

Option Anomalies and the Pricing Kernel*

Peter Christoffersen

Steven Heston

McGill University, CBS, and CREATES University of Maryland

Kris Jacobs

University of Houston, McGill University, and Tilburg University

January 12, 2010

Abstract

We provide a unified explanation for a number of index option anomalies: 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. We explain these anomalies in terms of a pricing kernel that depends on variance. Although the pricing kernel is a monotonic function of stock return and variance, it is U-shaped in returns after projecting variance on returns. This non-monotonicity is supported by semi-parametric evidence from returns and option data. We incorporate this feature into the Heston-Nandi (2000) model and estimate the resulting model using a loss function with an options component and a returns component. The model significantly improves on the fit of a model with a traditional Black-Scholes-Rubinstein pricing kernel, and the differences between physical and risk-neutral moments are similar to those of an ad-hoc model that is designed to fit both distributions separately.

JEL Classification: G12

Keywords: Pricing kernel; stochastic volatility; overreaction; variance risk.

*Christoffersen and Jacobs want to thank FQRSC, IFM² and SSHRC for financial support. We are grateful to Gurdip Bakshi for helpful comments. Chayawat Ornthanalai, Mehdi Karoui, and Nick Pan provided expert research assistance. Correspondence to: Steven Heston, R.H. Smith School of Business, University of Maryland, 4447 Van Munching Hall, College Park, MD 20742; Tel: (301) 405-9686; E-mail: sheston@rhsmith.umd.edu.

1 Introduction

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. More than a decade ago, 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 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. The inability of existing pricing kernels to explain option prices may therefore suggest arbitrage, and more general pricing kernels that can capture these stylized facts are needed.

Moreover, several persistent empirical puzzles have emerged from the options literature. The most often discussed puzzle is that volatilities implied from option prices tend to exceed realized volatility. This puzzle is well-known and understood in terms of a negative price of variance risk. For instance, Bakshi and Kapadia (2003) show that average returns on variance-sensitive option portfolios are indeed negative. Another variance puzzle is the expectations puzzle, which concerns the relationship of implied volatility to expected future volatility. Generally, implied variances do not provide an unbiased forecast of subsequent variance.¹ Moreover, Stein (1989) shows 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 not necessarily explicitly linked them to Bates’ 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 evidences 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 (2001), Bates (2008), and Bakshi, Madan, and Panayotov (2009).

Together the expectations puzzles, the overreaction puzzle, and the pricing kernel puzzle pose

¹See 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.

a collective challenge to option models. This paper attempts to provide a unified explanation for these puzzles by studying a pricing kernel that is more general than standard pricing kernels that are monotonic in market returns, and analyzing its implications for each of the stylized facts mentioned above.

We start by presenting new semi-parametric evidence on the conditional pricing kernel. The natural logarithm of the conditional pricing kernel appears to be a U-shaped function of returns, and that this function is relatively stable over time. We address this by incorporating variance risk in the pricing kernel. In our model, the pricing kernel specification is monotonic in returns and also monotonic in volatility. The key is that volatility tends to be high when the stock return is very large or very negative. Consequently the pricing kernel is a U-shaped function of the stock return after projecting the volatility effect onto returns. The pricing kernel also explains the other anomalies in terms of a volatility risk premium. A negative risk 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) risk 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 therefore can explain why long-maturity implied volatilities appear to overreact to temporary increases in volatility, thereby explaining Stein's (1989) observation.

These analytical 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, as well as to verify 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), we conduct an empirical analysis that uses an objective function with a return component and an option component. We use an option dataset that is substantially larger than existing studies, maximizing a likelihood function that consists of the sum of the return likelihood and a likelihood based on option prices. We generalize the Heston-Nandi (2000) GARCH model to include a more general pricing kernel that generates a variance premium. The variance risk premium contains two components: one is related to the equity risk premium, while the other is an independent component that originates in volatility preferences. The discrete-time structure of the model greatly facilitates the filtering problem, and makes it possible to maximize the likelihood with the return and option components. In this model the logarithm of the pricing kernel is a quadratic function of the market return.

We first estimate the model without imposing the restrictions between the physical and risk-neutral parameters, in order to have an ad-hoc benchmark that fits both components of the data as well as possible. We then estimate the model while imposing the restrictions from our 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

risk premium dramatically improves model fit compared to a traditional pricing kernel with equity risk only. The new model reduces valuation biases across strike price and maturity, and the resulting fit is reasonably close 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. While the estimated persistence of risk-neutral variance is larger than the persistence of physical variance, the difference is smaller than in the ad-hoc model, indicating that the new pricing kernel qualitatively captures this stylized fact, but quantitatively falls short. Presumably this is due to the fact that capturing variance persistence is not heavily weighted in the likelihood.

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 instance Bates (2008), Pan (2002), Benzoni, Collin-Dufresne, and Goldstein (2005), 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 non-monotonic pricing kernels. Rosenberg and Engle (2002) and Chernov (2003) document nonmonotonicities in pricing kernels using parametric assumptions on the underlying returns. Chabi-Yo (2009) uses Taylor series expansions of marginal utilities and documents nonmonotonicities after projecting on the market return. Brown and Jackwerth (2001) 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 (2009) show that the prices of S&P500 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 paper is organized as follows. Section 2 discusses a number of existing stylized facts and also presents new evidence on the shape of the conditional pricing kernel. Section 3 presents the theory of the variance risk premium and the corresponding pricing kernel. Section 4 presents a new discrete-time model incorporating this kernel, which is used in our empirical work. Section 5 presents the empirical results and Section 6 concludes. The Appendix collects proofs of propositions.

2 Stylized Facts in Index Option Markets

We start by documenting and analyzing a number of well-known and lesser-known stylized facts in option markets. We pay particular attention to the shape of the pricing kernel implied by option data. We first discuss the option and return data used in the empirical analysis.

2.1 Data

Our empirical analysis uses out-of-the-money S&P500 call and put options for the 1996-2004 period from OptionMetrics. 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, the optimization problems we conduct are very time-intensive. The size of our option sample is unique in the literature, and limiting the number of options reduces the computational burden. Picking one day per week is 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 OTM call, and when F/X is larger than one, the contract is an OTM put. The out-of-the-money put prices were converted into call prices using put-call parity. The sample includes a total of 21,391 option contracts with an average mid-price of \$28.42 and average implied volatility of 21.47%. The implied volatility is largest for the OTM put options, reflecting the well-known volatility smirk in index options. The average implied volatility term structure is roughly flat during the period.

Table 1 also presents descriptive statistics for the return sample. The return sample is from January 1, 1990 to December 31, 2005. 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 16.08%, is substantially smaller than the average option-implied volatility, at 21.47%. 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 December 31, 2004, which matches the option sample. In comparison to the 1990-2005 sample, the standard deviation is somewhat higher. Average returns, skewness and kurtosis in the subsample are very similar to the 1990-2005 sample.

2.2 Returns on Straddles

It is well-known that on average, risk-neutral volatility exceeds physical volatility.² Several authors have argued that the risk premium that explains this difference makes it interesting to short sell straddles.³ Figures 1 and 2 illustrate these stylized facts using the 1996-2004 option sample from Table 1. Figure 1 illustrates that risk-neutral volatility exceeds physical volatility, as filtered by a GARCH process. This stylized fact is robust to a large number of variations in the empirical setup, such as for instance measuring the physical volatility using a different filter, or using realized volatility instead of GARCH volatility. Figure 2 illustrates the returns and cumulative returns of a short straddle strategy, which for simplicity are computed using the nearest to at-the-money nearest to 30-day maturity call and put option on the third Friday of every month. The options are held until maturity, the cash accounts earns the risk-free rate, and the index starts out with \$100 in cash on January 1, 1996. The dashed line in Figure 2 plots the S&P500 monthly closing price normalized to 100 in January 1996 for comparison. It is obvious from Figure 2 that the short straddle strategy was very rewarding in the 1996-2004 period, especially in periods when the S&P500 performed well. 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.

2.3 The Overreaction Hypothesis

Stein (1989) documents another stylized fact in option markets that is equally robust, but 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

$$E_t [(IV_{t+(LT-ST)}^{ST} - IV_t^{ST}) - 2(IV_t^{LT} - IV_t^{ST})] = 0, \quad (1)$$

where IV_t^{LT} is the implied volatility of a long-term option and IV_t^{ST} 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, then future implied volatilities tend to be

²See for instance Bates (2000, 2003), Broadie, Chenov, and Johannes (2007), Chernov and Ghysels (2000), Eraker (2004), Heston and Nandi (2000), Jones (2003), and Pan (2002).

³See among others Coval and Shumway (2001), Bondarenko (2003), and Driessen and Maenhout (2007).

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 (1), using weekly time series of one-month and two-month implied volatilities. The regression is

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

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

Table 2 presents the results for the Stein regression using the 1996-2004 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 contracts 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.

Table 2 demonstrates convincingly just how robust Stein’s results are. We run the regressions first for the full sample 1996-2004, and subsequently for nine sub-samples, one for each of the years in the sample. We find a highly significant negative sign in all ten 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 that this robust stylized fact does not signal an anomaly. In fact, it is entirely consistent with rational behavior, and moreover it is perfectly consistent with a wealth of other existing empirical evidence. We discuss this further below when we introduce the model.

2.4 Fat Tails and Fatter Tails

The most challenging outstanding puzzle about equity index option markets concerns the relation of the risk-neutral distribution of the stock price to the physical distribution. Stochastic volatility models emerged as workhorses in the option literature because they were able to model fat tails and volatility smiles and smirks.⁴ However, Bates (1996b, 2003), Ait-Sahalia and Lo (1998, 2000), and Jackwerth (2000) document that the risk-neutral distribution has fatter tails than the empirical distribution. Many studies attempt to address this stylized fact with increasingly complex models for the underlying dynamic, including non-normal innovations, jump processes,

⁴See Hull and White (1987), Melino and Turnbull (1990), Wiggins (1987), and Heston (1993) for examples of option valuation with stochastic volatility.

and multiple factors, but there seems to be consensus in the literature that these models do not necessarily resolve the problem that the fatness of the tails dramatically differs between the physical and risk-neutral distribution.⁵ Despite these recent innovations in the modeling of underlying index returns, we do not have a model that convincingly captures the stylized facts of the physical as well as the risk-neutral distributions.

In order to capture this stylized fact, what is needed is risk-neutral conditional and unconditional skewness and kurtosis that exceed their physical counterparts. While it may be possible to do so with a more general or flexible return specification, we argue in this paper that it is more productive to directly focus on the pricing kernel that links the physical and risk-neutral distributions.

Our strategy is to guide the search for the appropriate pricing kernel by another robust stylized fact, which has not yet been extensively explored in the literature. We document that the conditional pricing kernel is U-shaped as a function of returns. The logic underlying the Capital Asset Pricing Model explains why the pricing kernel is high when stock index returns are low, but is incompatible with a non-monotonic stochastic discount factor. Yet the logic underlying the Capital Asset Pricing Model is implicitly present in almost all existing empirical studies of index option pricing.

