Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Econometrics 116 (2003) 181–224
JOURNAL OF Econometrics
www.elsevier.com/locate/econbase

The dynamics of stochastic volatility: evidence from underlying and options markets

Christopher S. Jones*

Marshall School of Business, University of Southern California, Los Angeles, CA 90089, USA

Abstract

This paper proposes and estimates a more general parametric stochastic variance model of equity index returns than has been previously considered using data from both underlying and options markets. I conclude that the square root stochastic variance model of Heston (Rev. Financial Stud. 6 (1993) 327) is incapable of generating realistic returns behavior, and that the data are better represented by a stochastic variance model in the CEV class or a model with a time-varying leverage effect. As the level of market variance increases, the volatility of market variance increases rapidly and the leverage effect becomes substantially stronger. The heightened heteroskedasticity in market variance that results causes returns to display unconditional skewness and kurtosis much closer to their sample values, while the model falls short of explaining the implied volatility smile for short-dated options and conditional higher moments in returns.

JEL classification: G12; C11

Keywords: Continuous-time estimation; Stochastic volatility; Option pricing; Bayesian analysis; Stock market crashes

1. Introduction

The dynamic nature of asset price volatility has long been a popular subject in the empirical finance literature, and even greater interest in volatility dynamics has followed the development of stochastic volatility models of option pricing. These models attribute the higher prices of options far from "at-the-money" to return non-normality generated

* Corresponding author.
E-mail address: christopher.jones@marshall.usc.edu (C.S. Jones).

0304-4076/03/$ - see front matter
doi:10.1016/S0304-4076(03)00107-6

by time-varying volatility. Because much of the interest in options focuses on the options that are furthest from at-the-money and therefore have positive payoffs only in extremely rare circumstances, precise characterizations of volatility dynamics are particularly important.

While the need to better model extreme events for the purposes of option pricing made the task of understanding volatility more exacting, the existence of options has made the researcher's data set much richer, so that more finely tuned models may be investigated with greater hope of obtaining accurate results.

Traditionally, volatility dynamics in asset prices have been explored using the time series of returns on the assets being studied. Following the work of Engle (1982) and Bollerslev (1986), models in the ARCH class have been studied extensively for this purpose and remain popular today. Perhaps because of the rapid growth in derivatives markets, continuous-time stochastic volatility models have recently become popular as well, and a growing body of research is concerned with fitting these models to asset returns. Two recent examples are the papers by Gallant and Tauchen (1997) and Andersen et al. (2002).

An alternative approach is to infer the risk-neutralized volatility dynamics from the prices of options. While methods of inferring risk-neutral returns probabilities from the cross section of options prices are well-known following the seminal work of Breeden and Litzenberger (1978), more recent studies such as Bates (2000) and Bakshi et al. (1997) fit parametric stochastic volatility models to option prices to learn about the parameters of these models.

But since the risk-neutral and objective measures are not wholly dissimilar, there must be an advantage to using both sets of data (the time series of the underlying's prices and the prices of options on it) to infer both measures simultaneously, a point made forcefully by Chernov and Ghysels (2000a). In addition, since it is the difference between the two measures that defines risk premia, it may further be advantageous to estimate the two measures together so that estimation errors relating to these differences can be treated in a sensible way. In fact, a number of recent papers have pursued this idea. Pan (2002), Poteshman (1998), and Chernov and Ghysels (2000a) all use some combination of a time series of the underlying price and one or more time series of liquid, short-term, near-the-money option prices to infer the dynamics of the underlying volatility process under both the objective and risk-neutral measures. [1]

In its use of option implied volatilities for the purpose of estimating more realistic volatility dynamics, the paper is related to a growing literature on alternative volatility proxies. The use of high-frequency realized volatilities (often calculated at 5-min intervals) is increasingly popular, pursued by Andersen and Bollerslev (1997) and Andersen et al. (2001) among many others. The range, or the high price minus the low price over some interval, has been employed recently by Gallant et al. (1999) and Alizadeh et al. (2001) for similar purposes. It is possible that implied volatility provides a proxy that is less noisy than the range and easier to work with

[1] Studies by Benzoni (2001) and Jiang and van der Sluis (1998) have similar objectives, but both use options data only to estimate the parameters that are not identified by data on the underlying. Additional literature is reviewed in Chernov and Ghysels (2000b).

than high-frequency realized volatility, although this paper provides no evidence for or against this conjecture.

In this paper, I consider a more general form of the one-factor stochastic variance model than has been considered previously and estimate the model in a Bayesian framework from a bivariate time series of returns and option implied volatilities. The analysis results in parameter estimates for both the objective and risk-neutral processes. The results show that the model I propose generates more realistic dynamics for stochastic variance and for equity index returns than does the standard square root model used by Heston (1993) and many others. I find that as variance increases, its *own* volatility increases at an even faster rate. Furthermore, the so-called "leverage effect," or the negative correlation between the price and instantaneous variance processes, becomes substantially stronger at higher levels of variance. A positive shock to market variance therefore increases the probability that market variance may grow even larger. The compound effect of several of these shocks can be large, and extreme returns may result. Because the price and variance processes are negatively correlated, particularly in high volatility states, this extreme return will more often than not be negative, frequently of a magnitude that the square root process is incapable of generating.

I focus on the implications of these models for higher moments of asset returns, since many of the outstanding puzzles in options markets and other areas of empirical asset pricing depend crucially on violations of the normality of returns. I find a realistic degree of unconditional non-normality implied by the new models, while the higher moments generated by the square root model fall far short of the levels observed in actual data. When conditioned on a value for the market variance, very short-horizon returns have a distribution that is much closer to Gaussian. For a monthly horizon, however, the new models generate conditional distributions with fatter tails than even a stochastic volatility jump-diffusion estimated from a similar data set. Unfortunately, model-implied options prices, while greatly improved over the square root specification, still fail to match observed prices of short-dated options.

Section 2 of this paper motivates the econometric specification and outlines a few of its properties. The estimation method used in the paper is described in Section 3. Section 4 contains the results, including parameter estimates and the models' implications for higher moments and option prices. Section 5 summarizes and discusses the results.

2. The model

By far the most popular recent formulations of stochastic volatility in continuous time have been variants of the stochastic variance model of Heston (1993). Using a square root process to represent the dynamics of instantaneous variance, Heston assumes the process

$$\begin{aligned} \mathrm{d}S_t &= \mu S_t\,\mathrm{d}t + \sqrt{V_t}S_t\,\mathrm{d}Z_t^{(1)}, \\ \mathrm{d}V_t &= (\alpha + \beta V_t)\,\mathrm{d}t + \sigma\sqrt{V_t}\,\mathrm{d}Z_t^{(2)}, \end{aligned} \tag{1}$$

where $dZ_t^{(1)}$ and $dZ_t^{(2)}$ are increments of Brownian motion with correlation ρ. Heston demonstrates how Fourier inversion techniques may be used to produce an analytical stochastic variance option pricing formula.

For option pricing, much interest centers around the values of β, σ, and ρ, which determine the ways in which the distribution of S_t departs from the lognormal. Kurtosis depends in large part on the magnitude of σ relative to that of β. If σ is relatively large, more "volatile" variance will lead to fat tails, raising the prices of all options far from at-the-money. Skewness depends in addition on ρ, and it has often been argued that the typical downward slope of Black–Scholes implied volatilities of equity options across moneyness (the "leverage" effect) is evidence of a negative value of ρ.

Unfortunately, the square root specification is generally rejected as a model of stock index returns. While Andersen et al. (2002) find that the rejection is due primarily to the insufficient kurtosis generated by the model, Pan (2002) rejects the square root model in part based on its implications for the term structure of volatility. Other rejections are reported by Chernov and Ghysels (2000a), Benzoni (2001), and less formally by Bakshi et al. (1997).

Subsequent generalizations to the model have included stochastic interest rates (Scott, 1997; Bakshi et al., 1997), stochastic dividends (Pan, 2002), and jumps (Bates, 2000; Bakshi et al., 1997; Scott, 1997; Duffie et al., 2000; Pan, 2002; Eraker et al., 2003). Of the three, only the inclusion of jumps has shown much promise for reversing the rejections of the model.

For several reasons, however, the econometric analysis of jumps in option pricing has proven to be problematic. First, small sample problems may be severe because jumps that are large enough to have a substantial effect on options prices must be infrequent in order to be consistent with the underlying returns data. These types of large jumps are seldom found in empirical analysis of jump-diffusions applied to broad market indexes.[2] Secondly, although jumps may help fit the prices of options, including them is not necessarily beneficial for the purpose of option hedging. Bakshi et al. (1997, 2000) find that adding jumps to the square root stochastic variance model has little effect on pricing or hedging longer maturity options and actually worsens hedging performance for short maturities. This result is somewhat surprising since adding jumps leads to the greatest improvement in the pricing of short maturity options, and it raises the possibility that including jumps amounts to overfitting.

In this paper, I focus on stochastic variance alone as the source of non-Gaussian return dynamics. While including jumps will naturally improve the fit of options prices, it seems reasonable to investigate more general forms of the one-factor stochastic variance model before pursuing additional extensions such as jumps, particularly because of the problems with jumps and the soon-to-be obvious misspecification of the square root model.[3]

[2] When jump processes are estimated using the time series of returns only, the jump intensity estimate tends to be high and the jump size small. Studies that find relatively small jumps include Andersen et al. (2002), Ball and Torous (1985), Jorion (1989), and Pan (2002). Using a Bayesian approach, Eraker et al. (2003) find larger jumps, with a more negative mean, than any of these papers. The implications of EJP's estimates are considered below.

[3] Poteshman (1998) also has this goal.

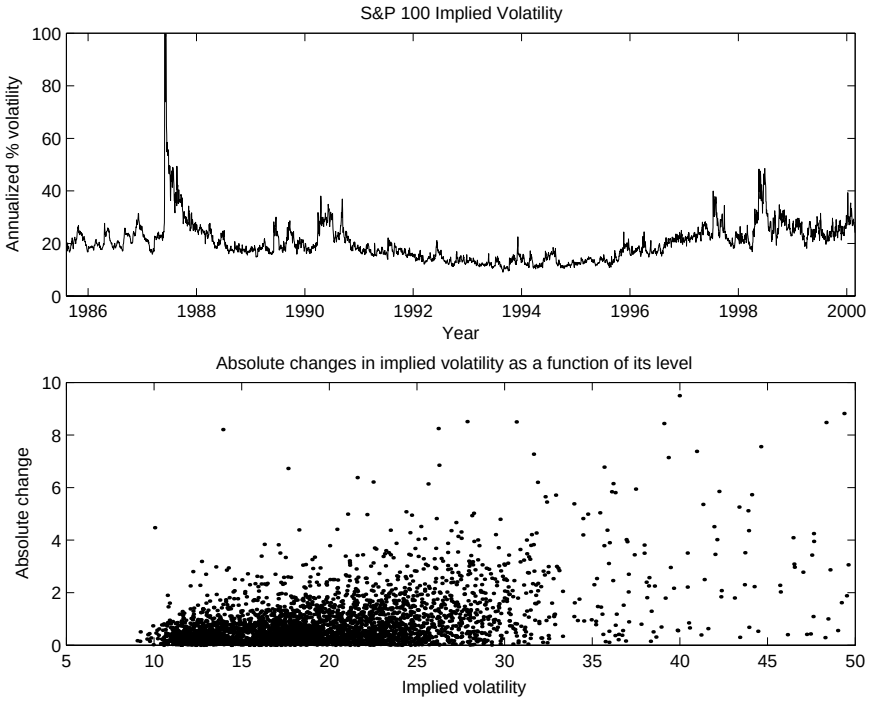

Fig. 1. Properties of implied volatility.

2.1. Stochastic specification

The goal in this section is to propose a more general parametric form of stochastic variance process than has been considered previously. The motivation is empirical, and it is based on the a casual inspection of the top panel of Fig. 1, which plots the VIX index, an index of the Black–Scholes implied volatility of S&P 100 index options, from January 2, 1986, to June 2, 2000. The plot is truncated at 100% for visual clarity; in the 2 weeks following the crash of October 1987, implied volatilities as high as 150% were recorded. The lower panel of Fig. 1 plots the absolute change in the VIX index versus its level at the previous close over the same period.

From panels it appears that periods of high volatility coincide with periods of volatile volatility, an observation that appears to be at odds with Heston's square root formulation. We can see this by deriving the stochastic volatility (as opposed to variance) process implied by Heston's model.

Let v_t denote the time t instantaneous volatility, so $v_t = \sqrt{V_t}$. By Ito's lemma,

$$\mathrm{d}v_t = -\frac{\beta}{2v_t}\left(-\frac{\alpha}{\beta} + \frac{\sigma^2}{4\beta} - v_t^2\right)\mathrm{d}t + \frac{1}{2}\sigma\,\mathrm{d}Z_t^{(2)}. \tag{2}$$

The equation shows that the volatility of instantaneous volatility (as opposed to the volatility of instantaneous variance) is not level dependent under the square root variance model, at least over short intervals.

Because level dependence does seem to be a feature of the data, this paper will consider a somewhat broader class of models for stochastic variance. The first generalization I consider is identical to the one offered for interest rate processes by Chan et al. (1992). The square root in the variance diffusion term is simply replaced by an exponent of undetermined magnitude, so that we have

$$\begin{aligned} dS_t &= \mu S_t\,dt + \sqrt{V_t} S_t\,dZ_t^{(1)}, \\ dV_t &= (\alpha + \beta V_t)\,dt + \sigma V_t^{\gamma_1}\,dZ_t^{(2)}, \end{aligned} \tag{3}$$

where again $\mathrm{Corr}(dZ_t^{(1)}, dZ_t^{(2)}) = \rho$.

Following Chan et al., this constant of elasticity of variance (CEV) specification has been investigated in interest rates in a number of papers, and has been applied to stochastic variance processes by Chacko and Viceira (2001) and Lewis (2000). In empirical analysis of US short-term interest rates, starting with Chan et al., typical estimates of γ_1 are around 1.5. Conley et al. (1997) recently noted that such processes will have stationary distributions even when β is positive, making mean reverting drift an unnecessary condition for stationarity.

It will be convenient to reparameterize the model to work with uncorrelated Brownian motions $B_t^{(1)}$ and $B_t^{(2)}$. Henceforth I will maintain $\mathrm{Corr}(dB_t^{(1)}, dB_t^{(2)}) = 0$ and introduce correlation between S_t and V_t by rewriting the model as

$$\begin{aligned} dS_t &= \mu S_t\,dt + \sqrt{V_t} S_t\,dB_t^{(1)}, \\ dV_t &= (\alpha + \beta V_t)\,dt + \sigma_1 V_t^{\gamma_1}\,dB_t^{(1)} + \sigma_2 V_t^{\gamma_1}\,dB_t^{(2)}. \end{aligned} \tag{4}$$

This is equivalent to the previous parameterization (3) once we set $\sigma = \sqrt{\sigma_1^2 + \sigma_2^2}$ and $\rho = \sigma_1 / \sqrt{\sigma_1^2 + \sigma_2^2}$.

Writing the model as in (4) suggests a second generalization of the variance process that falls outside the CEV class:

$$dV_t = (\alpha + \beta V_t)dt + \sigma_1 V_t^{\gamma_1}\,dB_t^{(1)} + \sigma_2 V_t^{\gamma_2}\,dB_t^{(2)}. \tag{5}$$

By introducing separate power parameters on the two random shocks to instantaneous variance, the elasticity of variance is no longer constant, but depends instead on the level of the variance process. More importantly, by allowing γ_1 and γ_2 to differ the model allows the correlation of the price and variance processes to depend on the level of instantaneous variance. This correlation will be equal to

$$\rho(V) = \frac{\sigma_1 V_t^{\gamma_1}}{\sqrt{\sigma_1^2 V_t^{2\gamma_1} + \sigma_2^2 V_t^{2\gamma_2}}}. \tag{6}$$

One can easily see that the correlation will increase in magnitude as V_t gets larger if $\gamma_1/\gamma_2 > 1$. I will call this specification the 2GAM model.

