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

June 8, 2010

Abstract

We generalize the Heston-Nandi (2000) GARCH model to a discrete-time analog of the Heston (1993) stochastic volatility model with a variance risk premium. We show the pricing kernel in these models generalizes the risk-neutralization in Rubinstein (1976) and Brennan (1979). While it is monotonic in the stock return and volatility, its projection onto the stock return is nonmonotonic. A negative risk premium makes it appear U-shaped. We present new semi-parametric evidence to confirm this U-shaped relationship between the risk-neutral and physical probability densities. The new pricing kernel substantially improves our ability to reconcile the time series properties of stock returns with the cross-section of option prices. It provides a unified explanation for the implied volatility puzzle, the overreaction of long-term options to changes in short-term variance, and the fat tails of the risk-neutral return distribution relative to the physical distribution.

JEL Classification: G12

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

*Christoffersen and Jacobs want to thank FQRSC, IFM2, and SSHRC for financial support. We are grateful to Gurdip Bakshi, Garland Durham, Stylianos Perrakis, and participants in seminars at the University of Houston, UC Irvine, Concordia University, Luxembourg Institute of Finance, University of Southern California, and the HEC Montreal Applied Financial Time Series Workshop 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 framework. 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 the empirical 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. Moreover, Stein (1989) and Poteshman (2001) show that long-term implied variance overreacts to changes in short-term variance. This puzzle involves movements in the term structure of implied volatility and is related to the expectations puzzle. Taken together, these anomalies indicate misspecification in the dynamic relationship between option values and the time series of spot returns. They are usually not discussed in the context of a parametric framework, and therefore the literature has 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 (2010).

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

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. In contrast to the risk-neutral valuation relationship used by Rubinstein (1976) and Brennan (1979), the natural logarithm of the conditional pricing kernel appears to be a U-shaped function of returns, and 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 insight 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. Several authors have noted that a negative risk premium for volatility can explain why implied volatilities are high and average option returns are low. However, 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 also explain why long-maturity implied volatilities appear to overreact to temporary increases in volatility, thereby explaining Stein’s (1989) observation.

Our analytical results indicate that the suggested pricing kernel is able to qualitatively account for a number of important puzzles. 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. Initially we derive these results in a continuous-time stochastic volatility framework. We then show that it is possible to develop a discrete-time pricing kernel with similar features in a GARCH framework. Specifically, we generalize the Heston-Nandi (2000) GARCH model to include a more general pricing kernel that generates a variance premium. In this model the logarithm of the pricing kernel is a quadratic function of the market return. The GARCH model directly provides tomorrow’s conditional variance from today’s observed return, and so this approach is very useful econometrically because it avoids having to apply particle filtering or MCMC techniques to extract the spot volatility path.²

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. The discrete-time GARCH structure of the model greatly facilitates the filtering problem, and its numerical efficiency makes it possible to maximize the likelihood with the return and option components in a large-scale empirical exercise.

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 Rubinstein (1976) and Brennan (1979) 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.

In summary, the key contributions of our paper are as follows: First, we provide model-free evidence on the prevalence of a U-shaped pricing kernel. Second, we show how this U-shaped kernel naturally emerges from monotonic preferences in both continuous-time stochastic volatility models and a discrete-time GARCH framework. Third, the discrete-time GARCH approach makes it feasible to use a large option and returns dataset in our empirical work, and we show how the pricing kernel helps explain the stylized facts of returns and options by jointly maximizing the likelihood of index returns and the fit of options.

(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 (2010) 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 GARCH 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 out-of-the-money (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 Fat Tails and Fatter Tails

The most challenging 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. 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, 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. Despite these recent innovations in the modeling of underlying index returns, we therefore do not have a model that convincingly captures the features of the physical as well as the risk-neutral distributions.

---

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

In order to capture these stylized facts, risk-neutral conditional and unconditional skewness and kurtosis have to 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 pricing kernel. 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 pricing kernel 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 pricing kernel will be an exponentially increasing function of future variance. In other words the bivariate pricing kernel 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 pricing kernel 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 pricing kernel 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.

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.
2.3 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 (2010) present the most closely related evidence on the conditional pricing kernel, but our conditioning approach is very different.

