t-1)-\gamma^*\sqrt{h^*(t-1)}\right)^2,\end{aligned} \tag{10}$$

where $z^*(t)$ has a standard normal distribution and

$$\begin{aligned}h^*(t)&=h(t)/(1-2\alpha\xi),\\ \omega^*&=\omega/(1-2\alpha\xi),\\ \alpha^*&=\alpha/(1-2\alpha\xi)^2,\\ \gamma^*&=\gamma-\phi.\end{aligned} \tag{11}$$

**Proof.** See Appendix B. ■

The risk-neutral dynamics differ from the physical dynamics through the effect of the equity premium parameter $\mu$ and scaling factor $(1-2\alpha\xi)$. Conditional on the parameters characterizing the physical dynamic, these risk-neutral dynamics are therefore implied by the values of the kernel parameters $\phi$ and $\xi$ in Equation (9).[7] The intuition is similar to the continuous-time case in (4), where the values of the equity premium and volatility risk premium parameters $\mu$ and $\lambda$ are implied by the values of the kernel parameters $\phi$ and $\xi$.

### 1.3 Implications of the variance-dependent pricing kernel

It can be seen from (11) that a nonzero $\xi$ parameter has important implications, because it influences the level, persistence, and volatility of the variance. In contrast to the Heston-Nandi (2000) model, the risk-neutral variance $h^*(t)$ differs from the physical variance $h(t)$.[8] When $\alpha\xi > 0$, the pricing kernel puts more weight on the tails of innovations and the risk-neutral variance $h^*(t)$ exceeds the physical variance $h(t)$. The Heston-Nandi (2000) model corresponds to the special case of $\xi = 0$. The variance premium also affects the risk-neutral drift of $h^*(t)$

$$E^*_{t-1}(h^*(t+1)) = \left(\beta+\alpha^*\gamma^{*2}\right)h^*(t) + \left(1-\beta-\alpha^*\gamma^{*2}\right)E^*(h^*(t)), \qquad (12)$$

where $E^*(h^*(t)) = (\omega^*+\alpha^*)/(1-\beta-\alpha^*\gamma^{*2})$. The risk-neutral autocorrelation equals $\beta+\alpha^*\gamma^{*2}$, and a negative variance premium ($\xi > 0$) increases the risk-neutral persistence as well as the level of the future variance.

Comparison of physical parameters with risk-neutral parameters shows that if the correlation between returns and variance is negative ($\gamma > 0$), the equity premium is positive ($\mu > 0$, which corresponds to $\phi < 0$), and the variance premium is negative ($\xi > 0$), then the risk-neutral mean reversion will be smaller than the actual mean reversion. Finally, note that the variance premium alters the conditional variance of the risk-neutral variance process

$$Var^*_{t-1}(h^*(t+1)) = 2\alpha^{*2} + 4\alpha^{*2}\gamma^{*2}h^*(t). \qquad (13)$$

If the correlation between returns and variance is negative ($\gamma > 0$), the equity premium is positive ($\mu > 0$), and the variance premium is negative ($\xi > 0$), then substituting the risk-neutral parameters $\alpha^*$ and $\gamma^*$ from (10) shows that the risk-neutral variance of variance is greater than the actual variance of variance. Furthermore, we can define the risk-neutral conditional covariance,

$$Cov^*_{t-1}(R(t), h^*(t+1)) = -2\alpha^*\gamma^*h^*(t). \qquad (14)$$

The following corollary summarizes the results for this discrete-time GARCH model, which parallel those of the continuous-time model.

---

[7] The mapping between $\mu$ and $\phi$ is contained in Appendix B. With an annual U.S. equity premium $\mu h(t)$ of around 8% and variance $h(t)$ of $20\%^2$, it can be inferred that the value of the equity premium parameter $\mu$ is small, around 2.

[8] In diffusion models, the instantaneous variance is identical under the physical and risk-neutral measures, but the risk-neutral variance will differ from the physical variance over a discrete interval such as one day.

**Corollary 1.** If the equity premium is positive ($\mu > 0$), the independent variance premium is negative ($\xi > 0$), and variance is negatively correlated with stock returns ($\gamma > 0$), then (1) the risk-neutral variance $h^*(t)$ exceeds the physical variance $h(t)$; (2) the risk-neutral expected future variance exceeds the physical expected future variance; (3) the risk-neutral variance process is more persistent than the physical process; and (4) the risk-neutral variance of variance exceeds the physical variance of variance.

Corollary 1 summarizes how a premium for volatility can explain a number of puzzles concerning the level and movement of implied option variance compared with observed time-series variance. The final puzzle concerns the stylized fact pointed out by Bates (1996b), and more recently by Broadie, Chernov, and Johannes (2007), that the physical and risk-neutral volatility smiles differ. Our model captures this stylized fact through a U-shaped pricing kernel. Interestingly, even though the pricing kernel in (9) is a monotonic function of the stock price and variance, the projection of the pricing kernel onto the stock price alone can have a U shape. The following corollary formalizes this relationship.

**Corollary 2.** The logarithm of the pricing kernel is a quadratic function of the stock return.

$$\ln\left(\frac{M(t)}{M(t-1)}\right) = \frac{\xi\alpha}{h(t)}(R(t)-r)^2 - \mu(R(t)-r) + \left(\eta + \xi(\beta-1) + \xi\alpha\left(\mu - \tfrac{1}{2} + \gamma\right)^2\right)h(t) + \delta + \xi\omega + \phi r, \tag{15}$$

where $R(t) = \ln(S(t)/S(t-1))$.

**Proof.** See Appendix C. ■

In other words, the pricing kernel is a parabolic curve when plotted in log–log space. Note that whether this shape is a positive smile or a negative frown depends on the independent variance premium $\xi$, not on the total variance premium. Due to the component of the variance premium that is correlated with equity risk, it is conceivable that the total variance premium could have a different sign from the independent negative component. A negative independent variance premium ($\xi > 0$) corresponds to a U-shaped pricing kernel and thus a strong option smile. The shape of the option smile therefore provides a revealing diagnostic on the underlying preferences.

**Corollary 3.** When the independent variance premium is negative ($\xi > 0$), the pricing kernel has a U shape.

Figure 1 provides a graphical illustration of this corollary. It shows the log pricing kernel at various maturities by reporting the risk-neutral to physical

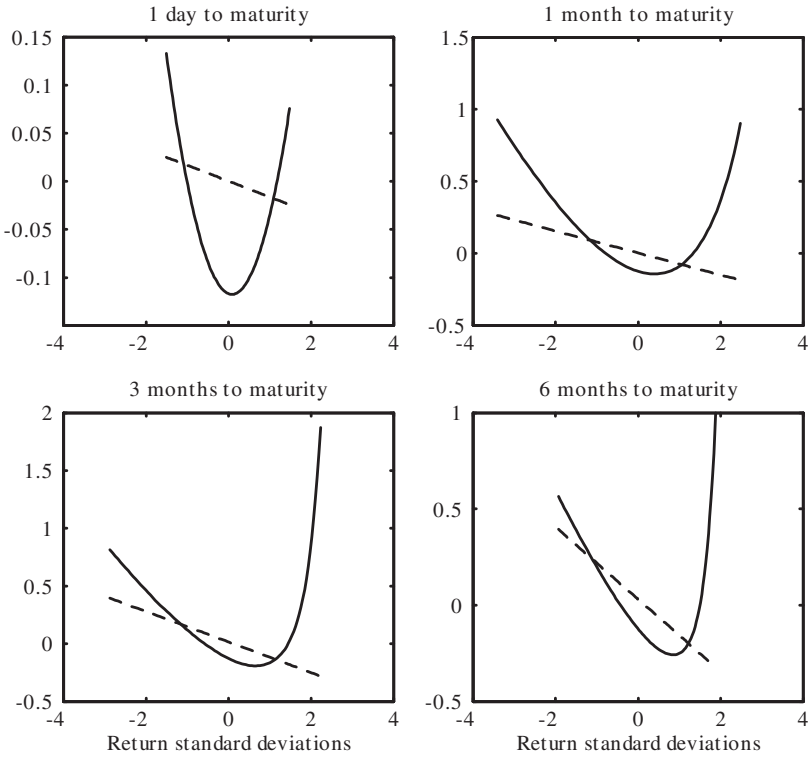


**Figure 1**
**Model-implied log ratio of the risk-neutral and physical density: Various maturities**
We plot the natural logarithm of the ratio of the risk-neutral and physical conditional densities implied by the new GARCH model using solid lines and the special case of a zero variance premium using dashed lines. Each panel corresponds to a different maturity. We use the parameters in the third Column of Table 4. The conditional variance is set to its unconditional level. The risk-free rate is set to 5% per year.

log density ratio for the new model (solid) and the special Rubinstein (1976) and Brennan (1979) case with a zero variance premium (dashed). We set the conditional variance to its unconditional level, fix the risk-free rate at 5% per year, and use our empirical estimates from the third column of Table 4, to be discussed in more detail below. Figure 1 illustrates that the log pricing kernel is U shaped, not only at the one-day maturity addressed by the corollary, but also across other maturities.

In summary, the model allows option values to display an implied variance process that is larger, more persistent, and more volatile than observed variance. The resulting risk-neutral distribution will have higher variance and fatter tails than the physical distribution. This increases the values of all options, particularly long-term options and out-of-the-money options. A negative premium for variance therefore potentially explains a number of puzzles regarding the cross-section of option prices, the relationship between physical

volatility and option-implied volatility, and the relative thickness of the tails of the physical and risk-neutral distribution.

Note that option valuation with this model is straightforward. Following Heston and Nandi (2000), the value of a call option at time $t$ with strike price $X$ maturing at $T$ is equal to

$$C(S(t),h^*(t+1),X,T)=S(t)P_1(t)-X\exp(-r(T-t))P_2(t), \qquad (16)$$

where $P_1(t)$ and $P_2(t)$ depend on the conditional moment-generating function for the risk-neutral process in (10).[9] Heston and Nandi (2000) provide a closed-form solution, which we include in Appendix D. Put options can be valued using put-call parity.

### 1.4 Variance term structures in GARCH and SV models

Before proceeding to the empirics, we now investigate the variance term structure in the new GARCH model with a variance-dependent pricing kernel and compare it to the Heston SV model. From (1), we have the stochastic variance dynamic

$$dv(t)=\kappa(\theta-v(t))dt+\sigma\sqrt{v(t)}z_v(t),$$

where $z_v(t)=\rho dz_1(t)+\sqrt{1-\rho^2}dz_2(t)$. From this we can define integrated variance as

$$VAR_0(T)\equiv E_0\left[\int_0^T v(t)dt\right]=\theta T+(v_0-\theta)\frac{\left(1-e^{-\kappa T}\right)}{\kappa}. \qquad (17)$$

The volatility term structure ($VTS$) is given by

$$VTS_{SV}\equiv\sqrt{VAR_0(T)/T}=\sqrt{\theta+\frac{(v_0-\theta)}{T}\frac{\left(1-e^{-\kappa T}\right)}{\kappa}}.$$

The risk-neutral $VTS$ is similarly

$$VTS^*_{SV}=\sqrt{\theta\frac{\kappa}{(\kappa+\lambda)}+\frac{\left(v_0-\theta\frac{\kappa}{\kappa+\lambda}\right)}{T}\frac{\left(1-e^{-(\kappa+\lambda)T}\right)}{(\kappa+\lambda)}},$$

where the price of variance risk $\lambda<0$.

In the GARCH model we have

$$h(t)=\omega+\beta h(t-1)+\alpha\left(z(t-1)-\gamma\sqrt{h(t-1)}\right)^2,$$

[9] The option valuation formula in (16) assumes no dividends. For each option on each day we use the risk-free rate to discount dividends realized during the maturity of the option from the current underlying index value. We then use this adjusted index value when computing option prices using (16).

so

$$VAR_0(T) \equiv E_0\left[\sum_{t=1}^{T} h(t)\right] = T \cdot E(h(t)) + (h(1) - E(h(t)))\frac{1-\left(\beta+\alpha\gamma^2\right)^T}{1-\left(\beta+\alpha\gamma^2\right)},$$

where $E(h(t)) = (\omega+\alpha)/(1-\beta-\alpha\gamma^2)$. We can use this to construct the GARCH volatility term structure

$$VTS \equiv \sqrt{VAR_0(T)/T} = \sqrt{E(h(t)) + \frac{(h(1) - E(h(t)))}{T}\frac{1-\left(\beta+\alpha\gamma^2\right)^T}{1-\left(\beta+\alpha\gamma^2\right)}}.$$

Using the new quadratic pricing kernel with a variance premium ($VP$), we get the risk-neutral variance term structure

$$VTS^* = \sqrt{E^*(h^*(t)) + \frac{(h^*(1) - E^*(h^*(t)))}{T}\frac{1-\left(\beta+\alpha^*\gamma^{*2}\right)^T}{1-\left(\beta+\alpha^*\gamma^{*2}\right)}},$$

where the parameters in (11) are used and $E^*(h^*(t)) = (\omega+\alpha^*)/(1-\beta-\alpha^*\gamma^{*2})$. Note that in the new model the variance premium affects the current risk-neutral spot variance, $h^*(1)$, as well as the long-run variance level, $E^*(h^*(t))$.

Figure 2 plots physical and risk-neutral volatility term structures using these formulas. We base the plots on the parameter estimates obtained in our empirical work in Section 3.3. The left column considers a low initial variance, $h(1) = E(h(t))/4$, and the right column considers the case of a high initial variance, $h(1) = 4E(h(t))$. In the top panels, we plot the GARCH model with a variance-dependent pricing kernel. In the bottom panels, we calibrate the (annual) SV parameters as follows:

$$\begin{aligned} \kappa &= \left(1-\beta-\alpha\gamma^2\right)\cdot 252, \\ \theta &= E(h(t))\cdot 252, \\ \lambda &= -\kappa\frac{E^*(h^*(t)) - E(h(t))}{E^*(h^*(t))}. \end{aligned} \tag{18}$$

This calibration ensures that the SV model matches the physical GARCH variance persistence, the physical unconditional variance, and the gap between the risk-neutral and physical unconditional variance. To match the initial conditions of the GARCH case as well, the left panel sets $v(0) = \theta/4$ and the right panel sets $v(0) = 4\theta$ .