As other studies have done, I impose an admittedly ad hoc form for the variance risk premium and assume that the form of the risk-neutralized process is identical to that under the true measure. Replacing the stock price drift with the instantaneous interest rate r and replacing the variance drift parameter β with its risk-neutral counterpart β^*, we have

$$\begin{aligned} dS_t &= rS_t\,dt + \sqrt{V_t}S_t\,dB_t^{(1)*}, \\ dV_t &= (\alpha + \beta^* V_t)\,dt + \sigma_1 V_t^{\gamma_1}\,dB_t^{(1)*} + \sigma_2 V_t^{\gamma_2}\,dB_t^{(2)*}. \end{aligned} \tag{7}$$

The implied price of variance risk is therefore

$$\frac{(\beta - \beta^*)V}{\sqrt{\sigma_1^2 V^{2\gamma_1} + \sigma_2^2 V^{2\gamma_2}}}. \tag{8}$$

Section C.8 verifies the Novikov condition for the CEV model, although a proof for the 2GAM model is unavailable.

2.2. Regularity conditions

One drawback to the variance processes in (3) and (5) is that they do not generally satisfy commonly used sufficient conditions for a number of important results. One problem stems from the fact that global growth and Lipschitz conditions may be violated by the diffusion function depending on the values of γ_1 and γ_2. For many purposes (such as proving existence and uniqueness of solutions) violations of these conditions are not necessarily problematic, as shown in Appendix C. They do, however, raise concerns about the assumed convergence of the Euler approximation scheme used below. For the CEV process, the growth condition will be violated as $V \to \infty$ if $\gamma_1 > 1$, while the Lipschitz condition is violated as $V \to 0$ if $\gamma_1 < 1$. For the 2GAM model, the two conditions can be violated simultaneously if $\gamma_1 < 1 < \gamma_2$ or $\gamma_2 < 1 < \gamma_1$.

An additional issue, which will be shown below, is that for some values of $\gamma_1 > 1$ the CEV variance process can lack all unconditional moments other than the first. This in turn implies potentially infinite skewness and kurtosis of market returns. In fact, the parameter estimates reported in Table 1 suggest this behavior for the CEV process. While analytical results are not available for the 2GAM model, it is likely that it also lacks these moments given the parameter estimates in Table 1.

To differing degrees, the Bayesian approach addresses both issues. First, the existence of moments is irrelevant for Bayesian inference, which only requires the existence of transition densities, a much weaker condition. Secondly, violations of growth and Lipschitz conditions outside the range of data observed in the sample are less critical to the Bayesian, whose calculations are all conditional on the observed sample, than they are to the frequentist, who must consider other hypothetical samples.[4] Nevertheless, violations of these conditions prompt the investigation of the validity of the Euler approximation, in particular if the scheme is used to simulate relatively long paths for

[4] There are some important exceptions to this statement. Bayesian analysis under the Jeffreys prior, for instance, requires the integration across alternative realizations of the data.

the purpose of options pricing. The results of several tests, described in Section C.10, are highly supportive of the accuracy of the approach taken.

2.3. Expected average variance and option price approximations

The cost of the model generalizations proposed is the loss of closed form option pricing formulas. While not offering closed form prices, the CEV and 2GAM models may still provide accurate approximate solutions for the prices of relatively short-term at-the-money options.

The approximation is based on Hull and White's (1987) argument that when the price and variance processes are uncorrelated, the price of an option under any stochastic variance model could be computed as the expectation of the Black–Scholes price with the variance argument replaced by the random future realization of average variance over the life of the option. If variance is not priced, as Hull and White assume, then the expectation is taken under the true measure. In general, however, if $\beta \neq \beta^*$ then the expectation would be taken under the risk-neutral measure.

Let $\bar{V}$ denote the average variance over the life of the option, or

$$\bar{V}_{t,t+\tau} = \frac{1}{\tau} \int_t^{t+\tau} V_s \, \mathrm{d}s. \tag{9}$$

Because the Black–Scholes formula is close to linear in its variance argument for short-term at-the-money options, approximating $\mathrm{E}^Q[\mathrm{BS}(\bar{V})]$ by $\mathrm{BS}(\mathrm{E}^Q[\bar{V}])$ may not result in large approximation errors, as the Jensen's inequality error will tend to be minor.

Correlation between the price and variance process presents an additional reason to invalidate this approximation, but the practical importance of correlation is unclear for at-the-money options. To the extent that price-variance correlation and skewness are related, the analysis of Jarrow and Rudd (1982) suggests that for these options the additional approximation error caused by non-zero correlation is likely to be small.[5]

Approximating the Black–Scholes implied variance by the expected average future variance is especially attractive for stochastic variance models such as CEV or 2GAM because expected average variance can be accurately and easily approximated. Given its linear drift, the expected average variance is approximately linear in the current state of instantaneous variance, or

$$\mathrm{E}_t^Q[\bar{V}_{t,t+\tau}] \approx A + BV_t,$$

$$\text{where } B = \frac{1}{\beta^*\tau}(\mathrm{e}^{\beta^*\tau} - 1) \quad \text{and} \quad A = -\frac{\alpha}{\beta^*}(1 - B). \tag{10}$$

By using Black–Scholes implied variance as a proxy for the quantity on the left-hand side, $\mathrm{E}_t^Q[\bar{V}_{t,t+\tau}]$, the relation can be used to provide some information about the otherwise unobservable process V_t. Even once error is introduced into the equation, the

[5] Results in an unpublished appendix, available from the author, verify the validity of the approximation under a wide range of parameter values.

equation can be used to draw more precise inference about the path of V_t than would be possible using returns alone. This, in turn, will lead to more precise inference about the parameters that determine the dynamics of V_t, namely α, β, σ_1, σ_2, γ_1, γ_2, and ρ. Furthermore, this relation between implied and instantaneous variance can also be used to infer the value of the otherwise unidentified risk-neutral parameter β^*.

3. Estimation strategy

This paper will adopt a Bayesian perspective and will use the procedure developed in Jones (1998) to compute posterior distributions for the parameters in (5) and (7). The Bayesian framework is used primarily for reasons of efficiency and tractability. It is a likelihood-based method and should therefore make full use of the information in the sample, providing exact finite-sample even for very complicated models. Although Bayesian analysis is naturally suited to situations in which the researcher has some prior information, the priors used in this analysis will be nearly uninformative, so this is of limited importance.

In this section I will describe the data, specify the econometric model, review the data augmentation procedure of Jones (1998), and propose an extension to this method that will allow the information in implied volatilities to be extracted in a simple way.

3.1. The data

The models are estimated using daily S&P 100 index returns and daily implied volatilities from the S&P 100 options market. The time period considered extends from January 2, 1986 to June 2, 2000, a sample size of 3537 observations. In many cases I will consider a "post-crash" subsample beginning on July 1, 1988, which contains 2907 observations.

The S&P 100 was chosen over the S&P 500 partly because of the S&P 100's more liquid options market during the 1980s. The more compelling reason to analyze the S&P 100 index, however, is the availability of the Chicago Board Options Exchange Market Volatility Index (VIX). This index represents an average of eight near-term near-the-money Black–Scholes implied volatilities from options on the S&P 100 index. The construction of the VIX index is described in detail in Whaley (1993).

Use of the VIX index could be criticized on the grounds that it is a hypothetical measure, since the index does not generally coincide with the implied volatility of any particular option traded that day. Similar criticism could be aimed at term structure research that uses bootstrapped and interpolated values to construct constant maturity zero coupon yields. Partly because of these concerns, I will allow for the possibility that the VIX index makes random deviations from the implied volatility of the hypothetical option that it is designed to mimic. An assessment of the magnitude of this error will be an output of the estimation.

This relaxation of the purely deterministic link between data from the underlying and options markets represents a departure from previous studies that have combined data

from the two markets. In theory, given the synchronous observation of the underlying asset's price and perhaps some other market fundamentals, option prices can be inverted to yield the underlying asset's exact instantaneous volatility.[6] Relaxing the strictness of this relation reflects a purely empirical perspective and is similar to the allowance of additional error terms in term structure research, where theory implies, for example, that bond yields of all maturities are deterministically linked in a single-factor term structure model. The realities of asset price data, which include sources of error such as asynchronous data and bid-ask spreads, make deterministic relations as unrealistic in options.[7]

3.2. The econometric model of implied volatility

The full model to be estimated is described by the SDEs that govern the dynamics of S_t and V_t under the objective measure, given by (3) or (5), and a regression-type equation that links the unobservable V_t to an observed implied variance,

$$IV_t = A + BV_t + \varepsilon_t. \tag{11}$$

Eq. (11) builds on (10) with the addition of an error term and the replacement of $\mathrm{E}_t^Q[\bar{V}_{t,t+\tau}]$ by its proxy, IV_t. The time series IV_t is a simple transformation of the VIX index: it is converted to decimals, squared to change it from a volatility to a variance, then divided by 264 to convert to a daily basis. Since the VIX index reflects a 22-day implied volatility, I set $\tau = 22$.

Because of the apparent heteroskedasticity in IV_t demonstrated in Fig. 1, the regression error ε_t will be assumed be heteroskedastic as well, with a standard deviation proportional to V_t. We therefore have $\varepsilon_t \sim \text{i.i.d.N}(0, \xi^2 V_t^2)$, with ξ a parameter to be estimated. The magnitude of ξ will reflect the accuracy of the proxy of $\mathrm{E}^Q[\bar{V}]$ by IV.

Allowing this form of heteroskedasticity should mitigate the concern that the level effect in the volatility of volatility apparent in Fig. 1 is solely due to measurement error. If this is the case, then we may still find low values for γ_1 and γ_2.

3.3. Computing posteriors by data augmentation

Estimation of these models using standard methods is difficult for several reasons. First, the dynamics of S_t and V_t are represented as a diffusion process, making standard maximum likelihood or moment-based approaches impractical. The latency of stochastic variance further limits the effectiveness of these approaches. Lastly, the method used must simultaneously estimate the parameters of the SDE (4) or (5) and the regression equation (11).

---

[6] This was proved formally by Bergman et al. (1996).

[7] Christensen and Prabhala (1998) have argued recently that these types of measurement errors are severe enough to bias tests of the informational content of implied volatility. Additional sources of error could arise from simplifications made about the dividend process (Harvey and Whaley, 1991) or errors in the market's inference about the current state of market volatility.

The approach taken uses the data augmentation procedure of Jones (1998) and is similar to the methods of Eraker (2001) and Elerian et al. (2000), although the latter's method applies only to univariate processes. The method is based on the idea that the Euler approximation, a discrete time Gaussian model, can be used to approximate the transition density of a diffusion process. The Euler approximation has been used in many papers, and its use in likelihood-based inference has been justified by Pedersen (1995) and Brandt and Santa-Clara (2002). Under regularity conditions, the Euler approximation converges to the diffusion as the time step of the Euler process shrinks to zero.

It is sufficient for convergence that growth and Lipschitz conditions are satisfied, both of which the unconstrained model may violate. Results in Section C.10 suggest that this is unlikely to be a problem in the present context. Simulation results in Jones (2002) suggest that the Euler approximation is accurate even in more nonlinear models.

The Euler approximation for the 2GAM stochastic variance model is

$$\begin{aligned} S_{(k+1)h} &= S_{kh} + h\mu S_{kh} + \sqrt{h}\sqrt{V_{kh}} S_{kh} \eta_{k+1}^{(1)}, \\ V_{(k+1)h} &= V_{kh} + h(\alpha + \beta V_{kh}) + \sqrt{h}\sigma_1 V_{kh}^{\gamma_1} \eta_{k+1}^{(1)} + \sqrt{h}\sigma_2 V_{kh}^{\gamma_2} \eta_{k+1}^{(2)}, \end{aligned} \tag{12}$$

where $\eta_k^{(1)}$ and $\eta_k^{(2)}$ are independent standard normals and $k = 0, \ldots, K$.

In the estimation of the model, the stock price process S_t will be observed at the end of each trading day. Without loss of generality, assume that these end-of-day times correspond to $t \in \{0, 1, 2, \ldots, T\}$. A discrete time stochastic variance model may therefore be formulated by setting $h = 1$. One can see, however, that the 1-day return would then have a conditionally Gaussian distribution. To induce non-normality at this horizon, h must be small.

For convergence of the Euler approximation and for the generation of non-normality in conditional one-period returns, it is therefore imperative to allow h to be small in the estimation process, even if this means that the Euler approximation interval h is considerably smaller than the time interval at which the data are observed. This creates a problem for likelihood and moment-based inference, since neither the moments nor the transition densities of the Euler approximation are known over intervals larger than h. The strategy proposed by Pedersen (1995) and Brandt and Santa-Clara (2002) is to compute the likelihoods by simulation. In moment-based techniques such as Gallant and Tauchen's (1996) efficient method of moments, moments are computed by simulating long paths of the Euler approximation.

Since estimation of (12) essentially amounts to the analysis of a conditionally Gaussian model with a large number of missing data points, this paper will make use of Tanner and Wong's (1987) data augmentation algorithm in a manner similar to the discrete time stochastic volatility analysis of Jacquier et al. (1994). Although we do not see the augmented data, we often know their distribution conditional on the model parameters and the data that are observed. Data augmentation expands the parameter vector to include all augmented data in addition to the structural parameters (e.g. α and β) and then seeks to characterize the posterior distribution of this vector.

Since posteriors of such high dimension can seldom be derived analytically, Markov chain Monte Carlo (MCMC) methods are used, in which many random draws are generated from the posterior distribution rather than deriving the distribution analytically. Given enough draws from the posterior, sample moments, quantiles, and even densities are easy to calculate.

Rather than analyzing the entire parameter vector as a whole, the Gibbs sampler is used to break it down into a number pieces, each drawn conditional on the previous draws of all the others. Eventually, the Markov chain formed by this procedure converges to the true posterior.

More formally, suppose θ is a multidimensional random vector that can be partitioned into I subvectors $\{\theta^{(1)}, \theta^{(2)}, \theta^{(3)}, \ldots, \theta^{(I)}\}$ of possibly different sizes, and suppose the goal is to generate a sequence of random draws from the distribution $p(\theta|X)$, where X represents some data set. Choose some arbitrary initial value θ_0 and then iteratively form a chain of $\theta_n = \{\theta_n^{(1)}, \theta_n^{(2)}, \theta_n^{(3)}, \ldots, \theta_n^{(I)}\}$. For i between 1 and I, let $\theta_n^{(i)}$ be a random draw of $\theta^{(i)}$ from its conditional distribution $p(\theta^{(i)}|\theta_n^{(-i)}, X)$, where

$$\theta_n^{(-i)} = \{\theta_n^{(j)}, j < i\} \cup \{\theta_{n-1}^{(j)}, j > i\}. \tag{13}$$

The result of the Gibbs sampler is that under extremely mild conditions the Markov chain θ_n converges in distribution to $p(\theta|X)$.

For the stochastic variance process, the partitions $\theta^{(i)}$ will be defined as follows:

- $\theta^{(1)} = \{\mu, \alpha, \beta\}$,
- $\theta^{(2)} = \{\sigma_1, \sigma_2, \gamma_1, \gamma_2\}$,
- $\theta^{(3)} = \{A, B\}$,
- $\theta^{(4)} = \{\xi\}$,
- $\{\theta^{(5)}, \ldots, \theta^{T+5}\} = \{V_t, t = 0, 1, \ldots, T\}$,
- $\{\theta^{(T+6)}, \ldots, \theta^{I}\} = \{(S_t, V_t), t \notin \text{integers}\}$.

The data set X corresponds to the observations of the stock price S_t and the implied variance index IV_t for $t = 0, 1, \ldots, T$.

One can think about the procedure as "cycling" through the partitions $\theta^{(i)}$. Each cycle (n) begins with a draw of $\theta^{(1)}$, followed by draws of $\theta^{(2)}$, $\theta^{(3)}$, and $\theta^{(4)}$. The set of all end-of-day values of V_t are cycled through next. Lastly, the algorithm cycles through the pairs of (S_t, V_t) that are realized at times between the end-of-days. [8]

As opposed to the simulated maximum likelihood methods of Pedersen (1995) and Brandt and Santa-Clara (2002), which simulate the Euler approximation forward, the draws of high-frequency data here merely "bridge" the observed low frequency data

[8] Although the number of partitions $I = (T/h) + 6$ can be very large, often close to half of the I draws can be performed simultaneously, reducing computing time considerably. Because the Euler approximation is a Markov process, only adjacent observations ($t - h$ and $t + h$) are relevant for each draw of V_t or (S_t, V_t). We can therefore draw every other $\theta^{(i)}$ (for $i > 4$) simultaneously and alternate between just a few matrix draws rather than I individual draws. Describing the method as "cycling" through the I draws is for expositional convenience only.