We therefore present an option pricing model with a modified pricing kernel. The underlying motivation and logic are simple. In the presence of stochastic volatility, the stochastic discount factor is an exponential-linear function of the variance. In contrast to the price of equity risk, we anticipate the price of volatility risk to be negative. In this case the stochastic discount factor will be an exponentially increasing function of future variance. In other words the bivariate stochastic discount factor is monotonic in both the stock return and in the variance. However, to address the Bates (1996a) puzzle, we turn our attention to the projection of the stochastic discount factor on the univariate stock return. When the variance is large, stock index returns tend to be very high or very low. This means that the projection of the variance risk premium component of the stochastic discount factor onto the stock returns may be a U-shaped function of the stock return. We document that this is indeed the case empirically, and that allowing for this possibility radically improves the model's ability to jointly fit the physical and the risk-neutral distribution.

⁵Andersen, 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 (2009), Carr and Wu (2004) and Huang and Wu (2004) investigate infinite-activity Levy processes. Bates (2000), Christoffersen, Heston, and Jacobs (2009) and Christoffersen, Jacobs, Wang, and Ornthanalai (2008) investigate multifactor stochastic volatility models. Christoffersen, Heston, and Jacobs (2006) propose heteroskedastic models with non-normal innovations.

It is useful to contrast this underlying logic with the intuition underlying stochastic volatility models, which are able to accommodate fat tails. High volatility tends to be associated with very high or very low stock returns. Therefore stochastic volatility models tend to produce fat tails, but they do so for both the physical and the risk-neutral distribution of stock returns. This produces the familiar “volatility smile” in option values relative to the Black-Scholes formula. Our approach does not merely address the stylized fact of fat tails. It explains why the tails of the risk-neutral distribution are fatter than the tails of the physical distribution. In other words, it explains the magnitude of the risk-neutral smile, and decomposes it into a component that is also present in the tails of the physical volatility and a component that is explained by the pricing kernel.

2.5 Retrieving Conditional Densities and Pricing Kernels

We now document the shape of the conditional pricing kernel using semiparametric methods. 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. Several papers 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, p. 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 (2009) 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 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 non-parametric estimates of the risk-neutral density or state-price density. This necessitates combining option data on different days, because non-parametric methods are very data intensive. Other papers, such as Jackwerth and Rubinstein (1996), 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 polynomial function of implied Black-Scholes volatility as a function of moneyness and maturity. Using this

estimated polynomial, we then generate a grid of at-the-money implied volatilities for 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.

$$\hat{C}(S(t), X, \tau, r) = C_{BS}(S(t), X, \tau, r; \hat{\sigma}(S(t), X, \tau)). \quad (2)$$

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

$$\hat{f}_t^*(S(T)) = \exp(r) \left[\frac{\partial^2 \hat{C}(S(t), X, \tau, r)}{\partial X^2} \right]_{|X=S(T)}. \quad (3)$$

We calculate this derivative numerically 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

$$\hat{f}_t^*(R(t, T)) = \frac{\partial}{\partial u} \Pr \left(\ln \left(\frac{S(T)}{S(t)} \right) \leq u \right) = S(t) \exp(u) \hat{f}_t^*(S(t) \exp(u)). \quad (4)$$

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 worried about the conditional nature of the estimates. Jackwerth (2000) uses one month 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, which is used in the empirical work, but the method can easily be applied for 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 30, 2005. A kernel-smoothed 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, $\hat{f}_t(R(t, T))$, by normalizing the return series by dividing by the conditional volatility on that day, as measured by the VIX.

A subset of the resulting physical and risk-neutral conditional densities are given in Figure

3. Recall that our sample consists of nine years worth of option data, for 1996-2004, 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. In order to show the time variation in the conditional densities, and the appeal of our method, Figure 3 presents nine physical and nine risk-neutral conditional densities, one for the first Wednesday of each year in our sample. The sample year is indicated in the title to each graph. The horizontal axis indicates annualized log returns.

These results are interesting from two perspectives. First, they illustrate that the conditional densities significantly change through time. The shapes of both the physical and the risk-neutral densities vary substantially over the years. Second, given these changes, it is remarkable that for each of the nine cases, the physical and the risk neutral density track each other rather closely. Given that estimation of the risk-neutral conditional density is straightforward, this result indicates the appeal of our simple estimation procedure for the physical conditional density.

Figure 3 clearly demonstrates the fat left tail of the risk-neutral conditional density, compared with that of the physical density. This finding is robust despite the fact that the conditional densities look very different across the years. This stylized fact gives rise to risk neutral model estimates that display excess kurtosis and excess negative skewness in comparison to physical estimates.

Figure 4 depicts the natural logarithm of the ratio of the weekly conditional 1-month risk-neutral and conditional physical density. We want to investigate the natural logarithm of the pricing kernel at different levels of return. As in Figure 3, we present nine sets of results, one for each year of the sample. Recall that in Figure 3 we only present results for the first week of each year, in order to illustrate the time-varying nature of the conditional density. Plotting the densities for all 52 weeks in a given year would make the figure unwieldy. In Figure 4, because the densities move together, we are able to present more information and plot results for all weeks of the year on each picture. Specifically, we plot

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

In each week we trimmed 3% of observations in the left and right tails, because these observations are sometimes very noisy. In order to let the data speak we did not kernel-smooth the historical return distributions as in Figure 3.

Three 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 natural logarithm of the pricing kernel can be well

approximated by a quadratic function of stock returns, supporting the approach taken in this paper. Third, 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. For instance, the 1998 kernel is different from the 1996 kernel, and by 2004 the kernel again looks similar to the 1996 kernel. But we are able to draw the fifty-two pricing kernels generated for a given year on one picture to clearly illustrate the quadratic nature of the kernel. If the kernel varied more within the year, Figure 4 would contain nothing but a cloudy scatterplot without much structure. In summary, Figure 4 clearly illustrates that the logarithm of the pricing kernel is roughly quadratic as a function of the return, and that this pricing relationship is stable over time.

3 Theory of the Variance Risk Premium

Consider the Heston (1993) model for the dynamics of the spot price $S(t)$

$$\begin{aligned} dS(t) &= (r + \mu v(t))S(t) dt + \sqrt{v(t)}S(t) dz_1(t), \\ dv(t) &= \kappa(\theta - v(t))dt + \sigma\sqrt{v(t)}\left(\rho dz_1(t) + \sqrt{1 - \rho^2}dz_2(t)\right), \end{aligned} \quad (5)$$

where r is the risk-free interest rate and where $z_1(t)$ and $z_2(t)$ are independent Wiener processes. The notation in (5) emphasizes the separate sources of equity risk, $z_1(t)$, and independent volatility risk, $z_2(t)$. An important aspect of our analysis will be the separate premia for these risks.

In this model, the instantaneous variance $v(t)$ reverts to a long-run mean of θ with a speed of κ . The expected future variance is a linear function of current variance

$$E_t(v(t + \Delta)) = \exp^{-\kappa\Delta} v(t) + (1 - \exp^{-\kappa\Delta})\theta. \quad (6)$$

Most of the existing literature follows Heston (1993) and specifies the following risk-neutral dynamic

$$\begin{aligned} dS(t) &= rS(t) dt + \sqrt{v(t)}S(t) dz_1^*(t), \\ dv(t) &= (\kappa(\theta - v(t)) - \lambda v(t))dt + \sigma\sqrt{v(t)}(\rho dz_1^*(t) + \sqrt{1 - \rho^2}dz_2^*(t)), \end{aligned} \quad (7)$$

where $z_1^*(t)$ and $z_2^*(t)$ are independent Wiener processes under the risk-neutral measure. The variance process is usually re-written as

$$dv(t) = (\kappa^*(\theta^* - v(t)))dt + \sigma\sqrt{v(t)}(\rho dz_1^*(t) + \sqrt{1 - \rho^2}dz_2^*(t)), \quad (8)$$

where $\theta^* = \kappa\theta/(\kappa + \lambda)$ and $\kappa^* = \kappa + \lambda$. Therefore θ^* is the risk-neutral long-run variance, and

κ^* is the risk-neutral mean reversion. The risk-neutral specification depends on the separate premium for volatility risk of the form $\lambda v(t)$. Note that this risk premium may be negative. Our focus is on the components and empirical implications of this variance risk premium.

We begin the analysis by characterizing the stochastic discount factor $M(t)$ underlying this risk-neutralization. Proposition 1 expresses the stochastic discount factor in terms of the parameters governing the dynamics of variance, κ , θ , ρ , and σ , the equity premium μ , and the variance premium λ .

Proposition 1 *The unique stochastic discount factor inherent in equations (5) and (7) takes the form*

$$M(t) = M(0) \left(\frac{S(t)}{S(0)} \right)^\phi \exp \left(\delta t + \eta \int_0^t v(s) ds + \xi (v(t) - v(0)) \right), \quad (9)$$

where

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

Proof. See Appendix. ■

When variance is constant, (9) amounts to the familiar power utility from Rubinstein's (1976) and Brennan's (1979) preference-based derivation of the Black-Scholes model. But the stochastic discount factor also depends on the variance $v(t)$.⁶ This has distinctive implications for option valuation.

Several papers have noted the importance of the stochastic discount factor for pricing options. Chernov and Ghysels (2000) emphasized that the risk-neutral specification must be incompatible with arbitrage. Bates (2006), Pan (2002), Liu, Pan, and Wang (2004), Broadie, Chernov, and Johannes (2007), and Santa-Clara and Yan (2009) have incorporated increasingly complex model features, with separate components in the stochastic discount factor for stochastic volatility and jumps. The purpose of this paper is to derive distinctive implications for option values from a comparatively simple stochastic discount factor. In particular, the stochastic discount factor in equation (9) can explain a number of outstanding option anomalies, even in a single-factor diffusion model or a discrete model with Gaussian innovations.

⁶This could result, for instance, if $v(t)$ governs the variance of aggregate production in a Cox-Ingersoll-Ross (1985) model with non-logarithmic utility. It could also result from the model of Benzoni, Collin-Dufresne, and Goldstein (2009) where uncertainty directly affects preferences. Bakshi, Madan, and Panayotov (2009) relate it to short-sale constraints.