It is relatively straightforward to estimate the risk-neutral conditional density of returns using option data, harnessing the insights of Breeden and Litzenberger (1978) and Banz and Miller (1978), and there is an extensive empirical literature reporting on this. Ait-Sahalia and Lo (2000) obtain 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 second-order polynomial function for implied Black-Scholes volatility as a function of moneyness and maturity. Using this estimated polynomial, we then generate a grid of 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.

$$\hat{C} (S(t), X, \tau, r) = C_{BS} (S(t), X, \tau, r; \hat{\sigma} (S(t), X, \tau)) .$$ (1)

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

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

(2)

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

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

(3)

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 histogram based on this time series is effectively an estimate of the unconditional physical density of one-month log returns. We obtain a conditional density estimate for a given day by first standardizing the monthly return series by the sample mean \( \bar{R} \) and the conditional volatility on that day, as measured by the VIX. This provides a series of return shocks \( Z = (R - \bar{R}) / VIX \). We then construct a conditional histogram for a given day \( t \) day using the VIX for that day, \( VIX(t,T) \), and the historical series of shocks, \( Z \). We write this estimate of the conditional physical distribution as

\[ \hat{f}_t (R(t,T)) = \hat{f} (VIX(t,T)Z + \bar{R}) \]

A subset of the resulting estimates of physical and risk-neutral conditional densities are given in Figure 1. 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 1 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 1 clearly demonstrates the fat left tail of the estimate of the risk-neutral conditional density, compared to 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 1 indicates that it is difficult to draw any definitive conclusion regarding the relative thickness of the right tail of the risk-neutral and physical density. The estimate of the right tail of the physical density is somewhat noisy, but more critically the right tail of the risk-neutral density is much harder to estimate than the left tail, because of the relative scarceness of traded out-of-the-money call options. Moreover, those options are usually thinly traded. Fortunately, the relative thickness of the right tail is inconsequential for establishing the nonlinearity of the logarithm of the ratio of the densities.

Figure 2 depicts the natural logarithm of the ratio of the estimates 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 1, we present nine sets of results, one for each year of the sample. Recall that in Figure 1 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 2, 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

$$\ln \left( \frac{\hat{f}_t^p (R(t,T))}{\hat{f}_t^0 (R(t,T))} \right), \quad \text{for } t = 1, 2, \ldots, 52$$

In each week we trimmed 3% of observations in the left and right tails, because these observations are sometimes very noisy.
Two very important conclusions obtain. First, the pricing kernel is clearly not a monotonic function of returns, rejecting a hypothesis implicit in the Black-Scholes model and much of the option pricing literature. Second, the shape of the pricing kernel is remarkably stable across time. It is evident that the shape of the pricing kernel varies somewhat across certain years. 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 nonlinear nature of the logarithm of the kernel. If the kernel varied more within the year, Figure 2 would contain nothing but a cloudy scatter without much structure. Whether the logarithm of the kernel is exactly a quadratic function of stock returns is perhaps less obvious, because there is some noise in the estimates of the densities’ right tail. However, it is clear that the relationship is nonlinear.

In summary, Figure 2 illustrates that the logarithm of the pricing kernel is nonlinear and roughly quadratic as a function of the return, and that this pricing relationship is relatively stable over time. This motivates us to look beyond the Rubinstein (1976) and Brennan (1979) pricing kernel which is linear in log-log space.

2.4 Returns on Straddles

A successful model for index options also has to address a number of other stylized facts and anomalies. It is well-known that on average, risk-neutral volatility exceeds physical volatility.\(^5\) Several authors have argued that the risk premium that explains this difference makes it interesting to short sell straddles.\(^6\) Figures 3 and 4 illustrate these stylized facts using the 1996-2004 option sample from Table 1. Figure 3 illustrates that risk-neutral volatility exceeds physical volatility, when both are 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, using realized volatility instead of GARCH volatility, or measuring risk-neutral volatility using the VIX. Figure 4 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 4 plots the S&P500 monthly closing price normalized to 100 in January 1996 for comparison. It is obvious from Figure 4 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


\(^6\)See among others Coval and Shumway (2001), Bondarenko (2003), and Driessen and Maenhout (2007).
would be approximately zero, and the strategy’s returns would not be correlated with market returns.

### 2.5 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

\[
E_t \left[ (IV_{t+(LT-ST)}^{ST} - IV_t^{ST}) - 2(IV_t^{LT} - IV_t^{ST}) \right] = 0, \tag{4}
\]

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 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 (4), 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.

3 The Pricing Kernel in a Continuous-Time Stochastic Volatility Model

Given statistical estimates of the physical process, we can use a pricing kernel to value options. This section shows there is a unique pricing kernel implicit in the Heston (1993) model. This provides the basis for subsequent valuation analysis in a GARCH model.

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

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

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 \( \theta \) with a speed of \( \kappa \). The expected future variance is a linear function of current variance

\[
    E_t(v(t + \Delta)) = \exp^{-\kappa \Delta} v(t) + (1 - \exp^{-\kappa \Delta}) \theta.
\]

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

\[
\begin{align*}
    dS(t) &= r S(t) dt + \sqrt{v(t)} S(t) dz_1^*(t), \\
    dv(t) &= (\kappa(\theta - v(t)) - \lambda v(t)) dt + \sigma \sqrt{v(t)}(\rho dz_1^*(t) + \sqrt{1 - \rho^2} dz_2^*(t)),
\end{align*}
\]

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

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

(8)

where \( \theta^* = \kappa \theta / (\kappa + \lambda) \) and \( \kappa^* = \kappa + \lambda \). Therefore \( \theta^* \) is the risk-neutral long-run variance, and \( \kappa^* \) 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 pricing operator \( M(t) \) underlying this risk-neutralization. Proposition 1 expresses \( M(t) \) in terms of the parameters governing the dynamics of variance, \( \kappa, \theta, \rho, \) and \( \sigma \), the equity premium \( \mu \), and the variance premium \( \lambda \).

