Volatility Dynamics for the S&P500: Evidence from Realized Volatility, Daily Returns, and Option Prices

Peter Christoffersen
McGill University, CBS, and CREATES

Kris Jacobs
University of Houston, McGill University, and Tilburg University

Karim Mimouni
American University in Dubai

Most recent empirical option valuation studies build on the affine square root (SQR) stochastic volatility model. The SQR model is a convenient choice, because it yields closed-form solutions for option prices. We investigate alternatives to the SQR model, by comparing its empirical performance with that of five different but equally parsimonious stochastic volatility models. We provide empirical evidence from three different sources: realized volatilities, S&P500 returns, and an extensive panel of option data. The three sources of data all point to the same conclusion: the best volatility specification is one with linear rather than square root diffusion for variance. This model captures the stylized facts in realized volatilities, it performs well in fitting various samples of index returns, and it has the lowest option implied volatility mean squared error in and out of sample. (*JEL* G12)

1. Introduction

The empirical literature on stochastic volatility modeling can be broadly categorized in three groups. One set of studies searches across volatility specifications to find the model that provides the best description of the conditional distribution of daily returns.[1] Another literature uses intraday returns to construct daily realized volatilities that are then used to assess the distribution of

Christoffersen and Jacobs are also affiliated with CIRANO and CIREQ and want to thank FQRSC, IFM^2, and SSHRC for financial support. Mimouni was supported by IFM^2 and FQRSC. We are grateful for helpful comments from Yacine Aït-Sahalia and Raman Uppal (the Editors), Torben Andersen, Mikhail Chernov, Jerome Detemple, Jin Duan, Bjorn Eraker, Bob Goldstein, Jeremy Graveline, Mark Kamstra, Nour Meddahi, Marcel Rindisbacher, Stephen Taylor, and two anonymous referees. Send correspondence to Peter Christoffersen, Desautels Faculty of Management, McGill University, 1001 Sherbrooke Street West, Montreal, Quebec, Canada, H3A 1G5; telephone: (514) 398-2869; fax: (514) 398-3876. E-mail: peter.christoffersen@mcgill.ca.

[1] Andersen, Benzoni, and Lund (2002), Chacko and Viceira (2003), and Chernov et al. (2003) are prominent papers in this literature.

doi:10.1093/rfs/hhq032 Advance Access publication April 29, 2010

volatility and to forecast future realized volatility.[2] A third group of papers uses stochastic volatility models to fit option prices across strike prices, maturities, and time. Bates (1996) and Bakshi, Cao, and Chen (1997) are important early contributions in the third category.

While the first two groups of papers investigate affine and nonaffine volatility dynamics, the papers in the third category have almost exclusively modeled stochastic volatility using an affine or "square root" structure. In particular, the Heston (1993) model, which accounts for time-varying volatility and a leverage effect, has been implemented in a large number of empirical studies.[3] Henceforth, we will refer to this model as the square root (SQR) model. The SQR model is convenient, because it leads to (quasi) closed-form solutions for prices of European options. It is not surprising that the choice of model is partly driven by convenience: the estimation of option valuation models on large option data sets is computationally burdensome because of the need to filter the latent stochastic volatility. As the solution to the Heston (1993) SQR model relies on univariate numerical integration, it is relatively easier to estimate than many other models that require Monte Carlo simulation to compute option prices.

The SQR model cannot capture important stylized facts. In order to address the limitations of the square root structure, the model is often combined with models of jumps in returns and/or volatility, or multiple volatility components are used.[4] However, relatively few studies analyze nonaffine stochastic volatility models, and therefore, much less is known about the empirical biases that result from imposing the affine square root structure on the stochastic volatility dynamic in the first place. Notable exceptions that investigate option valuation using nonaffine stochastic volatility models are Aït-Sahalia and Kimmel (2007), Jones (2003), and Benzoni (2002). The nonaffine model in Benzoni (2002) does not improve on the performance of the Heston (1993) model. Jones (2003) analyzes the more general constant elasticity of variance (CEV) model using a bivariate time series of returns and an at-the-money short maturity option, and a number of his specification tests favor the nonaffine CEV model over the SQR model. Aït-Sahalia and Kimmel (2007) also estimate different models using joint time series of the underlying return and a short-dated at-the-money option price or implied volatility. They also find that the SQR model is misspecified, but the nature of the misspecification and the empirical findings are different from those described in Jones (2003).

---

[2] See Andersen et al. (2001).

[3] The leverage effect was first characterized in Black (1976). For empirical studies that emphasize the importance of volatility clustering and the leverage effect for option valuation, see, among others, Benzoni (2002), Chernov and Ghysels (2000), Eraker (2004), Heston and Nandi (2000), and Pan (2002).

[4] For empirical studies that implement the Heston (1993) model by itself or in combination with different types of jump processes, see, for example, Andersen, Benzoni, and Lund (2002), Bakshi, Cao, and Chen (1997), Bates (1996, 2000), Broadie, Chernov, and Johannes (2007), Benzoni (2002), Chernov and Ghysels (2000), Huang and Wu (2004), Pan (2002), Eraker (2004), and Eraker, Johannes, and Polson (2003). Bates (2000) also introduces an SQR model with two volatility components.

This article further investigates the empirical implications of adopting an SQR stochastic volatility model for the modeling of the return distribution and option valuation. We compare the empirical performance of the Heston (1993) SQR stochastic volatility model with that of five simple alternatives. In contrast with Aït-Sahalia and Kimmel (2007) and Jones (2003), who compare the SQR model with more richly parameterized stochastic volatility models, we deliberately choose alternative specifications with the same number of parameters as the SQR model. Our approach also differs from that of several other studies that enrich the SQR model by adding jumps to the basic model, which also increases the number of model parameters. We instead investigate if simple alternatives to the SQR stochastic volatility model work better empirically when modeling realized volatilities, returns, and option prices. Part of our motivation for keeping the number of parameters constant is our desire to make the in-sample comparison between models as meaningful as possible. In-sample model comparisons favor relatively richly parameterized models, but these in-sample model rankings can easily be reversed in out-of-sample experiments.

We conduct an extensive empirical investigation on the six stochastic volatility specifications using several different data sources. First, we use realized S&P500 volatilities and the VIX option implied volatility index to assess the distributional properties of the SQR model and to guide us in the search for alternative specifications. Second, we estimate the model parameters using maximum likelihood on index returns only. Third, we employ nonlinear least squares on a time series of cross-sections of out-of-the-money (OTM) S&P500 option data with extensive cross-sectional variation across moneyness and maturity. Fourth, we value options using realized volatility and VIX data as proxies for spot volatility. Finally, we investigate the robustness of the option valuation results with respect to the filtering method used. All evidence points to the same conclusion: the affine SQR model is outperformed by the simple nonaffine models we consider. Overall, the best of the alternative volatility specifications is a model we refer to as the ONE model, which is of the GARCH diffusion type. This model captures the stylized facts in realized volatilities, and it provides the best fit for two of the three return samples. When estimating the models using option data, it has the lowest option price mean squared errors in and out of sample.

Our estimation on option contracts uses a rich panel of data. It therefore critically differs from that of Aït-Sahalia and Kimmel (2007) and Jones (2003), who estimate model parameters using a bivariate time series. Indeed, to the best of our knowledge, our analysis of the SQR model uses substantially larger cross-sections of option contracts than any other available study that explicitly filters volatility using returns. Existing studies either filter volatility using returns and use a small cross-section of option data or use a large cross-section and do not filter volatility on returns. For example, Aït-Sahalia and Kimmel (2007), Chernov and Ghysels (2000), Eraker (2004), and Jones (2003) explicitly filter volatility on returns using a relatively limited cross-section of op-

tions. Studies that investigate a wider cross-section include Bakshi, Cao, and Chen (1997), who estimate parameters one day at a time, and Bates (2000) and Huang and Wu (2004), who estimate a separate volatility parameter for each day without filtering on returns.

We are able to estimate a large number of stochastic volatility models using comprehensive option data sets because of our use of the particle filter (PF). This methodology is extensively used in the engineering literature, and it has also recently been used for financial applications.[5] Particle filtering provides a convenient tool for the analysis of latent factor models such as stochastic volatility models. It is easy to implement, and it can be adapted to provide the best possible fit to any objective function and any model of interest. In our opinion, the trade-off with other empirical methods is a very favorable one in the context of computationally intensive option valuation problems.

The article proceeds as follows. In Section 2, we discuss the benchmark Heston (1993) model in light of the distributional evidence from realized volatilities. We also use realized volatilities and the VIX to help guide the search for alternative specifications. Section 3 discusses PF-based estimation on returns and options. Section 4 presents the empirical results obtained when estimating the SV models, as well as extensions including return jumps, on index return data. Section 5 contains the empirical results from estimating the models on index option data. Section 6 concludes. Evidence on the finite sample properties of the PF-based estimator and on Monte Carlo-based option pricing are provided in the Appendix.

2. Stochastic Volatility Specifications

This section discusses alternatives to the SQR volatility specification in Heston (1993), which has become the standard building block for more elaborate models in the option valuation literature. Note that it is relatively straightforward to write down more heavily parameterized models that outperform the SQR model. Trivially, a more heavily parameterized model will always outperform a simpler nested model in sample. Moreover, Bates (2003) points out that even in short-horizon out-of-sample experiments, the richer models may outperform nested models, because the "smile" or "smirk" patterns in option prices are very persistent. This result is more likely if the models are reestimated frequently, say daily, as in the classic study by Bakshi, Cao, and Chen (1997).

Thus, we deliberately confine attention to models that have the same number of parameters as Heston's SQR model, and we estimate on long samples of data. The models we consider can be thought of as alternative building blocks for more elaborate models containing jumps in returns and volatility, as well as

[5] See Gordon, Salmond, and Smith (1993) and Pitt and Shephard (1999). See Johannes, Polson, and Stroud (2009) for an application to returns data.

multiple volatility factors. Although we consider jump models below, the main objective of our article is to establish a well-specified volatility dynamic, which in our opinion ought to precede the development of more elaborate versions of a possibly misspecified basic model.

2.1 Realized variance: implications for the SQR model

The Heston (1993) SQR model assumes that the instantaneous change in variance, V, has the following dynamics,

$$\text{SQR: } \mathrm{d}V = \kappa(\theta - V)\mathrm{d}t + \sigma\sqrt{V}\mathrm{d}w, \tag{1}$$

where κ denotes the speed of mean reversion, θ is the unconditional variance, and σ determines the variance of variance. The variance innovation $\mathrm{d}w$ is a Brownian motion, and we have $corr(\mathrm{d}z, \mathrm{d}w) = \rho$, where $\mathrm{d}z$ is the innovation in the underlying spot price, S, given by

$$\mathrm{d}S = \mu S\mathrm{d}t + \sqrt{V}S\mathrm{d}z, \tag{2}$$

where μ is the mean of the instantaneous rate of return. We will use the return specification (2) in each of the volatility models considered below.

In order to explore the implications of the SQR specification, consider the instantaneous volatility dynamic implied by the model. Using Ito's lemma, we can write

$$\text{SQR: } \mathrm{d}\sqrt{V} = \mu(V)\mathrm{d}t + \frac{1}{2}\sigma\mathrm{d}w, \tag{3}$$

where the volatility drift $\mu(V)$ is a function of the variance level. Note that the SQR model implies that the instantaneous change in volatility should be Gaussian and homoscedastic: V does not show up in any way in the diffusion term for $\mathrm{d}\sqrt{V}$. This is a strong implication, which can be quite easily evaluated empirically.

As a first piece of empirical evidence, we construct daily realized variances, RV_t, from intraday returns. We obtain intraday S&P500 quotes for 1996 through 2004, from which we construct a grid of two-minute returns. From these two-minute returns, we construct robust realized variances using the two-scale estimator from Zhang, Mykland, and Aït-Sahalia (2005).[6] The two-scale estimator is defined as

$$RV_t^{TS} = RV_t^{Avr} - 0.0614RV_t^{All},$$

where RV_t^{Avr} is the average of all the possible low-frequency RV estimates that use (in our case) thirty-minute squared returns on the two-minute grid, and where RV_t^{All} is the high-frequency estimator summing over all the available

[6] We scale up the open-to-close realized variance estimates to match the overall variance of close-to-close daily returns for the sample period.

two-minute squared returns.[7] The coefficient on RV_t^{All} depends on the frequency of the sparse estimators (thirty minutes here) and of the high-frequency estimator (two minutes here). Zhang, Mykland, and Aït-Sahalia (2005) show that even in the presence of market microstructure noise, this estimator converges to the true integrated variance defined by $\int_{t-1}^{t} V_\tau \mathrm{d}\tau$.

Consider the left column of figure 1. Using daily realized volatilities from 1996 through 2004, the top-left panel of figure 1 shows a quantile–quantile (QQ) plot of the daily realized volatility changes compared with the Gaussian distribution. The deviation of the data points from the straight line indicates that the Gaussian distribution is not a good assumption for daily changes in volatility. The observed tails (both left and right) are considerably fatter than the normal distribution would suggest. Now, consider the middle-left panel in figure 1, which scatter plots the daily volatility changes against the daily volatility level. According to the SQR model, this scatter plot should not display any systematic patterns. However, as the volatility level increases on the horizontal axis, a cone-shaped pattern in the daily volatility changes on the vertical axis is apparent. The bottom-left panel of figure 1 confirms this finding: it scatter plots the absolute daily volatility changes against the daily volatility level. A simple OLS regression line is shown for reference. Notice the apparent positive relationship between the volatility level and the magnitude of the volatility changes. This pattern is in conflict with the homoscedastic volatility implication of the SQR model.

Using Ito's lemma, the dynamics for log variance in the SQR model can be written as

$$\text{SQR: } \mathrm{d}\ln(V) = \mu(V)\mathrm{d}t + \sigma\frac{1}{\sqrt{V}}\mathrm{d}w. \tag{4}$$

The right column of figure 1 summarizes the empirics of the logarithms of the realized variances. The top panel shows that daily changes in the realized log variances follow the Gaussian distribution closely. The result clearly differs from the corresponding top-left panel for the daily volatility changes. Furthermore, the scatter of daily log variance changes against the log variance level in the middle-right panel of figure 1 does not reveal a cone-shaped pattern. The bottom-right panel in figure 1 confirms this result: it shows a virtually horizontal line when regressing absolute changes in log variance on log variance levels. The absence of a clear relationship between changes in the log variances and the log variance level in figure 1 casts further doubt on the SQR specification, for which the instantaneous changes in log variances are heteroscedastic, as is evident from Equation (4).

[7] See Aït-Sahalia, Mykland, and Zhang (2005) for a discussion of the optimal sampling frequency.

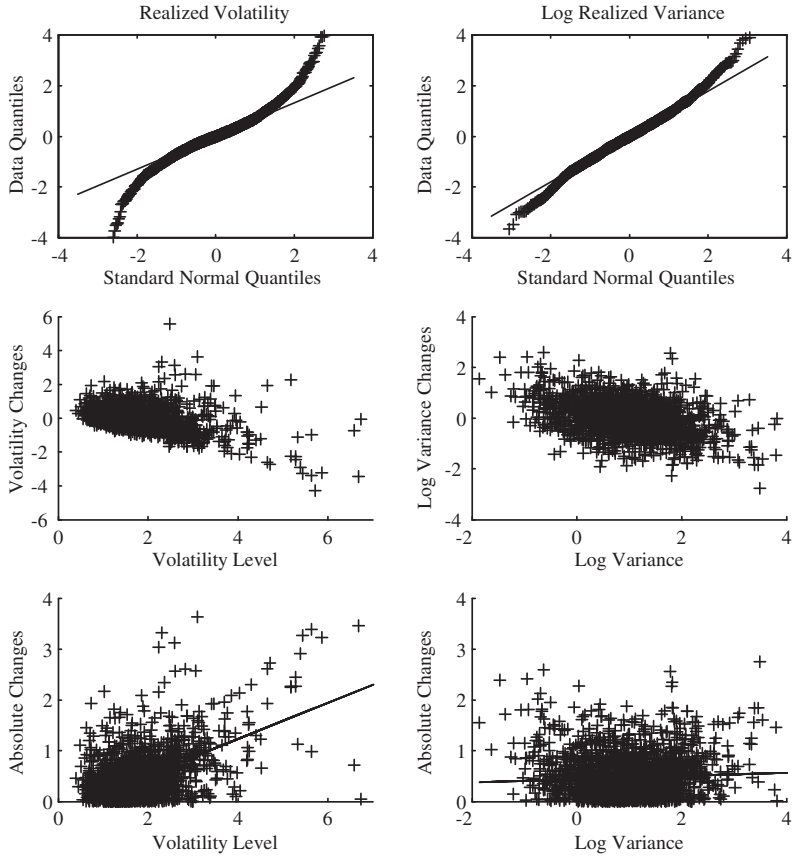

Figure 1
Diagnostics of realized volatility and log realized variance, 1996–2004
Using realized volatilities (left column) and log realized variances (right column) from 1996 through 2004, the top panels plot the quantiles of the daily changes against quantiles from the normal distribution. The middle panels scatter plot the daily changes against the daily levels. The bottom panels scatter plot the absolute daily changes against the daily levels and show an OLS regression line for reference.