Proposition 1 mechanically illustrates the restrictions imposed on the parameters of the stochastic discount factor (9) by the choice of physical dynamic (5) and the choice of risk premium in (7). The risk-free rate r restricts the time-preference parameters δ and η . The equity premium parameter μ determines the risk aversion parameter ϕ , and the variance premium λ determines the variance preference ξ . It is therefore also instructive to re-arrange Proposition 1 to express the equity premium μ and variance premium λ parameters in terms of the underlying preference parameters ϕ and ξ . This gives

$$\begin{aligned}\mu &= -\phi - \xi\sigma\rho, \\ \lambda &= \rho\sigma\mu - (1 - \rho^2)\sigma^2\xi \\ &= -\rho\sigma\phi - \sigma^2\xi.\end{aligned}\tag{10}$$

This demonstrates that both the equity risk premium μ and the variance risk premium λ can be thought of as having two distinct components originating in preferences. One is related to the risk-aversion parameter ϕ and the other one to the variance preference parameter ξ . We can therefore use economic intuition to sign the equity premium and the variance premium. If the stochastic discount factor 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 stochastic discount factor to be increasing in the volatility, $\xi > 0$. Empirically the correlation between stock market returns and variance ρ is strongly negative. Therefore from (10) the equity premium μ must be positive. The variance premium λ has a component based on covariance with equity risk, and a separate independent component based on the variance preference ξ . With a negative correlation ρ , we see that λ must be negative.

It is critical to note that the conventional assumption on the risk-neutral dynamics of variance (8) does not distinguish whether the variance risk premium emanates exclusively from ϕ (and therefore indirectly from the equity premium μ) or whether it has an independent component ξ . In other words, assuming $\xi = 0$ in (9) is consistent with a nonzero variance risk premium λ , as can be seen from (10). 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 (5) and the risk-neutral dynamic (8). It is not sufficient to simply state that (7) holds for arbitrary (negative) λ , because this assumption is consistent with the stochastic discount factor (9) but also with the special case (9) with $\xi = 0$, and the economic implications of those sets of assumptions are very different. This paper explores the distinct implications of variance premiums $\xi \neq 0$ for option prices.

We will investigate the pricing implications of these alternative assumptions on the stochastic discount factor below in the context of a discrete-time framework, which has similar intuition

but is easier to implement. For now it is instructive to relate our framework to the empirical findings in Section 2.5. If there is no independent component to the variance premium ($\xi = 0$), the ratio of the risk-neutral stock price density to the true density will be linear with a slope of ϕ when plotted in log-log space. This corresponds to the conventional power utility assumption. If there is an independent negative component to the variance premium ($\xi > 0$), the ratio of the risk-neutral stock price density to the true density will be a U-shaped “smile”. In contrast if $\xi < 0$ then this ratio will have a hump shaped “frown”. The evidence in Figure 4 in Section 2.5 therefore favors $\xi > 0$. We analyze this issue further below in the context of our discrete-time framework.

Finally, note that the risk-neutral expected variance is obtained by inserting the risk-neutral parameters of equation (8) into equation (6)

$$E_t^*(v(t + \Delta)) = \exp^{-\kappa^* \Delta} v(t) + (1 - \exp^{-\kappa^* \Delta}) \theta^*. \quad (11)$$

Recall that the risk-neutral long-run variance $\theta^* = \kappa\theta/(\kappa + \lambda)$ and the risk-neutral mean-reversion $\kappa^* = \kappa + \lambda$. The variance premium affects both quantities. Given that we know that $\lambda < 0$, the implications are that the risk-neutral expected variance will exceed the actual expected variance, and that the risk-neutral mean-reversion will be smaller than the physical mean reversion. Hence a negative risk premium for variance explains not only why implied volatility exceeds physical volatility, but also Stein’s (1989) finding of overreaction of long-term options to short-term fluctuations in volatility.

4 The Variance Risk Premium in Discrete Time

The objective of our empirical exercise is to investigate Bates’ (1996b) observation that the differences between the physical and risk-neutral distribution are the achilles’ heel of state-of-the-art option valuation models, which still applies today. Estimation using option data only and/or returns only can diagnose this problem, but it cannot assess a model’s ability to resolve it. It is therefore critically important to estimate and evaluate models using loss functions that include returns data as well as options data.

While it is feasible to do this using the Heston (1993) model in the previous section, or one of its extensions, the computational complexity involved when implementing this model is very high. While continuous-time models are often better suited to provide theoretical insights, discrete-time models may be easier to estimate using daily data. This section introduces the variance-dependent stochastic discount factor (9) into the Heston and Nandi (2000) GARCH model. This generates an additional premium for variance risk analogous to similar to the premium in Heston’s

(1993) square-root volatility model. The model mechanics can be empirically investigated using either the continuous-time model in the previous section or the model below. We use the discrete-time model because we present a rather computationally-intensive empirical exercise, and the discrete model is easier to implement. One may interpret the GARCH model as either an internally consistent discrete model or a daily approximation of the continuous-time model.

We begin with the Heston-Nandi (2000) physical GARCH process

$$\begin{aligned}\ln(S(t)) &= \ln(S(t-1)) + r + (\mu - \frac{1}{2})h(t) + \sqrt{h(t)}z(t), \\ h(t) &= \omega + \beta h(t-1) + \alpha(z(t-1) - \gamma\sqrt{h(t-1)})^2,\end{aligned}\tag{12}$$

where r is the daily continuously compounded interest rate and $z(t)$ has a standard normal distribution. Heston and Nandi (2000) show this process is a discrete analog of the square-root volatility process and can converge to the diffusion process (5). We will implement this model using daily data, and we are therefore interested in its predictions for a fixed daily interval.

Just as in the diffusion model (5), the expected future variance is a linear function of current variance

$$E_{t-1}(h(t+1)) = (\beta + \alpha\gamma^2)h(t) + (1 - \beta - \alpha\gamma^2)E(h(t)),\tag{13}$$

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.

$$Var_{t-1}(h(t+1)) = 2\alpha^2 + 4\alpha^2\gamma^2h(t).\tag{14}$$

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

$$Cov_{t-1}(R(t), h(t+1)) = -2\alpha\gamma h(t)\tag{15}$$

The data robustly indicate sizeable negative correlation, which means that γ must be positive.

We can value securities in this discrete-time model using the same stochastic discount factor (5) as the diffusion model. Recall that in the diffusion model, the variance process follows square-root dynamics with different parameters in the physical and risk-neutral measures. Proposition 2 shows an analogous result in the discrete model—the risk-neutral process remains in the same GARCH class.

Proposition 2 *The risk-neutral stock price corresponding to the physical Heston-Nandi GARCH*

process in (12) and the stochastic discount factor

$$M(t) = M(0) \left(\frac{S(t)}{S(0)} \right)^\phi \exp \left(\delta t + \eta \sum_{s=1}^t h(s) + \xi (h(t+1) - h(1)) \right) \quad (16)$$

follows the GARCH process

$$\begin{aligned} \ln(S(t)) &= \ln(S(t-1)) + r - \frac{1}{2}h^*(t) + \sqrt{h^*(t)}z^*(t), \\ h^*(t) &= \omega^* + \beta h^*(t-1) + \alpha^*(z^*(t-1) - \gamma^* \sqrt{h^*(t-1)})^2, \end{aligned} \quad (17)$$

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

$$\begin{aligned} h^*(t) &= h(t) / (1 - 2\alpha\xi), \\ \omega^* &= \omega / (1 - 2\alpha\xi), \\ \alpha^* &= \alpha / (1 - 2\alpha\xi)^2, \\ \gamma^* &= \gamma - \phi. \end{aligned} \quad (18)$$

Proof. See Appendix. ■

The risk-neutral dynamics differ from the physical dynamics through the effect of the equity premium parameter μ 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 ϕ and ξ in equation (16).⁷ The intuition is similar to the continuous-time case in Section 3, where the values of the equity premium and volatility risk premium parameters μ and λ are implied by the values of the kernel parameters ϕ and ξ . 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 μ is small, around 2.

The Heston-Nandi (2000) model corresponds to the special case of $\xi = 0$. It can be seen from (18) that the implications of this parameter are important because it influences the level, persistence, and volatility of the variance. Note that a negative variance risk premium ($\xi > 0$) implies that the risk-neutral variance $h^*(t)$ exceeds the physical variance $h(t)$ in the empirically relevant case where $\alpha > 0$. The variance risk premium also affects the risk-neutral drift of $h^*(t)$

$$E_{t-1}^*(h^*(t+1)) = (\beta + \alpha^* \gamma^{*2})h^*(t) + (1 - \beta - \alpha^* \gamma^{*2})E^*(h^*(t)), \quad (19)$$

where $E^*(h^*(t)) = (\omega^* + \alpha^*) / (1 - \beta - \alpha^* \gamma^{*2})$. The risk-neutral autocorrelation equals $\beta + \alpha^* \gamma^{*2}$,

⁷The mapping between μ and ϕ is contained in the Appendix.

and a positive equity premium 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 ($\gamma > 0$), if the equity premium is positive ($\mu > 0$, which corresponds to $\phi < 0$) and if 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

$$Var_{t-1}^*(h^*(t+1)) = 2\alpha^{*2} + 4\alpha^{*2}\gamma^{*2}h^*(t). \quad (20)$$

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 α^* and γ^* from (17) shows that risk-neutral variance of variance is greater than the actual variance of variance. Furthermore we can define the risk neutral conditional covariance

$$Cov_{t-1}^*(R(t), h^*(t+1)) = -2\alpha^*\gamma^*h^*(t) \quad (21)$$

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

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:*

- *The risk-neutral variance $h^*(t)$ exceeds the physical variance $h(t)$,*
- *The risk-neutral expected future variance exceeds the physical expected future variance,*
- *The risk-neutral variance process is more persistent than the physical process, and*
- *The risk-neutral variance of variance exceeds the physical variance of variance.*

The corollary summarizes how a risk premium for volatility can explain a number of puzzles concerning the level and movement of implied option variance compared to 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, which corresponds to risk-neutral skewness and kurtosis exceeding physical skewness and kurtosis. Our model captures this stylized fact through a U-shaped pricing kernel. Interestingly, even though the stochastic discount factor is a monotonic function of the stock price and variance (9), the projection of the stochastic discount factor onto the stock price alone can have a U-shape. The following proposition formalizes this relationship.

Corollary 2 *The logarithm of the stochastic discount factor is a quadratic function of the stock return.*

$$\ln \left(\frac{M(t)}{M(t-1)} \right) = \frac{\xi \alpha}{h(t)} (R(t) - r)^2 - \mu (R(t) - r) + \left(\eta + \xi (\beta - 1) + \xi \alpha \left(\mu - \frac{1}{2} + \gamma \right)^2 \right) h(t) + \delta + \xi \omega + \phi r, \quad (22)$$

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

In 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 ξ , not on the total variance premium. Due to the component of variance premium that is correlated with equity risk, it is conceivable that the total variance premium could have a different sign than the independent negative component. A negative independent variance premium ($\xi > 0$) corresponds to the empirical smile shown in Figure 4. Thus the strong U-shaped ratios in Figure 4 constitute 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.*

In summary, option values should reflect an implied variance process that is larger, more persistent, and more volatile than observed variance. Then the 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 risk therefore explains a number of puzzles concerning the relationship between physical volatility and volatilities implied from option prices.

