

NBER WORKING PAPER SERIES

MAXIMUM LIKELIHOOD ESTIMATION
OF LATENT AFFINE PROCESSES

David S. Bates

Working Paper 9673
<http://www.nber.org/papers/w9673>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2003

I am grateful for comments on earlier versions of this paper from Luca Benzoni, Michael Chernov, Frank Diebold, Garland Durham, George Jiang, Michael Johannes, George Tauchen, and Chuck Whiteman, and from seminar participants at Colorado, Iowa, and NYU. The views expressed herein are those of the authors and not necessarily those of the National Bureau of Economic Research.

©2003 by David S. Bates. All rights reserved. Short sections of text not to exceed two paragraphs, may be quoted without explicit permission provided that full credit including ©notice, is given to the source.

Maximum Likelihood Estimation of Latent Affine Processes
David S. Bates
NBER Working Paper No. 9673
May 2003
JEL No. C1

ABSTRACT

This article develops a direct filtration-based maximum likelihood methodology for estimating the parameters and realizations of latent affine processes. The equivalent of Bayes' rule is derived for recursively updating the joint characteristic function of latent variables and the data conditional upon past data. Likelihood functions can consequently be evaluated directly by Fourier inversion. An application to daily stock returns over 1953-96 reveals substantial divergences from EMM-based estimates: in particular, more substantial and time-varying jump risk.

David S. Bates
Henry B. Tippie College of Business
University of Iowa
Iowa City, IA 52242-1000
and NBER
david-bates@uiowa.edu

“The Lion in Affrik and the Bear in Sarmatia are Fierce,
but Translated into a Contrary Heaven, are of less Strength and Courage.”

Jacob Zeigler; translated (1555) by Richard Eden

While models proposing time-varying volatility of asset returns have been around for thirty years, it has proven extraordinarily difficult to estimate the parameters of the underlying volatility process, and the current volatility level conditional on past returns. It has been especially difficult to estimate the continuous-time stochastic volatility models that are best suited for pricing derivatives such as options and bonds. Recent models suggest that asset prices jump and that the jump intensity is itself time-varying, creating an additional latent state variable to be inferred from asset returns.

The problem of estimating unobserved and time-varying volatility and jump risk from stock returns is an example of the general problem of inferring latent variables from observed data.

Currently, three approaches are commonly used for such state space problems:

- 1) analytically tractable specifications, such as Gaussian or regime-switching models;
- 2) derivatives-based methods that infer latent variables' realizations directly from prices of bonds or options; and
- 3) simulation methods.

Each approach has had some difficulties when applied to the problem of estimating stochastic volatility parameters and realizations. The few analytically tractable state space models available appear to be somewhat poor approximations of the relationship between stock returns and underlying volatility. The joint evolution of asset prices and volatility is not Gaussian, nor does volatility take on only a finite number of discrete values. Nevertheless, some papers have attempted to use such models as approximations for volatility evolution and estimation. Ruiz (1994) and Harvey, Ruiz, and Shephard (1994) have explored the Kalman filtration associated with Gaussian models, while Fridman and Harris (1998) essentially use a constrained regime-switching model with a large number of states as an approximation to an underlying stochastic volatility process.

Derivatives-based methods such as Pearson and Sun (1994) and Pan (2002) eliminate the latency of underlying state variables, thereby permitting direct estimation of state variable processes. However, such methods place considerable faith in market participants' ability to assess the values of latent state variables, as well as considerable faith in the pricing models used.

Simulation methods are currently the most actively researched approach for estimating discrete- and continuous-time stochastic volatility processes. Provided the model and parameters are correct, simulated data should possess roughly the same moments as the actual data – thereby permitting estimation of model parameters by a GMM procedure. At present, there is an extraordinary diversity of proposed moment conditions: Simulated Method of Moments of Duffie and Singleton (1993), empirical characteristic functions (Singleton, 2001), and spectral method of moments, to name a few. While such approaches can estimate process parameters, they are not especially suited to the filtration problem of estimating latent variable realizations.

One currently popular simulation method is the “Efficient Method of Moments” approach of Gallant and Tauchen (2001). That approach derives moment conditions from an auxiliary discrete-time model, simulates sample paths from the hypothesized continuous-time process for given parameter estimates via Monte Carlo, and then chooses those parameters that best match the moment conditions. Estimating the latent volatility realizations requires an additional “reprojection” step of identifying an appropriate rule for inferring volatility from past returns, given the observed relationship between the two series in the many simulations. The particle filter approach is another simulation methodology that is also receiving considerable attention currently.¹

This article explores an alternate recursive maximum likelihood methodology applicable whenever the (discrete-time) data and the latent variable(s) have an affine joint characteristic function conditional upon the latent variable. The major innovation is to work almost entirely in the transform space of characteristic functions, rather than working with probability densities. I derive the equivalent of Bayes' rule for updating the characteristic

¹See Johannes, Polson, and Stroud (2002), and the references therein.

function of the latent variable conditional upon observed data. Given this and the affine structure, recursively updating the conditional characteristic function of the data is straightforward. The probability densities necessary for maximum likelihood estimation can then be evaluated numerically by Fourier inversion.

The approach can be viewed as an extension of the Kalman filtration methodology used with Gaussian state space models – which indeed are included in the class of affine processes. In Kalman filtration, the multivariate normality of the data and latent variable(s) is exploited to update the estimated mean $\hat{x}_{t|t}$ and variance $P_{t|t}$ of the latent variable realization conditional on past data. Given normality, the conditional distribution of the latent variable is fully summarized by those moment estimates, while the associated moment generating function is of the simple form $G_{t|t}(\psi) = \exp[\hat{x}_{t|t}\psi + \frac{1}{2}P_{t|t}\psi^2]$. My approach generalizes the recursive updating of $G_{t|t}(\psi)$ to other affine processes that lack the analytic conveniences of multivariate normality.

The approach is limited at present to affine processes, whether discrete- or continuous-time. However, the affine class of processes is a broad and interesting one, and is extensively used in pricing bonds and options. In particular, jumps in returns and/or in the state variable can be accommodated, whereas simulation methods can have difficulty with rare events. Furthermore, some recent interesting expanded-data approaches also fit within the affine structure; e.g., the intradaily “realized volatility” used by Andersen, Bollerslev, Diebold, and Ebens (2001), or the use of implicit variances from options. Finally, some discrete-time non-affine processes can be re-expressed as affine processes after appropriate data transformations; e.g., the log stochastic volatility model examined *inter alia* by Harvey, Ruiz, and Shephard (1994) and Jacquier, Polson, and Rossi (1994).

A further advantage of the approach relative to moment-based simulation methods is that filtered estimates of the latent state variable are directly obtained from the recursive updating of the conditional characteristic function, while the precision of the estimate can also be easily estimated. Furthermore, the filtration algorithm can be examined more directly than the “kitchen sink” regression approach used in EMM reprojection. In particular, it is possible to directly compare the filtration with those used in GARCH volatility assessments. Finally,

the approach is a direct maximum likelihood approach using observed data, obviating the estimation of an auxiliary discrete-time model.

Section 1 below derives the basic algorithm for arbitrary affine processes. Section 2 runs diagnostics, using a simplified discrete-time affine stochastic volatility process and simulated data. Section 3 provides estimates of some affine continuous-time stochastic volatility/jump-diffusion models previously estimated by Andersen, Benzoni and Lund (2002) and Chernov, Gallant, Ghysels, and Tauchen (2002). For direct comparison with EMM-based estimates, I use the Andersen *et al* data set of daily S&P 500 returns over 1953-1996, which were graciously provided by Luca Benzoni. Section 4 concludes.

1. Recursive evaluation of likelihood functions for affine processes

Let \mathbf{y}_t denote an $(L \times 1)$ vector of variables observed at date t . Let \mathbf{x}_t represent an $(M \times 1)$ vector of latent state variables affecting the dynamics of \mathbf{y}_t . The following assumptions are made:

- 1) $\mathbf{z}_t \equiv (\mathbf{y}_t, \mathbf{x}_t)$ is assumed to be Markov;
- 2) the latent variables \mathbf{x}_t are assumed stationary; and
- 3) the characteristic function of $\mathbf{z}_{t+\tau}$ conditional upon \mathbf{z}_t is exponentially affine in the latent variables \mathbf{x}_t :

$$\begin{aligned} F_{\mathbf{y}, \mathbf{x}}(i\Phi, i\Psi; \mathbf{z}_t) &\equiv E\left[e^{i\Phi' \mathbf{y}_{t+\tau} + i\Psi' \mathbf{x}_{t+\tau}} \mid \mathbf{z}_t\right] \\ &= \exp[\mathbf{C}(\tau; i\Phi, i\Psi; \mathbf{y}_t) + \mathbf{D}(\tau; i\Phi, i\Psi; \mathbf{y}_t)' \mathbf{x}_t]. \end{aligned} \quad (1)$$

For simplicity, I will focus on the most common case of one data source and one latent variable: $L = M = 1$. Generalizing to higher-dimensional data and/or multiple latent variables is theoretically straightforward, but numerically more complex.

Equation (1) is satisfied by the general class of *affine processes*. The best-known example of this class is the Gaussian state-space system discussed in Hamilton (1994, Ch. 13), for which the conditional density function $p(\mathbf{z}_{t+1} \mid \mathbf{z}_t)$ is multivariate normal. As described in Hamilton, a recursive structure exists in this case for updating the conditional Gaussian densities of x_t over time based upon observing y_t . Given the Gaussian structure, it suffices to update the mean and variance of the latent x_t , which is done by Kalman filtration.

More general affine processes typically lack analytic expressions for the conditional density functions needed in maximum likelihood estimation. Their popularity for bond and option pricing models lies in the ability to compute those quantities numerically from characteristic functions. In essence, the characteristic function is the Fourier transform of the probability density function, while the density function is the inverse Fourier transform of the characteristic function. Using the notation $G(\Phi)$ for the moment generating function and $G(i\Phi)$ for the characteristic function, the latter is:

$$\begin{aligned} G(i\Phi) &\equiv E[e^{i\Phi y}] \\ &= \int e^{i\Phi y} p(y) dy \end{aligned} \quad (2)$$

$$p(y) = \frac{1}{2\pi} \int G(i\Phi) e^{-i\Phi y} d\Phi. \quad (3)$$

The existence of analytic solutions for the moment generating and characteristic functions then gets the problem halfway to a solution. Distribution functions and option prices can also be evaluated from the characteristic function by Fourier inversion.