2.2 Alternative specifications for variance dynamics

The evidence in figure 1 suggests specifying the variance dynamic as follows:

$$\text{ONE: } \mathrm{d}V = \kappa(\theta - V)\mathrm{d}t + \sigma V \mathrm{d}w, \tag{5}$$

which has the property that $\ln(V)$ is homoscedastic.

$$\mathrm{d}\ln(V) = \mu(V)\mathrm{d}t + \sigma \mathrm{d}w.$$

Heston (1997) and Lewis (2000) suggest another alternative to the SQR model: the "three halves" model defined by

$$3/2\text{N: } dV = \kappa V(\theta - V)dt + \sigma V^{3/2}dw. \quad (6)$$

In addition to a different power on V in the diffusion term, the 3/2N model has the interesting implication that the variance drift is nonlinear in the level of the variance. We highlight this feature by denoting the model "3/2N". This allows for a potentially fast reduction in volatility following unusually large volatility spikes, as in October 1987, and for slow increases in volatility during prolonged low-volatility episodes, as occurred in the mid-1990s. The 3/2N model also leads to a closed-form solution for European call option prices (see Lewis 2000). However, to the best of our knowledge, the model has not yet been implemented using this solution.

So far, we have considered three different specifications of the diffusion and two different specifications for the drift of V. We can think of the three models considered so far as belonging to the class

$$dV = \kappa V^a(\theta - V)dt + \sigma V^b dw, \text{ for } a = \{0, 1\} \text{ and } b = \{1/2, 1, 3/2\}. \quad (7)$$

This framework encompasses a total of six models, which we denote

a	b	Name
0	1/2	SQR
1	1/2	SQRN
0	1	ONE
1	1	ONEN
0	3/2	3/2
1	3/2	3/2N

All of these models will be estimated below.[8]

Regarding the properties of these stochastic volatility models, the literature contains several good discussions of the properties of stochastic differential equations (SDEs). See, for instance, Karlin and Taylor (1981). Aït-Sahalia (1996) contains a discussion of the properties of general interest rate processes, and Lewis (2000) contains a treatment of certain variance processes. Jones (2003) provides an excellent discussion of the CEV stochastic variance process. He shows that for a model with $a = 0$ and $b > 1$, zero and plus infinity are inaccessible values for the stochastic variance. He also shows that a solution to the variance SDE exists and is unique, and that the stationary distribution exists. Thus, Jones (2003) covers the model we refer to as 3/2. Similar

[8] Although it would seem natural from the realized variance analysis above, we do not consider the homoscedastic log SV model in which $d\ln(V) = \kappa(\theta - \ln(V)) + \sigma dw$, because it is not nested in the class of models we consider, and it typically performs very similar to—but slightly worse than—the ONE model in Equation (5).

arguments can be used to show that these results hold also for the ONE model, the ONEN model, and the 3/2N model. For the SQRN model, all but the existence of the stationary distribution can be shown using similar arguments as well.

2.3 Volatility risk premia

In order to compute option prices, which depend on the price of volatility risk, we need to specify a volatility risk premium relationship for each of the above six models. Heston (1993) assumes that the volatility risk premium $\lambda(S, V, t)$ is equal to λV, so that the risk-neutral dynamic for the SQR model is

$$\text{SQR: } \mathrm{d}S = rS\mathrm{d}t + \sqrt{V}S\,\mathrm{d}z^* \tag{8}$$

$$\begin{aligned}\mathrm{d}V &= (\kappa(\theta - V) + \lambda V)\,\mathrm{d}t + \sigma\sqrt{V}\mathrm{d}w^* \\ &= (\kappa - \lambda)(\kappa\theta/(\kappa - \lambda) - V)\mathrm{d}t + \sigma\sqrt{V}\mathrm{d}w^*,\end{aligned} \tag{9}$$

where $\mathrm{d}z^*$ and $\mathrm{d}w^*$ are Brownian motions under the risk-neutral measure, with $corr(\mathrm{d}z^*, \mathrm{d}w^*) = \rho$. Notice that the variance dynamic takes the same form under the physical and risk-neutral measures.

To provide some intuition for the properties of risk-neutral volatility, figure 2 plots various stylized facts for the model-free, option-implied VIX volatility index, provided by the CBOE. The VIX is measured in annualized percentage standard deviations. Similar to figure 1, we report the QQ plot of the VIX in the top-left panel. The daily VIX changes are scattered against the VIX level in the center-left panel, and the daily absolute VIX changes are scattered against the VIX level in the bottom-left panel. In the right column, we provide the same analysis using the natural logarithm of the VIX rather than the VIX level.[9] The similarities to the RV-based plots in figure 1 are quite striking. The level of the VIX displays strong evidence of nonnormality and heteroscedasticity. The distribution of the logarithm of the VIX appears to be much closer to the normal distribution, and shows much less evidence of heteroscedasticity.

A key lesson from figure 2 is that the (model-free) risk-neutral volatility distribution has similar features to the (model-free) physical volatility distribution in figure 1. Therefore, when deciding on the risk premium specification for the nonaffine models, we choose a form that preserves the functional form of the spot variance process under the two measures. In particular, we use the following functional form for the variance risk premium:

$$\lambda(S, V, t) = \lambda V^{a+1}. \tag{10}$$

[9] Note that the analysis for the natural logarithm of the VIX is equivalent to the analysis of the natural logarithm of the squared VIX.

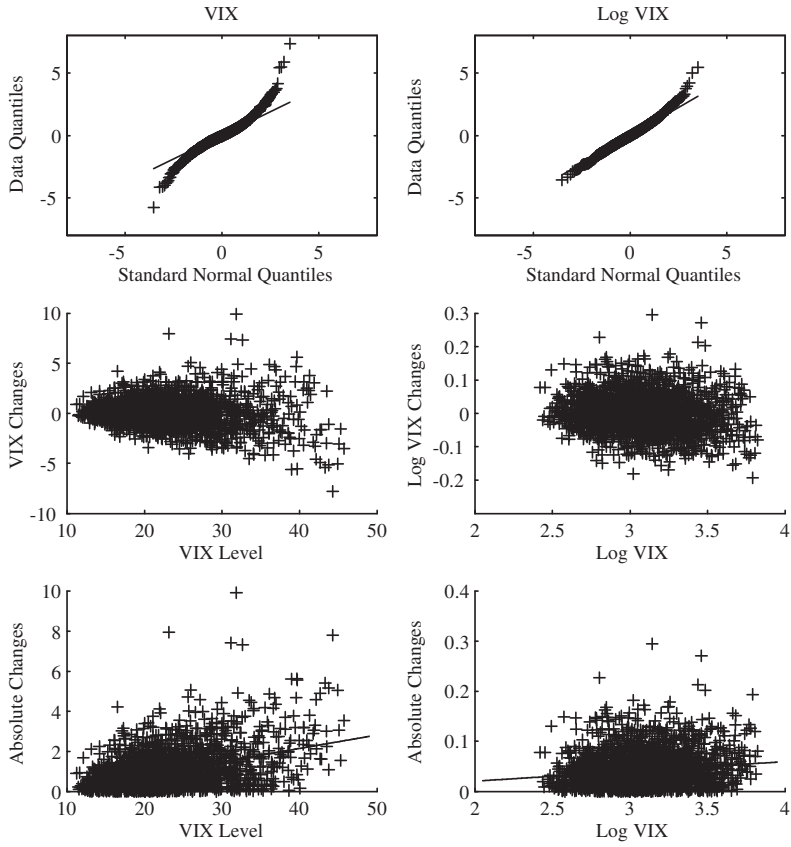

Figure 2
Diagnostics of option implied volatility (VIX) and log VIX, 1996–2004
Using VIX (left column) and log VIX (right column) from 1996 through 2004, the top panels plot the quantiles of the daily changes against quantiles from the normal distribution. The middle panels scatter plot the daily changes against the daily levels. The bottom panels scatter plot the absolute daily changes against the daily levels and show an OLS regression line for reference.

This specification ensures that the variance dynamic takes the same form under the two measures, as is the case in most available empirical studies that estimate the Heston (1993) SQR model. Under the risk-neutral measure, we have

$$\begin{aligned} \mathrm{d}S &= rS\mathrm{d}t + \sqrt{V}S\mathrm{d}z^* \\ \mathrm{d}V &= (\kappa - \lambda)V^a(\kappa\theta/(\kappa - \lambda) - V)\mathrm{d}t + \sigma V^b \mathrm{d}w^* \end{aligned} \tag{11}$$

with $corr(\mathrm{d}z^*, \mathrm{d}w^*) = \rho$.

In all six models we investigate, the risk neutralization can be obtained using a no-arbitrage argument. A utility-based argument behind the risk neutraliza-

tion can be explicitly characterized for some cases when assuming log utility. See Lewis (2000) for a thorough discussion of these issues.

Note that the risk-neutral process in Equation (11) can be rewritten

$$\begin{aligned} dS &= rSdt + \sqrt{V}Sdz^* \\ dV &= \kappa^* V^a(\theta^* - V)dt + \sigma V^b dw^*, \end{aligned} \tag{12}$$

where $\kappa^* = (\kappa - \lambda)$, and $\theta^* = \kappa\theta/\kappa^*$. This representation highlights that λ cannot be identified in any of the models when using only options in estimation.

2.4 Computing option prices

Heston (1993) demonstrates that the SQR model admits a closed-form solution for option prices, which can be written as

$$C(V_t) = S_t P_{1,t} - Xe^{-r\tau} P_{2,t}, \tag{13}$$

where τ is the option's time to maturity, X is the strike price, and $P_{1,t}$ and $P_{2,t}$ are computed using numerical integration of the conditional characteristic function.

Heston (1997) and Lewis (2000) show that for the 3/2N model, a similar closed-form solution is available, but in this case, the characteristic function involves the gamma and confluent hypergeometric function with complex arguments, which is very numerically intensive. Barone-Adesi, Rasmussen, and Ravanelli (2005) and Sabanis (2003) develop approximate analytical solutions for the ONE model, but as far as we know, an exact solution is not available.

In order to ensure that differences across models are not driven by a particular numerical technique, we compute all model option prices using Monte Carlo. Thus, the call option prices are computed via the Monte Carlo sample analogue of the discounted expectation

$$C(V_t) = e^{-r\tau} E_t^* \left[S_{t+\tau} - X, 0 \right], \tag{14}$$

where the expectation is calculated using the risk-neutral measure.

In order to calculate Monte Carlo prices, we use one thousand simulated paths and a number of numerical techniques to increase numerical efficiency—namely, the empirical martingale method of Duan and Simonato (1998), stratified random numbers, antithetic variates, and a Black–Scholes control variate technique. In the Appendix, we use the SQR model to demonstrate that the Monte Carlo prices match closely the prices computed via numerical integration.

3. Estimation Using the Particle Filter

The SQR model has been investigated empirically in a large number of studies. It is often used as a building block, together with models of jumps in return

and volatility. For our purpose, it is important to note that when estimating the model on option data, a number of different techniques can be used. First, the model's parameters can be estimated using a single cross-section of option prices treating V_t as a parameter. This is done, for example, in Bakshi, Cao, and Chen (1997). A second type of implementation of the SQR model uses multiple cross-sections of option prices but does not use the information in the underlying asset returns. Instead, for every cross-section, a different initial volatility is estimated, leading to a highly parameterized problem. Bates (2000) and Huang and Wu (2004), for instance, take this approach. A third group of papers provides a likelihood-based analysis of the stochastic volatility model. See, for example, Aït-Sahalia and Kimmel (2007), Bates (2006), Jones (2003), and Eraker (2004), who provides a Markov chain Monte Carlo analysis. A likelihood-based approach can combine information from the option data and the underlying returns, and impose consistency between the physical and risk-neutral dynamics in estimation. Both the return data and the option data carry a certain weight in the objective function. Finally, Chernov and Ghysels (2000) use the efficient method of moments, and Pan (2002) uses a method of moments technique as well. These methods can also combine information in option data with the information in underlying returns.

The empirical challenge in stochastic volatility models is that the spot volatility V_t is an unobserved latent factor. Thus, in order to estimate the parameters in stochastic volatility models, we need to apply a filtering technique using observed index returns. The last two groups of papers discussed above explicitly filter volatility using returns. The explicit filtering of volatility is also considered in some papers that use return data to estimate continuous-time stochastic volatility models, such as Chernov et al. (2003).

We provide an alternative implementation of the filtering problem for the six stochastic volatility models under investigation using the PF algorithm. As shown by Gordon, Salmond, and Smith (1993), the PF offers a convenient filter for nonlinear models such as the stochastic volatility models we consider here. The PF has recently been applied by Johannes, Polson, and Stroud (2009), who use return data to estimate continuous-time stochastic volatility models with jumps. But because the PF procedure is relatively new in finance, we discuss the implementation of this method in some detail.

We use the PF both for maximum likelihood estimation on returns and for nonlinear least squares estimation on option prices. Below, we describe first the PF, then the maximum likelihood importance sampling (MLIS) estimation methodology we use to estimate the models using return data, and finally the nonlinear least squares importance sampling (NLSIS) estimation methodology we use to estimate the models using option data.

3.1 Volatility discretization

Working with log returns, we can write the generic SV process as

$$\mathrm{d}\ln(S) = \left(\mu - \frac{1}{2}V\right)\mathrm{d}t + \sqrt{V}\mathrm{d}z \tag{15}$$

$$\mathrm{d}V = \kappa V^a(\theta - V)\mathrm{d}t + \sigma V^b \mathrm{d}w. \tag{16}$$

Note that Equations (15) and (16) specify how the unobserved state is linked to observed stock prices. This relationship allows us to infer the volatility path using returns data. We first need to discretize Equations (15)–(16). There are different discretization methods, and every scheme has certain advantages and drawbacks. We use the Euler scheme, which is easy to implement and has been found to work well for this type of application.[10] Discretizing Equations (15)–(16) gives

$$\ln(S_{t+1}) = \ln(S_t) + \left(\mu - \frac{1}{2}V_t\right) + \sqrt{V_t}z_{t+1} \tag{17}$$

$$V_{t+1} = V_t + \kappa V_t^a(\theta - V_t) + \sigma V_t^b w_{t+1} \tag{18}$$

with z_{t+1} and w_{t+1} standard normal. We implement the discretized model in Equations (17) and (18) using daily returns, but all parameters will be expressed in annual units below.

3.2 Volatility filtering

The PF algorithm relies on the approximation of the true density of the state V_{t+1} by a set of N discrete points or particles that are updated iteratively through Equations (17) and (18). The filter is implemented using particles or draws, $\left\{V_{t+1}^j\right\}_{j=1}^N$, from the empirical distribution of V_{t+1}, conditional on particles or draws, $\left\{V_t^j\right\}_{j=1}^N$, from the empirical distribution of V_t. Our particular implementation of the PF is referred to as the sampling-importance-resampling (SIR) PF and follows Pitt (2002).

For a sample size T, we recursively proceed through the following three steps for $t = 1, ..., T-1$.

3.2.1 Step 1: simulating the state forward: sampling. We first simulate the state forward by computing N raw particles $\{\tilde{V}_{t+1}^j\}_{j=1}^N$ from the set of smooth resampled particles $\{V_t^j\}_{j=1}^N$ using Equation (18), taking the correlation be-

[10] See, for example, Eraker (2001).

tween returns and variance into account. We fix the particles in the first period as $V_1^j = \theta$, for all j.[11]

We can then get correlated shocks via

$$z_{t+1}^j = \left(\ln(S_{t+1}) - \ln(S_t) - \left(\mu - \frac{1}{2}V_t^j\right)\right)/\sqrt{V_t^j} \tag{19}$$
$$w_{t+1}^j = \rho z_{t+1}^j + \sqrt{1-\rho^2}\varepsilon_{t+1}^j,$$

where ε_{t+1}^j are independent random draws from the standard normal so that $corr(z_{t+1}^j, \varepsilon_{t+1}^j) = 0$ and $corr(z_{t+1}^j, w_{t+1}^j) = \rho$. Substituting Equation (19) into Equation (18), we get

$$\tilde{V}_{t+1}^j = V_t^j + \kappa\left(V_t^j\right)^a\left(\theta - V_t^j\right) + \sigma\left(V_t^j\right)^b w_{t+1}^j. \tag{20}$$

This simulates N raw particles and thus provides a set of possible values of V_{t+1}.

