

Maximum likelihood estimation of stochastic volatility models[☆]

Yacine Aït-Sahalia*, Robert Kimmel

Department of Economics and Bendheim Center for Finance, Princeton University, Princeton, NJ, 08540, USA

Received 8 June 2004; received in revised form 23 September 2005; accepted 10 October 2005

Available online 11 September 2006

Abstract

We develop and implement a method for maximum likelihood estimation in closed-form of stochastic volatility models. Using Monte Carlo simulations, we compare a full likelihood procedure, where an option price is inverted into the unobservable volatility state, to an approximate likelihood procedure where the volatility state is replaced by proxies based on the implied volatility of a short-dated at-the-money option. The approximation results in a small loss of accuracy relative to the standard errors due to sampling noise. We apply this method to market prices of index options for several stochastic volatility models, and compare the characteristics of the estimated models. The evidence for a general CEV model, which nests both the affine Heston model and a GARCH model, suggests that the elasticity of variance of volatility lies between that assumed by the two nested models.

© 2006 Elsevier B.V. All rights reserved.

JEL classifications: G12; C22

Keywords: Closed-form likelihood expansions; Volatility proxies; Heston model; GARCH model; CEV model

[☆]We are especially grateful to Bill Schwert (the Editor) and an anonymous referee for comments and suggestions that greatly improved the paper. Financial support from the NSF under grant SBR-0350772 is also gratefully acknowledged.

*Corresponding author.

E-mail address: yacine@princeton.edu (Y. Aït-Sahalia).

1. Introduction

In this paper, we develop and implement a technique for the estimation of stochastic volatility models of asset prices. In the early option pricing literature, such as Black and Scholes (1973) and Merton (1973), equity prices followed a univariate Markov process, usually a geometric Brownian motion. The instantaneous relative volatility of the equity price is then constant. Evidence from the time-series of equity returns against this type of model was noted at least as early as Black (1976), who commented on the fat tails of the returns distribution. Evidence from option prices also calls this type of model into question; if equity prices follow a geometric Brownian motion, the implied volatility of options should be constant through time, across strike prices, and across maturities. These predictions can easily be shown to be false; see, e.g., Stein (1989), Aït-Sahalia and Lo (1998), or Bakshi et al. (2000).¹ An alternative is offered by true stochastic volatility models, such as Stein and Stein (1991) or Heston (1993), in which innovations to volatility need not be perfectly correlated with innovations to the price of the underlying asset. Such models can explain some of the empirical features of the joint time-series behavior of stock and option prices, which cannot be captured by the more limited models.

However, estimating stochastic volatility models poses substantial challenges. One challenge is that the transition density of the state vector is hardly ever known in closed form for such models; some moments may or may not be known in closed form, depending on the model. Furthermore, the additional state variables that determine the level of volatility are not all directly observed. The estimation of stochastic volatility models when only the time-series of stock prices is observed is essentially a filtering problem, which requires the elimination of the unobservable variables.²

Alternately, the value of the additional state variables can be extracted from the observed prices of options. This extraction can be through a proxy used in place of the unobservable volatility, for instance the Black-Scholes implied volatility of an at-the-money short-maturity option, as in e.g., Ledoit et al. (2002), which can be further refined in the case of the specific CEV model: see Lewis (2000). Other simple approximation techniques (including one we propose below) are possible and broadly applicable.

A potentially more accurate procedure is to calculate option prices for a variety of levels of the volatility state variables, and use the observed option prices to infer the current levels of those state variables; see, e.g., Pan (2002). The first method has the virtue of simplicity, but is an approximation that does not permit identification of the market price of risk parameters for the volatility state variable; the second method is more complex,

¹One class of models that attempts to model equity prices more realistically takes the approach of having instantaneous volatility be time-varying and a function of the stock price. Models of this type include Derman and Kani (1994), Dupire (1994), and Rubinstein (1995). Such time-inhomogeneous models are often able to match an observed cross-section of option prices (across different strike prices and possibly also across maturities) perfectly. However, empirical studies such as Dumas et al. (1998) have found that they perform poorly in explaining the joint time series behavior of the stock and option prices.

²This can be achieved by computing an approximate discrete time density for the observable quantities by integrating out the latent variables as in e.g., Ruiz (1994), or the derivation of additional quantities such as conditional moments of the integrated volatility to be approximated by their discrete high frequency versions as in e.g., Bollerslev and Zhou (2002). For some specific models, typically those in the affine class, other relevant theoretical quantities, such as the characteristic function, as in Chacko and Viceira (2003), Jiang and Knight (2002), Singleton (2001), or the density derived numerically from the inverse characteristic function as in Bates (2002), can be calculated and matched to their empirical counterparts.

but allows full identification of all model parameters. Whichever method is used to extract the implied time-series observations of the state vector, subsequent estimation in practice has typically been simulation-based, relying either on Bayesian methods as in e.g., [Jacquier et al. \(1994\)](#), [Kim et al. \(1999\)](#) and [Eraker \(2001\)](#), or on the efficient method of moments of [Gallant and Tauchen \(1996\)](#), as in e.g., [Andersen and Lund \(1997\)](#).

In this paper, we develop a method that employs maximum likelihood, using closed form approximations to the true (but unknown) likelihood function of the joint observations on the underlying asset and either option prices (when the exact technique described above is used) or the volatility state variables themselves (when the approximation technique described above is used). The statistical efficiency of maximum likelihood is well-known, but in financial applications likelihood functions are often not known in closed form for the model of interest, since the state variables of the underlying continuous-time theoretical model are observed only at discrete time intervals. Our solution to this problem relies on the approach of [Aït-Sahalia \(2001\)](#), who develops series approximations to the likelihood function for arbitrary multivariate continuous-time diffusions at discrete intervals of observations; see [Aït-Sahalia \(2002\)](#) for the univariate theory. This method has been shown to be very accurate, even when the series are truncated after only a few terms, for a variety of diffusion models at least in the univariate case: see [Aït-Sahalia \(1999\)](#), [Jensen and Poulsen \(2002\)](#), [Stramer and Yan \(2005\)](#) and [Hurn et al. \(2005\)](#).

In all cases, we rely on observations on the joint time-series of the underlying asset price and either an option price or a proxy for the unobserved volatility extracted from the implied volatility of a short-dated at-the-money option. By comparing the results we obtain from the exact procedure (where the option pricing model is inverted to produce an estimate of the unobservable volatility state variable from the observed option price) to those of the approximate procedure (where the implied volatility from a short-dated at-the-money option is used to construct a proxy for the volatility state variable), we can assess the effect of that approximation. We find that the error introduced by the approximation is often smaller than the sampling noise inherent in the estimation of the parameters, so that using an implied volatility proxy does not have adverse consequences, other than not allowing the identification of the market prices of volatility risk, and results in a large computational efficiency gain. As to the specific proxy used, we find that the use of an unadjusted Black-Scholes implied volatility proxy for volatility introduces significant bias to some parameter estimates. For those cases, we propose a modified proxy that accounts for expected changes in volatility over the life of the option, and can be computed as a closed form adjustment to the Black-Scholes implied volatility. Even in those cases where the use of an unadjusted Black-Scholes implied volatility proxy does cause significant bias, the modified proxy largely eliminates this error.

The closed form feature of our approach offers considerable benefits: for example, estimation is quick enough that large numbers of Monte Carlo simulations can be run to test its accuracy, and we do so in this paper. For most other methods, large numbers of simulations are already required for a single estimation; simulating on top of simulations to run large numbers of Monte Carlos with these techniques is so time-consuming as to be practically infeasible, and we are not aware of evidence on their small-sample behavior. By contrast, we demonstrate that our technique is quite feasible for typical stochastic volatility models, even if option prices rather than implied volatilities are used. Evidence from the included Monte Carlo simulations shows that the sampling distribution of the estimates is

fairly well predicted by standard statistical asymptotic theory, as it applies to the maximum likelihood estimator.

We illustrate our method on actual data using several typical models, including the affine model of Heston (1993), a GARCH stochastic volatility model as in Nelson (1990) and Meddahi (2001), and a CEV model as in, e.g., Jones (2003). An early summary of some of the models we use as examples, as well as several others, can be found in Taylor (1994). However, it is also important to note that our method is applicable to arbitrary diffusion-based stochastic volatility models; the only requirement is that either the specification of the model be sufficiently tractable for option prices to be mapped into the state variables at a reasonable computational cost, or that a tractable proxy based on implied volatility be available.

The rest of this paper is organized as follows. In Section 2, we discuss a general class of stochastic volatility models for asset prices. Section 3 presents our estimation method in detail, showing how to apply it to the class of models of the previous section. In Section 4, we show how to apply this method to the three models cited above, developing the explicit closed form likelihood expressions, and extracting the state vector from option prices. In Section 5, we discuss different implied volatility proxies for the purpose of extracting the volatility state variable. Section 6 tests the accuracy of the method by performing Monte Carlo simulations for the model of Heston (1993) and the CEV model, assessing the accuracy of the estimates, the degree to which their sampling distributions conform to asymptotic theory, and the effect of using the implied volatility proxies. In Section 7, we apply this estimation method to real S&P 500 option data for the three stochastic volatility models, and analyze and compare the results. Section 8 concludes and sketches out an extension of the method to jump-diffusions. The appendix contains the closed form likelihood expansion for the three models under consideration.

2. Stochastic volatility models

We consider stochastic volatility models for asset prices and in this section briefly review them and establish our notation. Although we refer to the asset as a “stock” throughout, the models described can just as easily be applied to other classes of financial assets, such as foreign currencies or futures contracts. A stochastic volatility model for a stock price is one in which the price is a function of a vector of state variables X_t that follows a multivariate diffusion process:

$$dX_t = \mu^P(X_t) dt + \sigma(X_t) dW_t^P, \quad (1)$$

where X_t is an m -vector of state variables, W_t^P is an m -dimensional canonical Brownian motion under the objective probability measure P , $\mu^P(\cdot)$ is an m -dimensional function of X_t , and $\sigma(\cdot)$ is an $m \times m$ matrix-valued function of X_t . The stock price is given by $S_t = f(X_t)$ for some function $f(\cdot)$, but usually either the stock price or its natural logarithm is taken to be one of the state variables. We take the stock price itself to be the first element of X_t , and write $X_t = [S_t; Y_t]'$, with Y_t an N -vector of other state variables, $N = m - 1$.

From the well-known results of Harrison and Kreps (1979) and Harrison and Pliska (1981), and many extensions since then, the existence of an equivalent martingale measure Q guarantees the absence of arbitrage among a broad class of admissible trading

strategies.³ Under the measure Q , the state vector follows the process

$$dX_t = \mu^Q(X_t)dt + \sigma(X_t)dW_t^Q, \quad (2)$$

where W_t^Q is an m -dimensional canonical Brownian motion under Q , and $\mu^Q(\cdot)$ is an m -dimensional function of X_t . The stock itself, since it is a traded asset, must satisfy

$$dS_t = (r - \delta)S_t dt + \sigma_1(X_t)dW_t^Q, \quad (3)$$

where r is the risk-free rate, δ is the dividend yield on the stock (both taken to be constant for simplicity only), and $\sigma_1(X_t)$ denotes the first row of the matrix $\sigma(X_t)$. In other words, under the measure Q , an investment in the stock must have an instantaneous expected return equal to the risk-free interest rate. The instantaneous mean (under Q) of the stock price is therefore dependent only on the stock price itself, but its volatility can depend on any of the state variables including, but not limited to, S_t itself.

The price $\phi(t, X_t)$ of a derivative security that does not pay a dividend must satisfy the Feynman–Kac differential equation

$$\frac{\partial \phi(t, X_t)}{\partial t} + \sum_{i=1}^m \frac{\partial \phi(t, X_t)}{\partial X_t(i)} \mu_i^Q(X_t) + \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \frac{\partial^2 \phi(t, X_t)}{\partial X_t(i) \partial X_t(j)} \sigma_{ij}^2(X_t) - r\phi(t, X_t) = 0, \quad (4)$$

where $\mu_i^Q(X_t)$ denotes element i of the drift vector $\mu^Q(X_t)$, and $\sigma_{ij}^2(X_t)$ denotes the element in row i and column j of the diffusion matrix $\sigma(X_t)\sigma'(X_t)$. The price of an ordinary derivative security with a European-style exercise convention must satisfy the boundary condition

$$\phi(T, X_T) = g(X_T), \quad (5)$$

where T is the maturity date of the derivative and $g(X_T)$ is its final payoff. Usually, the derivative payoff is a function only of the stock price

$$g(X_T) = h(S_T) \quad (6)$$

for some function h ; for standard options, such as puts and calls, this condition is always satisfied.

The nature of a solution to Eq. (4) depends critically on the volatility specification in Eq. (3). If σ_1 satisfies

$$\sigma_1(X_t)\sigma_1'(X_t) = \sigma_S(S_t) \quad (7)$$

for some function $\sigma_S(S_t)$, then the stock price is a univariate process under the measure Q (although not necessarily under P because of the potential dependence of $\mu^P(X_t)$ on state variables other than S_t). In this case, the price of any European-style derivative with a final payoff of the type specified in Eq. (6) can be expressed as $\phi(t, X_t) = \xi(t, S_t)$, and Eq. (4) simplifies to

$$\frac{\partial \xi(t, S_t)}{\partial t} + \frac{\partial \xi(t, S_t)}{\partial S_t} (r - \delta)S_t + \frac{1}{2} \frac{\partial^2 \xi(t, S_t)}{\partial S_t^2} \sigma_S^2(S_t) - r\xi(t, S_t) = 0 \quad (8)$$

with the consequence that the instantaneous changes in prices of all derivative securities are perfectly correlated with the instantaneous price change of the stock itself. In this case,

³The definition of admissibility in the literature varies. It is usually either an integrability restriction on the trading strategy, which requires that the Radon–Nikodym derivative of Q with respect to P have finite variance, or a boundedness restriction on the deflated wealth process, which imposes no such restriction on dQ/dP .

knowledge of S_t and the parameters of the model are sufficient to price any derivative with final payoff of the type in Eq. (6); any additional state variables are either wholly irrelevant, or affect the stock price dynamics only under the measure P , and are therefore irrelevant for derivative pricing purposes. (Of course, if the application at hand is something other than derivative pricing, the dynamics under the P measure could be relevant.) Models of this type usually allow explicit time dependency by replacing $\sigma_S(S_t)$ with $\sigma_S(t, S_t)$; see, e.g., Derman and Kani (1994), Dupire (1994), and Rubinstein (1995), who develop univariate models (or, more precisely, discrete-time approximations to continuous-time univariate models) that have the ability to match an observed cross-section of option prices perfectly. Some of these techniques are also able to match observed prices of a term structure (with respect to maturity) of option prices as well. Such models are usually calibrated from the cross-section and possibly the term structure of option prices observed at a single point in time, rather than estimated from time-series observations of the stock price itself. Calibration methods specify dynamics under the measure Q only, leaving the dynamics under P unspecified. Such methods are therefore able to accurately reflect a number of empirical regularities, such as volatility smiles and smirks, but cannot tell us anything about risk premia of the state variables in the model (if our method is implemented using proxies derived from implied volatilities instead of option prices, the risk premia will not be fully identified either).

Despite this ability to match a cross-section, and often a term structure, of observed option prices perfectly, Dumas et al. (1998) find that univariate calibrated models imply a joint time-series behavior for the stock price and option prices that is not consistent with the observed price processes. Consequently, such models require periodic recalibration, in which the volatility function $\sigma_S(t, S_t)$ is changed to match the new observed cross-section and term structure of option prices. The need for such recalibration shows that the price process implied by such models cannot be the true price process, and the implications of such models with respect to derivatives pricing, hedging, etc., are therefore suspect. Stochastic volatility models, in which Eq. (7) is not satisfied, offer an alternative. Having the volatility of the stock depend on a set of state variables that can have variation independent of the stock price itself permits more flexible time-series modeling than is possible with the univariate calibrated type of model. Furthermore, stochastic volatility models are able to generate volatility smiles and smirks, although they are not able to match a cross-section of options perfectly, as are the calibrated models. Nonetheless, a stochastic volatility model with one or more elements in Y_t provides considerable flexibility in modeling. In all the specific models we consider in Sections 4 and 6, volatility depends on a single state variable (i.e., Y_t has a single element).