**Proposition 1** The pricing operator inherent in equations (5) and (7) takes the form

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

(9)

where

\[ \delta = - (1 + \phi) \rho - \xi \kappa, \]
\[ \eta = - \phi \mu + \frac{1}{2} \phi + \xi \kappa - \frac{1}{2} (\phi^2 + 2 \phi \xi \sigma \rho + \xi^2 \sigma^2), \]
\[ \phi = - \mu - \xi \sigma \rho, \]
\[ \xi = \frac{\mu \sigma \rho - \lambda}{\sigma^2 (1 - \rho^2)}. \]

**Proof.** See Appendix. \( \blacksquare \)

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 pricing operator also depends on the variance \( v(t) \). This has distinctive implications for option valuation.

Several papers have noted the importance of the pricing kernel 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 pricing kernel for stochastic volatility and jumps. The purpose of this paper is to derive distinctive implications for option values from a comparatively

\[ \text{This could result, for instance, if } v(t) \text{ governs the variance of aggregate production in a Cox-Ingersoll-Ross (1985) model with non-logarithmic utility. It could also result from the model of Benzoni, Collin-Dufresne, and Goldstein (2009) where uncertainty directly affects preferences. Bakshi, Madan, and Panayotov (2010) relate it to short-sale constraints.} \]
simple pricing kernel. In particular, the pricing kernel 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.

Proposition 1 mechanically illustrates the restrictions imposed on the parameters of the pricing kernel (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 $\delta$ and $\eta$. The equity premium parameter $\mu$ determines the risk aversion parameter $\phi$, and the variance premium $\lambda$ determines the variance preference $\xi$. It is therefore also instructive to re-arrange Proposition 1 to express the equity premium $\mu$ and variance premium $\lambda$ parameters in terms of the underlying preference parameters $\phi$ and $\xi$. This gives

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

This demonstrates that both the equity risk premium $\mu$ and the variance risk premium $\lambda$ can be thought of as having two distinct components originating in preferences. One is related to the risk-aversion parameter $\phi$ and the other one to the variance preference parameter $\xi$. We can therefore use economic intuition to sign the equity premium and the variance premium. If the pricing kernel is decreasing in the spot price, we have $\phi < 0$, because marginal utility is a decreasing function of stock index returns. If hedging needs increase in times of uncertainty then we anticipate the pricing kernel to be increasing in the volatility, $\xi > 0$. Empirically the correlation between stock market returns and variance $\rho$ is strongly negative. Therefore from (10) the equity premium $\mu$ must be positive. The variance premium $\lambda$ has a component based on covariance with equity risk, and a separate independent component based on the variance preference $\xi$. With a negative correlation $\rho$, we see that $\lambda$ 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 $\phi$ (and therefore indirectly from the equity premium $\mu$) or whether it has an independent component $\xi$. In other words, assuming $\xi = 0$ in (9) is consistent with a nonzero variance risk premium $\lambda$, 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) $\lambda$, because this assumption is consistent with the pricing kernel (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 pricing kernel below in the context of a discrete-time GARCH 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 2.3. 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 $\phi$ 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 2 in Section 2.3 therefore favors $\xi > 0$. We analyze this issue further below in the context of our discrete-time GARCH framework.

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

$$E^*_t(v(t + \Delta)) = \exp^{-\kappa^* \Delta} v(t) + (1 - \exp^{-\kappa^* \Delta}) \theta^*.$$  

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 Incorporating the Pricing Kernel into a GARCH Model

The objective of our empirical exercise is to address differences between the physical and risk-neutral distributions. Estimation using only option data or only return data can diagnose this problem, but it cannot assess a model’s ability to resolve it. Therefore it is critically important to evaluate the models’ joint implications for return series as well as option prices. This requires specification of a nonnegative pricing kernel that relates the observed physical probabilities to the risk-neutral probabilities in option prices.

This section uses the variance-dependent pricing kernel (9) from the previous section to generate an additional risk premium in the Heston and Nandi (2000) GARCH model. This generalization makes it a complete analog of the Heston (1993) model. While the Heston (1993)
formula is quite similar to the GARCH formula, the GARCH approach has three advantages for empirical work. First, while it may be considered an accurate approximation to the continuous-time model, the GARCH model provides an internally consistent framework that makes exact predictions at a daily frequency, while avoiding discretization bias. Second, these predictions are exactly compatible with the GARCH filter used to extract variance. In other words, it enforces relations between option prices and the time series of spot returns. Third, the GARCH model is computationally fast, allowing us to use a long time-series of returns with a large cross-section of option prices.