3.2.2 Step 2: computing and normalizing the weights. At this point, we have a vector of N possible values of V_{t+1}, and we know according to Equation (17) that given the other available information, V_{t+1} is sufficient to generate $\ln(S_{t+2})$. Therefore, Equation (17) offers a simple way to evaluate the likelihood that the observation S_{t+2} has been generated by V_{t+1}. Hence, we are able to compute the weight given to each particle (or the likelihood or probability that the particle has generated S_{t+2}). The weight is computed as follows:

$$\tilde{W}_{t+1}^j = \frac{1}{\sqrt{2\pi\,\tilde{V}_{t+1}^j}}\exp\left(-\frac{1}{2}\frac{\left(\ln\left(\frac{S_{t+2}}{S_{t+1}}\right) - \left(\mu - \frac{1}{2}\tilde{V}_{t+1}^j\right)\right)^2}{\tilde{V}_{t+1}^j}\right) \tag{21}$$

for $j = 1, .., N$. Finally, we normalize the weights via

$$\breve{W}_{t+1}^j = \frac{\tilde{W}_{t+1}^j}{\sum_{j=1}^N \tilde{W}_{t+1}^j}. \tag{22}$$

3.2.3 Step 3: smooth resampling. The set $\left\{\breve{W}_{t+1}^j\right\}_{j=1}^N$ can be viewed as a discrete probability distribution of $\tilde{V}_{t+1}$ from which we can resample. Let the ordered particles be defined by $\tilde{V}_{t+1}^{(j)}$, then the corresponding cumulated sum of ordered weights, $\breve{W}_{t+1}^{(j)}$, gives the discrete, step-function CDF corresponding

[11] In the returns-based MLIS estimation, $t = 1$ is simply the first day of observed returns. In the options-based NLSIS estimation, $t = 1$ is one year prior to the first available option quote.

to the empirical distribution of V_{t+1}. Pitt (2002) proposes smoothing the $\breve{W}_{t+1}^{(j)}$-based CDF in order to ensure that the subsequent likelihood function is smooth in the model parameters, which renders numerical optimization feasible. The smooth bootstrap in Efron and Tibshirani (1993) offers a possible but computationally expensive solution to this problem. Fortunately, Pitt (2002) provides a very efficient smoothing approach that is crucial for our extremely computationally intensive option valuation application.

Pitt's (2002) approach is simple and intuitive: the discrete, step-function empirical CDF implied by $\breve{W}_{t+1}^{(j)}$ can be converted to a smooth continuous CDF by partitioning the sample space for V_{t+1} into regions delimited by the observed ordered particles. Let region j be defined by $B_j = [\tilde{V}_{t+1}^{(j)}, \tilde{V}_{t+1}^{(j+1)}]$. Now, define the probability of V_{t+1} being in region j as

$$\Pr\left(V_{t+1} \in B_j\right) = \tfrac{1}{2}\left(\breve{W}_{t+1}^{(j)} + \breve{W}_{t+1}^{(j+1)}\right), \quad j = 2, 3, \ldots N-2 \tag{23}$$

and define the conditional density for region j by

$$g\left(V_{t+1} | V_{t+1} \in B_j\right) = \frac{1}{\left(\tilde{V}_{t+1}^{(j+1)} - \tilde{V}_{t+1}^{(j)}\right)}, \quad j = 2, 3, \ldots N-2, \tag{24}$$

where N again is the number of particles.

By appropriately defining the probability distribution in the end regions—that is, for $j = 1$ and $j = N-1$—we ensure that the probabilities sum to one and that the continuous CDF corresponding to Equation (24) passes through a midpoint of each step in the discrete CDF step function. As N becomes large, the continuous resampled CDF will converge to the discrete CDF, which in turn will converge to the true CDF. We refer to Pitt (2002) and Smith and Gelfand (1992) for the details.

Note from Equation (24) that we are assuming a uniform distribution on each region of the sample space. Because the uniform distribution is easy to invert, this is a computationally efficient way to get a set of smooth particles. We resample the particles a total of N times in proportion to their smooth likelihood in Equations (23) and (24) to get the smooth resampled particles, $\left\{V_{t+1}^{j}\right\}_{j=1}^{N}$.

The filtering for period $t+1$ is now done. The filtering for period $t+2$ starts in step 1 above by computing shocks for period $t+2$ in Equation (19) using the smooth resampled particles, V_{t+1}^{j}, as inputs. Repeating steps 1–3 for all t yields a discrete distribution of filtered spot variances on each day. In all our implementations of the PF below, we use five hundred particles.

3.3 Maximum likelihood estimation

Our first empirical strategy uses a long sample of daily S&P500 index returns to estimate each of the SV models by maximizing the likelihood. To this end,

we need an estimation methodology that allows for models with a latent volatility factor.

Pitt (2002) builds on Gordon, Salmond, and Smith (1993) to show that the parameters of latent factor models in general, and of the SV model in Equations (17) and (18) in particular, can be estimated by maximizing the MLIS criterion, which is simply defined by

$$MLIS(\mu, \kappa, \theta, \rho, \sigma) = \sum_{t=1}^{T} \ln\left(\frac{1}{N}\sum_{j=1}^{N} \tilde{W}_t^j\right). \tag{25}$$

As described in the previous section, the particle weights, $\tilde{W}_t^j$, are determined via the conditional likelihood of particle j at time t. These individual likelihoods in turn are given from the model specification in Equations (17) and (18), taking into account that z_t and w_t are correlated normal random variables. The MLIS criterion then simply averages the particle weights across particles, takes logs, and sums over time to create a log likelihood function.

A key challenge in the use of the MLIS for estimation and inference is that it is not generally smooth in the underlying parameters. However, as discussed above, Pitt's (2002) ingenious implementation of the PF, where the resampling in step 3 is done in a smooth fashion, ensures that the MLIS criterion is smooth in the parameters. This smoothing drastically improves the numerical optimization performance, and it enables us to compute reliable parameter standard errors using conventional first-order techniques.

In the Appendix, we provide a small-scale Monte Carlo experiment comparing the MLIS estimator to other estimators available in the literature. We find that the MLIS estimator performs well in comparison with much more computationally intensive methods.

3.4 Nonlinear least squares estimation

Our second empirical strategy is to take a large panel of options traded on the S&P500 index and estimate each of the SV models by minimizing the option pricing errors on this sample. For all the SV models, our implementation uses the NLSIS estimation technique, which minimizes the following mean squared implied volatility error:

$$IVMSE(\mu, \kappa, \theta, \rho, \sigma, \lambda) = \frac{1}{N^T}\sum_{t,i}\left(IV_{i,t} - BS^{-1}\left\{C_i(\bar{V}_t)\right\}\right)^2, \tag{26}$$

where $IV_{i,t}$ is the Black–Scholes implied volatility corresponding to the market price of option i quoted on day t. $C_i\left(\bar{V}_t\right)$ is the model price evaluated at the filtered variance, and $BS^{-1}\left\{C_i(\bar{V}_t)\right\}$ denotes the Black–Scholes inversion of

the model option price.[12] The total number of options in the sample is denoted by $N^T = \sum_{t=1}^{T} N_t$, where T is the total number of days included in the options sample and where N_t is the number of options with various strike prices and maturities included in the sample at date t. The filtered spot variance $\bar{V}_t$ is simply the average of the smooth resampled particles,

$$\bar{V}_t = \frac{1}{N} \sum_{j=1}^{N} V_t^j. \tag{27}$$

Our implementation minimizing Equation (26) is different from existing studies that estimate SV models from option data while filtering volatility using returns. One advantage of our approach is that it is relatively straightforward and fast, therefore allowing us to estimate the models using much more extensive cross-sections of options. In our opinion, estimation using Equation (26) has an additional advantage, because matching the objective function used in parameter estimation with the function subsequently used to evaluate the models ensures the best possible performance of the models in and out of sample. This is motivated by the insights of Granger (1969), who demonstrates that the choice of objective function (also labeled loss function) is an integral part of model specification. It follows that estimating a model using one objective function and evaluating it using another one amounts to a suboptimal choice of objective function. Christoffersen and Jacobs (2004) demonstrate that this issue is empirically relevant for the estimation of the deterministic volatility functions in Dumas, Fleming, and Whaley (1998). We thus choose to implement the SV models in a way that is consistent with these insights. Notice, however, that our use of the particle-filtering algorithm is completely general: we can apply this technique to any volatility model and using any well-behaved loss function involving option prices and underlying returns, including a likelihood function.

Our optimization algorithm minimizes Equation (26) using an iterative procedure on the structural parameters. At each iteration, the volatility is filtered through time using the current parameter values and the information embedded in observed returns. Using the filtered volatility and the structural parameters, option prices are computed, and the IVMSE is calculated. The procedure searches in the structural parameter space until an optimum is reached.

4. Empirical Results Using MLIS Estimation on Returns

This section presents the empirical results obtained when estimating the models on daily index returns. We first discuss the return-based MLIS estimation

[12] Notice that we could alternatively compute the model price as the weighted average of the option prices computed for each particle. However, such an approach would be computationally very costly. Note that if the distribution of particles is centered around the mean, the two approaches will yield very similar results.

results and the models' relative fit. We then discuss the differences in volatility sample paths across models and investigate whether the models can capture the stylized facts in the volatility data. Finally, we investigate the robustness of our conclusions when jumps in returns are added to the stochastic volatility specifications.

4.1 MLIS estimation on index returns

Our first empirical strategy uses a long sample of daily S&P500 index returns and estimates each of the SV models by maximizing the model fit for this sample. First, recall the SV model specifications we consider:

$$dV = \kappa V^a(\theta - V)dt + \sigma V^b dw, \text{ for } a = \{0, 1\} \text{ and } b = \{1/2, 1, 3/2\}.$$

We use daily S&P500 returns from CRSP and estimate the physical parameters by maximizing

$$MLIS(\kappa, \theta, \rho, \sigma) = \sum_{t=1}^{T} \ln\left(\frac{1}{N}\sum_{j=1}^{N} \breve{W}_t^j\right).$$

In the MLIS optimization, the return drift μ is fixed at the sample average daily return in all models. The optimal parameters as well as the MLIS optimum values and the volatility properties for each model are given in table 1. Consider first the MLIS objective values. We report three sets of values. First, we present results for the period 1996–2004, which matches the sample period we will use in the subsequent option estimation analysis. Second, we present results for 1989–2004, which is a longer sample period that does not include the 1987 crash, and finally results for 1985–2004, which includes the crash. Note first that the ranking of models is very stable across the various samples. The ONE ($a = 0, b = 1$) model or the 3/2N ($a = 1, b = 3/2$) model is always best or second best. The SQR ($a = 0, b = 1/2$) model is ranked fourth or fifth, and the SQRN specification ($a = 1, b = 1/2$) is always worst. Although we do not have inference procedures for these MLIS objective values for nonnested models, recall that in standard LR tests, adding one parameter to a model is significant at the 5% level if the log-likelihood increases by approximately two points. The differences in objective values across these models that have the same number of parameters therefore appear to be quite large.

Table 1 also contains the parameter estimates for the 1996–2004 sample period, and the rightmost four columns of table 1 provide the first four sample moments of the filtered volatility. Note that the mean filtered volatility is similar across models. Note also that the κ and θ parameters are not directly comparable between linear ($a = 0$) and nonlinear ($a = 1$) drift specifications. Figure 3, therefore, plots the drift function for all models (solid lines) as a function of the level of variance. The linear drift specifications are given in the left column and the nonlinear drifts in the right column. The square root diffusions

Table 1
Parameter estimates from MLIS on returns

Model			Parameter estimates for 1996–2004				MLIS objective values			Filtered volatility, 1996–2004			
Name	a	b	κ	θ	σ	ρ	1996–2004	1989–2004	1985–2004	Mean	StdDev	Skewness	Kurtosis
SQR	0	1/2	6.5200	0.0352	0.4601	−0.7710	7,064.7	13,359.3	16,595.0	17.698	5.166	0.487	0.086
Standard errors:			1.1096	0.0026	0.0309	0.0375							
SQRN	1	1/2	100.0291	0.0457	0.3425	−0.7527	7,045.1	13,328.9	16,532.2	17.495	5.172	0.043	−0.411
Standard errors:			14.4534	0.0038	0.0193	0.0234							
ONE	0	1	3.9248	0.0408	2.7790	−0.7876	7,074.5	13,372.5	16,631.7	17.673	6.146	1.302	2.120
Standard errors:			1.1392	0.0067	0.1949	0.0345							
ONEN	1	1	133.9347	0.0560	2.4188	−0.7559	7,066.1	13,362.5	16,615.4	17.708	5.569	0.581	0.316
Standard errors:			25.3311	0.0053	0.1654	0.0417							
3/2	0	3/2	1.0852	0.0633	11.9534	−0.7411	7,064.9	13,352.9	16,625.8	17.294	5.863	1.989	6.341
Standard errors:			0.8260	0.0351	0.9125	0.0432							
3/2N	1	3/2	60.1040	0.0837	12.4989	−0.7591	7,068.8	13,362.9	16,638.4	17.683	5.893	1.616	4.659
Standard errors:			23.9651	0.0247	0.8575	0.0426							

The volatility dynamics are modeled as $dV = \kappa V^a(\theta - V)dt + \sigma V^b dw$, where a and b are fixed as indicated in the table. For each model, we estimate the remaining parameters using MLIS on daily returns. The reported parameter estimates use data from January 4, 1996, to December 31, 2004. The return drift parameter μ is fixed at the sample average return of 0.091 for all models. The parameters are reported in annual units. Standard errors are computed using the outer product of the gradient at the optimal parameter values. The MLIS objective value is reported for estimations on longer samples as well. The first four moments of the distribution of the annualized filtered volatility are reported in the last four columns of the table. Kurtosis is reported in excess of three.

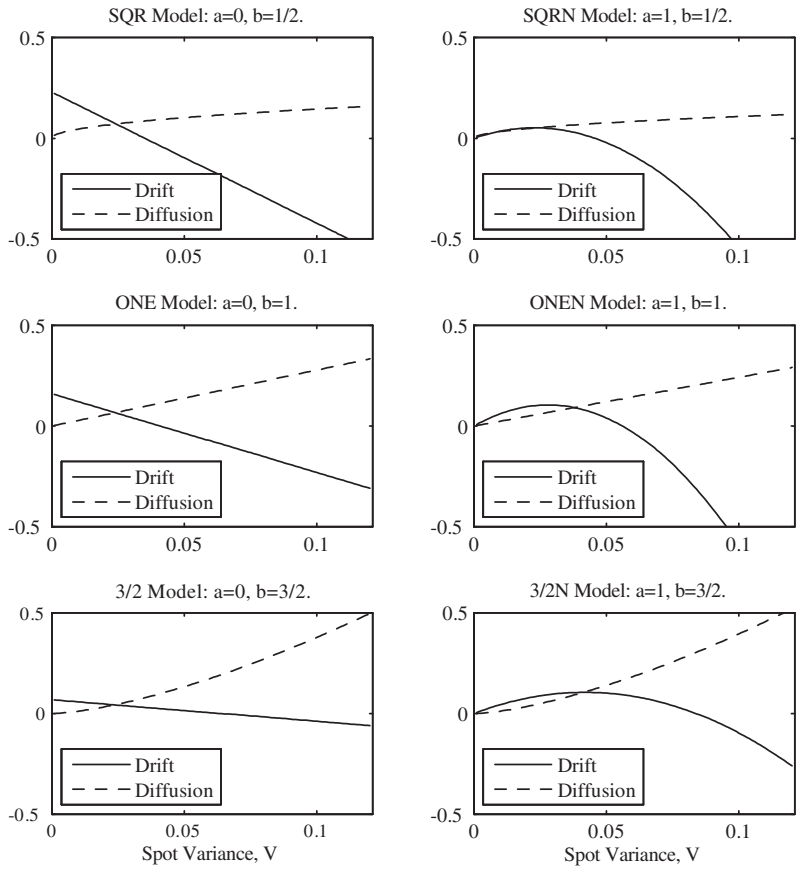

Figure 3
Drift and diffusion functions for various SV models
The solid lines denote the drift function $\kappa V^a(\theta - V)$ plotted against the level of spot variance, V. The dashed lines denote the diffusion function σV^b plotted against the level of spot variance. The parameter estimates are from table 1, using daily returns from 1996 through 2004.

are in the top row, the linear diffusions in the middle row, and the 3/2N diffusions in the bottom row. The differences between linear and nonlinear drift specifications are most evident for large values of the spot variance where the mean reversion in the nonlinear models is much stronger.