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

$$C(S(t), h(t+1), X, T) = S(t) \left(\frac{1}{2} + \int_{\ln(X)}^{\infty} \operatorname{Re} \left[\frac{X^{-i\varphi} g_t^*(i\varphi + 1)}{i\varphi} \right] d\varphi \right) - X \exp^{-r(T-t)} \left(\frac{1}{2} + \int_{\ln(X)}^{\infty} \operatorname{Re} \left[\frac{X^{-i\varphi} g_t^*(i\varphi)}{i\varphi} \right] d\varphi \right). \quad (23)$$

where $g_t^*(\cdot)$ is the conditional generating function for the risk neutral process (17). The functional form for the physical $g_t(\cdot)$ is provided in Heston and Nandi (2000). Together with the mapping in (18) this yields $g_t^*(\cdot)$.

5 Empirical Results

We now present an empirical investigation of the model outlined in Section 4. It is important to realize that the model’s success in addressing some of the stylized facts we discuss in Section 2 can only be evaluated 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 only been attempted 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-section of options while side-stepping the issue of return fit completely by parameterizing the volatility state variable.⁸ 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,⁹ 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.¹⁰ Eraker (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 which 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 (2009) estimate parameters using a likelihood which contains a returns component and an options component.

The literature also contains comparisons of the risk-neutral and physical distribution. 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.

Our empirical setup is most closely related to Santa-Clara and Yan (2009). We use a joint likelihood consisting of an option-based component and a return-based component which is relatively easy in discrete time. Note that the conditional density of the daily return is normal so

⁸See for instance the seminal paper by Bakshi, Cao and Chen (1997)

⁹See for instance Bates (2000), Christoffersen, Heston, and Jacobs (2009), and Huang and Wu (2004).

¹⁰See for instance Christoffersen and Jacobs (2004).

that

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

The return log likelihood is therefore

$$\ln L^R \propto -\frac{1}{2} \sum_{t=1}^T \{\ln(h(t)) + (R(t) - r - \mu h(t))^2 / h(t)\}. \quad (24)$$

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

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

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 that the option log likelihood is

$$\ln L^O \propto -\frac{1}{2} \sum_{i=1}^N \{\ln(s_\varepsilon^2) + \varepsilon_i^2 / s_\varepsilon^2\}. \quad (25)$$

where we can concentrate out s_ε^2 using the sample analogue $\hat{s}_\varepsilon^2 = \frac{1}{N} \sum_{i=1}^N \varepsilon_i^2$. 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. See for instance Carr and Wu (2007) and Trolle and Schwartz (2009) for applications of BSV^{Mkt} weighted option errors.

We can now solve the following optimization problem

$$\max_{\Theta, \Theta^*} \ln L^R + \ln L^O, \quad (26)$$

where $\Theta = \{\omega, \alpha, \beta, \gamma, \mu\}$ denotes the physical parameters and Θ^* denotes the risk-neutral parameters which are mapped from Θ using (18). The riskless rate r in (24) set to 5 percent, and we use the term structure of interest rates from OptionMetrics when pricing options in (25).

To demonstrate the usefulness and implications of the stochastic discount factor (9), we conduct four different empirical exercises. The first exercise is intended as a benchmark. We maximize the joint log likelihood in (26) with respect to nine parameters: the five physical parameters $\omega, \alpha, \beta, \gamma, \mu$ as well as the four risk-neutral parameters $\omega^*, \alpha^*, \beta^*, \gamma^*$. This exercise

has no value from an economic perspective, it is merely a fitting exercise. We refer to it as the ad-hoc model. Because there are no built-in restrictions between the physical and risk-neutral parameters, we effectively fit both the risk-neutral and physical distribution as well as possible, which will serve as a benchmark for three other models which impose economic restrictions between the physical and the risk-neutral distribution.

The first of the three models we consider contains no risk premia. This amounts to setting $\mu = 0$ and $\xi = 0$, so that we have $h^*(t+1) = h(t+1)$. The resulting model has only four free parameters: ω, α, β , and γ . The second model is the Heston-Nandi (2000) model that allows for equity risk. This effectively leaves the equity premium μ (or equivalently the risk-aversion ϕ) as a free parameter, yielding five parameters in total. The most general economic model allows both equity risk and independent volatility risk by allowing ξ to be a free parameter.

Table 3 presents the empirical results, which are quite striking.¹¹ First, consider the likelihood functions at the optimum for the four different models. Notice that at the optimum, the contribution from the options part of the likelihood is equal 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, and by implication sacrificing some goodness of fit in the return component of the likelihood, which is in all cases smaller than in the ad-hoc specification. This occurs despite using a return sample, 1990-2005, which is longer than the option sample, 1996-2004.

It is instructive to start by inspecting some of the properties of the ad-hoc model, which does not impose restrictions across the physical and risk-neutral parameters. The resulting properties are therefore entirely determined by the data, and serve as a useful benchmark for the three other models. We observe four very important features. First, the average risk-neutral volatility is much higher than the average physical volatility. This captures the first stylized fact discussed in Section 1. Second, the risk-neutral variance persistence, 0.985, is higher than the physical variance persistence, 0.962. This is the stylized fact underlying Stein's overreaction regressions. Third, the average annualized risk neutral volatility of variance in (20) is much larger than the physical volatility of variance in (14). Fourth, (the absolute value of) the average of the risk neutral leverage correlation defined as

$$\frac{Cov_{t-1}^*(R(t), h^*(t+1))}{\sqrt{h^*(t)Var_{t-1}^*(h^*(t+1))}} = \frac{-2\alpha^*\gamma^*h^*(t)}{\sqrt{h^*(t)(2\alpha^{*2} + 4\alpha^{*2}\gamma^{*2}h^*(t))}}$$

annualized and evaluated at $h^*(t) = E[h^*(t)]$, exceeds its physical counterpart. The last two empirical features are of course critical in ensuring that the risk-neutral distribution has fatter

¹¹We impose $\omega^* = \omega = 0$ in estimation because the nonnegativity constraint is binding and it is a necessary condition for positive variances.

tails than the physical distribution.

The model without risk premia serves as another benchmark. The model’s physical properties are identical to its risk-neutral properties. Allowing a nonzero equity premium results in a rather small increase in the total likelihood (from 55,366.1 to 55,368.4). 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 modest, as the risk-neutral average volatility, variance persistence, and volatility of variance are not very different from their physical counterparts. While these quantitative effects are small, it is reassuring that qualitatively all effects go in the expected direction, as emphasized in Section 3. 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 source of volatility risk to the specification, the likelihood function improves spectacularly (from 55,368.4 to 55,485.5). Perhaps even more pertinently, some of the model properties for the model with equity and volatility risk are very similar to those of the ad-hoc model, which provides benchmark properties that are completely data-driven. The reported parameter combination $1/(1 - 2\alpha\xi)$ shows the ratio of the risk-neutral variance $h^*(t)$ to the physical variance $h(t)$. The difference between the long-run physical and risk neutral volatilities is almost exactly the same as in the ad-hoc model, and the physical volatility of variance is a very good approximation of the ad-hoc benchmark. However, while the physical leverage correlation is smaller (in absolute value) than the risk-neutral leverage correlation, and even though the physical variance persistence is smaller than the risk-neutral variance persistence, the difference between the two persistence measures is not nearly large enough. The model therefore explains only part of the Stein puzzle. To understand why, note that the model’s parsimonious specification forces the same parameters to capture a large number of stylized facts. In maximizing the likelihood, the relative fit of these stylized facts is traded off, and it is well-known that it is difficult to precisely estimate the mean-reversion of highly persistent processes. The procedure will put weight on capturing the second, third and fourth moments, and Nelson and Foster (1994) show that the covariance between stock returns and variance is particularly important in this regard. Adding components to the GARCH model would have increased flexibility to fit these dynamics as in Christoffersen et al (2008).

Figure 5 further explores the models’ performance in matching the physical and risk neutral volatility of variance. It plots the square root of the risk-neutral $Var_{t-1}^*(h^*(t + 1))$ and the physical counterpart $Var_{t-1}(h(t + 1))$ over time for the four specifications in Table 3. Note that the internally consistent model with equity and volatility risk in the bottom left panel tracks quite closely the ad-hoc model in the bottom right panel. The risk-neutral volatility of variance exceeds its physical counterpart in both cases. In contrast, in the top two panels the models

without volatility risk have virtually identical physical and risk-neutral volatility of variance.

Table 4 explores further the option valuation performance of the new model that allows for equity and volatility risk by providing measures of the IV RMSE and Bias by moneyness and maturity. The first row in each panel of Table 4 is marked “ERP and VRP” and reports the IV RMSE and Bias corresponding to the parameters estimated in the column labeled “Equity and Volatility Risk” in Table 3. The second row in Table 4 labeled “ERP only” uses the same physical parameters as the first row but forces $\xi = 0$ so that only the equity risk premium (ERP) remains. The third row in each panel labeled “No RP” further forces $\mu = \xi = 0$ thus eliminating both risk premia but again keeping all other physical parameter values from the first two rows.

Comparing the first and second rows in Table 4 shows that the volatility risk premium offers a dramatic improvement in option fit—both in terms of IV RMSE and Bias. When comparing the second and third rows we see that the equity risk premium plays a much smaller role in improving option fit when comparing with a model with no risk premia. The parameterizations with no volatility risk premia imply a strong positive bias implying that on average the models underprice options when the volatility risk premium is excluded. This bias is virtually eliminated—and the RMSE is radically improved—when the volatility risk premium is incorporated into the model.

6 Conclusion

We suggest a general pricing kernel for the purpose of index option valuation. 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. This model feature is consistent with semi-parametric evidence from returns and options which reveals that the conditional pricing kernel is U-shaped in returns, and is relatively stable over time. We then demonstrate that this pricing kernel can account for a number of puzzles in the option pricing literature, using a negative price of variance risk.

We estimate a generalization of the Heston-Nandi (2000) option pricing model that incorporates the negative price of variance risk into the pricing kernel. 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 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 two im-

portant differences between physical and risk-neutral moments are very similar to the differences obtained for the ad-hoc model.

The results and empirical exercises in this paper can be 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. Alternatively, it may prove interesting to investigate the implications of more general pricing kernels for option valuation in the presence of richer return dynamics, for instance with jumps in returns and/or volatility, or multiple volatility components. 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, whether these models can improve option fit; on the other hand, whether they can describe 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.

7 Appendix

Proof of Proposition 1

We impose the condition that the product of any traded asset and the stochastic discount factor 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 (9) we get the dynamic of $M(t)$

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

This gives

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

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 (5). Note that we have suppressed dependence on t in the notation.

Again by Itô's lemma, we have

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

From the condition that $B(t)M(t)$ is a martingale, i.e. its drift is equal to zero, we deduce the restrictions on δ and η .