It is also possible to numerically evaluate joint density functions by Fourier inversion:

$$p(x, y) = \frac{1}{(2\pi)^2} \iint F_{x,y}(i\psi, i\Phi) e^{-i\psi x - i\Phi y} d\psi d\Phi. \quad (4)$$

The following key proposition indicates that characteristic functions for conditional distributions can be evaluated from a partial inversion of F .

Proposition 1. *Let*

$$\begin{aligned} F_{x,y}(i\psi, i\Phi) &\equiv E[e^{i\psi x + i\Phi y}] \\ &= \iint e^{i\psi x + i\Phi y} p(x, y) dx dy \end{aligned} \quad (5)$$

*be the joint characteristic function of the random variables (x, y) . The characteristic function of x **conditional** upon observing y is*

$$G_{x|y}(i\psi; y) = \frac{\frac{1}{2\pi} \int_{-\infty}^{\infty} F_{x,y}(i\psi, i\Phi) e^{-i\Phi y} d\Phi}{p(y)} \quad (6)$$

where

$$p(y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F_{x,y}(0, i\Phi) e^{-i\Phi y} d\Phi \quad (7)$$

is the marginal density of y .

Proof: By Bayes' law, the conditional characteristic function $G_{x|y}$ can be written as

$$\begin{aligned} G_{x|y}(i\psi; y) &= \int e^{i\psi x} p(x|y) dx \\ &= \frac{1}{p(y)} \int e^{i\psi x} p(x, y) dx. \end{aligned} \quad (8)$$

$F_{x,y}(\bullet, i\Phi)$ is therefore the Fourier transform of $G_{x|y}(\bullet, y)p(y)$:

$$\begin{aligned}
F_{x,y}(i\psi; i\Phi) &= \int e^{i\Phi y} [G_{x|y}(i\psi, y) p(y)] dy \\
&= \iint e^{i\Phi y + i\psi x} p(x, y) dx dy.
\end{aligned} \tag{9}$$

Consequently, $G_{x|y}(i\psi; y) p(y)$ is the *inverse*(Φ, y) Fourier transform of $F_{x,y}(i\psi, i\Phi)$, yielding (6) above. ■

Let $\mathbf{Y}_t \equiv \{y_1, \dots, y_t\}$ be the data observed up through date t . Define $G_{t|s}(i\psi) \equiv E[e^{i\psi x_t} | \mathbf{Y}_s]$ as the characteristic function (CF) of the latent variable x_t conditional upon observing \mathbf{Y}_s . Given Proposition 1 and the affine structure, the filtered CF $G_{t|t}(i\psi)$ can be recursively updated as follows:

Step 0: At time $t = 0$, initialize $G_{t|t}(i\psi) = G_{0|0}(i\psi)$ at the CF of the unconditional density of the latent variable. From equation (1), this has an analytic solution of the form

$$G_{0|0}(i\psi) = \exp[C(\tau = \infty, \Phi = 0, i\psi)]. \tag{10}$$

Step 1: Given $G_{t|t}$, the joint characteristic function of next's period's $\mathbf{z}_{t+\tau} = \{y_{t+\tau}, x_{t+\tau}\}$ conditional on data observed through date t can be evaluated by iterated expectations, exploiting the special structure of affine characteristic and moment generating functions given in (1) above:

$$\begin{aligned}
F_{y,x|t}(i\Phi, i\psi) &= E_t \left[E \left(e^{i\Phi y_{t+\tau} + i\psi x_{t+\tau}} \mid x_t \right) \right] \\
&= E \left[e^{C(\tau, i\Phi, i\psi) + D(\tau, i\Phi, i\psi) x_t} \mid \mathbf{Y}_t \right] \\
&= e^{C(\tau, i\Phi, i\psi)} G_{t|t}[D(\tau; i\Phi, i\psi)].
\end{aligned} \tag{11}$$

Step 2: The conditional density function of next period's datum $y_{t+\tau}$ conditional upon data observed through date t can be evaluated by Fourier inversion of its characteristic function:

$$p(y_{t+1} \mid \mathbf{Y}_t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F_{y,x|t}(i\Phi, 0) e^{-i\Phi y_{t+1}} d\Phi. \tag{12}$$

Step 3: Using Proposition 1, the conditional characteristic function of next period's latent variable is

$$\begin{aligned}
 G_{t+\tau|t+\tau}(i\psi) &= \frac{\frac{1}{2\pi} \int_{-\infty}^{\infty} F_{y,x|t}(i\Phi, i\psi) e^{-i\Phi y_{t+\tau}} d\Phi}{p(y_{t+\tau}|t)} \\
 &= \frac{\frac{1}{2\pi} \int_{-\infty}^{\infty} G_{t|t}[D(\tau; i\Phi, i\psi)] e^{C(\tau; i\Phi, i\psi) - i\Phi y_{t+\tau}} d\Phi}{p(y_{t+\tau}|t)}.
 \end{aligned} \tag{13}$$

Step 4: Repeat steps 1-3 for subsequent values of t . Given underlying parameters θ , the likelihood function used in maximum likelihood estimation is $L(\mathbf{Y}_T|\theta) = \prod_{t=0}^{T-1} p(y_{t+1}|Y_t, \theta)$.

$G_{t|t}(i\psi)$ is the time- t prior characteristic function of the latent variable x_t , while $F_{y,x|t}$ in equation (11) is the time- t joint prior CF for (y_{t+1}, x_{t+1}) . Step 3 is the equivalent of Bayesian updating for characteristic functions, and yields the posterior CF of latent x_{t+1} -- which is also the time- $(t+1)$ prior CF for the next time step. The equivalent steps for updating moment generating functions and the associated conditional density functions are given in Table 1. For notational simplicity, a standardized time gap $\tau = 1$ is used, and the dependency of C and D on the time gap is suppressed. However, the algorithm can easily cope with irregular time gaps.

Filtered estimates of next period's latent variable realization and the accompanying precision can be directly computed from derivatives of the moment generating function $G_{t+1|t+1}(\psi)$ in (13):

$$\begin{aligned}
 \hat{x}_{t+1|t+1} &= G_{t+1|t+1}'(0) \\
 P_{t+1|t+1} &\equiv \text{Var}_{t+1}(x_{t+1}) \\
 &= G_{t+1|t+1}''(0) - \hat{x}_{t+1|t+1}^2.
 \end{aligned} \tag{14}$$

I.B Implementation

The recursion in (10) - (13) indicates that for a given conditional characteristic function $G_{t|t}(i\psi)$ of latent x_t and an observed datum $y_{t+\tau}$, it is possible to compute an updated CF $G_{t+\tau|t+\tau}(i\psi)$ that fully summarizes the filtered distribution of latent $x_{t+\tau}$. To implement the recursion, it is necessary to temporarily store the entire function $G_{t|t}(i\psi)$ in some fashion. This is an issue of *approximating functions* -- a subject extensively treated in Press *et al* (1992, Ch. 5) and Judd (1998, Ch.6). Using atheoretic methods such as splines or Chebychev polynomials, it is possible to achieve arbitrarily precise approximations to $G_{t|t}(i\psi)$.

However, it may not be easy to impose the appropriate shape restrictions that insure a given atheoretic approximating function $\hat{G}_{t|t}(i\psi)$ is indeed a legitimate characteristic function. A simple illustration of potential pitfalls arises with the symmetric Edgeworth distribution, with unitary variance and an excess kurtosis of κ_4 . The associated density and characteristic functions are

$$p(x) = \left[1 + \frac{\kappa_4}{24}(x^4 - 6x^2 + 3) \right] n(x) \tag{15}$$

$$G(i\psi) = e^{-\frac{1}{2}\psi^2} \left(1 + \frac{\kappa_4}{24}\psi^4 \right)$$

where $n(x)$ is the standard normal density function. The Edgeworth distribution requires $\kappa_4 \leq 4$ to preclude negative probabilities. And yet it is not obvious from inspecting $G(i\psi)$ that $\kappa_4 = 4$ is a critical value, and that using an approximating function equivalent to a numerical value of $\hat{\kappa}_4 = 4.005$ would generate invalid densities.

To avoid such potential problems, it appears safer to use approximating characteristic functions of latent realizations that are generated directly from distributions with known properties. As the recursion in (10) - (13) is just Bayesian updating, any legitimate approximate prior characteristic function $\hat{G}_{t|t}(i\psi)$ from a known distribution will generate a legitimate posterior characteristic function $\hat{G}_{t+\tau|t+\tau}(i\psi)$. In particular, mixtures of distributions are a standard method of approximating arbitrary distributional forms. The applications below use the gamma distribution, to enforce nonnegativity constraints on the latent stochastic volatility realizations. Mixtures of gammas will be explored in future extensions if necessary.

2. A Monte Carlo examination of the approach

To illustrate the approach and to test the accuracy of the approximating functions, the following simplified discrete-time stochastic volatility process for log-differenced asset prices was simulated:

$$\begin{aligned} y_{t+1} &\sim N[0, V_t \Delta t] \\ dV &= (\alpha - \beta V_t) dt + \sigma \sqrt{V} dW \end{aligned} \quad (16)$$

with asset returns and variance shocks assumed independent. This process is the affine equivalent of the benchmark discrete-time stochastic volatility model studied by Harvey, Ruiz, and Shephard (1994), Jacquier, Polson and Rossi (1994), and Kim, Shephard and Chib (1998). Rather than the conditionally Gaussian AR(1) process for log volatility of those papers,² however, the square-root process for volatility implies a noncentral chi-squared transition density in discrete time:

$$\frac{2V_{t+1}}{K} \mid V_t \sim \chi^2 \left(\frac{4\alpha}{\sigma^2}, \frac{2V_t}{K} e^{-\beta \Delta t} \right) \text{ for } K \equiv \frac{\sigma^2}{2\beta}. \quad (17)$$

Given the assumed conditional independence of volatility increments and asset returns, the joint conditional characteristic function in (1) above takes the separable form

$$\begin{aligned} F_{y,V}(i\Phi, i\psi \mid V_t) &\equiv E \left[e^{i\Phi y_{t+1} + i\psi V_{t+1}} \mid V_t \right] \\ &= \exp \left[-\frac{2\alpha}{\sigma^2} \ln(1 - iK\psi) + \left(-\frac{1}{2} \Phi^2 \Delta t + \frac{e^{-\beta \Delta t} i\psi}{1 - iK\psi} \right) V_t \right]. \end{aligned} \quad (18)$$

The unconditional density of V_0 has a gamma distribution, with associated characteristic function

$$G_{0|0}(i\psi) = (1 - i\kappa_0 \psi)^{-\nu_0} \quad (19)$$

² For a discrete-time log volatility process, the data transformation $z_{t+1} \equiv \ln |y_{t+1} - E(y_{t+1})|$ generates an affine state space system in z and log volatility for which this article's methodology can be used. The square-root variance process (16) is used instead as a simple discrete-time approximation to the continuous-time affine specifications examined below.

for $v_0 \equiv 2\alpha/\sigma^2$ and $\kappa_0 = \sigma^2/2\beta$. The unconditional mean and variance of V_0 are $\kappa_0 v_0 = \alpha/\beta$ and $\kappa_0^2 v_0 = (\alpha/\beta)(\sigma^2/2\beta)$, respectively.