3. The estimation method

In stochastic volatility models, part of the state vector X_t is not directly observed. There are two fundamentally different approaches to dealing with this issue in estimation. One approach is to assume that we observe only a time-series of observations of the stock price S_t , and apply a filtering technique. The elements of X_t , other than S_t , are considered unobserved, and, since S_t is not a Markov process, the likelihood of an observation of S_t depends not only on the last observation S_{t-1} , but on the entire history of the stock price. Such an approach is taken by Bates (2002). This approach does not fully identify all of the parameters of the Q -measure dynamics. The model offers as many as m independent

sources of risk, but the stock price instantaneously depends on only one of these sources. Consequently, only the first element of $\mu^Q(\cdot)$ can be identified. If the dynamics under the measure P are the object of interest, then this approach has some advantages; for example, an incorrect specification of the Q -measure dynamics does not taint the P -measure estimation. However, if the Q -measure dynamics are the objective, then clearly another approach must be taken.

A second approach, which we adopt, is to assume that a time-series of observations of both the stock price S_t and a vector of option prices (which, for simplicity, we take to be call options) C_t is observed. The time-series of Y_t can then be inferred from the observed C_t . If Y_t is multidimensional, sufficiently many options are required with varying strike prices and maturities to allow extraction of the current value of Y_t from the observed stock and call prices. Otherwise, only a single option is needed. This approach has the advantage of using all available information in the estimation procedure, but the disadvantage that option prices must be calculated for each parameter vector considered, in order to extract the value of volatility from the call prices.

There are two distinct methods for extracting the value of Y_t from the observed option prices. One method is to calculate option prices explicitly as a function of the stock price and of Y_t , for each parameter vector considered during the estimation procedure. This approach has the advantage of permitting identification of all parameters under both the P and Q measures. An alternative is to use the Black-Scholes implied volatility of an at-the-money short-maturity option as a proxy for the instantaneous volatility of the stock. This approach has the virtue of simplicity, but can only be applied when there is a single stochastic volatility state variable. The Q -measure parameters are not fully identified when this method is employed. We use both of these approaches, and an additional method we develop, in Section 7 and compare them.

For reasons of statistical efficiency, we seek to determine the joint likelihood function of the observed data, as opposed to, for example, conditional or unconditional moments. We proceed as follows to determine this likelihood function. Since, in general, the transition likelihood function for a stochastic volatility model is not known in closed form, we apply the closed form approximation method of Aït-Sahalia (2001) which yields in closed form the joint likelihood function of $[S_t; Y_t]'$. From there, the joint likelihood function of the observations on $G_t = [S_t; C_t]'$ is obtained simply by multiplying the likelihood of $X_t = [S_t; Y_t]'$ by a Jacobian term. If a proxy for Y_t is used, this last step is not necessary.

We now examine each of these steps in turn: first, the determination of an explicit expression for the likelihood function of X_t ; second, the identification of the state vector X_t from the observed market data on G_t ; and third, a change of variable to go back from the likelihood function of X_t to that of G_t . We present in this section the method in full generality, before specializing and applying the results to the four specific stochastic volatility models we consider.

3.1. Closed-form likelihood expansions

The second step in our estimation method requires that we derive an explicit expression for the likelihood function of the state vector $X_t = [S_t; Y_t]'$ under P . Specifically, consider the stochastic differential equation describing the dynamics of the state vector X_t under the measure P , as specified by (1). Let $p_X(\Delta, x|x_0; \theta)$ denote its transition function, that is, the

conditional density of $X_{t+\Delta} = x$ given $X_t = x_0$, where θ denotes the vector of parameters for the model.

Rather than the likelihood function, we approximate the log-likelihood function, $l_X \equiv \ln p_X$. We now turn to the question of constructing closed form expansions for the function l_X of an arbitrary multivariate diffusion. The expansion of the log likelihood in Aït-Sahalia (2001) takes the form of a power series (with some additional leading terms) in Δ , the time interval separating observations:

$$l_X^{(J)}(\Delta, x|x_0; \theta) = -\frac{m}{2} \ln(2\pi\Delta) - D_v(x; \theta) + \frac{C_X^{(-1)}(x|x_0; \theta)}{\Delta} + \sum_{k=0}^J C_X^{(k)}(x|x_0; \theta) \frac{\Delta^k}{k!}, \quad (9)$$

where

$$D_v(x; \theta) \equiv \frac{1}{2} \ln(\det[v(x; \theta)]) \quad (10)$$

and $v(x) \equiv \sigma(x)\sigma'(x)$. The series can be calculated up to arbitrary order J . The unknowns so far are the coefficients $C_X^{(k)}$ corresponding to each Δ^k , $k = -1, 0, \dots, J$. We then calculate a Taylor series in $(x - x_0)$ of each coefficient $C_X^{(k)}$, at order j_k in $(x - x_0)$, which will turn out to be fully explicit. Such an expansion will be denoted by $C_X^{(j_k, k)}$, and is taken at order $j_k = 2(J - k)$.

The resulting expansion is then

$$\tilde{l}_X^{(J)}(\Delta, x|x_0; \theta) = -\frac{m}{2} \ln(2\pi\Delta) - D_v(x; \theta) + \frac{C_X^{(j_{-1}, -1)}(x|x_0; \theta)}{\Delta} + \sum_{k=0}^J C_X^{(j_k, k)}(x|x_0; \theta) \frac{\Delta^k}{k!} \quad (11)$$

and Aït-Sahalia (2001) shows that the coefficients $C_X^{(j_k, k)}$ can be obtained in closed form for arbitrary specifications of the dynamics of the state vector X_t by solving a system of linear equations.

The system of linear equations determining the coefficients is obtained by forcing the expansion (9) to satisfy, to order Δ^J , the forward and backward Fokker-Planck-Kolmogorov equations, either in their familiar form for the transition density p_X or in their equivalent form for $\ln p_X$. For instance, the forward equation for $\ln p_X$ is of the following form:

$$\begin{aligned} \frac{\partial l_X}{\partial \Delta} = & - \sum_{i=1}^m \frac{\partial \mu_i^P(x)}{\partial x_i} + \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \frac{\partial^2 v_{ij}(x)}{\partial x_i \partial x_j} - \sum_{i=1}^m \mu_i^P(x) \frac{\partial l_X}{\partial x_i} + \sum_{i=1}^m \sum_{j=1}^m \frac{\partial v_{ij}(x)}{\partial x_i} \frac{\partial l_X}{\partial x_j} \\ & + \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m v_{ij}(x) \frac{\partial^2 l_X}{\partial x_i \partial x_j} + \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \frac{\partial l_X}{\partial x_i} v_{ij}(x) \frac{\partial l_X}{\partial x_j}. \end{aligned} \quad (12)$$

In the appendix, we give the resulting coefficients $C_X^{(j_k, k)}$ in closed form for the stochastic volatility model of Heston (1993), and two other related stochastic volatility models. While the expressions at first look daunting, they are in fact quite simple to implement in practice. First, the calculations yielding the coefficients in formula (11) are performed using a symbolic algebra package such as *Mathematica*. Second, and most important, for a given

model, the expressions need to be calculated *only once*. So, to estimate, for instance, the model of Heston (1993) (or one of the other models considered), the expressions in the appendix are all that is needed for that model. The reader can then safely ignore the general method that gives rise to these expressions and simply plug the coefficients $C_X^{(j,k)}$ we give in the appendix into formula (11).

3.2. Identification of the state vector

When Y_t contains a single element, that is $N = 1$, one possible identification approach is to use the Black-Scholes implied volatility of an at-the-money short-maturity option as a proxy for the instantaneous relative standard deviation of the stock. From Eq. (3), the instantaneous relative volatility of the stock is given by $\sqrt{\sigma_0(X_t)\sigma'_0(X_t)}/S_t$. Since the stock price is observed and there is only one degree of freedom remaining in determining the instantaneous relative standard deviation, the stock and the implied volatility of a single option are sufficient to identify all elements of X_t . Such an approach is based on the theoretical observation that the implied volatility of an at-the-money option converges to the instantaneous volatility of the stock as the maturity of the option goes to zero. This approach has several advantages, but some disadvantages as well. First, it does not fully identify the Q -measure parameters. Second, this approach cannot be taken if Y_t has more than one element; in this case, multiple options are needed to identify the elements of Y_t , and simple approximation rules similar to that used for the univariate case are not available. Third, as we will see below, there are situations (such as the CEV model) where it is not sufficiently accurate.

If this approach is not possible or desirable, the elements of Y_t can be inferred from observed option prices C_t by calculating true (i.e., not dependent on the above approximation) option prices. Monte Carlo simulations in Section 6 below assess the effect of making this approximation on the overall quality of the estimates. Since the potential for simplification by using the approximation technique is substantial—in effect, rendering the option pricing model unnecessary—it is indeed worth investigating the tradeoff between the accuracy of the estimates and the effort involved in dealing with the option pricing model.

Clearly, to identify the N elements of Y_t requires observation of at least N option prices. If the mapping from the N elements of Y_t to prices of N options C_t with given strike prices and maturities has a unique inverse, then these options suffice to identify the state vector. If the inverse mapping is not unique, additional options are required, leading to a stochastic singularity problem. In this case, some or all of the options must be assumed to be observed with error. Whether the mapping from Y_t to the option prices is invertible must be verified for each specific model considered. In the specific models we use in our empirical application, $N = 1$ and this is not an issue.

For each time period in a data sample, we therefore need not only observations of the stock price S_t , but also at least N option prices of varying strikes and/or maturities. We denote the time of maturity and strike price of element i of C_t as T_i and K_i , respectively. The value of each element of C_t thus depends on the time to maturity $T_i - t$, the stock price S_t , the values of the other state variables Y_t , and the option strike price K_i ; these inputs form an $(N + 3)$ -dimensional space. As always, it is useful to reduce the dimensionality of the space of inputs as much as possible. We propose a number of approaches for achieving a low dimensionality, as follows. Holding $T_i - t$ constant for

each of the N options throughout the data sample reduces the dimensionality by one; we must then consider each of the N option inputs as occupying an $(N + 2)$ -dimensional space. We might be inclined to hold the strike price K_i constant throughout the data sample as well, although such a choice is usually not practical; if the stock price exhibits considerable variation over the data sample, it is unlikely that option prices with any fixed strike price K_i are observed in the market for the entire data sample. If, however, in addition to holding the time to maturity constant for each of the N options, we also hold moneyness (i.e., the ratio of S_t to K_i) constant, then the dimensionality of the input space is reduced to $N + 1$; each of the N options must be calculated for a variety of values of S_t and Y_t , but the time to maturity $T_i - t$ is held fixed for each option, and the strike price K_i is a simple function of the stock price for each option. In fact, option markets usually provide a reasonable range of moneyness traded at each point in time—introducing new options if necessary—thereby insuring that such data are always available. It should be noted, given these choices, that each $C_t(i)$ is not simply a time-series of observations of the same call throughout the data sample: the time to maturity remains constant, and the moneyness also remains constant even as the stock price changes through the sample.

A further reduction in dimensionality of the input space is possible if the stochastic volatility model satisfies a homogeneity property. Note that the payoff of a European call option is first-order homogeneous in the stock price and strike price. Denoting the call price C as a function of time of maturity, stock price, strike price, and Y_t , we have

$$C(T, \alpha S_T, \alpha K, Y_T) = (\alpha S_T - \alpha K)^+ = \alpha(S_T - K)^+ = \alpha C(T, S_T, K, Y_T). \quad (13)$$

In general, the price of an option is not first-order homogeneous prior to T , unless additional restrictions are placed on the model. The following conditions are sufficient:

$$\sigma_1(X_t)\sigma'_1(X_t) = \varphi_{11}(Y_t)S_t^2,$$

$$\sigma_1(X_t)\sigma'_i(X_t) = \varphi_{1i}(Y_t)S_t = \varphi_{i1}(Y_t)S_t, \quad i > 1,$$

$$\sigma_i(X_t)\sigma'_j(X_t) = \varphi_{ij}(Y_t), \quad i > 1, \quad j > 1,$$

$$\mu_i^Q(X_t) = \psi_i(Y_t), \quad i > 1 \quad (14)$$

for some set of functions $\varphi_{ij}(Y_t)$, $1 \leq i, j \leq m$, and $\psi_i(Y_t)$, $2 \leq i \leq m$. In this case, we can express the call price as

$$C(t, S_t, K, Y_t) = S_t H(t, m_t, Y_t), \quad (15)$$

where m_t is the logarithmic moneyness of the option:

$$m_t = \ln S_t - \ln K. \quad (16)$$

Substituting this expression into Eq. (4), we find that the pricing partial differential equation simplifies to

$$\begin{aligned}
 0 = & \frac{\partial H(t, m_t, Y_t)}{\partial t} - H(t, m_t, Y_t)\delta + \frac{\partial H(t, m_t, Y_t)}{\partial m_t}(r - \delta) \\
 & + \sum_{i=2}^m \frac{\partial H(t, m_t, Y_t)}{\partial Y_t(i)} \psi_i(Y_t) + \frac{1}{2} \left[\frac{\partial H(t, m_t, Y_t)}{\partial m_t} + \frac{\partial^2 H(t, m_t, Y_t)}{\partial m_t^2} \right] \varphi_{11}(Y_t) \\
 & + \sum_{i=2}^m \left[\frac{\partial H(t, m_t, Y_t)}{\partial Y_t(i)} + \frac{\partial^2 H(t, m_t, Y_t)}{\partial Y_t(i) \partial m_t} \right] \varphi_{i1}(Y_t) \\
 & + \frac{1}{2} \sum_{i=2}^m \sum_{j=2}^m \frac{\partial^2 H(t, m_t, Y_t)}{\partial Y_t(i) \partial Y_t(j)} \varphi_{ij}(Y_t). \quad (17)
 \end{aligned}$$

Note that the solution $H(t, m_t, Y_t)$ cannot depend on S_t , but this does not present a problem, since S_t has been eliminated from the coefficients of the partial differential equation. Furthermore, the strike price does not appear in the PDE or in the scaled option payoff:

$$H(T, m_T, Y_T) = (1 - e^{-m_T})^+. \quad (18)$$

The option price therefore inherits the homogeneity of its payoff. Thus, by calculating scaled option prices (i.e., option prices divided by the stock price), the dimensionality of the input space can be reduced to $m - 1$.

Provided the stochastic volatility model under consideration satisfies the homogeneity conditions of Eq. (14), scaled option prices with $m - 1$ distinct combinations of time to maturity $T_t - t$ and moneyness S_t/K_t must be calculated for varying values of Y_t . The time-series of values of Y_t can then be inferred by comparing the calculated option prices to the observed option prices. Once these values have been calculated for a given value of the parameter vector, the joint likelihood of the time-series of observations of S_t and Y_t must be calculated.

A variety of techniques exist for calculating option prices, and the most appropriate method in general depends on the specific stochastic volatility model under question. For instance, if the characteristic function of the transition likelihood is known in closed form (as is sometimes the case even when the likelihood itself is not known), options can be priced through a variety of Fourier transform methods.

A natural alternative is to apply the density expansion technique described in Section 3.1 to the risk-neutral dynamics (2), yielding at order J the transition function $q_X^{(J)}(\Delta, x|x_0; \theta)$, or, after expanding the state vector $X = [S; Y]'$, $q_X^{(J)}(\Delta, s, y|s_0, y_0; \theta)$. Denote the corresponding marginal with respect to s as $q_X^{(J)}(\Delta, s|s_0, y_0; \theta)$. Then an expansion of the call option price at order J is given by the Feynman–Kac formula

$$C^{(J)}(t, S_t, K, Y_t) = \exp(-r(T - t)) \int_0^{+\infty} (S_T - K)^+ q_X^{(J)}(T - t, S_T|S_t, Y_t; \theta) dS_T. \quad (19)$$

Since the specification of the market prices of risk is often such that the dynamics of the process under P and Q involve only adjustments to the parameter values, but no change of functional form for the drift and diffusion, the expression for $q_X^{(J)}$ follows directly from those already derived for $p_X^{(J)}$. If there is a change of functional form of the drift vector and/or the diffusion matrix due to the market prices of risk, then the same method described in

Section 3.1 can be applied to the new specification to obtain $q_X^{(J)}$. And note that the expression (19) is truly a one-dimensional integral because there is no need to obtain $q_X^{(J)}(\Delta, s|s_0, y_0; \theta)$ as the integral over the forward volatility variable of $q_X^{(J)}(\Delta, s, y|s_0, y_0; \theta)$; rather, integration over y of the backward PDE satisfied by $q_X^{(J)}(\Delta, s, y|s_0, y_0; \theta)$ shows that $q_X^{(J)}(\Delta, s|s_0, y_0; \theta)$ solves the same equation. A closed form expansion for $q_X^{(J)}(\Delta, s|s_0, y_0; \theta)$ is therefore obtained directly by solving the backward equation using the same functional form, with the coefficients obtained explicitly by constraining the solution to also satisfy the forward equation.