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

\[
\ln(S(t)) = \ln(S(t-1)) + r + (\mu - \frac{1}{2})h(t) + \sqrt{h(t)}\sigma(t), \tag{12}
\]
\[
h(t) = \omega + \beta h(t-1) + \alpha (z(t-1) - \gamma \sqrt{h(t-1)})^2,
\]

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.

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

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

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

The data robustly indicate sizeable negative correlation, which means that \( \gamma \) must be positive.

We can value securities in this discrete-time GARCH model using the same pricing kernel (5)

---


9GARCH departs from a stochastic volatility framework by making variance a function of the historical path of the spot price. By omitting a separate source of variance risk, it allows variance to be filtered directly from stock returns. Proper calibration produces similar probability predictions for both types of models. Using identical pricing kernels in stochastic volatility and GARCH models then ensures that the risk-neutral probabilities are also similar.
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 pricing operator

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

(16)

follows the GARCH process

\[ \ln(S(t)) = \ln(S(t-1)) + r - \frac{1}{2} h^*(t) + \sqrt{h^*(t)z^*(t)}, \]  
\[ h^*(t) = \omega^* + \beta h^*(t-1) + \alpha^*(z^*(t-1) - \gamma^* \sqrt{h^*(t-1)})^2, \]  

(17)

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

\[ h^*(t) = h(t)/(1 - 2\alpha\xi), \]  
\[ \omega^* = \omega/(1 - 2\alpha\xi), \]  
\[ \alpha^* = \alpha/(1 - 2\alpha\xi)^2, \]  
\[ \gamma^* = \gamma - \phi. \]  

(18)

**Proof.** See Appendix. ■

The risk-neutral dynamics differ from the physical dynamics through the effect of the equity premium parameter \( \mu \) and scaling factor \( 1 - 2\alpha\xi \). Conditional on the parameters characterizing the physical dynamic, these risk-neutral dynamics are therefore implied by the values of the kernel parameters \( \phi \) and \( \xi \) in equation (16).\(^{10}\) The intuition is similar to the continuous-time case in Section 3, where the values of the equity premium and volatility risk premium parameters \( \mu \) and \( \lambda \) are implied by the values of the kernel parameters \( \phi \) and \( \xi \). With an annual U.S. equity premium \( \mu h(t) \) of around 8% and variance \( h(t) \) of 20%\(^2\), it can be inferred that the value of the equity premium parameter \( \mu \) is small, around 2.

In contrast to the Heston-Nandi (2000) model, the risk-neutral variance \( h^*(t) \) differs from the physical variance \( h(t) \).\(^{11}\) When \( \alpha\xi > 0 \), the pricing kernel puts more weight on the tails of

---

\(^{10}\) The mapping between \( \mu \) and \( \phi \) is contained in the Appendix.

\(^{11}\) In diffusion models, the instantaneous variance is identical under the physical and risk-neutral measures, but the risk-neutral variance will differ from the physical variance over a discrete interval such as one day.
innovations and the risk-neutral variance \( h^*(t) \) exceeds the physical variance \( h(t) \). The Heston-Nandi (2000) model corresponds to the special case of \( \xi = 0 \). 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. The variance risk premium also affects the risk-neutral drift of \( h^*(t) \)

\[
E_{t-1}^*(h^*(t+1)) = (\beta + \alpha^* \gamma^* \sigma^2)h^*(t) + (1 - \beta - \alpha^* \gamma^* \sigma^2)E^*(h^*(t)),
\]

(19)

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

Comparison of physical parameters with risk-neutral parameters shows that if the correlation between returns and variance is negative \( (\gamma > 0) \), 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

\[
Var_{t-1}^*(h^*(t+1)) = 2\alpha^* \sigma^2 + 4\alpha^* \gamma^* \sigma^2 h^*(t).
\]

(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 \( \alpha^* \) and \( \gamma^* \) 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

\[
Cov_{t-1}^*(R(t), h^*(t+1)) = -2\alpha^* \gamma^* h^*(t)
\]

(21)

The following corollary summarizes the results for this discrete-time GARCH 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 negative skewness and kurtosis exceeding physical negative skewness and kurtosis. Our model captures this stylized fact through a U-shaped pricing kernel. Interestingly, even though the pricing kernel is a monotonic function of the stock price and variance (9), the projection of the pricing kernel onto the stock price alone can have a U-shape. The following proposition formalizes this relationship.

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

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

where \( R(t) = \ln \left( \frac{S(t)}{S(t-1)} \right) \).

Proof. See Appendix.  

