NBER WORKING PAPER SERIES

MAXIMUM LIKELIHOOD ESTIMATION OF STOCHASTIC VOLATILITY MODELS

Yacine Ait-Sahalia
Robert Kimmel

Working Paper 10579
http://www.nber.org/papers/w10579

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2004

Financial support from the NSF under grant SES-0111140 is gratefully acknowledged. The views expressed herein are those of the author(s) and not necessarily those of the National Bureau of Economic Research.

Maximum Likelihood Estimation of Stochastic Volatility Models
Yacine Ait-Sahalia and Robert Kimmel
NBER Working Paper No. 10579
June 2004
JEL No. G0

ABSTRACT

We develop and implement a new 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 the implied volatility of a short dated at-the-money option. We find that the approximation results in a negligible loss of accuracy. 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 model of Heston (1993) and a GARCH model, suggests that the elasticity of variance of volatility lies between that assumed by the two nested models.

Yacine Ait-Sahalia
Department of Economics
and Bendheim Center for Finance
Princeton University
Princeton, NJ 08540
and NBER
yacine@princeton.edu

Robert Kimmel
Department of Economics
and Bendheim Center for Finance
Princeton University
Princeton, NJ 08540
rkimmel@princeton.edu.

1. Introduction

In this paper, we develop and implement a new 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 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, for example, Stein (1989), Aït-Sahalia and Lo (1998) or Bakshi et al. (2000). One 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. These state-dependent, time-varying, volatility models represent a limited form of stochastic volatility; the stock price still follows a (time-inhomogeneous) Markov process. Models of this type include Derman and Kani (1994), Dupire (1994), and Rubinstein (1995). Such 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. 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 which 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.[1]

Alternately, the value of the additional state variables can be extracted from the observed prices of options.

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

This extraction can be through an approximation technique, such as that of Ledoit et al. (2002), in which the implied volatility (under the lognormal assumptions of Black and Scholes (1973)) of an at-the-money short-maturity option is taken as a proxy for the instantaneous volatility (under the stochastic volatility model) of the stock price. A more difficult, but 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, for example, 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, 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 has typically been simulation-based, relying either on Bayesian methods (as in Jacquier et al. (1994), Kim et al. (1999) and Eraker (2001)) or on the efficient method of moments of Gallant and Tauchen (1996).

In this paper, we develop a new 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 (2002) and Aït-Sahalia (2001), who develops series approximations to the likelihood function for arbitrary multivariate continuous time diffusions at discrete intervals of observations. This technique has been shown to be very accurate, even when the series are truncated after only a few terms, for a variety of diffusion models (see Aït-Sahalia (1999) and Jensen and Poulsen (2002)).

In all cases, we rely on observations on the joint time series of the underlying asset price and either an option price or a short dated at the money implied volatility. 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 as 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 much 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).

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 is used, and up to the option pricing model linking the state vector to observed option prices if those are used.

The closed form feature offers considerable benefits: for example, estimation is quick enough that large numbers of Monte Carlo simulations can be run to test its accuracy, as we do 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 well predicted by standard statistical asymptotic theory, as it applies to the maximum likelihood estimator.

We illustrate our method using several typical models, including the affine model of Heston (1993), and a GARCH model (see, for example, Meddahi (2001)), a lognormal model (see, for example, Scott (1987), Wiggins (1987), Chesney and Scott (1989), Scott (1991), and Andersen et al. (2002b)), and a CEV model (see, for example, Jones (2003)).[2] However, it is also important to note that our technique is applicable to arbitrary diffusion-based stochastic volatility models; the only requirement is that the model (i.e., its risk premia, etc.) be sufficiently tractable for option prices to be mapped into the state variables.

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 technique 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 technique to the four models cited above, developing the explicit closed-form likelihood expressions, and extracting the state vector from option prices or directly using an implied volatility proxy. Section 5 tests the accuracy of our technique by performing Monte Carlo simulations for the model of Heston (1993), assessing in particular the accuracy of the estimates, the degree to which their sampling distributions conform to asymptotic theory and the effect of using an implied volatility proxy in lieu of option prices. In Section 6, we apply our technique to real index option prices for four different stochastic volatility models, and analyze and compare the results. Section 7 shows how to extend the method to jump-diffusions. Finally, Section 8 concludes.

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 may just as easily be applied to other classes of financial assets, such as, for example, foreign currencies or futures contracts. A

[2] An early summary of some of the models we use as examples, as well as several others, may be found in Taylor (1994).

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:

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

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 $X_t = [S_t; Y_t]^T$, with Y_t a $N-$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:

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

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:

$$dS_t = (r_t - d_t)\,S_t dt + \sigma_1(X_t)\,dW_t^Q \tag{3}$$

where d_t is the instantaneous dividend yield on the stock 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:

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

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 $\sigma(X_t)\,\sigma^T(X_t)$. The price of a derivative security with a European-style exercise convention must satisfy the boundary condition:

$$\phi(T, X_T) = g(X_T) \tag{5}$$

[3] The definition of admissibility appearing 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.

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:

$$g(X_T) = h(S_T) \tag{6}$$

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

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

$$\sigma_1(X_t)\,\sigma_1^T(X_t) = \sigma_S(S_t) \tag{7}$$

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 equation (6) can be expressed as $\phi(t, X_t) = \xi(t, S_t)$ and equation (4) simplifies to:

$$\frac{\partial \xi(t, S_t)}{\partial t} + \frac{\partial \xi(t, S_t)}{\partial S_t}(r_t - d_t)\,S_t + \frac{1}{2}\frac{\partial^2 \xi(t, S_t)}{\partial S_t}\sigma_S^2(S_t) - r_t \xi(t, S_t) = 0 \tag{8}$$

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, knowledge of S_t and the parameters of the model are sufficient to price any derivative with final payoff of the type in equation (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 may be relevant.) Models of this type usually allow explicit time dependency by replacing $\sigma_S(S_t)$ with $\sigma_S(t, S_t)$; see, for example, 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 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 reflect accurately 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.

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 equation (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 5, volatility depends on a single state variable (i.e., Y_t has a single element).

Although stochastic volatility models offer considerable advantages in modeling, they do present some estimation challenges. The next section presents a method for performing maximum likelihood estimation of a stochastic volatility model for equity prices.

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 only on 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, it 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. As an alternative, one can use the method of Ledoit et al. (2002), in which the Black-Scholes implied volatility of an at-the-money short-maturity option is taken as a proxy for the instantaneous volatility of the stock, can be applied. 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 in Section 6 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 employ the closed-form approximation technique of Aït-Sahalia (2001) which yields to us in closed form the joint likelihood function of $[S_t; Y_t]^T$. From there, the joint likelihood function of the observations on $G_t = [S_t; C_t]^T$ is obtained simply by multiplying the likelihood of $X_t = [S_t; Y_t]^T$ by a Jacobian term. (If the approximation method of Ledoit et al. (2002) 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; 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 $X_t = [S_t; Y_t]^T$ 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:

$$l_X^{(K)}\left(\Delta, x|x_0;\theta\right) = -\frac{m}{2}\ln\left(2\pi\Delta\right) - D_v\left(x;\theta\right) + \frac{C_X^{(-1)}\left(x|x_0;\theta\right)}{\Delta} + \sum\nolimits_{k=0}^{K} C_X^{(k)}\left(x|x_0;\theta\right)\frac{\Delta^k}{k!}. \tag{9}$$

where

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

and $v\left(x\right) \equiv \sigma\left(x\right)\sigma^T\left(x\right)$. The series can be calculated up to arbitrary order K. The unknowns so far are the coefficients $C_X^{(k)}$ corresponding to each Δ^k, $k = -1, 0, ..., K$. 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 $j_k = 2(K - k)$.