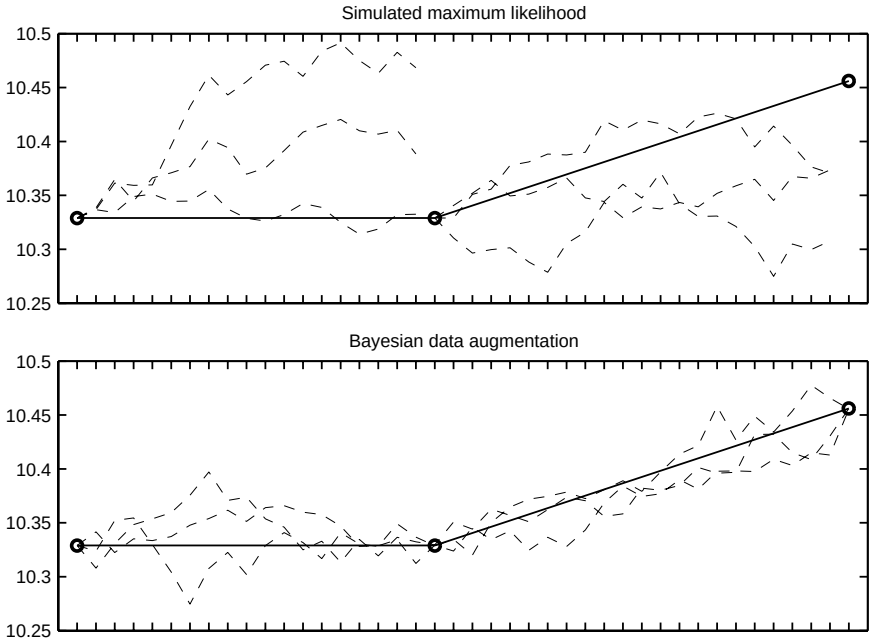

Fig. 2. Methods of simulation for likelihood-based inference of diffusions.

with short paths of high frequency data. Fig. 2 illustrates the difference between the two approaches. The advantage of the latter method is that the variance of the simulated data is smaller, since it must be consistent with both the start and end points. Since both methods must effectively integrate out the dependence on particular paths of simulated data, the smaller variance makes this integration less computationally demanding. This is particularly apparent when considering latent variable models, such as stochastic volatility, where simulated maximum likelihood is sometimes infeasible.

The methods for performing the draws of the structural parameters ($\theta^{(1)}$ to $\theta^{(4)}$) are contained in Appendix A, but the most important details can be described briefly. In drawing these parameters, virtually "flat" priors are used exclusively for all parameters except ξ, for which a small amount of prior information was found to be sometimes necessary. Each parameter is assumed independent of the others in the prior, with the joint prior given by

$$p(\mu, \alpha, \beta, \sigma_1, \sigma_2, \gamma_1, \gamma_2, \alpha^*, \beta, \xi) \propto \frac{1}{\sigma_2} f(\xi), \tag{14}$$

where $f(\xi)$ is the inverted gamma distribution with 15 degrees of freedom and gamma parameter 40/3.[9] The posterior distribution is computed to be proportional to the prior

[9] This inverted gamma distribution would result as the posterior distribution following the observation of 15 data points with a sample standard deviation of 0.1 with the diffuse prior $p(\xi) \propto \frac{1}{\xi}$. The sample sizes used in the analysis are all over 2,500, making $f(\xi)$ a relatively uninformative prior. With a truly diffuse prior, however, the algorithm occasionally approached a fixed point at $\xi = 0$.

times the complete likelihood, which includes both the product of transition densities and the likelihood of the initial observation (S_0, V_0). Since S_0 is non-stationary, its density is assumed to be diffuse, and so the joint density of (S_0, V_0) is replaced with the unconditional density of just V_0.

In drawing the augmented price and variance data, two algorithms are required. The first is applicable at non-integer t, which represent times at which neither S_t or V_t (nor implied volatility) is observed. Our goal for these t is therefore to draw S_t and V_t from

$$p(S_t, V_t|\theta^S, \mathbf{S}_{-t}, \mathbf{V}_{-t}, \mathbf{IV}), \tag{15}$$

where $\mathbf{W}_{-t}$ refers to all realizations of the time series W_s save the value at $s = t$ and where θ^S refers to the parameters of the stock price/variance process.

Since we are already conditioning on $\mathbf{V}_{-t}$, the additional conditioning on $\mathbf{IV}$ is irrelevant, since past and future values of IV just provide noisy signals about the contemporaneous values of V. The Markovian nature of the model further makes observations of S and V at times before $t-h$ and after $t+h$ irrelevant as well. The draw therefore reduces to the somewhat simpler

$$p(S_t, V_t|\theta^S, S_{t-h}, V_{t-h}, S_{t+h}, V_{t+h}). \tag{16}$$

Invoking Bayes rule, this can be shown to be proportional to

$$p(S_t, V_t|\theta^S, S_{t-h}, V_{t-h})\, p(S_{t+h}, V_{t+h}|\theta^S, S_t, V_t), \tag{17}$$

which is the product of a bivariate Gaussian density *for* (S_t, V_t) and a bivariate Gaussian density *conditional on* (S_t, V_t). Their product, unfortunately, is not a standard density for (S_t, V_t).

To draw the pair (S_t, V_t), a Metropolis–Hastings algorithm is proposed in which candidate draws are generated by $p(S_t, V_t|\theta^S, S_{t-h}, V_{t-h})$. The probability of accepting a candidate draw (S_t^*, V_t^*) over the previous draw (S_t, V_t) will therefore be equal to

$$\min\left\{\frac{p(S_{t+h}, V_{t+h}|\theta^S, S_t^*, V_t^*)}{p(S_{t+h}, V_{t+h}|\theta^S, S_t, V_t)},\ 1\right\}. \tag{18}$$

Thus, candidate draws are generated to be consistent with (S_{t-h}, V_{t-h}) and accepted based on there consistency with (S_{t+h}, V_{t+h}).

The second main type of data augmentation draw is for integer t, in which S_t is observed but V_t is not. A contemporaneous value of implied variance IV_t is observed as well. Slightly different from before, we would like to draw from

$$p(V_t|S_t, \theta^S, \mathbf{S}_{-t}, \mathbf{V}_{-t}, \mathbf{IV}). \tag{19}$$

In this case, the implied variance time series is informative, because IV_t provides information about V_t, and V_t is not contained in $\mathbf{V}_{-t}$. As before, however, past and future values of IV are irrelevant given past and future values of V. In addition, the Markovian nature of the model again makes observations of S and V at times before $t-h$ and after $t+h$ irrelevant. The density therefore simplifies to

$$p(V_t|\theta^S, S_{t-h}, V_{t-h}, S_t, S_{t+h}, V_{t+h}, IV_t), \tag{20}$$

which by Bayes rule (after eliminating irrelevant conditioning variables) is proportional to

$$p(V_t|\theta^S, S_{t-h}, V_{t-h}, S_t)\, p(IV_t|\theta^S, V_t)\, p(S_{t+h}, V_{t+h}|\theta^S, S_t, V_t). \tag{21}$$

Details of the Metropolis step used to draw V_t are relegated to Appendix B.

In practice, the acceptance rates of both Metropolis draws are typically around 50% for the bivariate draw of (S_t, V_t) and 67% for the univariate draw of V_t.

4. Results

Three versions of the model will be considered. The unrestricted version is the 2GAM model as it was described in (5). The first restricted version is the CEV model, in which the constraint $\gamma_1 = \gamma_2$ is imposed. The final version (the SQRT model) imposes the additional restriction that $\gamma_1 = \frac{1}{2}$, and is therefore identical to the square root model of Heston (1993). Both versions will be estimated over the full sample of January 1986 to June 2000 and the post-crash sample, which begins in July 1988. Six hundred thousand iterations of the estimation algorithm are made, with the first hundred thousand discarded to eliminate dependence on the initial conditions.

In this section I will begin with a summary of the posterior distributions of the model parameters, followed by an assortment of model diagnostics. The focus will then shift to the implications of the results for the distribution of returns, and I will examine both the higher moments and the options prices generated by each model.

In examining the implications of the models, I focus on parameter estimates obtained from the 1988–2000 sample. My intention is to examine whether the behavior of stock prices and implied volatilities over a period that excludes the stock market crashes of 1929 or 1987 offers any suggestion that such extreme events are possible. In other words, is there anything fundamentally different about the biggest market crashes, or are they simply extreme cases of more frequently observed phenomena?

4.1. Parameter posteriors

Table 1 presents summary statistics on the posterior distributions of the model parameters computed using $h = \frac{1}{10}$.[10]

The first two columns of the table contain posterior means and standard deviations of the parameters of the SQRT model. The second two columns report the same statistics for the CEV model, and the last two contain results for the 2GAM model. I report posterior statistics for the 10 parameters of the model as well as two simple transformations of these parameters. The first transformation, $\rho(0.0001)$, is the correlation between the price and variance processes when $V = 0.0001$, computed from (6).[11] This value of V corresponds to a daily volatility of 1%, a fairly typical day for the S&P

[10] Similar results are obtained by setting $h = \frac{1}{3}$, but results are noticeably different from those obtained using $h = 1$, suggesting that discretization bias is potentially a problem but that a relatively small amount of augmented data leads to the approximate convergence of posteriors.

[11] For the CEV and SQRT models, this correlation does not depend on V.

100 index. Second, I report $\sigma(0.0001)$, the volatility of instantaneous variance given $V = 0.0001$.

For all three models, results vary noticeably across the two samples. This demonstrates the large effect the crash of October 1987 has on the estimates, since the second sample only differs from the first in its exclusion of the months prior to July 1988. Differences are particularly large in the parameters that determine the volatility of instantaneous variance, namely σ_1, σ_2, γ_1, and γ_2. For the SQRT model, σ_1 and σ_2 are much larger in the sample that includes October 1987, attesting to the massive volatility spike that occurred at the time of the crash. These higher volatilities should generate much fatter tails and also alter the implied volatility smile. Differences in the parameters of the CEV and 2GAM diffusions are more difficult to interpret because of the interaction of the different variance parameters.

The first main empirical finding of the paper concerns the CEV model and is found in the two middle columns of Table 1. In both the whole sample and the post-July 1988 subsample, I find precise estimates of γ_1 that are greater than one, capturing the level dependence of the volatility of implied volatility that was evident in Fig. 1. A high variance elasticity it also found by Chacko and Viceira (2001) and by Poteshman (1998), the latter using a non-parametric approach to estimate a diffusion function that is convex, at least for low to moderate levels of V. The level dependence of the volatility of volatility that is implied by the CEV model is somewhat similar to the so-called "GARCH diffusion" model, in which $\gamma_1 = \gamma_2 = 1$, and to the EGARCH models proposed by Nelson (1991) and favored by Pagan and Schwert (1990). It is, in addition, a characteristic of the GARCH specification estimated by Hentschel (1995).

The second main result is that the additional generalization to the 2GAM model appears to be important. The last columns of Table 1 show that the γ_1 parameter is substantially higher than γ_2 for both sample periods. Since $\sigma_1 < 0$, this will have the effect of making the leverage effect stronger (or the correlation between the price and variance processes more negative) as the level of variance increases. In many other aspects, the CEV and 2GAM models are quite similar, with drift parameters that are extremely close to zero over both samples.

For both the CEV and 2GAM models, the volatility of instantaneous variance on a typical day, $\sigma(0.0001)$, is lower than it is for the SQRT model and is somewhat closer across the two sample periods. For all three models, strong negative correlations are typical between the price and variance process.

One potential concern is that β is essentially indistinguishable from zero for the CEV and 2GAM models. Since β determines the degree of mean reversion in the variance process drift, these results might lead one to conclude that the variance processes is explosive. Conley et al. (1997) show, however, that this conclusion is false, because stationarity will be generated through the diffusion term as long as either γ_1 or γ_2 are greater than one, regardless of the sign of β. These results are reviewed in Appendix C.

For the SQRT model, further parameter instability is evidenced in the estimates of β^*. Since this parameter is important in determining the relation between implied and instantaneous variances, this instability is somewhat consistent with the observation of Bates (2000) that the implied volatility smile of S&P 500 index options appeared to experience a regime switch at the time of the crash of 1987. In contrast, the CEV and

Table 1
Posterior means and standard deviations (in parentheses)

Parameter	SQRT 1986–2000	SQRT 1988–2000	CEV 1986–2000	CEV 1988–2000	2GAM 1986–2000	2GAM 1988–2000
μ	4.82×10^{-4} (1.52×10^{-4})	1.75×10^{-4} (1.19×10^{-4})	4.60×10^{-4} (1.42×10^{-4})	4.50×10^{-4} (1.46×10^{-4})	5.35×10^{-4} (1.37×10^{-4})	4.58×10^{-4} (1.40×10^{-4})
α	3.86×10^{-6} (0.04×10^{-6})	2.41×10^{-6} (0.07×10^{-6})	0.58×10^{-6} (0.09×10^{-6})	0.53×10^{-6} (0.10×10^{-6})	0.51×10^{-6} (0.10×10^{-6})	0.89×10^{-6} (0.11×10^{-6})
β	-1.47×10^{-2} (0.31×10^{-2})	-1.81×10^{-2} (0.22×10^{-2})	-0.03×10^{-2} (0.22×10^{-2})	0.05×10^{-2} (0.25×10^{-2})	-0.06×10^{-2} (0.23×10^{-2})	-0.48×10^{-2} (0.25×10^{-2})
σ_1	-1.99×10^{-3} (0.07×10^{-3})	-0.97×10^{-3} (0.03×10^{-3})	-1.76 (0.35)	-0.43 (0.10)	-8.16 (2.56)	-3.67 (0.89)
σ_2	2.63×10^{-3} (0.04×10^{-3})	1.05×10^{-3} (0.03×10^{-3})	1.74 (0.33)	0.38 (0.08)	7.81×10^{-1} (2.36×10^{-1})	0.18×10^{-1} (0.07×10^{-1})
γ_1	0.50 N/A	0.50 N/A	1.33 (0.02)	1.17 (0.02)	1.50 (0.03)	1.39 (0.03)
γ_2					1.24 (0.03)	0.84 (0.03)
β^*	-4.73×10^{-2} (0.14×10^{-2})	1.13×10^{-2} (0.19×10^{-2})	3.14×10^{-2} (0.16×10^{-2})	3.40×10^{-2} (0.22×10^{-2})	3.80×10^{-2} (0.27×10^{-2})	3.55×10^{-2} (0.31×10^{-2})
ξ	4.49×10^{-2} (0.16×10^{-2})	6.55×10^{-2} (0.27×10^{-2})	8.18×10^{-2} (0.25×10^{-2})	6.14×10^{-2} (0.42×10^{-2})	8.43×10^{-2} (0.35×10^{-2})	5.59×10^{-2} (0.31×10^{-2})
$\rho(0.0001)$[a]	-0.60 (0.01)	-0.68 (0.01)	-0.71 (0.01)	-0.75 (0.01)	-0.70 (0.01)	-0.78 (0.01)
$\sigma(0.0001)$[a]	3.30×10^{-5} (0.06×10^{-5})	1.43×10^{-5} (0.03×10^{-5})	1.14×10^{-5} (0.02×10^{-5})	1.22×10^{-5} (0.03×10^{-5})	1.13×10^{-5} (0.02×10^{-5})	1.27×10^{-5} (0.02×10^{-5})

[a] $\rho(0.0001)$ is the correlation of the price and volatility processes given an instantaneous variance of 0.0001, corresponding to a daily volatility of 1%. $\sigma(0.0001)$ is the volatility of instantaneous variance given the same condition.

2GAM models show relative stability for β^* across sample periods. As in the results of Pan (2002), this parameter is positive, implying a lack of mean reversion in the variance process drift. In the SQRT model, this causes the variance process to explode, while volatility-induced stationarity again prevents explosion in the CEV and 2GAM models.

4.2. Diagnostics

As in Eraker (2001), model diagnostics may be developed by examining the residuals defined implicitly by the Euler approximation of the SDE. At each iteration of the Markov chain, given the current draw of the parameter vector and the augmented data, these residuals can be evaluated. Under the SQRT model, for instance, these residuals are defined as

$$\eta_k^a = (S_k - S_{k-1} - h\mu S_{k-1})/(\sqrt{h}\sqrt{V_{k-1}}S_{k-1}),$$

$$\eta_k^b = (V_k - V_{k-1} - h(\alpha + \beta V_{k-1}))/(\sqrt{h}\sqrt{\sigma_1^2 + \sigma_2^2}\sqrt{V_{k-1}}). \tag{22}$$

Since η_k^a and η_k^b are assumed to be standard normals, correlated with each other but serially independent, the examination of low-order moments and autocorrelations is a convenient way to detect model misspecification. Additional tests make use of the residuals of the implied variance, or "VIX", Eq. (11), which implies that $(IV_t - A - BV_t)/(\xi V_t) \sim$ i.i.d. $N(0,1)$.