The simplest approximating function in subsequent periods to assume a gamma distribution:

$$\hat{G}_{t|i}(i\psi) = (1 - i\kappa_t \psi)^{-v_t}. \quad (20)$$

This approximation is exact for $t = 0$, and imposes appropriate nonnegativity constraints for $t > 0$. The density is summarized at each date t by the parameter pair (κ_t, v_t) . Furthermore, updated estimates (κ_{t+1}, v_{t+1}) can be generated from the moment conditions from equations (13) and (14):

$$\begin{aligned} \kappa_{t+1} v_{t+1} &= \hat{V}_{t+1|t+1} \\ &= G_{t+1|t+1}'(0) \\ \kappa_{t+1}^2 v_{t+1} &= \text{Var}_{t+1}(V_{t+1}) \\ &= G_{t+1|t+1}''(0) - \hat{V}_{t+1|t+1}^2. \end{aligned} \quad (21)$$

Further details are in the appendix. The resulting updated approximation $\hat{G}_{t+1|t+1}(\psi) \equiv (1 - \kappa_{t+1} \psi)^{-v_{t+1}}$ matches the first and second moments of the true $G_{t+1|t+1}(\psi)$ computed from (13) for different values of ψ , and is therefore comparable to a second-order Taylor approximation.

This approximation methodology somewhat resembles that of Ruiz (1994) and Harvey, Ruiz, and Shephard (1994). The HRS approach assumes an approximate Gaussian distribution for log absolute returns, and uses Kalman filtration to update the mean and variance of log volatility. The HRS approach is consistent, but efficiency is reduced by the fact that the distribution of log absolute returns is poorly approximated by the Gaussian distribution. Here, the exact distributional properties of observed data are used in (13) when updating the posterior conditional mean and variance of the latent variable. Approximation enters in the specification of the prior density for latent variance, which affects the relative weights of the data and the prior.

2.B. Simulations

Data were simulated based upon the stochastic volatility estimates of Andersen, Benzoni and Lund (2002) from daily data over 1953-1996: $\alpha = .0438$, $\beta = 3.2508$, and $\Delta t = 1/252$, in annualized units. Andersen *et al* (henceforth ABL) estimated the average annualized variance at $(0.116)^2$, while variance shocks mean-revert with an estimated half-life of 2.56 months. The volatility of volatility parameter was set to $\sigma = \sqrt{4\alpha/5} = .187$, instead of the ABL estimate of .185. This produces an integer number (five) of degrees of freedom for the noncentral χ^2 density in (17), permitting exact Monte Carlo generation of noncentral χ^2 shocks from the sum of squared normal shocks.

The initial variance V_0 was independently drawn from the unconditional gamma density for each sample path, and 1000 variance sample paths $\{V_t\}_{t=0}^{T-1}$ were independently generated for various values of T . Normally distributed log-differenced returns were generated given the variance realizations. The parameters $\hat{\theta} = \{\hat{\alpha}, \hat{\beta}, \hat{\sigma}\}$ for each set of returns were estimated by maximizing the likelihood function described above, with true parameter values used as starting values. For comparison, the parameters $\hat{\theta}_V$ were also estimated by direct maximum likelihood *conditional* upon observing the $\{V_t\}_{t=0}^{T-1}$ sample path, using the noncentral χ^2 transition densities and initial gamma distribution.

Table 2 summarizes the results. The estimation methodology is consistent, with the estimated bias (average bias rows) for all parameters and parameter transformations decreasing with longer data samples. The fastest convergence is for the average variance α/β and the volatility of variance parameter σ . The slowest is for the parameter β that determines serial correlation and the associated half-life estimates; some bias remains even for 48 years' of data. Those biases reflect in magnified fashion the small-sample biases for a persistent series that would exist even if the $V(t)$ series were directly observed, as illustrated in the second set of $\hat{\beta}$ estimates.³ Here, of course, the values of that series must be inferred from noisy returns data, which almost doubles the magnitude of the bias.

³See Nankervis and Savin (1988) and Stambaugh (1999) for a discussion of these biases.

The estimation procedure for the parameters of the latent stochastic variance process performs surprisingly well. Squared daily returns are very noisy signals of daily latent variance, and yet the RMSE of returns-based parameter estimates is typically less than double that of estimates conditioned on actually observing the underlying variance. Furthermore, parameter estimates for α and β are highly correlated with the estimates conditional on directly observing $V(t)$ data.

An interesting exception is the volatility of variance estimator $\hat{\sigma}$. Were $V(t)$ data observed, its volatility parameter σ would be pinned down quite precisely even for relatively small data sets; e.g., a RMSE of .004 on a data set of 1000 observations. By contrast, when $\{V(t)\}$ must be inferred from noisy stock returns, the imprecision of the $\hat{V}_{t|t}$ estimates increases the RMSE of the $\hat{\sigma}$ estimate tenfold.

2.C Filtration

Figures 1 and 2 illustrate the accuracy of the *volatility* filtration $E_t\sqrt{V_t}$ conditional upon using the true parameters, for the first 1000 observations (four years) of a generated sample of 12,000 observations. Since the conditional distribution of latent variance V_t is gamma (or scaled χ^2), the conditional distribution of latent volatility has a scaled χ distribution, with noncentral moments

$$E\left[\left(\sqrt{V_t}\right)^n \mid \mathbf{Y}_t\right] = \frac{\kappa_t^{n/2} \Gamma(v_t + 1/2)}{\Gamma(v_t)}, \quad (22)$$

where $\Gamma(\bullet)$ is the gamma function.

The volatility estimate begins at the unconditional mean of 11.0% at time 0, but converges rapidly towards the true value as returns are observed. Changes in the filtered volatility perform lag behind changes in the true volatility, since the filtered estimate is inferred from past squared returns. The absolute divergence is typically less than 4%, but is occasionally larger. To put this error in perspective: a volatility estimate (towards the end of the sample) of 11% when the true volatility is 16% is a substantial error from an option pricing perspective. The magnitude of this error reflects the low informational content of squared daily returns for estimating latent volatility and variance.

A key issue is whether the filtered gamma densities are indeed reasonably accurate approximations to the true conditional distributions of latent variances. On simulated data, this can be directly assessed by examining the frequency with which V_t realizations fall within the quantiles of the conditional $V_{t|t}$ gamma distributions. The realized frequencies over a run of 12,000 observations indicate the gamma approximation is indeed accurate :

Quantile p :	.05	.10	.25	.5	.75	.90	.95
$Pr\hat{ob}[V(t) < \hat{V}_p(t) I_t]$:	.043	.090	.243	.492	.743	.897	.951

where $\hat{V}_p(t)$ is the filtered gamma quantile with lower tail probability p at time t . Consequently, there appears to be no need to use more complicated approximating functions such as a mixture of gammas.

3. Estimates from stock index returns

There is a growing body of work on the estimation of continuous-time affine and non-affine stochastic volatility/jump models for stock market returns by simulation methods. Three recent contributions are Andersen, Benzoni, and Lund (2002), Chernov, Gallant, Ghysels, and Tauchen (2002), and Eraker, Johannes, and Polson (2000). The first two papers (henceforth ABL and CGGT, respectively) use the SNP/EMM methodology of Gallant and Tauchen for daily stock index returns over 1953-96 and 1953-99, respectively. Eraker *et al* use Bayesian MCMC methods for daily S&P returns over 1980-99, as well as NASDAQ returns over 1985-99. Included among the affine specifications considered in those papers is the following continuous-time affine model:

$$\begin{aligned}
 dS/S &= (\mu_0 + \mu_1 V - \lambda_t \bar{k}) dt + \sqrt{V} dW + (e^\gamma - 1) dN \\
 dV &= (\alpha - \beta V) dt + \sigma \sqrt{V} dW_V \\
 \text{Corr}[dW, dW_V] &= \rho \\
 \text{Prob}[dN = 1] &= (\lambda_0 + \lambda_1 V) dt \\
 \gamma &\sim N[\bar{\gamma}, \delta^2]
 \end{aligned} \tag{23}$$

where $\bar{k} = \exp[\bar{\gamma} + \frac{1}{2}\delta^2] - 1$. The latter two papers also examine the interesting affine specification in which there are jumps in latent variance that may be correlated with stock market jumps.

This article uses the 11,076 daily S&P 500 returns over 1953 through 1996 that formed the basis for Andersen, Benzoni and Lund's (2002) EMM/SNP estimates.⁴ I will not repeat the data description in that article, but two comments are in order. First, Andersen *et al* prefiltered the data to remove an MA(1) component that may be attributable to nonsynchronous trading in the underlying stocks. Second, there were three outliers that created some numerical difficulties for likelihood evaluation: the -22% stock market crash of October 19, 1987, the 7% drop on September 26, 1955 that followed reports of President Eisenhower's heart attack, and the 6% mini-crash on October 13, 1989. Details of computing probability densities for such outliers are given in the appendix, as is the functional form of

⁴I am indebted to Luca Benzoni for providing the data.

the joint likelihood function.