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

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

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

Similarly, from the condition that $S(t)M(t)$ is a martingale, i.e. its drift is equal to zero, we deduce a restriction on ϕ . 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 ξ

$$\begin{aligned} 0 = & \left[M \left(U_t + r S U_S + \mu v S U_S + \kappa(\theta - v) U_v + \frac{1}{2} v S^2 U_{SS} + v S \sigma \rho U_{Sv} \right. \right. \\ & \left. \left. + \frac{1}{2} v \sigma^2 U_{vv} \right) - r U M + M v (\phi + \xi \sigma \rho) (S U_S + \sigma \rho U_v) + M \xi v \sigma^2 (1 - \rho^2) U_v \right] dt. \end{aligned} \quad (\text{A1})$$

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

$$\lambda v U_v + r U = U_t + r S U_S + \kappa(\theta - v) U_v + \frac{1}{2} v S^2 U_{SS} + v S \sigma \rho U_{Sv} + \frac{1}{2} v \sigma^2 U_{vv}. \quad (\text{A2})$$

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

$$\begin{aligned} D^{UM} &= [M (\lambda v U_v + r U + \mu v S U_S) - r U M \\ &\quad + M v (\phi + \xi \sigma \rho) (S U_S + \sigma \rho U_v) + M \xi v \sigma^2 (1 - \rho^2) U_v] dt \\ &= 0. \end{aligned}$$

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

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

and therefore

$$\xi = \frac{\mu\sigma\rho - \lambda}{\sigma^2(1 - \rho^2)}. \quad (\text{A3})$$

Using (A3) we also obtain

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

Proof of Proposition 2

The discrete-time form of the stochastic discount factor (9) is

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

and therefore

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

The summations in (A4) are equivalent to the integrals in the continuous-time form (9) under the standard GARCH convention that variance is constant throughout the day, and changes discretely overnight. We shall show that the stochastic discount factor in (A4) is consistent with the Heston-Nandi GARCH dynamic (12) for the following parameter mapping

$$\begin{aligned} \delta &= -(\phi + 1)r - \xi\omega + \frac{1}{2} \ln(1 - 2\xi\alpha), \\ \eta &= -(\mu - \frac{1}{2})\phi - \xi\alpha\gamma^2 + (1 - \beta)\xi - \frac{(\phi - 2\xi\alpha\gamma)^2}{2(1 - 2\xi\alpha)}, \\ \phi &= -(\mu - \frac{1}{2} + \gamma)(1 - 2\alpha\xi) + \gamma - \frac{1}{2}. \end{aligned}$$

From the GARCH dynamic (12) we can write

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

Substituting these into (A5) gives

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

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

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

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.

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

Note that

$$\begin{aligned} E_{t-1} \left[\frac{M(t)}{M(t-1)} \right] &= \exp \left(\phi r + \delta + \xi \omega + \left[\phi \left(\mu - \frac{1}{2} \right) + \eta + \xi (\beta - 1) + \xi \alpha \gamma^2 \right] h \right) \\ &\quad * E \left(\exp \left([\phi - 2\xi \alpha \gamma] \sqrt{h} z + [\xi \alpha] z^2 \right) \right). \end{aligned}$$

We need the following result

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

For our application we have

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

and thus

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

Therefore

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

and

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

Rearranging and using (A6)

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

Therefore we must have

$$\begin{aligned} \delta &= -(\phi + 1) r - \xi \omega + \frac{1}{2} \ln (1 - 2\xi \alpha), \\ \eta &= -(\mu - \frac{1}{2}) \phi - \xi \alpha \gamma^2 + (1 - \beta) \xi - \frac{(\phi - 2\xi \alpha \gamma)^2}{2(1 - 2\xi \alpha)}. \end{aligned}$$

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 ϕ is replaced by $\phi + 1$, thus we can use the expression for $E_{t-1} \left[\frac{M(t)}{M(t-1)} \right]$ to write

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

Taking logs, setting equal to zero and using the above solutions for δ and η gives

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

Solving this for ϕ yields

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

To find the risk-neutral dynamic, note that the risk-neutral density is proportional to the physical density times the stochastic discount factor

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

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

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

The risk-neutral dynamics ((17)) can then be derived by substituting the risk-neutral innovation (A7) into the physical GARCH process (12).

Proof of Corollary 2

We want to show that the logarithm of the stochastic discount factor takes the following quadratic form in the stock return

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

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

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

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

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

so that

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

From this we get

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

expanding the square and collecting terms yields

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

From the equation for ϕ we have

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

and the result obtains.

References

- [1] Ait-Sahalia, Y., and A. Lo (1998), Nonparametric Estimation of State-Price Densities Implicit in Financial Asset Prices, *Journal of Finance*, 53, 499-547.
- [2] Ait-Sahalia, Y., and A. Lo (2000), Nonparametric Risk Management and Implied Risk Aversion, *Journal of Econometrics*, 94, 9-51.
- [3] Andersen, T.G., L. Benzoni, and J. Lund (2002), Estimating Jump-Diffusions for Equity Returns, *Journal of Finance*, 57, 1239-1284.
- [4] Bakshi, G., C. Cao, and Z. Chen (1997), Empirical Performance of Alternative Option Pricing Models, *Journal of Finance*, 52, 2003-2049.
- [5] Bakshi, G., and N. Kapadia (2003), Delta Hedged Gains and the Negative Market Volatility Risk Premium, *Review of Financial Studies*, 16, 527-566.
- [6] Bakshi, G. D. Madan, and G. Panayotov (2009), Returns of Claims on the Upside and the Viability of U-Shaped Pricing Kernels, Forthcoming in the *Journal of Financial Economics*.
- [7] Banz, R. and M. Miller (1978), Prices for State-Contingent Claims: Some Estimates and Applications, *Journal of Business*, 51, 653-672.
- [8] Bates, D. (1996a), Testing Option Pricing Models, in *Handbook of Statistics, Statistical Methods in Finance*, G.S. Maddala and C.R. Rao (eds.), 567-611. Amsterdam: Elsevier.
- [9] Bates, D. (1996b), Jumps and Stochastic Volatility: Exchange Rate Processes Implicit in Deutsche Mark Options, *Review of Financial Studies*, 9, 69-107.
- [10] Bates, D. (2000), Post-'87 Crash Fears in the S&P 500 Futures Option Market, *Journal of Econometrics*, 94, 181-238.
- [11] Bates, D. (2003), Empirical Option Pricing: A Retrospection, *Journal of Econometrics*, 116, 387-404.
- [12] Bates, D. (2006), Maximum Likelihood Estimation of Latent Affine Processes, *Review of Financial Studies*, 19, 909-965.
- [13] Bates, D. (2008), The Market for Crash Risk, *Journal of Economic Dynamics and Control*, 32, 2291-2321.
- [14] Bates, D. (2009), Estimation and Filtration of Time-Changed Levy Processes, Working Paper, University of Iowa.

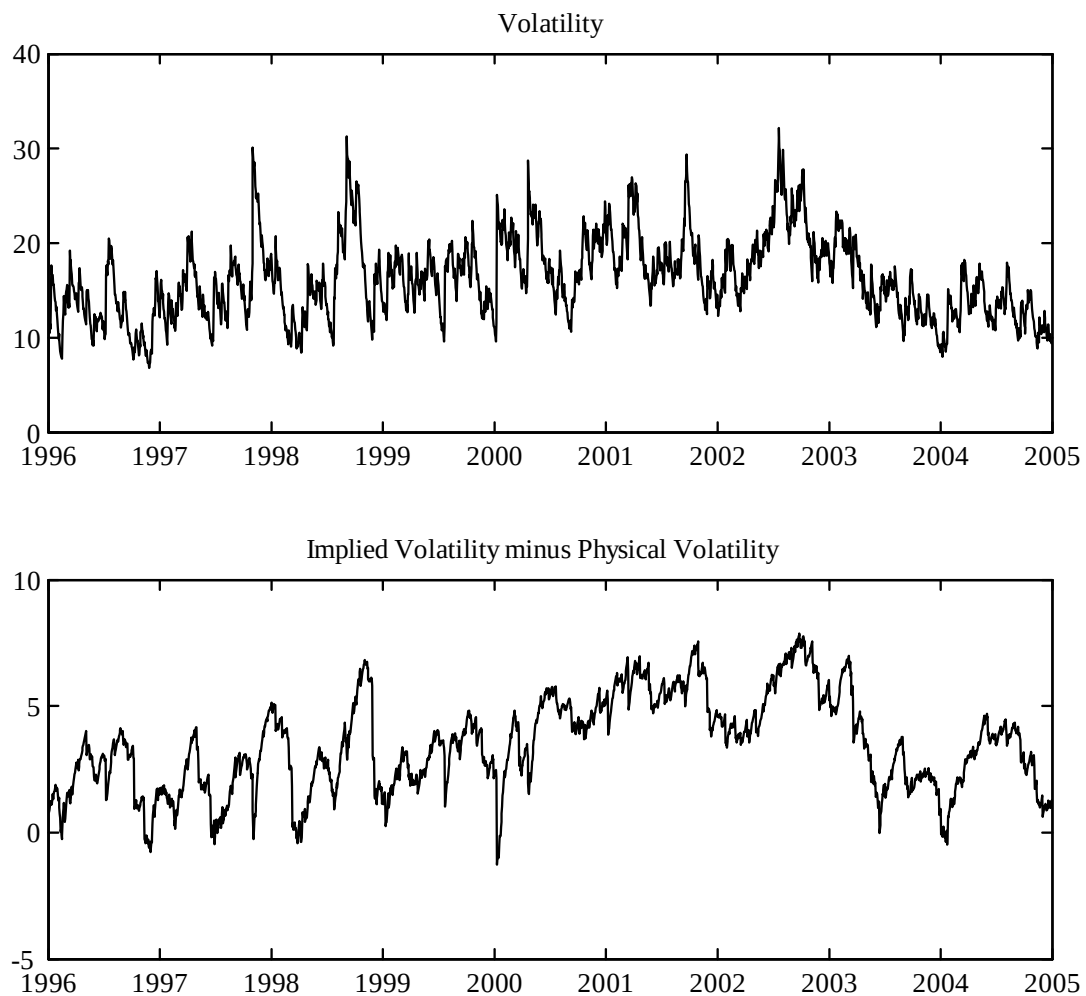
- [15] Benzoni, L. (2002), Pricing Options Under Stochastic Volatility: An Empirical Investigation, Working Paper, University of Minnesota.
- [16] Benzoni, L., P. Collin-Dufresne, and R. Goldstein (2005), Can Standard Preferences Explain the Prices of Out-of-the-Money S&P 500 Put Options?, Working Paper, Columbia University.
- [17] Black, F. (1976), Studies of Stock Price Volatility Changes, in: Proceedings of the 1976 Meetings of the Business and Economic Statistics Section, American Statistical Association, 177-181.
- [18] Black, F., and M. Scholes (1973), The Pricing of Options and Corporate Liabilities, Journal of Political Economy, 81, 637-659.
- [19] 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.
- [20] Bliss, R. and N. Panigirtzoglou (2004), Option-Implied Risk Aversion Estimates, Journal of Finance, 59, 407-446.
- [21] Bollerslev, Tim, George Tauchen, and Hao Zhou, 2009, Expected Stock Return and Variance Risk Premia, Review of Financial Studies, 22, 4463-4492.
- [22] Bondarenko, O. (2003), Why Are Put Options So Expensive?, Working Paper, University of Illinois at Chicago.
- [23] Breeden, D. and R. Litzenberger (1978), Prices of State-Contingent Claims Implicit in Option Prices, Journal of Business, 51, 621-651.
- [24] Brennan, M. (1979), The Pricing of Contingent Claims in Discrete-Time Models, Journal of Finance, 34, 53-68.
- [25] Broadie, M., M. Chernov, and M. Johannes (2007), Model Specification and Risk Premiums: Evidence from Futures Options, Journal of Finance, 62, 1453-1490.
- [26] Brown, D., and J. Jackwerth (2001), The Pricing Kernel Puzzle: Reconciling Index Option Data and Economic Theory, Working Paper, University of Wisconsin.
- [27] Canina, L., and S. Figlewski. (1993), The Informational Content of Implied Volatility, Review of Financial Studies, 6, 659-681.