Following Zellner (1975), we may view these residuals as model parameters or as latent variables. Like other parameters, posterior distributions can be constructed for functions of these parameters simply by evaluating the function given the residuals computed at each step of the Markov chain. For the purpose of model diagnostics, these functions will consist of the sample mean, standard deviation, skewness, kurtosis, and two autocorrelations. The Euler approximation residuals are situated in time h days apart from another, so the "intra-day" autocorrelation is simply their first-order autocorrelation. For these residuals, the "daily" autocorrelation is therefore the $(1/h)$th order autocorrelation. The VIX equation produces only daily residuals.

Table 2 reports the posterior medians and 95% confidence intervals for these statistics over both sample periods. For the CEV and 2GAM models, residual means and standard deviations are indistinguishable from their theoretical values of zero and one, respectively. Some failures are observed for the SQRT model, however, most importantly in the standard deviation of the price equation residuals over the 1986–2000 sample. Since the standard deviation is less than one, residuals are on average were too small, indicating that the model-implied returns variance was often too high, most likely caused by an attempt to fit the crash of 1987.

The CEV and 2GAM models also perform well in matching the theoretical skewness and kurtosis of zero and three, respectively, for the price and variance equation residuals over both samples. The SQRT model, in contrast, fits only the post-crash sample, as its variance equation residuals over the 1986–2000 sample display substantial excess kurtosis. Autocorrelations in price and variance equation residuals are also smaller for the CEV and 2GAM models over both samples, although a small amount of negative autocorrelation is detectible in the CEV price equation residuals.

Where the CEV and 2GAM models fail is in describing the behavior of the VIX index. Over the 1986–2000 sample, residuals from the VIX equation display pronounced skewness and excess kurtosis as well as substantial positive residual autocorrelations. While the SQRT model displays problems here as well, they are not as severe. The results differ appreciably in the post-crash sample, in which the 2GAM model displays only marginal misspecification and the other two models show significant if modest deviations from normality.

To summarize, the SQRT model fails to describe neither the price, variance, nor VIX dynamics over the full sample period, but performs reasonably well in the post-crash sample. The CEV and 2GAM models appear to describe the actual price and variance dynamics well, even over the full 1986–2000 sample period, but are inconsistent with the dynamics of the VIX index. The 2GAM model offers moderate improvements over the CEV process on a number of measures.

One possible explanation for these results is that the price of variance risk is misspecified, with the true price of risk more time-varying than any of the models allow. In October of 1987 the VIX index peaked at 150%, equivalent to a one-month implied

Table 2
Posterior medians and 95% confidence intervals for model diagnostics[a]

	Mean	Standard deviation	Skewness	Kurtosis	Intra-day autocorrelation	Daily autocorrelation
Price equation, 1986–2000 sample						
SQRT	−0.004	0.975	−0.005	3.006	−0.047	0.000
	(−0.012, 0.004)	(0.967, 0.982)	(−0.030, 0.021)	(2.956, 3.060)	(−0.057, −0.037)	(−0.011, 0.010)
CEV	0.003	0.994	−0.006	3.016	−0.011	−0.001
	(−0.006, 0.013)	(0.987, 1.001)	(−0.032, 0.019)	(2.965, 3.069)	(−0.022, −0.002)	(−0.011, 0.009)
2GAM	0.001	0.997	−0.006	3.017	−0.005	−0.001
	(−0.008, 0.011)	(0.989, 1.005)	(−0.032, 0.019)	(2.967, 3.070)	(−0.016, 0.006)	(−0.011, 0.009)
Price equation, 1988–2000 sample						
SQRT	0.013	0.994	−0.003	3.014	−0.011	−0.001
	(0.004, 0.022)	(0.986, 1.002)	(−0.031, 0.025)	(2.957, 3.073)	(−0.022, 0.000)	(−0.012, 0.010)
CEV	0.004	0.993	−0.006	3.011	−0.012	−0.001
	(−0.007, 0.014)	(0.985, 1.001)	(−0.034, 0.022)	(2.955, 3.070)	(−0.024, −0.001)	(−0.012, 0.010)
2GAM	0.004	0.997	−0.006	3.012	−0.006	−0.001
	(−0.007, 0.014)	(0.988, 1.005)	(−0.034, 0.022)	(2.956, 3.070)	(−0.018, 0.006)	(−0.012, 0.010)
Variance equation, 1986–2000 sample						
SQRT	0.009	0.991	0.462	9.021	−0.014	0.006
	(0.001, 0.017)	(0.983, 0.998)	(0.413, 0.528)	(8.289, 10.102)	(−0.025, −0.004)	(−0.004, 0.017)
CEV	−0.001	0.997	0.022	3.033	−0.005	0.001
	(−0.011, 0.009)	(0.990, 1.004)	(−0.004, 0.047)	(2.981, 3.089)	(−0.016, 0.005)	(−0.009, 0.011)
2GAM	−0.001	0.999	0.019	3.030	−0.002	0.002
	(−0.011, 0.009)	(0.992, 1.006)	(−0.006, 0.045)	(2.978, 3.085)	(−0.012, 0.009)	(−0.008, 0.012)
Variance equation, 1988–2000 sample						
SQRT	−0.014	0.996	0.021	3.030	−0.008	0.002
	(−0.023, −0.005)	(0.987, 1.004)	(−0.007, 0.049)	(2.973, 3.090)	(−0.020, 0.003)	(−0.009, 0.013)
CEV	−0.003	0.996	0.018	3.028	−0.007	−0.004
	(−0.014, 0.008)	(0.988, 1.004)	(−0.010, 0.046)	(2.971, 3.089)	(−0.019, 0.004)	(−0.015, 0.007)
2GAM	−0.006	0.998	0.020	3.036	−0.003	−0.004
	(−0.017, 0.004)	(0.990, 1.006)	(−0.008, 0.048)	(2.978, 3.097)	(−0.015, 0.008)	(−0.016, 0.007)
VIX equation, 1986–2000 sample						
SQRT	0.003	0.991	0.210	8.428		0.064
	(−0.030, 0.036)	(0.968, 1.015)	(0.059, 0.734)	(7.226, 14.082)		(0.032, 0.095)
CEV	0.000	0.999	−1.335	21.188		0.152
	(−0.033, 0.033)	(0.976, 1.022)	(−1.539, −1.128)	(18.157, 24.285)		(0.116, 0.186)
2GAM	0.000	0.999	−1.365	24.053		0.141
	(−0.033, 0.033)	(0.976, 1.022)	(−1.582, −1.164)	(20.861, 27.802)		(0.111, 0.171)
VIX equation, 1988–2000 sample						
SQRT	0.000	0.996	0.111	3.599		0.047
	(−0.036, 0.037)	(0.971, 1.022)	(−0.009, 0.227)	(3.288, 4.081)		(0.011, 0.083)
CEV	0.000	0.995	−0.093	4.050		0.040
	(−0.036, 0.037)	(0.970, 1.021)	(−0.755, 0.102)	(3.133, 13.236)		(−0.002, 0.092)
2GAM	0.000	0.994	0.006	3.254		0.016
	(−0.036, 0.037)	(0.969, 1.020)	(−0.120, 0.119)	(2.994, 3.961)		(−0.020, 0.052)

[a]A correct specification implies that the residual mean is zero, the standard deviation is one, skewness is zero, kurtosis is three, and autocorrelations are zero.

volatility of 43%, roughly ten times the average monthly volatility of the S&P 100 index. The magnitude of this value suggests the that the difference between the objective and risk-neutral volatility forecasts may have been unusually large. Time variation in risk premia should be expected to produce both outliers and positive autocorrelations in the VIX residuals.

4.3. Dynamic behavior of implied volatility

To illustrate the kinds of implied volatility paths one might expect to see under each model, I graph several simulated trajectories of implied volatility. Each path of IV_t is simulated using the post-July 1988 posterior means as the true parameters and represents a time series of 3537 days, the length of the entire sample from January 1986 to June 2000.

Fig. 3 plots eight simulations performed under each model, where the implied volatilities are plotted on an annualized percentage basis. The SQRT simulations on the left side are fairly uniform, displaying relatively regular and moderate fluctuations. In the middle and right-hand columns, the simulations of the CEV and 2GAM models exhibit behavior more consistent with the actual data, plotted in Fig. 1. Long periods of low and non-volatile volatility are interrupted by occasional volatility spikes, which tend to die out quickly, especially for the 2GAM model. This is the same characteristic that Schwert (1990) noted of S&P 500 index volatilities around the crash of October 1987.

This fast decay following volatility spikes might be somewhat surprising since the values of β used in the CEV and 2GAM simulations are extremely close to zero, which in linear Gaussian models would correspond to autoregressive coefficients in excess of 0.995. For nonlinear models, however, Conley et al. (1997, p. 532) argue that focusing on the drift alone "abstracts from the role of uncertainty," and they demonstrate that mean reversion can be due to either the diffusion or the drift. Intuitively, when the variance elasticity is large, then large movements in volatility are likely when volatility is high. When the move is upward, the volatility of volatility increases even more, furthering the chance of another large move. But when these movements are downward, the volatility of the next movement in volatility is reduced dramatically, and the process tends to stay down, sometimes for very long periods.

4.4. The leverage effect

For the 2GAM model, Table 1 showed that $\gamma_1 > \gamma_2$ with extremely high posterior probability. Since $\sigma_1 < 0$, the effect of this is to make the correlation between price and variance processes inversely related to the level of instantaneous variance. This is clear from (6).

Fig. 4 shows this correlation as a function of instantaneous volatility. On a typical day, the volatility of the S&P 100 is just under 1% per day, and the figure shows that all models imply relatively similar correlations. It is when markets become turbulent that the correlation implied by the 2GAM model becomes even more negative. In the most extreme cases, when daily volatility exceeds 3%, this correlation is fairly close to negative one. Through this mechanism, extreme shocks to volatility are likely to be associated with large negative returns. This will substantially increase the skewness of returns under the 2GAM model relative to the CEV model, as will be shown next.

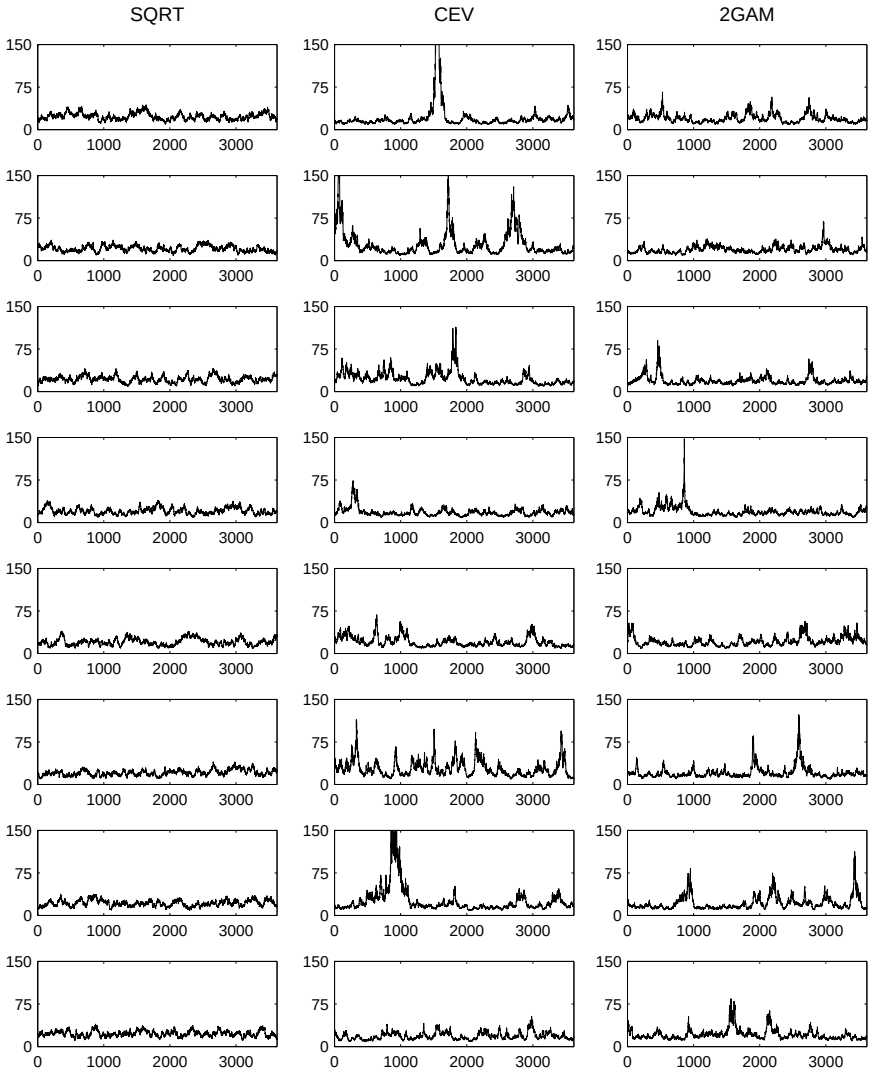

Fig. 3. Simulated volatility paths.

4.5. Unconditional skewness and kurtosis of returns and implied volatilities

The ability to generate realistic higher moments in stock returns is one of a stochastic volatility model's most crucial requirements. For several reasons, I focus not on the population moments implied by the various models, but on the distributions of sample moments generated by each model. These distributions supply a more informative description of the return characteristics implied by each model than do population moments since they address more directly what values of higher moments are likely to

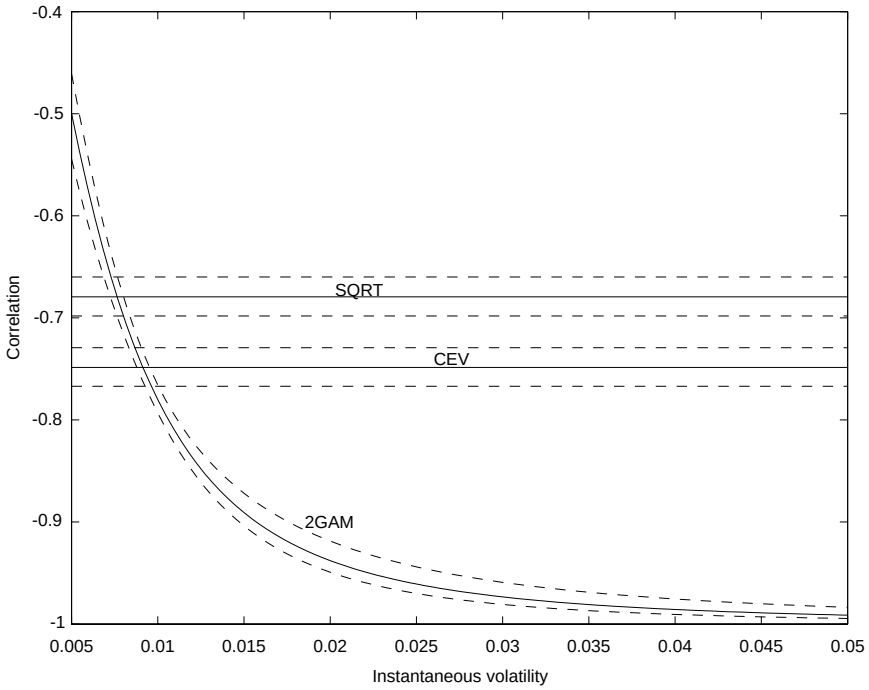

Fig. 4. Correlation between price and instantaneous variance processes (posterior means and 90% confidence intervals).

be found in a finite sample of data.[12] For this reason, the focus on sample moments offers a more parallel basis to compare observed and model-implied moments.

The focus on model-implied distributions of sample moments is also motivated by the fact that some model-implied population moments may not exist, as shown in Section C.6. For example, given the parameter estimates reported in Table 1, CEV model returns have finite mean and variance, but not skewness or kurtosis, since the variance process has no finite moments higher than the first. Distributions of sample moments remain well-defined and finite even when their probability limits are not, and they provide an intuitive means to measure the degree of non-normality that each model is capable of generating.