3.3. Change of variables: from state to observed variables

We have now obtained an expansion of the joint likelihood of observations on $X_t = [S_t; Y_t]'$ in the form (11). If a proxy based on an option's implied volatility has been used to identify Y_t , then this likelihood can be maximized directly; provided the instantaneous interest rate and dividend yield are observed rather than estimated, then the identification of Y_t does not depend in any way on the model parameters. The value of X_t therefore remains constant as the model parameters are varied during a likelihood search. When the true option prices are calculated, this is no longer the case; as the model parameters are varied during a likelihood search, the implied values of X_t do not remain constant. Estimation by maximization of the likelihood of X_t is therefore not possible; rather, estimation requires maximization of the likelihood of the observed market prices, $G_t = [S_t; C_t]'$.

The third and last step of our method is therefore moving from X_t to the time-series observations on G_t , and this step requires only that the likelihood of X_t be multiplied by a Jacobian term. This term is a function of the partial derivatives of the X_t with respect to S_t and C_t ; these derivatives are arranged in a matrix, and the Jacobian term is the determinant of this matrix. Because S_t is itself an element of X_t , the determinant takes on a particularly simple form:

$$\begin{aligned}
 J_t = \det \begin{bmatrix} \frac{\partial S_t}{\partial S_t} & \frac{\partial S_t}{\partial Y_t(1)} & \cdots & \frac{\partial S_t}{\partial Y_t(N)} \\ \frac{\partial C_t(1)}{\partial S_t} & \frac{\partial C_t(1)}{\partial Y_t(1)} & \cdots & \frac{\partial C_t(1)}{\partial Y_t(N)} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial C_t(N)}{\partial S_t} & \frac{\partial C_t(N)}{\partial Y_t(1)} & \cdots & \frac{\partial C_t(N)}{\partial Y_t(N)} \end{bmatrix} &= \det \begin{bmatrix} 1 & 0 & \cdots & 0 \\ \frac{\partial C_t(1)}{\partial S_t} & \frac{\partial C_t(1)}{\partial Y_t(1)} & \cdots & \frac{\partial C_t(1)}{\partial Y_t(N)} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial C_t(N)}{\partial S_t} & \frac{\partial C_t(N)}{\partial Y_t(1)} & \cdots & \frac{\partial C_t(N)}{\partial Y_t(N)} \end{bmatrix} \\
 &= \det \begin{bmatrix} \frac{\partial C_t(1)}{\partial Y_t(1)} & \cdots & \frac{\partial C_t(1)}{\partial Y_t(N)} \\ \vdots & \ddots & \vdots \\ \frac{\partial C_t(N)}{\partial Y_t(1)} & \cdots & \frac{\partial C_t(N)}{\partial Y_t(N)} \end{bmatrix}. \tag{20}
 \end{aligned}$$

It is therefore only necessary to calculate partial derivatives of the option prices C_t with respect to the state variables Y_t ; these derivatives are the stochastic multivariate analog of

the familiar vega of Black-Scholes option prices. The delta coefficients of the option prices do not appear in the Jacobian term. When we calculate the option prices to identify the state vector X_t (as per Section 3.2), the derivatives are also calculated as a by-product.

Once the state vector is identified and the Jacobian term from the change of variables formula computed, the transition function of the observed asset prices (the stock and options), $G_t = [S_t; C_t]'$, can be derived from the transition function of the state vector $X_t = [S_t; Y_t]'$. Specifically, consider the stochastic differential equation describing the dynamics of the state vector X_t under the measure P , as specified by (1). Let $p_X(\Delta, x|x_0; \theta)$ denote its transition function, that is the conditional density of $X_{t+\Delta} = x$ given $X_t = x_0$, where θ denotes the vector of parameters for the model. Let $p_G(\Delta, g|g_0; \theta)$ similarly denote the transition function of the vector of the asset prices G observed Δ units apart.

We now express the stock and option prices as functions of the state vector, $G_{t+\Delta} = f(X_{t+\Delta}; \theta)$. Defining the inverse of this function to express the state as a function of the observed asset prices, $X_{t+\Delta} = f^{-1}(G_{t+\Delta}; \theta)$, we have the following for the conditional density of $G_{t+\Delta} = g$ given $G_t = g_0$:

$$\begin{aligned} p_G(\Delta, g|g_0; \theta) &= \det \left(\frac{\partial f(f^{-1}(g; \theta))}{\partial x} \right)^{-1} p_X(\Delta, f^{-1}(g; \theta) | f^{-1}(g_0; \theta); \theta) \\ &= J_t(\Delta, g|g_0; \theta)^{-1} p_X(\Delta, f^{-1}(g; \theta) | f^{-1}(g_0; \theta); \theta), \end{aligned} \quad (21)$$

where $J_t(\Delta, g|g_0; \theta)$ is the determinant defined in (20).

Then, recognizing that the vector of asset prices is Markovian and applying Bayes' Rule, we see that the log-likelihood function for discrete data on the asset prices vector g_t sampled at dates t_0, t_1, \dots, t_n has the simple form

$$\ell_n(\theta) \equiv n^{-1} \sum_{i=1}^n l_G(t_i - t_{i-1}, g_{t_i} | g_{t_{i-1}}; \theta), \quad (22)$$

where

$$l_G(\Delta, g|g_0; \theta) \equiv \ln p_G(\Delta, g|g_0; \theta) = -\ln J_t(\Delta, g|g_0; \theta) + l_X(\Delta, f^{-1}(g; \theta) | f^{-1}(g_0; \theta); \theta)$$

with l_X obtained in Section 3.1, and we are done.

We assume in this paper that the sampling process is deterministic. Indeed, in typical practical situations, and in our Monte Carlo experiments below, these types of models are estimated on the basis of daily or weekly data, so that $t_i - t_{i-1} = \Delta = 7/365$ or $t_i - t_{i-1} = \Delta = 1/252$ is a fixed number. Ait-Sahalia and Mykland (2003) provide a treatment of maximum likelihood estimation in the case of randomly spaced sampling times. Maximum likelihood estimation of the parameter vector θ then involves maximizing expression (22), evaluated at the observations $g_{t_0}, g_{t_1}, \dots, g_{t_n}$ over the parameter values.

4. Example: the Heston model

In what follows, we apply the method described above to the prototypical stochastic volatility model, that of Heston (1993). Under the Q measure, S_t and Y_t follow the dynamics

$$dX_t = d \begin{bmatrix} S_t \\ Y_t \end{bmatrix} = \begin{bmatrix} (r-d)S_t \\ \kappa'(\gamma' - Y_t) \end{bmatrix} dt + \begin{bmatrix} \sqrt{(1-\rho^2)Y_t}S_t & \rho\sqrt{Y_t}S_t \\ 0 & \sigma\sqrt{Y_t} \end{bmatrix} d \begin{bmatrix} W_1^Q(t) \\ W_2^Q(t) \end{bmatrix}. \quad (23)$$

Note that Y_t is a local variance rather than a local standard deviation; while keeping this in mind, we will continue to refer to Y_t as the stochastic volatility variable. Y_t follows the square root process of Feller (1951), and is bounded below by zero. The boundary value zero cannot be achieved if Feller's condition, $2\kappa'\gamma' \geq \sigma^2$, is satisfied. If we instead restate the dynamics in terms of the logarithmic stock price $s_t = \ln S_t$, we have

$$d \begin{bmatrix} s_t \\ Y_t \end{bmatrix} = \begin{bmatrix} r - d - \frac{1}{2} Y_t \\ \kappa'(\gamma' - Y_t) \end{bmatrix} dt + \begin{bmatrix} \sqrt{(1 - \rho^2)Y_t} & \rho\sqrt{Y_t} \\ 0 & \sigma\sqrt{Y_t} \end{bmatrix} d \begin{bmatrix} W_1^Q(t) \\ W_2^Q(t) \end{bmatrix}. \quad (24)$$

The log stock price s_t has volatility that is an affine function of Y_t , and the covariance between s_t and Y_t is also affine in Y_t itself. The model of Black and Scholes (1973) is obviously a special case of the model of Heston (1993), in which $\sigma = 0$ and $Y_0 = \gamma'$ so that Y_t is constant. The likelihood function for the model of Heston (1993) is not known in closed form,⁴ unless we impose parameter restrictions that in effect make the model equivalent to that of Black and Scholes (1973); hence the need for methods such as ours to estimate models of this type by maximum likelihood.

The market price of risk specification in the model is $\Lambda = [\lambda_1 \sqrt{(1 - \rho^2)Y_t}, \lambda_2 \sqrt{Y_t}]'$. The joint dynamics of s_t and Y_t under the objective measure P are then

$$d \begin{bmatrix} s_t \\ Y_t \end{bmatrix} = \begin{bmatrix} a + b Y_t \\ \kappa(\gamma - Y_t) \end{bmatrix} dt + \begin{bmatrix} \sqrt{(1 - \rho^2)Y_t} & \rho\sqrt{Y_t} \\ 0 & \sigma\sqrt{Y_t} \end{bmatrix} d \begin{bmatrix} W_1^P(t) \\ W_2^P(t) \end{bmatrix}, \quad (25)$$

where

$$a = r - d, \quad b = \lambda_1(1 - \rho^2) + \lambda_2\rho - \frac{1}{2}, \quad \kappa = \kappa' - \lambda_2\sigma, \quad \gamma = \left(\frac{\kappa + \lambda_2\sigma}{\kappa} \right) \gamma'. \quad (26)$$

When the volatility state variable Y_t is not observable, its value must be extracted from option prices as discussed above in order to carry out the maximum likelihood estimation of the model's parameters, $\theta = [\kappa; \gamma; \sigma; \rho; \lambda_1; \lambda_2]'$. Since the price of a call option is a monotonically increasing function of the level of volatility, the value of Y_t can be determined from the price of a single option. We therefore take as given a joint time-series of observations of the log stock price s_t and the price of an at-the-money, constant-maturity option C_t . In principle, any option can be used, but this choice has three advantages. First, at-the-money and short-dated options are likely to be the most actively traded and liquid options, so their prices are least affected by microstructure and other such issues. Second, at-the-money options are highly sensitive to changes in volatility, so small observation errors in the price will have minimal effect on the implied level of volatility. Finally, as described in Section 3.2, the use of options with constant moneyness and time to maturity considerably simplifies the extraction of volatility from the vector of observed option prices. Note that this model satisfies the homogeneity requirements of (14), so that only the value of Y_t need be varied when computing option prices.

To calculate option prices in this model, we can use either the Feynman-Kac based on an expansion at order J of the density function, as given in Eq. (19), or characteristic functions as in Heston (1993), modified by Carr and Madan (1998), exploiting the fact that

⁴See Lamoureux and Paseka (2004) for an expression of the density of the Heston model using Fourier inversion of the characteristic function, which reduces the dimensionality of the required integration to a one-dimensional integral. The remaining integral is over a modified Bessel function of noninteger order.

this particular model is affine under the Q measure (it is also affine under P but this is irrelevant). The former method can be applied to all stochastic volatility models (provided only that they imply option prices that are martingales and not just local martingales), since the expansion $q_X^{(J)}$ can be calculated without restrictions, whereas the latter is specific to models that admit a closed form characteristic function; see other examples beyond Heston's model in Lewis (2000). However, the expansion-based pricing method will be most accurate for short-maturity options, since the density expansion is a Taylor expansion in the time variable.

In either case, we start with the option price expressed as

$$C(s_t, Y_t, K, \Delta) = E^Q[[\exp(s_{t+\Delta}) - K]^+ | s_t, Y_t], \quad (27)$$

where K is the strike price of the option, and Δ is the time remaining until maturity. Heston (1993) provides a Fourier transform method for calculating the option price; however, with this method, the characteristic function of the option is singular at the origin, making numeric integration difficult. Carr and Madan (1998) present an alternate Fourier transform procedure that avoids this difficulty. Rather than computing the option price directly, we calculate the option price scaled by the current price of the stock:

$$c(s_t, Y_t, K, \Delta) = \exp[-s_t] C(s_t, Y_t, K, \Delta). \quad (28)$$

It is then convenient to express the scaled option price in terms of the logarithmic moneyness of the option rather than the raw value of the strike price, $m_t = s_t - \ln K$. This scaled option price is given by

$$c(s_t, Y_t, K, \Delta) = \int_0^{+\infty} \operatorname{Re} \left[\frac{\exp[w_0 + w_1 \cdot m_t + w_2 \cdot Y_t]}{\alpha(\alpha + 1) - u^2 + (2\alpha + 1)iu} \right] du, \quad (29)$$

where α is an arbitrary scaling parameter and

$$\begin{aligned} w_0 &= \Delta((r - d)(\alpha + 1) - r + (r - d)iu) + \frac{\kappa' \gamma'}{\sigma^2} (\Delta \gamma_1 - 2 \ln(\gamma_2)), \\ w_1 &= iu + \alpha, \quad w_2 = (-u^2 + (2\alpha + 1)iu + \alpha(\alpha + 1)) \left(1 - \frac{1}{\gamma_2} \right) \frac{1}{\gamma_1}, \\ \gamma_0 &= \sqrt{c_0 + c_1 u + c_2 u^2}, \\ \gamma_1 &= \kappa - (iu + \alpha + 1)\rho\sigma + \gamma_0, \quad \gamma_2 = 1 + \left(\frac{\gamma_1}{\gamma_0} \right) \left(\frac{\exp(\Delta \gamma_0) - 1}{2} \right), \\ c_0 &= (\kappa')^2 - \sigma(\alpha + 1)(2\kappa'\rho - \sigma) - \sigma^2(\alpha + 1)^2(1 - \rho^2), \\ c_1 &= -i\sigma(2\sigma(\alpha + 1)(1 - \rho^2) + 2\kappa'\rho - \sigma), \quad c_2 = \sigma^2(1 - \rho^2). \end{aligned}$$

This expression can be evaluated quickly, since it is a one-dimensional integral. Heston (1993) even refers to similar one-dimensional integrals as “closed form.” Since we use options with constant moneyness and time to maturity, the integral above need only be calculated for each parameter vector evaluated during a likelihood search and over a one-dimensional grid of values of Y_t . By the above procedure, we can find the values of s_t and Y_t as functions of S_t and C_t . As discussed in Section 3.1, we then derive the likelihood f_{sY} of s_t and Y_t explicitly. The log-likelihood formulas, made specific for this particular model, are given in the appendix.

In the case where volatility is unobservable, the dependence of the joint likelihood function of $X_t = [S_t; Y_t]'$ under P on the full set of market price of risk parameters is introduced by the Jacobian term, itself resulting from the transformation from $[S_t; C_t]'$ to $[S_t; Y_t]'$ as described in Section 3.3. In the unobservable volatility case, the separate identification of the two market price of risk parameters is tenuous (see the Monte Carlo experiments below).

5. Volatility proxies

If, on the other hand, we have available a proxy for the state volatility variable, then maximum likelihood estimation of the vector θ can proceed directly without the need for option prices. Note, however, that the dynamics under P of the process $[S_t; Y_t]'$, or $[s_t; Y_t]'$ as given in Eq. (25), will only permit identification of the parameters $[\kappa; \gamma; \sigma; \rho; b]'$ or equivalently $[\kappa; \gamma; \sigma; \rho; \lambda_1]'$, since both components of the observed vector are viewed under P . In that situation, we will (arbitrarily) treat the λ_2 parameter as fixed at 0, and given the other identified parameters, translate the estimated value of b into an estimate for λ_1 .

The first proxy we use is an unadjusted Black-Scholes proxy in which the implied volatility of a short-maturity at-the-money option is used in place of the true instantaneous volatility state variable. The use of this proxy is justified in theory by the fact that the implied volatility of such an option converges to the instantaneous volatility of the logarithmic stock price as the maturity of the option goes to zero. The Monte Carlo simulations of Section 6 show that, for some models, the use of this proxy in place of an exact option pricing method still produces accurate estimates for the Heston model. Option pricing formulae are known in closed form for only a few models, and tractable numeric procedures such as those based on characteristic functions can be used only for a limited class of models. The ability to use a numerically tractable proxy is thus a significant advantage. However, as we will see below for the more general CEV model, we find that this simple proxy method introduces significant bias in the estimation of the elasticity of volatility parameter (simulations not shown). To remedy this problem, we develop an alternate proxy method.