The bottom panels in Figure 2 show that the impact of $\lambda$ in the SV model is to drive a wedge between the physical and risk-neutral term structures. However, the wedge is zero at very short maturities and then slowly rises to its unconditional level.

The new GARCH model in the top panel displays a nonzero wedge at all maturities. When the spot variance is low (left panel) the wedge is increasing

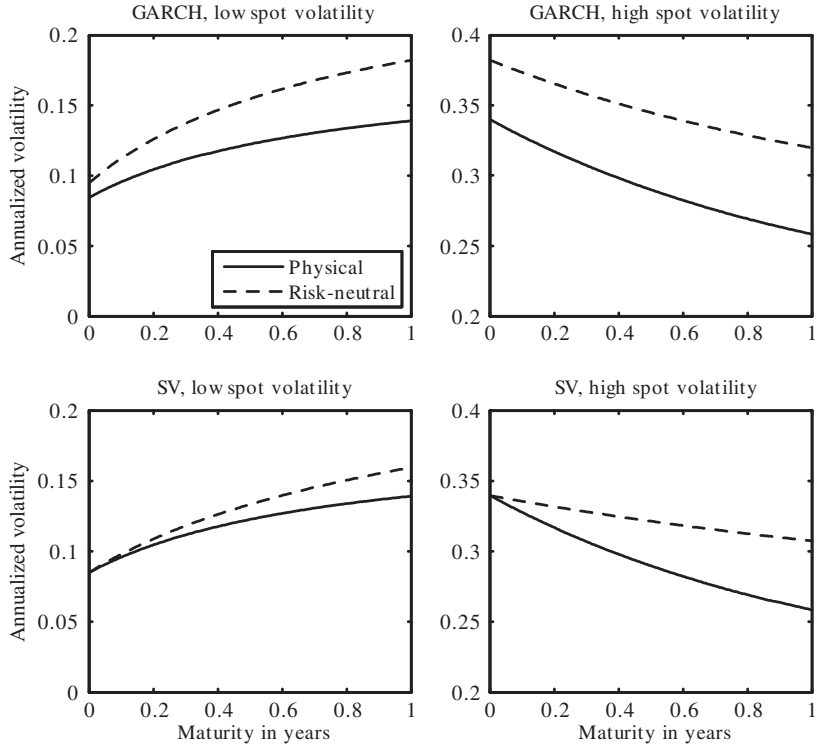


**Figure 2**
**Volatility term structures in GARCH and stochastic volatility models**
The figure shows the physical and risk-neutral variance term structures in GARCH and SV models. We use the parameter estimates from the variance premium GARCH model in Table 4 and calibrate the SV model to match initial and long-run physical variance using Equation (18). In the left column, we consider a low initial variance, $h(1)=E(h(t))/4$, and in the right column, we take a high initial variance, $h(1)=4E(h(t))$.

with maturity, but when the spot variance is high (right panel) the wedge is relatively constant across maturities. Notice that by construction the physical volatility term structures are the same across the two models. In summary, Figure 2 demonstrates that the new GARCH model and the SV model display nontrivial economic differences.

## 2. Stylized Facts in Index Option Markets

In this section, we first discuss the option and return data used in the empirical analysis, and then document and analyze a number of well-known and lesser-known stylized facts in index option markets. We pay particular attention to the shape of the pricing kernel implied by option data. Subsequently, we discuss how the new model addresses these stylized facts.

Our new model nests the existing Heston-Nandi model, and so when fitted to the data it will trivially perform better in-sample. Bates (2003) has argued that even traditional out-of-sample evaluations tend to favor more heavily parameterized models because the empirical patterns in option prices are so persistent over time. Before fitting the model to the data, we therefore assess the model by providing some relatively model-free empirical evidence on the pricing kernel implied in index returns and index option prices. We then compare this evidence with the model properties outlined in the above corollaries. We estimate the model parameters and assess in more detail the ability of the model to simultaneously fit the physical and risk-neutral distributions in Section 3.

### 2.1 Data

Our empirical analysis uses out-of-the-money S&P 500 call and put options for the January 1, 1996 to October 28, 2009 period. This is the full sample period available from OptionMetrics at the time of writing. We use only options with positive trading volume and with maturity between two weeks and one year. For each available maturity, we keep the six most actively traded strike prices and we apply the filters proposed by Bakshi, Cao, and Chen (1997), as well as other consistency checks. Rather than using a short time series of daily option data, we use an extended time period, but we select option contracts for one day per week only. This choice is motivated by two constraints. On the one hand, it is important to use as long a time period as possible, in order to be able to identify key aspects of the model. See, for instance, Broadie, Chernov, and Johannes (2007) for a discussion. On the other hand, despite the numerical efficiency of our model, the optimization problems we conduct are very time intensive, because we use very large cross-sections of option contracts. Selecting one day per week over a long time period is therefore a useful compromise. We use Wednesday data, because it is the day of the week least likely to be a holiday. It is also less likely than other days, such as Monday and Friday, to be affected by day-of-the-week effects. Moreover, following the work of Dumas, Fleming, and Whaley (1998) and Heston and Nandi (2000), several studies have used a long time series of Wednesday contracts.

Table 1 presents descriptive statistics for the option data by moneyness and maturity. Moneyness is defined as implied futures price $F$ divided by strike price $X$. When $F/X$ is smaller than one, the contract is an out-of-the-money (OTM) call, and when $F/X$ is larger than one, the contract is an OTM put. The OTM put prices were converted into call prices using put-call parity. The sample includes a total of 21,709 option contracts with an average midprice of \$33.58 and average implied volatility of 21.69%. The implied volatility is largest for the OTM put options, reflecting the well-known volatility smirk in index options. The implied volatility term structure is roughly flat on average during the sample period.

**Table 1**
**Returns and options data**

Panel A: Return characteristics (annualized)

| | 1990–2010 | 1996–2009 |
|---|---|---|
| Mean (%) | 7.90 | 5.46 |
| Standard deviation (%) | 18.28 | 20.66 |
| Skewness | −0.201 | −0.180 |
| Kurtosis | 12.258 | 10.981 |

Panel B: Option data by moneyness

| | $F/X \leq 0.96$ | $0.96 < F/X \leq 0.98$ | $0.98 < F/X \leq 1.02$ | $1.02 < F/X \leq 1.04$ | $1.04 < F/X \leq 1.06$ | $F/X > 1.06$ | All |
|---|---|---|---|---|---|---|---|
| Number of contracts | 3,032 | 1,778 | 6,287 | 2,315 | 1,744 | 6,553 | 21,709 |
| Average IV (%) | 19.81 | 18.74 | 18.71 | 20.88 | 22.45 | 26.32 | 21.69 |
| Average price | 27.90 | 32.63 | 40.87 | 41.25 | 36.97 | 25.86 | 33.58 |
| Average spread | 1.66 | 1.74 | 1.92 | 1.76 | 1.65 | 1.49 | 1.70 |

Panel C: Option data by maturity

| | $DTM \leq 30$ | $30 < DTM \leq 60$ | $60 < DTM \leq 90$ | $90 < DTM \leq 120$ | $120 < DTM \leq 180$ | $DTM > 180$ | All |
|---|---|---|---|---|---|---|---|
| Number of contracts | 968 | 4,824 | 3,710 | 1,769 | 3,074 | 7,364 | 21,709 |
| Average IV (%) | 20.91 | 20.93 | 21.38 | 23.35 | 22.21 | 21.84 | 21.69 |
| Average price | 12.92 | 19.54 | 26.97 | 32.54 | 33.76 | 49.00 | 33.58 |
| Average spread | 0.89 | 1.29 | 1.64 | 1.84 | 1.76 | 2.05 | 1.70 |

We present descriptive statistics for daily return data from January 1, 1990 to December 31, 2010, as well as daily return data from January 1, 1996 to December 31, 2009. We use Wednesday closing OTM options contracts from January 10, 1996 to October 28, 2009.

Table 1 also presents descriptive statistics for the return sample. The return sample is constructed from the S&P 500 total return index, which includes dividends. The return sample dates from January 1, 1990 to December 31, 2010. It is longer than the option sample, in order to give returns more weight in the optimization, as explained in more detail below. The standard deviation of returns, at 18.28%, is substantially smaller than the average option-implied volatility, at 21.69%. The higher moments of the return sample are consistent with return data in most historical time periods, with a very small negative skewness and substantial excess kurtosis. Table 1 also presents descriptive statistics for the return sample from January 1, 1996 to October 28, 2009, which matches the option sample. In comparison to the 1990–2010 sample, the standard deviation is somewhat higher, and average returns are somewhat lower. Average skewness and kurtosis are relatively similar to the 1990–2010 sample.

### 2.2 Fat tails and fatter tails

We now document the shape of the conditional pricing kernel using semi-parametric methods for estimating the risk-neutral and physical conditional densities. The literature does not contain a wealth of evidence on this issue. Much of what we know is either entirely (see, for instance, Bates 1996b) or partly (Rosenberg and Engle 2002) filtered through the lens of a parametric model.

Among the papers that study risk-neutral and physical densities, Jackwerth (2000) focuses on risk aversion instead of the (obviously related) shape of the pricing kernel. Ait-Sahalia and Lo (2000, 36) provide a picture of the pricing kernel as a by-product of their analysis of risk aversion, but because of their empirical technique, their estimate is most usefully interpreted as an unconditional pricing kernel. Our focus is on the conditional pricing kernel. Shive and Shumway (2006) and Bakshi, Madan, and Panayotov (2010) present the most closely related evidence on the conditional pricing kernel, but our conditioning approach is very different.

It is relatively straightforward to estimate the risk-neutral conditional density of returns using option data, harnessing the insights of Breeden and Litzenberger (1978) and Banz and Miller (1978), and there is an extensive empirical literature reporting on this. Ait-Sahalia and Lo (2000) obtain nonparametric estimates of the risk-neutral density or state-price density. This necessitates combining option data on different days, because nonparametric methods are very data intensive. Other papers, such as Jackwerth (2000), Rubinstein (1994), Bliss and Panigirtzoglou (2004), Rosenberg and Engle (2002), and Rompolis and Tzavalis (2008), use option data on a single day to infer risk-neutral densities, using a variety of methods.

Our objective is to stay as nonparametric as possible, but to provide evidence on the conditional density. We therefore need to impose a minimum

of parametric assumptions. We proceed as follows. Using the entire cross-section of options on a given day, we first estimate a second-order polynomial function for implied Black-Scholes volatility as a function of moneyness and maturity. Using this estimated polynomial, we then generate a grid of implied volatilities corresponding to a desired grid of strikes. Call these generated implied volatilities $\hat{\sigma}(S(t), X, \tau)$. Call prices can then be obtained using the Black-Scholes functional form.

$$\hat{C}(S(t), X, \tau, r) = C_{BS}(S(t), X, \tau, r; \hat{\sigma}(S(t), X, \tau)). \tag{19}$$

Following Breeden and Litzenberger (1978), the risk-neutral density for the spot price on the maturity date $T = t + \tau$ is calculated as a simple function of the second derivative of the semiparametric option price with respect to the strike price

$$\hat{f}_t^*(S(T)) = \exp(r\tau)\left[\frac{\partial^2 C_{BS}(S(t), X, \tau, r; \hat{\sigma}(S(t), X, \tau))}{\partial X^2}\right]_{|X=S(T)}. \tag{20}$$

We calculate this derivative across a grid of strike prices for each horizon, setting the current interest rate to its average sample value.

Finally, in order to plot the density against log returns rather than future spot prices, we use the transformation

$$\hat{f}_t^*(R(t,T)) = \frac{\partial}{\partial u}\Pr\left(\ln\left(\frac{S(T)}{S(t)}\right) \leq u\right) = S(t)\exp(u)\,\hat{f}_t^*(S(t)\exp(u)). \tag{21}$$

The resulting densities are truly conditional because they only reflect option information for that given day.

It is much more challenging to construct the conditional physical density of returns. Available studies walk a fine line between using short samples of daily returns, which makes the estimate truly conditional, and using longer samples, which improves the precision of the estimates. Ait-Sahalia and Lo (1998) use a relatively long series because they are less concerned about the conditional nature of the estimates. Jackwerth (2000) uses one month's worth of daily return data because he wants to illustrate the time-varying nature of the conditional density. We use a somewhat different approach. We discuss the case of monthly returns, because this is consistent with the maturity of the options used in the empirical work, but the method can easily be applied to shorter- or longer-maturity returns.

Because we want to estimate the tails of the distribution as reliably as possible, we use a long daily time series of the natural logarithm of one-month returns, from January 1, 1990, to December 31, 2010. A histogram based on this time series is effectively an estimate of the unconditional physical density of one-month log returns. We obtain a conditional density estimate for a given day by first standardizing the monthly return series by the sample mean $\bar{R}$ and the conditional one-month variance on that day, $h(t,T)$, as implied

by the daily GARCH model in (5). This provides a series of return shocks $Z(t,T)=(R(t,T)-\bar{R})/\sqrt{h(t,T)}$. We then construct a conditional histogram for a given day $t$ using the conditional variance for that day, $h(t,T)$, and the historical series of monthly shocks, $Z$. We write this estimate of the conditional physical distribution as

$$\hat{f}_t(R(t,T))=\hat{f}\left(\bar{R}+\sqrt{h(t,T)}Z\right).$$

Recall that our sample consists of fourteen years' worth of option data, for 1996–2009, and that we use Wednesday data only when we estimate the models. We conduct the estimation of the conditional densities for each of the Wednesdays in our sample, which is straightforward to execute. We cannot report all these results because of space constraints.