The sample moments of the levels and changes in the VIX index are also statistics that can be compared with their model-implied counterparts. Examination of these moments could reveal whether model misspecification is caused by unreasonable volatility dynamics or an incorrect characterization of the distribution of returns conditional on volatility (e.g. jumps or no jumps).

[12] For example, they will reveal whether higher moments are driven primarily by large returns occurring somewhat infrequently (a high median, low variance distribution of sample kurtosis) or by enormous returns occurring with minute probability (a low median, high variance, and right skewed distribution of sample kurtosis).

Table 3
Distribution percentiles of 13-year sample moments

		Returns		VIX Levels		VIX Changes	
		Skewness	Kurtosis	Skewness	Kurtosis	Skewness	Kurtosis
Actual sample moments							
Post-crash sample (1988–2000)		−0.299	8.058	1.102	4.975	0.795	15.426
Whole sample (1986–2000)		−2.230	49.218	4.627	54.696	17.201	990.817
Model-implied sample moment distributions							
SQRT	1st	−0.247	3.570	0.029	2.263	−0.099	3.025
	5th	−0.189	3.720	0.144	2.392	−0.067	3.096
	50th	−0.055	4.156	0.469	2.886	0.008	3.292
	95th	0.067	4.851	0.885	4.057	0.089	3.554
	99th	0.118	5.273	1.106	4.909	0.121	3.699
CEV	1st	−1.255	4.137	0.935	3.584	−2.892	4.791
	5th	−0.491	4.644	1.167	4.454	−1.071	5.731
	50th	0.012	8.779	2.276	10.074	−0.010	14.564
	95th	1.014	48.031	5.143	38.554	0.921	98.284
	99th	3.576	140.375	7.388	71.721	2.297	238.658
2GAM	1st	−2.384	3.868	0.892	3.732	−5.833	4.556
	5th	−0.702	4.177	1.125	4.473	−1.734	5.461
	50th	−0.097	6.205	2.184	10.305	−0.024	14.922
	95th	0.140	24.930	5.797	56.651	1.356	192.174
	99th	0.460	68.774	9.548	144.287	4.062	604.541
EJP-SVJ[a]	1st	−2.147	5.062	0.056	2.324	−0.139	3.272
	5th	−1.636	6.150	0.174	2.460	−0.101	3.383
	50th	−0.692	11.108	0.502	2.970	−0.015	3.704
	95th	−0.074	22.430	0.965	4.259	0.073	4.248
	99th	0.182	31.252	1.266	5.287	0.112	4.590

[a]The EJP-SVJ model is Eraker et al. (2003) model of stochastic volatility with jumps in the price process.

To account for the uncertainty of the parameter estimates, distributions of sample moments are computed using many different draws from the posterior distribution of the parameter vector. Specifically, I draw five thousand parameter vectors at random from each model's posterior distribution. For each parameter draw, I simulate a 13-year sample of returns and variance data, starting with a randomly drawn initial level of V. Given the level of instantaneous variance V at the end of each day, I draw a corresponding value of the VIX index from (11). For each sample I compute skewness and kurtosis for daily returns (simple, not logarithmic), VIX levels, and VIX changes. Various percentiles of these distributions are reported in Tables 3.

For comparison, I also include the results generated by Eraker et al. (2003, EJP) jump-diffusion ("SVJ") model, which combines the square root stochastic variance process with a Poisson jump. Because the full posterior distribution is not available for

this model, distributions of sample moments are computed using the point estimates reported in Table 3 of their paper. These estimates are calculated from daily S&P 500 returns from the 1980–1999 period, which is notable since it contains the 1987 crash. [13]

The simulated distribution for the SQRT, CEV, and 2GAM models represent a type of predictive distribution that is not conditioned on the terminal values of the sample data, since the initial value of volatility in each simulation is drawn from the unconditional distribution of the process rather than the filtered distribution of volatility on the last day of the sample. The simulations for EJP's SVJ model, since they are based on a single set of point estimates, generate what is commonly called a parametric bootstrap distribution.

Table 3 compares model-implied distributions of 13-year sample moments with actual sample moments calculated over the 1986–2000 and 1988–2000 periods. [14] It is immediately apparent that the 2GAM and CEV models typically generate far greater deviations from normality in returns than does the SQRT model. The table shows that for both sample periods the sample skewness and kurtosis lie beyond the 1st and 99th percentiles of the distributions implied by the SQRT model. The 1986–2000 sample moments are particularly hard to reconcile with the square root volatility process.

By adding jumps, the SVJ model comes close to capturing the negative skewness in returns observed around the 1987 crash. It is not as successful as the CEV or 2GAM models at generating the required level of kurtosis, however. This is despite the fact that the SVJ parameters are estimated on a sample of returns that includes the 1987 crash, while the other models are based on post-crash posterior distributions.

By allowing the leverage effect to increase in magnitude as the level of variance rises, the 2GAM model is able to produce substantially more negative skewness than the CEV model while producing more moderate levels of kurtosis. While the skewness falls short of being consistent with the 1986–2000 sample moments, it is on par with that generated by the model that includes jumps.

Additional insight is gained from looking at the sample and model-implied moments of the levels and daily changes in the VIX index. In both sample periods, both levels and changes were positively skewed with substantial excess kurtosis. The level of non-normality is extreme in the 1986–2000 sample, which is largely due to the increase from 36% to 150% in the VIX index that occurred on the day of the 1987 crash. While none of the models is capable of reproducing the skewness or kurtosis of VIX changes, the 2GAM model comes closest. Only the CEV and 2GAM models are successful in matching the higher moments in the levels of the VIX index.

[13] Simulations of the VIX index under the SVJ model are performed by numerically establishing the relation between V_t and the 22-day ATM call price, which is then converted into an implied volatility. Since no volatility or jump risk premia are estimated by EJP, these parameters are assumed to be zero. Finally, a proportional error of 6% (in line with the results for the CEV and 2GAM models from Table 1) is added to the simulated implied volatility.

[14] The average length of these two samples is about 13 years.

Table 4
Conditional moments

		Returns		VIX Levels	
		Skewness	Kurtosis	Skewness	Kurtosis
Moments conditional on $V_0 = 0.0001$					
SQRT	$\tau = 1$	−0.100	3.009	0.106	3.017
	$\tau = 5$	−0.238	3.105	0.155	2.978
	$\tau = 22$	−0.414	3.317	0.326	2.960
CEV	$\tau = 1$	−0.096	3.013	0.238	3.115
	$\tau = 5$	−0.240	3.145	0.588	3.654
	$\tau = 22$	−0.508	3.675	1.398	7.072
2GAM	$\tau = 1$	−0.107	3.020	0.261	3.135
	$\tau = 5$	−0.267	3.194	0.646	3.843
	$\tau = 22$	−0.568	3.954	1.786	11.898
EJP-SVJ	$\tau = 1$	−0.559	8.491	0.089	3.014
	$\tau = 5$	−0.304	4.117	0.162	3.027
	$\tau = 22$	−0.215	3.299	0.248	3.053
Moments conditional on $V_0 = 0.001$					
SQRT	$\tau = 1$	0.053	3.001	0.003	3.012
	$\tau = 5$	0.105	2.989	0.061	3.004
	$\tau = 22$	0.190	3.014	0.065	2.997
CEV	$\tau = 1$	−0.089	3.021	0.357	3.227
	$\tau = 5$	−0.235	3.133	0.874	4.508
	$\tau = 22$	−0.462	3.558	2.271	14.445
2GAM	$\tau = 1$	−0.241	3.126	0.655	3.853
	$\tau = 5$	−0.602	3.825	2.043	14.720
	$\tau = 22$	−1.049	5.252	5.350	105.352
EJP-SVJ[a]	$\tau = 1$	0.072	3.091	0.021	3.002
	$\tau = 5$	0.162	3.050	0.081	2.995
	$\tau = 22$	0.320	3.161	0.173	3.050

[a]The EJP-SVJ model is Eraker et al. (2003) model of stochastic volatility with jumps in the price process.

4.6. Conditional skewness and kurtosis of returns and implied volatilities

It is well-known (see Das and Sundaram, 1999, for example) that although stochastic volatility models may produce large unconditional deviations from normality, their conditional distributions often look relatively Gaussian. Table 4 presents the higher conditional moments of the same models considered previously, where the sole conditioning variable is the level of the instantaneous variance process V_t. One, 5- and 22-day time horizons are considered.

For simplicity, conditional population moments are calculated using the posterior mean parameter values taken from the 1988–2000 sample, and are computed from five hundred thousand independent simulated paths of prices and instantaneous variances. As before, Eraker et al. (2001) SVJ model uses parameter values estimated from S&P 500 returns from 1980–1999. Since the conditional moments of the levels and changes in the VIX index are nearly identical, I report only the results for the former.

Table 4 shows that conditional 1-day return distributions under the SQRT, CEV, and 2GAM models are very nearly Gaussian, in stark contrast to the unconditional distributions. Given a relatively typical initial variance of 0.0001 (a volatility of 1% per day), return distributions of all three models are only slightly left-skewed with almost no excess kurtosis. The SVJ model of Eraker et al. (2001) produces substantially greater deviations from normality at the 1-day horizon, with a conditional kurtosis almost as high as the median unconditional level from Table 3. At longer horizons, however, the ordering is reversed, with the three continuous models producing the greatest levels of kurtosis for 22-day returns. The 2GAM model, for instance, has a 22-day return kurtosis of 3.95, while the kurtosis of the SVJ model is just 3.30.

When volatility is unusually high ($V_0 = 0.001$, corresponding to a volatility of about 3.2% per day), the 2GAM model offers the highest conditional kurtosis at all time horizons. Furthermore, the two specifications based on the square root volatility process (the SQRT and SVJ models) imply positive skewness in the returns distribution. One interpretation is that because the square root variance diffusion function is concave in V_t, the volatility of the variance process decreases relative to its level, causing the randomness in the variance process to becomes less important as V_t increases. As a result, the process approximates more closely the lognormal model, implying the right skew in simple returns observed in Table 4. The SVJ model exhibits similar behavior because at high levels of diffusive volatility, the jump component becomes relatively unimportant.

Similarly to the unconditional moments in Table 3, the conditional distributions of the VIX process in Table 4 are substantially less Gaussian for the 2GAM and CEV processes. While intuition about the "right" levels of conditional skewness and kurtosis in volatility is less readily available than it is for returns, these moments perhaps most clearly distinguish the various specifications from one another.

4.7. Option prices

If the CEV and 2GAM models produce substantially greater non-normality in returns, at least over longer horizons, they should have an advantage over the SQRT model in explaining the volatility smile, or the tendency of options far from the money to sell at premiums relative to a Black–Scholes benchmark calculated using the implied volatility of an at-the-money option. I will therefore examine how the different models price puts and calls of different maturities with varying degrees of moneyness. I calculate these prices by simulation under the risk-neutral measure, again using the post-July 1988 posterior means as the parameter values for simulation. For simplicity, I assume that the interest and dividend rates are zero.

Table 5
Model-implied put and call prices[a]

		Puts				ATM	Calls			
	Strike	80	90	95	99	100	101	105	110	120
$\tau = 5$,	SQRT	0.0000	0.0001	0.0223	0.5252	0.9259	0.4960	0.0069	0.0000	0.0000
$V_0 = 0.0001$	CEV	0.0000	0.0001	0.0243	0.5286	0.9288	0.4988	0.0074	0.0000	0.0000
	2GAM	0.0000	0.0002	0.0268	0.5350	0.9339	0.5015	0.0072	0.0000	0.0000
$\tau = 5$,	SQRT	0.0030	0.2386	1.0023	2.3852	2.8651	2.4004	1.0623	0.2965	0.0094
$V_0 = 0.001$	CEV	0.0207	0.3530	1.1213	2.4337	2.8901	2.4030	0.9968	0.2306	0.0030
	2GAM	0.0484	0.4267	1.1658	2.4011	2.8348	2.3249	0.8669	0.1422	0.0002
$\tau = 22$,	SQRT	0.0029	0.1382	0.6391	1.7566	2.1917	1.6980	0.4515	0.0365	0.0000
$V_0 = 0.0001$	CEV	0.0131	0.2004	0.7204	1.8148	2.2400	1.7408	0.4619	0.0357	0.0000
	2GAM	0.0206	0.2320	0.7687	1.8639	2.2859	1.7771	0.4761	0.0367	0.0000
$\tau = 22$,	SQRT	0.6294	2.4101	4.0695	5.8383	6.3434	5.8735	4.2372	2.6998	0.9499
$V_0 = 0.001$	CEV	1.3255	3.0786	4.5571	6.1290	6.5822	6.0574	4.2280	2.5147	0.6906
	2GAM	1.2805	2.8082	4.1204	5.5521	5.9724	5.4191	3.4959	1.7684	0.2538
$\tau = 66$,	SQRT	0.4527	1.7044	2.9967	4.5010	4.9524	4.4323	2.6957	1.2541	0.1589
$V_0 = 0.0001$	CEV	1.1603	2.5670	3.7987	5.1711	5.5796	5.0198	3.0819	1.4078	0.1343
	2GAM	1.1338	2.5198	3.7481	5.1232	5.5332	4.9702	3.0402	1.3772	0.1254
$\tau = 66$,	SQRT	4.4168	7.9026	10.1267	12.1362	12.6699	12.2145	10.5199	8.6627	5.7329
$V_0 = 0.001$	CEV	6.7271	9.6490	11.4727	13.1332	13.5781	13.0387	10.9941	8.7315	5.1907
	2GAM	4.1383	6.6201	8.2791	9.8487	10.2778	9.7215	7.6608	5.4667	2.3888

[a]Assuming an initial stock price of \$100 and an interest rate of zero.

Each option price is calculated from one-hundred thousand simulations assuming an initial stock price of \$100. Two variance reduction techniques are used: antithetic variables and conditional simulations, whereby only V_t is simulated. The latter technique is developed by Willard (1997) and implemented recently by Poteshman (1998). For the 2GAM model, Willard's method is inapplicable since the price/variance correlation is time-varying, so I instead use the extension of Romano and Touzi (1997).

Table 5 shows the prices of puts and calls generated by each model for a variety of strike prices, times to maturity, and initial variance levels. To highlight the difference between the models, no in-the-money options were included. Standard errors of the put and call prices are not reported but are less than 2% of the option price for all options whose values exceed 0.0001.

Because the drift of the variance process under the risk-neutral measure is everywhere positive for all three models, option prices in Table 5 tend to be high. For the shorter term options, however, this effect is relatively minor. While all models generate comparable prices for options that are approximately at-the-money, substantial differences exist for options that are far from at-the-money. Deep out-of-the-money put options are valued higher by the CEV and 2GAM models than by the SQRT model. In some

cases these options are over five times more expensive under the CEV model and over 10 times more expensive under the 2GAM model. In contrast, the deep out-of-the-money call options are generally more expensive under the SQRT model, reflecting the less negative skewness of returns for that model.

For longer term options, the large positive drift in the CEV variance process implies very expensive options of all degrees of moneyness. The plausibility of these prices is undermined by the implication that the implied volatility term structure is unreasonably upward sloping. Given an initial level of variance equal to 0.0001 (a volatility of 1% per day), the 5-day, 22-day, and 66-day ATM implied volatilities are 17%, 19.5%, and 28%, respectively, for the CEV model. While implied volatility curves are often upward sloping, consistent upward slopes of this magnitude are unreasonable, as Pan (2002) argues. Misspecification in the assumed price of volatility risk is a possible explanation for these results.

4.8. Implied volatility smiles

To determine whether the implied volatility smiles generated by the CEV and SQRT models are consistent with options price data, I engage in a preliminary analysis using closing prices of S&P 100 equity options from July 1988 through August 1996. Black–Scholes implied volatilities are calculated using the method of Aït-Sahalia and Lo (1998). Since this method is only applicable to European-style options, the results here are open to criticism. It will be apparent below, however, that the early exercise premium would have to be extremely large to alter the results. In addition, the qualitative results are identical if S&P 500 options are used instead, and these options are European.

Using this method, simultaneously observed puts and calls with the same strike price and maturity are used to infer the implied index forward price. This circumvents the need for an accurate dividend forecast and a concurrent value for the true value of the index, but allows only one implied volatility to be extracted from every pair of options.

I define "at-the-money" as the equality of the strike price and the implied forward price. I discard all option prices that are not part of matched pairs as well as all option prices recorded on days without at least two matched pairs. On each day, at-the-money implied volatilities for each maturity are calculated by interpolating (or occasionally extrapolating) the corresponding volatility smiles on that day.