The resulting expansion is then:

$$\tilde{l}_X^{(K)}\left(\Delta, x|x_0;\theta\right) = -\frac{m}{2}\ln\left(2\pi\Delta\right) - D_v\left(x;\theta\right) + \frac{C_X^{(j_{-1},-1)}\left(x|x_0;\theta\right)}{\Delta} + \sum\nolimits_{k=0}^{K} C_X^{(j_k,k)}\left(x|x_0;\theta\right)\frac{\Delta^k}{k!}. \tag{11}$$

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^K, 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 form:

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

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 three other related stochastic volatility models. While the expressions may 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 importantly, for a given model, the expressions need to be calculated *only once*. So, if one is interested in estimating, for instance, the model of Heston (1993) (or any of the other three 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-in the coefficients $C_X^{(j_k,k)}$ 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 equation (3), the instantaneous relative volatility of the stock is given by $\sqrt{\sigma_0 (X_t) \sigma_0^T (X_t)}/S_t$. 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 has 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.

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 5 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 trade off 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 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 price for the entire data sample. However, if, in addition to holding 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 time to maturity $T_i - t$ is held fixed for each option, and strike price K_i is a simple function of stock price for each option. In fact, option markets usually provide a reasonable range of moneynesses 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 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:

$$C\left(T, \alpha S_T, \alpha K, Y_T\right) = \left(\alpha S_T - \alpha K\right)^+ = \alpha\left(S_T - K\right)^+ = \alpha C\left(T, S_T, K, Y_T\right) \tag{13}$$

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:

$$\begin{aligned} \sigma_1\left(X_t\right)\sigma_1^T\left(X_t\right) &= \varphi_{11}\left(Y_t\right)S_t^2 & & \\ \sigma_1\left(X_t\right)\sigma_i^T\left(X_t\right) &= \varphi_{1i}\left(Y_t\right)S_t = \varphi_{i1}\left(Y_t\right)S_t & & i>1 \\ \sigma_i\left(X_t\right)\sigma_j^T\left(X_t\right) &= \varphi_{ij}\left(Y_t\right) & & i>1,\ j>1 \\ \mu_i^Q\left(X_t\right) &= \psi_i\left(Y_t\right) & & i>1 \end{aligned} \tag{14}$$

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

$$C\left(t, S_t, K, Y_t\right) = S_t H\left(t, m_t, Y_t\right) \tag{15}$$

where m_t is the logarithmic moneyness of the option:

$$m_t = \ln S_t - \ln K \tag{16}$$

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

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

Note that the solution $H\left(t,m_t,Y_t\right)$ 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:

$$
H\left(T,m_T,Y_T\right) = \left(1 - e^{-m_T}\right)^{+} \tag{18}
$$

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 $m-1$.

Thus, provided the stochastic volatility model under consideration satisfies the homogeneity conditions of equation (14), scaled option prices with $m-1$ distinct combinations of time to maturity $T_i - t$ and moneyness S_t/K_i 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.

3.3. *Change of Variables: From State to Observed Variables*

We have now obtained an expansion of the joint likelihood of observations on $X_t = \left[S_t; Y_t\right]^T$ in the form (11). If the method of Ledoit et al. (2002) 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, $G_t = [S_t; C_t]^T$.

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:

$$\begin{aligned} J_t &= \det \begin{bmatrix} \frac{\partial S_t}{\partial S_t} & \frac{\partial S_t}{\partial Y_t(1)} & \cdots & \frac{\partial S_t}{\partial Y_t(N)} \\ \frac{\partial C_t(1)}{\partial S_t} & \frac{\partial C_t(1)}{\partial Y_t(1)} & \cdots & \frac{\partial C_t(1)}{\partial Y_t(N)} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial C_t(N)}{\partial S_t} & \frac{\partial C_t(N)}{\partial Y_t(1)} & \cdots & \frac{\partial C_t(N)}{\partial Y_t(N)} \end{bmatrix} = \det \begin{bmatrix} 1 & 0 & \cdots & 0 \\ \frac{\partial C_t(1)}{\partial S_t} & \frac{\partial C_t(1)}{\partial Y_t(1)} & \cdots & \frac{\partial C_t(1)}{\partial Y_t(N)} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial C_t(N)}{\partial S_t} & \frac{\partial C_t(N)}{\partial Y_t(1)} & \cdots & \frac{\partial C_t(N)}{\partial Y_t(N)} \end{bmatrix} \\ &= \det \begin{bmatrix} \frac{\partial C_t(1)}{\partial Y_t(1)} & \cdots & \frac{\partial C_t(1)}{\partial Y_t(N)} \\ \vdots & \ddots & \vdots \\ \frac{\partial C_t(N)}{\partial Y_t(1)} & \cdots & \frac{\partial C_t(N)}{\partial Y_t(N)} \end{bmatrix} \end{aligned} \tag{19}$$

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), $G_t = [S_t; C_t]^T$ can be derived from the transition function of the state vector $X_t = [S_t; Y_t]^T$. 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 for the conditional density of $G_{t+\Delta} = g$ given $G_t = g_0$:

$$\begin{aligned} p_G(\Delta, g|g_0; \theta) &= \det \left(\frac{\partial f\left(f^{-1}(g; \theta)\right)}{\partial x} \right)^{-1} p_X(\Delta, f^{-1}(g; \theta) | f^{-1}(g_0; \theta); \theta) \\ &= J_t(\Delta, g|g_0; \theta)^{-1} \; p_X(\Delta, f^{-1}(g; \theta) | f^{-1}(g_0; \theta); \theta) \end{aligned} \tag{20}$$

where $J_t(\Delta, g|g_0;\theta)$ is the determinant defined in (19).

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

$$\ell_n(\theta) \equiv n^{-1} \sum\nolimits_{i=1}^{n} l_G\left(t_i - t_{i-1}, g_{t_i} | g_{t_{i-1}}; \theta\right) \tag{21}$$

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 (see however Aït-Sahalia and Mykland (2003) for 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 the expression (21), evaluated at the observations $g_{t_0}, g_{t_1}, ...g_{t_n}$ over the parameter values.

4. Example: The Heston Model

In what follows, we apply our 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

$$dX_t = d\begin{bmatrix} S_t \\ Y_t \end{bmatrix} = \begin{bmatrix} (r-d)\,S_t \\ \kappa'(\gamma' - Y_t) \end{bmatrix} dt + \begin{bmatrix} \sqrt{(1-\rho^2)\,Y_t}S_t & \rho\sqrt{Y_t}S_t \\ 0 & \sigma\sqrt{Y_t} \end{bmatrix} d\begin{bmatrix} W_1^Q(t) \\ W_2^Q(t) \end{bmatrix} \tag{22}$$

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 0 cannot be achieved if Feller's condition, $2\kappa'\gamma' \geq \sigma^2$, is satisfied. If we restate the dynamics in terms of the logarithmic stock price $s_t = \ln S_t$ instead, we have:

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

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, 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: $\Lambda = \left[\lambda_1 \sqrt{(1-\rho^2) Y_t},\ \lambda_2 \sqrt{Y_t}\right]^T$. The joint dynamics of s_t and Y_t under the objective measure P are then:

$$d \begin{bmatrix} s_t \\ Y_t \end{bmatrix} = \begin{bmatrix} a + bY_t \\ \kappa (\gamma - Y_t) \end{bmatrix} dt + \begin{bmatrix} \sqrt{(1-\rho^2) Y_t} & \rho\sqrt{Y_t} \\ 0 & \sigma\sqrt{Y_t} \end{bmatrix} d \begin{bmatrix} W_1^P(t) \\ W_2^P(t) \end{bmatrix} \tag{24}$$

where

$$a = r - d, \quad b = \lambda_1 \left(1 - \rho^2\right) + \lambda_2 \rho - \frac{1}{2}, \quad \kappa = \kappa' - \lambda_2 \sigma, \quad \gamma = \left(\frac{\kappa + \lambda_2 \sigma}{\kappa}\right) \gamma'. \tag{25}$$

4.1. Unobservable Volatility

When the volatility state variable Y_t is not observable, its value must be backed out from option prices as discussed above in order to carry out the maximum likelihood estimation of the model's parameters, $\theta = [\kappa; \gamma; \sigma; \rho; \lambda_1; \lambda_2]^T$. 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 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, we use characteristic functions (as in Heston (1993), modified by Carr and Madan (1998)), exploiting the fact that this particular model is affine under the Q measure (it is also affine under P but this is irrelevant). The option price can be expressed as:

$$C(s_t, Y_t, K, \Delta) = E^Q \left[\left[\exp(s_{t+\Delta}) - K\right]^+ \mid s_t, Y_t \right] \tag{26}$$

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:

$$c\left(s_t, Y_t, K, \Delta\right) = \exp\left[-s_t\right] C\left(s_t, Y_t, K, \Delta\right) \tag{27}$$

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:

$$c\left(s_t, Y_t, K, \Delta\right) = \int_0^{+\infty} \mathrm{Re}\left[\frac{\exp\left[w_0 + w_1 \cdot m_t + w_2 \cdot Y_t\right]}{\alpha\left(\alpha+1\right) - u^2 + \left(2\alpha+1\right) iu}\right] du \tag{28}$$

where α is an arbitrary scaling parameter and

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

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.

4.2. Using a Volatility Proxy

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 $[S_t; Y_t]^T$, or $[s_t; Y_t]^T$ as given in (24), will only permit identification of the parameters $[\kappa; \gamma; \sigma; \rho; b]^T$ or equivalently $[\kappa; \gamma; \sigma; \rho; \lambda_1]^T$, 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.

In the case where volatility is unobservable, the dependence of the joint likelihood function of $X_t = [S_t; Y_t]^T$ under P on the full set of market price of risk parameters is introduced by the Jacobian term, itself resulting from the transformation from $[S_t; C_t]^T$ to $[S_t; Y_t]^T$ as described in Section 3.3. But this suggests that, in the unobservable volatility case, the separate identification of the two market price of risk parameters is likely to be tenuous, a fact confirmed by the Monte Carlo experiments below.

5. Monte Carlo Results

One major advantage of our method 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.

5.1. Small Sample Distributions

We perform 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 proxy of Ledoit et al. (2002). We use sample lengths of $n = 500$, $5,000$, and $10,000$ transitions, at the daily ($\Delta = 1/252$) and weekly ($\Delta = 7/365$) sampling intervals. The parameter values for κ, γ, and σ used in the simulations are 3.0, 0.10, and 0.25, which are similar to the values obtained from the empirical application in Section 6. A value of -0.8 was chosen for ρ, to reflect the empirical regularity that innovations to volatility and stock price are generally negatively correlated; this value is also similar to the value estimated in 6. The values of the instantaneous interest rate and dividend yield, r and d, were held fixed at 0.04 and 0.015. The risk premia parameters, λ_1 and λ_2, were set to 4.0 and 0.0, respectively, in the simulations.

For each batch of simulations, we generate $1,000$ sample path samples using an Euler discretization of the process, using thirty sub-intervals per sampling interval; twenty nine out of every thirty 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 then estimate the model parameters using the method described above. When simulating the joint dynamics of the state vector $X_t = [S_t; Y_t]^T$, 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. 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 for $1,000$ data series, each containing 500, $5,000$ or $10,000$ observations at the daily frequency, or 500 weekly observations, all with observed volatility. Table 2 reports results for 500 weekly observations, 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 $G_t = [S_t; C_t]^T$ as the observed vector). The mean difference between the estimates and the true values of the parameter (i.e., those 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 quite small relative to the parameter values. The γ parameter fares only slightly 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; for example, with 500 daily observations and volatility observed, the true value of κ is 3.0, but the standard error is 1.6.

The use of otherwise similar batches of simulations with differing numbers of daily observations in each simulated series provides some insight into how closely 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 the asymptotic distribution.

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.

Also of note is the large standard error for the γ parameter when volatility is unobserved. This result may 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 may 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. With observed volatility, the standard deviation of this product with 500 weekly observations is 0.31; when volatility is unobserved, this standard deviation increases to 0.38. The large 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.

5.2. *The Effects of Using a Volatility Proxy*

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 volatility. The results in Table 1 are based on an 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.

Table 3 compares the use of Fourier inversion to determine the level of stochastic volatility to the use of an implied volatility proxy. The simulations are the same as those used in Table 2 at the weekly frequency. The λ_2 parameter is held fixed at the data generating value of -6.0, since it is unidentified when a proxy is used; holding this parameter fixed in both sets of estimates makes the results comparable.

A relevant metric to assess the effect is to examine whether the use of the proxy introduces enough additional noise to be noticeable at the scale of the standard error of the estimators due to the sampling noise. The answer is "no," since the estimators are on average more accurate when treating volatility as observable.

A more subtle point is that the design where the value of λ_2 is known does bias the results in favor of the use of the proxy. Indeed, while we can estimate the parameter $b = \lambda_1 \left(1 - \rho^2\right) + \lambda_2\rho - 1/2$ in (25) quite accurately, the resulting estimate for λ_1 is necessarily dependent upon the assumption made on λ_2. When using real data, one does 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 a input above. This said, it is important to note that this feature may 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.

5.3. *Comparing Small Sample to Asymptotic Distributions*

We can also use these Monte Carlos simulations to assess whether the predicted asymptotic behavior of maximum-likelihood estimators is matched in small samples by our maximum-likelihood estimator. Table 4 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:

$$n^{1/2}\left(\hat{\theta}-\theta\right) \rightarrow N(0, F^{-1}) \tag{29}$$

where $F = -E\left[\partial l_G / \partial\theta\partial\theta'\right]$ 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 4 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 very close to the asymptotic distribution with as few as 500 daily observations (of course, anything can happen in finite samples and it is possible for a different estimator to beat maximum likelihood in finite samples).

5.4. *Conclusions from the Monte Carlo Simulations*

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

1. The use of the 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 no proxy is

used; it seems therefore that using the proxy is a reasonable trade-off to make if we can live without the full identification or make an arbitrary assumption on one of market price of risk parameters.

2. The small sample distributions of the maximum-likelihood estimators are well approximated by their asymptotic counterparts.

6. Four 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 $G_t = [S_t; C_t]^T$ 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 30 calendar days to expiration (or 22 trading days). This constant maturity - 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 directly the VIX as the proxy discussed above for Y_t (or rather $\sqrt{Y_t}$, since Y_t is the local variance).