5.1. The integrated volatility proxy

Our alternate proxy, which we name the integrated volatility proxy, corrects for the effect of mean reversion in volatility during the life of an option. If Y_t is the instantaneous variance of the logarithmic stock price, we can express the average integrated variance from time t to T as $V(t, T)$:

$$V(t, T) = \frac{1}{T-t} \int_t^T Y_u du. \quad (30)$$

If the volatility process is instantaneously uncorrelated with the logarithmic stock price process, we can then calculate option prices by taking the expected value of the Black-Scholes option price with $V(t, T)$ as the implied variance over the probability distribution of $V(t, T)$: see Hull and White (1987). If the two processes are correlated, then the price of the option is a weighted average of Black-Scholes prices evaluated at different stock prices and volatilities: see Romano and Touzi (1997).

Our proxy is determined by calculating the expected value of $V(t, T)$ first, and substituting this value into the Black-Scholes formula as implied variance. This proxy is model-free, in that it can be calculated whether or not an exact volatility can be computed, and results in a straightforward estimation procedure. On the other hand, this procedure is in general approximate, first because the volatility process is unlikely to be instantaneously uncorrelated with the logarithmic stock price process, and second because the expectation is taken before substituting $V(t, T)$ into the Black-Scholes formula rather than after. We examine below in Monte Carlo simulations the respective impact of these two approximations, with the objective of determining whether the tradeoff involved between simplicity and exactitude is worthwhile.

The idea is to adjust the Black-Scholes implied volatility for the effect of mean reversion in volatility, essentially undoing the averaging that takes place in Eq. (30). Specifically, if the Q -measure drift of Y_t is of the form $a + bY_t$ (as it is in all the models we examine), then the expected value of $V(t, T)$ is given by

$$E_t[V(t, T)] = \left(\frac{e^{b(T-t)} - 1}{b(T-t)} \right) \left(Y_t + \frac{a}{b} \right) - \frac{a}{b}. \quad (31)$$

(A similar expression can be derived in the special case where $b = 0$.) By taking the expected value on the left-hand side to be the observed implied variance $V_{\text{imp}}(t, T)$ of a short-maturity T at-the-money option, our adjusted proxy is then given by

$$Y_t \approx \frac{bV_{\text{imp}}(t, T) + a(T-t)}{e^{b(T-t)} - 1} - \frac{a}{b}. \quad (32)$$

To examine the accuracy of this proxy, we perform the following experiment for a range of different parameter values θ : given Y_t , generate at-the-money short-maturity (30 days as is the case for the VIX data; see Section 7 below) option prices using the exact CEV model, compute the options' Black-Scholes implied volatility $V_{\text{imp}}(t, T)$, then back out Y_t using Eq. (32) and plot the resulting value of Y_t (proxy-implied) against the original value of Y_t . If the overall procedure is accurate, then the plot should be close to a 45° line. The result is reported in Fig. 1, which is to be compared to Fig. 2 for the same experiment but using the unadjusted Black-Scholes proxy, where one simply sets $Y_t \approx V_{\text{imp}}(t, T)$ instead of our

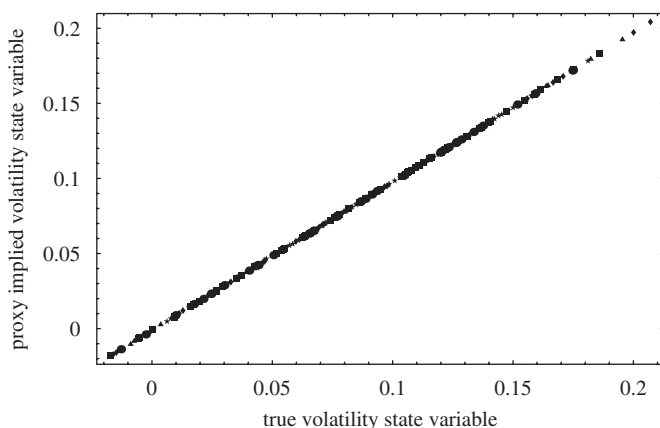


Fig. 1. From implied volatility proxy back to latent volatility state variable: integrated volatility proxy.

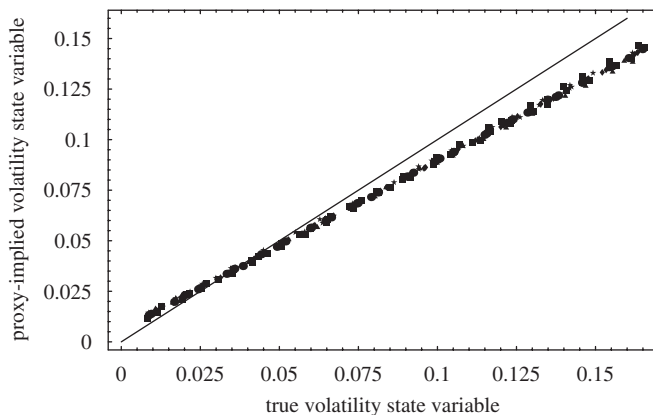


Fig. 2. From implied volatility proxy back to latent volatility state variable: Black-Scholes proxy.

proxy where one sets $E_t[V(t, T)] \approx V_{\text{imp}}(t, T)$. Each point corresponds to a different value of Y_t , and one of ten sets of parameter values chosen to be realistic for the model. As is clear from comparing the two figures, adjusting for the mean-reverting behavior of the volatility state variable results in a marked improvement in our ability to produce an accurate proxy-implied Y_t from an option-implied volatility. By observing the departures from the 45° line in Fig. 2, we see that the Black-Scholes proxy tends to overstate (understate) Y_t when volatility is low (high). This is as expected, given the mean reversion in volatility. Our adjusted proxy corrects for this, and recovers the unobservable Y_t with much greater accuracy.

The Black-Scholes proxy $Y_t \approx V_{\text{imp}}(t, T)$ does not depend on any of the model's parameters (except the instantaneous interest rate and dividend yield, which we treat as observed); by contrast, our integrated volatility proxy depends on the parameters of the drift of volatility. But, given the explicit inversion formula (32), we can simply take $[S_t; V_{\text{imp}}(t, T)]'$ as the state vector, write its likelihood from that of $[S_t; Y_t]'$ using a Jacobian term for the change of variable (32), and estimate all the parameters in one stage.⁵

5.2. The Lewis proxy for the CEV model

We also consider a more general model, which nests the Heston model as well as the GARCH model we will investigate empirically below. Under the Q measure, the state variables S_t and Y_t follow a process of the following form:

$$dX_t = d \begin{bmatrix} S_t \\ Y_t \end{bmatrix} = \begin{bmatrix} (r - d)S_t \\ \kappa'(\gamma' - Y_t) \end{bmatrix} dt + \begin{bmatrix} \sqrt{(1 - \rho^2)Y_t}S_t & \rho\sqrt{Y_t}S_t \\ 0 & \sigma Y_t^\beta \end{bmatrix} d \begin{bmatrix} W_1^Q(t) \\ W_2^Q(t) \end{bmatrix}, \quad (33)$$

⁵If a different proxy is used, where the inversion of $V_{\text{imp}}(t, T)$ back into Y_t is not available in closed form, one can employ a two-stage estimation procedure to speed computations. In the first stage, estimate only the parameters of the univariate volatility process, using the Black-Scholes proxy. Then use the first-stage estimated parameters to calculate the volatility proxy, which is then used in a second-stage estimate of all model parameters.

where we constrain the parameter $\beta \geq 1/2$, as well as $\beta \leq 1$ (to retain the uniqueness of option prices). This model is considered by Jones (2003) among others and nests both the models of Heston (1993) ($\beta = 1/2$) and the GARCH model ($\beta = 1$). Note, however, that the special properties of these models could still warrant separate investigation, despite their being nested by the model of (33); for example, the Fourier inversion method for option pricing is feasible for the model of Heston (1993), where $\beta = 1/2$, but not for the general CEV model. The state variable Y_t has a boundary at zero, but this boundary can only be achieved when $\beta = 1/2$, and even then only for certain values of the model parameters. Proceeding as before, we can express these dynamics in terms of the logarithmic stock price rather than the stock price itself:

$$d \begin{bmatrix} s_t \\ Y_t \end{bmatrix} = \begin{bmatrix} r - d - \frac{1}{2} Y_t \\ \kappa'(\gamma' - Y_t) \end{bmatrix} dt + \begin{bmatrix} \sqrt{(1 - \rho^2)Y_t} & \rho\sqrt{Y_t} \\ 0 & \sigma Y_t^\beta \end{bmatrix} d \begin{bmatrix} W_1^Q(t) \\ W_2^Q(t) \end{bmatrix}. \quad (34)$$

We again set at zero the component of the market price of risk vector corresponding to Y_t , and specify $\Lambda = [\lambda_1 \sqrt{(1 - \rho^2)Y_t}, 0]'$.

The P -measure dynamics of the state variables are then

$$d \begin{bmatrix} s_t \\ Y_t \end{bmatrix} = \begin{bmatrix} r - d - \frac{1}{2} Y_t + \lambda_1(1 - \rho^2)Y_t \\ \kappa(\gamma - Y_t) \end{bmatrix} dt + \begin{bmatrix} \sqrt{(1 - \rho^2)Y_t} & \rho\sqrt{Y_t} \\ 0 & \sigma Y_t^\beta \end{bmatrix} d \begin{bmatrix} W_1^P(t) \\ W_2^P(t) \end{bmatrix}, \quad (35)$$

where $\kappa' = \kappa$ and $\gamma' = \gamma$. The Q -measure dynamics are not affine in Y_t . The corresponding log-likelihood expansion can be found in the appendix.

It is possible to refine the implied volatility proxy by expressing it in the form of a Taylor series in the “volatility of volatility” parameter σ in the case of the CEV model, where the Q -measure drift of Y_t is of the form $a + bY_t$, and the Q -measure diffusion of Y_t is of the form σY_t^β . Lewis (2000) showed that in that case

$$V_{\text{imp}}(t, T) = E_t[V(t, T)] + \sigma(T - t)^{-1} \left(\frac{1}{2} - \frac{\ln(S_t/K) + (r - \delta)(T - t)}{Y_t(T - t)} \right) \\ \times \frac{\rho}{b} \int_t^T (e^{b(T-s)} - 1) \left(e^{b(s-t)} \left(Y_t + \frac{a}{b} \right) - \frac{a}{b} \right)^{\beta+1/2} ds + O(\sigma^2), \quad (36)$$

where $E_t[V(t, T)]$ is given in (31) as a function of Y_t . While Eq. (36) should be expected to be more accurate in the case of the CEV model, provided that the parameter σ is indeed small, it is unfortunately not that useful in our context because the relation between the observed $V_{\text{imp}}(t, T)$ and the latent Y_t is not invertible without numerical computation of the parameter-dependent integral in Eq. (36), followed by the numerical inversion of the nonlinear relation from $V_{\text{imp}}(t, T)$ back to Y_t . This somewhat defeats the purpose of using a volatility proxy for our estimation purposes, since option prices are themselves integrals of the state variables: inferring Y_t from option prices C_t directly, through Eq. (19), also requires inversion of a parameter-dependent integral.

Said differently, if we are willing to solve numerically a parameter-dependent integral as part of the maximum likelihood search, then we might as well use the exact option prices C_t to infer Y_t directly, as in the general inference method we discuss above, and dispense with the additional approximation involved in the small-volatility-of-volatility Taylor series, or the use of implied volatility altogether. This is not to say that expression (36)

cannot be useful in different contexts, such as figuring out the implications for the implied volatility smile of given parameter values, for instance, which is an application that uses the formula in the direction from Y_t to $V_{\text{imp}}(t, T)$, but not for our purpose of inverting it in the direction from $V_{\text{imp}}(t, T)$ to Y_t . But the inversion required to use the proxy (36) is of the same degree of complexity as that of using option prices directly, thereby nullifying the only advantage of using a proxy in our context, which is to tradeoff some accuracy in exchange for a greatly simplified search for the likelihood maximum.

Even if we estimate the parameters based on a change of state variables from $[S_t, Y_t]'$ to $[S_t, V_{\text{imp}}(t, T)]'$, applying Itô's Lemma to (36) in order to find out the Q -dynamics of $V_{\text{imp}}(t, T)$ implied by the Q -dynamics of Y_t does not eliminate the need to compute that integral for each pair of observations and parameter values along the maximization path of the likelihood. For these reasons, we suggest the immediately computable proxy (32). Furthermore, that proxy has the advantage of being applicable to all stochastic volatility models with affine drift, not just the CEV model.⁶

6. Monte Carlo results

One major advantage of the proposed estimation method for stochastic volatility models is that it is numerically tractable, so that large numbers of Monte Carlo simulations can be conducted to determine the small-sample distribution of the estimators, examine the effect of replacing the unobservable volatility variable Y_t by a proxy, and compare the small-sample behavior of the estimators to their predicted asymptotic behavior.

6.1. Estimators for the Heston model

We start with simulations for the model of Heston (1993) for a variety of assumptions about sample length, time between observations, and observability of the volatility state variable. This model is a natural choice, since option prices can be calculated easily through Fourier inversion of the characteristic function; it is possible therefore to compare results obtained with the exact option pricing formula to those obtained using the unadjusted Black-Scholes proxy discussed above. We use sample lengths of $n = 500$, 5,000, and 10,000 transitions at the daily ($\Delta = 1/252$) frequency, and $n = 500$ transitions at the weekly ($\Delta = 7/365$) frequency. The parameter values used in the simulations are similar to the values obtained from the empirical application in Section 7, and are reported in Table 1. Note in particular that a value of -0.8 is chosen for ρ , to reflect the empirical regularity that innovations to volatility and stock price are generally strongly negatively

⁶The integrated volatility proxy requires us to calculate the expectation of $V(t, T)$. This is an exact calculation, with result given in (31), for all models whose volatility has an affine drift. For models in which the drift of the volatility state variable is not affine, this expectation cannot in general be found in closed form. One possible way to apply the integrated volatility proxy to such models is to use a local linear approximation, in which the drift of volatility is replaced by the first two terms of a power series expansion of the drift about the current value of volatility. The accuracy of such an approximation for short maturities in a term structure context is confirmed by Takamizawa and Shoji (2003). However, since in the models we consider, the drift of volatility is affine, we do not need to employ this. And, in any event, in the absence of an easily computable formula for inverting $V_{\text{imp}}(t, T)$ into Y_t , we might as well use option prices C_t directly, as discussed above, and bypass the use of a proxy for Y_t altogether.

Table 1
Monte Carlo simulations with observed volatility—Heston model

Parameter	True value	500 Daily obs.		5,000 Daily obs.		10,000 Daily obs.		500 Weekly obs.	
		Bias	Std. dev.	Bias	Std. dev.	Bias	Std. dev.	Bias	Std. dev.
κ	3.00	0.86	1.57	0.068	0.38	0.033	0.25	0.159	0.56
γ	0.10	0.0005	0.022	0.0001	0.0057	−0.0001	0.004	0.0001	0.0084
σ	0.25	0.0002	0.006	0.0000	0.0020	0.0000	0.0014	0.0004	0.006
ρ	−0.8	−0.0002	0.013	−0.0001	0.0042	−0.0001	0.003	−0.0015	0.013
λ_1	4.0	0.92	6.5	0.07	1.9	0.0977	1.4	0.2	2.9

This table shows the results of 1,000 Monte Carlo simulations for the Heston model, with observed volatility and 500, 5,000, and 10,000 daily observations (i.e., $\Delta = 1/252$) and 500 weekly observations (i.e., $\Delta = 7/365$). The second column shows the true value θ_0 of the parameters used to generate the simulated sample paths. The “Bias” column shows the mean bias of the estimated parameter vector, i.e., the difference between the estimated parameters and the true values. The “Std. dev.” column shows the standard deviation of the parameter estimates.

The bias and standard deviation quantify the distance $\|\hat{\theta}_{n,\Delta}^{(J,\text{instant})} - \theta_0\|$ for the various number of observations and sampling frequency. The market price of risk of the stochastic volatility variable, λ_2 , is not identified when volatility is treated as observed. The instantaneous interest rate and the instantaneous dividend yield of the stock are held fixed at the values of 4% and 1.5%, respectively.

correlated. The values of the instantaneous interest rate and dividend yield, r and d , are held fixed.