The diffusion parameter σ is comparable within a given diffusion specification (i.e., for a given value of b) but not across diffusion specifications. In order to facilitate comparisons of the models, figure 3 shows the diffusion functions (dashed lines) plotted against the value of the spot variance V. Notice that in comparison with the ONE diffusion specification ($b = 1$) in the middle row and the 3/2 specifications ($b = 3/2$) in the bottom row, the SQR and SQRN

specifications ($b = 1/2$) in the top row enable much less diffusion in the variance when the variance level is large.

The final parameter estimate is that of ρ, which captures the correlation between the shocks to return and variance. Ranging from -0.7876 to -0.7411, the estimate of ρ is stable across models and relatively large in magnitude. However, recent studies, including Jones (2003) and Aït-Sahalia and Kimmel (2007), have obtained similarly large correlation estimates.

4.2 Properties of the filtered spot volatilities

We now present some more empirical evidence on the differences in model performance over time. We first plot the path of filtered volatility over time in figure 4. Subsequently, we plot the path of the conditional volatility of variance in figure 5. Finally, in figures 6–8 we assess the distributional characteristics of the filtered spot volatilities for the three models with linear drift and compare them to the properties of the realized volatilities in figure 1.

Figure 4 plots the filtered volatility paths $\sqrt{\bar{V}_t}$ for each model during the 1996 to 2004 sample period. All volatility paths are shown on the same scale, going from 0% to 70% volatility in annual terms. Naturally, the overall pattern in volatility over time is similar across models. However, when volatility increases, it tends to do so much more sharply in the 3/2 models and somewhat more sharply in the linear (ONE) diffusion models when compared with the SQR diffusion models. The ONE and 3/2 diffusions thus exhibit more spikes in volatility when compared with the SQR model. Examples of this include September 1998 (LTCM/Russia default), September 2001 (9/11), and July 2002.

The path of spot variances in figure 4 is of course crucial for the models' option valuation performance over time. However, the time paths of the higher moments are equally important. We can define the model-based conditional variance of variance as

$$Var_t(V_{t+1}) = \sigma^2 \bar{V}_t^{2b}. \tag{28}$$

Figure 5 shows the path of the annualized conditional volatility of variance for each model—that is, the square root of the expression in Equation (28). We again use the same scale for all the plots so as to emphasize the differences across models. Notice how high-volatility episodes such as the September 1998 LTCM debacle and Russia default, as well as the July 2002 stock market decline, led to pronounced differences between the benchmark models—namely, the SQR model in the top-left panel, the ONE model in the middle-left panel, and the 3/2N model in the bottom-right panel.

In figures 6–8, we report on the distributional properties of the levels of the filtered spot volatilities (left column) and the natural logarithms of the filtered variances (right column), for the SQR, ONE, and 3/2 models with linear drifts. As in figures 1 and 2, we report the QQ plots in the top row, the

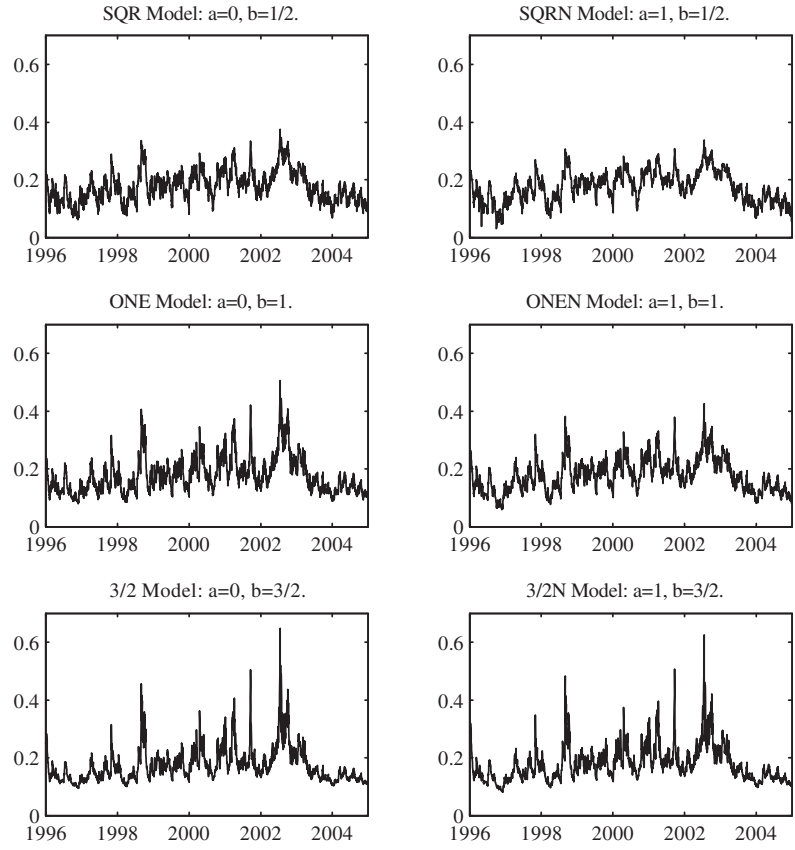

Figure 4
Spot volatility paths for various SV models, 1996–2004
For each SV model of the form $dV = \kappa V^a(\theta - V)dt + \sigma V^b dw$, we plot the annualized daily filtered spot volatility path, $\sqrt{\overline{V}_t}$, during 1996–2004. The parameters are from the MLIS estimation on daily S&P500 returns from 1996 through 2004, as reported in table 1.

daily changes scattered against the volatility (or log variance) in the middle row, and the daily absolute changes scattered against volatility (or log variance) in the bottom row. The differences across the three models are quite striking. The QQ plots reveal that the SQR model in figure 6 has close to normally distributed spot volatilities but nonnormal log spot variances. The ONE model in figure 7 has nonnormal volatilities but approximately normal log variances, and the 3/2 model in figure 8 has strong nonnormality in both the spot volatilities and the log variances. The scatter plots are also quite revealing. Notice that the SQR model in figure 6 has close to homoscedastic volatility changes, but heteroscedastic log variance changes. The ONE model in figure 7 has heteroscedastic volatility changes but close to homoscedastic log variance

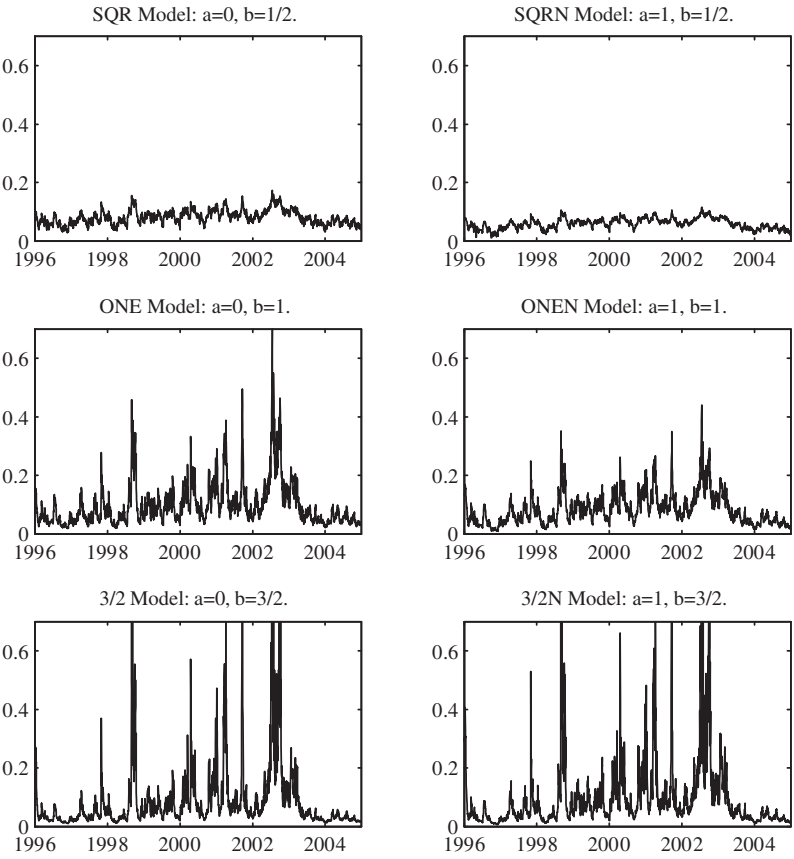

Figure 5
Conditional volatility of variance paths for various SV models, 1996–2004
For each SV model of the form $\mathrm{d}V = \kappa V^a(\theta - V)\mathrm{d}t + \sigma V^b \mathrm{d}w$, we plot the annualized daily conditional volatility of variance path defined as the square root of the conditional variance of the variance of returns, $\sigma \bar{V}_t^b$. The parameters are from the MLIS estimation on daily S&P500 returns from 1996 through 2004, as reported in table 1. The vertical axes have been truncated for the 3/2 diffusions in order to facilitate comparisons with the other models.

changes. The 3/2 model in figure 8 is characterized by heteroscedastic changes in both volatility and log variance.

Properties of the distribution of annualized filtered volatilities can also be inferred from the standard deviation, skewness, and kurtosis of the $\sqrt{\bar{V}_t}$ paths provided in table 1 for all six models. Not only is the standard deviation of volatility different across models but the ONE, 3/2, and 3/2N models are also characterized by much larger skewness and excess kurtosis of volatility compared with the SQR model. This of course confirms the findings in figures 6–8.

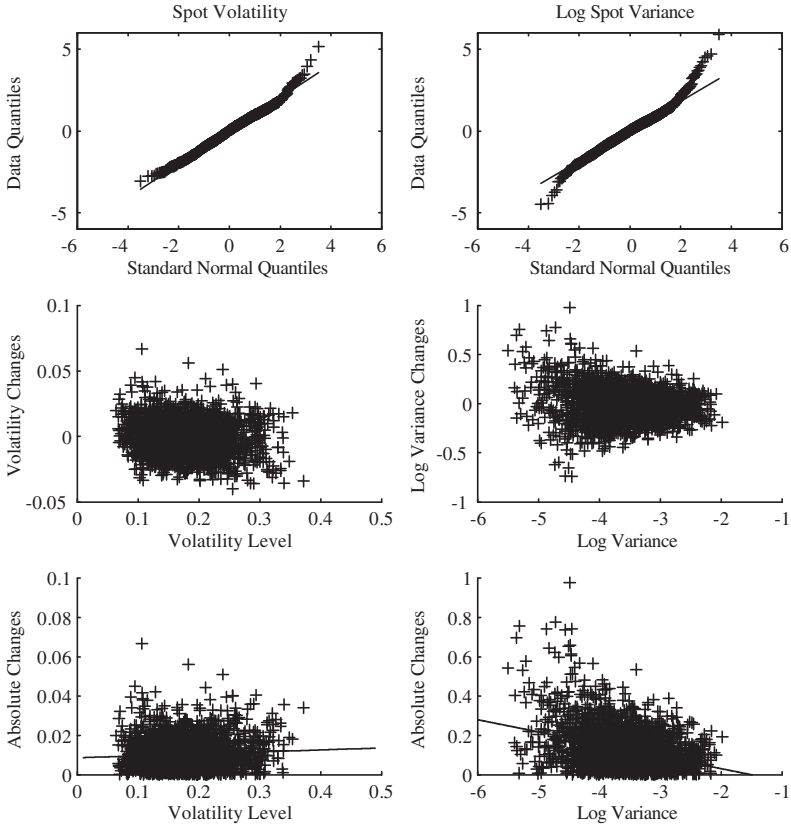

Figure 6
Diagnostics of spot volatility from SQR model, 1996–2004
Using the filtered $\bar{V}_t^{1/2}$ (left column) and $\ln(\bar{V}_t)$ (right column) from 1996 through 2004, the top panels plot the quantiles of the daily changes against quantiles from the normal distribution. The middle panels scatter plot the daily changes against the daily levels. The bottom panels scatter plot the absolute daily changes against the daily levels and show an OLS regression line for reference.

Finally, note that when comparing the model-based spot volatilities and log variances in figures 6–8 to the model-free realized volatilities and log variances in figure 1, the ONE model seems to best capture the distribution of volatility.

4.3 Allowing for jumps in returns

Several papers, including Bates (2000), Pan (2002), Eraker (2004), and Broadie, Chernov, and Johannes (2007), have suggested adding jumps to the basic Heston (1993) SV model. As a robustness check of our results, we therefore estimate the six stochastic volatility models, allowing for Poisson

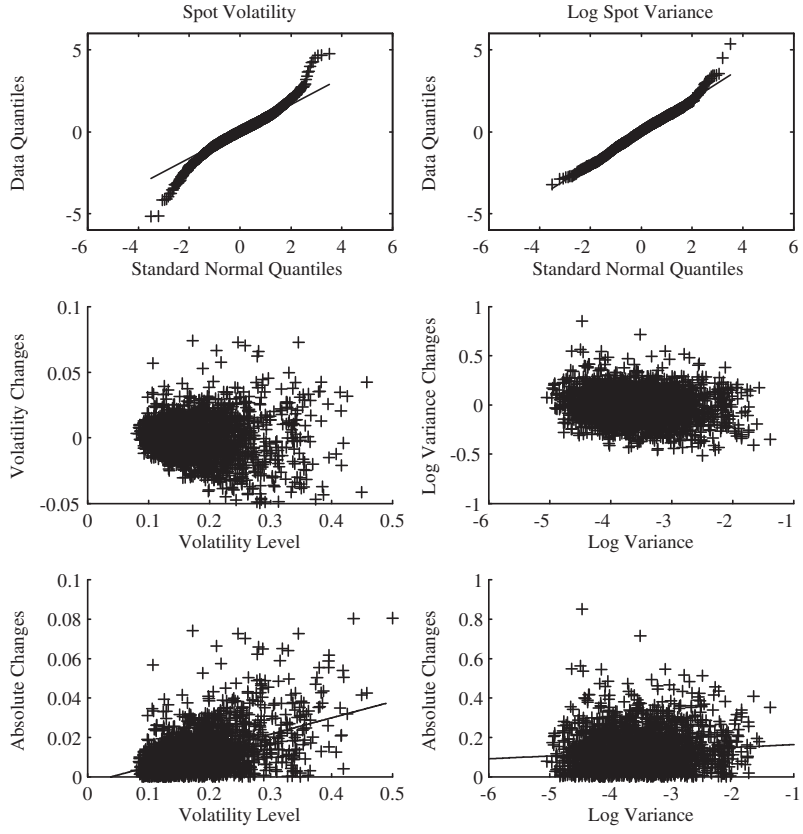

Figure 7
Diagnostics of spot volatility from ONE model, 1996–2004

Using the filtered $\bar{V}_t^{1/2}$ (left column) and $\ln\left(\bar{V}_t\right)$ (right column) from 1996 through 2004, the top panels plot the quantiles of the daily changes against quantiles from the normal distribution. The middle panels scatter plot the daily changes against the daily levels. The bottom panels scatter plot the absolute daily changes against the daily levels and show an OLS regression line for reference.

jumps in returns (SVJ). Working with log returns, we can write the SVJ processes as

$$\mathrm{d}\ln(S) = \left(\mu - \frac{1}{2}V - \gamma_J\overline{\mu}_J\right)\mathrm{d}t + \sqrt{V}\mathrm{d}z + J\mathrm{d}N \tag{29}$$

$$\mathrm{d}V = \kappa V^a(\theta - V)\mathrm{d}t + \sigma V^b\mathrm{d}w. \tag{30}$$

As in most studies, we assume that N is a Poisson jump counter with intensity γ_J so that $\Pr(\mathrm{d}N = 1) = \gamma_J\mathrm{d}t$ and that the jump size J is normally distributed, $J \sim N\left(\mu_J, \sigma_J^2\right)$. The jump compensation term in the return drift takes the form $\gamma_J\overline{\mu}_J$, where $\overline{\mu}_J = \exp\left(\mu_J + \frac{1}{2}\sigma_J^2\right) - 1$.