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 \( \xi \), 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 2. Thus the strong U-shaped ratios in Figure 2 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

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

\[
- X \exp^{-r(T-t)} \left( \frac{1}{2} + \int_{\ln(X)}^{\infty} \text{Re} \left[ \frac{X^{-i\varphi} g_t^* (i\varphi)}{i\varphi} \right] d\varphi \right).
\]

where \( g_t^* (.) \) is the conditional generating function for the risk-neutral process (17). The functional form for the physical \( g_t (.) \) is provided in Heston and Nandi (2000). Together with the mapping in (18) this yields \( g_t^* (.) \). Put options can be valued using the put-call parity.

## 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.\(^{12}\) When estimating multiple cross-sections, one can parameterize the volatility state variable in the same way, at the cost of estimating a high number of parameters,\(^{13}\) or one can filter the volatility from underlying returns, using a variety of filters. Some papers take into account returns through the filtering exercise, but do not explicitly take into account returns in the objective function.\(^{14}\) 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.

---

\(^{12}\) See for instance the seminal paper by Bakshi, Cao and Chen (1997).

\(^{13}\) See for instance Bates (2000), Christoffersen, Heston, and Jacobs (2009), and Huang and Wu (2004).

\(^{14}\) See for instance Christoffersen and Jacobs (2004).

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 a discrete time GARCH setting. Note that the conditional density of the daily return is normal so 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

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

(24)

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

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

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

\[
\ln L^O \propto -\frac{1}{2} \sum_{i=1}^{N} \left\{ \ln (s_\varepsilon^2) + \varepsilon_i^2 / s_\varepsilon^2 \right\}.
\]

(25)

where we can concentrate out \(s_\varepsilon^2\) 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 \(BSV_i^{Mkt}\) weighted option
We can now solve the following optimization problem

$$\max_{\Theta, \Theta^*} \ln L^R + \ln L^O,$$  \hspace{1cm} (26)

where $\Theta = \{\omega, \alpha, \beta, \gamma, \mu\}$ denotes the physical parameters and $\Theta^*$ denotes the risk-neutral parameters which are mapped from $\Theta$ 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 pricing kernel (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: $\omega, \alpha, \beta, \gamma$. The second model is the Heston-Nandi (2000) model that allows for equity risk. This effectively leaves the equity premium $\mu$ (or equivalently the risk-aversion $\phi$) as a free parameter, yielding five parameters in total. The most general economic model allows both equity risk and independent volatility risk by allowing $\xi$ to be a free parameter.

Table 3 presents the empirical results, which are quite striking.\footnote{We impose $\omega^* = \omega = 0$ in estimation because the nonnegativity constraint is binding and it is a necessary condition for positive variances.} 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.

The most important feature of the new GARCH model is that it captures the U-shape in the log ratio of the risk-neutral and physical densities found in Figure 2. In order to show this we plot the model-implied log density ratio in Figure 5 for various maturities. The solid lines plot the
log density ratio for the new model and the dashed lines show the special Rubinstein (1976) and Brennan (1979) case of a variance premium of zero. Note that the 30-day maturity used in Figure 2 corresponds to the upper-right panel of Figure 5. In Figure 5 we set the conditional variance to its unconditional level and we fix the risk-free rate at 5% per year. We use the parameters from the third column of Table 3. The new model therefore captures the first stylized fact discussed in Section 2.

To explore other features of the new GARCH model, it is instructive to inspect 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, Table 3 shows that the average risk-neutral volatility is much higher than the average physical volatility. Second, the risk-neutral variance persistence, 0.985, is higher than the physical variance persistence, 0.962 in Table 3. This is the stylized fact underlying Stein’s overreaction regressions.

Third, the average annualized risk-neutral volatility of risk-neutral variance in (20) is much larger than the physical volatility of physical variance in (14). Fourth, (the absolute value of) the average of the risk-neutral leverage correlation defined as

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

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 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 6 further explores the models’ performance in matching the physical and risk-neutral volatility of variance. It plots the square root of the risk-neutral $Var_t^* (h^*(t + 1))$ 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 implied volatility root mean squared error (IV RMSE) and IV 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 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 important 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 pricing operator is a martingale under the physical probability measure. Let $B(t)$ be the risk-free bond ($\frac{dB(t)}{B(t)} = r dt$) 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

$$d \log(M) = \phi d \log(S) + \delta dt + \eta v dt + \xi dv$$

$$= \left[ \phi \left( r + \mu v - \frac{1}{2} v \right) + \delta + \eta v + \xi \kappa (\theta - v) \right] dt +$$

$$\left[ \phi \sqrt{v} + \xi \sigma \rho \sqrt{v} \right] d z_1 + \left[ \xi \sigma \sqrt{1 - \rho^2} \sqrt{v} \right] d z_2,$$

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

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

$$+ \left[ \phi \sqrt{v} + \xi \sigma \rho \sqrt{v} \right] d z_1 + \left[ \xi \sigma \sqrt{1 - \rho^2} \sqrt{v} \right] d z_2.$$

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

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

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

$$\delta = -(1 + \phi) r - \xi \kappa \theta.$$ 

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

27
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 \( \phi \). Using the fact that the drift of \( M \) is equal to \( -rMdt \), we have
\[
\mu v + (\phi + \xi \sigma) v = 0,
\]
or equivalently \( \phi = -\mu - \xi \sigma \rho \).