The first three columns of Table 3 present estimates of the stochastic volatility model without jumps (SV) from Chernov *et al*, Andersen *et al*, and the methodology of this paper.⁵ As discussed in CGGT, estimating the parsimonious stochastic volatility model without jumps creates conflicting demands for the volatility mean reversion parameter β and the volatility of volatility parameter σ . Extreme outliers such as the 1987 crash can be explained by highly volatile volatility that mean-reverts within days, whereas standard volatility persistence suggest lower volatility of volatility and slower mean reversion. In CGGT's estimates, the former effect dominates; in ABL's estimates, the latter dominates.

My estimates are affected by both phenomena, but matching the volatility persistence clearly dominates. While constraining σ to the CGGT estimate of 1.024 substantially raises the likelihood of the outliers in 1955, 1987, and 1989, this is more than offset by likelihood reductions for the remainder of the data. The overall log likelihood falls from 39,234 to 39,049 – a strong rejection of the constraint with an associated P-value less than 10^{-16} . And although the CGGT data set includes a few outliers in 1997-99 that are not in the ABL data set used here, the likelihood impact per outlier of a larger $\hat{\sigma}$ seems insufficient to explain the difference in results.⁶

Although my stochastic volatility parameter estimates are qualitatively similar to those of ABL on the same data set, there are substantial differences. In particular, I estimate a higher volatility of volatility (.315 instead of .197) and faster volatility mean reversion (half-life of 1.4 months, instead of 2.1 months). The estimate of the average annualized level of variance is also higher: $(.125)^2$, rather than $(.112)^2$. The estimates of the correlation between volatility and return shocks are comparable. The substantial reduction in log likelihood of the six ABL parameter estimates is strongly significant statistically, with a P-value of

⁵The ABL estimates are from their Table IV, converted to an annualized basis.

⁶It is possible the difference in estimates is attributable to how different specifications of the SNP discrete-time auxiliary model interact with outliers. ABL specify an EGARCH-based auxiliary model to capture the correlation between return and volatility shocks. CGGT use a GARCH framework, and capture the volatility-return correlation through terms in the Hermite polynomials.

9×10^{-16} . It appears that the two-stage SNP/EMM methodology used by Andersen *et al* generates a objective function for parameter estimation that is substantially different from my maximum likelihood methodology.

As found in the earlier studies, adding a jump component substantially improves the overall fit. As indicated in the middle three columns of Table 2, I estimate a more substantial, less frequent jump component than previous studies: three jumps every four years, of average size -1.0% and standard deviation 5.2%. As outliers are now primarily explained by the jump component, the parameters governing volatility dynamics parameters are modified: σ drops, and the half-life of volatility shocks lengthens. The divergence of parameter estimates from the ABL estimates is again strongly significant statistically.

Bates (2000) showed that a volatility-dependent jump intensity component $\lambda_1 V_t$ helps explain the cross-section of stock index option prices. Some time series evidence for the specification was provided in Bates and Craine (1999), while Eraker et al (2000) found stronger empirical support. In contrast to the results in ABL, the final two columns of Table 2 indicate that jumps are indeed more likely when volatility is high. The hypothesis that $\lambda_1 = 0$ is rejected at a P-value of 5×10^{-8} . The time-invariant jump component λ_0 ceases to be statistically significant when λ_1 is added.

Standard maximum likelihood diagnostics dating back to Pearson (1933) can be used to assess model specification. As in Bates (2000), I use the normalized transition density

$$z_{t+1} = N^{-1}[CDF(y_{t+1} | \mathbf{Y}_t, \hat{\theta})] \quad (24)$$

where N^{-1} is the inverse of the cumulative normal distribution function, and the cumulative distribution function CDF is evaluated from the conditional characteristic function by Fourier inversion given parameter estimates $\hat{\theta}$. Under correct specification, the z 's should be independent and identical draws from a normal distribution. The approach is equivalent to a standard Qq plot of the CDF realizations; the additional inverse normal transformation usefully highlights outliers.

The histogram of normalized returns in Figure 3 indicates the SVJ1 model fits most returns quite well. A major remaining outlier is, however, the 1987 crash, which is interpreted as

a five standard deviation draw from a jump-contingent normal distribution with standard deviation 3.9%, and has a z -score of -5.52. As z -scores less than -5 should be observed only once every 14,000 years, that single 1987 outlier constitutes substantial evidence against the model.

To address this issue, an additional stochastic volatility/jump model is estimated in which jumps in log-differenced prices are drawn from a mixture of normals. The resulting estimates for the stochastic volatility component are roughly unchanged; the jump parameters become

$$Prob[dN = 1] = 133.44 V_t, \quad \ln(1 + \tilde{k}) \sim \begin{cases} N[.001, (.029)^2] & \text{with prob. .982} \\ N[-.222, (.007)^2] & \text{with prob. .018} \end{cases} \quad (25)$$

The conditional distribution of daily returns is approximately a mixture of three normals: normal daily volatility that varies stochastically over a range of .2% - 1.8%, infrequent symmetric jumps with a larger standard deviation of 2.9% and time-varying arrival rate that averages 1.85 jumps per year, and an extremely infrequent crash corresponding to the 1987 outlier. Log likelihood rises from 39,309.51 to 39,317.81 -- an improvement almost entirely attributable to a better fit for the 1987 crash. Correspondingly, the z -score on that day drops in magnitude, to a more plausible value of -3.50. The increase in log likelihood has a marginal significance level of .0008 under a likelihood ratio test, given 3 additional parameters.

3.B. Filtration

Figure 4 illustrates the filtered estimates of latent volatility $\sqrt{V_t}$ from the SVJ1 model, and the difference between SVJ1 and SV volatility estimates. Those estimates are generally almost identical, except following large positive or negative stock returns. For instance, the 1955 and 1987 crashes have much more of an impact on volatility assessments under the SV model than under the jump models.

Figure 5 illustrates the typical impact of asset returns on volatility assessments, using the “news impact curves” $E_{t+1}\sqrt{V_{t+1}} - E_t\sqrt{V_{t+1}}$ employed by Hentschel (1995) when assessing GARCH models. The volatility estimates were computed from the conditional moments of

latent variance,⁷ using the second-order Taylor approximation

$$E[\sqrt{V}] \approx \sqrt{E[V]} \left[1 - \frac{1}{8} \frac{\text{Var}[V]}{E[V]^2} \right]. \quad (26)$$

The estimates were calibrated from a median volatility day (October 14, 1960) with a volatility estimate of 11.4%, and initial filtered gamma distribution parameters $(\kappa_t, \nu_t) = (.00229, 5.89)$.

All news impact curves are tilted, with negative returns having a larger impact on volatility assessments than positive returns. All models process the information in small asset returns similarly. The most striking result, however, is that taking jumps into account implies that volatility updating becomes a *non-monotonic* function of the magnitude of asset returns. Under the SVJ0 model, large moves indicate a jump has occurred, which totally obscures any information in returns regarding latent volatility for moves in excess of seven standard deviations. Under the SVJ1 model, the large-move implication that a jump has occurred still contains some information regarding volatility, given jump intensities are proportional to latent variance. Neither case, however, resembles the U- and V-shaped GARCH news impact curves estimated by Hentschel (1995).

Even the SV filtration is not necessarily compatible with GARCH filtrations. In the stochastic volatility model, the standard deviation of the volatility prior varies stochastically over time over a range of 1 - 4%, in a fashion that is correlated with but not fully summarized by the volatility estimate $E_t\sqrt{V_t}$. The relative weighting of the information in asset returns and the volatility prior therefore changes over time in a fashion that cannot be captured by

⁷The mean and variance of $V_{t+1|t}$ can be derived from those of $V_{t|t}$ ($\kappa_t \nu_t$ and $\kappa_t \nu_t^2$, respectively) using the moments of the non-central χ^2 conditional transition density and the relationships

$$\begin{aligned} E_t(V_{t+1}) &= E_t[E(V_{t+1} | V_t)] \\ \text{Var}_t(V_{t+1}) &= E_t[\text{Var}(V_{t+1} | V_t)] + \text{Var}_t[E(V_{t+1} | V_t)]. \end{aligned}$$

The mean and variance of $V_{t+1|t+1}$ are $\kappa_{t+1} \nu_{t+1}$ and $\kappa_{t+1}^2 \nu_{t+1}$ respectively, updated from (κ_t, ν_t) conditional on the observed asset return using the algorithm in (11) - (13).

GARCH specifications. It is an open question as to whether this variable weighting improves volatility and variance assessments relative to GARCH estimates, and whether it outweighs the constraints imposed by the affine SV model on the functional form of the news impact curve. As discussed by Hentschel, GARCH specifications can be quite flexible regarding that functional form.

The number of jumps on any given day is also a latent variable that can be inferred from observed returns. It is shown in the appendix that the joint conditional distribution of log-differenced asset prices and the number of jumps $\Delta N \equiv N(t+1) - N(t)$ has an affine specification. The characteristic function $G_{\Delta N}(i\xi | \mathbf{Y}_{t+1}) \equiv E[e^{i\xi \Delta N_{t+1}} | y_{t+1}, \mathbf{Y}_t]$ can therefore be evaluated by Proposition 1.

While it is possible to evaluate the daily probability that n jumps occurred by Fourier inversion of $G_{\Delta N}$, it is simpler to estimate the number of jumps: $E[\Delta N_{t+1} | \mathbf{Y}_{t+1}] = G_{\Delta N}'(0 | \mathbf{Y}_{t+1})$. At the daily horizon, ΔN is essentially binomial, and the estimated number of jumps is approximately the probability that a jump occurred. Unsurprisingly from Figure 6, large moves are attributed to jumps and small moves are not. Intermediate moves of roughly three to five times the estimated latent standard deviation imply a small probability that a jump occurred. It is the accumulation of these small jump probabilities for the moderately frequent intermediate-sized moves that underpin the overall estimate of jump intensities; e.g., .744 jumps per year in the SVJ0 model.

4. Conclusions and extensions

This article has presented a new methodology for estimating continuous-time affine processes on discrete-time data: both parameter values and latent variable realizations. Somewhat unexpectedly, the parameter estimates of an affine stochastic volatility/jump process on the data set of Andersen, Benzoni and Lund (2002) data base are significantly different from the ABL estimates. In particular, I find 1) a generally higher volatility of volatility, 2) a more substantial jump component, and 3) strong support for the hypothesis that jumps are more likely when volatility is high.