I compute the difference between each matched pair's implied volatility and the contemporaneous at-the-money implied volatility for the same maturity. This difference, the "smile magnitude," will be plotted against a normalized strike price, which is equal to the true strike price over the implied forward price minus one. Looking at the difference between two implied volatilities calculated from the same model should reduce the extent with which the differences between the models are driven simply by different levels of risk-neutral drift.

Although we should expect this measure of the volatility smile to depend on both the time to maturity of the option and the level of at-the-money implied volatility, I group data only by the former. Three maturity groups are considered: (a) options with

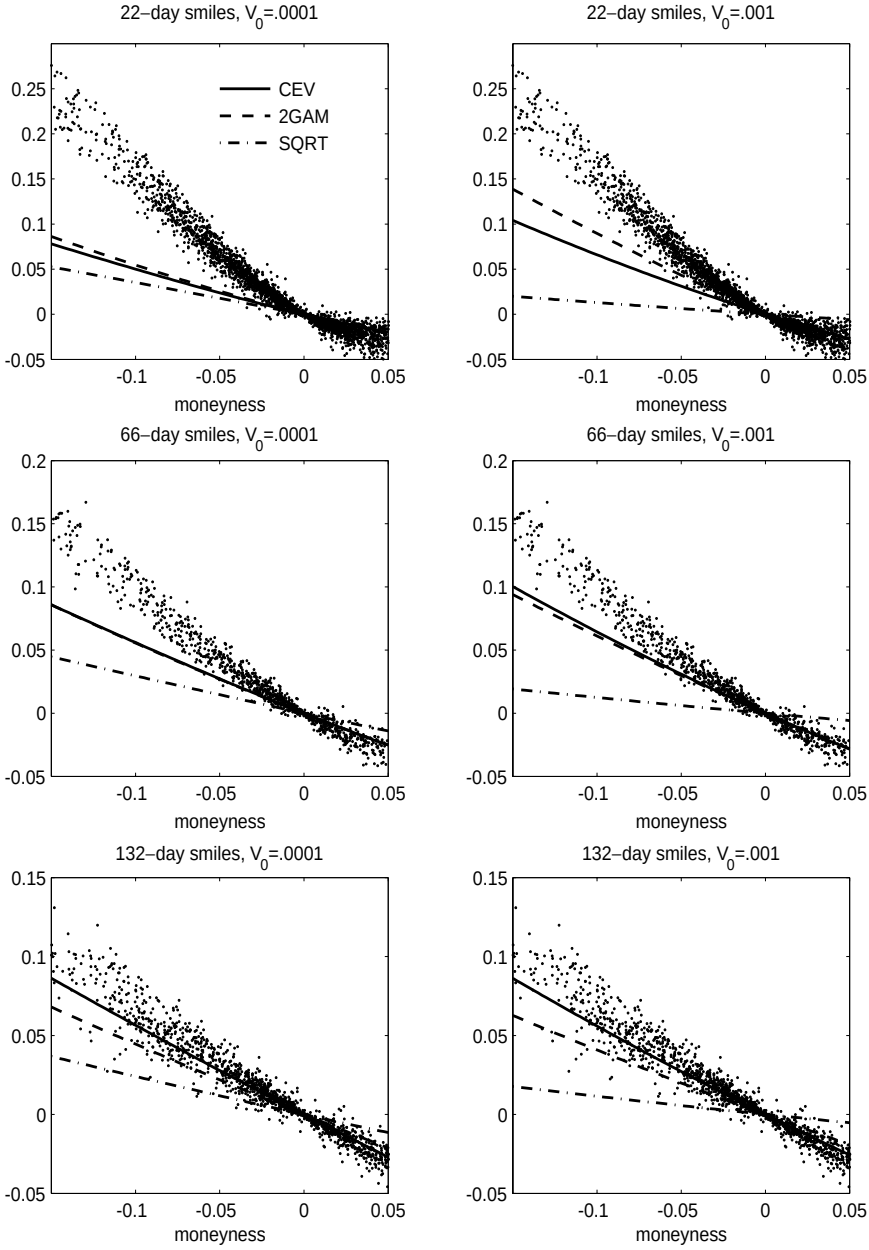

Fig. 5. Implied volatility smiles.

between 21 and 23 days to expiration, (b) options with between 64 and 68 days to expiration, and (c) options with between 120 and 144 days to expiration.[15]

The observed smile magnitudes (dots) are plotted in Fig. 5 along with the volatility smiles generated by the CEV (solid line), the 2GAM (dashed line), and SQRT (dot-dash line) models. Subplots on the left-hand side assume an initial level of instantaneous variance equal to 0.0001, while those on the right assume $V_0 = 0.001$. In general, the model-implied smiles for 22-day (1-month) options are far less severe than those observed in the data. For 3-month options, the fit of the CEV and 2GAM models is somewhat better, while for 6-month options the fit is quite good. In all cases, the SQRT model is clearly inadequate in explaining most in-the-money and out-of-the-money options prices.

5. Conclusion

As opposed to the GARCH literature on stock volatility, in which new models are primarily introduced to explain previously unnoticed features of price dynamics, the continuous-time literature has tended to favor models that generate closed form solutions for derivative prices. While such an emphasis is natural given the typical uses of these models, the inability of popular models to generate realistic price dynamics limits their usefulness.

The models introduced in this paper have obvious advantages over the square root model of Heston (1993) and others. While the square root model does produce non-Gaussian returns, the skewness and kurtosis it generates are far too small to be consistent with equity index returns. The unconditional skewness and kurtosis generated by the CEV and 2GAM models matched the sample moments remarkably well, even those moments calculated from a sample that included the crash of 1987. It is notable that the model-implied moments reported were implied by parameters estimated over a sample that did not include that crash. The results therefore suggest that the patterns of returns and volatility in relatively calm markets can be informative about less benign states as well. In other words, we may not need to see a crash to know the probability with which one will occur.

Unfortunately, while the option prices and volatility smiles generated by the CEV and 2GAM models are in many cases much more realistic that the prices implied by the square root model, they are still inconsistent with the volatility smile consistently observed in short-term equity index options. This suggests that while unconditional moments are well-described simply by modeling volatility as a more nonlinear process, conditional moments may not be. An analysis of conditional moments shows near-Gaussian behavior for all of the stochastic variance models, particularly at short horizons. Jumps, which substantially increase conditional non-normality, may therefore be an important component of returns, at least under the risk-neutral measure.

[15] In an earlier version of this paper, I examined the implied volatility smile of 5-day options as well. After observing that the bulk of the out-of-the-money option prices were no larger than a single tick, I excluded all very short-term options data from the paper.

The CEV and 2GAM models are also incapable of matching the time series behavior of the VIX index of implied volatility during the crash of 1987, although they offer tremendous improvement over models based on the square root variance process. One possible explanation is that the price of risk is misspecified. Another is that volatility itself might jump, an interesting idea introduced by Bates (2000) and Duffie et al. (2000) and carefully pursued by Eraker et al. (2003). A third is that volatility is better described by a two-factor model, which has been pursued by Alizadeh et al. (2001) and Chernov et al. (2003), among others.

Although option prices and conditional moments remain somewhat of a puzzle, many other aspects of the data are less perplexing. The results of this paper suggest a simple link between the observed intertemporal patterns in volatility and the existence of very rare and very large market crashes.

When volatility is low, it tends to stay low, sometimes over extremely long periods, since its drift is near zero and its own variance is small. When eventually there is a large positive shock to the volatility process, volatility itself rapidly becomes more volatile, and the probability of a larger subsequent rise in volatility increases dramatically. Successive positive shocks to volatility can therefore compound rapidly, and the eventual occurrence of such shocks makes the extreme return a rare but regular event. When the degree of negative correlation between prices and volatility increases, these extreme returns, more likely than not, are stock market crashes. Following the crash, these high volatility states tend to die out quickly because the mean reversion in volatility is itself volatility induced.

While there remain unanswered questions regarding the correct specification of conditional moments, risk premia, and options pricing, the models developed here show clear potential as simple yet much more realistic generalizations of existing models. With the growing popularity of derivatives written on realized volatilities rather than prices, future work using these derivatives' prices may provide a natural way to improve our understanding of volatility dynamics even further.

Acknowledgements

I thank David Bates, Greg Bauer, Mike Chernov, Mike Johannes, Nour Meddahi, Jun Pan, Neil Shephard, Ane Tamayo, and seminar participants at McGill, Rochester, the 2000 Meetings of the Western Finance Association, the 2000 World Congress of the Econometric Society, and the 2000 Duke-CIRANO Conference on Risk Neutral and Objective Probability Distributions for many helpful comments. I am especially grateful to Iulian Obreja for sharing his technical expertise, to the editor, George Tauchen, for many suggestions, and to an anonymous referee for an excellent report. All errors remain my own.

Appendix A. Drawing the parameters

In this appendix I describe draws of the four parameter blocks of $\theta=(\theta^{(1)},\theta^{(2)},\theta^{(3)},\theta^{(4)})$ given the entire set of high-frequency stock prices and variances, written in

shorthand as S and V. As an illustration, consider the draw of $\theta^{(3)}$. We would like to draw from the distribution $p(\theta^{(3)}|\theta^{(1)},\theta^{(2)},\theta^{(4)},S,V)$, which is, by Bayes rule, proportional to $p(S,V|\theta)p(\theta^{(3)})$, where $p(\theta^{(3)})$ is the prior on $\theta^{(3)}$. The likelihood function, $p(S,V|\theta)$, can be decomposed as the likelihood of the initial observation times the likelihood of all subsequent observations, or $p(S,V|\theta)=p(S_0,V_0|\theta)p(S,V|\theta,S_0,V_0)$. The second term here is the product of Gaussian transition probabilities, and were we to ignore the first non-Gaussian term many analytical results would be available about the conditional distributions of the parameters in $\theta^{(1)}$ to $\theta^{(4)}$.

In fact, the strategy in drawing parameters will be to ignore this first term and then make a correction for it later. Specifically, we generate a candidate set of parameter draws $\theta^{(1)}$ to $\theta^{(4)}$ from distributions such as $p(S,V|\theta,S_0,V_0)p(\theta^{(3)})$ and then accept or reject the collection of draws $\theta^{(1)}$ to $\theta^{(4)}$ depending on what they imply about the missing term, $p(S_0,V_0|\theta)$. Sections A.1–A.4 consider the draws of $\theta^{(1)}$ to $\theta^{(4)}$, while Section A.5 gives the criterion for acceptance.

A.1. Drawing $\theta^{(1)}=\{\mu,\alpha,\beta\}$

Given the parameters of the diffusion function, σ_1, σ_2, γ_1, and γ_2, and observation of the full price and variance paths, we can derive the conditional distribution of $\theta^{(1)}$ using the SUR methodology of Chib and Greenberg (1996). From the Euler approximation (22), the parameters μ, α, and β can be seen as the coefficients of a bivariate linear regression with a time-varying but known residual covariance matrix.

Inference about α is complicated, however, by the fact that additional information about α is also provided by the intercept of the regression equation for implied variance, $IV_t=A+BV_t+V_t v_t$, where $B=(1/\beta^*\tau)(e^{\beta^*\tau}-1)$, $A=-(\alpha/\beta^*)(1-B)$, and $v_t\sim N(0,\xi^2)$. Rearranging, we get

$$\frac{IV_t-BV_t}{V_t}=-\frac{1-B}{\beta^* V_t}\alpha+v_t. \tag{A.1}$$

Conditional on β^* and ξ and assuming a flat prior, this regression equation implies a normal distribution for α.

Now using this conditional posterior as a prior for α, and choosing extremely diffuse prior means and variances for μ and β, the method of Chib and Greenberg (1996) can be applied directly.

A.2. Drawing $\theta^{(2)}=\{\sigma_1,\sigma_2,\gamma_1,\gamma_2\}$

The parameters γ_1 and γ_2 enter the model in a nonlinear way, so drawing from their distribution presents an additional challenge. Because little is known about the distribution of γ_1 and γ_2, I will use the Metropolis–Hastings algorithm to generate their posterior. I will assume diffuse prior information about both, or $p(\gamma_1,\gamma_2)\propto 1$. As a candidate generating density I will use an independent bivariate random walk, so that the candidate draw is equal to the previous draw plus two independent Gaussian innovations with mean zero and standard deviation ζ_γ.

Efficiency will be improved if σ_1 and σ_2 are drawn jointly with γ_1 and γ_2, because the four parameters are highly correlated in the posterior. The candidate draws $\tilde{\sigma}_1$ and $\tilde{\sigma}_2$ will be conditioned on the candidate draw of $\tilde{\gamma}_1$ and $\tilde{\gamma}_2$. This conditional draw can be motivated by a rearrangement of the Euler approximation (12):

$$\frac{S_{(k+1)h} - (1 + h\mu)S_{kh}}{\sqrt{h}\sqrt{V_{kh}}S_{kh}} = \eta_{k+1}^{(1)},$$

$$\frac{V_{(k+1)h} - V_{kh} - h\alpha - hV_{kh}\beta}{\sqrt{h}V_{kh}^{\gamma_2}} = \sigma_1 V_{kh}^{\gamma_1-\gamma_2}\eta_{k+1}^{(1)} + \sigma_2\eta_{k+1}^{(2)}.$$

Note from the first equation that given μ and the time series of (S_k, V_k) the residual $\eta^{(1)}$ can be treated as an observed variable. The second equation can then be interpreted as a linear regression on an observed variable, $V_{kh}^{\gamma_1-\gamma_2}\eta_{k+1}^{(1)}$, with σ_1 as the regression coefficient and σ_2 as the standard deviation of the residual.[16]

With the flat prior $p(\sigma_1, \sigma_2) \propto 1/\sigma_2$, σ_2 has an inverted gamma distribution and σ_1 has a normal distribution given σ_2. This bivariate distribution is used to generate candidate pairs $(\tilde{\sigma}_1, \tilde{\sigma}_2)$ given the draw of γ_1 and γ_2. The complete candidate generating density is therefore given by this IG/normal density multiplied by the bivariate normal random walk density used to draw γ_1 and γ_2. Let $f_G(\sigma_1, \sigma_2, \gamma_1, \gamma_2)$ denote this density.

We can calculate the true conditional distribution of $(\sigma_1, \sigma_2, \gamma_1, \gamma_2)$ up to a constant of proportionality by an application of Bayes rule:

$$p(\sigma_1, \sigma_2, \gamma_1, \gamma_2 | \mu, \alpha, \beta, S, V, S_0, V_0)$$
$$\propto p(S, V | \mu, \alpha, \beta, \sigma_1, \sigma_2, \gamma_1, \gamma_2, S_0, V_0) p(\sigma_1, \sigma_2, \gamma_1, \gamma_2). \tag{A.2}$$

The first term on the right side is the completed likelihood function—the probability of generating the full set of high frequency S and V given the model parameters. This likelihood is calculated using the Euler approximation and is therefore the product of bivariate Gaussian transition densities. The second term is the prior, which is proportional to $1/\sigma_2$.

The Metropolis–Hastings algorithm dictates the probability with which the candidate draws must be accepted in order for convergence to the proper distribution to occur. The probability of moving from the previous draw $(\sigma_1, \sigma_2, \gamma_1, \gamma_2)$ to the candidate draw $(\tilde{\sigma}_1, \tilde{\sigma}_2, \tilde{\gamma}_1, \tilde{\gamma}_2)$ is

$$\min\left\{\frac{p(S, V | \mu, \alpha, \beta, \tilde{\sigma}_1, \tilde{\sigma}_2, \tilde{\gamma}_1, \tilde{\gamma}_2) p(\tilde{\sigma}_1, \tilde{\sigma}_2, \tilde{\gamma}_1, \tilde{\gamma}_2) f_G(\sigma_1, \sigma_2, \gamma_1, \gamma_2)}{p(S, V | \mu, \alpha, \beta, \sigma_1, \sigma_2, \gamma_1, \gamma_2) p(\sigma_1, \sigma_2, \gamma_1, \gamma_2) f_G(\tilde{\sigma}_1, \tilde{\sigma}_2, \tilde{\gamma}_1, \tilde{\gamma}_2)}, 1\right\}. \tag{A.3}$$

The resulting draws will converge to $p(\sigma_1, \sigma_2, \gamma_1, \gamma_2 | \mu, \alpha, \beta, S, V, S_0, V_0)$.