Finally, by equating the drift of \( U(t)M(t) \) to zero, we deduce the restriction on \( \xi \)
\[
0 = \left[ M \left( U_t + rSU_S + \mu vSU_S + \kappa(\theta - v)U_v + \frac{1}{2}vS^2U_{SS} + vS\sigma U_{Sv} \right) + \frac{1}{2}v\sigma^2U_{vv} \right) - rUM + Mv(\phi + \xi \sigma)(SU_S + \sigma U_v) + M\xi v\sigma^2(1 - \rho^2)U_v \right] dt.
\]
Since \( \frac{U(t)}{B(t)} \) is a martingale under the risk-neutral measure, we can also show that
\[
\lambda vU_v + rU = U_t + rSU_S + \kappa(\theta - v)U_v + \frac{1}{2}vS^2U_{SS} + vS\sigma U_{Sv} + \frac{1}{2}v\sigma^2U_{vv}.
\]
Substituting (A2) into (A1) we get for the drift \( D^{UM} \)
\[
D^{UM} = \left[ M \left( \lambda vU_v + rU + \mu vSU_S \right) - rUM 
+ Mv(\phi + \xi \sigma)(SU_S + \sigma U_v) + M\xi v\sigma^2(1 - \rho^2)U_v \right] dt
= 0.
\]
After simplification and using the fact that \( \phi + \xi \sigma \rho = -\mu \), we get
\[
\lambda vU_v - \mu v\sigma U_v + \xi v\sigma^2(1 - \rho^2)U_v = 0,
\]
and therefore
\[
\xi = \frac{\mu \sigma - \lambda}{\sigma^2(1 - \rho^2)}. \tag{A3}
\]
Using (A3) we also obtain
\[
\phi = -\mu + \lambda \sigma^{-1} \rho \frac{\rho}{1 - \rho^2}.
\]

Proof of Proposition 2

The discrete-time form of the pricing operator (9) is
\[
M(t) = M(0) \left( \frac{S(t)}{S(0)} \right)^\phi \exp \left( \delta t + \eta \sum_{s=1}^t h(s) + \xi (h(t + 1) - h(1)) \right), \tag{A4}
\]

28
and therefore

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

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

\[
\begin{align*}
\delta &= - (\phi + 1) r - \xi \omega + \frac{1}{2} \ln (1 - 2 \xi \alpha), \\
\eta &= - \left( \mu - \frac{1}{2} \right) \phi - \xi \alpha \gamma^2 + (1 - \beta) \xi - \frac{(\phi - 2 \xi \alpha \gamma)^2}{2 (1 - 2 \xi \alpha)}, \\
\phi &= -(\mu - \frac{1}{2} + \gamma)(1 - 2 \alpha \xi) + \gamma - \frac{1}{2}.
\end{align*}
\]

From the GARCH dynamic (12) we can write

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

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

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

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

\[
E_{t-1} \left[ \frac{M(t)}{M(t-1)} \right] = \exp(-r). \tag{A6}
\]
Note that
\[
E_{t-1} \left[ \frac{M(t)}{M(t-1)} \right] = \exp \left( \phi r + \delta + \xi \omega + \left[ \phi \left( \mu - \frac{1}{2} \right) + \eta + \xi (\beta - 1) + \xi \alpha^2 \right] h \right)
\]
\[
*E \left( \exp \left( \left[ \phi - 2\xi \alpha \gamma \right] \sqrt{h} z + [\xi \alpha] z^2 \right) \right).
\]

We need the following result
\[
E \left[ \exp \left( a z^2 + 2abz \right) \right] = \exp \left( -\frac{1}{2} \ln (1 - 2a) + \frac{2a^2b^2}{1 - 2a} \right).
\]

For our application we have
\[
a = \xi \alpha, \\
b = \left( \frac{\phi - 2\xi \alpha \gamma}{2\xi \alpha} \right) \sqrt{h},
\]
and thus
\[
2a^2b^2 = 2\xi^2 \alpha^2 \left( \frac{\phi - 2\xi \alpha \gamma}{2\xi \alpha} \right)^2 h = \frac{1}{2} (\phi - 2\xi \alpha \gamma)^2 h.
\]

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

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

Rearranging and using (A6)
\[
(\phi + 1) r + \delta + \xi \omega - \frac{1}{2} \ln (1 - 2\xi \alpha) + \left[ \phi \left( \mu - \frac{1}{2} \right) + \eta + \xi (\beta - 1) + \xi \alpha^2 + \frac{(\phi - 2\xi \alpha \gamma)^2}{2(1 - 2\xi \alpha)} \right] h = 0.
\]