Figure 3 depicts the natural logarithm of the ratio of the estimates of the weekly conditional one-month risk-neutral and conditional physical density. We want to investigate the natural logarithm of the pricing kernel at different levels of return. We present fourteen sets of results, one for each year of the sample. We plot results for all weeks of the year on each picture. Specifically, we plot

$$\ln\left(\hat{f}_t^*(R(t,T))/\hat{f}_t(R(t,T))\right), \quad \text{for } t=1,2,\ldots,52.$$

In each week we trimmed 5% of observations in the left and right tails, because these observations are sometimes very noisy.

Two very important conclusions obtain. First, the pricing kernel is clearly not a monotonic function of returns, rejecting a hypothesis implicit in the Black-Scholes model and much of the option pricing literature. Second, the shape of the pricing kernel is remarkably stable across time. It is evident that the shape of the pricing kernel varies somewhat across certain years. But, we are able to draw the fifty-two pricing kernels generated for a given year on one picture to clearly illustrate the nonlinear nature of the logarithm of the kernel. If the kernel varied more within the year, Figure 3 would contain nothing but a cloudy scatter without much structure. Whether the logarithm of the kernel is exactly a quadratic function of stock returns is perhaps less obvious, because there is some noise in the estimates of the densities' right tail. However, it is clear that the relationship is nonlinear.

Compare next the purely empirical, nonparametric pricing kernels in Figure 3 with the model-based and highly parametric pricing kernels in Figure 1. Note that the one-month maturity used in Figure 3 corresponds to the upper-right panel of Figure 1, and that the empirical U shapes in Figure 3 are matched well by Figure 1.

In summary, Figure 3 illustrates that the logarithm of the pricing kernel is nonlinear and roughly quadratic as a function of the return in most sample years, and that this pricing relationship is relatively stable over time.

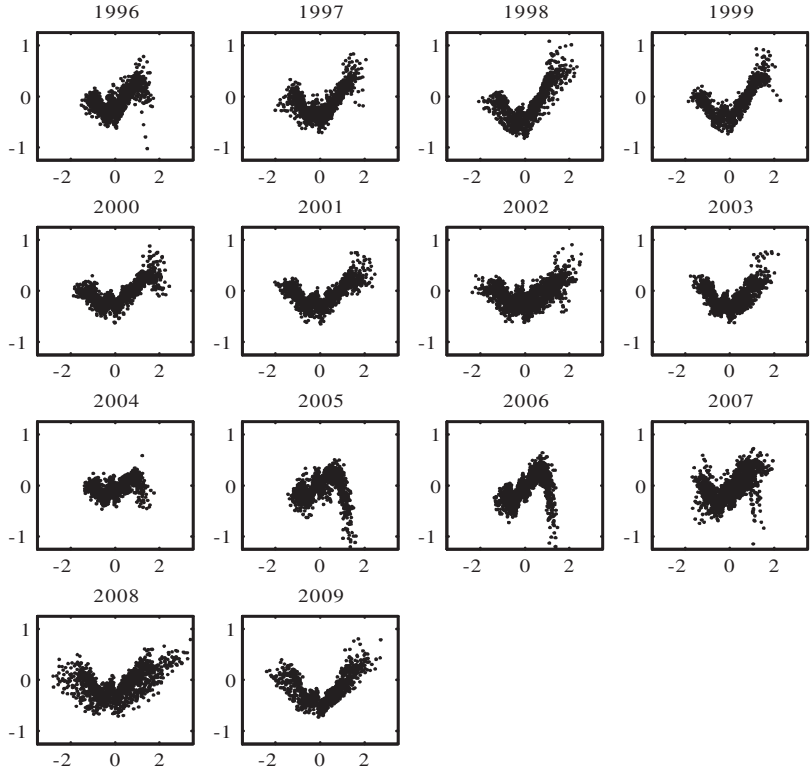


**Figure 3**
**Log ratios of risk-neutral one-month densities and physical one-month histogram**
We plot the natural logarithm of the ratio of the risk-neutral conditional densities and the physical conditional histogram. For each year in the option sample, we plot the ratios for each of the Wednesdays in that year. Log returns in monthly standard deviations are on the horizontal axis.

## 2.3 Returns on straddles

A successful model for index options also has to address a number of other stylized facts and anomalies. It is well-known that on average, risk-neutral volatility exceeds physical volatility.[10] Several authors have argued that the risk premium that explains this difference makes it interesting to short-sell straddles.[11]

Panel A in Figure 4 illustrates the cumulative log monthly returns of a short straddle strategy, which for simplicity are computed using the nearest to at-the-money, nearest to thirty-day maturity call and put option on the third Wednesday of every month. The options are held until maturity, the cash account earns the

[10] See for instance Bates (2000, 2003), Broadie, Chernov, and Johannes (2007), Chernov and Ghysels (2000), Eraker (2004), Heston and Nandi (2000), Jones (2003), and Pan (2002).

[11] See among others Coval and Shumway (2001), Bondarenko (2003), and Driessen and Maenhout (2007).

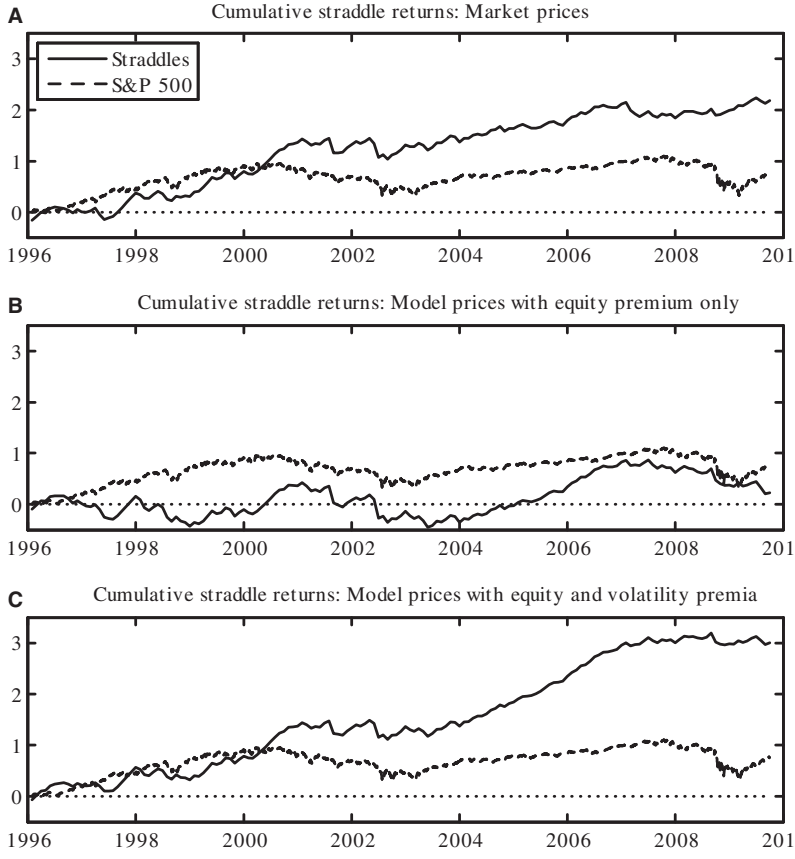


**Figure 4**
**Cumulative returns on short straddles and S&P 500 index, 1996–2009: Market and model prices**
The top panel plots the cumulative log returns of a short straddle strategy (solid) and the S&P 500 index (dashes). The middle panel uses fitted model prices allowing for only an equity premium. The bottom panel uses fitted option prices, allowing for both equity and volatility premia. We use the parameter estimates in Table 5.

risk-free rate, and the index starts out with \$100 in cash on January 1, 1996. Each month the straddle position is scaled so that the premiums collected correspond to 10% of accumulated wealth. The dashed lines in Figure 4 show the S&P 500 cumulative log monthly return for comparison.

It is obvious from Panel A in Figure 4 that the short straddle strategy was very rewarding in the 1996–2009 period. In the Black-Scholes model, the average return on this strategy would be approximately zero, and the strategy's returns would not be correlated with market returns. Panel A of Table 2 shows that the average monthly straddle log return is 1.32% with a standard deviation of 6.77%, skewness of −1.389, and excess kurtosis of 3.234. The conventional $t$-statistic testing a mean return of zero is 2.511, and thus rejects zero as

**Table 2**
**Straddle returns from market prices and model prices**

| | Panel A: Straddle returns using market prices | Panel B: Straddle returns using model prices | |
|---|---|---|---|
| Sample moments | | Equity premium only | Equity and volatility premia |
| Mean (%) | 1.32 | 0.14 | 1.82 |
| Standard deviation (%) | 6.77 | 7.79 | 6.38 |
| Skewness | −1.389 | −1.467 | −1.406 |
| Excess kurtosis | 3.234 | 3.792 | 3.463 |
| $t$-statistic | 2.511 | 0.226 | 3.666 |
| Percentile (%) | Bootstrapped critical values for the $t$-statistic | | |
| 1 | −2.173 | −2.158 | −2.148 |
| 2.5 | −1.845 | −1.838 | −1.829 |
| 5 | −1.557 | −1.551 | −1.550 |
| 10 | −1.228 | −1.211 | −1.220 |
| 90 | 1.387 | 1.373 | 1.380 |
| 95 | 1.811 | 1.805 | 1.792 |
| 97.5 | 2.177 | 2.169 | 2.169 |
| 99 | 2.599 | 2.644 | 2.610 |

We report sample statistics on 165 monthly straddle returns from January 1996 through October 2009. Model-based straddle returns rely on parameter values from sequential estimation in Table 5. Each sample of returns is bootstrapped 50,000 times to generate the reported finite sample critical values for the $t$-statistics.

a plausible value when using the standard asymptotic normal distribution. However, as forcefully pointed out by Broadie, Chernov, and Johannes (2009), straddle returns are very risky and non-Gaussian, so conventional asymptotic tests should be interpreted with care. We therefore include bootstrapped critical values for the $t$-statistic in Table 2. The critical values in Panel A show that the $t$-statistic is significant at the 5% level when using the bootstrapped critical values in a two-sided test. Indeed, the empirical value of 2.511 is close to the bootstrapped 99th percentile of 2.599.

The bootstrapped critical values can also be used to compute a finite sample confidence band around the empirical average straddle return of 1.32%. The 95% confidence band stretches from 0.351% to 2.471% per month.

### 2.4 The overreaction hypothesis

Stein (1989) documents another stylized fact in option markets that is equally robust, but which has attracted somewhat less attention. He demonstrates using a simple regression approach that longer-term implied volatility overreacts to changes in shorter-term implied volatility. Stein's most general empirical test, which is contained in Table V of his paper, is motivated by the restriction

$$E_t\left[\left(IV^{ST}_{t+(LT-ST)} - IV^{ST}_t\right) - 2\left(IV^{LT}_t - IV^{ST}_t\right)\right] = 0, \tag{22}$$

where $IV^{LT}_t$ is the implied volatility of a long-term option and $IV^{ST}_t$ is the implied volatility of a short-term option that has half the maturity of the long-term option. Intuitively, this says that the slope of the term structure of implied volatility is equal to one-half of the expected change in implied volatility. This restriction can be tested by regressing the time series in brackets on the

left-hand side on current information. Stein (1989) regresses on $IV_t^{ST}$ and finds a negative sign, which is consistent with his overreaction hypothesis, as well as with his other empirical results. When the term structure of implied volatility is steep, future implied volatilities tend to be below the forward forecasts implied by the term structure of volatility. In other words, long-term options seem to overreact to changes in short-term volatility.

We follow Stein's implementation of (22), using weekly time series of one-month and two-month implied volatilities. The regression is

$$\left(IV_{t+4}^{1M} - IV_t^{1M}\right) - 2\left(IV_t^{2M} - IV_t^{1M}\right) = a_0 + a_1 IV_t^{1M} + e_{t+4},$$

where $2M$ and $1M$ denote the two- and one-month maturity, and we test the null hypothesis that $a_1 = 0$.

Panel A in Table 3 presents the results for the Stein regression using the 1996–2009 option data. Remember that the frequency of the time series of implied volatilities is weekly, as in Stein (1989), making our results directly comparable to his. We use options that are at-the-money, according to the definition used in Table 1. Rather than averaging the two contracts that are closest to at-the-money, we fit a polynomial in maturity and moneyness to all option-implied volatilities on a given day, and then interpolate in order to obtain at-the-money implied volatility for the desired maturities. This strategy eliminates some of the noise from the data.

Panel A in Table 3 demonstrates convincingly just how robust Stein's results are. We run the regressions first for the full sample 1996–2009 and subsequently for fourteen subsamples, one for each of the years in the sample. We find a highly significant negative sign in all fifteen cases.

Stein (1989) interprets this stylized fact as an anomaly. Long-term options overreact to short-term fluctuations in implied volatility, even though volatility shocks decay very quickly. Stein (1989) therefore argues that this is a violation of rational expectations. We argue next that this robust stylized fact does not signal an anomaly but is entirely consistent with the new model developed in Section 1.

### 2.5 Stylized facts and the variance-dependent pricing kernel

In summary, this section has documented three stylized facts: First, the (log) pricing kernel appears to be quite robustly U shaped. Second, option-implied volatility is almost always higher than the physical volatility from index returns, so selling straddles is profitable on average. Third, long-term options tend to overreact to changes in short-term volatility.