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. Being near the money, the options entering the basket are the most liquid ones in existence. The VIX options are European, simplifying the analysis. For further details on the VIX, see Whaley (1993) and 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). For the weekly estimation, the data for each Wednesday is used. For the relatively small number of dates on which no trading took place on a Wednesday, the average of the Tuesday and Thursday prices is used. The results for each of the four models are discussed briefly below. Both point estimates and standard errors for each of the four models can be estimated quickly and easily, without the need for simulations; other models can be estimated as easily using the technique outlined above.

6.1. *The Heston Model*

In Table 5, we report the estimation results for the Heston (1993) model described above, treating volatility Y_t as observed in the form of a proxy (the VIX index). The Monte Carlo results suggest that the effect of replacing Y_t by this proxy are quite small – within the asymptotic standard errors based on Fisher's Information matrix. As expected, we find that the correlation parameter ρ between the innovations to stock price and stochastic volatility is strongly negative, hovering around -0.8. The long term value of the volatility $\gamma^{1/2}$ is estimated to be approximately 21% per year with a mean reversion coefficient between 3 and 5 depending on the sampling frequency. Comparing the results at the daily and weekly frequencies can be interpreted as a form of specification test: a well-specified model should yield similar estimates up to sampling noise.

The large uncertainty for the risk premia estimates are 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 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 arbitrarily precisely 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, hourly, etc. 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.

6.2. *The GARCH Stochastic Volatility Model*

We now turn to the GARCH stochastic volatility model (see Meddahi (2001)). Although some of the parameters of this model have been estimated in Andersen et al. (2002a), neither they nor Meddahi (2001) specify fully the form of the drift of the stock price. The estimation in Andersen et al. (2002a) uses asymptotics where the time interval goes to zero in order to estimate the parameters of the volatility process; as the time between observations of the stock price goes to zero, the effect of the drift of the stock price on the volatility estimates disappears. By contrast, the time between observations remains fixed in our estimation procedure, and the drift of the stock price remains relevant even asymptotically. Consequently, we must complete the model definition by specifying the drift of Y_t.

S_t and Y_t follow a process of the following form under the Q measure:

$$dX_t = d\begin{bmatrix} S_t \\ Y_t \end{bmatrix} = \begin{bmatrix} (r-d)\,S_t \\ \kappa'\left(\gamma' - Y_t\right) \end{bmatrix} dt + \begin{bmatrix} \sqrt{(1-\rho^2)\,Y_t}S_t & \rho\sqrt{Y_t}S_t \\ 0 & \sigma Y_t \end{bmatrix} d\begin{bmatrix} W_1^Q(t) \\ W_2^Q(t) \end{bmatrix} \tag{30}$$

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

$$d\begin{bmatrix} s_t \\ Y_t \end{bmatrix} = \begin{bmatrix} r - d - \frac{1}{2}Y_t \\ \kappa'(\gamma' - Y_t) \end{bmatrix} dt + \begin{bmatrix} \sqrt{(1-\rho^2)\,Y_t} & \rho\sqrt{Y_t} \\ 0 & \sigma Y_t \end{bmatrix} d\begin{bmatrix} W_1^Q(t) \\ W_2^Q(t) \end{bmatrix} \tag{31}$$

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 $\Lambda = \left[\lambda_1\sqrt{(1-\rho^2)\,Y_t},\ 0\right]^T$. With these assumptions, the dynamics of the state variables under the measure P are then:

$$d\begin{bmatrix} s_t \\ Y_t \end{bmatrix} = \begin{bmatrix} r - d + (\lambda_1\left(1-\rho^2\right) - \frac{1}{2})Y_t \\ \kappa(\gamma - Y_t) \end{bmatrix} dt + \begin{bmatrix} \sqrt{(1-\rho^2)\,Y_t} & \rho\sqrt{Y_t} \\ 0 & \sigma Y_t \end{bmatrix} d\begin{bmatrix} W_1^P(t) \\ W_2^P(t) \end{bmatrix} \tag{32}$$

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 Appendix B.

In Table 6, we report the estimation results for this model. Compared to the Heston model, the speed of mean reversion is estimated at a much lower value, and the unconditional mean of volatility is estimated at a much higher value in this model (note, however, that the standard errors for the speed of mean reversion are very large). The point estimates for the correlation parameter are 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.

The surprisingly large standard errors on the speed of mean reversion parameter suggest that there may be some model misspecification; particularly at the daily frequency, the speed of mean reversion estimated by maximum likelihood on the VIX proxy alone is normally much higher and much better estimated. The unconditional mean of volatility corresponds to an instantaneous standard deviation in the stock price of approximately 27% at the daily frequency and 33% at the weekly frequency, much higher than that estimated for any of the other models.

6.3. *The Lognormal Stochastic Volatility Model*

Another model we consider is from Scott (1987), Wiggins (1987), Chesney and Scott (1989), Scott (1991), and Andersen et al. (2002b), and is also examined in Meddahi (2001) and Andersen et al. (2002a) (who, as with the GARCH model, do not fully specify the drift of the stock price). Under the Q measure, the state variables S_t and Y_t follow a process of the following form:

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

Note that Y_t follows an Ornstein-Uhlenbeck process, and can take on any value, positive or negative; as there is no boundary, there are no parameter restrictions needed to prevent attainment of the boundary. Proceeding as before, we can express these dynamics in terms of the logarithmic stock price rather than the stock price itself:

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

Since the market price of risk parameters for Y_t are not identified when volatility is observable, we set this component of the market price of risk vector at 0, and specify $\Lambda = \left[\lambda_1\sqrt{(1-\rho^2)},\ 0\right]^T$.

The P-measure dynamics of the state variables are then:

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

where $\kappa' = \kappa$ and $\gamma' = \gamma$. As with the GARCH model, the Q-measure dynamics are not affine in Y_t. The log-likelihood expansion for this model is given in Appendix C.

In Table 7, we report the estimation results for the lognormal model, treating volatility Y_t as observed in the form a proxy (the VIX). The speed of mean reversion is much better estimated in this model than in the previous model, and has a higher value. The unconditional mean of the logarithmic volatility is about the same when estimated at both the daily and weekly frequency, and is approximately -3.3. This value corresponds to an instantaneous standard deviation of the stock price of approximately 0.2, slightly lower than the value implied by the estimates for the Heston model, and much lower than those implied by the GARCH model. The point estimates for the correlation are quite similar to those found in the previous two models.

6.4. The CEV Stochastic Volatility Model

We finally consider a more general model, which nests some of the previous examples. Under the Q measure, the the state variables S_t and Y_t follow a process of the following form:

$$dX_t = d\begin{bmatrix} S_t \\ Y_t \end{bmatrix} = \begin{bmatrix} (r-d)S_t \\ \kappa'(\gamma' - Y_t) \end{bmatrix} dt + \begin{bmatrix} \sqrt{(1-\rho^2)Y_t}S_t & \rho\sqrt{Y_t}S_t \\ 0 & \sigma Y_t^\beta \end{bmatrix} d\begin{bmatrix} W_1^Q(t) \\ W_2^Q(t) \end{bmatrix} \tag{36}$$

where we constrain the parameter $\beta \geq 1/2$. This model is considered by, for example, Jones (2003), 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 may still warrant separate investigation, despite their being nested by the model of (36); 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 (see the specific discussion of the Heston (1993) model). Proceeding as before, we can express these dynamics in terms of the logarithmic stock price rather than the stock price itself:

$$d \begin{bmatrix} s_t \\ Y_t \end{bmatrix} = \begin{bmatrix} r - d - \frac{1}{2} Y_t \\ \kappa' \left(\gamma' - Y_t \right) \end{bmatrix} dt + \begin{bmatrix} \sqrt{\left(1-\rho^2\right) Y_t} & \rho \sqrt{Y_t} \\ 0 & \sigma Y_t^{\beta} \end{bmatrix} d \begin{bmatrix} W_1^Q(t) \\ W_2^Q(t) \end{bmatrix} \tag{37}$$