A.3. Drawing $\theta^{(3)} = \beta^*$

Again recall the assumptions about the relationship between the instantaneous and option-implied variances: $IV_t = A + BV_t + V_t v_t$, where $B = (1/\beta^*\tau)(e^{\beta^*\tau} - 1)$, $A = -(\alpha/\beta^*)(1 - B)$, and $v_t \sim N(0, \xi^2)$.

[16] This approach is borrowed from Jacquier et al. (2001).

Conditional on all parameters save β^*, the normality of v_t enables the computation of a likelihood function for β^*. This likelihood is non-standard, however, and the conditional posterior for β^* is of an unknown form. A value of β^* is therefore drawn using the griddy Gibbs sampler of Ritter and Tanner (1992), where the Grid consists of ten thousand equally-spaced points between -0.1 and 0.2.

A.4. Drawing $\theta^{(4)} = \xi$

Given A and B, a time series of homoskedastic regression residuals is defined by $v_t = (IV_t - A - BV_t)/V_t$. The parameter ξ is the standard deviation of v_t, so given any prior of the inverted gamma form the posterior of ξ is also an inverted gamma.

A.5. Incorporating the information in V_0

Suppose we have drawn a candidate set of parameters θ using the steps described in B.1–B.4. Then these parameters are jointly drawn from a density proportional to $p(S, V|\theta, S_0, V_0)\,p(\theta)$, which ignores the information in the initial observation (S_0, V_0). In order to take the full likelihood into account, we require the posterior be given by $p(S, V|\theta, S_0, V_0)\,p(S_0, V_0|\theta)\,p(\theta)$.

We therefore correct the previous draws of $\theta = (\theta^{(1)}, \theta^{(2)}, \theta^{(3)}, \theta^{(4)})$ by rejecting the parameter draws that are least likely to generate V_0 in their stationary distribution. Given the nonstationarity of S_t, its initial value can be assumed diffuse, containing no information about the model parameters. This accept/reject step is accomplished again using the Metropolis–Hastings algorithm, with the draws described in B.1–B.4 jointly used as the candidate generator. The Metropolis–Hastings algorithm therefore dictates that the set of new draws $(\tilde{\theta}^{(1)}, \tilde{\theta}^{(2)}, \tilde{\theta}^{(3)}, \tilde{\theta}^{(4)})$ is accepted over the old set of draws $(\theta^{(1)}, \theta^{(2)}, \theta^{(3)}, \theta^{(4)})$ with probability

$$\min\left\{\frac{p(V_0|\tilde{\theta}^{(1)}, \tilde{\theta}^{(2)}, \tilde{\theta}^{(3)}, \tilde{\theta}^{(4)})}{p(V_0|\theta^{(1)}, \theta^{(2)}, \theta^{(3)}, \theta^{(4)})}, 1\right\}. \tag{A.4}$$

If the new draw is rejected, we repeat the old one.

Appendix B. The distribution of high-frequency variance

For integer t both S_t and IV_t are observed, so draws of the latent variable V_t should come from

$$p(V_t|\theta^S, S_{t-h}, V_{t-h}, S_t)\,p(IV_t|\theta^S, V_t)\,p(S_{t+h}, V_{t+h}|\theta^S, S_t, V_t), \tag{B.1}$$

where the first term is proportional to a the conditional density of a bivariate normal random variable with mean

$$M = V_{t-h} + h(\alpha + \beta V_{t-h}) + \rho(V_{t-h})\frac{\sigma(V_{t-h})}{\sqrt{V_{t-h}}}(S_t - h(1+\mu)S_{t-h}) \tag{B.2}$$

and standard deviation is $\sqrt{h}\sigma(V_{t-h})\sqrt{1-\rho^2(V_{t-h})}$, where $\rho(V) = (\sigma_1 V^{\gamma_1})/\sqrt{\sigma_1^2 V^{2\gamma_1} + \sigma_2^2 V^{2\gamma_2}}$ and $\sigma(V) = \sqrt{\sigma_1^2 V^{2\gamma_1} + \sigma_2^2 V^{2\gamma_2}}$. The second term, $p(IV_t|\theta^S, V_t)$, is proportional to

$$\frac{1}{V_t}\exp\left(-\frac{1}{2}\left(\frac{IV_t - A - BV_t}{\xi V_t}\right)^2\right). \tag{B.3}$$

Again using the Metropolis–Hastings algorithm, a candidate draw of V_t is generated from the convolution of the first term, $p(V_t|\theta^S, S_{t-h}, V_{t-h}, S_t)$, and an approximation to the second term, $p(IV_t|\theta^S, V_t)$, that is given by

$$\begin{aligned} p_a(IV_t|\theta^S, V_t) &= \frac{1}{\bar{V}}\exp\left(-\frac{1}{2}\left(\frac{IV_t - A - BV_t}{\xi\bar{V}}\right)^2\right) \\ &\propto \exp\left(-\frac{1}{2}\left(\frac{V_t - (IV_t - A)/B}{\xi\bar{V}/B}\right)^2\right), \end{aligned} \tag{B.4}$$

where $\bar{V} = \frac{1}{2}(V_{t-h} + V_{t+h})$. The approximate density therefore represents a normal distribution for V_t, making the candidate generating density

$$p(V_t|\theta^S, S_{t-h}, V_{t-h}, S_t)\, p_a(IV_t|\theta^S, V_t) \tag{B.5}$$

normal as well. The Metropolis acceptance probability, or the probability of replacing the previous draw V_t with the candidate draw V_t^*, is

$$\min\left\{\frac{p(S_{t+h}, V_{t+h}|\theta^S, S_t, V_t^*)\, p(IV_t|\theta^S, V_t^*)\, p_a(IV_t|\theta^S, V_t)}{p(S_{t+h}, V_{t+h}|\theta^S, S_t, V_t)\, p(IV_t|\theta^S, V_t)\, p_a(IV_t|\theta^S, V_t^*)}, 1\right\}. \tag{B.6}$$

Appendix C. Some results for the CEV process with $\gamma > 1$

The relatively simple specification of the CEV process makes it possible to prove a number of results that are difficult to show for the 2GAM process. Given the parameter estimates in Table 1, I focus on the case in which $\gamma > 1$.

C.1. Definition of the scale measure

The scale measure is defined as

$$\int_m^n s(x)\,\mathrm{d}x, \tag{C.1}$$

where $s(x)$, the scale density, is

$$s(x) = \exp\left(-\int^x \frac{2\mu(y)}{\sigma^2(y)}\,\mathrm{d}y\right). \tag{C.2}$$

For the CEV process $dV = (\alpha + \beta V)\,dt + \sigma V^{\gamma}\,dB$, the scale measure is equal to

$$\int_m^n \exp\left(\frac{2\alpha}{\sigma^2(2\gamma-1)}\frac{1}{x^{2\gamma-1}} + \frac{2\beta}{\sigma^2(2\gamma-2)}\frac{1}{x^{2\gamma-2}}\right) dx. \tag{C.3}$$

C.2. Boundary behavior at $+\infty$

Infinity is not attainable as long as

$$\int_m^{+\infty} s(x)\,dx = +\infty \tag{C.4}$$

for arbitrary m (see Karlin and Taylor, 1981, pp. 230–231).

For the CEV process with $\gamma > 1$, the result is obtained regardless of the signs of α and β. We can observe that as $x \uparrow +\infty$,

$$\exp\left(\frac{2\alpha}{\sigma^2(2\gamma-1)}\frac{1}{x^{2\gamma-1}} + \frac{2\beta}{\sigma^2(2\gamma-2)}\frac{1}{x^{2\gamma-2}}\right) \to 1, \tag{C.5}$$

since $1/x^{2\gamma-1}$ and $1/x^{2\gamma-2}$ both converge to zero. The integral therefore diverges in this case as well.

C.3. Boundary behavior at **0**

The lower bound of 0 is not attainable if

$$\int_0^n s(x)\,dx = +\infty \tag{C.6}$$

for arbitrary n.

For the CEV process with $\gamma > 1$,

$$\exp\left(\frac{2\alpha}{\sigma^2(2\gamma-1)}\frac{1}{x^{2\gamma-1}} + \frac{2\beta}{\sigma^2(2\gamma-2)}\frac{1}{x^{2\gamma-2}}\right) \approx \exp\left(\frac{2\alpha}{\sigma^2(2\gamma-1)}\frac{1}{x^{2\gamma-1}}\right) \tag{C.7}$$

as $x \downarrow 0$, since the $1/x^{2\gamma-2}$ term diverges to $+\infty$ at a slower rate.

If $\alpha > 0$ and $\gamma > 1$, then $2\gamma - 1 > 1$ and $\delta = 2\alpha/\sigma^2(2\gamma-1) > 0$. This implies that

$$\exp\left(\frac{2\alpha}{\sigma^2(2\gamma-1)}\frac{1}{x^{2\gamma-1}}\right) = \exp\left(\frac{\delta}{x^{2\gamma-1}}\right) > \exp\left(\frac{\delta}{x}\right) > \frac{\delta}{x} \tag{C.8}$$

holds for small enough x. Since $1/x$ is not integrable at 0, (C.6) is satisfied.

C.4. Existence and uniqueness of the solutions

We have proven that as long as $\gamma > 1$, the sign of β is irrelevant for the attainability of 0 and $+\infty$ by the CEV variance process. Furthermore, the sign of α is irrelevant for the attainability of $+\infty$. For the zero drift case, identical results are proved by Andersen and Andreasen (1998).

Together, the unattainability of 0 and $+\infty$ is equivalent to Assumption 1 of Conley et al. (1997). They note that this assumption is a sufficient condition for the existence and uniqueness of the solution of the variance SDE, according to Theorem 5.7 of Karatzas and Shreve (1991, p. 335).

Given the existence of the solution to the variance SDE, the solution of the price SDE can also be shown to exist using a theorem of Revuz and Yor (1999, p. 375). Defining $\mathrm{d}M = \mu\,\mathrm{d}t + \sqrt{V}\,\mathrm{d}B$ (which is well-defined due to the square integrability of $\sqrt{V}$, proved below), the price process can be written as $\mathrm{d}S = f(t,S)\,\mathrm{d}M$, where $f(t,S) = S_t$. Since f satisfies a Lipschitz condition and is locally bounded, the solution of the price SDE exists and is unique.

C.5. Stationarity

Given the unattainability of 0 and $+\infty$, the stationary density, if it exists, will be proportional to $1/(s(x)\sigma^2(x))$ (see Karlin and Taylor, 1981, p. 241). Existence of the stationary distribution is ensured as long as

$$\int_0^{+\infty} \frac{1}{s(x)\sigma^2(x)}\,\mathrm{d}x < \infty \tag{C.9}$$

(see Karlin and Taylor, 1981, p. 221).

When $\gamma \neq 1$, the stationary density must take the form

$$C\,\frac{1}{\sigma^2 x^{2\gamma}} \exp\left(-\frac{2\alpha}{\sigma^2(2\gamma-1)}\,\frac{1}{x^{2\gamma-1}} - \frac{2\beta}{\sigma^2(2\gamma-2)}\,\frac{1}{x^{2\gamma-2}}\right), \tag{C.10}$$

where C is an integrating constant.

If $\gamma > 1$, then as $x \uparrow +\infty$ this density approaches $C/(\sigma^2 x^{2\gamma})$ since the exponential converges to 1. Since

$$\int_m^{+\infty} \frac{1}{x^{2\gamma}}\,\mathrm{d}x < \infty \tag{C.11}$$

for $m > 0$, the right tail of the density is integrable regardless of the signs of α and β.

As $x \downarrow 0$, the behavior of the exponential is determined solely by the term inside the exponential that diverges at the fastest rate. When $\gamma > 1$, this is the $1/x^{2\gamma-1}$ term, making the density asymptotically proportional to

$$\frac{1}{x^{2\gamma}} \exp\left(-\frac{\kappa}{x^{2\gamma-1}}\right), \tag{C.12}$$

where $\kappa = 2\alpha/\sigma^2(2\gamma-1) > 0$ as long as $\alpha > 0$.

Defining a change of variable $u = x^{1-2\gamma}$, we have

$$\int_0^m \frac{1}{x^{2\gamma}} \exp\left(-\frac{\kappa}{x^{2\gamma-1}}\right)\,\mathrm{d}x = \int_{m^{1-2\gamma}}^{+\infty} \frac{1}{1-2\gamma} \exp(-u)\,\mathrm{d}u, \tag{C.13}$$

which is finite.

Thus for $\gamma > 1$ and $\alpha > 0$, both tails of the density are integrable. Given the absence of singularities inside $(0, +\infty)$, the whole density is therefore integrable, implying the existence of the stationary distribution. Again, the sign of β is irrelevant as long as $\gamma > 1$.

Following Conley et al. (1997), the stationarity of a process is said to be "volatility-induced" when a deterministic process with the same drift would be explosive. Since the parameter region defined by $\gamma > 1$, $\alpha > 0$, and $\beta \geqslant 0$ implies stationary behavior even though the drift is everywhere positive, stationarity in this region is volatility-induced. These are the conditions given by Conley et al. (1997, p. 539) for the cases considered here. [17]

C.6. Existence of moments

For $\gamma > 1$, the variance process will have a finite kth moment whenever the integral

$$\int_0^{+\infty} x^k \frac{1}{\sigma^2 x^{2\gamma}} \exp\left(-\frac{2\alpha}{\sigma^2(2\gamma-1)} \frac{1}{x^{2\gamma-1}} - \frac{2\beta}{\sigma^2(2\gamma-2)} \frac{1}{x^{2\gamma-2}}\right) \mathrm{d}x \tag{C.14}$$

is finite. As long as $k > 0$, the x^k term reduces the integrand as $x \downarrow 0$ (which was already shown to converge as long as $\alpha > 0$), so the integral converges at 0. Thus, only the upper limit of the integral must be shown to converge for a positive moment to exist.

Since the exponential term converges to 1 as $x \uparrow +\infty$, convergence of the integral just requires that

$$\int_0^{+\infty} \frac{x^{k-2\gamma}}{\sigma^2} \mathrm{d}x \tag{C.15}$$

is finite. This will be the case whenever $k < 2\gamma - 1$. Therefore if $\gamma > 1$, then V has a finite unconditional first moment, while a finite second moment requires that $\gamma > 1.5$ and a finite "$1\frac{1}{2}$" moment requires that $\gamma > 1.25$. As long as $\gamma > 1$, the sign of β is unimportant.

C.7. Expected future variance

Aït-Sahalia (1996, p. 532) argues that $\mathrm{E}[V_{t+\Delta}|V_t] = a + bV_t$ under conditions apparently satisfied by the CEV process. His result, however, makes use of Dynkin's formula (see Oksendal, 2000, p. 118), which is not necessarily applicable if V_t is only a local martingale. It is possible to demonstrate, however, that the linear form likely provides an accurate approximation for values of Δ from 1 to 22 days, the horizons for which it will be used.

For the CEV process, posterior means of α are very close to zero, hence highly suggestive results are available from examining the case of $\alpha = 0$, for which the

[17] Note that the γ parameter used by CHLS is equivalent to 2γ in the present notation. In addition, identification issues related to subordination require CHLS to set σ equal to 1, without loss of generality.

transition density of the CEV process has been derived by Cox (1996) among others. In this case, we can compute the expectation of $V_{t+\Delta}$ both using the linear approximation and using the actual density function. For the parameter values in Table 1, the two approaches yield very similar answers. For a 5-day horizon they are virtually identical, while for the 22-day horizon the approximate expectations are slightly higher for high initial values of V_t.

When $\beta > 0$, it is obvious that the approximate expectation must be biased upward, because the stationary variance of V_t is finite while the limit as $\Delta \to \infty$ of $a + bV_t$ is not (since $b \to \infty$). This approximation error is likely to bias β^* towards slightly lower values, will which tend to offset the upward bias in $\mathrm{E}[V_{t+\Delta}|V_t]$.

Taking the approximation as given and integrating $\mathrm{E}[V_{t+\Delta}]$ from $\Delta = 0$ to τ, Fubini's theorem implies $\mathrm{E}[\mathrm{V}_{t,t+\tau}|V_t] = A + BV_t$, where $B = (1/\beta\tau)(\mathrm{e}^{\beta\tau} - 1)$ and $A = (\alpha/\beta)(1 - B)$.