Qualitatively, these findings match the model predictions captured in the three corollaries in Section 1. Corollary 1 shows that if we assume that the equity premium is positive and the independent variance premium is negative, and that variance is negatively correlated with stock returns, then in the model the risk-neutral variance will exceed the physical variance. Under realistic

**Table 3**
**Stein regressions on market and model prices**

| | | | | Model prices | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Panel A: Market prices | | | Panel B: Equity premium only | | | Panel C: Equity and volatility premia | | |
| Sample | Coefficient | Std. error | $t$-statistic | Coefficient | Std. error | $t$-statistic | Coefficient | Std. error | $t$-statistic |
| Full sample | −0.237 | 0.0093 | −25.62 | −0.197 | 0.0063 | −31.42 | −0.169 | 0.0054 | −31.13 |
| 1996 | −0.326 | 0.0255 | −12.76 | −0.260 | 0.0224 | −11.60 | −0.214 | 0.0193 | −11.11 |
| 1997 | −0.179 | 0.0266 | −6.75 | −0.200 | 0.0246 | −8.13 | −0.167 | 0.0215 | −7.77 |
| 1998 | −0.272 | 0.0387 | −7.04 | −0.196 | 0.0274 | −7.17 | −0.167 | 0.0240 | −6.95 |
| 1999 | −0.276 | 0.0157 | −17.58 | −0.195 | 0.0181 | −10.78 | −0.165 | 0.0158 | −10.44 |
| 2000 | −0.210 | 0.0138 | −15.25 | −0.183 | 0.0178 | −10.27 | −0.157 | 0.0157 | −9.99 |
| 2001 | −0.225 | 0.0233 | −9.66 | −0.202 | 0.0253 | −7.96 | −0.173 | 0.0225 | −7.69 |
| 2002 | −0.184 | 0.0273 | −6.74 | −0.155 | 0.0226 | −6.87 | −0.134 | 0.0201 | −6.67 |
| 2003 | −0.234 | 0.0200 | −11.68 | −0.235 | 0.0169 | −13.94 | −0.199 | 0.0145 | −13.73 |
| 2004 | −0.361 | 0.0194 | −18.59 | −0.234 | 0.0160 | −14.56 | −0.194 | 0.0140 | −13.85 |
| 2005 | −0.439 | 0.0246 | −17.89 | −0.255 | 0.0144 | −17.67 | −0.211 | 0.0124 | −17.03 |
| 2006 | −0.538 | 0.0234 | −22.99 | −0.285 | 0.0184 | −15.44 | −0.242 | 0.0166 | −14.55 |
| 2007 | −0.388 | 0.0386 | −10.05 | −0.206 | 0.0277 | −7.41 | −0.173 | 0.0248 | −6.98 |
| 2008 | −0.160 | 0.0479 | −3.34 | −0.114 | 0.0251 | −4.55 | −0.101 | 0.0228 | −4.43 |
| 2009 | −0.228 | 0.0156 | −14.59 | −0.205 | 0.0225 | −9.12 | −0.180 | 0.0199 | −9.05 |

Using 1996–2009 option data, we run the forecasting regressions from Stein (1989, 1021). We use at-the-money, fixed-maturity options obtained by fitting a polynomial in maturity and moneyness on every day in the option sample. As in Stein (1989), we use one-month maturity for short-term options and two-month maturity for long-term options. We run the regressions for the full sample and for each year in the sample separately. In Panel A, we regress on observed market prices; in Panel B, we regress on fitted model prices allowing only for an equity premium; and in Panel C, we regress on model prices allowing for equity and volatility risk premia. Model parameter estimates are from Table 5.

assumptions, the model thus qualitatively captures the fact that profits from selling straddles tend to be positive.

Corollary 1 also shows that under the same assumptions the risk-neutral variance process will be more persistent than the physical variance process in the model. This will qualitatively produce the Stein finding of overreaction: high persistence in the risk-neutral variance will generate large reactions ("overreaction") in the model prices of long-term options when short-term volatility changes.

Corollary 2 above shows that the daily log pricing kernel will be quadratic in the model, and Corollary 3 shows that when the variance premium is negative the daily log pricing kernel will be U shaped. Thus, at least in a qualitative sense, and at the daily frequency, the new model matches this stylized fact. Assessing whether the model generates a pricing kernel that is adequate at the horizons of interest for option valuation requires estimation of the model's parameters. That is the topic to which we now turn.

## 3. Model Estimation and Fit

We now present a detailed empirical investigation of the model outlined in Section 1. It is important to realize that the model's success in quantitatively capturing some of the stylized facts we discuss in Section 2 can be evaluated only in an appropriately designed empirical experiment. Specifically, the model's ability to capture the differences between the physical and risk-neutral distributions requires fitting both distributions using the same, internally consistent set of parameters. Perhaps somewhat surprisingly, in the stochastic volatility option pricing literature such an exercise has been attempted only by a very limited number of studies. In order to understand the implications of our empirical results, a brief summary of the existing empirical literature on index options is therefore warranted.

While the theoretical literature on option valuation is grounded in an explicit description of the link between the risk-neutral and physical distribution, much of the empirical literature on index options studies the valuation of options without contemporaneously fitting the underlying returns. In fact, it is possible to fit separate cross-sections of options while side-stepping the issue of return fit completely by parameterizing the volatility state variable.[12] When estimating multiple cross-sections, one can parameterize the volatility state variable in the same way, at the cost of estimating a high number of parameters,[13] or one can filter the volatility from underlying returns, using a variety of filters. Some papers take into account returns through the filtering exercise, but do not explicitly take into account returns in the objective function. Eraker (2004)

[12] See for instance the seminal paper by Bakshi, Cao, and Chen (1997)

[13] See for instance Bates (2000), Christoffersen, Heston, and Jacobs (2009), and Huang and Wu (2004).

and Jones (2003) conduct a Bayesian analysis based on options and return data. A few studies take a frequentist approach using an objective function that contains an option data component as well as a return data component. Chernov and Ghysels (2000) and Pan (2002) do this in a method-of-moments framework, while Santa-Clara and Yan (2010) estimate parameters using a likelihood that contains a returns component and an options component.

The literature also contains comparisons of the risk-neutral and physical distributions. Bates (1996b) observes that parameters for stochastic volatility models estimated from option data cannot fit returns. Eraker, Johannes, and Polson (2003) show the reverse. Broadie, Chernov, and Johannes (2007) use parameters estimated from returns data, and subsequently estimate the jump risk premia needed to price options.

### 3.1 A joint likelihood function

Our empirical setup is most closely related to Santa-Clara and Yan (2010). We use a joint likelihood consisting of an option-based component and a return-based component, which is relatively easy in a discrete-time GARCH setting. Note that the conditional density of the daily return is normal, so

$$f(R(t)|h(t)) = \frac{1}{\sqrt{2\pi h(t)}} \exp\left( -\frac{\left(R(t) - r - \left(\mu - \frac{1}{2}\right)h(t)\right)^2}{2h(t)} \right).$$

The return log likelihood is therefore

$$\ln L^R \propto -\frac{1}{2} \sum_{t=1}^{T} \left\{ \ln(h(t)) + \left( R(t) - r - \left(\mu - \frac{1}{2}\right)h(t) \right)^2 / h(t) \right\}. \tag{23}$$

Define the Black-Scholes Vega (BSV) weighted option valuation errors as

$$\varepsilon_i = \left(C_i^{Mkt} - C_i^{Mod}\right) / BSV_i^{Mkt},$$

where $C_i^{Mkt}$ represents the market price of the $i^{th}$ option, $C_i^{Mod}$ represents the model price, and $BSV_i^{Mkt}$ represents the Black-Scholes vega of the option (the derivative with respect to volatility) at the market-implied level of volatility. Assume these disturbances are i.i.d. normal, so the option log likelihood is

$$\ln L^O \propto -\frac{1}{2} \sum_{i=1}^{N} \left\{ \ln\left(s_\varepsilon^2\right) + \varepsilon_i^2 / s_\varepsilon^2 \right\}, \tag{24}$$

where we can concentrate out $s_\varepsilon^2$ using the sample analog $\hat{s}_\varepsilon^2 = \frac{1}{N}\sum_{i=1}^{N} \varepsilon_i^2$. These vega-weighted option errors are very useful because it can be shown that they are an approximation to implied volatility-based errors, which have desirable statistical properties. Unlike implied volatility errors, they do not require Black-Scholes inversion of model prices at every step in the optimization, which is very costly in large-scale empirical estimation exercises such as ours.[14]

[14] See for instance Carr and Wu (2007) and Trolle and Schwartz (2009) for applications of $BSV^{Mkt}$ weighted option errors.

To implement the new model with a variance-dependent pricing kernel we solve the following joint optimization problem,

$$\max_{\Theta,\Theta^*} \ln L^R + \ln L^O, \tag{25}$$

where $\Theta = \{\omega, \alpha, \beta, \gamma, \mu\}$ denotes the physical parameters and $\Theta^*$ denotes the risk-neutral parameters, which are mapped from $\Theta$ using (11). The physical variance is filtered on returns as follows:

$$\begin{aligned} h(t) &= \omega + \beta h(t-1) + \alpha\left(z(t-1) - \gamma\sqrt{h(t-1)}\right)^2, \text{ where} \\ z(t) &= \left[R(t) - r - \left(\mu - \frac{1}{2}\right)h(t)\right]/\sqrt{h(t)}, \end{aligned} \tag{26}$$

and the risk-neutral variance is computed from

$$h^*(t) = h(t)/(1 - 2\alpha\xi).$$

The riskless rate $r$ in (23) is set to 5%, and we use the term structure of interest rates from OptionMetrics when pricing options in (24). For the three models with nonzero $\mu$, we use the value estimated from returns to ensure that the equity premium is properly calibrated.

We implement the joint optimization problem just described for the new model with a variance-dependent pricing kernel, as well as for two special cases that are nested in the new model. The first of the special cases we consider contains no risk premia. This amounts to setting $\mu = 0$ and $\xi = 0$. The second special case is the Heston-Nandi (2000) model that allows for equity risk only, so we have $\mu \neq 0$ and $\xi = 0$. Finally, the third model is the newly proposed model: It contains an independent volatility premium by allowing $\xi$ to be a free parameter, which is estimated by maximizing the joint likelihood in (25).

### 3.2 The ad hoc benchmark

To demonstrate the usefulness and implications of the new pricing kernel in (2), we consider a particular implementation that completely ignores the specification of the pricing kernel. This exercise is merely intended as a benchmark, and we refer to it as the ad hoc model. In this model, we estimate the physical and risk-neutral parameters separately. The physical parameters are estimated by maximizing the return likelihood only:

$$\Theta_{AdHoc} = \{\omega, \alpha, \beta, \gamma, \mu\} = \arg\max_{\Theta} \ln L^R, \tag{27}$$

where physical variance is filtered on returns, $R(t)$,

$$\begin{aligned} h(t) &= \omega + \beta h(t-1) + \alpha\left(z(t-1) - \gamma\sqrt{h(t-1)}\right)^2, \text{ where} \\ z(t) &= \left[R(t) - r - \left(\mu - \frac{1}{2}\right)h(t)\right]/\sqrt{h(t)}, \text{ and} \\ h(0) &= \frac{\omega + \alpha}{1 - \beta - \alpha\gamma^2}. \end{aligned} \tag{28}$$

The risk-neutral parameters are estimated by maximizing the option likelihood only:

$$\Theta^*_{AdHoc} = \left\{\omega^*, \alpha^*, \beta^*, \gamma^*\right\} = \arg\max_{\Theta^*} \ln L^O, \tag{29}$$

where the risk-neutral variance path is obtained from filtering on returns, $R(t)$, as well:

$$\begin{aligned} h^*(t) &= \omega^* + \beta^* h(t-1) + \alpha^* \left(z^*(t-1) - \gamma^* \sqrt{h^*(t-1)}\right)^2, \text{ where} \\ z^*(t) &= \left[R(t) - r + \frac{1}{2} h^*(t)\right] / \sqrt{h^*(t)}, \text{ and} \\ h^*(0) &= \frac{\omega^* + \alpha^*}{1 - \beta^* - \alpha^* \gamma^{*2}}. \end{aligned} \tag{30}$$

Note that in (30) we again use returns, $R(t)$, to filter the risk-neutral variances, $h^*(t)$, which are then used to compute option prices.

The starting point for the physical variance path, $h(0)$, is the first trading day of January 1990 when our return sample begins. The starting point for the ad hoc risk-neutral variance path, $h^*(0)$, is the first Wednesday of January 1996, when our option sample begins.

In the ad hoc model the mapping between $\Theta$ and $\Theta^*$ in (11) has been ignored completely. The risk-neutral parameters in $\Theta^*_{AdHoc}$ are free to vary independently of the physical parameters in $\Theta_{AdHoc}$, and $h^*(t)$ is therefore not restricted to be proportional to $h(t)$. Note that $h(t)$ and $h^*(t)$ are both filtered on returns, and neither is estimated as a separate parameter. While both variance processes are filtered from returns, they can follow quite distinct sample paths, as we will see in Figure 6.

Because there are no built-in restrictions between the physical and risk-neutral parameters in the ad hoc model, we separately fit the risk-neutral and physical distribution as well as possible. This serves as a benchmark for three other implementations that impose economic restrictions between the physical and the risk-neutral distribution.

### 3.3 Joint model fit and model properties

Table 4 presents the estimation results, which are quite striking.[15] Consider the likelihood functions for the different models. Notice that at the optimum, the contribution from the options part of the likelihood is very similar in all four cases. This indicates that the contribution of the option data to the likelihood is so important that the parameters adjust to fit the option data as in the benchmark ad hoc case in Column 4, by implication sacrificing some goodness of fit in the

[15] We impose $\omega^* = \omega = 0$ in estimation because the nonnegativity constraint–which ensures positive variances–is binding.