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

The P-measure dynamics of the state variables are then:

$$d \begin{bmatrix} s_t \\ Y_t \end{bmatrix} = \begin{bmatrix} r - d - \frac{1}{2} Y_t + \lambda_1 \left(1-\rho^2\right) Y_t \\ \kappa \left(\gamma - Y_t \right) \end{bmatrix} dt + \begin{bmatrix} \sqrt{\left(1-\rho^2\right) Y_t} & \rho \sqrt{Y_t} \\ 0 & \sigma Y_t^{\beta} \end{bmatrix} d \begin{bmatrix} W_1^P(t) \\ W_2^P(t) \end{bmatrix} \tag{38}$$

where $\kappa' = \kappa$ and $\gamma' = \gamma$. As with the GARCH and lognormal models, the Q-measure dynamics are not affine in Y_t. The corresponding log-likelihood expansion can be found in Appendix D.

In Table 8, we report the estimation results for the general CEV model, treating volatility Y_t as observed in the form a proxy (the VIX). The instantaneous volatility of the stock price at the unconditional mean of the volatility state variable is approximately 0.22 at the daily frequency and 0.27 at the weekly frequency, somewhat higher than that obtained for the Heston and lognormal models, but much lower than that obtained for the GARCH model. Of particular interest for the CEV model of the exponent β, which is estimated (at both sampling frequencies) above the Heston value of 0.5 but below the GARCH value of 1. At the daily frequency, either value can be rejected at the conventional 95% confidence level; at the weekly frequency, the hypothesis that $\beta = 1$ can be rejected. This finding stands in contrast to that of Jones (2003), who, using a Bayesian method, estimates this exponent above the GARCH value of 1. The point estimates for the correlation coefficient are almost identical for all four models.

6.5. Likelihood Ratio Tests for Nested Models

The CEV model nests the Heston ($\beta = 1/2$) and the GARCH ($\beta = 1$) models. The use of likelihood estimation makes it straightforward to calculate likelihood ratio statistics for the nested models. These statistics are shown in Table 9 for both models at the daily and weekly frequencies. All four combinations of model and sampling frequency are easily rejected at the conventional 95% confidence level. The statistic corresponding to the highest p-value is for the Heston model at the weekly frequency, but the likelihood ratio statistic of 15.4 is still more than four times the 95% cutoff value of 3.84. For the GARCH model at the weekly frequency, and for both models at the daily frequency, the statistic is anywhere from 40 to 180 times the 95% cutoff value, suggesting extremely strong rejection at any reasonable confidence level.

The point estimate of the β coefficient at both frequencies 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 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 and standard errors suggest that neither boundary value is appropriate, with the elasticity of variance lying between the two. At the daily frequency, either boundary value is strongly rejected.

7. Incorporating Jumps

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

$$dX_t = \mu^P (X_t)\, dt + \sigma (X_t)\, dW_t^P + J_t^P dN_t^P \tag{39}$$

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 $\nu(., \theta)$.

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

$$\begin{aligned} p_X^{(K)} (\Delta, x|x_0; \theta) \quad &= \quad \exp\left(-\frac{m}{2} \ln (2\pi\Delta) - D_v (x; \theta) + \frac{c_X^{(-1)} (x|x_0; \theta)}{\Delta} \right) \sum\nolimits_{k=0}^{K} c_X^{(k)} (x|x_0; \theta) \frac{\Delta^k}{k!} \\ &\quad + \sum\nolimits_{k=1}^{K} d_X^{(k)} (x|x_0; \theta) \frac{\Delta^k}{k!} \end{aligned} \tag{40}$$

Again, the series can be calculated up to arbitrary order K and the unknowns are the coefficients $c_X^{(k)}$ and $d_X^{(k)}$. The difference between the coefficients $c_X^{(k)}$ in (40) 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 $\ln(p_X^{(K)})$ in Δ are matched to the coefficients $C_X^{(k)}$ of the direct Taylor series $\ln p_X^{(K)}$. 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 sum of $d_X^{(k)}$ coefficients.) The coefficients can be computed analogously to the pure diffusive case.

8. Conclusions

We have described and implemented a technique for maximum likelihood estimation of models with stochastic volatility, or latent variables, and applied this technique to the models of Heston (1993) and three others with VIX data. We performed Monte Carlo simulations for the Heston model 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 additional to the statistical efficiency of maximum likelihood. When the observed vector consists of $G_t = [S_t; C_t]^T$, 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 an approximation technique such as Ledoit et al. (2002) 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. In this case, using such a proxy reduces the exercise to one of simply applying our likelihood expansion to the state vector $X_t = [S_t; Y_t]^T$. But even when that is not deemed desirable – or no such reasonable proxy exists – our method retains its high accuracy and computational efficiency as demonstrated by the simulations above.

We applied our method to the Heston (1993), GARCH, lognormal, 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 the stock level S_t and the volatility variable Y_t is consistently around -0.8 for all models. This negative correlation has long been noted (in the form of the "leverage effect"). This suggests that stochastic volatility models, pricing and/or estimation methods that rely on the assumption of uncorrelated shocks (such as Hull and White (1987) for instance) will be quite unrealistic in this context.

However, nothing 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, or for that matter models with other types of latent variables. In Section 3, we described our method without reference to any specific model. Provided that enough traded asset prices (such as the call options we used) or other observable quantities can be found to be mapped into the unobservable latent state vector, our method can then be applied.

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. Tech. rep., 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., Bollerslev, T., Meddahi, N., 2002a. Analytic evaluation of volatility forecasts. Tech. rep., Northwestern University.

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

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

Bakshi, G., Cao, C., Chen, Z., 2000. 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. Tech. rep., 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.

Chesney, M., Scott, L., 1989. Pricing European currency options: A comparison of the modified Black-Scholes model and a random variance model. Journal of Financial and Quantitative Analysis 24, 267–284.

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.

Harvey, A. C., Shephard, N., 1994. Estimation of an asymmetric stochastic volatility model for asset returns. Journal of Business and Economic Statistics 14, 429–434.

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. Tech. rep., University of Maryland.

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

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

Ledoit, O., Santa-Clara, P., Yan, S., 2002. Relative pricing of options with stochastic volatility. Tech. rep., University of California at Los Angeles.

Meddahi, N., 2001. An eigenfunction approach for volatility modeling. Tech. rep., Université de Montréal.

Merton, R. C., 1973. The theory of rational option pricing. Bell journal of Economics and Management Science 4, 141–183.

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

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.

Scott, L. O., 1987. Option pricing when the variance changes randomly: Theory, estimation, and an application. Journal of Financial and Quantitative Analysis 22, 419–438.

Scott, L. O., 1991. Random variance option pricing: Empirical tests of the model and delta-sigma hedging. Advances in Futures and Options Research 5, 113–135.

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.

Taylor, S. J., 1994. Modeling stochastic volatility: A review and comparative study. Mathematical Finance 4, 183–204.

Whaley, R. E., 1993. Derivatives on market volatility: Hedging tools long overdue. The journal of Derivatives 1, 71–84.

Whaley, R. E., 2000. The investor fear gauge. The journal of Portfolio Management 26, 12–17.