C.8. The Novikov condition

It is assumed that the process under the objective measure follows $\mathrm{d}S = \mu S\,\mathrm{d}t + \sqrt{V}\,\mathrm{d}Z^{(1)}$ with $\mathrm{d}V = (\alpha + \beta V)\,\mathrm{d}t + \sigma V^{\gamma}\,\mathrm{d}Z^{(2)}$. Under the risk-neutral measure it satisfies $\mathrm{d}S = rS\,\mathrm{d}t + \sqrt{V}\,\mathrm{d}Z^{(1)*}$ with $\mathrm{d}V = (\alpha + \beta^* V)\,\mathrm{d}t + \sigma V^{\gamma}\,\mathrm{d}Z^{(2)*}$.

These dynamics implicitly define the prices of risk. These functions, $\lambda_1(V)$ and $\lambda_2(V)$, are the solutions to $\sqrt{V}\lambda_1(V) = \mu - r$ and $\sigma V^{\gamma}\lambda_2(V) = \beta V - \beta^* V$. Thus

$$\lambda_1(V) = \frac{\mu - r}{\sqrt{V}} \quad \text{and} \quad \lambda_2(V) = cV^{1-\gamma}, \tag{C.16}$$

where $c = (\beta - \beta^*)/\sigma$.

The implied Radon–Nikodym derivative is well-defined and markets are free from arbitrage opportunities if the Novikov condition is satisfied:

$$\mathrm{E}\left[\exp\left(\frac{1}{2}\int_0^T (\lambda_1^2(V_t) + \lambda_2^2(V_t))\,\mathrm{d}t\right)\right] < \infty. \tag{C.17}$$

To check the condition, we may derive an upper bound for the expectation in (49) using Jensen's inequality and Fubini's theorem. Since the exponential is convex,

$$\mathrm{E}\left[\exp\left(\frac{1}{2}\int_0^T (\lambda_1^2(V_t) + \lambda_2^2(V_t))\,\mathrm{d}t\right)\right] \leqslant \int_0^T \mathrm{E}\left[\exp\left(\frac{1}{2}(\lambda_1^2(V_t) + \lambda_2^2(V_t))\right)\right]\mathrm{d}t. \tag{C.18}$$

It therefore suffices to show the finiteness of

$$\begin{aligned}&\mathrm{E}\left[\exp\left(\frac{1}{2}(\lambda_1^2(V_t) + \lambda_2^2(V_t))\,\mathrm{d}t\right)\right]\\ &\quad = \mathrm{E}\left[\exp\left(\frac{1}{2}((\mu - r)^2 V_t^{-1} + c^2 V_t^{2-2\gamma})\right)\right].\end{aligned} \tag{C.19}$$

Using the stationary density derived above, this requires that

$$\int_0^{+\infty} \exp\left(\frac{1}{2}\left((\mu - r)^2 x^{-1} + c^2 x^{2-2\gamma}\right)\right) \tag{C.20}$$

$$\times \frac{1}{\sigma^2 x^{2\gamma}} \exp\left(-\frac{2\alpha}{\sigma^2(2\gamma - 1)} x^{1-2\gamma} - \frac{2\beta}{\sigma^2(2\gamma - 2)} x^{2-2\gamma}\right) \mathrm{d}x < \infty. \tag{C.21}$$

As long as $\gamma > 1$, $\exp(\frac{1}{2}((\mu - r)^2 x^{-1} + c^2 x^{2-2\gamma}))$ is decreasing in x. With the previous result that the stationary density is integrable at $+\infty$, the integral in (53) must also converge at $+\infty$.

As $x \downarrow 0$, the behavior of the exponential terms is determined solely by the term inside the exponentials that diverges at the fastest rate. As long as $\gamma > 1$, the x^{-1} and $x^{2-2\gamma}$ terms diverge more slowly than $x^{1-2\gamma}$. Following the previous proof of the existence of a stationary distribution, the integral in (53) therefore converges at 0. Since the integrand lacks other singularities, (53) is satisfied.

Thus as long as $\gamma > 1$ and $\alpha > 0$, the Radon–Nikodym derivative is well-defined even though the prices of risk are unbounded.

C.9. Derivative pricing

The absence of arbitrage implies an equivalent probability measure Q under which discounted security prices are local martingales. Without additional restrictions on the drift and diffusion functions (such as growth and Lipschitz conditions), discounted prices will not necessarily be martingales. It therefore does not follow that the time-t price of a security with payoff $f(S_{t+\tau})$ can always be represented as $\mathrm{E}_t^Q[\mathrm{e}^{-r\tau} f(S_{t+\tau})]$. Lewis (2000) notes, however, that in certain cases this representation still holds. For example, the absence of arbitrage implies that the price of a European put option be bounded by the strike price, K. Since the discounted put price is a local martingale that is bounded below by 0 and above by K, it must be a martingale (see Oksendal, 2000, p. 126). The put price can therefore be represented as $\mathrm{E}_t^Q[\mathrm{e}^{-r\tau} \max\{K - S_{t+\tau}, 0\}]$.

While it is straightforward to value a call option by appealing put-call parity, Lewis (2000, p. 285–286) also proves that the call price is given by $\mathrm{E}_t^Q[\mathrm{e}^{-r\tau} \max\{S_{t+\tau} - K, 0\}]$ as long as the variance process is non-explosive under the objective and risk-neutral measures (which we have shown) and that the correlation between the price and volatility processes is negative (which we find empirically).

C.10. Validity of the Euler approximation

While the convergence of the Euler approximation may not be established analytically, experimentation can provide some evidence of its validity for specific cases. Establishing the accuracy of the Euler approximation's implied transition densities is problematic, however, because the true transition densities of the CEV process are unknown as long as $\alpha \neq 0$. For both sample periods, however, the posterior mean of α is extremely close to zero, and the transition density of the CEV process for $\alpha = 0$ has been derived by Cox (1996) among others. It is therefore relevant and convenient

to look at the convergence of the restricted CEV process, $dV = \beta V\,dt + \sigma V^{\gamma}\,dB$, as the discretization parameter h goes to zero.

For a given h, ten million 1-day paths variance paths were simulated from initial values of either 0.0001 or 0.001. Since the correct transition density of the diffusion process is known, a cumulative distribution function, denoted as $\Phi_D(V)$, may be calculated by numerical integration. Now let V_i denote the terminal value of the ith simulation and Φ_N the CDF of the standard normal. Under the null hypothesis that the simulated distribution matches the analytical one,

$$X_i \equiv \Phi_N^{-1}(\Phi_D(V_i)) \sim N(0,1). \tag{C.22}$$

$E \equiv \Sigma X_i^2$ therefore has a chi-squared distribution under the null and may be interpreted as a summary statistic for the accuracy of the Euler approximation.

Given the large number of simulations, the limiting behavior of E can be characterized with great accuracy. For either level of initial variance, and using either sets of CEV parameter estimates reported in Table 1 (but with $\alpha = 0$), the deviation of E from its expected value of ten million declines by more than half each time h is halved until this deviation is within the 95% confidence interval implied by the chi-squared distribution. While statistically significant deviations are observed for values of h as small as 0.01, the effects on posteriors of choosing a much larger h are extremely minor. Posteriors are not noticeably different for $h = 0.33$ or $h = 0.1$, for instance, though they are obviously different from those obtained with $h = 1$.

In option pricing applications, it is sometimes necessary to simulate the variance process over significantly longer intervals. In these applications, extremely small values of h appear to be necessary for accurate results, making the approximation very slow and possibly unreliable. Much better performance is obtained by simulating the Euler approximation of the log volatility process. For the CEV model, the log volatility process becomes $d\ln V = ((\alpha/V) + \beta - \sigma^2 V^{2\gamma-2})\,dt + \sigma V^{\gamma-1}\,dB$. While the transformation induces nonlinearity in the drift, the convexity of the diffusion is greatly reduced. The result is that option prices computed by simulating the log process are not noticeably different for values of h smaller than 0.1.

A formal test of accuracy of the Euler approximation over long horizons can be constructed based on the steady-state density of the CEV process calculated in Section C.5. Letting $\Phi_S(V)$ denote the true steady-state CDF of $\ln V$, the convergence of the Euler approximation implies

$$Y_i \equiv \Phi_N^{-1}(\Phi_S(\ln V_i)) \sim N(0,1), \tag{C.23}$$

where $\ln V_i$ is the terminal value of a sufficiently long simulation of the log process. Using one million 100-year simulated paths, ΣY_i^2 is observed to quickly converge to its expected value as h goes to zero. The result appears robust to the choice of parameter values and is obtained for the 2GAM model as well.

For further reading

The following reference may also be of interest to the reader: Black and Scholes (1973).

References

Aït-Sahalia, Y., 1996. Nonparametric pricing of interest rate derivative securities. Econometrica 64, 527–560.

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

Alizadeh, S., Brandt, M.W., Diebold, F.X., 2001. Range-based estimation of stochastic volatility models. Journal of Finance 57, 1047–1091.

Andersen, L., Andreasen, J., 1998. Volatility skews and extensions of the Libor market model. Working paper.

Andersen, T.G., Bollerslev, T., 1997. Heterogeneous information arrivals and return volatility dynamics: uncovering the long-run in high-frequency returns. Journal of Finance 52, 975–1005.

Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P., 2001. Modeling and forecasting realized volatility. Working paper.

Andersen, T.G., Benzoni, L., Lund, J., 2002. An empirical investigation of continuous-time equity return models. Journal of Finance 57, 1239–1284.

Bakshi, G., Cao, C., Chen, Z., 1997. Empirical performance of alternative option pricing models. Journal of Finance 52, 2003–2049.

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

Ball, C.A., Torous, W.N., 1985. On jumps in common stock prices and their impact on call option pricing. Journal of Finance 40, 155–173.

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

Benzoni, L., 2001. Pricing options under stochastic volatility: an empirical investigation. Working paper.

Bergman, Y.Z., Grundy, G.D., Wiener, Z., 1996. General properties of option prices. Journal of Finance 51, 1573–1610.

Black, F., Scholes, M., 1973. The pricing of options and corporate liabilities. Journal of Political Economy 81, 637–659.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31, 307–327.

Brandt, M.W., Santa-Clara, P., 2002. Simulated likelihood estimation of diffusions with an application to exchange rate dynamics in incomplete markets. Journal of Financial Economics 63, 161–210.

Breeden, D., Litzenberger, R., 1978. Prices of state-contingent claims implicit in options prices. Journal of Business 51, 621–651.

Chacko, G., Viceira, L.M., 2001. Spectral GMM estimation of continuous-time processes. Working paper.

Chan, K.C., Karolyi, G.A., Longstaff, F.A., Sanders, A.B., 1992. An empirical comparison of alternative models of the short-term interest rate. Journal of Finance 47, 1209–1227.

Chernov, M., Ghysels, E., 2000a. A study towards a unified approach to the joint estimation of objective and risk neutral measures for the purpose of options valuation. Journal of Financial Economics 56, 407–458.

Chernov, M., Ghysels, E., 2000b. Estimation of stochastic volatility models for the purpose of options pricing. In: Abu-Mostafa, Y.S., LeBaron, B., Lo, A.W., Weigend, A.S. (Eds.), Computational Finance 1999. MIT Press, Cambridge, pp. 567–581.

Chernov, M., Gallant, A.R., Ghysels, E., Tauchen, G.E., 2003. Alternative models for stock price dynamics. Journal of Econometrics, this issue.

Chib, S., Greenberg, E., 1996. Markov chain Monte Carlo simulation methods in econometrics. Econometric Theory 12, 409–431.

Christensen, B.J., Prabhala, N.R., 1998. The relation between implied and realized volatility. Journal of Financial Economics 50, 125–150.

Conley, T.G., Hansen, L.P., Luttmer, E.G.J., Scheinkman, J.A., 1997. Short-term interest rates as subordinated diffusions. Review of Financial Studies 10, 525–577.

Cox, J.C., 1996. The constant elasticity of variance option pricing model. Journal of Portfolio Management, special December issue, 15–17.

Das, S.R., Sundaram, R.K., 1999. Of smiles and smirks: a term structure perspective. Journal of Financial and Quantitative Analysis 34, 211–239.

Duffie, D., Pan, J., Singleton, K., 2000. Transform analysis and asset pricing for affine jump-diffusions. Econometrica 68, 1343–1376.

Elerian, O., Chib, S., Shephard, N., 2000. Likelihood inference for discretely observed non-linear diffusions. Econometrica 69, 959–993.

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

Eraker, B., 2001. MCMC analysis of diffusion models with application to finance. Journal of Business and Economic Statistics 19, 177–191.

Eraker, B., Johannes, M., Polson, N., 2003. The impact of jumps in volatility and returns. Journal of Finance 59, forthcoming.

Gallant, A.R., Tauchen, G.E., 1996. Which moments to match? Journal of Econometric Theory 12, 657–681.

Gallant, A.R., Tauchen, G.E., 1997. Estimation of continuous-time models for stock returns and interest rates. Macroeconomic Dynamics 1, 137–168.

Gallant, A.R., Hsu, C.T., Tauchen, G.E., 1999. Using daily range data to calibrate volatility diffusions and extract the forward integrated variance. Review of Economics and Statistics 81, 617–631.

Harvey, C.R., Whaley, R.E., 1991. Dividends and S& P 100 index option valuation. Journal of Futures Markets 12, 123–137.

Hentschel, L., 1995. All in the family: nesting symmetric and asymmetric GARCH models. Journal of Financial Economics 39, 71–104.

Heston, S., 1993. A closed-form solution for options with stochastic volatility with applications to bond and currency options. Review of Financial Studies 6, 327–343.

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

Jacquier, E., Polson, N.G., Rossi, P.E., 1994. Bayesian analysis of stochastic volatility models. Journal of Business and Economic Statistics 12, 371–389.

Jacquier, E., Polson, N.G., Rossi, P.E., 2001. Bayesian analysis of stochastic volatility with leverage effect and fat tails. Journal of Econometrics, forthcoming.

Jarrow, R., Rudd, A., 1982. Approximate option valuation for arbitrary stochastic processes. Journal of Financial Economics 10, 347–369.

Jiang, G., van der Sluis, P., 1998. Pricing stock options under stochastic volatility and interest rates with efficient method of moments estimation. Working paper.

Jones, C.S., 1998. Bayesian estimation of continuous-time finance models. Working paper.

Jones, C.S., 2002. Nonlinear mean reversion in the short-term interest rate. Review of Financial Studies, forthcoming.

Jorion, P., 1989. On jump processes in the foreign exchange and stock markets. Review of Financial Studies 1, 427–445.

Karatzas, I., Shreve, S.E., 1991. Brownian Motion and Stochastic Calculus. Springer, Berlin.

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

Lewis, A.L., 2000. Option Valuation Under Stochastic Volatility. Finance Press, Newport Beach.

Nelson, D.B., 1991. Conditional heteroskedasticity in asset returns: a new approach. Econometrica 59, 347–370.

Oksendal, B., 2000. Stochastic Differential Equations: an Introduction with Applications. Springer, Berlin.

Pagan, A.R., Schwert, G.W., 1990. Alternative models for conditional stock volatility. Journal of Econometrics 45, 267–290.

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

Pedersen, A.R., 1995. A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. Scandinavian Journal of Statistics 22, 55–71.

Poteshman, A.M., 1998. Estimating a general stochastic variance model from option prices. Working paper.

Revuz, D., Yor, M., 1999. Continuous Martingales and Brownian Motion. Springer, Berlin.

Ritter, C., Tanner, M.A., 1992. The Gibbs stopper and the griddy Gibbs sampler. Journal of the American Statistical Association 87, 861–868.

Romano, M., Touzi, N., 1997. Contingent claims and market completeness in a stochastic volatility model. Mathematical Finance 7, 399–412.

Schwert, G.W., 1990. Stock volatility and the crash of '87. Review of Financial Studies 3, 77–102.

Scott, L.O., 1997. Pricing stock options in a jump-diffusion model with stochastic volatility and interest rates: applications of Fourier inversion methods. Mathematical Finance 7, 413–426.

Tanner, M.A., Wong, W.H., 1987. The calculation of posterior distributions by data augmentation. Journal of the American Statistical Association 82, 528–549.

Whaley, R.E., 1993. Derivatives on market volatility: hedging tools long overdue. Journal of Derivatives 1, 71–84.

Willard, G., 1997. Calculating prices and sensitivities for path-independent derivative securities in multifactor models. Journal of Derivatives 5, 45–61.

Zellner, A., 1975. Bayesian analysis of regression error terms. Journal of the American Statistical Association 70, 138–144.