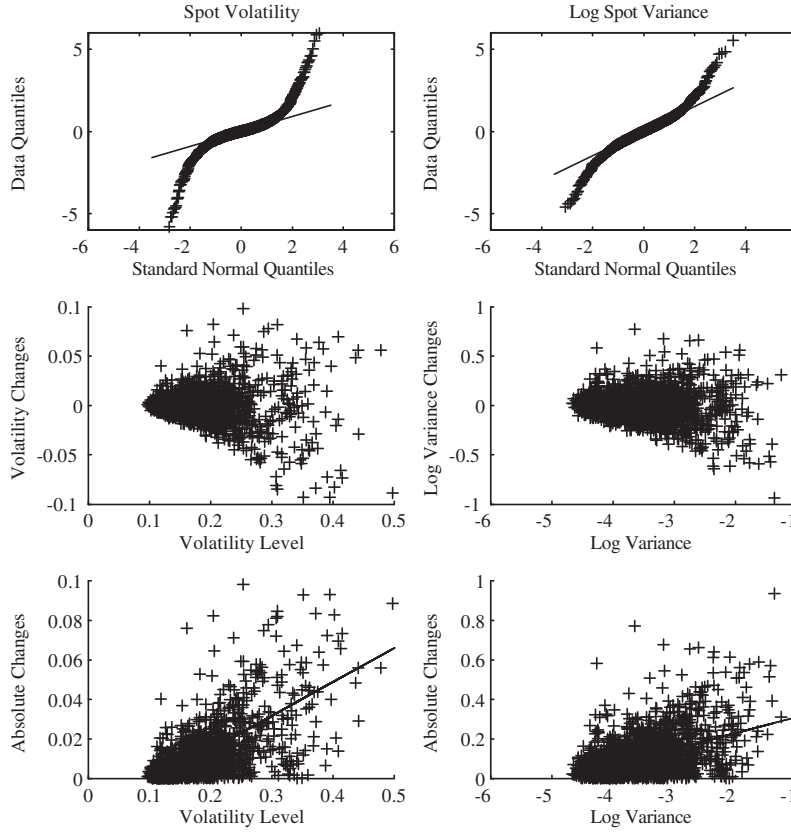

Figure 8
Diagnostics of spot volatility from 3/2 model, 1996–2004
Using the filtered $\bar{v}_t^{1/2}$ (left column) and $\ln(\bar{v}_t)$ (right column) from 1996 through 2004, the top panels plot the quantiles of the daily changes against quantiles from the normal distribution. The middle panels scatter plot the daily changes against the daily levels. The bottom panels scatter plot the absolute daily changes against the daily levels and show an OLS regression line for reference.

The estimation results are reported in table 2. The ranking of models is consistent with table 1: the ONE model provides the best fit to S&P500 index returns, and the two SQR diffusions fit returns the worst. Comparing with the SV estimates in table 1 shows that the likelihood increases by 2.4 for SQR, 3.6 for ONE, and 7.9 for the 3/2 model when including jumps.

As a rule, the stochastic volatility parameter estimates for the SVJ models in table 2 are close to the corresponding parameters for the pure SV models in table 1. The estimates of the mean-reversion parameter κ in the nonlinear drift models provide the only exception to this. The jump parameters in table 2 are annualized so that $\gamma_J = 2.79$ in the SQR model, for example, refers to approximately three jumps per year on average. The mean jump size is -3.24 per

Table 2
Jump model parameter estimates from MLIS on returns

Model			Parameter estimates for 1996–2004							MLIS
Name	a	b	κ	θ	σ	ρ	μ_J	σ_J	γ_J	objective
SQR	0	1/2	6.5889	0.0324	0.4500	−0.7767	−3.2441	0.2115	2.7902	7,067.1
Standard errors:			1.1136	0.0040	0.0382	0.0373	0.1510	0.6429	0.1192	
SQRN	1	1/2	65.0218	0.0362	0.3188	−0.7850	−1.6202	0.0885	2.0594	7,051.0
Standard errors:			19.1974	0.0057	0.0276	0.0398	0.7671	0.2943	0.1044	
ONE	0	1	3.7994	0.0346	2.6097	−0.8150	−0.8301	0.1192	1.9328	7,078.1
Standard errors:			1.0912	0.0079	0.3088	0.0399	0.8925	0.1347	0.2287	
ONEN	1	1	101.8510	0.0500	2.2839	−0.7949	−1.0057	0.1144	1.8999	7,070.1
Standard errors:			26.7143	0.0084	0.2738	0.0390	0.5094	0.1349	0.1411	
3/2	0	3/2	1.1652	0.0391	10.7654	−0.8095	−1.7266	0.1030	2.4126	7,072.8
Standard errors:			0.8332	0.0330	1.4334	0.0740	0.5928	0.1151	0.8252	
3/2N	1	3/2	53.5322	0.0606	11.5979	−0.7799	−2.8341	0.0422	1.0356	7,072.8
Standard errors:			24.2491	0.0196	1.3625	0.0418	0.3997	0.1280	0.1493	

For each model, we estimate the parameters using MLIS on daily returns. The reported parameter estimates use data from January 4, 1996, to December 31, 2004. The return drift parameter μ is fixed at the sample average return of 0.091 for all models. The parameters are reported in annual units. Standard errors are computed using the outer product of the gradient at the optimal parameter values.

year in the SQR model, which corresponds to −1.29% per day. The standard deviation of the jump size, σ_J, is 0.212 per year, corresponding to 1.32% per day. These parameter values are all in line with estimates found in the literature. See, for example, Eraker (2004). Estimates of the jump parameters for the other models are of a similar order of magnitude.

The estimation results for the SVJ model demonstrate that the relative performance of affine versus nonaffine SV models is robust to the inclusion of jumps in returns. Further improvements on the models could potentially be obtained if jumps with time-varying intensity and/or jumps in volatility were considered, as done, for example, in Eraker (2004). However, an exhaustive analysis of all possible jump specifications in combination with our six stochastic volatility specifications is beyond the scope of this article.

5. Empirical Results Using NLSIS Estimation on Options

The properties of realized volatility and VIX in Sections 2.1 and 2.3 suggest that the ONE model matches the distribution of volatility. The MLIS estimation results furthermore suggest that this model also provides a good fit to daily returns. We now turn our attention to an analysis of the models' ability to price index options. We present results using NLSIS estimation on options, while simultaneously using returns to filter the latent volatility. First, we introduce the option data. Second, we present the parameter estimates and overall option fit, and we also analyze the valuation errors in more detail. Third, we check the robustness of our results to the use of different volatility proxies and different volatility filtering techniques. Finally, we augment the SV models, allowing for jumps in returns.

5.1 Option data

We conduct our empirical option-based analysis using OTM S&P500 index option data for the 1996–2004 period. We use only Wednesday and Thursday option data. The decision to pick two days every week is to some extent motivated by computational constraints. The optimization problems in the next section are fairly time intensive, and limiting the number of options reduces the computational burden. Using only Wednesday data in estimation—and Thursdays for out-of-sample analysis—allows us to study a long time series, which is useful considering the highly persistent volatility processes. An additional motivation for using only Wednesday data is that following the work of Dumas, Fleming, and Whaley (1998), several studies, including Heston and Nandi (2000), have used this setup. On each Wednesday and Thursday, we remove the least liquid contracts, keeping only the seven highest-volume strike prices at each maturity.

Panel A of table 3 presents descriptive statistics for the Wednesday closing OTM call and put option data for 1996–2004 sorted by moneyness (defined

Table 3
S&P500 index option data, 1996–2004

Panel A. Option data characteristics by moneyness and maturity: Wednesday OTM data

By moneyness	$F/X < 0.96$	$0.96 < F/X < 0.98$	$0.98 < F/X < 1.02$	$1.02 < F/X < 1.04$	$1.04 < F/X < 1.06$	$F/X > 1.06$	All
Number of contracts	1,985	1,617	3,034	2,925	1,916	3,351	14,828
Average price	29.26	34.89	38.06	31.93	26.96	22.40	30.35
Average implied volatility	19.45	19.42	19.59	20.58	22.45	25.52	21.46
Average bid–ask spread	1.519	1.534	1.625	1.535	1.401	1.344	1.491
By maturity	DTM<30	30<DTM<60	60<DTM<90	90<DTM<180	180<DTM<270	DTM>270	All
Number of contracts	695	3,476	2,551	3,416	2,885	1,805	14,828
Average price	12.11	17.97	24.47	31.30	41.25	50.35	30.35
Average implied volatility	20.69	21.04	21.62	21.82	21.63	21.37	21.46
Average bid–ask spread	0.830	1.184	1.452	1.580	1.714	1.864	1.491

Panel B. Option data characteristics by moneyness and maturity: Thursday OTM data

By moneyness	$F/X < 0.96$	$0.96 < F/X < 0.98$	$0.98 < F/X < 1.02$	$1.02 < F/X < 1.04$	$1.04 < F/X < 1.06$	$F/X > 1.06$	All
Number of contracts	2,283	1,162	3,902	1,713	1,295	4,220	14,575
Average price	24.84	30.65	37.03	33.11	30.70	25.98	30.39
Average implied volatility	19.38	18.99	19.60	21.08	22.19	24.93	21.47
Average bid–ask spread	1.449	1.504	1.640	1.484	1.431	1.395	1.492
By maturity	DTM<30	30<DTM<60	60<DTM<90	90<DTM<180	180<DTM<270	DTM>270	All
Number of contracts	1,276	2,710	2,518	3,373	2,829	1,869	14,575
Average price	12.94	18.73	24.09	31.44	41.03	49.70	30.39
Average implied volatility	20.87	21.18	21.59	21.79	21.64	21.28	21.47
Average bid–ask spread	0.876	1.227	1.437	1.579	1.720	1.867	1.492

We use Wednesday and Thursday closing out of the money (OTM) call and put option data from OptionMetrics from January 1, 1996, through December 31, 2004. F denotes the implied futures price of the S&P500 index, X denotes the strike price, and DTM denotes the number of calendar days to maturity. The average bid–ask spread is reported in dollars.

as the implied futures price over strike price, F/X, and maturity. When F/X is below 1, the contract is an OTM call, and when F/X is above 1, the contract is an OTM put. The sample includes a total of 14,828 Wednesday option contracts with an average midprice of \$30.35 and average implied volatility of 21.46%. The implied volatility is largest for the OTM put options where F/X is large, reflecting the well-known volatility smirk in index options. The average implied volatility term structure is roughly flat during the sample period. The average bid–ask spread is \$1.491.

Panel B of table 3 provides the option data characteristics for OTM options recorded at the close of each Thursday. We observe a total of 14,575 option contracts with an average midprice of \$30.39 and an average implied volatility of 21.47%. The average bid–ask spread is \$1.492. The implied volatility smirk is clearly evident in the Thursday options as well.

5.2 NLSIS estimation on index options

Table 4 contains the parameter values obtained from minimizing the option implied volatility mean squared error, defined above as

$$IVMSE\,(\kappa, \theta, \rho, \sigma, \lambda) = \frac{1}{N^T} \sum_{t,i} \left(IV_{i,t} - BS^{-1} \left\{ C_i(\bar{V}_t) \right\} \right)^2. \qquad (31)$$

In the NLSIS optimization, the return drift, μ, is fixed at the sample average daily return in all models, as was the case in MLIS estimation.

Note that we are using daily returns to create the volatility path, $\bar{V}_t$, whereas option values are computed each Wednesday. Ideally, we would minimize IVMSE using option data from all the days of the week, but computationally, this is infeasible for a nine-year sample. Letting the data set span a long time period is desirable because of the well-known slow mean reversion in market volatility. Using daily returns for volatility filtering and weekly data for options provides a useful compromise between the length of the data sample and computational feasibility.

Consider first the in-sample implied volatility root mean squared error (IVRMSE) column in table 4, which uses Wednesday OTM options to compare the models. The 3/2 model performs the worst with an IVRMSE of 3.450%, the benchmark SQR model has an IVRMSE of 3.216%, and the ONE model performs the best with an IVRMSE of 2.900%. The root mean squared error (RMSE) thus differs by about 9.8% between the benchmark SQR and the ONE model. Note that this improvement in fit is obtained without adding any parameters to the model. The ONE model significantly outperforms the SQR model, and the 3/2 model is significantly worse than the SQR model when judged by the Diebold and Mariano (1995) (DM) test. The other models are not significantly different from the SQR model in this exercise.

Consider next the out-of-sample column in table 4, which uses Thursday OTM options. We use the same calendar period for the out-of-sample study so

Table 4
NLSIS estimates and IVRMSE, 1996–2004

Model			NLSIS estimation on option implied volatilities					Model	IVRMSE		IVRMSE, DTM>30	
Name	a	b	κ	θ	σ	λ	ρ	name	In-sample Wednesday	Out-of-sample Thursday	In-sample Wednesday	Out-of-sample Thursday
SQR	0	1/2	2.8791	0.0631	0.5368	8.69E−05	−0.7042	SQR	3.216	3.247	3.212	3.255
Standard errors			2.47E−03	4.68E−05	2.31E−03	1.88E−04	1.93E−04	Ratio	1.000	1.000	1.000	1.000
SQRN	1	1/2	90.3775	0.0786	0.4873	6.51E−04	−0.6228	SQRN	3.355	3.422	3.329	3.390
Standard errors			5.54E−02	1.05E−05	1.09E−04	3.04E−03	4.23E−04	Ratio	1.043	1.054	1.036	1.041
								DM test	1.324	1.651	1.097	1.379
ONE	0	1	1.7869	0.0648	1.5040	5.65E−03	−0.7581	ONE	2.900	2.913	2.894	2.887
Standard errors			1.44E−03	7.44E−06	4.91E−04	1.04E−05	1.93E−04	Ratio	0.902	0.897	0.901	0.887
								DM test	**−3.023**	**−2.789**	**−2.784**	**−2.774**
ONEN	1	1	68.8798	0.0917	2.3009	4.98E−02	−0.7181	ONEN	3.194	3.230	2.926	2.927
Standard errors			5.72E−02	6.55E−05	7.96E−04	1.20E−01	3.27E−04	Ratio	0.993	0.995	0.911	0.899
								DM test	−0.367	−0.171	**−2.451**	**−2.443**
3/2	0	3/2	1.7504	0.0588	8.2703	1.36E+00	−0.6634	3/2	3.450	3.495	3.363	3.401
Standard errors			5.13E−03	4.37E−04	5.93E−03	3.83E−03	2.96E−04	Ratio	1.073	1.076	1.047	1.045
								DM test	**3.544**	**3.557**	**2.093**	1.904
3/2N	1	3/2	32.8837	0.1147	7.9115	3.99E−04	−0.7321	3/2N	3.193	3.220	3.107	3.136
Standard errors			7.85E−02	1.56E−04	4.10E−03	3.66E−02	2.43E−04	Ratio	0.993	0.992	0.967	0.963
								DM test	−0.372	−0.378	−1.359	−1.436

Model parameters are estimated using NLSIS on the 14,828 Wednesday closing OTM option quotes observed from January 4, 1996, to December 31, 2004. Standard errors are computed using the outer product of the gradient at the optimal parameter values. Out-of-sample refers to the 14,575 Thursday closing OTM option prices observed during the same period. The Black–Scholes benchmark yields an RMSE of 5.28% for the Wednesday OTM sample and 5.42% for the Thursday OTM sample. The *ad hoc* Black–Scholes benchmark uses the average IV for the same weekday of the previous week and has an RMSE of 3.18% for the Wednesday sample and 3.24% for the Thursday sample. The Diebold–Mariano test is computed on the weekly mean squared errors. Bold type indicates significance at the 5% level. The last two columns use estimates from Wednesday data with more than thirty days to maturity.

as to avoid the impact of structural breaks in volatility. Because we are comparing equally parsimonious models estimated on a large sample, the Bates (2003) critique mentioned in Section 2, that more heavily parameterized models are favored in such out-of-sample experiments, does not apply. The ONE model outperforms the SQR model by 10.3%, and the ONE model is again the only model that significantly outperforms the SQR model according to the Diebold–Mariano test.

The overall improvement in IVRMSE provided by the ONE model over the benchmark SQR model is significant and roughly similar in and out of sample. It is also consistent with the evidence from returns, realized variances, and the VIX discussed earlier.

When comparing the option-based NLSIS estimates in table 4 with the return-based estimates in table 1, we find that the κ is smaller in table 4 for most cases, indicating a slower mean reversion of variance when options are driving the parameters. The θ parameter is larger for most models in table 4 compared with table 1. The volatility of variance parameter σ is generally lower in table 4, the exception being the two square root models. The volatility risk premium parameter λ is generally small but positive for all models as expected. Finally, the correlation coefficient, ρ, is slightly smaller (in absolute value) in all models when estimated using both options and returns.

The footnote to table 4 reports the IVRMSE from two benchmark models. First, the simple Black–Scholes model in which volatility is kept constant at the average implied volatility across all options and across the entire sample period. The resulting IVRMSE is 5.28% for the Wednesday sample and 5.42% for the Thursday sample. Note that, not surprisingly, all the SV models in table 4 outperform the simple Black–Scholes model by a wide margin both in and out of sample. The second benchmark is the so-called *ad hoc* Black–Scholes model in which the implied volatility for each option is set to the average implied volatility across options on the same weekday of the previous week. This *ad hoc* model has an IVRMSE of 3.18% for the Wednesday sample and 3.24% for the Thursday sample. Thus, only the ONE model outperforms this benchmark significantly in both samples. This result may be surprising, but note that the *ad hoc* model allows for the estimation of 9*52 = 468 parameters versus only five parameters in the SV models. These results are not easily comparable with other recent studies, because the weekly estimated *ad hoc* benchmark is not often compared with SV models estimated on a long sample. Heston and Nandi (2000) provide a similar benchmark, but their GARCH models are estimated on much shorter samples.