Wiggins, J., 1987. Option values under stochastic volatility: Theory and empirical estimates. Journal of Financial Economics 19, 351–372.

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.

Appendix: The Log-Likelihood Expansion for Stochastic Volatility Models

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

A. The Heston Model

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

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

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

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

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

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

$$\begin{aligned} C_X^{(1)}\left(x|x_0;\theta\right) = &\frac{\rho\sigma a_2 b_1+\rho\sigma a_1 b_2-\sigma^2 a_1 b_1-a_2 b_2}{(1-\rho^2)\sigma^2} + \frac{\left(2\rho\sigma b_1 b_2-\sigma^2 b_1^2-b_2^2\right)x_{20}}{2(1-\rho^2)\sigma^2} \\ &-\frac{\sigma^4-\rho^2\sigma^4+6\sigma^2 a_1^2-6\sigma^2 a_2+6\rho^2\sigma^2 a_2-12\rho\sigma a_1 a_2+6a_2^2}{12(1-\rho^2)\sigma^2 x_{20}} \end{aligned}$$

Note that the behavior at the 0 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.

B. The GARCH Stochastic Volatility Model

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

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

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

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

$$
\begin{aligned}
C_X^{(1)}(x|x_0;\theta) = {} & -\frac{x_{20}\left(1+4\lambda_1^2\left(1-\rho^2\right)\right)}{8(1-\rho^2)} - \frac{\gamma^2\kappa^2}{2(1-\rho^2)\sigma^2 {x_{20}}^2} + \frac{(r-d)\gamma\kappa\rho}{(1-\rho^2)\sigma {x_{20}}^{\frac{3}{2}}} + \frac{\rho\left(4\kappa+\sigma^2\right)\sqrt{x_{20}}}{8(1-\rho^2)\sigma} \\
& - \frac{\rho\left(2\gamma\kappa-d\left(4\kappa+\sigma^2\right)+r\left(4\kappa+\sigma^2\right)\right)}{4(1-\rho^2)\sigma\sqrt{x_{20}}} - \frac{\left(4(r-d)\sigma\sqrt{x_{20}}+\rho\left(4\kappa+\sigma^2\right)x_{20}-2\sigma x_{20}^{3/2}-4\gamma\kappa\rho\right)\lambda_1}{4(1-\rho^2)\sigma\sqrt{x_{20}}} \\
& + \frac{\gamma\kappa\left(4\kappa+\left(4-3\rho^2\right)\sigma^2\right)-2(r-d)^2\sigma^2}{4(1-\rho^2)\sigma^2 x_{20}} - \frac{48\kappa^2+24\left(2(d-r+\kappa)-\kappa\rho^2\right)\sigma^2+\left(13-10\rho^2\right)\sigma^4}{96(1-\rho^2)\sigma^2}
\end{aligned}
$$

Note that the behavior at the 0 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.

C. The Lognormal Stochastic Volatility Model

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

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

$$
\begin{aligned}
C_X^{(-1)}(x|x_0;\theta) = {} & \frac{\sigma^2(x_1-x_{10})^4}{96e^{2x_{20}}(1-\rho^2)^2} + \frac{(x_1-x_{10})^2(x_2-x_{20})}{e^{x_{20}}4(1-\rho^2)} - \frac{\rho\sigma(x_1-x_{10})^3(x_2-x_{20})}{24e^{3x_{20}/2}(1-\rho^2)^2} \\
& - \frac{\rho(x_1-x_{10})(x_2-x_{20})^2}{e^{x_{20}/2}(1-\rho^2)4\sigma} + \frac{\left(5\rho^2-2\right)(x_1-x_{10})^2(x_2-x_{20})^2}{48e^{x_{20}}(1-\rho^2)^2} - \frac{\rho^3(x_1-x_{10})(x_2-x_{20})^3}{24e^{x_{20}/2}\sigma(1-\rho^2)^2} \\
& + \frac{\rho^2(x_2-x_{20})^4}{96(1-\rho^2)^2\sigma^2} - \frac{\sigma^2(x_1-x_{10})^2+e^{x_{20}}(x_2-x_{20})^2+2e^{x_{20}/2}\rho\sigma(x_1-x_{10})(-x_2+x_{20})}{2e^{x_{20}}(1-\rho^2)\sigma^2}
\end{aligned}
$$

$$
\begin{aligned}
C_X^{(0)}(x|x_0;\theta) = {} & -\frac{(x_1-x_{10})\left(2e^{x_{20}}\sigma+4(d-r)\sigma+e^{x_{20}/2}\left(4\theta\kappa\rho+\rho\sigma^2-4\left(1-\rho^2\right)\sigma\lambda_1-4\kappa\rho x_{20}\right)\right)}{4e^{x_{20}}(1-\rho^2)\sigma} \\
& - \frac{(x_2-x_{20})\left(-2e^{x_{20}}\rho\sigma+4(r-d)\rho\sigma+e^{x_{20}/2}\left(-4\theta\kappa-\sigma^2+4\rho\left(1-\rho^2\right)\sigma\lambda_1+4\kappa x_{20}\right)\right)}{4e^{x_{20}/2}(1-\rho^2)\sigma^2} \\
& - \frac{\sigma^2(x_1-x_{10})^2}{48e^{x_{20}}(1-\rho^2)} - \frac{(x_2-x_{20})^2\left(-6e^{x_{20}}\rho\sigma+12(d-r)\rho\sigma+e^{x_{20}/2}\left(24\kappa+\sigma^2\right)\right)}{48e^{x_{20}/2}(1-\rho^2)\sigma^2} \\
& - \frac{(x_1-x_{10})(x_2-x_{20})\left(24(r-d)\sigma+e^{x_{20}/2}\left(\sigma\left(-5\rho\sigma+12\left(1-\rho^2\right)\lambda_1\right)+12\kappa\rho(-2-\theta+x_{20})\right)\right)}{48e^{x_{20}}(1-\rho^2)\sigma}
\end{aligned}
$$

$$
\begin{aligned}
C_X^{(1)}(x|x_0;\theta) = {} & \kappa - \frac{e^{x_{20}}}{8(1-\rho^2)} - \frac{(d-r)^2}{2e^{x_{20}}(1-\rho^2)} - \frac{e^{x_{20}/2}\rho\left(4\theta\kappa+\sigma^2-4\kappa x_{20}\right)}{8(1-\rho^2)\sigma} + -\frac{\sigma^4(1+2\rho^2)}{96(1-\rho^2)\sigma^2} \\
& - \frac{(d-r)\rho\left(4\theta\kappa+\sigma^2-4\kappa x_{20}\right)}{4e^{x_{20}/2}(1-\rho^2)\sigma} + \frac{e^{x_{20}/2}\lambda_1}{2} + \frac{(d-r)\lambda_1}{e^{x_{20}/2}} + \frac{\rho\left(4\theta\kappa+\sigma^2-4\kappa x_{20}\right)\lambda_1}{4\sigma} \\
& - \frac{2\kappa^2 x_{20}^2-\kappa\rho^2\sigma^2 x_{20}-4\theta\kappa^2 x_{20}}{4(1-\rho^2)\sigma^2} - \frac{2\theta^2\kappa^2-2(r-d)\sigma^2+2\kappa\sigma^2(1-\rho^2)+\theta\kappa\rho^2\sigma^2+2\sigma^2\lambda_1^2(1-\rho^2)^2}{4(1-\rho^2)\sigma^2}
\end{aligned}
$$

Note that x_2 has no boundary; positivity of the variance is ensured because x_2 is not the instantaneous variance of the stock, but rather its natural logarithm.