**Table 4**
**Joint maximum likelihood estimation on returns and options**

| | | | | Ad hoc model |
|---|---|---|---|---|
| Physical parameters | No premia | Equity premium only | Equity and volatility premia | Return-based estimation |
| $\omega$ | 0 | 0 | 0 | 0 |
| $\alpha$ | $1.410\times10^{-6}$ | $1.410\times10^{-6}$ | $8.887\times10^{-7}$ | $3.364\times10^{-6}$ |
| $\beta$ | 0.755 | 0.755 | 0.756 | 0.838 |
| $\gamma$ | 411.19 | 409.63 | 515.57 | 196.82 |
| $\mu$ | 0 | 1.594 | 1.594 | 1.594 |
| Risk-neutral parameters | | | | Option-based estimation |
| $(1-2\alpha\xi)^{-1}$ | 1 | 1 | 1.2638 | |
| $\omega^*$ | 0 | 0 | 0 | 0 |
| $\alpha^*$ | $1.410\times10^{-6}$ | $1.410\times10^{-6}$ | $1.419\times10^{-6}$ | $1.417\times10^{-6}$ |
| $\beta^*$ | 0.755 | 0.755 | 0.756 | 0.755 |
| $\gamma^*$ | 411.19 | 411.23 | 409.32 | 410.36 |
| Total likelihood | 56,403.5 | 56,410.7 | 56,480.9 | 56,578.5 |
| From returns | 17,673.7 | 17,681.0 | 17,749.2 | 17,846.9 |
| From options | 38,729.7 | 38,729.8 | 38,731.6 | 38,731.7 |
| Physical properties | | | | |
| Long-run volatility | 0.239 | 0.210 | 0.170 | 0.165 |
| Daily autocorrelation, $h$(t) | 0.994 | 0.992 | 0.992 | 0.969 |
| Annualized volatility of $h$(t) | 0.070 | 0.062 | 0.040 | 0.058 |
| Correlation ($R$(t), $h$(t+1)) | −0.994 | −0.992 | −0.992 | −0.945 |
| Risk-neutral properties | | | | |
| Long-run volatility | 0.239 | 0.239 | 0.240 | 0.240 |
| Daily autocorrelation, $h^*$(t) | 0.994 | 0.994 | 0.994 | 0.994 |
| Annualized volatility of $h^*$(t) | 0.070 | 0.070 | 0.071 | 0.071 |
| Correlation ($R$(t), $h^*$(t+1)) | −0.994 | −0.994 | −0.994 | −0.994 |

Parameter estimates are obtained by optimizing a joint likelihood on returns and options. Parameters and autocorrelations are daily. The returns and option samples are described in Table 1. For each model, we report the total likelihood value at the optimum as well as the value of the returns component at the optimum and the option component at the optimum. We estimate four models. The no premia model has four parameters, with $\mu=0$ and $\xi=0$. The model with equity premium only has five parameters. It imposes $\xi=0$. The model with equity and volatility premia has six parameters. In the ad hoc model, the physical and risk-neutral parameters are not linked. This model has nine parameters. All volatility parameters are constrained to be positive, which guarantees positive variance.

return component of the likelihood, which is in all cases much smaller than in the ad hoc specification. This occurs despite using a return sample, 1990–2010, that is longer than the option sample, 1996–2009.

The ad hoc case in Column 4, which is not constrained by any mapping between physical and risk-neutral parameters, serves as a reference point. The resulting physical and risk-neutral properties are entirely determined by the returns and option data respectively, and serve as a benchmark for the stylized facts the other three models need to match.

Column 4 of Table 4 shows that in the ad hoc case, the average (long-run) risk-neutral volatility, 0.240, is much higher than the average physical volatility, 0.165. The average annualized risk-neutral volatility of risk-neutral variance in (13) is 0.071, which is larger than the physical volatility of physical variance, 0.058. The (absolute value of the) average of the risk-neutral leverage

correlation defined as

$$\frac{Cov^*_{t-1}(R(t),h^*(t+1))}{\sqrt{h^*(t)Var^*_{t-1}(h^*(t+1))}}=\frac{-2\alpha^*\gamma^*h^*(t)}{\sqrt{h^*(t)\left(2\alpha^{*2}+4\alpha^{*2}\gamma^{*2}h^*(t)\right)}},$$

annualized and evaluated at $h^*(t)=E[h^*(t)]$, is $-0.994$, which in absolute value exceeds its physical counterpart, $-0.945$. These empirical features are of course critical in ensuring that the risk-neutral distribution has fatter tails than the physical distribution. Table 4 contains preliminary information on the overreaction puzzle as well, which is determined by persistence. In Column 4, the daily risk-neutral variance persistence, 0.994, is higher than the daily physical variance persistence, 0.969.

Now consider our new model with a variance-dependent pricing kernel in Column 3. How does the wedge between physical and risk-neutral properties for this model relate to the ad hoc case in Column 4, where this wedge is entirely data-driven? Note that the most important determinant of these differences is the parameter combination $(1-2\alpha\xi)^{-1}$, which shows the ratio of the risk-neutral variance $h^*(t)$ to the physical variance $h(t)$. Our estimate of this parameter combination is 1.2638, indicating a substantial difference between physical and risk-neutral volatility.

In Column 3, the difference between the long-run physical, 0.240, and risk-neutral volatility, 0.170, is almost exactly the same as in the ad hoc case in Column 4. The physical volatility of variance, 0.040, is a good approximation of the benchmark as is its risk-neutral counterpart, 0.071. However, while the physical leverage correlation, $-0.992$, is smaller (in absolute value) than the risk-neutral leverage correlation, $-0.994$, the difference is very small. Similarly, even though the physical variance persistence, 0.992, is smaller than the risk-neutral variance persistence, 0.994, the difference between the two persistence measures is much smaller than in Column 4. However, note that in the new model, risk-neutral persistence differs from physical persistence, but spot volatility differs under the two measures as well. The model's ability to capture Stein's puzzle depends on both spot volatility and volatility persistence. We investigate this issue in more detail below.

The model without risk premia in Column 1 serves as another interesting benchmark. The model's physical properties are identical to its risk-neutral properties. Allowing for a nonzero equity premium in Column 2 results in a rather small increase in the total likelihood (from 56,403.5 to 56,410.7). Asymptotically, twice the difference in likelihood values has a chi-square distribution with one degree of freedom. Therefore, this improvement is statistically insignificant. It can be seen that in economic terms the improvements are also modest, as the risk-neutral average volatility, variance persistence, and volatility of variance are not dramatically different from their physical counterparts. However, while these effects are quantitatively small, it is reassuring that qualitatively all effects go in the expected direction, confirming our results in Section 1. The risk-neutral volatility exceeds the

physical volatility, the risk-neutral variance persistence is higher, and the risk-neutral tails are fatter.

When adding an independent volatility premium to the specification in Column 3, the likelihood function improves spectacularly (from 56,410.7 to 56,480.9). Perhaps even more pertinently, the physical and risk-neutral model properties for the new model in Column 3 match the ad hoc case in Column 4 substantially better than the model in Column 2, which only has an equity premium.

### 3.4 Sequential parameter estimation

In the joint estimation exercise above, the option prices drive the parameter estimates. Option fit is virtually identical across models, whereas the return fit differs substantially across models. In this section, we instead follow Broadie, Chernov, and Johannes (2007), who first estimate each model on returns only, and then subsequently assess the fit of each model to option prices in a second step. In the new model, only the $\xi$ parameter is estimated in the second step. In the ad hoc model, as before, all parameters are fit in the second step. Note that the ad hoc model estimates are both joint and sequential because the parameters are not linked between the two measures. In the model with only an equity premium and in the model with no premia at all, no parameters are estimated in the second step; the option likelihood is simply evaluated using the risk-neutral parameters.

Table 5 contains the estimation results from this sequential estimation exercise. Note that, as expected, the return fit is now virtually identical across models, whereas the option fit differs substantially. The option likelihood increases from 26,009.7 in the model with no premia to 27,424.3 in the model with equity premium only and to 32,644.7 in the new model with a variance-dependent pricing kernel. Table 5 also shows that the sequentially estimated ratio $(1-2\alpha\xi)^{-1} = 1.2836$ is close to the value of 1.2638 from joint estimation in Table 4. Finally, the ad hoc model, as in Table 4, has an option likelihood of 38,731.7. Using conventional $\chi^2$ critical values, we would thus strongly reject the model that only contains an equity premium in favor of the new model, but we would also reject the new model in favor of the ad hoc model. We investigate the ad hoc model further below.

In summary, we conclude that the new model with independent variance premium improves on existing models in both a statistical and an economic sense.

## 4. Model Evaluation Based on the Stylized Facts

While the above model diagnostics are insightful, ultimately we want to assess if the estimated model can capture the model-free stylized facts documented in Section 2.

**Table 5**
**Sequential maximum likelihood estimation on returns and options**

| Physical parameters | No premia | Equity premium only | Equity and volatility premia | Ad hoc model |
|---|---|---|---|---|
| $\omega$ | 0 | 0 | 0 | 0 |
| $\alpha$ | $3.364\times10^{-6}$ | $3.364\times10^{-6}$ | $3.364\times10^{-6}$ | $3.364\times10^{-6}$ |
| $\beta$ | 0.838 | 0.838 | 0.838 | 0.838 |
| $\gamma$ | 196.82 | 196.82 | 196.82 | 196.82 |
| $\mu$ | 0 | 1.594 | 1.594 | 1.594 |
| Risk-neutral parameters | | | | |
| $(1-2\alpha\xi)^{-1}$ | 1 | 1 | 1.2836 | |
| $\omega^*$ | 0 | 0 | 0 | 0 |
| $\alpha^*$ | $3.364\times10^{-6}$ | $3.364\times10^{-6}$ | $5.543\times10^{-6}$ | $1.417\times10^{-6}$ |
| $\beta^*$ | 0.838 | 0.838 | 0.838 | 0.755 |
| $\gamma^*$ | 196.82 | 198.42 | 154.69 | 410.36 |
| Total likelihood | 43,854.1 | 45,271.2 | 50,491.6 | 56,578.5 |
| From returns | 17,844.3 | 17,846.9 | 17,846.9 | 17,846.9 |
| From options | 26,009.7 | 27,424.3 | 32,644.7 | 38,731.7 |
| Physical properties | | | | |
| Long-run volatility | 0.165 | 0.165 | 0.165 | 0.165 |
| Daily autocorrelation, $h$(t) | 0.969 | 0.969 | 0.969 | 0.969 |
| Annualized volatility of $h$(t) | 0.058 | 0.058 | 0.058 | 0.058 |
| Correlation ($R$(t), $h$(t+1)) | −0.945 | −0.945 | −0.945 | −0.945 |
| Risk-neutral properties | | | | |
| Long-run volatility | 0.165 | 0.171 | 0.220 | 0.240 |
| Daily autocorrelation, $h^*$(t) | 0.969 | 0.971 | 0.971 | 0.994 |
| Annualized volatility of $h^*$(t) | 0.058 | 0.060 | 0.100 | 0.071 |
| Correlation ($R$(t), $h^*$(t+1)) | −0.945 | −0.949 | −0.950 | −0.994 |

Parameter estimates are obtained by optimizing first the likelihood on returns and then the likelihood on options. Parameters and autocorrelations are daily. The returns and option samples are described in Table 1. For each model, we report the total likelihood value at the optimum as well as the value of the returns component at the optimum and the option component at the optimum. We estimate four models. The no premia model has four parameters, with $\mu=0$ and $\xi=0$. The model with equity premium only has five parameters. It imposes $\xi=0$. The model with equity and volatility premia has six parameters. In the ad hoc model, the physical and risk-neutral parameters are not linked. This model has nine parameters. All volatility parameters are constrained to be positive, which guarantees positive variance.

### 4.1 Fat tails and fatter tails in model prices

Figure 3 presents strong model-free evidence of a U-shaped pricing kernel. In Figure 5, we once again present the log ratios of risk-neutral and physical densities, but we use the fitted model prices from the new model with estimates from Table 5 instead of the market prices used in Figure 3. The similarity between the density ratios in Figures 3 and 5 is quite striking. Recall that each plot contains an entire year's worth of Wednesday closing option prices. We conclude that the new model captures the first stylized fact discussed in Section 2.

### 4.2 Returns on straddles from model prices

Panel A of Figure 4 indicates that selling index straddles on average has been a very profitable, albeit risky, strategy during the 1996–2010 period. We now investigate whether the new model can produce this pattern of straddle profits.

Panel B of Figure 4 plots the cumulative monthly log returns of a short straddle strategy (solid line) and the S&P 500 index (dashes). Rather than

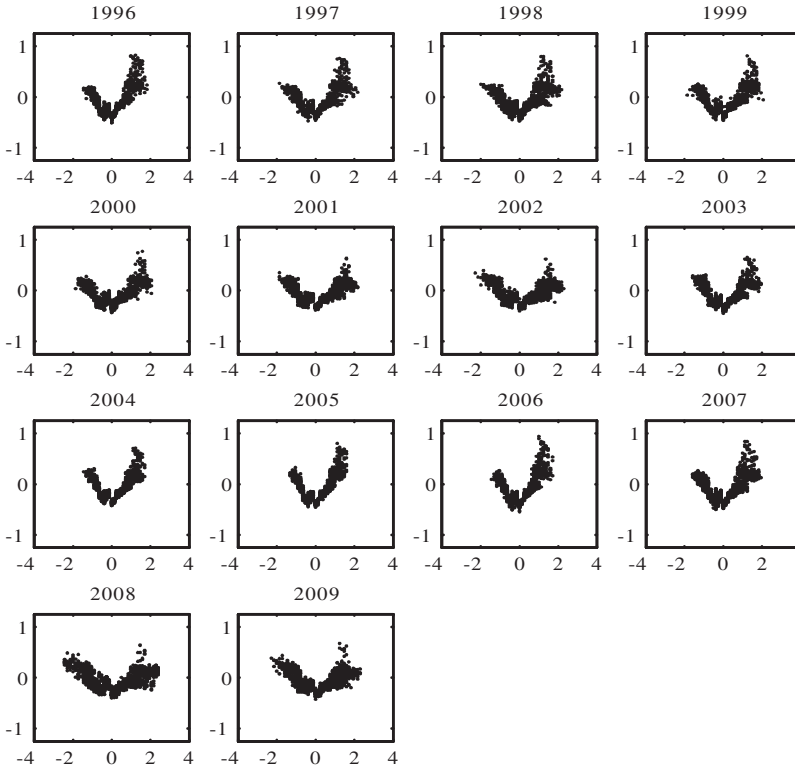


**Figure 5**
**Log ratios of risk-neutral and physical distributions from model option prices**
We plot the natural logarithm of the ratio of the risk-neutral conditional densities and the physical conditional histogram. For each year in the option sample, we plot the ratios for each of the Wednesdays in that year. Log returns in monthly standard deviations are on the horizontal axis. This figure uses model prices rather than market prices. We use the parameter estimates in Table 5.

selling at the market prices, the strategy now sells at the model prices implied by the parameter estimates in Column 2 of Table 5 corresponding to the model that has only an equity premium and no volatility premium. As in Panel A, the straddle position is scaled so that the premiums collected correspond to 10% of accumulated wealth. Panel B of Figure 4 shows that the model without an equity premium is not able to produce the straddle returns seen in the market data in Panel A. Panel B of Table 2 shows that the average monthly log straddle return is 0.14% with a standard deviation of 7.79%, implying a conventional $t$-statistic of 0.226. Return skewness is $-1.467$, and excess kurtosis is 3.792.