Therefore we must have
\[
\delta = -(\phi + 1) r - \xi \omega + \frac{1}{2} \ln (1 - 2\xi \alpha),
\]
\[
\eta = -(\mu - \frac{1}{2}) \phi - \xi \alpha^2 + (1 - \beta) \xi - \frac{(\phi - 2\xi \alpha \gamma)^2}{2(1 - 2\xi \alpha)}.
\]
Now we use the Euler equation for the underlying index

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

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

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

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

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

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

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

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

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

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 pricing kernel takes the following quadratic form
in the stock return

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

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

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

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

$$\ln \left( \frac{M(t)}{M(t-1)} \right) = \phi R(t) + \delta + \eta h(t) + \xi (h(t + 1) - h(t))$$

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

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

expanding the square and collecting terms yields

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

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

From the equation for $\phi$ we have

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

and the result obtains.
References


Figure 1: One-Month Physical Conditional Histograms and Risk-Neutral Conditional Densities

Notes to Figure: We plot physical conditional histograms and risk-neutral 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 histogram, we use return innovations for 1990-2004, scale the time series by the Wednesday’s conditional volatility, and draw the histogram. On the horizontal axis are annualized returns in percent.
Figure 2: Log Ratios of Risk-Neutral One-Month Densities and Physical One-Month Histogram

Notes to Figure: We plot the natural logarithm of the ratio of the risk-neutral conditional densities and the physical conditional histogram. For each year in the option sample, we plot the ratios for each of the Wednesdays in that year. On the horizontal axis are annualized returns in percent.
Figure 3: 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.
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 5: Model Implied Log Ratio of the Risk-Neutral and Physical Density. Various Maturities.

Notes to Figure: We plot the natural logarithm of the ratio of the risk-neutral and physical conditional densities implied by the new GARCH model using solid lines and the special case of a zero variance premium using dashed lines. Each panel corresponds to a different maturity. We use the parameters in the third column of Table 3. The conditional variance is set to its unconditional level. The risk-free rate is set to 5% per year.
Figure 6: 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)

<table>
<thead>
<tr>
<th></th>
<th>1990-2005</th>
<th>1996-2004</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>7.77</td>
<td>7.52</td>
</tr>
<tr>
<td>St. Deviation</td>
<td>16.08</td>
<td>19.01</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.102</td>
<td>-0.094</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>6.786</td>
<td>5.696</td>
</tr>
</tbody>
</table>

Panel B. Option Data by Moneyness

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

Panel C. Option Data by Maturity

<table>
<thead>
<tr>
<th></th>
<th>DTM&lt;30</th>
<th>30&lt;DTM&lt;60</th>
<th>60&lt;DTM&lt;90</th>
<th>90&lt;DTM&lt;120</th>
<th>120&lt;DTM&lt;180</th>
<th>DTM&gt;180</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Contracts</td>
<td>1,069</td>
<td>5,717</td>
<td>3,996</td>
<td>1,726</td>
<td>3,253</td>
<td>5,630</td>
<td>21,391</td>
</tr>
<tr>
<td>Average Price</td>
<td>12.58</td>
<td>17.49</td>
<td>23.61</td>
<td>26.34</td>
<td>31.81</td>
<td>44.62</td>
<td>28.42</td>
</tr>
<tr>
<td>Average Spread</td>
<td>0.87</td>
<td>1.14</td>
<td>1.36</td>
<td>1.44</td>
<td>1.53</td>
<td>1.76</td>
<td>1.41</td>
</tr>
</tbody>
</table>

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

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

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.