For each batch of simulations, we generate 1,000 sample paths using an Euler discretization of the process, using 30 sub-intervals per sampling interval; 29 out of every 30 observations are then discarded, leaving only observations at either a daily or weekly frequency. Each simulated data series is initialized with the volatility state variable at its unconditional mean, and the stock at 100. An initial 500 observations are generated and then discarded; the last of these observations is then taken as the starting point for the simulated data series. We then generate 500, 5,000, or 10,000 additional observations.

We next estimate the model parameters using the method described above. When simulating the joint dynamics of the state vector $X_t = [S_t; Y_t]'$, we have the luxury of deciding whether Y_t is observable or not; we can determine the effect of ignoring the difference between the (unobservable) stochastic volatility variable Y_t , and an (observable) proxy, namely the implied volatility of a short-dated at-the-money option, or an affine function of the implied volatility of such an option. Our method can be applied to either situation: treating Y_t as unobservable or replacing it by an observable proxy, which, as discussed in Section 3, eliminates the need for the third step of our method, and greatly simplifies the second.

Table 1 reports results, all with observed volatility. Table 2 reports results, but with volatility not directly observed, that is, employing the full estimation procedure where we use the model to generate simulated option prices, i.e., observations on C_t , then use $G_t = [S_t; C_t]'$ as the observed vector. In each case, the mean difference between the estimates and the true values of the parameter (used in the data-generation procedure) over the simulated paths is reported as the bias of the estimation procedure. The standard deviation of each parameter is computed accordingly and reported.

Throughout, the best estimates are for the σ and ρ parameters. Regardless of sampling frequency and whether or not volatility is observed, both the biases and standard errors of the estimates are small relative to the parameter values. The γ parameter fares only slightly

Table 2
Monte Carlo simulations with unobserved volatility—Heston model

Parameter	True value	500 Weekly obs. Bias	Using option prices Std. dev.
(1)	(2)	(3)	(4)
κ	3.00	−0.59	1.59
γ	0.10	−3.04	14.5
σ	0.25	−0.013	0.03
ρ	−0.8	−0.0006	0.026
λ_1	−7.0	−0.54	4.9
λ_2	−6.0	−0.67	3.8

This table shows the results of 1,000 Monte Carlo simulations for the Heston model with 500 weekly observations (i.e., $\Delta = 7/365$) and unobserved volatility, but observed option prices. The data are pairs of values of $[S_t; C_t]'$, and we use Fourier inversion to back out the unobservable value Y_t of the state variable. Column 2 shows the parameters used to generate the simulated sample paths. This table quantifies the distance $\|\hat{\theta}_{n,\Delta}^{(J,\text{price})} - \theta_0\|$. The instantaneous interest rate and the instantaneous dividend yield of the stock are held fixed at the values of 4% and 1.5%, respectively.

worse when volatility is observed; when volatility is unobserved, the standard deviation of γ is much larger, for reasons discussed below. The κ and λ_1 parameters are estimated with less accuracy. The bias in the κ estimates decreases substantially with the length of the sample; this is possibly indicative that the source of the bias identified in Li et al. (2004) could be relevant for sample sizes of 500 daily observations. As expected, we typically overestimate in small-samples the speed of mean reversion when it is low (so the bias on the κ parameter is positive).

The use of otherwise similar batches of simulations with differing numbers of daily observations in each simulated series provides some insight into how fast the small-sample distribution of the estimated parameters approaches the asymptotic distribution. As the number of observations in each simulated data series increases, we would expect the standard errors of the parameter estimates to decrease at a rate inversely proportional to the square root of the number of observations. The decreases in standard errors are approximately what one would expect from asymptotic theory; for example, in Table 1, the small-sample standard errors for all parameters except κ are very close to the asymptotic standard errors. The small-sample standard error for κ is larger than the asymptotic standard error for 500 daily observations, but is much closer for 5,000 and 10,000 daily observations. The standard errors for all parameters decrease with sample size at roughly the rate one would predict from asymptotic theory, i.e., by a factor of the square root of ten when increasing from 500 to 5,000 observations, and by a factor of the square root of two when increasing from 5,000 to 10,000 observations. These results suggest that the distribution of the estimates is approaching its asymptotic limit.

When the value of the volatility state variable is determined through the use of an option price C_t , rather than observed directly, the identification of λ_2 relies exclusively on the introduction of the Jacobian term in the likelihood function of the observables. As expected given this tenuous dependence of the likelihood on the second market price of risk parameter, that parameter is generally identified quite poorly. The strong correlation between the two Brownian motions driving state variable evolution confounds this problem; as shown in Table 2, the standard errors for both market price of risk parameters

are large. This is a feature of the sampling noise inherent in the estimation, not of the approximation involved in the likelihood function.

Also of note is the large standard error for the γ parameter when volatility is unobserved. This result might seem surprising, given the relatively accurate estimation of this parameter when volatility is observed. However, the results are not comparable. Note that γ is always multiplied by κ to determine the constant term in the drift of Y_t . When volatility is treated as observed, the market price of risk parameter λ_2 is held fixed, so that the value of κ' (i.e., the P -measure speed of mean reversion) is constrained by the value of κ (i.e., the Q -measure speed of mean reversion). However, when volatility is treated as unobserved, κ' and κ can vary independently. Consequently, κ is estimated more poorly when volatility is unobserved, and this has an effect on the estimation of γ . If we consider the product $\kappa\gamma$, we find it is estimated only slightly worse when volatility is unobserved. The increase in the standard error of γ when volatility is unobserved is therefore largely a by-product of the increase in volatility of κ , rather than a result of any severe deterioration of our ability to estimate the constant term in the drift of the volatility state variable.

6.2. The effects of using a volatility proxy in the Heston model

Of course, volatility is not observed. Consequently, of particular interest are the results for the Monte Carlo simulations with the same sampling frequency and number of observations, but different methods of determining the level of the volatility state variable, Y_t , using indirect market data such as an option price, a simple Black-Scholes implied volatility from a short-term at-the-money option, or our adjusted integrated volatility proxy.

When using a proxy, one way to assess its effect is to examine whether the proxy for Y_t introduces enough additional noise to be noticeable at the scale of the standard error of the estimators due to the sampling noise. Indeed, for a given sample size, the best possible estimator is the true (but effectively incomputable) maximum likelihood estimator; call it $\hat{\theta}_{n,\Delta}^{(\text{true})}$ when n data points sampled at interval Δ are used. Around the true parameter vector θ_0 , its sampling distribution is $\hat{\theta}_{n,\Delta}^{(\text{true})} - \theta_0$. Because of the Cramer-Rao lower bound, the distance $\|\hat{\theta}_{n,\Delta}^{(\text{true})} - \theta_0\|$ is as close as one can get to θ_0 using a consistent estimator. Because $\hat{\theta}_{n,\Delta}^{(\text{true})}$ is incomputable, we consider alternative estimators $\hat{\theta}_{n,\Delta}^{(J)}$ from an expansion of the log-likelihood function of order J ; $J = 1$ in these simulations. There are three versions of the latter, in simulations: $\hat{\theta}_{n,\Delta}^{(J,\text{instant})}$ if the (unobservable in reality, but not in simulations) instantaneous volatility is used, $\hat{\theta}_{n,\Delta}^{(J,\text{proxy})}$ when using a volatility proxy instead of the unobservable instantaneous volatility and $\hat{\theta}_{n,\Delta}^{(J,\text{price})}$ if we use an option price C_t to back out by Fourier inversion the implied value of the volatility state variable. One estimator we propose for empirical work is $\hat{\theta}_{n,\Delta}^{(J,\text{proxy})}$. To evaluate the performance of a proposed estimator such as $\hat{\theta}_{n,\Delta}^{(J,\text{proxy})}$, we would like to compare $\|\hat{\theta}_{n,\Delta}^{(J,\text{proxy})} - \hat{\theta}_{n,\Delta}^{(\text{true})}\|$ to $\|\hat{\theta}_{n,\Delta}^{(\text{true})} - \theta_0\|$, or measure the increase in $\|\hat{\theta}_{n,\Delta}^{(J,\text{proxy})} - \theta_0\|$ relative to $\|\hat{\theta}_{n,\Delta}^{(\text{true})} - \theta_0\|$. Even though we cannot compute $\hat{\theta}_{n,\Delta}^{(\text{true})}$, we can estimate $\|\hat{\theta}_{n,\Delta}^{(\text{true})} - \theta_0\|$ using the inverse of Fisher's information matrix—a useful by-product of using a likelihood-based estimation method. In other words, we can assess how much farther from θ_0 we end up because of the dual

approximations (using a proxy instead of the instantaneous volatility and an expansion of the log-likelihood instead of the true log-likelihood), recalling that there is nothing we can do about the true MLE sampling error $\|\hat{\theta}_{n,\Delta}^{(\text{true})} - \theta_0\|$.

Then, since $E[\|\hat{\theta}_{n,\Delta}^{(J,\text{proxy})} - \hat{\theta}_{n,\Delta}^{(\text{true})}\|^2]$ is the root mean squared error of the estimator, we can compare the bias and variance of $\hat{\theta}_{n,\Delta}^{(J,\text{proxy})} - \hat{\theta}_{n,\Delta}^{(\text{true})}$ to those of $\hat{\theta}_{n,\Delta}^{(\text{true})} - \theta_0$. Note that this measures directly the effect of the estimator using the volatility proxy. We can also compare $\|\hat{\theta}_{n,\Delta}^{(J,\text{price})} - \theta_0\|$ to $\|\hat{\theta}_{n,\Delta}^{(J,\text{proxy})} - \theta_0\|$, and $\|\hat{\theta}_{n,\Delta}^{(J,\text{instant})} - \theta_0\|$ to $\|\hat{\theta}_{n,\Delta}^{(J,\text{proxy})} - \theta_0\|$ in order to isolate in simulations the effect of the approximation introduced by the volatility proxy, keeping in mind that, just like $\hat{\theta}_{n,\Delta}^{(\text{true})}$, $\hat{\theta}_{n,\Delta}^{(J,\text{instant})}$ too is an infeasible estimator since instantaneous volatility is not in general observable.

The results in Table 1 are based on the assumption that volatility is observed, whereas the results in Table 2 are based on volatility extracted from option prices. At the daily frequency, the standard errors for λ_1 are roughly similar; the standard error for κ is substantially smaller when volatility is observed through a proxy. There is no question that when volatility is observed, the estimation error is going to be smaller, and the table confirms that.

But in the real world volatility is not observed. So comparing the estimation accuracy with and without volatility being observed is useful for benchmarking—which is why we include it—but not for judging an estimation method in relative terms. The relevant comparison is among different methods that do not assume that volatility is observed. For that purpose, Table 3 compares the use of the observed volatility state variable (infeasible

Table 3
Effect of volatility proxy—Heston model

Parameter	True value	Observed vol.		Black-Scholes implied vol.		Integrated vol. proxy	
		Bias	Std. dev.	Bias	Std. dev.	Bias	Std. dev.
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
κ	3.00	0.17	0.56	0.16	0.55	0.16	0.53
γ	0.10	−0.0002	0.008	−0.0007	0.008	−0.001	0.007
σ	0.25	0.0000	0.003	−0.03	0.005	0.0009	0.006
ρ	−0.8	0.0000	0.006	−0.0009	0.006	0.0006	0.006
λ_1	4.0	0.19	2.8	0.21	2.8	0.25	2.7

This table shows the results of 1,000 Monte Carlo simulations for the Heston model with 2,500 daily observations (i.e., $\Delta = 1/252$) comparing the estimator when volatility is treated as observed (infeasible outside of simulations) to those obtained using two different proxies for the instantaneous volatility (the implied volatility of a short dated at-the-money option and our adjusted integrated volatility proxy). Column 2 shows the parameters used to generate the simulated sample paths; the simulations are the same as those used in Table 2. Column 3 shows the mean bias of the estimated parameter vector, i.e., the difference between the estimated parameters and the values shown in the second column, when the volatility is observed. Column 4 shows the standard deviation of the parameter estimates, also using Fourier inversion. These two columns quantify $\|\hat{\theta}_{n,\Delta}^{(J,\text{instant})} - \theta_0\|$. Columns 5 and 6 show the same information, but using the implied volatility of an at-the-money short-maturity option to determined the level of the stochastic volatility variable. These two columns quantify $\|\hat{\theta}_{n,\Delta}^{(J,\text{proxy})} - \theta_0\|$, for the unadjusted implied volatility proxy. Columns 7 and 8 report the comparable numbers when our adjusted or integrated volatility proxy is used. The instantaneous interest rate and the instantaneous dividend yield of the stock are held fixed at the values of 4% and 1.5%, respectively.

outside of simulations) to that of two alternative proxies to determine the level of stochastic volatility: the Black-Scholes, unadjusted, implied volatility proxy, and our adjusted integrated volatility proxy. The λ_2 parameter is held fixed at its data generating value, since it is unidentified when a proxy is used; holding this parameter fixed in both sets of estimates makes the results comparable.

The table shows that the adjusted proxy and the implied volatility give comparable results for the Heston model, with a slight advantage for the adjusted proxy. Note in particular that the implied volatility proxy results in a slight bias in the estimated σ parameter, which is typically estimated very precisely. In this particular model, we are only attempting to estimate a single “volatility of volatility” parameter. But this bias already announces the markedly different results—and issues with the use of an unadjusted implied volatility proxy—that we will see below for the CEV model, where volatility of volatility is controlled by two separate parameters.

A different but related point is that the design where the value of λ_2 is known biases the results in favor of the use of the proxy. Indeed, while we can estimate the parameter $b = \lambda_1(1 - \rho^2) + \lambda_2\rho - 1/2$ in (26) quite accurately, the resulting estimate for λ_1 is necessarily dependent upon the assumption made about λ_2 . When using real data, we do not have the luxury of knowing the value of λ_2 . One solution is simply to focus on the parameter b alone, but this is of little use if the objective is to price derivatives (in which case we need the parameters under Q). Another solution is to first estimate the process using the full procedure, and use the resulting value of λ_2 as an input above. This said, it is important to note that this feature could well be shared by other methods designed to estimate stochastic volatility models, but their numerical intensity makes simulating them impractical so it is difficult to know precisely how they behave.

6.3. Simulations for the CEV stochastic volatility model

Similar simulations to those of the Heston model for the CEV model show that, when an unadjusted Black-Scholes implied volatility proxy is used, significant bias results in the estimation of the elasticity parameter β in particular. The estimates of β will often bump against the exogenously imposed boundary of 1 (there to ensure that the deflated stock price is a martingale, not just a local martingale.) We therefore focus on the effect of using the integrated volatility proxy, and compare that to the Lewis proxy (36). Table 4 shows the parameter estimates and standard errors obtained for the CEV model using the same procedure as that described in Section 4. Obviously, the use of a proxy induces estimation error relative to the (infeasible) use of the instantaneous volatility, but the estimation biases and variances remain small.

As noted above, the computational times for the Lewis proxy are substantially higher than with the simple integrated volatility proxy (32) due to the need to numerically invert the mapping from the observed implied volatility to the volatility state variable, and the further nonlinearity of the relation between $V_{\text{imp}}(t, T)$ and Y_t which cannot be inverted explicitly. Furthermore, despite the fact that it should in theory be more accurate, the table shows that the additional complexity results in no marked improvement over our simple integrated volatility proxy, despite its much higher computational burden. In fact, for two of the main parameters of the model, σ and β , the Lewis proxy results in a decrease in accuracy relative to the simpler integrated volatility proxy. Results for the other parameters are comparable for the two proxies.

Table 4
Effect of volatility proxy—CEV model

Parameter	True value	Observed volatility		Integrated volatility proxy		Lewis volatility proxy	
		Bias	Std. dev.	Bias	Std. dev.	Bias	Std. dev.
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
κ	4.00	1.07	2.10	1.11	2.23	1.09	2.11
γ	0.05	0.0014	0.018	0.0008	0.015	0.0006	0.014
σ	0.75	0.0004	0.15	−0.048	0.14	0.13	0.10
ρ	−0.75	−0.0004	0.017	0.0016	0.02	−0.006	0.02
β	0.80	−0.0065	0.065	−0.055	0.088	0.19	0.029
λ_1	4.0	1.28	7.56	1.15	7.45	1.35	7.66

This table shows the results of 1,000 Monte Carlo simulations of the CEV model with 500 daily observations (i.e., $\Delta = 1/252$), comparing the situation where volatility is observed to that where our integrated volatility proxy is used. Column 2 shows the parameters used to generate the simulated sample paths; the simulations are the same as those used in Table 2. Column 3 shows the mean bias of the estimated parameter vector, i.e., the difference between the estimated parameters and the values shown in Column 2, when the volatility is observed directly. Column 4 shows the standard deviation of the parameter estimates, also with observed volatility. These two columns quantify $\|\hat{\theta}_{n,\Delta}^{(J,\text{instant})} - \theta_0\|$. Columns 5 and 6 show the corresponding information using the implied volatility of an at-the-money short-maturity option to determine the level of the stochastic volatility variable. These two columns quantify $\|\hat{\theta}_{n,\Delta}^{(J,\text{proxy})} - \theta_0\|$. Columns 7 and 8 repeat the experiment for the Lewis proxy. The λ_2 parameter is not identified and held fixed at its data-generating value $\lambda_2 = 0$. The instantaneous interest rate and the instantaneous dividend yield of the stock are held fixed at the values of 4% and 1.5%, respectively.