- [28] Carr, P. and L. Wu (2004), Time-Changed Levy Processes and Option Pricing, *Journal of Financial Economics*, 17, 113-141.
- [29] Carr, P., and L. Wu (2007), Stochastic Skew in Currency Options, *Journal of Financial Economics*, 86, 213-247.
- [30] Chabi-Yo, Fousseni, (2009), Pricing Kernels with Stochastic Skewness and Volatility Risk, Working Paper, Ohio State University.
- [31] Chernov, M. (2003), Empirical Reverse Engineering of the Pricing Kernel, *Journal of Econometrics*, 116, 329-364.
- [32] Chernov, M. (2007), On the Role of Risk Premia in Volatility Forecasting, *Journal of Business and Economic Statistics*, 25, 411-426.
- [33] 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, 407-458.
- [34] Christoffersen, P., S. Heston, and K. Jacobs (2006), Option Valuation with Conditional Skewness, *Journal of Econometrics*, 131, 253-284.
- [35] Christoffersen, P., S. Heston, and K. Jacobs (2009), The Shape and Term Structure of the Index Option Smirk: Why Multifactor Stochastic Volatility Models Work so Well, *Management Science*, 55, 1914-1932.
- [36] Christoffersen, P., and K. Jacobs (2004), Which GARCH Model for Option Valuation?, *Management Science*, 50, 1204-1221.
- [37] Christoffersen, P., K. Jacobs, C. Ornathanalai, and Y. Wang (2008), Option Valuation with Long-Run and Short-Run Volatility Components, *Journal of Financial Economics*, 90, 272-297.
- [38] Coval, J. and T. Shumway (2001), Expected Option Returns, *Journal of Finance*, 56, 983-1009.
- [39] Cox, J., J. Ingersoll, and S. Ross (1985), An Intertemporal General Equilibrium Model of Asset Prices, *Econometrica*, 53, 363-384.
- [40] Day, T. and C. Lewis (1992), Stock Market Volatility and the Information Content of Stock Index Options, *Journal of Econometrics*, 52, 267-287.

- [41] Driessen, J. and P. Maenhout (2007), An Empirical Portfolio Perspective on Option Pricing Anomalies, *Review of Finance*, 11, 561-603.
- [42] Dumas, B., J. Fleming, and R. Whaley (1998), Implied Volatility Functions: Empirical Tests, *Journal of Finance*, 53, 2059-2106.
- [43] Eraker, B. (2004), Do Stock Prices and Volatility Jump? Reconciling Evidence from Spot and Option Prices, *Journal of Finance*, 59, 1367-1403.
- [44] Eraker, B., M. Johannes, and N. Polson (2003), The Role of Jumps in Returns and Volatility, *Journal of Finance*, 58, 1269-1300.
- [45] Fleming, J. (1998), The Quality of Market Volatility Forecasts Implied by S&P 100 Index Option Prices, *Journal of Empirical Finance*, 5, 317-345.
- [46] Heston, S. (1993), A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options, *Review of Financial Studies*, 6, 327-343.
- [47] Heston, S. and S. Nandi (2000), A Closed-Form GARCH Option Pricing Model, *Review of Financial Studies*, 13, 585-626.
- [48] Huang, J.-Z. and L. Wu (2004), Specification Analysis of Option Pricing Models Based on Time-Changed Levy Processes, *Journal of Finance*, 59, 1405-1439.
- [49] Hull, J., and A. White, 1987, The Pricing of Options with Stochastic Volatilities, *Journal of Finance*, 42, 281-300.
- [50] Jackwerth, J. (2000), Recovering Risk Aversion from Option Prices and Realized Returns, *Review of Financial Studies*, 13, 433-451.
- [51] Jones, C. (2003), The Dynamics of Stochastic Volatility: Evidence from Underlying and Options Markets, *Journal of Econometrics*, 116, 181-224.
- [52] Jorion, P. (1995), Predicting Volatility in the Foreign Exchange Market, *Journal of Finance*, 50, 507-528.
- [53] Lamoureux, C., and W. Lastrapes (1993), Forecasting Stock-Return Variance: Toward an Understanding of Stochastic Implied Volatilities, *Review of Financial Studies*, 6, 293-326.
- [54] Lewis, A. (2000), *Option Valuation Under Stochastic Volatility*, Finance Press, Newport Beach, CA.

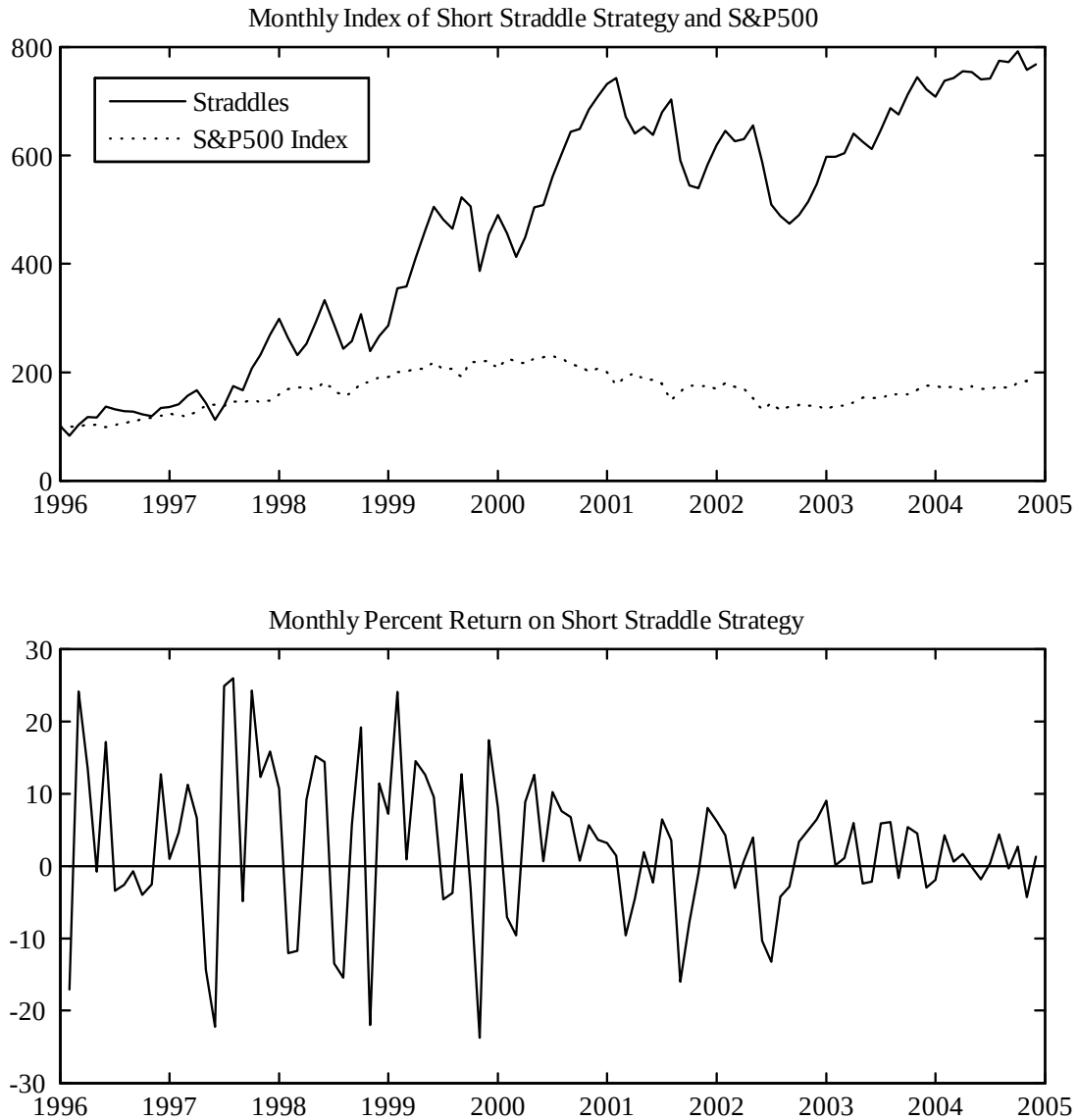
- [55] Liu, J., J. Pan, and T. Wang (2004), An Equilibrium Model of Rare Event Premia, *Review of Financial Studies*, 18, 131-164.
- [56] Melino, A., and S. Turnbull (1990), Pricing Foreign Currency Options with Stochastic Volatility, *Journal of Econometrics*, 45, 239-265.
- [57] Nelson, D., and D. Foster (1994), Asymptotic Filtering Theory for Univariate ARCH Models, *Econometrica*, 62, 1-41.
- [58] Pan, J. (2002), The Jump-Risk Premia Implicit in Options: Evidence from an Integrated Time-Series Study, *Journal of Financial Economics*, 63, 3-50.
- [59] Rompolis, L. and E. Tzavalis (2008), Recovering Risk Neutral Densities from Option Prices: A New Approach, *Journal of Financial and Quantitative Analysis*, 43, 1037-1054.
- [60] Rosenberg, J. and R. Engle (2002), Empirical Pricing Kernels, *Journal of Financial Economics* 64, 341-372.
- [61] Rubinstein, M. (1976), The Valuation of Uncertain Income Streams and the Pricing of Options, *Bell Journal of Economics*, 7, 407-425.
- [62] Rubinstein, M. (1994), Implied Binomial Trees, *Journal of Finance*, 49, 771-818.
- [63] Santa-Clara, P., and S. Yan (2009), Crashes, Volatility, and the Equity Premium, forthcoming in the *Review of Economics and Statistics*.
- [64] Shive, S., and T. Shumway (2006), Is the Pricing Kernel Monotonic?, Working Paper, Notre Dame University.
- [65] Stein, J. (1989), Overreactions in the Options Markets, *Journal of Finance*, 44, 1011-1023.
- [66] Trolle, A., and E. Schwartz (2009), Unspanned Stochastic Volatility and the Pricing of Commodity Derivatives, *Review of Financial Studies*, 22, 4423-4461.
- [67] Wiggins, J. (1987), Option Values Under Stochastic Volatility: Theory and Empirical Evidence, *Journal of Financial Economics*, 19, 351-372.