Panel C of Figure 4 uses model prices including equity and volatility premia and shows that the model is capable of generating option prices that broadly produce the level of risk and return found in the actual straddle strategy in Panel A of Figure 4. Panel B of Table 2 shows that the average monthly

log straddle return is 1.82% with a standard deviation of 6.38%, implying a conventional $t$-statistic of 3.666. Return skewness is $-1.406$, and excess kurtosis is 3.463. The straddle returns from the new model match the empirical moments in Section 2.3 quite well.

Using the bootstrapped critical values for the $t$-statistics in Panel B of Table 2, we see that the value of 0.226 for the model with only an equity premium is not significant, whereas the $t$-statistic for the new model with a variance premium is significant at all conventional levels.

Recall from Section 2.3 that the 95% confidence band around the empirical mean straddle return went from 0.351% to 2.471% per month. The model-based mean of 1.82% in the new model is thus well within the 95% confidence band. Note also that the straddle return mean of 0.14% in the model with only an equity premium is not within the 95% confidence band of the empirical mean.

### 4.3 The overreaction hypothesis with model prices

Panel A in Table 3 replicates Stein's (1989) finding of overreaction by running the weekly ATM implied volatility regression

$$\left(IV_{t+4}^{1M} - IV_t^{1M}\right) - 2\left(IV_t^{2M} - IV_t^{1M}\right) = a_0 + a_1 IV_t^{1M} + e_{t+4},$$

for each year in our sample.

Panels B and C in Table 3 repeat the Stein regressions but use the fitted model prices rather than market prices. Just as in Section 2.4, we fit a polynomial in maturity and moneyness to all option-implied volatilities on a given day and then interpolate in order to obtain at-the-money implied volatility for the desired maturities. Table 3 shows that the model IVs yield similar findings. The $a_1$ slope coefficient on current short-run volatility is significantly negative in each year of the 1996–2010 sample when using model IVs instead of market IVs.

Table 3 shows that both the existing model (Panel B) with only equity premium and the new model (Panel C) with equity and volatility premia are capable of generating highly persistent implied volatilities at the two-month horizon in spite of the small gap between daily physical and risk-neutral persistence documented in Tables 4 and 5.

### 4.4 Model fit across moneyness and maturity

Table 6 further explores the option valuation performance of our new model with independent variance premium. It presents implied volatility root-mean-squared error (IV RMSE) and IV bias by moneyness and maturity. We use the model estimates from Table 5.

When comparing the first and second rows in Table 6, we see that the equity premium plays a relatively small role in improving option fit compared to a model with no premia.

A comparison of the second and third rows of Table 6 shows that the independent volatility premium yields a pervasive and substantial improvement in option fit—both in terms of IV RMSE and bias. The parameterizations with

**Table 6**
**IV RMSE and bias (in percent) by moneyness and maturity, 1996–2009, using sequentially estimated parameters**

| Model | $F/X \leq 0.96$ | $0.96 < F/X \leq 0.98$ | $0.98 < F/X \leq 1.02$ | $1.02 < F/X \leq 1.04$ | $1.04 < F/X \leq 1.06$ | $F/X > 1.06$ | All |
|---|---|---|---|---|---|---|---|
| Panel A: IV RMSE by moneyness | | | | | | | |
| No premia | 6.8283 | 5.6924 | 5.7626 | 6.5941 | 7.0885 | 9.6589 | 7.4563 |
| Equity premium only | 6.4638 | 5.3941 | 5.4992 | 6.2679 | 6.7228 | 9.2202 | 7.1021 |
| Equity and volatility premia | 5.0671 | 4.7315 | 5.0024 | 5.1364 | 5.0785 | 6.7271 | 5.5865 |
| Ad hoc | 3.6690 | 3.6747 | 3.8085 | 3.8521 | 3.8237 | 5.2592 | 4.2767 |
| Panel B: IV bias by moneyness (market minus model) | | | | | | | |
| No premia | 4.4884 | 3.1217 | 2.6990 | 3.9055 | 4.8356 | 7.3502 | 4.6864 |
| Equity premium only | 4.0137 | 2.6804 | 2.2900 | 3.4684 | 4.3880 | 6.8361 | 4.2278 |
| Equity and volatility premia | 0.0590 | −0.9241 | −1.1645 | −0.0958 | 0.7252 | 2.8330 | 0.4981 |
| Ad hoc | −0.4450 | −1.1417 | −0.9284 | −0.0489 | 0.5995 | 2.2217 | 0.2885 |
| Model | $DTM \leq 30$ | $30 < DTM \leq 60$ | $60 < DTM \leq 90$ | $90 < DTM \leq 120$ | $120 < DTM \leq 180$ | $DTM > 180$ | All |
| Panel C: IV RMSE by maturity | | | | | | | |
| No premia | 6.6097 | 6.7522 | 7.0608 | 8.5752 | 7.7639 | 7.7678 | 7.4563 |
| Equity premium only | 6.3609 | 6.4455 | 6.7288 | 8.1550 | 7.3950 | 7.3852 | 7.1021 |
| Equity and volatility premia | 5.3207 | 5.1211 | 5.2516 | 6.1861 | 5.7475 | 5.8504 | 5.5865 |
| Ad hoc | 5.1807 | 4.5744 | 4.1802 | 4.1503 | 4.2775 | 4.0131 | 4.2767 |
| Panel D: IV bias by maturity (market minus model) | | | | | | | |
| No premia | 3.2681 | 4.0271 | 4.4477 | 5.8684 | 5.0785 | 4.9767 | 4.6864 |
| Equity premium only | 2.9170 | 3.6462 | 4.0239 | 5.3842 | 4.5902 | 4.4543 | 4.2278 |
| Equity and volatility premia | 0.0662 | 0.5506 | 0.5288 | 1.6367 | 0.6182 | 0.1814 | 0.4981 |
| Ad hoc | −0.1722 | 0.4435 | 0.4808 | 0.8048 | 0.5389 | −0.0777 | 0.2885 |

We report IV-based RMSE and bias by moneyness and maturity using the parameter estimates from the sequential estimation exercise in Table 5.

no volatility premia display a strong positive bias, implying that on average the models underprice options when the volatility premium is ignored. This bias is virtually eliminated—and the RMSE is greatly improved—when the independent volatility premium is incorporated into the model.

Comparing the third and fourth rows, we see that the ad hoc model improves upon the new model throughout the option categories. While the biases are low for both models, the ad hoc model is able to capture more of the variation in option IVs than is the new model.

### 4.5 Deconstructing the ad hoc model

While the new model substantially improves upon the standard Heston-Nandi model's ability to fit returns and options, the ad hoc model fits the data even better, albeit with more parameters. The ad hoc model allows for the risk-neutral volatility to evolve differently from the physical volatility, which is a key advantage when fitting returns and options jointly. Figure 6 shows the ratio of risk-neutral and physical volatility in the ad hoc model. The horizontal line at one indicates the ratio imposed in the Heston-Nandi model. The horizontal line at $\sqrt{1.26} \approx 1.12$ indicates the ratio imposed by the new model with a variance-dependent pricing kernel. The horizontal line at $\sqrt{h^*/h} \approx 1.18$ indicates the square root of the ratio of the sample average risk-neutral and physical variance paths in the ad hoc model. Figure 6 suggests that while the new model matches the average ad hoc volatility ratio quite well, a more general pricing kernel allowing for a time-varying proportional volatility premium would fare even better. We leave this research topic for future work.

## 5. Conclusion

We develop a new GARCH option model for the purpose of index option valuation by specifying a more general pricing kernel. Unlike the traditional Black-Scholes (1973) and Rubinstein (1976) pricing kernel, which is a function of the index return only, we specify that the pricing kernel is also a function of the return variance. Although the pricing kernel is specified as monotonic in the index return, the projection of the pricing kernel onto returns is U shaped. With a negative variance premium, this model feature is consistent with semiparametric evidence from returns and options, which reveals that the conditional pricing kernel is U shaped in returns and is relatively stable over time.

The new model generalizes the Heston-Nandi (2000) model by allowing the risk-neutral variance, persistence, and volatility of variance to differ from their physical counterparts. We demonstrate that the model can qualitatively and quantitatively account for a number of important puzzles in the option pricing literature. In order to demonstrate that the more general pricing kernel can reconcile the return distributions implicit in the time series of returns and option prices, we implement the model by maximizing the sum of the

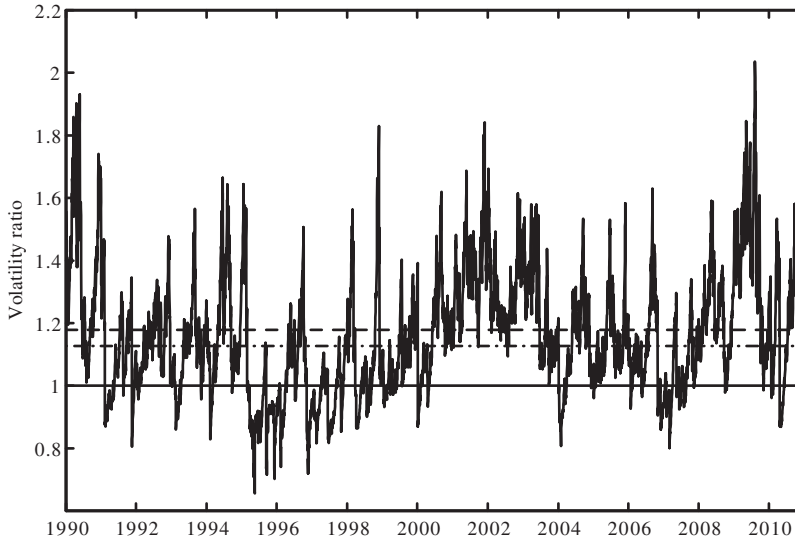


**Figure 6**
**Risk-neutral to physical volatility ratio in the ad hoc model**

We plot the ratio $\sqrt{h_t^*/h_t}$ in the ad hoc model. The horizontal dashes mark the square root of the average variance ratio (1.18) in the ad hoc model. The horizontal dash-dots mark the $(1-2\alpha\xi)^{-1/2}$ ratio (1.12) in the new model with a variance-dependent pricing kernel. The model with only an equity risk premium has a ratio of one marked by the solid horizontal line.

return likelihood and a likelihood based on successive cross-sections of option prices. We benchmark the model's performance to an ad hoc model that does not impose restrictions across the physical and risk-neutral parameters. We find that the fit of the model with the new pricing kernel is dramatically better than the fit resulting from the traditional pricing kernel, and that several important differences between physical and risk-neutral moments are similar to the differences obtained for the ad hoc model. We use fitted model prices to show that the model can quantitatively capture the stylized facts found in our initial model-free analysis based on market prices.

The results and empirical exercises in this paper can be extended and generalized in a number of ways. First, while different models may be needed, alternative loss functions may emphasize different moments and therefore yield different results. Second, we have focused on European-style options, and extensions to American-style options (Ritchken and Trevor 1999; Duan and Simonato 2001) would be useful. Third, it may prove interesting to investigate the implications of more general pricing kernels for option valuation in the presence of richer return dynamics with fat tails or jumps in returns and/or volatility (Bollerslev 1987; Christoffersen, Heston, and Jacobs 2006; Barone-Adesi, Engle, and Mancini 2008), multiple volatility components (Engle and Lee 1999; Christoffersen et al. 2008), long memory in volatility

(Bollerslev and Mikkelsen 1996), or economic drivers of uncertainty (David and Veronesi 2011). Our results indicate that it is critical to evaluate such models using a loss function similar to the one in this paper, with a return component as well as an option component. Two questions arise in this regard: On the one hand, can these models improve option fit? And, on the other hand, can they capture a richer link between the physical and risk-neutral distributions with intuitively plausible prices of risk? Models that can reconcile the differences between physical and risk-neutral persistence, as well as physical and risk-neutral leverage correlation, would be of particular interest.

## Appendix

### A. Risk neutralization of the Heston (1993) SV model

We show that the physical SV process in (1) and the risk-neutral SV process in (3) are linked by the pricing kernel in (2) by imposing the condition that the product of any traded asset and the pricing operator is a martingale under the physical probability measure.