6.4. Comparing small-sample to asymptotic distributions

We can also use Monte Carlo simulations to assess whether the predicted asymptotic behavior of maximum likelihood estimators is matched in small-samples by our maximum likelihood estimator. Table 5 compares the asymptotic standard deviations of the estimates obtained from the approximate likelihood function with the empirical standard deviations obtained from the Monte Carlo simulations. As shown, the two versions of the standard deviations converge as the sample size gets larger, suggesting not only that the likelihood approximations are quite accurate but also that standard statistical theory, namely

$$n^{1/2}(\hat{\theta} - \theta) \rightarrow N(0, F^{-1}), \tag{37}$$

where $F = -E[\partial^2 l_G / \partial \theta \partial \theta']$ is Fisher’s information matrix, works well in this context. As is well known, the Cramer-Rao lower bound states that F^{-1} is the lowest possible asymptotic variance achievable by a consistent estimator of θ .

Table 5 shows that we are close to the efficient asymptotic standard errors for all parameters despite the finite sample sizes, as the finite sample distribution appears close to the asymptotic distribution with as few as 500 daily observations (of course, anything can happen in finite samples and it is theoretically possible for a different estimator to beat maximum likelihood in finite samples).

Table 5
Asymptotic and small-sample variances of estimates with observed volatility

Parameter	True value	500 Daily obs.		5,000 Daily obs.		10,000 Daily obs.	
		ASE	SSSE	ASE	SSSE	ASE	SSSE
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
κ	3.00	1.14	1.57	0.36	0.38	0.25	0.25
γ	0.10	0.019	0.022	0.0059	0.0057	0.0042	0.0041
σ	0.25	0.0061	0.0062	0.0019	0.0020	0.0014	0.0014
ρ	−0.8	0.0133	0.0134	0.0042	0.0042	0.003	0.003
λ_1	4.0	6.2	6.5	1.97	1.91	1.40	1.42

This table shows the standard deviations of the parameter estimates $\hat{\theta}_{n,d}^{(J,\text{instant})}$ for the Heston model, calculated both analytically and from the Monte Carlo simulations. All values are based on daily observations. Column 2 shows the true values of the parameter vector used to generate the sample paths for the Monte Carlo simulations and to calculate the standard deviations from the likelihood expressions. Column 3, marked ASE for Asymptotic Standard Error, shows the asymptotic standard deviations of each parameter when the data series contains 500 daily observations. These values were obtained by computing the expected value of the second derivatives of the log likelihood in the form of an integral. Column 4, marked SSSE for Small Sample Standard Error, shows the standard deviations of the parameter estimates from the Monte Carlo simulations, i.e., the same information as in the corresponding column of Table 1. Columns 5 and 6 show the same information as the third and fourth columns, but with 5,000 daily observations instead of 500. Finally, Columns 7 and 8 show the same information but with 10,000 daily observations. The market price of risk for the stochastic volatility variable is not identified when volatility is observed; the instantaneous interest rate and dividend yield of the stock are held fixed at 4% and 1.5%/year, respectively. Note that, with this number of observations, the standard deviations calculated analytically from the approximate likelihood function are quite close to those observed in the Monte Carlo simulations.

6.5. Conclusions from the Monte Carlo simulations

We therefore leave our Monte Carlo analysis with the following conclusions:

1. The use of an implied volatility of a short-dated at-the-money option as a proxy for the unobservable volatility variable Y_t means that one market price of risk parameter is not identifiable, but on the other hand the separate identification of the two market price of risk parameters is poor when option prices are used and computational times are greatly increased due to the need to numerically compute the option prices. It seems therefore that using a proxy is a reasonable tradeoff to make if we can live without the full identification or make an arbitrary assumption about one of the market price of risk parameters.
2. Black-Scholes implied volatility works reasonably well as a proxy in simulations for the Heston model. For the more general CEV model, however, the use of this unadjusted implied volatility leads to significant bias in the elasticity of variance parameter β (which of course is held fixed and not estimated in the Heston model). The integrated volatility proxy takes the Black-Scholes implied volatility and adjusts it for the effect of mean reversion in volatility during the life of an option. This is only a partial adjustment to what makes the Black-Scholes implied volatility different from the exact unobservable instantaneous volatility, but our simulations show that this simple adjustment goes a long way towards eliminating the bias. For the CEV model, the Lewis proxy does not

provide an improvement over the integrated volatility proxy, while necessitating major computational effort due to the need to recover the latent volatility variable numerically rather than explicitly.

3. The small-sample distributions of the maximum likelihood estimators are well approximated by their asymptotic counterparts, even in samples as small as 500 daily observations.

7. Three models and the data

Given the guidance provided by the Monte Carlo simulations above, we are now ready to tackle real data. When applying our method to real data, we use direct observations on asset prices $G_t = [S_t; C_t]'$ with S_t representing the S&P 500 Index and C_t the price of a short-maturity at-the-money option. The option price is computed from its implied volatility, itself measured as the Chicago Board Options Exchange (CBOE) Volatility Index (VIX). We use the VIX data computed using the methodology introduced by the CBOE on September 22, 2003, which is an implied volatility index based on the European S&P 500 options as opposed to the American S&P 100 options (whose implied volatility index symbol is now VXO).

The VIX is an estimate of the implied volatility of a basket of S&P 500 Index Options (SPX) constructed from different traded options in such a way that at any given time it represents the implied volatility of a hypothetical at-the-money option with $T = 30$ calendar days to expiration (or 22 trading days). This constant maturity and constant moneyness feature of the data matches nicely with the assumptions we have made to reduce the dimensionality of the option pricing problem (see Section 3.2). In what follows, we will use the VIX as our measure of the implied volatility, $VIX_t^2 = V_{\text{imp}}(t, T)$, from which we can compute our proxy for Y_t through (32).

The anticipated daily cash dividends of the S&P 500 are forecast by the CBOE. These forecasts are generally very accurate in light of the short time span as well as the averaging effect of a large stock index. The VIX options are European, simplifying the analysis. For further details on the VIX, see [Whaley \(2000\)](#).

We use daily data from January 2, 1990 until September 30, 2003. Each trading day is considered to be $\Delta = 1/252$ after the previous day, regardless of the calendar time passed (i.e., weekends and holidays receive no special consideration). The results for each of the three models are discussed briefly below. Both point estimates and standard errors for each of the three models can be estimated quickly and easily, without the need for simulations; other models can be estimated as easily using the technique outlined above. [Figs. 3 and 4](#) show respectively the SPX and VIX for the period under consideration.

7.1. The Heston model

In [Table 6](#), we report the estimation results for the [Heston \(1993\)](#) model described in Eq. (24), treating volatility Y_t as observed in the form of a proxy (the VIX index), and also using the two-stage estimation procedure, where the univariate CEV model is used in the first stage to construct an integrated volatility proxy. The two methods produce similar results. The Monte Carlo results suggest that the effect of replacing Y_t by the simple implied volatility proxy is quite small—within the asymptotic standard errors based on Fisher's Information matrix. As expected, we find that the correlation parameter ρ

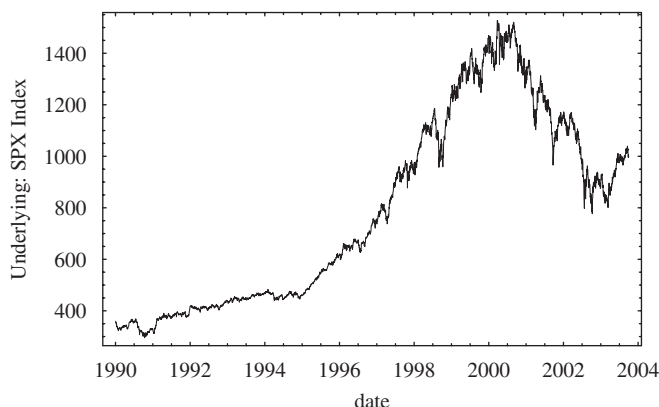


Fig. 3. The SPX (S&P 500) index represents the value of the underlying asset.

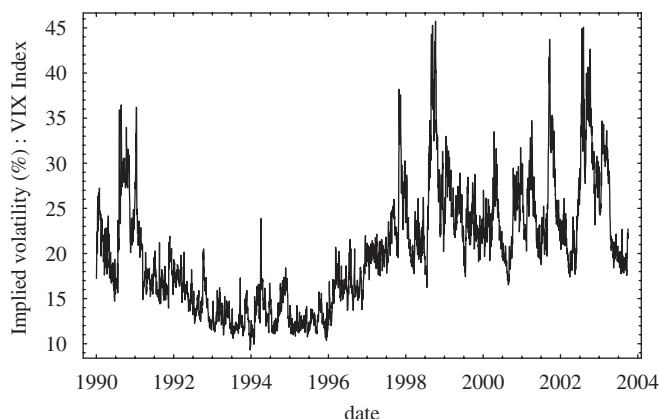


Fig. 4. The VIX index represents the value of the implied volatility of a basket of short-maturity at-the-money options on the S&P 500 index.

between the innovations to stock price and stochastic volatility is strongly negative, hovering around -0.75 . The long-term value of the volatility $\gamma^{1/2}$ is estimated to be approximately 21% per year with a speed of mean reversion coefficient of approximately 5.

The large uncertainty for the risk premia estimates is perhaps not surprising, given that the sample period is 13 years long, and that risk premia are typically poorly estimated even in much longer samples. These parameters pertain to the drift, and the quality of the estimates of drift parameters typically depends only on the length of the sample, and not the sampling frequency. (To take an extreme case, consider an arithmetic or geometric Brownian motion. The volatility can be estimated to an arbitrary degree of precision by sampling frequently enough, but the drift estimate is independent of sampling frequency. The first and last observations provide as good an estimate of the drift as weekly, daily or hourly observations.) Given the length of the available data, there is little that can be done to improve the quality of the λ_1 estimate, apart from waiting for more data to accrue.

Table 6
Parameter estimates for the Heston model

(1)	Heston model for (SPX,VIX)			
	Black-Scholes implied volatility proxy		Integrated volatility proxy	
	Estimate (2)	Standard error (3)	Estimate (4)	Standard error (5)
κ	5.07	0.68	5.13	0.71
γ	0.0457	0.0065	0.0436	0.0065
σ	0.48	0.0036	0.52	0.0033
ρ	−0.767	0.0056	−0.754	0.0054
λ_1	3.9	4.3	3.9	4.1

This table shows the estimated parameter values for the Heston stochastic volatility model using the (SPX,VIX) dataset. Column 2 shows the parameter estimates for daily observations with observed volatility. Standard errors are shown in Column 3. Columns 2 and 3 use the unadjusted VIX data directly as a proxy for instantaneous volatility. Columns 4 and 5 report the corresponding results, also from daily observations, when using our adjusted integrated volatility proxy (constructed from the VIX data) as a proxy for instantaneous volatility.

7.2. The GARCH stochastic volatility model

We now turn to the GARCH stochastic volatility model; see Nelson (1990) and Meddahi (2001). S_t and Y_t follow a process of the following form under the Q measure:

$$d\begin{bmatrix} S_t \\ Y_t \end{bmatrix} = \begin{bmatrix} (r - d)S_t \\ \kappa'(\gamma' - Y_t) \end{bmatrix} dt + \begin{bmatrix} \sqrt{(1 - \rho^2)Y_t}S_t & \rho\sqrt{Y_t}S_t \\ 0 & \sigma Y_t \end{bmatrix} d\begin{bmatrix} W_1^Q(t) \\ W_2^Q(t) \end{bmatrix}. \tag{38}$$

Note that Y_t , which we take to be positive, has a boundary at zero, and this boundary is never achieved as long as $\kappa'\gamma' \geq 0$. The dynamics of $s_t = \ln S_t$ under Q are independent of the stock level

$$d\begin{bmatrix} s_t \\ Y_t \end{bmatrix} = \begin{bmatrix} r - d - \frac{1}{2}Y_t \\ \kappa'(\gamma' - Y_t) \end{bmatrix} dt + \begin{bmatrix} \sqrt{(1 - \rho^2)Y_t} & \rho\sqrt{Y_t} \\ 0 & \sigma Y_t \end{bmatrix} d\begin{bmatrix} W_1^Q(t) \\ W_2^Q(t) \end{bmatrix}. \tag{39}$$

With the VIX series as our volatility proxy, we assume that the market price of risk of the volatility state variable is zero, and that for the stock price is proportional to the stock price itself, and to the square root of the volatility state variable, or $\Lambda = [\lambda_1\sqrt{(1 - \rho^2)Y_t}, 0]'$. With these assumptions, the dynamics of the state variables under the measure P are then

$$d\begin{bmatrix} s_t \\ Y_t \end{bmatrix} = \begin{bmatrix} r - d + (\lambda_1(1 - \rho^2) - \frac{1}{2})Y_t \\ \kappa(\gamma - Y_t) \end{bmatrix} dt + \begin{bmatrix} \sqrt{(1 - \rho^2)Y_t} & \rho\sqrt{Y_t} \\ 0 & \sigma Y_t \end{bmatrix} d\begin{bmatrix} W_1^P(t) \\ W_2^P(t) \end{bmatrix}, \tag{40}$$

where $\kappa' = \kappa$ and $\gamma' = \gamma$. The Q -measure dynamics are not affine in Y_t but, as noted earlier, this is not a problem for our technique which is applicable to all diffusion

specifications. The log-likelihood formula corresponding to this model is given in the appendix.

In Table 7, we report the estimation results for this model. Compared to the Heston model, the unconditional mean of volatility is estimated at a much higher value in this model, or about 27% (note, however, that the standard error for this parameter is large). The point estimates for the correlation parameter are roughly similar to those in the Heston model. The volatility of volatility and risk premia parameters are not directly comparable, owing to the differing model specifications.

7.3. The CEV stochastic volatility model

In Table 8, we report the estimation results for the general CEV model described above, using the two-stage estimation procedure, where the first-stage estimates are used to construct an integrated volatility proxy. From the first-stage estimation, the proxy is given by

$$Y_t = -0.0061 + 1.1308\text{VIX}_t, \quad (41)$$

where VIX_t is the implied variance of the short-maturity at-the-money option. At an implied variance of 0.0463, the proxy is equal to the implied variance; at higher values, the

Table 7
Parameter estimates for the GARCH model

Parameter	GARCH model for (SPX,VIX)	
	Estimate	Standard error
κ	1.62	1.1
γ	0.074	0.04
σ	2.204	0.016
ρ	-0.754	0.0056
λ_1	2.4	3.8

This table shows the estimated parameter values for the GARCH stochastic volatility model using the (SPX,VIX) dataset. The results are for daily observation frequency. Standard errors are in the right column.

Table 8
Parameter estimates for the CEV model

Parameter	CEV model for (SPX,VIX)		CEV univariate model for VIX	
	Estimate	Standard error	Estimate	Standard error
(1)	(2)	(3)	(4)	(5)
κ	4.1031	0.89	2.2	0.92
γ	0.0451	0.009	0.0528	0.016
σ	0.8583	0.012	1.79	0.063
β	0.6545	0.0026	0.94	0.0097
ρ	-0.760	0.005	—	—
λ_1	3.9	4.1	—	—

Columns 2 and 3 show the estimated parameter values and standard errors for the general CEV stochastic volatility model using the (SPX,VIX) dataset at the daily frequency. Columns 4 and 5 show the results for the univariate CEV stochastic volatility model fitted to the integrated volatility proxy.

proxy exceeds implied variance by about 13% of the difference between implied variance and 0.0463; similarly, at lower levels, the proxy is lower than implied variance by a comparable amount. From the second-stage estimates, shown in Table 8, the instantaneous standard deviation of the stock price at the unconditional mean of the volatility state variable is approximately 0.21, which is about the same as in the Heston model, but much lower than that obtained for the GARCH model. Of particular interest for the CEV model is the exponent β , which is estimated at 0.65, above the Heston value of 0.5 but below the GARCH value of 1. Either value can be rejected at the conventional 95% confidence level (see below). This finding stands in contrast to that of Jones (2003) who, using a Bayesian method, estimates this exponent above the GARCH value of 1. Note however that the univariate estimate of β from fitting our integrated volatility proxy as a univariate process, namely the second equation in (35), produces a higher estimate of β , 0.94. The first-stage estimate of β from the unadjusted VIX data is right at the boundary value of 1. Under the assumptions of the model, this univariate estimate is legitimate, as the second equation in (35) is the proper diffusion marginal for the variable Y_t . The discrepancy between the point estimates obtained from the full (SPX,VIX) dataset and the marginal VIX dataset is evidence of some degree of misspecification of the CEV model.