<table>
<thead>
<tr>
<th>Physical Parameters</th>
<th>No Risk Premium</th>
<th>Equity Risk Only</th>
<th>Equity and Volatility Risk</th>
<th>Ad Hoc Specification</th>
</tr>
</thead>
<tbody>
<tr>
<td>ω</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>α</td>
<td>3.248E-06</td>
<td>3.249E-06</td>
<td>1.546E-06</td>
<td>3.452E-06</td>
</tr>
<tr>
<td>β</td>
<td>0.826</td>
<td>0.826</td>
<td>0.826</td>
<td>0.875</td>
</tr>
<tr>
<td>γ</td>
<td>220.9</td>
<td>219.4</td>
<td>318.5</td>
<td>158.7</td>
</tr>
<tr>
<td>µ</td>
<td>0</td>
<td>1.501</td>
<td>1.543</td>
<td>1.954</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>RN Parameters</th>
<th>1/(1−2αξ)</th>
<th>1</th>
<th>1</th>
<th>1.2039</th>
</tr>
</thead>
<tbody>
<tr>
<td>ω*</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>α*</td>
<td>3.248E-06</td>
<td>3.249E-06</td>
<td>3.248E-06</td>
<td>3.248E-06</td>
</tr>
<tr>
<td>β*</td>
<td>0.826</td>
<td>0.826</td>
<td>0.826</td>
<td>0.826</td>
</tr>
<tr>
<td>γ*</td>
<td>220.93</td>
<td>220.93</td>
<td>220.78</td>
<td>220.93</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Likelihood</th>
<th>Total</th>
<th>55,366.1</th>
<th>55,368.4</th>
<th>55,485.5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>From returns</td>
<td>13,188.3</td>
<td>13,190.6</td>
<td>13,307.7</td>
</tr>
<tr>
<td></td>
<td>From options</td>
<td>42,177.8</td>
<td>42,177.8</td>
<td>42,177.8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Physical Properties</th>
<th>Long run volatility</th>
<th>0.234</th>
<th>0.219</th>
<th>0.153</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Daily autocorrelation, h(t)</td>
<td>0.985</td>
<td>0.983</td>
<td>0.983</td>
</tr>
<tr>
<td></td>
<td>Annualized volatility of h(t)</td>
<td>0.087</td>
<td>0.081</td>
<td>0.039</td>
</tr>
<tr>
<td></td>
<td>Correlation(R(t),h(t+1))</td>
<td>-0.977</td>
<td>-0.974</td>
<td>-0.974</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>RN Properties</th>
<th>Long run volatility</th>
<th>0.234</th>
<th>0.234</th>
<th>0.232</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Daily autocorrelation, h*(t)</td>
<td>0.985</td>
<td>0.985</td>
<td>0.985</td>
</tr>
<tr>
<td></td>
<td>Annualized volatility of h*(t)</td>
<td>0.087</td>
<td>0.087</td>
<td>0.086</td>
</tr>
<tr>
<td></td>
<td>Correlation(R(t),h*(t+1))</td>
<td>-0.977</td>
<td>-0.977</td>
<td>-0.977</td>
</tr>
</tbody>
</table>

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 µ=0 and ξ=0. The model with "Equity Risk Only" has five parameters. It imposes ξ=0. The model with "Equity and Volatility Risk" has six parameters, and estimates µ and ξ. All volatility 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.

| Model            | F/X<0.96 | 0.96<|F/X<1.02 | 1.02<|F/X<1.04 | 1.04<|F/X<1.06 | F/X>1.06 | All     |
|------------------|----------|------|-----------|---------|-----------|---------|---------|---------|
| 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  |

| Model            | F/X<0.96 | 0.96<|F/X<1.02 | 1.02<|F/X<1.04 | 1.04<|F/X<1.06 | F/X>1.06 | All     |
|------------------|----------|------|-----------|---------|-----------|---------|---------|---------|
| 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  |

<table>
<thead>
<tr>
<th>Model</th>
<th>DTM&lt;30</th>
<th>30&lt;DTM&lt;60</th>
<th>60&lt;DTM&lt;90</th>
<th>90&lt;DTM&lt;120</th>
<th>120&lt;DTM&lt;180</th>
<th>DTM&gt;180</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>ERP and VRP</td>
<td>3.3513</td>
<td>3.4181</td>
<td>3.4614</td>
<td>2.9799</td>
<td>3.4059</td>
<td>3.3447</td>
<td>3.3685</td>
</tr>
<tr>
<td>ERP only</td>
<td>4.7976</td>
<td>5.4619</td>
<td>5.8224</td>
<td>5.6526</td>
<td>5.8944</td>
<td>5.9440</td>
<td>5.7111</td>
</tr>
<tr>
<td>No RP</td>
<td>5.0606</td>
<td>5.7668</td>
<td>6.1823</td>
<td>6.0429</td>
<td>6.3122</td>
<td>6.4292</td>
<td>6.0986</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>DTM&lt;30</th>
<th>30&lt;DTM&lt;60</th>
<th>60&lt;DTM&lt;90</th>
<th>90&lt;DTM&lt;120</th>
<th>120&lt;DTM&lt;180</th>
<th>DTM&gt;180</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>ERP and VRP</td>
<td>0.1206</td>
<td>0.7148</td>
<td>0.5890</td>
<td>0.4085</td>
<td>0.0273</td>
<td>-0.5019</td>
<td>0.2121</td>
</tr>
<tr>
<td>ERP only</td>
<td>3.7334</td>
<td>4.4842</td>
<td>4.8342</td>
<td>4.9162</td>
<td>4.9451</td>
<td>5.0505</td>
<td>4.7661</td>
</tr>
<tr>
<td>No RP</td>
<td>4.0889</td>
<td>4.8651</td>
<td>5.2597</td>
<td>5.3583</td>
<td>5.4299</td>
<td>5.5997</td>
<td>5.2190</td>
</tr>
</tbody>
</table>

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.