Let $B(t)$ be the risk-free bond ($\frac{dB(t)}{B(t)}=rdt$) and let $U(t)$ be an asset that depends on the spot price, $S(t)$, and the volatility, $v(t)$. From (2) we get the dynamic of $M(t)$

$$\log(M(t))=\log(M(0))+\phi\log(S(t))-\phi\log(S(0))+\delta t+\eta\int_0^t v(s)ds+\xi[v(t)-v(0)].$$

This gives

$$\begin{aligned}d\log(M)&=\phi d\log(S)+\delta dt+\eta v dt+\xi dv\\&=\left[\phi\left(r+\mu v-\frac{1}{2}v\right)+\delta+\eta v+\xi\kappa(\theta-v)\right]dt\\&\quad+\left[\phi\sqrt{v}+\xi\sigma\rho\sqrt{v}\right]dz_1+\left[\xi\sigma\sqrt{1-\rho^2}\sqrt{v}\right]dz_2,\end{aligned}$$

where we use the fact that $d\log(S)=(r+\mu v-\frac{1}{2}v)dt+\sqrt{v}dz_1$, which is obtained by applying Itô's lemma to Equation (1). Note that we have suppressed dependence on $t$ in the notation. Again by Itô's lemma, we have

$$\begin{aligned}\frac{dM}{M}&=\left[\phi\left(r+\mu v-\frac{1}{2}v\right)+\delta+\eta v+\xi\kappa(\theta-v)+\frac{1}{2}\phi^2 v+\phi\xi\sigma\rho v+\frac{1}{2}\xi^2\sigma^2 v\right]dt\\&\quad+\left[\phi\sqrt{v}+\xi\sigma\rho\sqrt{v}\right]dz_1+\left[\xi\sigma\sqrt{1-\rho^2}\sqrt{v}\right]dz_2.\end{aligned}$$

From the condition that $B(t)M(t)$ is a martingale, that is, its drift is equal to zero, we deduce the restrictions on $\delta$ and $\eta$.

$$\phi\left(r+\mu v-\frac{1}{2}v\right)+\delta+\eta v+\xi\kappa(\theta-v)+\frac{1}{2}\phi^2 v+\phi\xi\sigma\rho v+\frac{1}{2}\xi^2\sigma^2 v=-r.$$

This must hold for $v=0$ and for $v=+\infty$, giving the restrictions on $\delta$ and $\eta$

$$\begin{aligned}\delta&=-(1+\phi)r-\xi\kappa\theta,\\\eta&=-\phi\mu+\frac{1}{2}\phi+\xi\kappa-\frac{1}{2}\left(\phi^2+2\phi\xi\sigma\rho+\xi^2\sigma^2\right).\end{aligned}$$

Similarly, from the condition that $S(t)M(t)$ is a martingale, we deduce a restriction on $\phi$. Using the fact that the drift of $M$ is equal to $-rMdt$, we have

$$\mu v+(\phi+\xi\sigma\rho)v=0,$$

or equivalently $\phi=-\mu-\xi\sigma\rho$.

Finally, by equating the drift of $U(t)M(t)$ to zero, we deduce the restriction on $\xi$

$$0=\Bigg[M\left(U_t+rSU_S+\mu vSU_S+\kappa(\theta-v)U_v+\frac{1}{2}vS^2U_{SS}+vS\sigma\rho U_{Sv}+\frac{1}{2}v\sigma^2U_{vv}\right) \\ -rUM+Mv(\phi+\xi\sigma\rho)(SU_S+\sigma\rho U_v)+M\xi v\sigma^2\left(1-\rho^2\right)U_v\Bigg]dt. \tag{A1}$$

Since $\frac{U(t)}{B(t)}$ is a martingale under the risk-neutral measure, we can also show that

$$\lambda vU_v+rU=U_t+rSU_S+\kappa(\theta-v)U_v+\frac{1}{2}vS^2U_{SS}+vS\sigma\rho U_{Sv}+\frac{1}{2}v\sigma^2U_{vv}. \tag{A2}$$

Substituting (A2) into (A1), we get the drift $D^{UM}$

$$D^{UM}=\Big[M(\lambda vU_v+rU+\mu vSU_S)-rUM+Mv(\phi+\xi\sigma\rho)(SU_S+\sigma\rho U_v)+M\xi v\sigma^2\left(1-\rho^2\right)U_v\Big]dt$$

$$=0.$$

After simplification and using the fact that $\phi+\xi\sigma\rho=-\mu$, we get

$$\lambda vU_v-\mu v\sigma\rho U_v+\xi v\sigma^2\left(1-\rho^2\right)U_v=0,$$

and therefore

$$\xi=\frac{\mu\sigma\rho-\lambda}{\sigma^2\left(1-\rho^2\right)}. \tag{A3}$$

Using (A3), we also obtain

$$\phi=\frac{-\mu+\lambda\sigma^{-1}\rho}{\left(1-\rho^2\right)}.$$

**B. Proof of Proposition 1**

The discrete-time form of the pricing operator (2) is

$$M(t)=M(0)\left(\frac{S(t)}{S(0)}\right)^{\phi}\exp\left(\delta t+\eta\sum_{s=1}^{t}h(s)+\xi(h(t+1)-h(1))\right), \tag{A4}$$

and therefore

$$\frac{M(t)}{M(t-1)}=\left(\frac{S(t)}{S(t-1)}\right)^{\phi}\exp(\delta+\eta h(t)+\xi(h(t+1)-h(t))). \tag{A5}$$

The summations in (A4) are equivalent to the integrals in the continuous-time form (2) under the standard GARCH convention that variance is constant throughout the day, and changes discretely overnight. We shall show that the pricing kernel in (A4) is consistent with the Heston-Nandi GARCH dynamic (5) for the following parameter mapping:

$$\delta=-(\phi+1)r-\xi\omega+\tfrac{1}{2}\ln(1-2\xi\alpha),$$

$$\eta=-\left(\mu-\tfrac{1}{2}\right)\phi-\xi\alpha\gamma^2+(1-\beta)\xi-\frac{(\phi-2\xi\alpha\gamma)^2}{2(1-2\xi\alpha)},$$

$$\phi=-\left(\mu-\tfrac{1}{2}+\gamma\right)(1-2\alpha\xi)+\gamma-\tfrac{1}{2}.$$

From the GARCH dynamic (5), we can write

$$\frac{S(t)}{S(t-1)}=\exp\left(r+\left(\mu-\tfrac{1}{2}\right)h(t)+\sqrt{h(t)}z(t)\right),$$

$$h(t+1)-h(t)=\omega+(\beta-1)h(t)+\alpha\left(z(t)-\gamma\sqrt{h(t)}\right)^2.$$

Substituting these into (A5) gives

$$\frac{M(t)}{M(t-1)}=\exp\left(\phi r+\phi\left(\mu-\tfrac{1}{2}\right)h+\phi\sqrt{h}z+\delta+\eta h+\xi\omega+\xi(\beta-1)h+\xi\alpha\left(z-\gamma\sqrt{h}\right)^2\right),$$

where we have dropped the time subscripts for $z$ and $h$. Expanding the square and collecting terms gives

$$\frac{M(t)}{M(t-1)}=\exp\left(\phi r+\delta+\xi\omega+\left[\phi\left(\mu-\tfrac{1}{2}\right)+\eta+\xi(\beta-1)+\xi\alpha\gamma^2\right]h+[\phi-2\xi\alpha\gamma]\sqrt{h}z+[\xi\alpha]z^2\right).$$

First, we use the fact that for any initial value $h(t)$, the parameters must be consistent with the Euler equation for the riskless asset.

$$E_{t-1}\left[\frac{M(t)}{M(t-1)}\right]=\exp(-r). \tag{A6}$$

Note that

$$E_{t-1}\left[\frac{M(t)}{M(t-1)}\right]=\exp\left(\phi r+\delta+\xi\omega+\left[\phi\left(\mu-\tfrac{1}{2}\right)+\eta+\xi(\beta-1)+\xi\alpha\gamma^2\right]h\right)$$

$$*E\left(\exp\left([\phi-2\xi\alpha\gamma]\sqrt{h}z+[\xi\alpha]z^2\right)\right).$$

We need the following result:

$$E\left[\exp\left(az^2+2abz\right)\right]=\exp\left(-\tfrac{1}{2}\ln(1-2a)+\frac{2a^2b^2}{1-2a}\right).$$

For our application, we have

$$a=\xi\alpha,$$

$$b=\left(\frac{\phi-2\xi\alpha\gamma}{2\xi\alpha}\right)\sqrt{h},$$

and thus

$$2a^2b^2=2\xi^2\alpha^2\left(\frac{\phi-2\xi\alpha\gamma}{2\xi\alpha}\right)^2h=\tfrac{1}{2}(\phi-2\xi\alpha\gamma)^2h.$$

Therefore,

$$E\left(\exp\left([\phi-2\xi\alpha\gamma]\sqrt{h}z+\xi\alpha z^2\right)\right)=\exp\left(-\tfrac{1}{2}\ln(1-2\xi\alpha)+\frac{(\phi-2\xi\alpha\gamma)^2}{2(1-2\xi\alpha)}h\right),$$

and

$$E_{t-1}\left[\frac{M(t)}{M(t-1)}\right]=\exp\left(\begin{array}{c}\phi r+\delta+\xi\omega+\left[\phi\left(\mu-\tfrac{1}{2}\right)+\eta+\xi(\beta-1)+\xi\alpha\gamma^2\right]h\\ -\tfrac{1}{2}\ln(1-2\xi\alpha)+\frac{(\phi-2\xi\alpha\gamma)^2}{2(1-2\xi\alpha)}h\end{array}\right).$$

Rearranging and using (A6), we get

$$(\phi+1)r+\delta+\xi\omega-\tfrac{1}{2}\ln(1-2\xi\alpha)+\left[\phi\left(\mu-\tfrac{1}{2}\right)+\eta+\xi(\beta-1)+\xi\alpha\gamma^2+\frac{(\phi-2\xi\alpha\gamma)^2}{2(1-2\xi\alpha)}\right]h=0.$$

Therefore, we must have

$$\delta=-(\phi+1)r-\xi\omega+\tfrac{1}{2}\ln(1-2\xi\alpha),$$

$$\eta=-\left(\mu-\tfrac{1}{2}\right)\phi-\xi\alpha\gamma^2+(1-\beta)\xi-\frac{(\phi-2\xi\alpha\gamma)^2}{2(1-2\xi\alpha)}.$$

Now we use the Euler equation for the underlying index

$$E_{t-1}\left[\frac{S(t)}{S(t-1)}\frac{M(t)}{M(t-1)}\right]=1.$$

First, note that $\frac{S(t)}{S(t-1)}\frac{M(t)}{M(t-1)}$ is equal to $\frac{M(t)}{M(t-1)}$ in (A5) with $\phi$ is replaced by $\phi+1$; thus, we can use the expression for $E_{t-1}[\frac{M(t)}{M(t-1)}]$ to write

$$E_{t-1}\left[\frac{S(t)}{S(t-1)}\frac{M(t)}{M(t-1)}\right]=\exp\left(\begin{array}{c}(\phi+1)r+\delta+\xi\omega+\left[(\phi+1)\left(\mu-\frac{1}{2}\right)+\eta+\xi(\beta-1)+\xi\alpha\gamma^2\right]h\\ -\frac{1}{2}\ln(1-2\xi\alpha)+\frac{(\phi+1-2\xi\alpha\gamma)^2}{2(1-2\xi\alpha)}h\end{array}\right).$$

Taking logs, setting equal to zero and using the above solutions for $\delta$ and $\eta$ gives

$$\left[\mu-\tfrac{1}{2}+\frac{1+2\phi-4\xi\alpha\gamma}{2(1-2\xi\alpha)}\right]h=0.$$

Solving this for $\phi$ yields

$$\phi=-\left(\mu-\frac{1}{2}+\gamma\right)(1-2\alpha\xi)+\gamma-\tfrac{1}{2}.$$

To find the risk-neutral dynamic, note that the risk-neutral density is proportional to the physical density times the pricing kernel

$$f^*_{t-1}(S(t))=\frac{f_{t-1}(S(t))M(t)}{E_{t-1}(M(t))}.$$

Tedious integration shows that $z(t)$ is normally distributed under the risk-neutral measure, but with a different mean and variance. This is a direct implication of the form of the pricing kernel. It is therefore convenient to define a standardized risk-neutral innovation

$$z^*(t)=\sqrt{1-2\alpha\xi}\left(z(t)+\left(\mu+\frac{\alpha\xi}{1-2\alpha\xi}\right)\sqrt{h(t)}\right). \tag{A7}$$

The risk-neutral dynamics in Equation (10) can then be derived by substituting the risk-neutral innovation (A7) into the physical GARCH process in Equation (5).

### C. Proof of Corollary 2

We want to show that the logarithm of the pricing kernel takes the quadratic form

$$\ln\left(\frac{M(t)}{M(t-1)}\right)=\frac{\xi\alpha}{h(t)}(R(t)-r)^2-\mu(R(t)-r)+\left(\eta+\xi(\beta-1)+\xi\alpha\left(\mu-\tfrac{1}{2}+\gamma\right)^2\right)h(t)+\delta+\xi\omega+\phi r,$$

where $R(t)=\ln(S(t)/S(t-1))$. First recall that

$$\frac{M(t)}{M(t-1)}=\left(\frac{S(t)}{S(t-1)}\right)^{\phi}\exp(\delta+\eta h(t)+\xi(h(t+1)-h(t))), \tag{A8}$$

and that

$$h(t+1)-h(t)=\omega+(\beta-1)h(t)+\alpha\left(z(t)-\gamma\sqrt{h(t)}\right)^2.$$

We also have

$$R(t)=r+\left(\mu-\tfrac{1}{2}\right)h(t)+\sqrt{h(t)}z(t),$$

so

$$z(t)=\frac{R(t)-r-\left(\mu-\frac{1}{2}\right)h(t)}{\sqrt{h(t)}}.$$

From this we get

$$\begin{aligned}\ln\left(\frac{M(t)}{M(t-1)}\right)&=\phi R(t)+\delta+\eta h(t)+\xi(h(t+1)-h(t))\\&=\phi R(t)+\delta+\xi\omega+(\eta+\xi(\beta-1))h(t)+\xi\alpha\left(z(t)-\gamma\sqrt{h(t)}\right)^2\\&=\phi R(t)+\delta+\xi\omega+(\eta+\xi(\beta-1))h(t)+\frac{\xi\alpha}{h(t)}\left(R(t)-r-\left(\mu-\tfrac{1}{2}+\gamma\right)h(t)\right)^2.\end{aligned}$$

Expanding the square and collecting terms yields

$$\begin{aligned}\ln\left(\frac{M(t)}{M(t-1)}\right)=\frac{\xi\alpha}{h(t)}(R(t)-r)^2+\left(\phi-2\xi\alpha\left(\mu-\tfrac{1}{2}+\gamma\right)\right)(R(t)-r)\\+\left(\eta+\xi(\beta-1)+\xi\alpha\left(\mu-\tfrac{1}{2}+\gamma\right)^2\right)h(t)+\delta+\xi\omega+\phi r.\end{aligned}$$

From the equation for $\phi$ we have

$$\phi-2\alpha\xi\left(\mu-\tfrac{1}{2}+\gamma\right)=-\mu,$$

and the result obtains.