The point estimate for the correlation coefficient is roughly the same as in the other models. This parameter is consistently estimated at around -0.75 for the three models under consideration. One consequence is that theoretical developments—whether theoretical option pricing models or econometric methods for volatility estimation—that assume that the two Brownian motions W_1^P and W_2^P are uncorrelated are likely to be badly misspecified.

In Fig. 5, we report the implied volatility smile predicted by the CEV model at the parameter values just estimated using the joint (SPX,VIX) data. The volatility state variable is estimated at its unconditional mean γ . The results look broadly consistent in shape and magnitude with the typical patterns documented for the implied volatility smile as in e.g., Aït-Sahalia and Lo (1998).

Finally, we compute tests for parameter stability across subsamples, and the tests reject the null hypothesis of equal parameter values. It is not clear however that finding statistically significant differences in parameter estimates across different time periods should be interpreted (in economic terms as opposed to statistical terms) as evidence against the model instead of evidence in favor of regime switches in volatility, similar to what is observed empirically with interest rates. Fig. 4 shows clearly identifiable volatility peaks and low-volatility periods.

Similar simulations (not shown) for the CEV model show a significant bias in the estimation of the elasticity parameter. We therefore examine the effect of using the integrated volatility proxy with simulations based on the CEV model (which nests two of the other models used). Table 4 shows the parameter estimates and standard errors obtained using the two-stage procedure described in Section 4. As shown, the use of the integrated volatility proxy does not result in a substantial degradation in the quality of the estimates; although some measures have become worse, others have improved, and the differences are usually quite small.

7.4. Likelihood ratio tests

The CEV model nests the Heston ($\beta = 1/2$) and GARCH ($\beta = 1$) models. The use of likelihood estimation makes it straightforward to calculate likelihood ratio statistics for the

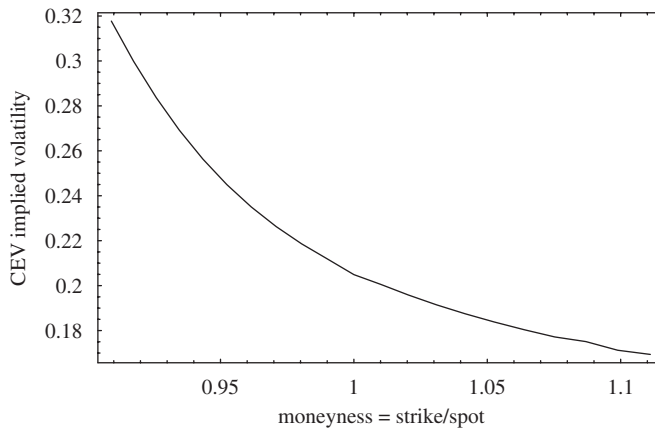


Fig. 5. This plot reports the implied volatility smile predicted by fitting the CEV model to the (SPX,VIX) data, fitted using our adjusted integrated volatility proxy as described in the text. The state variables are held fixed at their unconditional estimated means for the purpose of computing this implied volatility smile.

nested models. The statistics for both models tested against the CEV model at the daily frequency are 782 and 122, respectively, using as above the (SPX,VIX) dataset and our proxy construction. To make the likelihoods directly comparable, all estimates are based on the two-stage procedure with the CEV model used in the first stage to construct the integrated volatility proxy. By this method, all three models are then using the same data as input to the second-stage estimation. Both nested models are easily rejected at the conventional 95% confidence level. In both cases, the likelihood ratio statistic is many times (and in one case, hundreds of times) the 95% cutoff value of 3.84, implying strong rejection at any reasonable confidence level.

The point estimate of the β coefficient lies between the Heston value of $1/2$ and the GARCH value of 1. Both of these values are boundary cases in an appropriate sense; for values of β below $1/2$, the boundary of zero is achievable, so that the stock price can be instantaneously deterministic. For values of β above 1, the deflated stock price is a local martingale, but not a martingale, and there exists a replicating portfolio for the stock that is cheaper than the stock itself, see e.g., [Heston et al. \(2004\)](#). Although violation of either bound does not result in arbitrage opportunities, both situations could be considered undesirable modeling properties. The two boundary values of $1/2$ and 1 are commonly used in stochastic volatility models, owing to their tractability, but the point estimates, standard errors, and likelihood ratios suggest that neither boundary value is appropriate, with the elasticity of variance lying between the two. As we show, either boundary value is very strongly rejected.

8. Extensions and conclusions

In conclusion, the main advantage of our approach is twofold: we provide a *maximum likelihood* estimator for the parameters of the underlying model, with all its associated desirable statistical properties, and we do it in *closed form*, fully if an implied volatility proxy (either a simple proxy or our modified proxy) is used, and up to the option pricing model linking the state vector to observed option prices if those are used.

One advantage of our methodology is that it extends readily to the situation where the underlying asset price and/or the volatility state variable(s) can jump. Suppose that, instead of (1), X_t follows under P the dynamics

$$dX_t = \mu^P(X_t) dt + \sigma(X_t) dW_t^P + J_t^P dN_t^P, \quad (42)$$

where the pure jump process N^P has stochastic intensity $\lambda(X_t, \theta)$ and jump size 1. The jump size J_t^P is independent of the filtration generated by the X process at time $t-$, and has probability density $v(\cdot, \theta)$.

This setup incorporates the stochastic volatility with jump models that have been proposed in the literature, such as Bates (2000), Bakshi et al. (1997), and Pan (2002). It is possible to extend the basic likelihood expansion described in Section 3.1 to cover such cases. The expression, due to Yu (2003), is

$$\begin{aligned} p_X^{(J)}(\Delta, x|x_0; \theta) = & \exp\left(-\frac{m}{2}\ln(2\pi\Delta) - D_v(x; \theta) + \frac{c_X^{(-1)}(x|x_0; \theta)}{\Delta}\right) \sum_{k=0}^J c_X^{(k)}(x|x_0; \theta) \frac{\Delta^k}{k!} \\ & + \sum_{k=1}^J d_X^{(k)}(x|x_0; \theta) \frac{\Delta^k}{k!}. \end{aligned} \quad (43)$$

Again, the series can be calculated up to arbitrary order J and the unknowns are the coefficients $c_X^{(k)}$ and $d_X^{(k)}$. The difference between the coefficients $c_X^{(k)}$ in (43) and $C_X^{(k)}$ in (9) is due to the fact that the former is written for $\ln p_X$ while the latter is for p_X itself; the two coefficients families match once the terms of the Taylor series of $\ln(p_X^{(J)})$ in Δ are matched to the coefficients $C_X^{(k)}$ of the direct Taylor series $\ln p_X^{(J)}$. The coefficients $d_X^{(k)}$ are the new terms needed to capture the presence of the jumps in the transition function. The latter terms are needed to capture the different behavior of the tails of the transition density when jumps are present. These tails are not exponential in x , hence the absence of a the factor $\exp(c_X^{(-1)}\Delta^{-1})$ in front of the summation of $d_X^{(k)}$ coefficients. The coefficients can be computed analogously to the pure diffusive case.

We perform Monte Carlo simulations for the Heston and CEV models to assess the accuracy of the technique, and find that it not only produces accurate estimates, but can also be implemented efficiently. Computational time for estimation is of the order of a few minutes on a standard PC using Matlab when volatility is treated as unobserved, and considerably less when a proxy is used. This is a major advantage of our method, in addition to the statistical efficiency of maximum likelihood. When the observed vector consists of $G_t = [S_t; C_t]'$, we can fully identify all the parameters of the model, including the market prices of risk, provided an option pricing technique is included in the estimation procedure. Use of either a simple Black-Scholes implied volatility or our integrated volatility proxy simplifies estimation, but does not permit identification of the market price of risk for the volatility state variable. The asymptotic variances calculated from the approximate likelihood expressions are close to those found empirically from the Monte Carlo simulations. We find that the use of the implied volatility of at-the-money short-maturity options as a proxy for the true stochastic volatility results in reasonable estimates for the Heston model, and the integrated volatility proxy performs similarly for the CEV model. In this case, using such a proxy reduces the exercise to one of simply applying our likelihood expansion to the state vector $X_t = [S_t; Y_t]'$.

We apply our method to the Heston, GARCH, and CEV stochastic volatility models. One of the findings in our empirical analysis across models is the fact that the estimated correlation coefficient ρ between the shocks to the stock level S_t and the volatility variable Y_t is consistently in the neighborhood of -0.75 for all models. This negative correlation has long been noted (in the form of the “leverage effect”). The robustness of this finding across models suggests that stochastic volatility models, pricing and/or estimation methods that rely on the assumption of uncorrelated shocks, such as Hull and White (1987), will be quite unrealistic in this context. Note that while our integrated volatility proxy does not adjust for this correlation, the likelihood itself does and the Monte Carlo results suggest that this is the first-order consideration.

However, little in our estimation procedure depends on the specific properties of these models. It is in fact applicable to a wide variety of diffusion-based stochastic volatility models (where Y_t is one-dimensional, and if a simple proxy for volatility is to be used, has linear drift), or for that matter models with other types of latent variables. In Section 3, we describe our method without reference to any specific model. Provided that traded asset prices (such as the call options we use) or other observable quantities can be found to be mapped into the unobservable latent state vector, or simply using our adjusted volatility proxy, the method can then be employed.

Appendix A. The log-likelihood expansion for stochastic volatility models

In this appendix, we give the coefficients of the likelihood expansion at order $J = 1$ corresponding to each of the models considered. These expressions, as well as higher-order expansions, are available upon request from the authors in computer form.

A.1. The Heston model

At order $J = 1$, with $x_1 = s$ and $x_2 = Y$ and the indirect parameters

$$\begin{aligned} a_1 &= r - d, & a_2 &= \kappa' \gamma', \\ b_1 &= \rho \lambda_2 + (1 - \rho^2) \lambda_1 - \frac{1}{2}, & b_2 &= \lambda_2 \sigma - \kappa', \end{aligned}$$

the expressions for the coefficients appearing in formula (11) are given by

$$\begin{aligned} D_v(x; \theta) &= \frac{1}{2} \ln((1 - \rho^2) \sigma x_2), \\ C_X^{(-1)}(x|x_0; \theta) &= - \frac{(x_2 - x_{20})^2 - 2\rho\sigma(x_2 - x_{20})(x_1 - x_{10}) + \sigma^2(x_1 - x_{10})^2}{2(1 - \rho^2)\sigma^2 x_{20}} \\ &\quad + \frac{(x_2 - x_{20})^3}{4(1 - \rho^2)\sigma^2 x_{20}^2} - \frac{\rho(x_2 - x_{20})^2(x_1 - x_{10})}{2(1 - \rho^2)\sigma x_{20}^2} \\ &\quad + \frac{(x_2 - x_{20})(x_1 - x_{10})^2}{4(1 - \rho^2)x_{20}^2} + \frac{(7\rho - 8\rho^3)(x_2 - x_{20})^3(x_1 - x_{10})}{24(1 - \rho^2)^2 \sigma x_{20}^3} \\ &\quad - \frac{(7 - 10\rho^2)(x_2 - x_{20})^2(x_1 - x_{10})^2}{48(1 - \rho^2)^2 x_{20}^3} \\ &\quad - \frac{\rho\sigma(x_2 - x_{20})(x_1 - x_{10})^3}{24(1 - \rho^2)^2 x_{20}^3} + \frac{\sigma^2(x_1 - x_{10})^4}{96(1 - \rho^2)^2 x_{20}^3} - \frac{(15 - 16\rho^2)(x_2 - x_{20})^4}{96(1 - \rho^2)^2 \sigma^2 x_{20}^3}, \end{aligned}$$

$$\begin{aligned}
C_X^{(0)}(x|x_0; \theta) &= \frac{(x_2 - x_{20})(x_{20}b_2 - \rho\sigma x_{20}b_1 - \rho\sigma a_1 + a_2)}{(1 - \rho^2)\sigma^2 x_{20}} \\
&\quad - \frac{(x_1 - x_{10})(\rho x_{20}b_2 - \sigma x_{20}b_1 - \sigma a_1 + \rho a_2)}{(1 - \rho^2)\sigma x_{20}} \\
&\quad - \frac{\sigma^2(x_1 - x_{10})^2}{24(1 - \rho^2)x_{20}^2} - \frac{(x_2 - x_{20})^2(\sigma(\sigma - 12\rho a_1) + 12a_2)}{24(1 - \rho^2)\sigma^2 x_{20}^2} \\
&\quad + \frac{(x_2 - x_{20})(x_1 - x_{10})(\rho\sigma^2 - 6\sigma a_1 + 6\rho a_2)}{12\sigma x_{20}^2(1 - \rho^2)}, \\
C_X^{(1)}(x|x_0; \theta) &= \frac{\rho\sigma a_2 b_1 + \rho\sigma a_1 b_2 - \sigma^2 a_1 b_1 - a_2 b_2}{(1 - \rho^2)\sigma^2} + \frac{(2\rho\sigma b_1 b_2 - \sigma^2 b_1^2 - b_2^2)x_{20}}{2(1 - \rho^2)\sigma^2} \\
&\quad - \frac{\sigma^4 - \rho^2\sigma^4 + 6\sigma^2 a_1^2 - 6\sigma^2 a_2 + 6\rho^2\sigma^2 a_2 - 12\rho\sigma a_1 a_2 + 6a_2^2}{12(1 - \rho^2)\sigma^2 x_{20}}.
\end{aligned}$$

Note that the zero boundary of the state variable x_2 depends here upon the values of the parameters. This is due to the fact that x_2 in this model follows a square root process which is the limiting (and only) case for which such a phenomenon occurs. This can be seen through the presence in the coefficients of terms x_{20}^{-n} where n is an integer. Thus the behavior of the likelihood expansion near such a boundary is specified exogenously to match that of the assumed model—the unattainability of the zero boundary in this case—in the limit where x_{20} tends to zero; this is achieved by setting the log-likelihood expansion to an arbitrarily high negative value.