My methodology differs substantially from the SNP/EMM methodology used by Andersen *et al*, so the source of divergence is not immediately apparent. It appears that the SNP/EMM

methodology may be quite sensitive to precisely how the auxiliary discrete-time SNP model is specified -- especially in the presence of infrequent large outliers such as the 1987 crash. Chernov *et al* (2002) find very different estimates from Andersen *et al* for the stochastic volatility model, for a similar data set but a different auxiliary model. My methodology does not require an auxiliary discrete-time model. The resulting SV estimates are between the two sets of EMM-based estimates, but are substantially closer to those of Andersen *et al*.

This article has focused on classical maximum likelihood estimation. However, the recursive likelihood evaluation methodology presented here can equally be used in Bayesian analysis, when combined with a prior distribution on parameter values.

More recent research into volatility dynamics has focused on the additional information provided by alternate data sources: high-low ranges, “realized” intradaily variances, and implicit variances from option prices. The latter two approaches definitely possess the requisite affine structure, and can therefore be handled within this framework. For instance, the joint characteristic function of asset returns, latent variance, and integrated variance has an exponentially affine form

$$\begin{aligned} F_{y,V,\bar{V}}(i\Phi, i\psi, i\xi) &\equiv E\left[\exp\left(i\Phi y_{t+1} + i\psi V_{t+1} + i\xi \int_{\tau=t}^{t+1} V_{\tau} d\tau\right) \mid V_t\right] \\ &= \exp\left[C(i\Phi, i\psi, i\xi) + D(i\Phi, i\psi, i\xi) V_t\right] \end{aligned} \quad (27)$$

that can be identified by solving the relevant Feynman-Kac partial differential equation.

Such expanded-data approaches are feasible, but are numerically more complex. For instance, extracting the characteristic function of latent $V_{t+1|t+1}$ from observed (y_{t+1}, \bar{V}_{t+1}) requires bivariate integration. Furthermore, it appears from Alizadeh, Brandt and Diebold (2002) and Andersen, Bollerslev, Diebold and Ebens (2001) that the additional data are sufficiently informative about latent variance that single-factor models no longer suffice. The complexities of using additional data sources in conjunction with multi-factor models of latent variance will therefore be explored in future research.

Appendix

A.1 Joint moment generating functions from continuous-time affine processes

Let

$$F(\Phi, \psi, \xi | s_t, V_t, N_t, t) \equiv E_t[e^{\Phi s_T + \psi V_T + \xi N_T} | s_t, V_t, N_t] \quad (\text{A.1})$$

be the joint moment generating function of the future variables $\mathbf{z}_T \equiv (s_T, V_T, N_T)$ conditional upon observing \mathbf{z}_t today, where s_t is the log asset price, V_t is the instantaneous variance, and N_t is the number of jumps that have occurred. Since F is a conditional expectation, changes in F are unpredictable:

$$E_t[dF(\bullet | s_t, V_t, N_t, t)] = 0. \quad (\text{A.2})$$

Expanding this by the jump-diffusion generalization of Itô's lemma yields the backwards Kolmogorov equation that F must solve for a given stochastic process. For the affine processes in this article, the solution is exponentially affine in the state variables:

$$\begin{aligned} F(\Phi, \psi, \xi | s_t, V_t, N_t, \tau) \\ = \exp[\Phi s_t + \xi N_t + C(\tau; \Phi, \psi, \xi) + D(\tau; \Phi, \psi, \xi) V_t] \end{aligned} \quad (\text{A.3})$$

for $\tau \equiv T - t$.

By Itô's lemma, the stochastic volatility/jump-diffusion in equation (23) implies a log asset price evolution of the form

$$\begin{aligned} ds &= [\mu_0 + (\mu_1 - 1/2)V - (\lambda_0 + \lambda_1 V)\bar{k}] dt + \sqrt{V} dW + \gamma dN \\ dV &= (\alpha - \beta V) dt + \sigma\sqrt{V} dW_v \end{aligned} \quad (\text{A.4})$$

where $\text{Cov}(dW, dW_v) = \rho dt$, N_t is a Poisson counter with jump intensity $\lambda_0 + \lambda_1 V_t$, γ is normally distributed $N[\bar{\gamma}, \delta^2]$, and $\bar{k} \equiv e^{\bar{\gamma} + 1/2\delta^2} - 1$. The corresponding backwards Kolmogorov equation for F is

$$\begin{aligned} F_\tau &= [\mu_0 + (\mu_1 - 1/2)V - (\lambda_0 + \lambda_1 V)\bar{k}]F_s + (\alpha - \beta V)F_V \\ &+ 1/2V(F_{ss} + 2\rho\sigma_v F_{sV} + \sigma_v^2 F_{VV}) \\ &+ (\lambda_0 + \lambda_1 V)E[F(\bullet | s + \gamma, N + 1, V; \tau) - F] \end{aligned} \quad (\text{A.5})$$

which is solved subject to the boundary condition

$$F(\Phi, \psi, \xi | s, V, N; \tau = 0) = e^{\Phi s + \psi V + \xi N}. \quad (\text{A.6})$$

Plugging (A.3) into (A.5) yield a recursive system of ordinary differential equations that $C(\tau; \bullet)$ and $D(\tau; \bullet)$ must solve, subject to the boundary conditions $C(0; \Phi, \psi, \xi) = 0$ and $D(0; \Phi, \psi, \xi) = \psi$. The resulting solutions are:

$$E(\Phi, \xi) = e^{\xi + \bar{\gamma}\Phi + \frac{1}{2}\delta^2\Phi^2} - (1 + \bar{k}\Phi) \quad (\text{A.7})$$

$$\begin{aligned} C(\tau; \Phi, \psi, \xi) &= \mu_0\Phi T - \frac{\alpha\tau}{\sigma^2} (\rho\sigma\Phi - \beta - \gamma) + \lambda_0\tau E(\Phi, \xi) \\ &- \frac{2\alpha}{\sigma^2} \ln\left[1 + \frac{1}{2}(\rho\sigma\Phi - \beta - \gamma) \frac{1 - e^{\gamma\tau}}{\gamma}\right] - \frac{2\alpha}{\sigma^2} \ln[1 - K(\Phi, \xi)\psi] \end{aligned} \quad (\text{A.8})$$

$$D(\tau; \Phi, \psi, \xi) = \frac{-2(\mu_1 - \frac{1}{2})\Phi - \Phi^2}{\rho\sigma\Phi - \beta + \gamma \frac{1 + e^{\gamma\tau}}{1 - e^{\gamma\tau}}} + \frac{\Lambda(\Phi, \xi)\psi}{1 - K(\Phi, \xi)\psi} \quad (\text{A.9})$$

$$\gamma = \sqrt{(\rho\sigma\Phi - \beta)^2 - 2\sigma^2[\frac{1}{2}\Phi^2 + (\mu_1 - \frac{1}{2})\Phi] + \lambda_1 E(\Phi, \xi)} \quad (\text{A.10})$$

$$\Lambda(\Phi, \xi) = \frac{\left(\frac{e^{\gamma\tau} + 1}{e^{\gamma\tau} - 1}\right)^2 - 1}{\left(\frac{e^{\gamma\tau} + 1}{e^{\gamma\tau} - 1} + \frac{\beta - \rho\sigma\Phi}{\gamma}\right)^2} \quad (\text{A.11})$$

$$K(\Phi, \xi) = \frac{\sigma^2}{\gamma \frac{e^{\gamma\tau} + 1}{e^{\gamma\tau} - 1} + \beta - \rho\sigma\Phi}. \quad (\text{A.12})$$

The moment generating functions underlying the marginal transition densities of s_T , V_T , and N_T are of course given by $F(\Phi, 0, 0 | \bullet)$, $F(0, \psi, 0 | \bullet)$, and $F(0, 0, \xi | \bullet)$, respectively.⁸

Given (A.3), the joint moment generating function of log-differenced prices $y \equiv \Delta s = \ln(S_{t+\Delta t}/S_t)$, number of jumps $\Delta N \equiv N_{t+\Delta t} - N_t$, and future variance $V_{t+\tau}$ conditional upon observing V_t is

$$F_{y, V, \Delta N}(\Phi, \psi, \xi | V_t, \Delta t) = \exp[C(\Delta t; \Phi, \psi, \xi) + D(\Delta t; \Phi, \psi, \xi) V_t]. \quad (\text{A.13})$$

The case of $\xi \neq 0$ is used only for inferences about the occurrences of jumps. In variance filtration and maximum likelihood estimation, ξ is set to zero and the joint transform in equation (1) is

$$F_{y, V}(\Phi, \psi | V_t, \Delta t) = F_{y, V, \Delta N}(\Phi, \psi, 0 | V_t, \Delta t). \quad (\text{A.14})$$

As discussed above and illustrated in Table 1, the joint conditional characteristic function of (y_{t+1}, V_{t+1}) conditional upon data \mathbf{Y}_t observed through time t is

$$F_{y, V | t}(i\Phi, i\psi | \mathbf{Y}_t) = e^{C(i\Phi, i\psi, 0; \Delta t)} G_{t|t}[D(i\Phi, i\psi, 0; \Delta t)] \quad (\text{A.15})$$

where $G_{t|t}(\psi) \equiv E[e^{\psi V_t} | \mathbf{Y}_t]$ is the conditional moment generating function of V_t . As discussed in Section 2, this is approximated by the gamma MGF $\hat{G}_{t|t}(\psi) \equiv -v_t \ln(1 - \kappa_t \psi)$.

A.2 Filtration

Density evaluation and variance updating involves three numerical integrations at each date t . The density is evaluated by Fourier inversion of (A.15):

$$p(y_{t+1} | t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F_{y, V | t}[i\Phi, 0] e^{-i\Phi y_{t+1}} d\Phi. \quad (\text{A.16})$$

The noncentral moments $E_{t+1}[V_{t+1}^n]$ are evaluated by taking analytic derivatives of equation (13) with respect to ψ :⁹

⁸For $\Phi = \xi = 0$, $\gamma = \beta$, $\Lambda(0, 0) = e^{-\beta\tau}$, and $\mathbf{K}(0, 0) = \frac{1}{2}\sigma^2(1 - e^{-\beta\tau})/\beta$.