### D. Heston-Nandi option valuation

This appendix provides the moment-generating function (MGF) and option valuation formula for the GARCH(1,1) process used in this paper and derived in Heston and Nandi (2000). We have the risk-neutral dynamic

$$\ln(S(t))=\ln(S(t-1))+r-\frac{1}{2}h^*(t)+\sqrt{h^*(t)}z^*(t),$$

$$h^*(t+1)=\omega^*+\beta h^*(t)+\alpha^*\left(z^*(t)-\gamma^*\sqrt{h^*(t)}\right)^2.$$

The conditional MGF corresponding to this process is

$$g^*_{t,T}(\varphi)\equiv E^*_t[\exp(\varphi\ln(S(T)))]=\exp\left(\varphi\ln(S(t))+A_{t,T}(\varphi)+B_{t,T}(\varphi)h^*(t+1)\right).$$

As $S_T$ is known at time $T$, the MGF implies the terminal condition

$$A_{T,T}(\varphi)=B_{T,T}(\varphi)=0.$$

The $A_{t,T}(\varphi)$ and $B_{t,T}(\varphi)$ functions are defined via the following difference equations:

$$A_{t,T}(\varphi)=A_{t+1,T}(\varphi)+\varphi r+B_{t+1,T}(\varphi)\omega^*-\frac{1}{2}\ln\left(1-2B_{t+1,T}(\varphi)\alpha^*\right),$$

$$B_{t,T}(\varphi)=-\frac{1}{2}\varphi+B_{t+1,T}(\varphi)\beta+B_{t+1,T}(\varphi)\alpha^*\left(\gamma^*\right)^2 + \frac{\frac{1}{2}\varphi^2+2B_{t+1,T}(\varphi)\alpha^*\gamma^*\left(B_{t+1,T}(\varphi)\alpha^*\gamma^*-\varphi\right)}{1-2B_{t+1,T}(\varphi)\alpha^*}.$$

The conditional MGF $g^*_{t,T}(\varphi)$ is computed by recursing day-by-day from the terminal $A_{T,T}(\varphi)$ and $B_{T,T}(\varphi)$ to the current $A_{t,T}(\varphi)$ and $B_{t,T}(\varphi)$. Option values can now be computed using

$$C(S(t),h^*(t+1),X,T)=S(t)P_1(t)-X\exp(-r(T-t))P_2(t),$$

where we use numerical integration to compute

$$P_1(t)=\left(\frac{1}{2}+\frac{\exp(-r(T-t))}{\pi}\int_0^\infty \mathrm{Re}\left[\frac{X^{-i\varphi}g^*_{t,T}(i\varphi+1)}{i\varphi S(t)}\right]d\varphi\right),$$

$$P_2(t)=\left(\frac{1}{2}+\frac{1}{\pi}\int_0^\infty \mathrm{Re}\left[\frac{X^{-i\varphi}g^*_{t,T}(i\varphi)}{i\varphi}\right]d\varphi\right).$$

**References**


Ait-Sahalia, Y., and A. Lo. 1998. Nonparametric estimation of state-price densities implicit in financial asset prices. *Journal of Finance* 53:499–547.

———. 2000. Nonparametric risk management and implied risk aversion. *Journal of Econometrics* 94:9–51.

Andersen, T. G., L. Benzoni, and J. Lund. 2002. Estimating jump-diffusions for equity returns. *Journal of Finance* 57:1239–84.

Bakshi, G., C. Cao, and Z. Chen. 1997. Empirical performance of alternative option pricing models. *Journal of Finance* 52:2003–49.

Bakshi, G., and N. Kapadia. 2003. Delta hedged gains and the negative market volatility risk premium. *Review of Financial Studies* 16:527–66.

Bakshi, G., D. Madan, and G. Panayotov. 2010. Returns of claims on the upside and the viability of U-shaped pricing kernels. *Journal of Financial Economics* 97:130–54.

Banz, R., and M. Miller. 1978. Prices for state-contingent claims: Some estimates and applications. *Journal of Business* 51:653–72.

Barone-Adesi, G., R. Engle, and L. Mancini. 2008. A GARCH option pricing model with filtered historical simulation. *Review of Financial Studies* 21:1223–58.

Bates, D. 1996a. Testing option pricing models. Statistical methods in finance. In *Handbook of statistics*. Eds. G. S. Maddala and C. R. Rao, 567–611. Amsterdam: Elsevier.

———. 1996b. Jumps and stochastic volatility: Exchange rate processes implicit in Deutsche mark options. *Review of Financial Studies* 9:69–107.

———. 2000. Post-'87 crash fears in the S&P 500 futures option market. *Journal of Econometrics* 94:181–238.

———. 2003. Empirical option pricing: A retrospection. *Journal of Econometrics* 116:387–404.

———. 2006. Maximum likelihood estimation of latent affine processes. *Review of Financial Studies* 19:909–65.

———. 2008. The market for crash risk. *Journal of Economic Dynamics and Control* 32:2291–321.

———. 2012. U.S. stock market crash risk, 1926–2010. *Journal of Financial Economics* 105:229–59.

Benzoni, L., P. Collin-Dufresne, and R. Goldstein. 2011. Explaining asset pricing puzzles associated with the 1987 market crash. *Journal of Financial Economics* 101:552–73.

Black, F., and M. Scholes. 1973. The pricing of options and corporate liabilities. *Journal of Political Economy* 81:637–59.

Blair, B., S.-H. Poon, and S. Taylor. 2001. Forecasting S&P 500 volatility: The incremental information content of implied volatilities and high-frequency index returns. *Journal of Econometrics* 105:5–26.

Bliss, R., and N. Panigirtzoglou. 2004. Option-implied risk aversion estimates. *Journal of Finance* 59:407–46.

Bollerslev, T. 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31:307–27.

———. 1987. A conditionally heteroskedastic time series model for speculative prices and rates of return. *Review of Economics and Statistics* 69:542–47.

Bollerslev, T., and H. Mikkelsen. 1996. Modeling and pricing long memory in stock market volatility. *Journal of Econometrics* 73:151–84.

Bollerslev, T., G. Tauchen, and H. Zhou. 2009. Expected stock return and variance risk premia. *Review of Financial Studies* 22:4463–92.

Bondarenko, O. 2003. Why are put options so expensive? Working Paper, University of Illinois at Chicago.

Breeden, D., and R. Litzenberger. 1978. Prices of state-contingent claims implicit in option prices. *Journal of Business* 51:621–51.

Brennan, M. 1979. The pricing of contingent claims in discrete-time models. *Journal of Finance* 34:53–68.

Broadie, M., M. Chernov, and M. Johannes. 2007. Model specification and risk premiums: Evidence from futures options. *Journal of Finance* 62:1453–90.

———. 2009. Understanding index option returns. *Review of Financial Studies* 22:4493–529.

Brown, D. P., and J. C. Jackwerth. 2012. The pricing kernel puzzle: Reconciling index option data and economic theory. In *Derivative securities pricing and modelling.* Eds. J. A. Batten and N. Wagner, 155–83. West Yorkshire, UK: Emerald Group Publishing Limited.

Canina, L., and S. Figlewski. 1993. The informational content of implied volatility. *Review of Financial Studies* 6:659–81.

Carr, P., and L. Wu. 2004. Time-changed Levy processes and option pricing. *Journal of Financial Economics* 17:113–41.

———. 2007. Stochastic skew in currency options. *Journal of Financial Economics* 86:213–47.

Chabi-Yo, F. 2012. Pricing kernels with stochastic skewness and volatility risk. *Management Science* 58:624–40.

Chabi-Yo, F., R. Garcia, and E. Renault. 2008. State dependence can explain the risk aversion puzzle. *Review of Financial Studies* 21:973–1011.

Chernov, M. 2003. Empirical reverse engineering of the pricing kernel. *Journal of Econometrics* 116:329–64.

———. 2007. On the role of risk premia in volatility forecasting. *Journal of Business and Economic Statistics* 25:411–26.

Chernov, M., and E. Ghysels. 2000. A study towards a unified approach to the joint estimation of objective and risk neutral measures for the purpose of option valuation. *Journal of Financial Economics* 56:407–58.

Christoffersen, P., S. Heston, and K. Jacobs. 2006. Option valuation with conditional skewness. *Journal of Econometrics* 131:253–84.

———. 2009. The shape and term structure of the index option smirk: Why multifactor stochastic volatility models work so well. *Management Science* 55:1914–32.

Christoffersen, P., K. Jacobs, C. Ornthanalai, and Y. Wang. 2008. Option valuation with long-run and short-run volatility components. *Journal of Financial Economics* 90:272–97.

Coval, J., and T. Shumway. 2001. Expected option returns. *Journal of Finance* 56:983–1009.

Cox, J., J. Ingersoll, and S. Ross. 1985. An intertemporal general equilibrium model of asset prices. *Econometrica* 53:363–84.

David, A., and P. Veronesi. 2011. Investor and central bank uncertainty and fear measures embedded in index options. Working Paper, University of Chicago.

Day, T., and C. Lewis. 1992. Stock market volatility and the information content of stock index options. *Journal of Econometrics* 52:267–87.

Driessen, J., and P. Maenhout. 2007. An empirical portfolio perspective on option pricing anomalies. *Review of Finance* 11:561–603.

Duan, J. 1995. The GARCH option pricing model. *Mathematical Finance* 5:13–32.

Duan, J., and J.-G. Simonato. 2001. American option pricing under GARCH by a Markov chain approximation. *Journal of Economic Dynamics and Control* 25:1689–718.

Dumas, B., J. Fleming, and R. Whaley. 1998. Implied volatility functions: Empirical tests. *Journal of Finance* 53:2059–106.

Engle, R. 1982. Autoregressive conditional heteroskedasticity with estimates of the variance of UK inflation. *Econometrica* 50:987–1008.

Engle, R., and G. Lee. 1999. A permanent and transitory component model of stock return volatility. In *Cointegration, causality, and forecasting: A Festschrift in honor of Clive W. J. Granger*. Eds. R. Engle and H. White, 475–97. New York: Oxford University Press.

Eraker, B. 2004. Do stock prices and volatility jump? Reconciling evidence from spot and option prices. *Journal of Finance* 59:1367–403.

Eraker, B., M. Johannes, and N. Polson. 2003. The role of jumps in returns and volatility. *Journal of Finance* 58:1269–300.

Fleming, J. 1998. The quality of market volatility forecasts implied by S&P 100 index option prices. *Journal of Empirical Finance* 5:317–45.

Fleming, J., and C. Kirby. 2003. A closer look at the relation between GARCH and stochastic autoregressive volatility. *Journal of Financial Econometrics* 1:365–419.

Heston, S. 1993. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies* 6:327–43.

Heston, S., and S. Nandi. 2000. A closed-form GARCH option pricing model. *Review of Financial Studies* 13:585–626.

Huang, J.-Z., and L. Wu. 2004. Specification analysis of option pricing models based on time-changed Levy processes. *Journal of Finance* 59:1405–39.

Hull, J., and A. White. 1987. The pricing of options with stochastic volatilities. *Journal of Finance* 42:281–300.

Jackwerth, J. 2000. Recovering risk aversion from option prices and realized returns. *Review of Financial Studies* 13:433–51.

Jones, C. 2003. The dynamics of stochastic volatility: Evidence from underlying and options markets. *Journal of Econometrics* 116:181–224.

Jorion, P. 1995. Predicting volatility in the foreign exchange market. *Journal of Finance* 50:507–28.

Lamoureux, C., and W. Lastrapes. 1993. Forecasting stock-return variance: Toward an understanding of stochastic implied volatilities. *Review of Financial Studies* 6:293–326.

Lamoureux, C., and A. Paseka. 2009. Information in option prices and the underlying asset dynamics. Working Paper, University of Arizona.

Liu, J., J. Pan, and T. Wang. 2004. An equilibrium model of rare event premia. *Review of Financial Studies* 18:131–64.

Melino, A., and S. Turnbull. 1990. Pricing foreign currency options with stochastic volatility. *Journal of Econometrics* 45:239–65.

Pan, J. 2002. The jump-risk premia implicit in options: Evidence from an integrated time-series study. *Journal of Financial Economics* 63:3–50.

Poteshman, A. 2001. Underreaction, overreaction, and increasing misreaction to information in the options market. *Journal of Finance* 56:851–76.

Ritchken, P., and R. Trevor. 1999. Pricing options under generalized GARCH and stochastic volatility processes. *Journal of Finance* 54:377–402.

Rompolis, L., and E. Tzavalis. 2008. Recovering risk neutral densities from option prices: A new approach. *Journal of Financial and Quantitative Analysis* 43:1037–54.

Rosenberg, J., and R. Engle. 2002. Empirical pricing kernels. *Journal of Financial Economics* 64:341–72.

Rubinstein, M. 1976. The valuation of uncertain income streams and the pricing of options. *Bell Journal of Economics* 7:407–25.

———. 1994. Implied binomial trees. *Journal of Finance* 49:771–818.

Santa-Clara, P., and S. Yan. 2010. Crashes, volatility, and the equity premium. *Review of Economics and Statistics* 92:435–51.

Shive, S., and T. Shumway. 2006. Is the pricing kernel monotonic? Working Paper, Notre Dame University.

Stein, J. 1989. Overreactions in the options markets. *Journal of Finance* 44:1011–23.

Trolle, A., and E. Schwartz. 2009. Unspanned stochastic volatility and the pricing of commodity derivatives. *Review of Financial Studies* 22:4423–61.

Wiggins, J. 1987. Option values under stochastic volatility: Theory and empirical evidence. *Journal of Financial Economics* 19:351–72.