5.3 Decomposing the option valuation errors

To provide more insight into the models' performance, we decompose the in-sample option valuation errors along several dimensions. The out-of-sample results yield very similar conclusions regarding the models' relative performance and are thus omitted from the tables and figures. Option valuation mod-

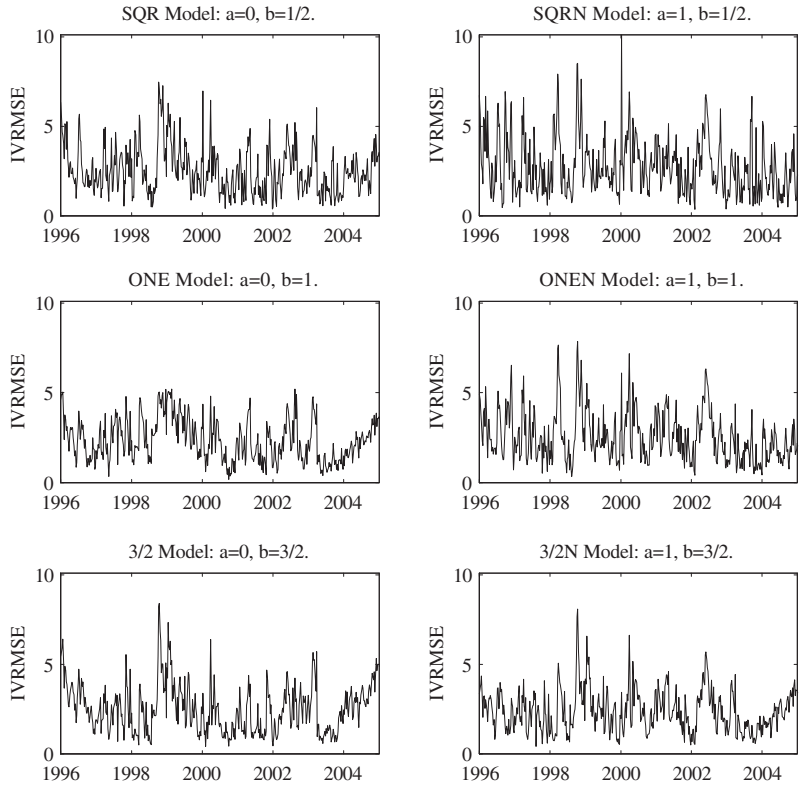

Figure 9
Weekly IVRMSE for at-the-money options
For each model, we plot the weekly IVRMSE for at-the-money options, defined as options with moneyness, F/X, between 0.98 and 1.02. The parameter estimates are from table 4.

els can fail for several reasons: they may imply a biased estimate of the latent volatility variable, which leads to biases in the valuation of at-the-money options. Alternatively, one volatility model may outperform another in the maturity dimension if it provides better estimates of the term structure of volatility, or in the moneyness dimension if it provides better estimates of the implied volatility smile. We investigate all these model aspects, and we also investigate the models' performance as a function of the level of market volatility, as measured by the VIX.

Figures 9 and 10 plot the performance of the models in valuing at-the-money options over time. Figure 9 plots the weekly IVRMSE for at-the-money options, defined as options with moneyness, F/X, between 0.98 and 1.02. The ONE model displays a relatively stable IVRMSE across time. The IVRMSE plots for the other models exhibit many more spikes.

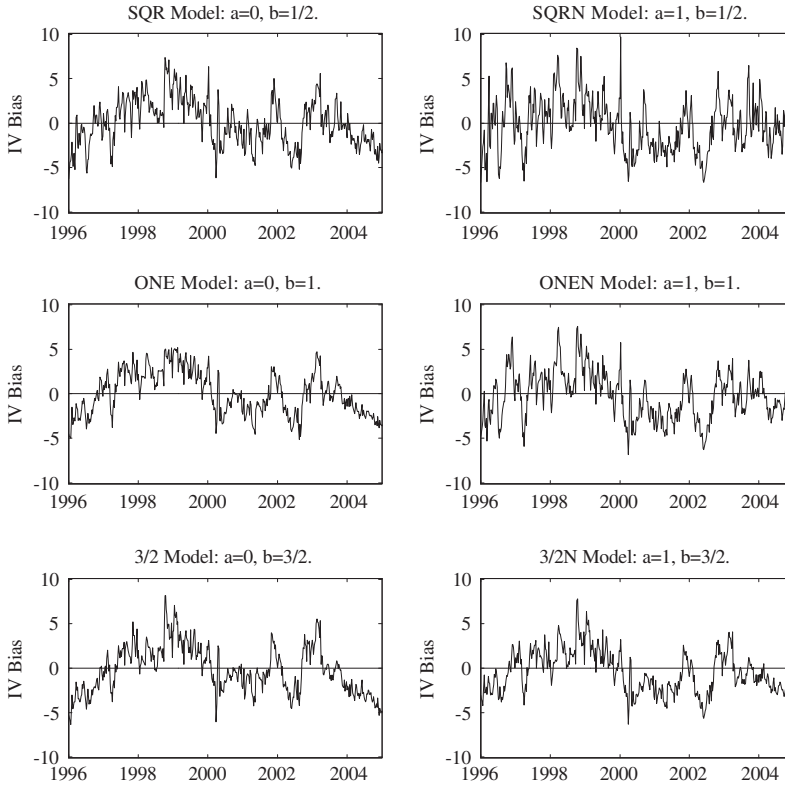

Figure 10
Weekly IV bias for at-the-money options
For each model, we plot the weekly IV bias for at-the-money options, defined as options with moneyness, F/X, between 0.98 and 1.02. The parameter estimates are from table 4.

The MSE-based objective function can be viewed as a sum of error variance and bias squared. Figure 10 therefore shows the weekly IV bias (average data IV less average model IV) for the at-the-money options. It is clear that some of the spikes in figure 9 are driven at least partly by spikes in the bias. This is particularly the case for the models with nonlinear drift. In addition to the high-frequency movements evident in figure 10, it also appears that all models display a somewhat persistent bias. This would indicate that volatility models with more flexible autocorrelation functions, as provided, for example, by the multiple volatility component models in Bates (2000), are needed.

In table 5, we report the in-sample IVRMSE by moneyness, maturity, and volatility level as measured by the VIX. Figure 11 depicts the IVRMSE results by moneyness, maturity, and volatility categories, for the three models with linear variance drift, $a = 0$, in the left column, and the three models with nonlinear drift, $a = 1$, in the right column. The impressive overall performance of the

Table 5
In-sample IVRMSE (%) by moneyness, maturity, and VIX level, 1996–2004

Model			Panel A. IVRMSE by moneyness						
Name	a	b	$F/X < 0.96$	$0.96 < F/X < 0.98$	$0.98 < F/X < 1.02$	$1.02 < F/X < 1.04$	$1.04 < F/X < 1.06$	$F/X > 1.06$	All
SQR	0	1/2	2.850	2.872	2.835	3.024	3.210	3.835	3.216
SQRN	1	1/2	3.012	3.286	3.149	3.104	3.198	3.837	3.355
ONE	0	1	2.636	2.590	2.624	2.790	2.964	3.346	2.900
ONEN	1	1	2.884	2.932	2.861	3.019	3.171	3.738	3.194
3/2	0	3/2	3.121	2.948	2.909	3.148	3.376	4.252	3.450
3/2N	1	3/2	3.012	2.807	2.655	2.802	3.032	3.947	3.193
Average			2.919	2.906	2.839	2.981	3.159	3.826	3.218

Model			Panel B. IVRMSE by maturity						
Name	a	b	DTM<30	30<DTM<60	60<DTM<90	90<DTM<180	180<DTM<270	DTM>270	All
SQR	0	1/2	3.646	3.336	3.204	3.086	3.094	3.253	3.216
SQRN	1	1/2	3.814	3.607	3.442	3.317	3.103	2.971	3.355
ONE	0	1	2.850	2.950	2.870	2.934	2.863	2.858	2.900
ONEN	1	1	3.575	3.406	3.230	3.134	2.998	2.971	3.194
3/2	0	3/2	3.533	3.436	3.293	3.484	3.458	3.583	3.450
3/2N	1	3/2	3.405	3.241	3.119	3.194	3.128	3.217	3.193
Average			3.470	3.329	3.193	3.191	3.107	3.142	3.218

Model			Panel C. IVRMSE by VIX level						
Name	a	b	VIX<15	15<VIX<20	20<VIX<25	25<VIX<30	30<VIX<35	VIX>35	All
SQR	0	1/2	3.592	2.800	2.793	3.702	4.324	4.279	3.216
SQRN	1	1/2	2.699	3.165	3.159	3.460	3.864	5.345	3.355
ONE	0	1	3.204	2.259	2.837	3.383	3.593	3.405	2.900
ONEN	1	1	2.560	2.904	3.049	3.462	3.838	4.617	3.194
3/2	0	3/2	4.423	2.733	2.869	4.083	4.860	4.910	3.450
3/2N	1	3/2	3.205	2.530	2.980	3.681	4.195	4.690	3.193
Average			3.281	2.732	2.948	3.628	4.112	4.541	3.218

We use the NLSIS estimates from table 4 to compute the option IVRMSE in percent for various moneyness, maturity, and volatility (VIX level) bins for each model. The contracts used in the table are for the 1996–2004 in-sample period, which consists of Wednesday closing OTM option quotes.

ONE model (denoted by "x" in the left column) clearly originates from good performance in virtually all moneyness, maturity, and volatility categories. The 3/2 model (denoted by "+" in the left column) and the SQRN model (denoted by "o" in the right column), however, generally perform worst or near worst in all moneyness, maturity, and volatility categories. Table 5 shows that the ONE model has the lowest IVRMSE in sixteen out of eighteen categories, and either the 3/2 or SQRN models have the highest IVRMSE in all eighteen categories.

The IVRMSE is generally increasing in moneyness for most models, which may be partly driven by the fact that the average implied volatility is increasing in moneyness (see table 3). Table 5 indicates that the ONE model performs best in all six moneyness categories.

Note also that IVRMSE tends to decrease with maturity for all models. This is only partially explained by the average implied volatility being larger for short-term options, because these differences are small (see table 3). Figure 11 shows clearly that the ONE model outperforms the other five models in all six maturity categories.

For the analysis by VIX level in panel C of table 5 and the bottom panels in figure 11, we split the sample days into six categories sorted by the VIX level on each day of the sample. The ONEN model performs the best at lowest volatility levels, whereas the ONE model performs the best for the three highest volatility levels. The SQR model performs best when VIX is between 20 and 25, but it is otherwise worse than average. Beyond the lowest level of volatility, all models tend to produce increasing IVRMSE when the VIX is increasing. This finding confirms that a more flexible volatility specification may be needed. Our findings strongly suggest that the search for such a specification should start from the nonaffine ONE model rather than the affine SQR model.

In table 6, we report on the relative pricing error of each model across moneyness, maturity, and VIX levels. We compute the square root of the average squared relative errors defined as

$$RRMSE = \sqrt{\frac{1}{N^T}\sum_{t,i}\left(\left(C_{i,t} - C_i(\bar{V}_t)\right)/C_{i,t}\right)^2}, \tag{32}$$

where $C_{i,t}$ denotes the market price of option i at time t. While the RRMSE has a nice, intuitive interpretation, it is not often used in the literature, because it implicitly puts a lot of weight on options with low prices, which tend to be relatively illiquid, short-term, deep OTM contracts for which the data are likely to be noisy. See Bates (2003) for a discussion. In spite of these caveats, table 6 shows that the ranking of models from table 5 is largely unchanged. The ONE model performs the best overall, and the 3/2 and SQRN models perform the worst. The ONE model performs the best in twelve out of eighteen categories, and either the 3/2 or SQRN models are worst in all categories. As in table 5, the RRMSE across models tends to be highest for deep OTM and for

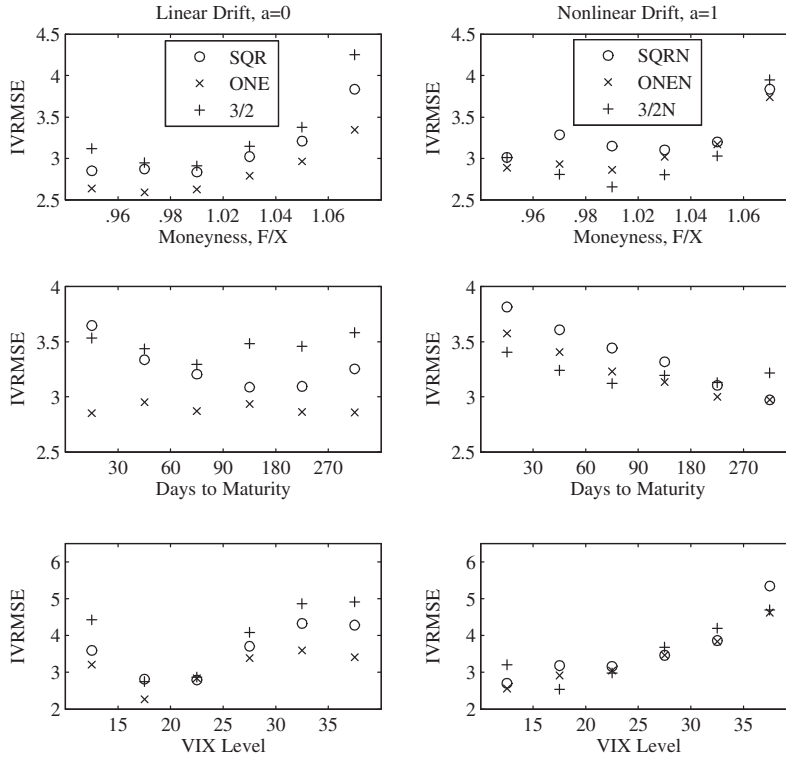

Figure 11
IVRMSE by moneyness, maturity, and VIX level
For each model, we plot the in-sample IVRMSE by moneyness (top row), maturity (middle row), and volatility level as measured by the VIX (bottom row). The left column shows the models with linear drift, and the right column shows the models with nonlinear drift in spot variance. The parameter estimates are from table 4.

short-maturity options. The RRMSE is also high during periods when VIX is low, suggesting that models with richer volatility dynamics, such as multiple-component volatility models, may improve performance.

5.4 Alternative data, filtering techniques, and volatility paths

The NLSIS estimation on option data leads to a number of interesting conclusions. Perhaps most important, the ONE model outperforms the SQR model in virtually all dimensions. These results confirm our earlier findings obtained using realized volatility diagnostics and returns-based MLIS estimation.

We now further investigate the robustness of these findings with respect to the choice of filtering technique and the data used in filtering. Although the PF used in the MLIS and NLSIS estimation is computationally efficient and reliable, it is of interest to see if similar results obtain with alternative filtering

Table 6
Relative pricing error (RRMSE) by moneyness, maturity, and VIX level, 1996–2004

Model			Panel A. RRMSE by moneyness						
Name	a	b	$F/X < 0.96$	$0.96 < F/X < 0.98$	$0.98 < F/X < 1.02$	$1.02 < F/X < 1.04$	$1.04 < F/X < 1.06$	$F/X > 1.06$	All
SQR	0	1/2	0.4152	0.2745	0.1843	0.2275	0.2684	0.3405	0.2961
SQRN	1	1/2	0.4595	0.3276	0.2110	0.2406	0.2712	0.3188	0.3093
ONE	0	1	0.4116	0.2843	0.1736	0.1987	0.2308	0.2949	0.2740
ONEN	1	1	0.4202	0.2827	0.1838	0.2268	0.2629	0.3183	0.2900
3/2	0	3/2	0.5415	0.3431	0.2031	0.2299	0.2659	0.3488	0.3363
3/2N	1	3/2	0.4886	0.3001	0.1743	0.1960	0.2315	0.3182	0.3008
Average			0.4561	0.3021	0.1883	0.2199	0.2551	0.3233	0.3011
Model			Panel B. RRMSE by maturity						
Name	a	b	DTM<30	30<DTM<60	60<DTM<90	90<DTM<180	180<DTM<270	DTM>270	All
SQR	0	1/2	0.3070	0.2903	0.2595	0.2878	0.3106	0.3397	0.2961
SQRN	1	1/2	0.4083	0.3366	0.3108	0.2994	0.2767	0.2741	0.3093
ONE	0	1	0.3382	0.2790	0.2568	0.2650	0.2644	0.2916	0.2740
ONEN	1	1	0.3603	0.3085	0.2803	0.2757	0.2750	0.2858	0.2900
3/2	0	3/2	0.3721	0.3105	0.2970	0.3251	0.3554	0.4042	0.3363
3/2N	1	3/2	0.3746	0.2975	0.2756	0.2843	0.3001	0.3385	0.3008
Average			0.3601	0.3037	0.2800	0.2896	0.2970	0.3223	0.3011
Model			Panel C. RRMSE by VIX level						
Name	a	b	VIX<15	15<VIX<20	20<VIX<25	25<VIX<30	30<VIX<35	VIX>35	All
SQR	0	1/2	0.5502	0.3261	0.2379	0.2532	0.2599	0.1959	0.2961
SQRN	1	1/2	0.4284	0.3522	0.2846	0.2740	0.2346	0.2335	0.3093
ONE	0	1	0.5171	0.2831	0.2410	0.2309	0.2269	0.1639	0.2740
ONEN	1	1	0.3931	0.3270	0.2698	0.2588	0.2334	0.2051	0.2900
3/2	0	3/2	0.7230	0.3583	0.2526	0.2691	0.2803	0.2167	0.3363
3/2N	1	3/2	0.5244	0.3132	0.2707	0.2633	0.2515	0.2082	0.3008
Average			0.5227	0.3267	0.2594	0.2582	0.2478	0.2039	0.3011

We report the relative option price root mean squared error (RRMSE) for various moneyness, maturity, and volatility (VIX level) bins for each model. The contracts used in the table are for the 1996–2004 in-sample period, which consists of Wednesday closing OTM option quotes.