Figure 1: Physical and Risk-Neutral Volatility on the S&P500. 1996-2004.



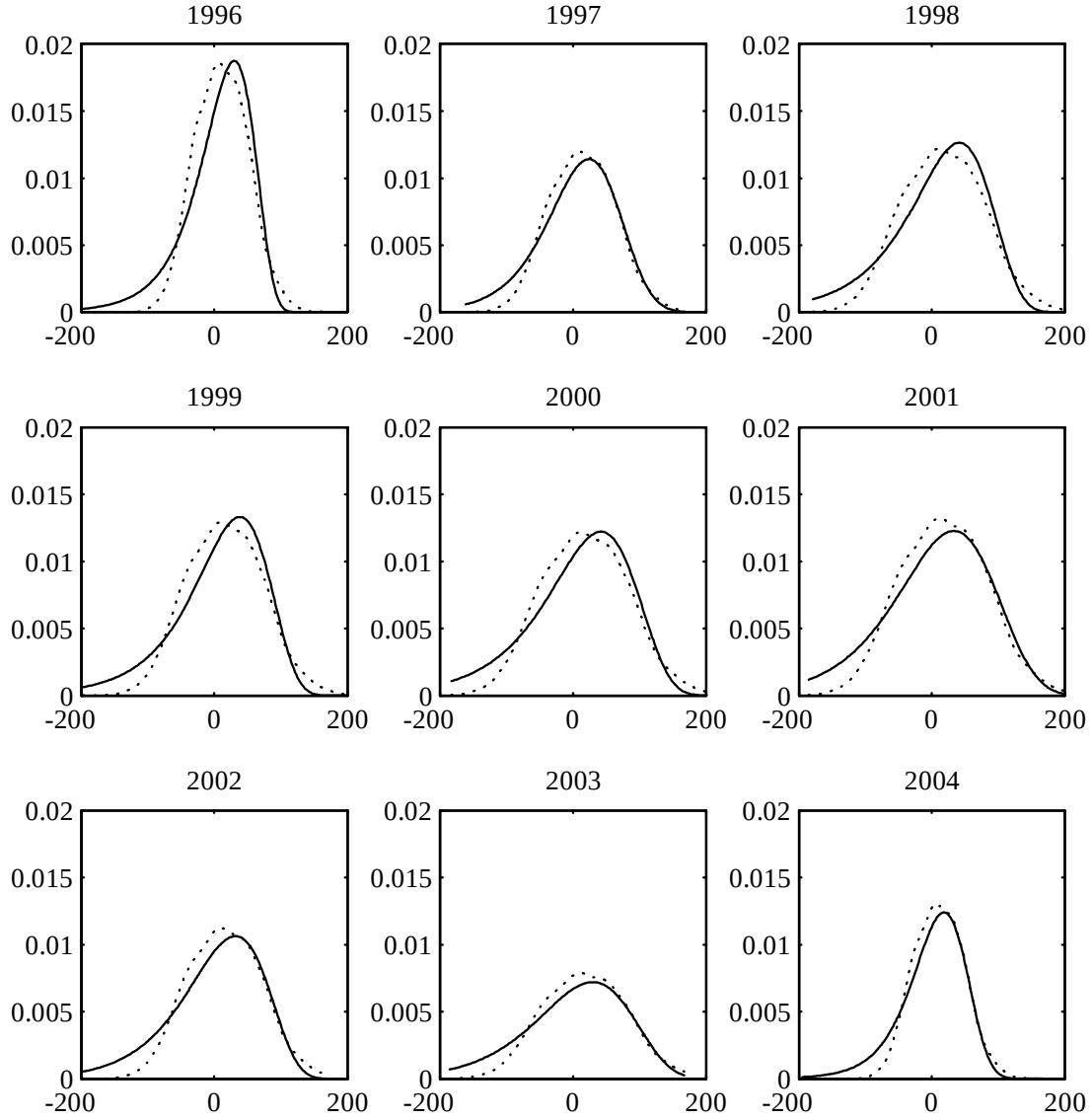
Notes to Figure: The top panel plots the annualized standard deviation of the daily S&P500 return from 1996 through 2004. This volatility is filtered through a GARCH process and is a physical volatility. The bottom panel plots the daily difference between the option-implied (risk-neutral) and physical volatilities.

Figure 2: Returns on Short Straddles. 1996-2004.



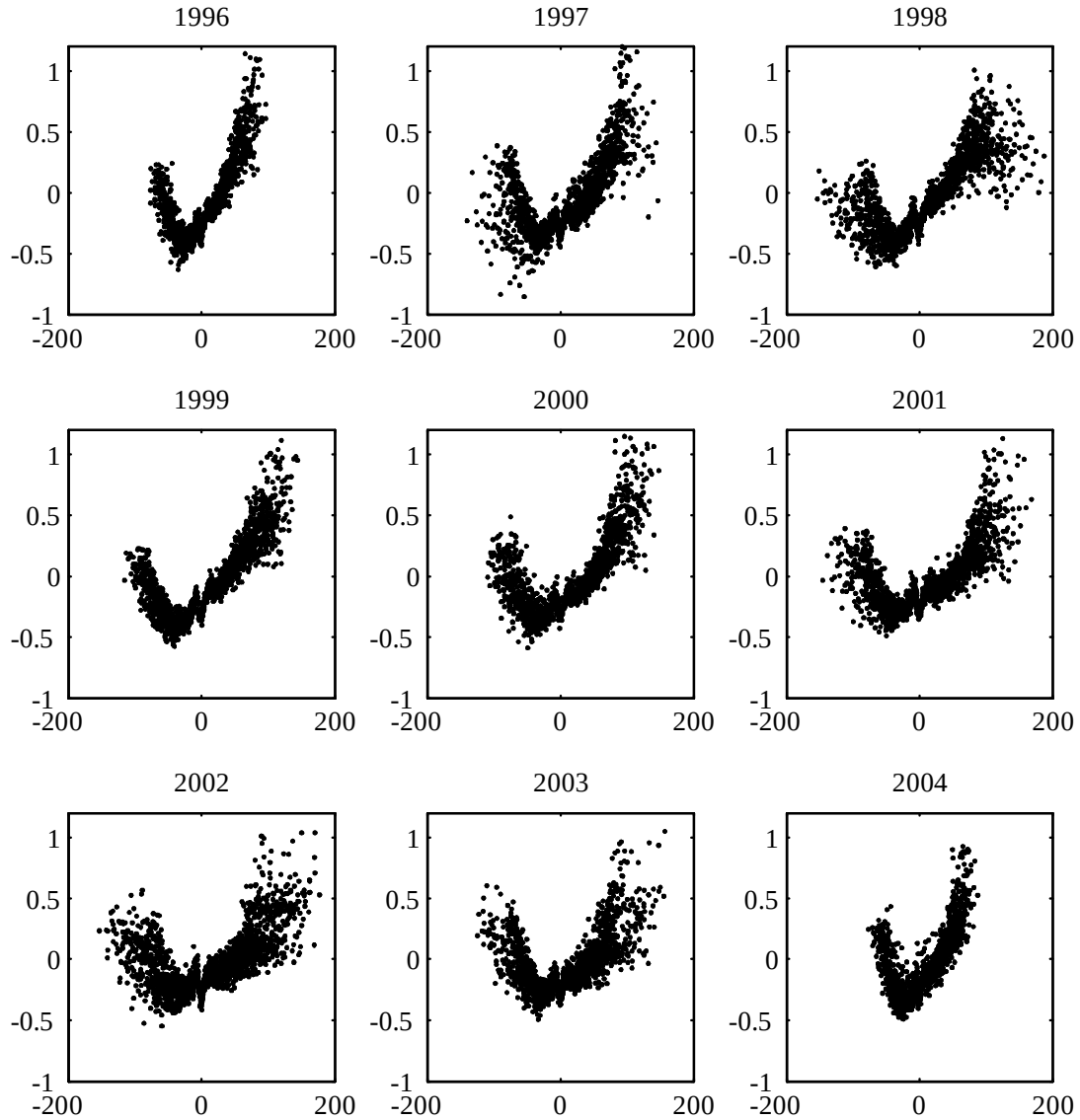
Notes to Figure: The top panel plots the monthly index of a short straddle strategy (solid line) and the S&P500 index (dashes) both normalized to 100 on January 1, 1996. The bottom panel plots the monthly returns on the straddle strategy. Straddle returns are computed using the nearest to at-the-money nearest to 30-day maturity call and put option on the third Friday of every month, with payoffs computed at maturity. The monthly index is computed starting with \$100 in cash, and keeping track of the cash account, with the cash account earning the risk-free rate.