⁹Numerical derivatives are also feasible, but reduce the accuracy of the numerically computed log likelihood gradient used in maximum likelihood estimation..

$$\begin{aligned}
E_{t+1} [V_{t+1}^n] &= \left. \frac{\partial^n G_{t+\Delta t|t+\Delta t}(\psi)}{\partial \psi^n} \right|_{\psi=0} \\
&= \frac{1}{2\pi p(y_{t+1}|t)} \int_{-\infty}^{\infty} \left. \frac{\partial^n F_{y,V|t}[i\Phi, \psi]}{\partial \psi^n} \right|_{\psi=0} e^{-i\Phi y_{t+1}} d\Phi \quad (\text{A.17}) \\
&= \frac{1}{2\pi p(y_{t+1}|t)} \int_{-\infty}^{\infty} \left. \frac{\partial^n e^{C(i\Phi, \psi)} G_{t|t}[D(i\Phi, \psi)]}{\partial \psi^n} \right|_{\psi=0} e^{-i\Phi y_{t+1}} d\Phi .
\end{aligned}$$

These moments were then used in equations (14) to update next period's conditional moment generating function $G_{t+1|t+1}(\psi)$.

For each integral, an upper limit Φ_{\max} was computed analytically for which truncation error would be less than $\varepsilon = 10^{-10}$. The integral was then computed numerically to 10^{-10} accuracy using IMSL's adaptive DQDAG integration routine over $(0, \Phi_{\max})$.

A.3 Inferring jumps

A similar procedure was used to infer whether a jumps occurred on a given day. From (A.13), the prior joint characteristic function of $(y_{t+1}, \Delta N_{t+1})$ conditional on data through date t is

$$F_{y, \Delta N|t}(i\Phi, i\xi | \mathbf{Y}_t) = e^{C(i\Phi, 0, i\xi; \Delta t)} G_{t|t}[D(i\Phi, 0, i\xi; \Delta t)]. \quad (\text{A.18})$$

Proposition 1 can then be invoked to compute the *ex post* characteristic function

$$G_{\Delta N|t+1}(i\xi) \equiv E[e^{i\xi \Delta N_{t+1}} | y_{t+1}, \mathbf{Y}_t]. \quad (\text{A.19})$$

The *ex post* moments of the number of jumps $E[(\Delta \tilde{N}_{t+1})^n | \mathbf{Y}_{t+1}]$ can be computed analogously to the approach in (A.17). In particular, since the number of jumps on a given day is approximately a $(0, 1)$ binomial variable, the estimated number of jumps $E[\Delta \tilde{N}_{t+1} | \mathbf{Y}_{t+1}]$ is approximately the probability that a jump occurred on date $t+1$. Since the analytic derivatives of (A.18) with respect to ξ are messy, numerical derivatives of (A.19) were used.

A.4 Outliers

For extreme outliers, $p(y)$ takes on near-zero values (e.g., 2×10^{-22} for the 1987 crash under the SV model), creating underflow problems for the numerical integration in (A.17). For such outliers, a *scaled* density function can be evaluated more accurately. The Fourier transform of the scaled density function $e^{ay^*} p(y^*)$ of the standardized return $y^* \equiv y_{t+1} / \sqrt{\hat{V}_{t|t} \Delta t}$ is

$$\begin{aligned} \int [e^{ay^*} p(y^*)] e^{i\Phi y^*} dy &= E[e^{(a+i\Phi)y^*}] \\ &= E\left[\exp\left(\frac{(a+i\Phi)y}{\sqrt{V\Delta t}}\right)\right] \\ &= \exp\left[C\left(\frac{a+i\Phi}{\sqrt{V\Delta t}}, 0\right)\right] G_{t|t}\left[D\left(\frac{a+i\Phi}{\sqrt{V\Delta t}}, 0\right)\right] \end{aligned} \quad (\text{A.20})$$

so the scaled density can be computed by Fourier inversion:

$$e^{ay^*} p(y^*) = \frac{1}{2\pi} \int \exp\left[C\left(\frac{a+i\Phi}{\sqrt{V\Delta t}}, 0\right)\right] G_{t|t}\left[D\left(\frac{a+i\Phi}{\sqrt{V\Delta t}}, 0\right)\right] e^{-i\Phi y^*} d\Phi. \quad (\text{A.21})$$

The density of observed data is then $p(y) = e^{-ay^*} [e^{ay^*} p(y^*)] / \sqrt{\hat{V}_{t|t} \Delta t}$. The transformation is the Fourier transform equivalent of importance sampling for low-probability events in Monte Carlo integration. Similar transformations can be used when updating the moments of latent variance in (A.17).

No transformation was necessary for $|y^*| < 6$, leaving only three observations for which numerical integration was occasionally problematic: the crashes of 1955, 1987, and 1989. How fast tail probabilities fall off depends on the model: rapidly for the stochastic volatility model, slower for jump models. With only three outliers, it was simplest to set the parameter a on a case by case basis:

- SV: $a = 2$ for the 1955 and 1989 crashes, $a = 4$ for the 1987 crash;
- SVJ0, SVJ1: $a = 1$ for the 1987 crash, 0 otherwise;
- SVJ2: $a = 0$ always.

Experimentation confirmed that density evaluations are insensitive to the precise value of a .

References

- Alizadeh, Sassan, Michael W. Brandt, and Francis X. Diebold (2002). "Range-based Estimation of Stochastic Volatility Models." *Journal of Finance* **57**, 1047-1091.
- Andersen, Torben G., Luca Benzoni, and Jesper Lund (2002). "An Empirical Investigation of Continuous-Time Equity Return Models." *Journal of Finance* **57**, 1239-1284.
- Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Heiko Ebens (2001). "The Distribution of Stock Return Volatility." *Journal of Financial Economics* **61**, 43-76.
- Bates, David S. (2000). "Post-'87 Crash Fears in the S&P 500 Futures Option Market." *Journal of Econometrics* **94**, 181-238.
- Bates, David S. and Roger Craine (1999). "Valuing the Futures Market Clearinghouse's Default Exposure During the 1987 Crash." *Journal of Money, Credit, and Banking* **31**, 248-272.
- Chernov, Mikhail, A. Ronald Gallant, Eric Ghysels, and George Tauchen (2002). "Alternative Models for Stock Price Dynamics." *Journal of Econometrics*, forthcoming.
- Eraker, Bjorn, Michael Johannes, and Nicholas G. Polson (2000). "Evidence for and the Impact of Jumps in Volatility and Returns." University of Chicago working paper.
- Fridman, Moshe and Lawrence Harris (1998). "A Maximum Likelihood Approach for Non-Gaussian Stochastic Volatility Models," *Journal of Business and Economic Statistics* **16**, 284-291.
- Gallant, A. Ronald and George Tauchen (2001). "Simulated Score Methods and Indirect Inference for Continuous-Time Models." University of North Carolina working paper.
- Hamilton, James D. (1994). *Time Series Analysis*, Princeton, NJ: Princeton University Press.
- Harvey, Andrew, Esther Ruiz, and Neil Shephard (1994). "Multivariate Stochastic Variance Models." *Review of Economic Studies* **61**, 247-264.
- Hentschel, Ludger (1995). "All in the Family: Nesting Symmetric and Asymmetric GARCH Models." *Journal of Financial Economics* **39**, 71-104.
- Jacquier, Eric, Nicholas G. Polson, and Peter E. Rossi (1994). "Bayesian Analysis of Stochastic Volatility Models." *Journal of Business and Economic Statistics* **12**, 1-19.

Johannes, Michael, Nicolas Polson, and Jonathan Stroud (2002). “Nonlinear Filtering of Stochastic Differential Equations with Jumps.” Columbia working paper, October.

Kim, Sanjoon, Neil Shephard, and Siddhartha Chib (1998). “Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models.” *Review of Economic Studies* **65**, 361-393.

McCullagh, Peter (1994). “Does the Moment-Generating Function Characterize a Distribution?” *The American Statistician* **48**, 208.

Nankervis, J. C. and N. E. Savin (1988). “The Exact Moments of the Least-Squares Estimator for the Autoregressive Model: Corrections and Extensions.” *Journal of Econometrics* **37**, 381-388.

Pearson, K. (1933). “On a Method of Determining whether a Sample of Size n Supposed to have been Drawn from a Parent Population Having a Known Probability Integral has Probably been Drawn at Random.” *Biometrika* **25**, 379-410.

Press, William H., Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery (1992). *Numerical Recipes in FORTRAN: The Art of Scientific Computing*, 2nd edition. Cambridge: Cambridge University Press.

Ruiz, Esther (1994). “Quasi-Maximum Likelihood Estimation of Stochastic Volatility Models,” *Journal of Econometrics* **63**, 289-306.

Stambaugh, Robert F. (1999). “Predictive Regressions.” *Journal of Financial Economics* **54**, 375-421.

Table 1: Fourier inversion approach to computing likelihood functions

Let

$$\begin{aligned} F(\Phi, \psi; y_t, x_t) &\equiv E[e^{\Phi y_{t+1} + \psi x_{t+1}} \mid y_t, x_t] \\ &= \exp[C(\Phi, \psi; y_t) + D(\Phi, \psi, y_t)x_t] \end{aligned}$$

be the (analytic) joint moment generating function of (y_{t+1}, x_{t+1}) conditional upon knowing (y_t, x_t) . Let

$$G_{t|s}(\psi) \equiv E[e^{\psi x_t} \mid \mathbf{Y}_s]$$

be the moment generating function (MGF) of x_t conditional on observing data $\mathbf{Y}_s = \{y_1, \dots, y_s\}$. Its initial value $G_{0|0}(\psi) = E[\exp(\psi x_0)]$ (the unconditional MGF of x_0) has an analytic solution. For a given parameter vector $\boldsymbol{\theta}$, subsequent MGF's $G_{t|t}$ and the likelihood function can be updated via the following recursion:

Densities	Associated moment generating functions
Conditional density $p(x_t \mid t)$	$G_{t t}(\psi)$
Joint density of datum and next period's latent variable $p(y_{t+1}, x_{t+1} \mid t)$ $= \int p(y_{t+1}, x_{t+1} \mid x_t) p(x_t \mid t) dx_t$	$\begin{aligned} F_{y,x t}(\Phi, \psi) &= E_t \left[E \left(e^{\Phi y_{t+1} + \psi x_{t+1}} \mid x_t \right) \right] \\ &= E \left[e^{C(\Phi, \psi) + D(\Phi, \psi)x_t} \mid \mathbf{Y}_t \right] \\ &= e^{C(\Phi, \psi)} G_{t t}[D(\Phi, \psi)] \end{aligned}$
Conditional density evaluation	
$p(y_{t+1} \mid t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F_{y,x t}[i\Phi, 0] e^{-i\Phi y_{t+1}} d\Phi$	
Updated conditional density of x_{t+1} $p(x_{t+1} \mid t+1) = \frac{p(y_{t+1}, x_{t+1} \mid t)}{p(y_{t+1} \mid t)}$	$\tilde{G}_{t+1 t+1}(\psi) = \frac{\frac{1}{2\pi} \int_{-\infty}^{\infty} F_{y,x t}(i\Phi, \psi) e^{-i\Phi y_{t+1}} d\Phi}{p(y_{t+1} \mid t)}$
Noncentral moments of x_{t+1} :	$E[x_{t+1}^n \mid t+1] = \left. \frac{\partial^n G_{t+1 t+1}(\psi)}{\partial \psi^n} \right _{\psi=0}$

Table 2. Estimates on simulated data.log-differenced returns: $y_{t+1} \sim N[0, V_t \Delta t]$, for $\Delta t = 1/252$ (one day).Variance innovations: $dV = (\alpha - \beta V_t)dt + \sigma\sqrt{V}dW$

1000 sample paths were simulated for data sets of 1,000 - 12,000 days; roughly 4 - 48 years of daily data. Parameters were estimated by maximum likelihood conditional on observed returns only (first columns), and conditional on also observing the latent variance realizations (last columns). All parameters are in annualized units *except* the half-life of variance shocks $HL \equiv 12 \ln 2 / \beta$, which is in months.

Average: $Avg(\hat{\theta})$; avg. bias: $Avg(\hat{\theta} - \theta)$; std. error: $\sqrt{Var(\hat{\theta}) / 1000}$; RMSE: $\sqrt{Avg[(\hat{\theta} - \theta)^2]}$. $Corr$ is the correlation $Corr(\hat{\theta}, \hat{\theta}_V)$ between the two sets of parameter estimates.

	T (days)	Estimates $\hat{\theta}$ from returns only					Estimates $\hat{\theta}_V$ if latent variance V_t were observed				
		α	β	σ	$\sqrt{\alpha/\beta}$	HL (mths)	α	β	σ	$\sqrt{\alpha/\beta}$	HL (mths)
True value:		0.044	3.25	0.187	0.116	2.56	0.044	3.25	0.187	0.116	2.56
Average:	1000	0.065	4.98	0.196	0.115	2.25	0.055	4.26	0.187	0.116	2.22
Average:	2000	0.053	4.06	0.190	0.116	2.37	0.049	3.74	0.187	0.116	2.40
Average:	4000	0.049	3.65	0.187	0.116	2.45	0.047	3.50	0.187	0.116	2.47
Average:	8000	0.047	3.48	0.185	0.116	2.48	0.046	3.39	0.187	0.116	2.50
Average:	12000	0.046	3.41	0.185	0.116	2.50	0.045	3.34	0.187	0.116	2.52
Avg bias:	1000	0.021	1.73	0.008	-0.001	-0.31	0.012	1.01	0.000	-0.001	-0.34
Avg bias:	2000	0.010	0.81	0.003	-0.001	-0.19	0.005	0.49	0.000	0.000	-0.16
Avg bias:	4000	0.005	0.40	-0.001	0.000	-0.11	0.003	0.25	0.000	0.000	-0.09
Avg bias:	8000	0.003	0.23	-0.002	0.000	-0.08	0.002	0.14	0.000	0.000	-0.05
Avg bias:	12000	0.002	0.16	-0.002	0.000	-0.06	0.001	0.09	0.000	0.000	-0.04
std.error:	1000	0.001	0.11	0.001	0.000	0.05	0.001	0.06	0.000	0.000	0.03
std.error:	2000	0.001	0.06	0.001	0.000	0.03	0.000	0.04	0.000	0.000	0.02
std.error:	4000	0.000	0.03	0.001	0.000	0.02	0.000	0.02	0.000	0.000	0.02
std.error:	8000	0.000	0.02	0.000	0.000	0.01	0.000	0.02	0.000	0.000	0.01
std.error:	12000	0.000	0.02	0.000	0.000	0.01	0.000	0.01	0.000	0.000	0.01
RMSE:	1000	0.043	3.41	0.046	0.013	1.49	0.022	1.82	0.004	0.013	0.89
RMSE:	2000	0.023	1.85	0.030	0.010	0.91	0.013	1.13	0.003	0.010	0.69
RMSE:	4000	0.013	1.07	0.020	0.007	0.69	0.008	0.74	0.002	0.007	0.50
RMSE:	8000	0.008	0.70	0.014	0.005	0.47	0.005	0.49	0.001	0.005	0.36
RMSE:	12000	0.007	0.56	0.012	0.004	0.40	0.004	0.39	0.001	0.004	0.29
Corr:	1000	0.62	0.63	0.13	0.95	0.53					
Corr:	2000	0.64	0.68	0.09	0.97	0.68					
Corr:	4000	0.62	0.71	0.15	0.97	0.71					
Corr:	8000	0.60	0.71	0.10	0.98	0.70					
Corr:	12000	0.61	0.71	0.12	0.98	0.69					

Table 3: Model estimates, and comparison with results from Chernov *et al* (2002) and Andersen *et al* (2002). New estimates in bold; other estimates and associated standard errors (in parentheses) from Chernov *et al* (2000) and Andersen *et al* (2002).

Model:

$$dS/S = (\mu_0 + \mu_1 V - \lambda_1 \bar{k}) dt + \sqrt{V} dW + (e^\gamma - 1) dN$$

$$dV = (\alpha - \beta V) dt + \sigma \sqrt{V} dW_V$$

$$\text{Corr}[dW, dW_V] = \rho, \quad \text{Prob}[dN = 1] = (\lambda_0 + \lambda_1 V) dt, \quad \gamma \sim N[\bar{\gamma}, \delta^2]$$

Chernov et al (CGGT) have an additional latent variable for the conditional mean.

Data: ABL and Bates: Daily S&P 500 returns, 1953 - 1996, 11,076 observations, prefiltered.

CGGT: Daily DJIA returns, 1953 - July 16, 1999; 11,717 observations.

All parameters in annualized units except the variance half-life $HL = 12 \ln 2 / \beta$, which is in months.

	SV			SVJ0, $\lambda_1 = 0$			SVJ1, $\lambda_1 \neq 0$		
	CGGT	ABL	Bates	CGGT	ABL	Bates	ABL	Bates	Bates
μ_0		.051 (.032)	.026 (.025)		.037 (.045)	.028 (.027)	.037 (.095)	.040 (.025)	.040 (.025)
μ_1		2.58 (2.82)	3.70 (1.98)		4.02 (3.89)	3.89 (2.19)	4.03 (5.77)	3.09 (2.16)	3.09 (2.16)
α	1.283	.051 (.010)	.093 (.011)	.044	.047 (.013)	.063 (.009)	.047 (.017)	.061 (.008)	.061 (.008)
β	137.87 (.17)	3.93 (.81)	5.94 (.81)	2.79 (.54)	3.70 (1.08)	4.38 (.70)	3.70 (1.71)	4.25 (.59)	4.25 (.59)
σ	1.024 (.030)	.197 (.018)	.315 (.018)	.207 (.02)	.184 (.019)	.244 (.016)	.184 (.019)	.237 (.015)	.237 (.015)
ρ	-1.99 (.000)	-.597 (.045)	-.579 (.031)	-.483 (.10)	-.620 (.067)	-.612 (.031)	-.620 (.086)	-.611 (.031)	-.611 (.031)
$\sqrt{\alpha/\beta}$.096	.114	.125 (.004)	.125	.113	.120 (.004)	.113	.119 (.004)	.119 (.004)
HL	0.06 (.00)	2.12 (.44)	1.40 (.19)	2.98 (.58)	2.25 (.66)	1.90 (.31)	2.25 (1.04)	1.96 (.27)	1.96 (.27)
λ_0				1.70	5.09 (.43)	.744 (.217)	5.09 (7.18)		.000 (.000)
λ_1							.70 (488.0)	93.4 (33.4)	93.4 (33.4)
$\bar{\gamma}$				-.030 (.002)		-.010 (.010)		-.002 (.006)	-.002 (.006)
δ				.008 (.001)	.012 (.001)	.052 (.009)	.012 (.001)	.039 (.008)	.039 (.008)
ln L		39,192.4 ^a	39,233.9		39,238.0 ^a	39,294.8	39,238.0 ^a	39,309.5	39,309.5

^aABL log likelihoods were evaluated using the ABL parameter estimates and data, and this paper's methodology.