The Review of Financial Studies v 23 n 8 2010

techniques. Moreover, while our results are largely consistent across the MLIS and NLSIS filtering exercises that use daily returns, it is of interest to investigate if similar results obtain when filtering the spot volatility from alternative data.

Table 7 reports on these robustness exercises. We consider two alternative data sources for filtering: the daily realized variance data for 1996–2004 considered in Section 2.1 and the daily VIX data from Section 2.3 for the same period. In panel A of table 7, we first value the Wednesday options using the raw realized variance and (squared) VIX data as proxies for the spot variance. Subsequently, in panel B, we filter the latent variance from these two series using the Kalman filter. We use the simplest possible setup to filter from the VIX and RV series, ignoring returns altogether. We limit our analysis to the SQR, ONE, and 3/2 models, because the measurement equations relating realized volatility and VIX to the latent volatility are not known for models with nonlinear drift. Valuation results are obtained using the NLSIS-based parameter estimates from table 4.

When filtering latent volatility from realized volatility, a sensible state-space representation for the SQR, ONE, and 3/2 models is

$$\begin{aligned} RV_{t,t+1} &= E[RV_{t,t+1}|V_t] + u_t \\ V_{t+1} &= V_t + \kappa(\theta - V_t) + \sigma V_t^b w_{t+1}. \end{aligned}$$

Using the results in Aït-Sahalia and Kimmel (2007), this can be written as

$$\begin{aligned} RV_{t,t+1} &= \theta + \left(\frac{\exp(-\kappa/252) - 1}{(-\kappa/252)}\right)(V_t - \theta) + u_t \\ V_{t+1} &= V_t + \kappa(\theta - V_t) + \sigma V_t^b w_{t+1}. \end{aligned} \tag{33}$$

For the VIX, we employ the following state-space representation:

$$\begin{aligned} VIX^2_{t,t+30} &= \theta + \left(\frac{\exp(-30\kappa/252) - 1}{(-30\kappa/252)}\right)(V_t - \theta) + u_t \\ V_{t+1} &= V_t + \kappa(\theta - V_t) + \sigma V_t^b w_{t+1}. \end{aligned} \tag{34}$$

Panels A and B of table 7 present the IVRMSEs, as well as the IVRMSE ratio between the ONE and 3/2 models and the SQR model. We also report the DM test for the ONE and 3/2 models against the SQR benchmark. The main conclusion is that the results confirm that the SQR model can be substantially improved on by the ONE model.

In general, the IVRMSEs using the raw RV data in panel A are large compared with the NLSIS-based IVRMSEs in table 4. This is not surprising, because these raw data contain more outliers than the variance data filtered from returns. Panel B shows that when filtering the RV using the Kalman filter setup in Equation (33), the IVRMSEs are somewhat lower than when using raw RV data.

Table 7
IVRMSEs using VIX, RV, and filtered returns, 1996–2004

Panel A: Using raw RV and VIX as spot volatility paths

Model			Wednesday OTM				Thursday OTM			
Name	*a*	*b*	RV-based	Ratio	VIX-based	Ratio	RV-based	Ratio	VIX-based	Ratio
SQR	0	1/2	10.847	1.000	9.171	1.000	10.030	1.000	9.168	1.000
ONE	0	1	7.112	0.656	4.032	0.440	6.142	0.612	4.026	0.439
DM test			**–14.574**		**–7.462**		**–16.350**		**–10.559**	
3/2	0	3/2	6.737	0.621	5.873	0.640	6.669	0.665	5.940	0.648
DM test			**–5.261**		**–5.886**		**–6.160**		**–7.806**	

Panel B: Using Kalman-filtered RV and VIX as spot volatility paths

Model			Wednesday OTM				Thursday OTM			
Name	*a*	*b*	RV-based	Ratio	VIX-based	Ratio	RV-based	Ratio	VIX-based	Ratio
SQR	0	1/2	9.841	1.000	8.184	1.000	8.752	1.000	7.969	1.000
ONE	0	1	7.084	0.720	4.018	0.491	6.143	0.702	4.012	0.503
DM test			**–13.967**		**–6.598**		**–14.151**		**–8.786**	
3/2	0	3/2	6.731	0.684	5.859	0.716	6.655	0.760	5.926	0.744
DM test			**–4.171**		**–4.881**		**–4.125**		**–5.748**	

Panel C: Using alternative model-based spot volatility paths

Model			Wednesday OTM: Return Data				Thursday OTM: Return Data			
Name	*a*	*b*	SQR-V	ONE-V	3/2-V	Average	SQR-V	ONE-V	3/2-V	Average
SQR	0	1/2	3.216	3.172	3.369	3.252	3.247	3.247	3.408	3.300
ONE	0	1	3.116	2.900	3.010	3.008	3.075	2.913	2.984	2.991
3/2	0	3/2	3.414	3.329	3.450	3.398	3.587	3.402	3.495	3.495
Average			3.248	3.134	3.276		3.303	3.187	3.295	

We use the NLSIS-based parameters for the three models with linear volatility drifts from table 4 to compute option prices and IVRMSEs for the two option data sets using different volatility paths. RV-based refers to using daily realized volatilities to obtain the volatility state variable used in pricing, and VIX-based refers to using the daily VIX. Panel A uses the raw RV and VIX data, and panel B uses RV and VIX filtered using the Kalman filter. The Diebold–Mariano test is computed on the weekly mean squared errors. Bold type indicates significance at the 5% level. Panel C uses each of the three linear-drift volatility paths estimated in table 4 in each of the three models. Gray shading indicates that the model's own volatility path is used.

The VIX data yield a lower IVRMSE than the realized volatility data. This is also not necessarily surprising, because this implementation gives the models crucial additional information on the current state of volatility as implied by the one-month options underlying VIX.[13] It is, however, quite striking that the ONE model performs so robustly well.

In panel C of table 7 we report on a complementary filtering exercise. For each of the three models with linear drift, we compute the option IVRMSE first using each model's own spot variance path and then using the variance path of each of the other two models. The objective of this exercise is to disentangle each model's ability to extract an adequate variance path from its ability to deliver adequate option prices for a given variance path. The rows in panel C represent the valuation model used, and the columns in panel C represent the

[13] Note that Aït-Sahalia and Kimmel (2007), as well as Jones (2003), use VIX in the filtering of volatility.

variance path, V, used. The shaded diagonal elements of panel C represent the case in which the model's own variance path is used.

Consider first each row of panel C in table 7. Note that for the SQR model, using the volatility path from the ONE model (ONE-V) instead of the model's own volatility path (SQR-V) actually gives a slightly lower IVRMSE in the Wednesday sample and a virtually identical IVRMSE in the Thursday sample. When the ONE model is used to price the options in the second row, the IVRMSE using the model's own ONE-V path yields the lowest IVRMSE in both samples. In the third row, the 3/2 model is used to price options, and again, we see that the ONE-V volatility path produces the lowest IVRMSE in both the Wednesday and Thursday samples. Looking at the averages across rows, we see that the ONE-V path delivers an IVRMSE of 3.13 for Wednesdays and 3.19 for Thursdays. The numbers for the SQR-V paths are 3.25 and 3.30, respectively. The 3/2-V paths give an average IVRMSE of 3.28 on Wednesdays and 3.30 on Thursdays. Although the differences across models are small, it is quite striking how the ONE-V path leads to the best option fit in all three valuation models.

When considering each column in panel C, we see that the ONE model delivers the lowest option IVRMSE regardless of which volatility path is used and regardless of whether we use the Wednesday or Thursday sample. Looking at the average across columns, we see that the ONE model delivers a substantially lower average IVRMSE across volatility paths. For the Wednesday data, the average IVRMSE is 3.01 for the ONE model, compared with 3.25 for the SQR model and 3.40 for the 3/2 model. For the Thursday data, the averages are 2.99 for the ONE model, compared with 3.30 for SQR, and 3.50 for the 3/2 model. These differences are quite substantial and strongly suggest that the ONE model provides the option valuation relationship for a given value of the spot volatility.

In summary, the results obtained using alternative data sources, filtering techniques, and model variance paths confirm that the ONE model offers substantial improvements over the benchmark affine SQR model. While the option-fitting exercise in table 4 shows that the SQR model can be estimated to fit option data reasonably well, table 7 shows that it is robustly outperformed by the ONE model when using alternative filtering techniques. Panel C suggests that the ONE model provides the best option valuation relationship for a given level of spot volatility as well as the best volatility path.

5.5 Considering jumps in returns

To judge the magnitude of the improvement in fit between the ONE and the benchmark SQR model, the most natural reference point is the rich literature that uses Poisson jumps in returns and/or volatility in conjunction with an SQR stochastic volatility model. When estimating using options, the evidence on the in-sample and especially the out-of-sample improvement provided by in-

cluding jump processes is inconclusive, with some studies finding moderate improvements, and others concluding that there is no improvement in fit.[14]

The models we considered in table 4 do not include jumps in returns. Standard jump specifications are most likely to affect options with short maturities. Thus, in order to assess the sensitivity of our NLSIS results to the potential misspecification from omitted jumps, we first reestimate all the models omitting options with less than thirty days to maturity. The results for the Wednesday in-sample and Thursday out-of-sample data sets are reported in the last two columns of table 4 labeled DTM>30. Comparing these results with the full sample results in table 4, we see that the results are very similar. The ONE model is still significantly better than the SQR model. The only substantial change is that the ONEN model now delivers significant improvements on the SQR model both in and out of sample as well.

In table 2, we estimated the jump models in Equations (29) and (30) on returns using MLIS. We now estimate the same jump models using the option-based NLSIS methodology that we applied to the diffusive models in table 4. The risk-neutral process for the jump models we use is

$$\mathrm{d}\ln(S) = \left(r - \frac{1}{2}V - \gamma_J \overline{\mu}_J^*\right)\mathrm{d}t + \sqrt{V}\mathrm{d}z^* + (J\mathrm{d}N)^* \tag{35}$$

$$\mathrm{d}V = \kappa^* V^a(\theta^* - V)\mathrm{d}t + \sigma V^b \mathrm{d}w^*. \tag{36}$$

We follow Eraker (2004) and assume that under the risk-neutral measure, the jump size J is distributed $J \sim N\left(\mu_J^*, \sigma_J^2\right)$ with $\mu_J^* = \mu_J + \lambda_J$, where λ_J is the jump risk premium. As in the diffusive models, we have that $\kappa^* = (\kappa - \lambda)$, and $\theta^* = \kappa\theta/\kappa^*$ where λ is the volatility risk premium. The jump compensation term in the return drift now takes the form $\gamma_J \overline{\mu}_J^*$, where $\overline{\mu}_J^* = \exp\left(\mu_J^* + \frac{1}{2}\sigma_J^2\right) - 1$ and where γ_J is the Poisson jump intensity.

The results for the option-based estimation of the jump models are reported in table 8. Note that we report the risk-neutral mean jump size, μ_J^*, instead of the jump risk premium, λ_J. Table 8 shows that for all models, the mean jump size is negative and larger in magnitude under the risk-neutral measure as expected. When comparing the IVRMSEs in table 8 with the in-sample Wednesday IVRMSEs for the diffusive models in table 4, we see that the improvement for each model is relatively modest. This is consistent with the results we found when comparing the jump models on returns in table 2 with the diffusive models in table 1. Comparing tables 8 and 4, we also see that the ranking of models is unchanged. Finally, note that in table 8, as in table 4, only the ONE model significantly outperforms the benchmark SQR model according to the Diebold–Mariano statistic.

[14] See, for instance, Bates (2000), Pan (2002), and Eraker (2004). In recent work, using a very different empirical setup, Broadie, Chernov, and Johannes (2007) find large improvements in sample when including Poisson jumps.

Table 8
Jump model estimates from returns and options, 1996–2004

Model			Parameter estimates for 1996–2004									Model	
Name	a	b	κ	θ	σ	ρ	μ_J	σ_J	γ_J	μ_J^*	λ	name	IVRMSE
SQR	0	1/2	2.6380	0.0627	0.4481	−0.7821	−3.6565	0.0935	2.8317	−4.1955	1.75E−04	SQR	3.194
Standard errors:			2.14E−03	4.70E−05	3.99E−03	2.03E−04	1.51E−01	6.43E−01	1.19E−01	2.50E−01	1.70E−04	Ratio	1.000
SQRN	1	1/2	90.8951	0.0783	0.4913	−0.6270	−1.9559	0.0979	2.1133	−1.9974	6.54E−04	SQRN	3.344
Standard errors:			4.17E−02	1.35E−05	1.95E−04	3.95E−04	9.43E−01	3.35E−01	1.27E−01	7.82E−01	3.20E−03	Ratio	1.047
												DM test	1.324
ONE	0	1	1.7893	0.0598	1.5423	−0.7627	−0.8503	0.1210	1.9512	−0.8545	5.67E−03	ONE	2.898
Standard errors:			1.49E−03	6.18E−06	7.97E−04	2.38E−04	3.89E−01	9.62E−01	3.50E−01	3.76E−01	1.88E−05	Ratio	0.907
												DM test	**−3.168**
ONEN	1	1	64.0969	0.0911	2.2133	−0.7215	−1.2644	0.0904	2.9055	−2.1165	2.68E−02	ONEN	3.180
Standard errors:			1.02E−01	4.74E−05	9.27E−04	3.14E−04	5.88E−01	8.50E−01	2.84E−01	5.16E−01	5.16E−01	Ratio	0.996
												DM test	−0.342
3/2	0	3/2	1.5819	0.0578	8.4387	−0.6769	−1.7618	0.1051	2.4617	−3.0910	1.39E+00	3/2	3.397
Standard errors:			3.62E−03	3.67E−04	5.07E−03	2.74E−04	9.33E−01	8.48E−01	6.71E−01	6.10E−01	2.10E+00	Ratio	1.064
												DM test	**2.858**
3/2N	1	3/2	27.3842	0.1237	7.7509	−0.7512	−3.7884	0.0464	1.4101	−4.0417	1.10E−04	3/2N	3.145
Standard errors:			5.77E−02	1.92E−04	3.42E−03	2.27E−04	1.18E+00	1.08E+00	4.04E−01	7.37E−01	4.31E−02	Ratio	0.985
												DM test	−0.845

The model parameters are estimated using NLSIS on the 14,828 Wednesday closing OTM option quotes observed from January 4, 1996, to December 31, 2004. Standard errors are computed using the outer product of the gradient at the optimal parameter values. The return drift parameter μ is fixed at the sample average return of 0.091 for all models. The parameters are reported in annual units.

In summary, although we have not undertaken an exhaustive investigation of jump dynamics and jump risk premium specifications, for the simple return jump, the results from the purely diffusive models are robust: the ONE model significantly outperforms the benchmark SQR model and no other models do.

6. Summary and Conclusions

This article provides an empirical comparison of the affine square root model of Heston (1993) with a range of nonaffine but equally parsimonious option valuation models. The main conclusion is that nonaffine specifications can substantially improve upon the affine SQR model. This finding is robust across a wide range of data sources used for volatility filtering, as well as across different estimation techniques. First, the misspecification of the SQR model is suggested by the stylized facts characterizing realized volatility data and the VIX. Second, an extensive estimation exercise on index returns confirms the presence of misspecification in the SQR model. Third, the finding that the SQR volatility dynamic can be improved upon in returns fitting is confirmed when allowing for jumps in returns. Fourth, the nonaffine ONE model leads to substantially smaller in-sample errors in option valuation when estimating using a data set consisting of long time series of cross-sections of OTM option data. These results are confirmed when using the estimates out of sample by pricing options on a different day and when including return jumps in the specification. Fifth, option pricing errors are also smaller for the ONE model when using realized volatility and the VIX as a proxy for spot volatility, and when using the Kalman filter to obtain spot volatility from realized volatility and the VIX.