Figure 3: Physical and Risk-Neutral One-Month Conditional Densities



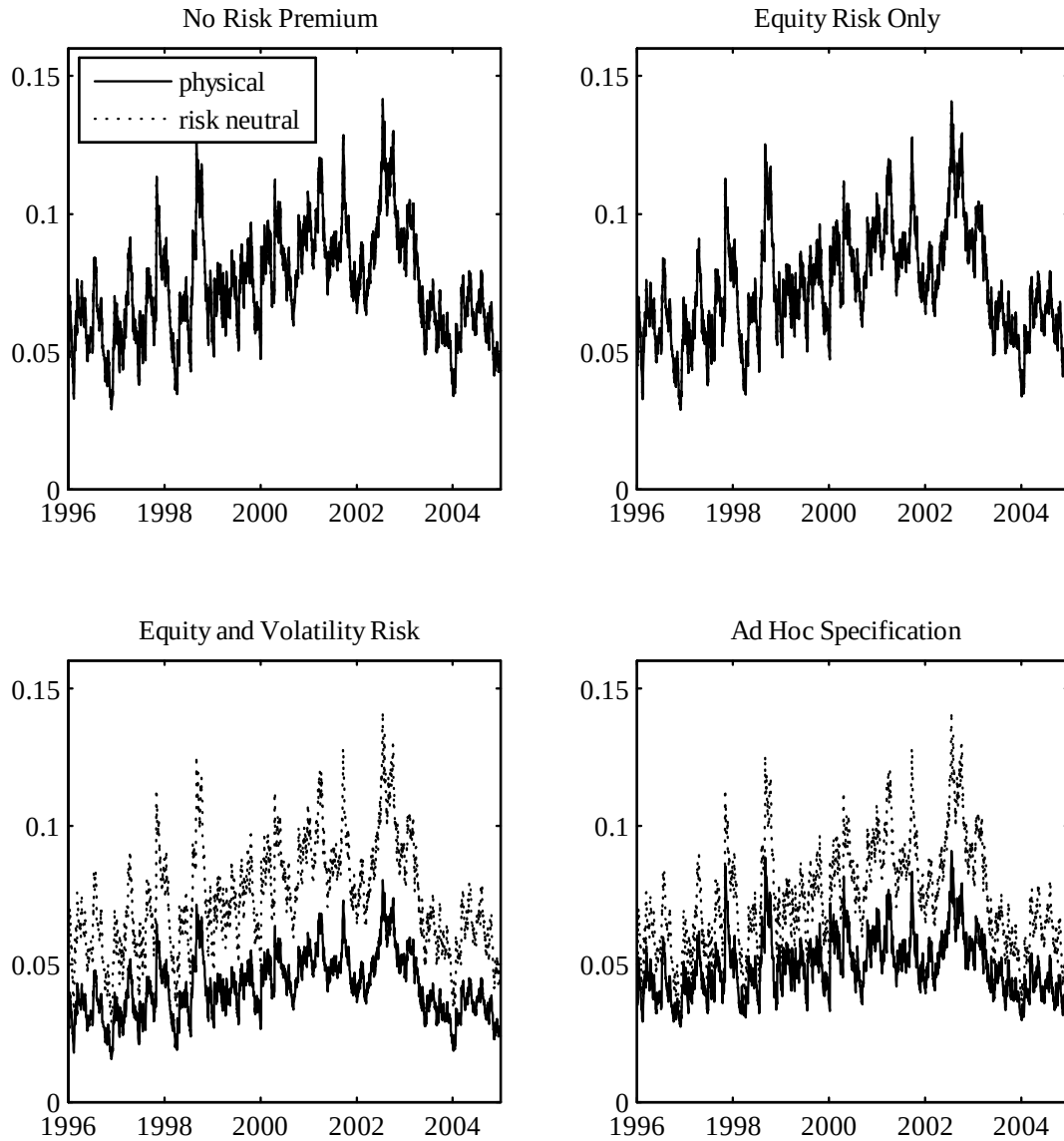
Notes to Figure: We plot physical (dashed line) and risk-neutral (solid line) conditional densities. The risk-neutral conditional density is obtained using the first Wednesday of option data for each of the nine years in our sample. For each Wednesday, a polynomial is fitted to the data, and the density is obtained from the polynomial fit using the method of Breeden and Litzenberger (1978). To construct the physical conditional density, we use returns for 1990-2004, scale the time series by the Wednesday's conditional volatility, and use a normal kernel. On the horizontal axis are annualized returns in percent.

Figure 4: Log Ratios of Risk-Neutral and Physical One-Month Densities



Notes to Figure: We plot the natural logarithm of the ratio of the risk-neutral and physical conditional densities. For each year in the option sample, we plot the ratios for each of the Wednesdays in that year. On the X-axis are annualized returns in percent.

Figure 5: Volatility of Variance, Annualized, 1996-2004.



Notes to Figure: We plot the annualized square root of the conditional variance of variance under the physical measure in solid lines and under the risk neutral measure in dashed lines. The parameters used are reported in Table 3.

Table 1: Returns and Options Data**Panel A: Return Characteristics (Annualized)**

	1990-2005	1996-2004
Mean	7.77	7.52
St. Deviation	16.08	19.01
Skewness	-0.102	-0.094
Kurtosis	6.786	5.696

Panel B. Option Data by Moneyness

	<u>F/X<0.96</u>	<u>.96<F/X<.98</u>	<u>.98<F/X<1.02</u>	<u>1.02<F/X<1.04</u>	<u>1.04<F/X<1.06</u>	<u>F/X>1.06</u>	<u>All</u>
Number of Contracts	3,118	1,925	5,688	2,498	1,979	6,183	21,391
Average IV	19.31	18.88	19.71	21.21	22.04	24.91	21.47
Average Price	24.02	27.58	35.28	31.34	28.42	23.40	28.42
Average Spread	1.39	1.40	1.57	1.44	1.35	1.29	1.41

Panel C. Option Data by Maturity

	<u>DTM<30</u>	<u>30<DTM<60</u>	<u>60<DTM<90</u>	<u>90<DTM<120</u>	<u>120<DTM<180</u>	<u>DTM>180</u>	<u>All</u>
Number of Contracts	1,069	5,717	3,996	1,726	3,253	5,630	21,391
Average IV	21.04	21.17	21.69	21.78	21.68	21.49	21.47
Average Price	12.58	17.49	23.61	26.34	31.81	44.62	28.42
Average Spread	0.87	1.14	1.36	1.44	1.53	1.76	1.41

Notes: We present descriptive statistics for daily return data from January 1, 1990 to December 31, 2005, as well as for daily return data from January 1, 1996 to December 31, 2004. We use Wednesday closing OTM options contracts from January 1, 1996 to December 31, 2004.

Table 2: Stein Regression of Overreaction

Sample Period	Coefficient	Standard Error	t-Statistic
Full Sample	-0.241	0.0091	-26.42
1996	-0.358	0.0284	-12.58
1997	-0.195	0.0269	-7.24
1998	-0.298	0.0410	-7.26
1999	-0.311	0.0169	-18.41
2000	-0.211	0.0141	-15.00
2001	-0.211	0.0241	-8.76
2002	-0.167	0.0274	-6.08
2003	-0.231	0.0203	-11.34
2004	-0.364	0.0207	-17.60

Notes: Using 1996-2004 option data, we run the forecasting regressions from Stein (1989, p. 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 1-month maturity for short-term options and 2-month maturity for long-term options. We run the regressions for the full sample as well as for each year in the sample separately.

Table 3: Parameter Estimation and Model Fit. Joint Estimation using Returns and Options.

		No Risk Premium	Equity Risk Only	Equity and Volatility Risk	Ad Hoc Specification
Physical Parameters	ω	0	0	0	0
	α	3.248E-06	3.249E-06	1.546E-06	3.452E-06
	β	0.826	0.826	0.826	0.875
	γ	220.9	219.4	318.5	158.7
	μ	0	1.501	1.543	1.954
RN Parameters	$1/(1-2\alpha\xi)$	1	1	1.2039	
	ω^*	0	0	0	0
	α^*	3.248E-06	3.249E-06	3.248E-06	3.248E-06
	β^*	0.826	0.826	0.826	0.826
	γ^*	220.93	220.93	220.78	220.93
Likelihood	Total	55,366.1	55,368.4	55,485.5	55,512.2
	From returns	13,188.3	13,190.6	13,307.7	13,334.4
	From options	42,177.8	42,177.8	42,177.8	42,177.8
Physical Properties	Long run volatility	0.234	0.219	0.153	0.151
	Daily autocorrelation, $h(t)$	0.985	0.983	0.983	0.962
	Annualized volatility of $h(t)$	0.087	0.081	0.039	0.046
	Correlation($R(t), h(t+1)$)	-0.977	-0.974	-0.974	-0.905
RN Properties	Long run volatility	0.234	0.234	0.232	0.234
	Daily autocorrelation, $h^*(t)$	0.985	0.985	0.985	0.985
	Annualized volatility of $h^*(t)$	0.087	0.087	0.086	0.087
	Correlation($R(t), h^*(t+1)$)	-0.977	-0.977	-0.977	-0.977

Notes: Parameter estimates are obtained by optimizing an joint likelihood on returns and options. Parameters as well as 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. In the "Ad Hoc Specification" the physical and risk-neutral parameters are not linked. This model has nine parameters. The "No Risk Premium Model" has four parameters, with $\mu=0$ and $\xi=0$. The model with "Equity Risk Only" has five parameters. It imposes $\xi=0$. The model with "Equity and Volatility Risk" has six parameters, and estimates μ and ξ . A parameters are constrained to be positive which is a sufficient condition for positive variance.

Table 4: IV RMSE and Bias (%) by Moneyness and Maturity. 1996-2004.

Panel A. IV RMSE by Moneyness							
<u>Model</u>	<u>F/X<0.96</u>	<u>.96<F/X<.98</u>	<u>.98<F/X<1.02</u>	<u>1.02<F/X<1.04</u>	<u>1.04<F/X<1.06</u>	<u>F/X>1.06</u>	<u>All</u>
ERP and VRP	2.9741	3.1391	3.1399	3.2364	3.2766	3.8717	3.3685
ERP only	4.8723	4.7336	5.1241	5.6927	5.9226	6.7394	5.7111
No RP	5.2882	5.1320	5.5185	6.0926	6.3163	7.1125	6.0986
Panel B. IV Bias by Moneyness (Data Less Model)							
<u>Model</u>	<u>F/X<0.96</u>	<u>.96<F/X<.98</u>	<u>.98<F/X<1.02</u>	<u>1.02<F/X<1.04</u>	<u>1.04<F/X<1.06</u>	<u>F/X>1.06</u>	<u>All</u>
ERP and VRP	-0.6011	-0.7463	-0.3734	0.1899	0.4802	1.3824	0.2121
ERP only	4.1721	3.7105	4.0581	4.7634	5.0777	5.9468	4.7661
No RP	4.6481	4.1875	4.5256	5.2185	5.5214	6.3695	5.2190
Panel C. IV RMSE by Maturity							
<u>Model</u>	<u>DTM<30</u>	<u>30<DTM<60</u>	<u>60<DTM<90</u>	<u>90<DTM<120</u>	<u>120<DTM<180</u>	<u>DTM>180</u>	<u>All</u>
ERP and VRP	3.3513	3.4181	3.4614	2.9799	3.4059	3.3447	3.3685
ERP only	4.7976	5.4619	5.8224	5.6526	5.8944	5.9440	5.7111
No RP	5.0606	5.7668	6.1823	6.0429	6.3122	6.4292	6.0986
Panel D. IV Bias by Maturity (Data Less Model)							
<u>Model</u>	<u>DTM<30</u>	<u>30<DTM<60</u>	<u>60<DTM<90</u>	<u>90<DTM<120</u>	<u>120<DTM<180</u>	<u>DTM>180</u>	<u>All</u>
ERP and VRP	0.1206	0.7148	0.5890	0.4085	0.0273	-0.5019	0.2121
ERP only	3.7334	4.4842	4.8342	4.9162	4.9451	5.0505	4.7661
No RP	4.0889	4.8651	5.2597	5.3583	5.4299	5.5997	5.2190

Notes: We report option implied volatility based RMSE and bias by moneyness and maturity using the parameter estimates from the column labeled "Equity and Volatility Risk" in Table 3. ERP denotes Equity Risk Premium and VRP denotes Volatility Risk Premium. In the row labeled "ERP only" we force the volatility risk premium to be zero and in the row labeled "No RP" we force both the volatility and the equity risk premium to be zero. The physical parameters are the same in all three rows.