D. The CEV Model

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

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

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

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

$$\begin{aligned}
C_X^{(-1)}\left(x|x_0;\theta\right) = & \frac{x_{20}^{-4+2\beta}\sigma^2(x_1-x_{10})^4}{96(1-\rho^2)^2} + \frac{(x_1-x_{10})^2(x_2-x_{20})}{4x_{20}^2(1-\rho^2)} - \frac{x_{20}^{-(7/2)+\beta}\rho\sigma(x_1-x_{10})^3(x_2-x_{20})}{24(1-\rho^2)^2} \\
& - \frac{(1+2\beta)x_{20}^{-(3/2)-\beta}\rho(x_1-x_{10})(x_2-x_{20})^2}{4(1-\rho^2)\sigma} + \frac{\left(9\rho^2-6-2\beta\left(1-\rho^2\right)\right)(x_1-x_{10})^2(x_2-x_{20})^2}{48x_{20}^3(1-\rho^2)^2} \\
& + \frac{\beta x_{20}^{-1-2\beta}(x_2-x_{20})^3}{2(1-\rho^2)\sigma^2} - \frac{x_{20}^{-(5/2)-\beta}\rho\left(3\rho^2-2-8\beta\left(1-\rho^2\right)-4\beta^2\left(1-\rho^2\right)\right)(x_1-x_{10})(x_2-x_{20})^3}{24(1-\rho^2)^2\sigma} \\
& + \frac{\left(\rho^2-16\beta\left(1-\rho^2\right)-28\beta^2\left(1-\rho^2\right)\right)(x_2-x_{20})^4}{96x_{20}^{2(1+\beta)}(1-\rho^2)^2\sigma^2} \\
& - \frac{x_{20}^{-1-2\beta}\left(x_{20}^3+2x_{20}^{3/2+\beta}\rho\sigma(x_1-x_{10})+x_{20}^{2\beta}\sigma^2(x_1-x_{10})^2-2\left(x_{20}^2+x_{20}^{1/2+\beta}\rho\sigma(x_1-x_{10})\right)x_2+x_{20}x_2^2\right)}{2(1-\rho^2)\sigma^2}
\end{aligned}$$

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

$$\begin{aligned}
C_X^{(1)}\left(x|x_0;\theta\right) = & \frac{\left(3-28\beta+12\beta^2-6\rho^2+40\beta\rho^2-24\beta^2\rho^2\right)\sigma^2x_{20}^{-2+2\beta}}{96(1-\rho^2)} - \frac{\kappa^2(x_{20}-\gamma)^2}{2(1-\rho^2)\sigma^2x_{20}^{2\beta}} \\
& + \frac{(1-2\beta)\rho\sigma x_{20}^{-3/2+\beta}(a+bx_{20})}{4(1-\rho^2)} - \frac{\kappa\rho x_{20}^{-1/2-\beta}(x_{20}-\gamma)(a+bx_{20})}{(1-\rho^2)\sigma} \\
& - \frac{2a^2-4\beta\gamma\kappa+\gamma\kappa\rho^2+2\beta\gamma\kappa\rho^2+4abx_{20}-2\kappa x_{20}+4\beta\kappa x_{20}+\kappa\rho^2x_{20}-2\beta\kappa\rho^2x_{20}+2b^2x_{20}^2}{4(1-\rho^2)x_{20}}
\end{aligned}$$

Note that the boundary at 0 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 for the Heston (1993) model).

Table 1. Monte Carlo Simulations with Observed Volatility

Parameter	True Value	500 Daily Obs. Bias	Std.Dev.	5,000 Daily Obs. Bias	Std.Dev.	10,000 Daily Obs. Bias	Std.Dev.	500 Weekly Obs. Bias	Std.Dev.
κ	3.00	0.8	1.6	0.06	0.4	0.02	0.3	0.1	0.5
γ	0.10	0.0005	0.022	0.0001	0.006	−0.0001	0.0041	0.0001	0.008
σ	0.25	0.0002	0.006	0.0000	0.002	0.0000	0.0014	0.0003	0.006
ρ	−0.8	−0.0002	0.013	−0.0001	0.004	−0.0001	0.003	−0.0013	0.013
λ_1	4.0	0.9	6.5	0.07	1.9	0.1	1.4	0.2	2.9

Note: This table shows the results of 1,000 Monte Carlo simulations with respectively 500, 5,000 and 10,000 daily observations (i.e., $\Delta = 1/252$) and observed volatility. The second column shows 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 market price of risk of the stochastic volatility variable, λ_2, is not identified when volatility is observed. The instantaneous interest rate and the instantaneous dividend yield of the stock were held fixed at the values of 4%, and 1.5% respectively.

Table 2. Monte Carlo Simulations with Unobserved Volatility

Parameter	True Value	500 Weekly Obs. Bias	Std.Dev.
κ	3.00	−0.6	1.6
γ	0.10	−2.6	11.9
σ	0.45	−0.014	0.03
ρ	−0.70	−0.0004	0.03
λ_1	−7.0	−0.5	4.9
λ_2	−6.0	−0.7	3.8

Note: This table shows the results of 1,000 Monte Carlo simulations with 500 weekly observations (i.e., $\Delta = 7/365$) and unobserved volatility. The second column shows 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 instantaneous interest rate and the instantaneous dividend yield of the stock were held fixed at the values of 4%, and 1.5% respectively.

Table 3. Effect of Volatility Proxy

Parameter	True Value	Actual		Proxy	
		Bias	Std.Dev.	Bias	Std.Dev.
κ	3.00	−0.4	0.61	0.3	0.66
γ	0.10	−3.1	8.8	−2.6	15.1
σ	0.45	−0.01	0.024	−0.018	0.013
ρ	−0.7	−0.003	0.023	−0.02	0.02
λ_1	−7.0	0.1	2.0	−0.6	2.0

Note: This table shows the results of 1,000 Monte Carlo simulations with 500 weekly observations (i.e., $\Delta = 7/365$) and unobserved volatility, using both Fourier inversion and an implied volatility proxy to determine the level of the stochastic volatility variable. The second column shows the parameters used to generate the simulated sample paths; the simulations are the same as those used in Table 2. The third column 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 determined by Fourier inversion. The fourth column shows the standard deviation of the parameter estimates, also using Fourier inversion. The fifth and sixth columns 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. The λ_2 parameter is unidentified and fixed at -6.0 when the implied volatility proxy is used. To make the results comparable, the λ_2 parameter was held fixed even when Fourier inversion is used. The instantaneous interest rate and the instantaneous dividend yield of the stock were held fixed at the values of 4% and 1.5%, respectively.

Table 4. Asymptotic Variance of Estimates with Observed Volatility

Parameter	True Value	500 Daily Obs. ASE	500 Daily Obs. SSSE	$5,000$ Daily Obs. ASE	$5,000$ Daily Obs. SSSE	$10,000$ Daily Obs. ASE	$10,000$ Daily Obs. SSSE
κ	3.00	1.136	1.554	0.359	0.377	0.254	0.253
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	4.00	6.24	6.49	1.97	1.91	1.40	1.42

Note: This table shows the standard deviations of the parameter estimates, calculated both analytically and from the Monte Carlo simulations. All values are based on daily observations. The second column shows the true values of the parameter vector; this value was used to generate the sample paths for the Monte Carlo simulations, and to calculate the standard deviations from the likelihood expressions. The third column, 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. The fourth column, 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. The fifth and sixth columns show the same information as the third and fourth columns, but with $5,000$ daily observations instead of 500. Finally, the seventh and eighth columns 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 were held fixed at 4% and 1.5% per 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.