A.2. The GARCH stochastic volatility model

At order $J = 1$, with $x_1 = s$ and $x_2 = Y$, the expressions for the coefficients appearing in formula (11) are given by

$$D_v(x; \theta) = \frac{1}{2} \ln(x_2^3(1 - \rho^2)\sigma^2),$$

$$\begin{aligned}
C_X^{(-1)}(x|x_0; \theta) &= \frac{(45\rho^2 - 44)(x_2 - x_{20})^4}{96(1 - \rho^2)^2\sigma^2 x_{20}^4} + \frac{(14\rho - 15\rho^3)(x_2 - x_{20})^3(x_1 - x_{10})}{24(1 - \rho^2)^2\sigma x_{20}^{7/2}} \\
&\quad - \frac{3\rho(x_2 - x_{20})^2(x_1 - x_{10})}{4\sigma(1 - \rho^2)x_{20}^{5/2}} + \frac{(x_2 - x_{20})^3}{2(1 - \rho^2)\sigma^2 x_{20}^3} \\
&\quad + \frac{(-8 + 11\rho^2)(x_2 - x_{20})^2(x_1 - x_{10})^2}{48(1 - \rho^2)^2 x_{20}^3} + \frac{(x_2 - x_{20})(x_1 - x_{10})^2}{4(1 - \rho^2)x_{20}^2} \\
&\quad - \frac{\rho\sigma(x_2 - x_{20})(x_1 - x_{10})^3}{24(1 - \rho^2)^2 x_{20}^{5/2}} + \frac{\sigma^2(x_1 - x_{10})^4}{96(1 - \rho^2)^2 x_{20}^2} \\
&\quad - \frac{x_2^2 - 2x_2\sqrt{x_{20}}(\sqrt{x_{20}} + \rho\sigma(x_1 - x_{10})) + x_{20}(x_{20} + 2\rho\sigma\sqrt{x_{20}}(x_1 - x_{10}) + \sigma^2(x_1 - x_{10})^2)}{2(1 - \rho^2)\sigma^2 x_{20}^2},
\end{aligned}$$

$$\begin{aligned}
C_X^{(0)}(x|x_0; \theta) &= -\frac{(x_2 - x_{20})(-40\kappa + 4(r-d)\rho\sigma\sqrt{x_{20}} + (4\kappa + \sigma^2)x_{20} + 2\rho\sigma x_{20}^{3/2}(2\sqrt{1-\rho^2}\lambda_1 - 1))}{4(1-\rho^2)\sigma^2 x_{20}^2} \\
&+ \frac{(x_2 - x_{20})^2(-48\theta\kappa + 36(r-d)\rho\sigma\sqrt{x_{20}} + 24\kappa x_{20} + 5\sigma^2 x_{20} + 6\rho\sigma x_{20}^{3/2}(2\sqrt{1-\rho^2}\lambda_1 - 1))}{48(1-\rho^2)\sigma^2 x_{20}^3} \\
&- \frac{(x_1 - x_{10})(x_2 - x_{20})(-36\theta\kappa\rho + 24(r-d)\sigma\sqrt{x_{20}} + \rho(12\kappa + \sigma^2)x_{20})}{48(1-\rho^2)\sigma x_{20}^{5/2}} - \frac{\sigma^2(x_1 - x_{10})^2}{48x_{20}(1-\rho^2)} \\
&+ \frac{(x_1 - x_{10})(-4\theta\kappa\rho + 4(r-d)\sigma\sqrt{x_{20}} + \rho(4\kappa + \sigma^2)x_{20} + 2\sigma x_{20}^{3/2}(2\sqrt{1-\rho^2}\lambda_1 - 1))}{4(1-\rho^2)\sigma x_{20}^{3/2}}, \\
C_X^{(1)}(x|x_0; \theta) &= -\frac{x_{20}(1 + 4\lambda_1^2(1-\rho^2))}{8(1-\rho^2)} - \frac{\gamma^2\kappa^2}{2(1-\rho^2)\sigma^2 x_{20}^2} + \frac{(r-d)\gamma\kappa\rho}{(1-\rho^2)\sigma x_{20}^{3/2}} + \frac{\rho(4\kappa + \sigma^2)\sqrt{x_{20}}}{8(1-\rho^2)\sigma} \\
&- \frac{\rho(2\gamma\kappa - d(4\kappa + \sigma^2) + r(4\kappa + \sigma^2))}{4(1-\rho^2)\sigma\sqrt{x_{20}}} \\
&- \frac{(4(r-d)\sigma\sqrt{x_{20}} + \rho(4\kappa + \sigma^2)x_{20} - 2\sigma x_{20}^{3/2} - 4\gamma\kappa\rho)\lambda_1}{4(1-\rho^2)\sigma\sqrt{x_{20}}} \\
&+ \frac{\gamma\kappa(4\kappa + (4-3\rho^2)\sigma^2) - 2(r-d)^2\sigma^2}{4(1-\rho^2)\sigma^2 x_{20}} \\
&- \frac{48\kappa^2 + 24(2(d-r+\kappa) - \kappa\rho^2)\sigma^2 + (13-10\rho^2)\sigma^4}{96(1-\rho^2)\sigma^2}.
\end{aligned}$$

Note that the behavior at the zero boundary of the state variable x_2 is unattainable, provided $\kappa\gamma$ is non-negative. The log-likelihood is set to an arbitrarily high negative value when this condition is violated.

A.3. The CEV model

At order $J = 1$, with $x_1 = s$ and $x_2 = Y$ and the indirect parameters

$$a = r - d, \quad b = (1 - \rho^2)\lambda_1 - \frac{1}{2},$$

the expressions for the coefficients appearing in formula (11) are given by

$$D_v(x; \theta) = \frac{1}{2} \ln(x_2^{1+2\beta}(1-\rho^2)\sigma^2),$$

$$\begin{aligned}
C_X^{(-1)}(x|x_0; \theta) &= \frac{x_{20}^{-4+2\beta}\sigma^2(x_1 - x_{10})^4}{96(1-\rho^2)^2} + \frac{(x_1 - x_{10})^2(x_2 - x_{20})}{4x_{20}^2(1-\rho^2)} - \frac{x_{20}^{-(7/2)+\beta}\rho\sigma(x_1 - x_{10})^3(x_2 - x_{20})}{24(1-\rho^2)^2} \\
&- \frac{(1+2\beta)x_{20}^{-(3/2)-\beta}\rho(x_1 - x_{10})(x_2 - x_{20})^2}{4(1-\rho^2)\sigma} + \frac{(9\rho^2 - 6 - 2\beta(1-\rho^2))(x_1 - x_{10})^2(x_2 - x_{20})^2}{48x_{20}^3(1-\rho^2)^2}
\end{aligned}$$

$$\begin{aligned}
& + \frac{\beta x_{20}^{-1-2\beta}(x_2 - x_{20})^3}{2(1 - \rho^2)\sigma^2} \\
& - \frac{x_{20}^{-(5/2)-\beta} \rho(3\rho^2 - 2 - 8\beta(1 - \rho^2) - 4\beta^2(1 - \rho^2))(x_1 - x_{10})(x_2 - x_{20})^3}{24(1 - \rho^2)^2\sigma} \\
& + \frac{(\rho^2 - 16\beta(1 - \rho^2) - 28\beta^2(1 - \rho^2))(x_2 - x_{20})^4}{96x_{20}^{2(1+\beta)}(1 - \rho^2)^2\sigma^2} \\
& - \frac{x_{20}^{-1-2\beta}(x_{20}^3 + 2x_{20}^{3/2+\beta}\rho\sigma(x_1 - x_{10}) + x_{20}^{2\beta}\sigma^2(x_1 - x_{10})^2 - 2(x_{20}^2 + x_{20}^{1/2+\beta}\rho\sigma(x_1 - x_{10}))x_2 + x_{20}x_2^2)}{2(1 - \rho^2)\sigma^2}, \\
\\
C_X^{(0)}(x|x_0; \theta) \\
& = - \frac{x_{20}^{-3/2-\beta}(4\gamma x_{20}\kappa\rho - 4x_{20}^2\kappa\rho - 4ax_{20}^{1/2+\beta}\sigma - 4bx_{20}^{3/2+\beta}\sigma + (1 - 2\beta)x_{20}^{2\beta}\rho\sigma^2)(x_1 - x_{10})}{4(1 - \rho^2)\sigma} \\
& + \frac{x_{20}^{-1-2\beta}(4\gamma x_{20}\kappa - 4x_{20}^2\kappa - 4ax_{20}^{1/2+\beta}\rho\sigma - 4bx_{20}^{3/2+\beta}\rho\sigma + (1 - 2\beta)x_{20}^{2\beta}\sigma^2)(x_2 - x_{20})}{4(1 - \rho^2)\sigma^2} \\
& + \frac{(2\beta - 3)x_{20}^{-3+2\beta}\sigma^2(x_1 - x_{10})^2}{48(1 - \rho^2)} \\
& + \frac{x_{20}^{-5/2-\beta}(12(1 + 2\beta)\gamma x_{20}\kappa\rho + 12(1 - 2\beta)x_{20}^2\kappa\rho - 24ax_{20}^{1/2+\beta}\sigma + (15 - 28\beta + 12\beta^2)x_{20}^{2\beta}\rho\sigma^2)(x_1 - x_{10})(x_2 - x_{20})}{48(1 - \rho^2)\sigma} \\
& - \frac{(48\beta\gamma x_{20}\kappa + 24(1 - 2\beta)x_{20}^2\kappa - 12a(1 + 2\beta)x_{20}^{1/2+\beta}\rho\sigma + 12b(1 - 2\beta)x_{20}^{3/2+\beta}\rho\sigma + (9 - 14\beta)x_{20}^{2\beta}\sigma^2)(x_2 - x_{20})^2}{48x_{20}^{2(1+\beta)}(1 - \rho^2)\sigma^2}, \\
\\
C_X^{(1)}(x|x_0; \theta) \\
& = \frac{(3 - 28\beta + 12\beta^2 - 6\rho^2 + 40\beta\rho^2 - 24\beta^2\rho^2)\sigma^2 x_{20}^{-2+2\beta}}{96(1 - \rho^2)} - \frac{\kappa^2(x_{20} - \gamma)^2}{2(1 - \rho^2)\sigma^2 x_{20}^{2\beta}} \\
& + \frac{(1 - 2\beta)\rho\sigma x_{20}^{-3/2+\beta}(a + bx_{20})}{4(1 - \rho^2)} - \frac{\kappa\rho x_{20}^{-1/2-\beta}(x_{20} - \gamma)(a + bx_{20})}{(1 - \rho^2)\sigma} \\
& - \frac{2a^2 - 4\beta\gamma\kappa + \gamma\kappa\rho^2 + 2\beta\gamma\kappa\rho^2 + 4abx_{20} - 2\kappa x_{20} + 4\beta\kappa x_{20} + \kappa\rho^2 x_{20} - 2\beta\kappa\rho^2 x_{20} + 2b^2 x_{20}^2}{4(1 - \rho^2)x_{20}}.
\end{aligned}$$

Note that the boundary at zero of the state variable x_2 cannot be achieved if $\beta > \frac{1}{2}$ and $2\kappa\gamma \geq \sigma^2$; whether the boundary is attainable when $\beta = 1/2$ depends on the other parameter values (see the discussion above for the Heston model).

References

- Aït-Sahalia, Y., 1999. Transition densities for interest rate and other nonlinear diffusions. *Journal of Finance* 54, 1361–1395.
- Aït-Sahalia, Y., 2001. Closed-form likelihood expansions for multivariate diffusions. Working Paper, Princeton University.
- Aït-Sahalia, Y., 2002. Maximum-likelihood estimation of discretely-sampled diffusions: a closed-form approximation approach. *Econometrica* 70, 223–262.
- Aït-Sahalia, Y., Lo, A., 1998. Nonparametric estimation of state-price-densities implicit in financial asset prices. *Journal of Finance* 53, 499–547.
- Aït-Sahalia, Y., Mykland, P.A., 2003. The effects of random and discrete sampling when estimating continuous-time diffusions. *Econometrica* 71, 483–549.
- Andersen, T.G., Lund, J., 1997. Estimating continuous time stochastic volatility models of the short term interest rate. *Journal of Econometrics* 77, 343–378.

- 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. Do call prices and the underlying stock always move in the same direction? *Review of Financial Studies* 13, 549–584.
- Bates, D.S., 2000. Post-'87 crash fears in the S&P 500 futures option market. *Journal of Econometrics* 94, 181–238.
- Bates, D.S., 2002. Maximum likelihood estimation of latent affine processes. Working Paper, University of Iowa.
- Black, F., 1976. Studies of stock price volatility changes. In: *Proceedings of the 1976 Meetings of the American Statistical Association*, pp. 171–181.
- Black, F., Scholes, M., 1973. The pricing of options and corporate liabilities. *Journal of Political Economy* 81, 637–654.
- Bollerslev, T., Zhou, H., 2002. Estimating stochastic volatility diffusions using conditional moments of integrated volatility. *Journal of Econometrics* 109, 33–65.
- Carr, P., Madan, D.B., 1998. Option valuation using the fast Fourier transform. *Journal of Computational Finance* 2, 61–73.
- Chacko, G., Viceira, L.M., 2003. Spectral GMM estimation of continuous-time processes. *Journal of Econometrics* 116, 259–292.
- Derman, E., Kani, I., 1994. Riding on the smile. *RISK* 7, 32–39.
- Dumas, B., Fleming, J., Whaley, R.E., 1998. Implied volatility functions: empirical tests. *Journal of Finance* 53, 2059–2106.
- Dupire, B., 1994. Pricing with a smile. *RISK* 7, 18–20.
- Eraker, B., 2001. MCMC analysis of diffusion models with application to finance. *Journal of Business and Economic Statistics* 19, 177–191.
- Feller, W., 1951. Two singular diffusion problems. *Annals of Mathematics* 54, 173–182.
- Gallant, A.R., Tauchen, G., 1996. Which moments to match? *Econometric Theory* 12, 657–681.
- Harrison, M., Kreps, D., 1979. Martingales and arbitrage in multiperiod securities markets. *Journal of Economic Theory* 20, 381–408.
- Harrison, M., Pliska, S., 1981. Martingales and stochastic integrals in the theory of continuous trading. *Stochastic Processes and Their Applications* 11, 215–260.
- Heston, S., 1993. A closed-form solution for options with stochastic volatility with applications to bonds and currency options. *Review of Financial Studies* 6, 327–343.
- Heston, S., Loewenstein, M., Willard, G.A., 2004. Options and bubbles. Technical Report, University of Maryland.
- Hull, J., White, A., 1987. The pricing of options on assets with stochastic volatilities. *Journal of Finance* 42, 281–300.
- Hurn, A.S., Jeisman, J., Lindsay, K., 2005. Seeing the wood for the trees: a critical evaluation of methods to estimate the parameters of stochastic differential equations. Working Paper, School of Economics and Finance, Queensland University of Technology.
- Jacquier, E., Polson, N.G., Rossi, P.E., 1994. Bayesian analysis of stochastic volatility models. *Journal of Business and Economic Statistics* 14, 429–434.
- Jensen, B., Poulsen, R., 2002. Transition densities of diffusion processes: numerical comparison of approximation techniques. *Journal of Derivatives* 9, 1–15.
- Jiang, G.J., Knight, J., 2002. Estimation of continuous-time processes via empirical characteristic function. *Journal of Business and Economic Statistics* 20, 198–212.
- Jones, C.S., 2003. The dynamics of stochastic volatility: evidence from underlying and options markets. *Journal of Econometrics* 116, 181–224.
- Kim, S., Shephard, N., Chib, S., 1999. Stochastic volatility: likelihood inference and comparison with ARCH models. *Review of Economic Studies* 65, 361–393.
- Lamoureux, C.G., Paseka, A., 2004. Information in option prices and the underlying asset dynamics. Working Paper, University of Arizona.
- Ledoit, O., Santa-Clara, P., Yan, S., 2002. Relative pricing of options with stochastic volatility. Working Paper, University of California at Los Angeles.
- Lewis, A.L., 2000. *Option Valuation under Stochastic Volatility*. Finance Press, Newport Beach, CA.
- Li, M., Pearson, N.D., Poteshman, A.M., 2004. Conditional estimation of diffusion processes. *Journal of Financial Economics* 74, 31–66.
- Meddahi, N., 2001. An eigenfunction approach for volatility modeling. Technical Report, Université de Montréal.

- Merton, R.C., 1973. The theory of rational option pricing. *Bell Journal of Economics and Management Science* 4, 141–183.
- Nelson, D.B., 1990. ARCH models as diffusion approximations. *Journal of Econometrics* 45, 7–38.
- Pan, J., 2002. The jump-risk premia implicit in options: evidence from an integrated time-series study. *Journal of Financial Economics* 63, 3–50.
- Romano, M., Touzi, N., 1997. Contingent claims and market completeness in a stochastic volatility model. *Mathematical Finance* 7, 399–412.
- Rubinstein, M., 1995. As simple as one, two, three. *RISK* 8, 44–47.
- Ruiz, E., 1994. Quasi-maximum likelihood estimation of stochastic volatility models. *Journal of Econometrics* 63, 289–306.
- Singleton, K., 2001. Estimation of affine asset pricing models using the empirical characteristic function. *Journal of Econometrics* 102, 111–141.
- Stein, E.M., Stein, J.C., 1991. Stock price distributions with stochastic volatility: an analytic approach. *Review of Financial Studies* 4, 727–752.
- Stein, J.C., 1989. Overreactions in the options market. *Journal of Finance* 44, 1011–1023.
- Stramer, O., Yan, J., 2005. On simulated likelihood of discretely observed diffusion processes and comparison to closed-form approximation. Working Paper, University of Iowa.
- Takamizawa, H., Shoji, I., 2003. Modeling the term structure of interest rates with general short-rate models. *Finance and Stochastics* 7, 323–335.
- Taylor, S.J., 1994. Modeling stochastic volatility: a review and comparative study. *Mathematical Finance* 4, 183–204.
- Whaley, R.E., 2000. The investor fear gauge. *The Journal of Portfolio Management* 26, 12–17.
- Yu, J., 2003. Closed-form likelihood estimation of jump-diffusions with an application to the realignment risk premium of the Chinese yuan. Ph.D. Thesis, Princeton University.