We therefore conclude that although the option valuation literature's focus on affine models is well motivated, because the resulting closed-form solutions are extremely convenient, this analytical convenience comes at a price, and nonaffine models need to be studied more extensively. Our empirical results provide useful guidance for the specification of nonaffine models. The ONE model consistently performs very well. It provides the best fit for two of the three return samples, and the second best fit using the third sample. When estimating the models using option data, it is the best-performing model in and out of sample.

At the methodological level, we use particle filtering, which is rather flexible and straightforward to implement. It can easily be used to investigate option data and underlying equity returns jointly for a wide range of objective functions as well as for a wide range of models. In our opinion, because of its simplicity and numerical efficiency, this method is attractive compared with other methods that have been used in this literature.

Our empirical results suggest a number of important extensions. First, while nonaffine models substantially outperform affine models, the biases of the affine model can also be addressed by alternative model specifications, such as

multiple volatility components, and/or Levy jumps in returns and volatility.[15] This study analyzes nonaffine specifications without considering these alternatives, but ultimately, it will be of interest to investigate the relative advantages of nonaffine specifications, multiple volatility components, and different types of jump processes. Second, the finding from the returns data that the 3/2N model is particularly useful in a sample that includes the 1987 crash indicates that another interesting avenue for future work is to focus more explicitly on the CEV model, and particularly to investigate its performance when periodically reestimating the CEV parameter. Third, the filtering analysis in Section 5.4 provides some interesting insights and may be worth extending, for instance by combining returns, options, VIX, and RV in the filter. Finally, all the models we consider contain a latent stochastic volatility factor, but none relate this factor to underlying economic variables. Understanding the links between fundamental economic variables and the stochastic volatility factor is left for future research.

A. Appendix

In this Appendix, we first study the properties of Monte Carlo-based option prices. We compare Monte Carlo prices for the Heston (1993) SQR model with prices computed via Fourier inversion of the conditional characteristic function. We then study the finite sample properties of the MLIS estimator, comparing it with other available estimators in the literature.

A.1 Option price computation by simulation

To assess the accuracy of our Monte Carlo setup, we compute a set of analytical option prices using Heston's (1993) model for various strike prices, moneyness, and maturities. We then compute Monte Carlo option prices for the same options. Using the numerical example in Heston (1993), the risk-neutral variance process is parameterized as follows:

$$dV = 2(.01 - V)dt + .2\sqrt{V}dw^*. \tag{A1}$$

The correlation ρ is set to -0.5, and the risk-free rate is set to zero.

Figure A1 reports the results. The Monte Carlo prices, denoted by "+," offer a very good approximation to the analytical prices, denoted by solid lines, in all cases. The figure contains two maturities: one month (left column) and three months (right column). The figure considers three spot variance levels: $V_0 = .005$, which is half the unconditional variance (top row); $V_0 = .01$, which is equal to the unconditional variance (middle row); and $V_0 = .02$, which is twice the unconditional variance (bottom row).

A.2 Finite sample performance of MLIS

We assess the finite sample performance of the relatively new MLIS estimation procedure based on the PF. Following Bates (2006), who introduces a new approximate ML estimation procedure, we

[15] See Carr and Wu (2004) and Huang and Wu (2004) for option valuation studies using Levy processes.

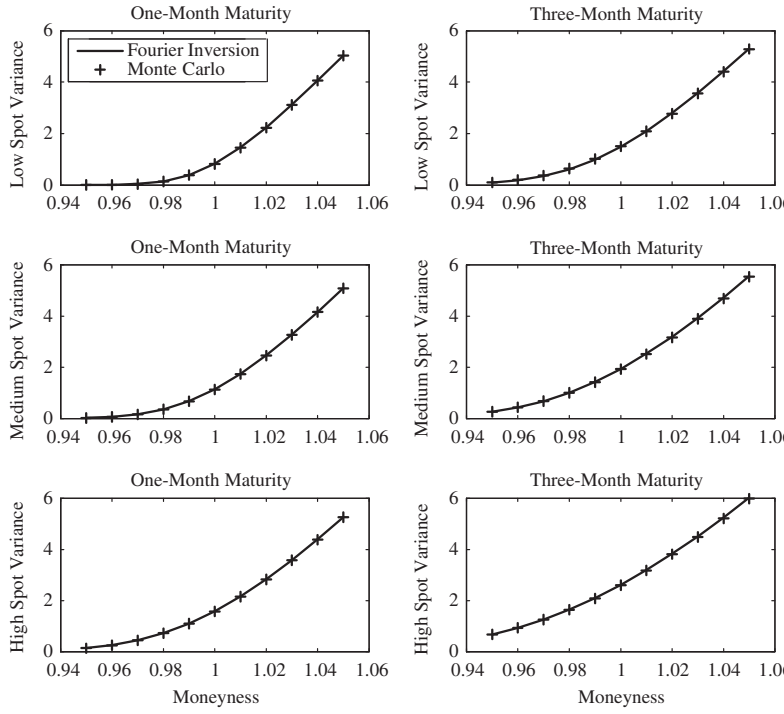

Figure A1
Analytical (Fourier inversion) and Monte Carlo prices for the Heston (1993) model
We first compute option prices from Heston's (1993) model using the Fourier inversion technique (solid) and then using Monte Carlo simulation ("+"). We consider different moneyness (F/X) on the horizontal axis. We consider two maturities: one month (left column) and three months (right column). We consider three spot variance levels: half the unconditional variance (top row), equal to the unconditional variance (middle row), and twice the unconditional variance (bottom row).

add the MLIS estimator to the large-scale Monte Carlo study in Andersen, Chung, and Sørensen (1999), henceforth ACS. ACS use the following simple SV model as a data generating process:

$$\ln(S_{t+1}) = \ln(S_t) + \sqrt{V_t} z_{t+1} \tag{A2}$$

$$\ln\left(V_{t+1}\right) = \omega + \phi \ln\left(V_t\right) + \sigma w_{t+1}, \tag{A3}$$

where z_{t+1} and w_{t+1} are uncorrelated. They consider a large number of estimators, including the QML method from Harvey, Ruiz, and Shephard (1994), the GMM from Andersen and Sørensen (1996), the MCMC from Jacquier, Polson, and Rossi (1994), as well as the EMM implementation suggested by ACS themselves. Table A.1 reproduces the results from ACS and includes the approximate maximum likelihood (AML) technique from Bates (2006). We of course also include the MLIS estimation procedure from Pitt (2002), which is used in this article. The various estimators are compared with the benchmark (but of course unrealistic) case in which the spot variance is observed and standard ML estimation is straightforward. We report results for the sample size $T = 2{,}000$, which is relevant for our subsequent empirical study.

Table A.1 reports parameter bias and RMSE for each estimator using five hundred Monte Carlo replications. The MCMC estimator has the lowest absolute bias for the ω parameter followed by the AML and the MLIS. The MLIS, MCMC, and AML estimators have the lowest absolute

Table A.1
Monte Carlo study of MLIS and alternative estimation methods

Estimators	
ML|V	ML conditional on observing the true spot variance
QML	Quasi ML as in Harvey, Ruiz, and Shephard (1994)
GMM	Andersen and Sørensen (1996)
EMM	Efficient method of moments as in Andersen, Chung, and Sørensen (1999)
AML	Approximate ML of Bates (2006)
MCMC	Markov chain Monte Carlo as in Jacquier, Polson, and Rossi (1994)
MLIS	Method used in this article; see Pitt (2002)

Parameter	ω	ϕ	σ
True values	−0.736	0.9	0.363
Estimators		**Bias**	
ML|V	−0.015	−0.002	0.000
QML	−0.117	−0.020	0.020
GMM	0.150	0.020	−0.080
EMM	−0.057	−0.007	−0.004
AML	−0.039	0.005	0.005
MCMC	−0.026	−0.004	−0.004
MLIS	**0.049**	**0.004**	**−0.017**
Estimators		**RMSE**	
ML|V	0.076	0.010	0.006
QML	0.460	0.060	0.110
GMM	0.310	0.040	0.120
EMM	0.224	0.030	0.049
AML	0.173	0.023	0.043
MCMC	0.150	0.020	0.034
MLIS	**0.157**	**0.020**	**0.046**

Following Bates (2006), we compare the MLIS estimator with those considered in the large-scale Monte Carlo study by Andersen, Chung, and Sørensen (1999). We generate five hundred Monte Carlo samples of two thousand returns with zero drift, and stochastic volatility following the logarithmic SV model in Equations (A2) and (A3) with constant term ω, persistence φ, and diffusion parameter σ. We then estimate the model using MLIS and compare the bias and RMSE from MLIS with the bias and RMSE reported in Bates (2006) and Andersen, Chung, and Sørensen (1999).

biases for ϕ. The AML, EMM, and MCMC estimators have the lowest biases for the σ parameter followed by the MLIS. The MCMC and MLIS estimators have the lowest RMSE for ω; AML, MCMC, and MLIS have the lowest RMSE for ϕ; and MCMC has the lowest RMSE for the σ parameter followed by AML and MLIS. Overall, the MLIS estimator seems to perform well. The fact that it is also easily implementable across a wide range of models and objective functions makes it a very suitable estimation method in our large-scale empirical study.

References

Aït-Sahalia, Y. 1996. Testing Continuous-Time Models of the Spot Interest Rate. *Review of Financial Studies* 9:385–426.

Aït-Sahalia, Y., and R. Kimmel. 2007. Maximum Likelihood Estimation of Stochastic Volatility Models. *Journal of Financial Economics* 83:413–52.

Aït-Sahalia, Y., P. Mykland, and L. Zhang. 2005. How Often to Sample a Continuous-Time Process in the Presence of Market Microstructure Noise. *Review of Financial Studies* 18:351–416.

Andersen, T. G., L. Benzoni, and J. Lund. 2002. Estimating Jump-Diffusions for Equity Returns. *Journal of Finance* 57:1239–84.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and H. Ebens. 2001. The Distribution of Realized Stock Return Volatility. *Journal of Financial Economics* 61:43–76.

Andersen, T. G., H. Chung, and B.E. Sørensen. 1999. Efficient Method of Moments Estimation of a Stochastic Volatility Model: A Monte Carlo Study. *Journal of Econometrics* 91:61–87.

Andersen, T. G., and B. E. Sørensen. 1996. GMM Estimation of a Stochastic Volatility Model: A Monte Carlo Study. *Journal of Business and Economic Statistics* 14:328–52.

Bakshi, G., C. Cao, and Z. Chen. 1997. Empirical Performance of Alternative Option Pricing Models. *Journal of Finance* 52:2003–49.

Barone-Adesi, G., H. Rasmussen, and C. Ravanelli. 2005. An Option Pricing Formula for the GARCH Diffusion Model. *Computational Statistics and Data Analysis* 49:287–310.

Bates, D. 1996. Jumps and Stochastic Volatility: Exchange Rate Processes Implicit in Deutsche Mark Options. *Review of Financial Studies* 9:69–107.

———. 2000. Post-'87 Crash Fears in the S&P 500 Futures Option Market. *Journal of Econometrics* 94:181–238.

———. 2003. Empirical Option Pricing: A Retrospection. *Journal of Econometrics* 116:387–404.

———. 2006. Maximum Likelihood Estimation of Latent Affine Processes. *Review of Financial Studies* 19:909–65.

Benzoni, L. 2002. Pricing Options under Stochastic Volatility: An Empirical Investigation. Manuscript, University of Minnesota.

Black, F. 1976. Studies of Stock Price Volatility Changes. *Proceedings of the 1976 Meetings of the Business and Economic Statistics Section, American Statistical Association*, 177–81.

Black, F., and M. Scholes. 1973. The Pricing of Options and Corporate Liabilities. *Journal of Political Economy* 81:637–59.

Broadie, M., Chernov, M., and M. Johannes. 2007. Model Specification and Risk Premiums: Evidence from Futures Options. *Journal of Finance* 62:1453–90.

Carr, P., and L. Wu. 2004. Time-Changed Levy Processes and Option Pricing. *Journal of Financial Economics* 17:113–41.

Chacko, G., and L. Viceira. 2003. Spectral GMM Estimation of Continuous-Time Processes. *Journal of Econometrics* 116:259–92.

Chernov, M., A. R. Gallant, E. Ghysels, and G. Tauchen. 2003. Alternative Models for Stock Price Dynamics. *Journal of Econometrics* 116:225–57.

Chernov, M., and E. Ghysels. 2000. A Study Towards a Unified Approach to the Joint Estimation of Objective and Risk Neutral Measures for the Purpose of Option Valuation. *Journal of Financial Economics* 56:407–58.

Christoffersen, P., and K. Jacobs. 2004. The Importance of the Loss Function in Option Valuation. *Journal of Financial Economics* 72:291–318.

Diebold, F. X., and R. S. Mariano. 1995. Comparing Predictive Accuracy. *Journal of Business and Economic Statistics* 13:253–65.

Duan, J.-C., and J.-G. Simonato. 1998. Empirical Martingale Simulation for Asset Prices. *Management Science* 44:1218–33.

Dumas, B., J. Fleming, and R. Whaley. 1998. Implied Volatility Functions: Empirical Tests. *Journal of Finance* 53:2059–106.

Efron, B., and R. Tibshirani. 1993. *An Introduction to the Bootstrap*. New York: Chapman and Hill.

Eraker, B. 2001. MCMC Analysis of Diffusion Models with Application to Finance. *Journal of Business and Economic Statistics* 19:177–91.

———. 2004. Do Stock Prices and Volatility Jump? Reconciling Evidence from Spot and Option Prices. *Journal of Finance* 59:1367–403.

Eraker, B., M. Johannes, and N. Polson. 2003. The Role of Jumps in Returns and Volatility. *Journal of Finance* 58:1269–300.

Gordon, N. J., D. J. Salmond, and A. F. M. Smith. 1993. A Novel Approach to Nonlinear/Non-Gaussian Bayesian State Estimation. *IEE-Proceedings F* 140:107–33.

Granger, C. 1969. Prediction with a Generalized Cost of Error Function. *Operations Research Quarterly* 20:199–207.

Harvey, A., E. Ruiz, and N. Shephard. 1994. Multivariate Stochastic Variance Models. *Review of Economic Studies* 61:247–64.

Heston, S. 1993. A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options. *Review of Financial Studies* 6:327–43.

———. 1997. A Simple New Formula for Options with Stochastic Volatility. Manuscript, John M. Olin School of Business, Washington University.

Heston, S., and S. Nandi. 2000. A Closed-Form GARCH Option Pricing Model. *Review of Financial Studies* 13:585–626.

Huang, J.-Z., and L. Wu. 2004. Specification Analysis of Option Pricing Models Based on Time-Changed Levy Processes. *Journal of Finance* 59:1405–39.

Jacquier, E., N. G. Polson, and P. E. Rossi. 1994. Bayesian Analysis of Stochastic Volatility Models. *Journal of Business and Economic Statistics* 12:1–19.

Johannes, M., N. Polson, and J. Stroud. 2009. Optimal Filtering of Jump-Diffusions: Extracting Latent States from Asset Prices. *Review of Financial Studies* 22:2759–99.

Jones, C. 2003. The Dynamics of Stochastic Volatility: Evidence from Underlying and Options Markets. *Journal of Econometrics* 116:181–224.

Karlin, S., and H. M. Taylor. 1981. *A Second Course in Stochastic Processes*. San Diego: Academic Press.

Lewis, A. 2000. *Option Valuation Under Stochastic Volatility*. Newport Beach, CA: Finance Press.

Pan, J. 2002. The Jump-Risk Premia Implicit in Options: Evidence from an Integrated Time-Series Study. *Journal of Financial Economics* 63:3–50.

Pitt, M. 2002. Smooth Particle Filters for Likelihood Evaluation and Maximization. Manuscript, University of Warwick.

Pitt, M., and N. Shephard. 1999. Filtering via Simulation: Auxiliary Particle Filters. *Journal of the American Statistical Association* 94:590–99.

Sabanis, S. 2003. Stochastic Volatility and the Mean Reverting Process. *Journal of Futures Markets* 23:33–47.

Smith, A., and A. Gelfand. 1992. Bayesian Statistics without Tears: A Sampling–Resampling Perspective. *American Statistician* 46:84–88.

Zhang, L., P. A. Mykland, and Y. Aït-Sahalia. 2005. A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Data. *Journal of the American Statistical Association* 100:1394–411.