Table 5. Parameter Estimates for the Heston Model

Parameter	Heston Model Daily	Weekly
κ	5.0	3.6
	(0.7)	(0.6)
γ	0.045	0.045
	(0.06)	(0.08)
σ	0.48	0.43
	(0.04)	(0.07)
ρ	-0.77	-0.79
	(0.06)	(0.14)
λ_1	3.9	4.2
	(4)	(5)

Note: This table shows the estimated parameter values for the Heston stochastic volatility model using the SPX-VIX dataset. The first column shows results for daily observations with observed volatility. The second column shows estimates also with observed volatility, but with weekly observations. Standard errors are shown in parentheses beneath each parameter estimate.

Table 6. Parameter Estimates for the GARCH Model

Parameter	GARCH Model Daily	Weekly
κ	1.6 (1.1)	1.0 (1.3)
γ	0.075 (0.1)	0.11 (1.3)
σ	2.20 (0.016)	3.37 (0.06)
ρ	-0.75 (0.06)	-0.74 (0.2)
λ_1	2.3 (3.9)	2.1 (3.8)

Note: This table shows the estimated parameter values for the GARCH stochastic volatility model using the VIX dataset. The results for daily and weekly observation frequencies are shown. Standard errors are shown in parentheses beneath each parameter estimate.

Table 7. Parameter Estimates for the Lognormal Model

Parameter	Log-Normal Model Daily	Weekly
κ	4.0 (0.6)	2.7 (0.5)
γ	-3.3 (0.5)	-3.3 (0.2)
σ	2.1 (0.02)	1.8 (0.04)
ρ	-0.76 (0.05)	-0.79 (0.13)
λ_1	0.4 (9)	0.4 (10)

Note: This table shows the estimated parameter values for the lognormal stochastic volatility model using the SPX-VIX dataset. The results for daily and weekly observation frequencies are shown. Standard errors are shown in parentheses beneath each parameter estimate.

Table 8. Parameter Estimates for the General CEV Model

Parameter	CEV Model Daily	Weekly
κ	3.0 (1.0)	3.3 (0.8)
γ	0.052 (0.07)	0.046 (0.07)
σ	1.37 (0.03)	0.58 (0.3)
β	0.85 (0.05)	0.59 (0.1)
ρ	-0.77 (0.05)	-0.79 (0.14)
λ_1	3.4 (4.2)	4.3 (4.9)

Note: This table shows the estimated parameter values for the general CEV stochastic volatility model using the SPX-VIX dataset. The results for daily and weekly observation frequencies are shown. Standard errors are shown in parentheses beneath each parameter estimate.

Table 9. Likelihood Ratio Tests for Nested Models

Model	Likelihood Ratio Statistic Daily Frequency	 Weekly Frequency
Heston	696.6	15.6
GARCH	131.6	258.8

Note: This table shows likelihood ratio statistics for the Heston and GARCH stochastic volatility models, relative to the CEV model (which nests both). The first column shows the likelihood ratio statistics for the estimated parameter values at a daily frequency, and the second column shows the results for weekly estimation. All estimates for all models are obtained using the implied volatility of an at-the-money short-maturity option as a proxy for the true level of volatility. In all cases, there is a single degree of freedom, so the 95% chi-squared cutoff value is 3.84. As shown, both nested models at both frequencies are rejected. The rejection is particularly strong for the Heston model, at both frequencies.

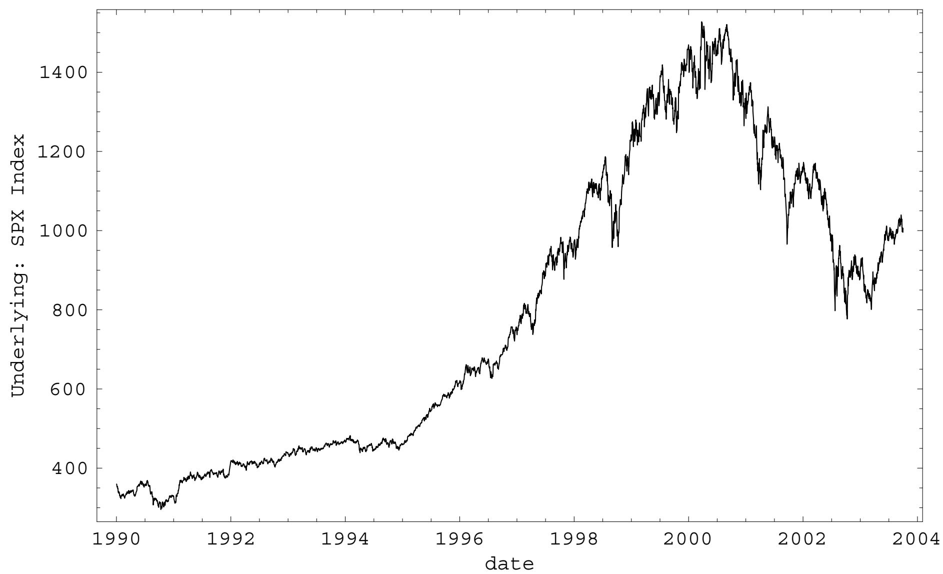

Fig. 1. The SPX (S&P 500) index represents the value of the underlying asset.

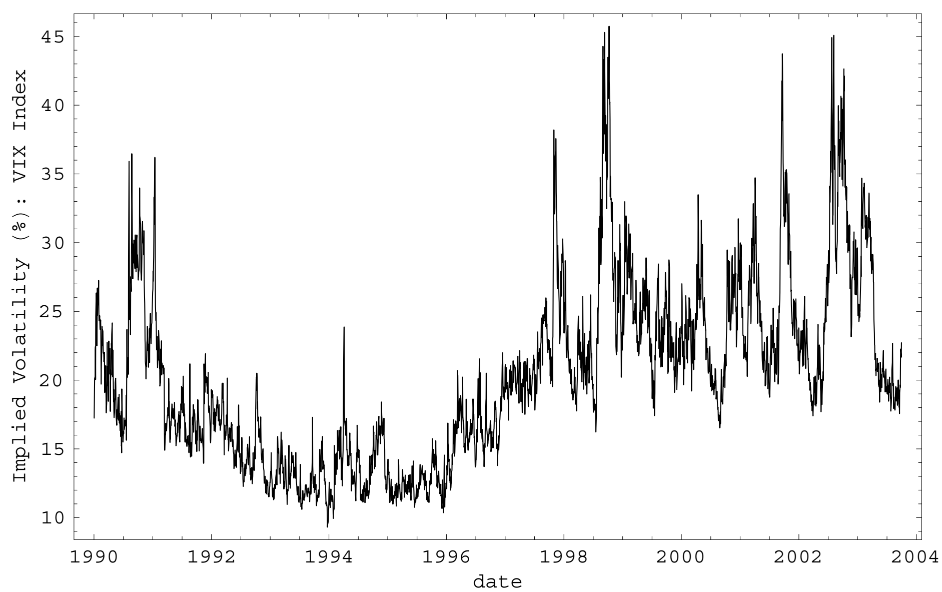

Fig. 2. The VIX index represents the value of the implied volatility of a basket of short maturity at-the-money options on the S&P 500 index.