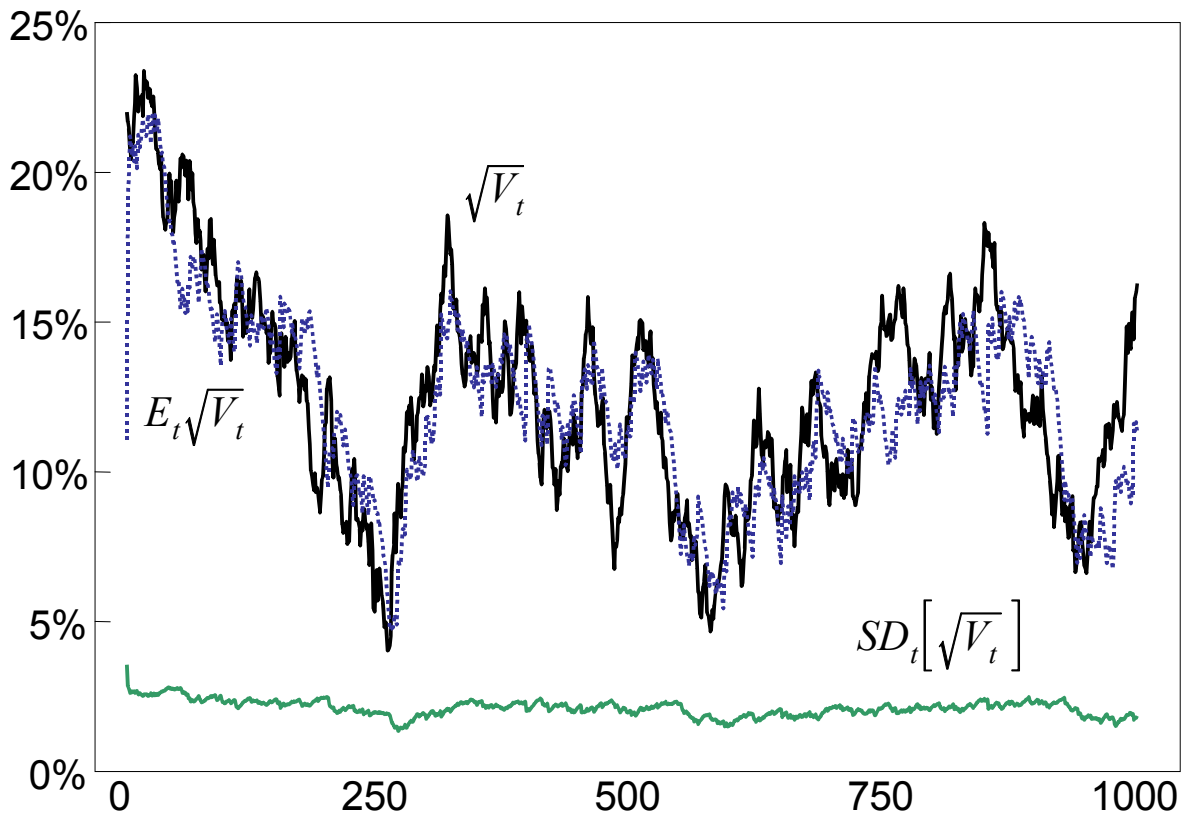


Figure 1: Actual and filtered volatility estimates from simulated data, and standard deviation of filtered estimates.

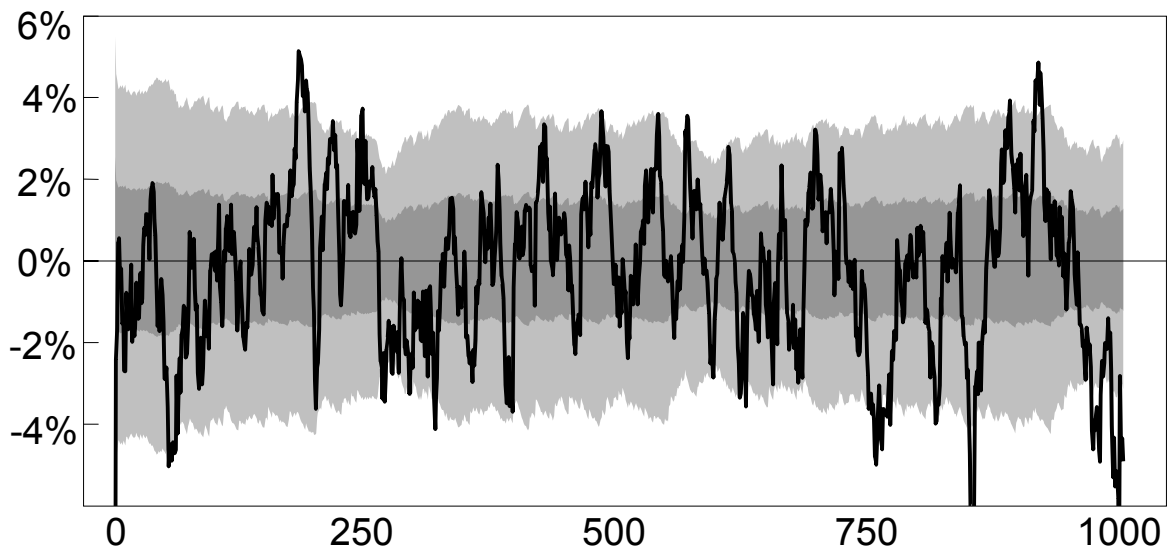


Figure 2: Volatility filtration error $E_t\sqrt{V_t} - \sqrt{V_t}$, and associated [5%, 25%, 75%, 95%] quantiles. 50% of the errors should fall in the dark gray area; 90% in the dark and light gray areas.

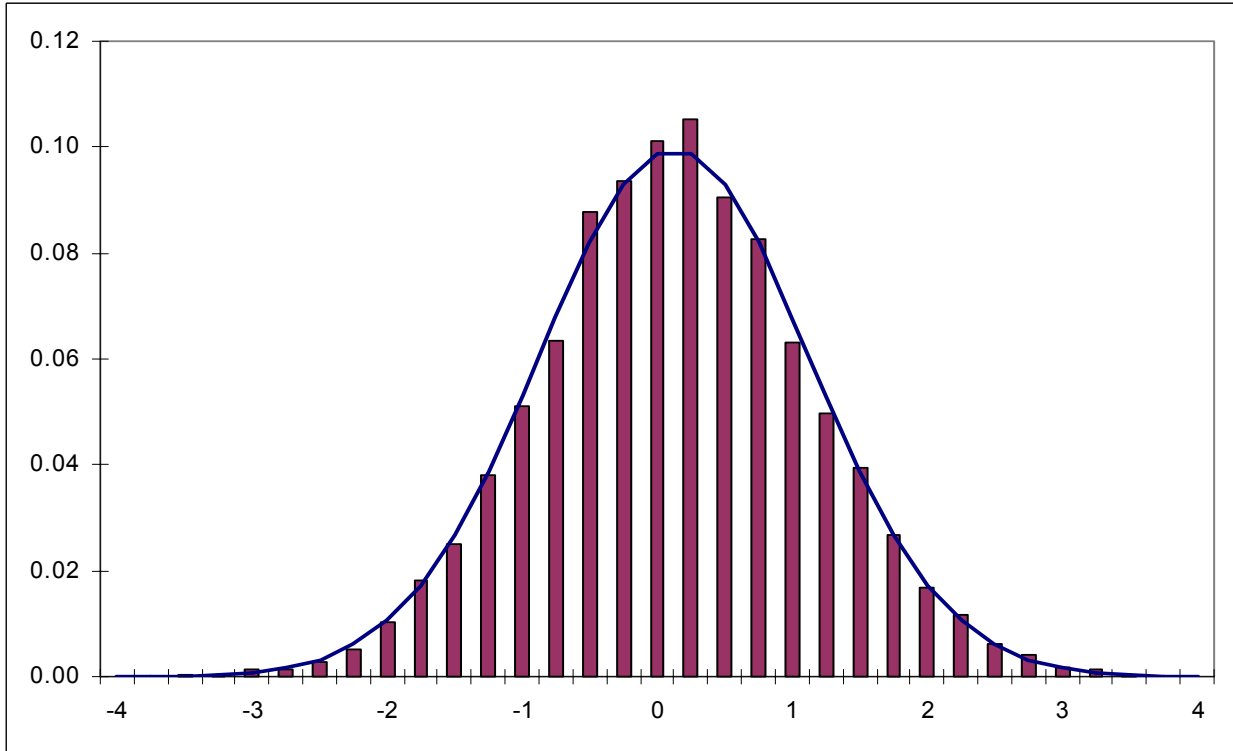


Figure 3: Distribution of normalized asset returns $N^{-1}[CDF(y_{t+1} | I_t)]$ from the stochastic volatility/jump model SVJ1, and the theoretical normal equivalents. Not shown on the graph: one outlier (out of 11,076 observations) with a z-value of -5.52.

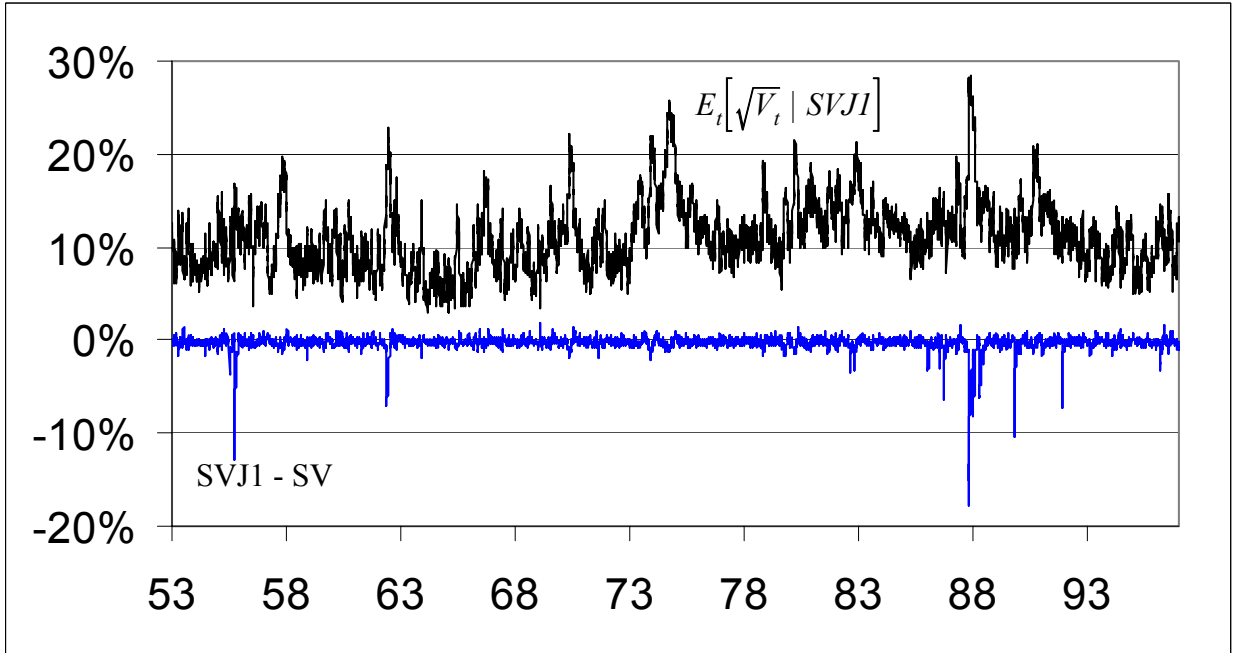


Figure 4: Filtered volatility estimate $E_t[\sqrt{V_t}]$ from the stochastic volatility/jump model SVJ1, and divergence from SV volatility estimates.

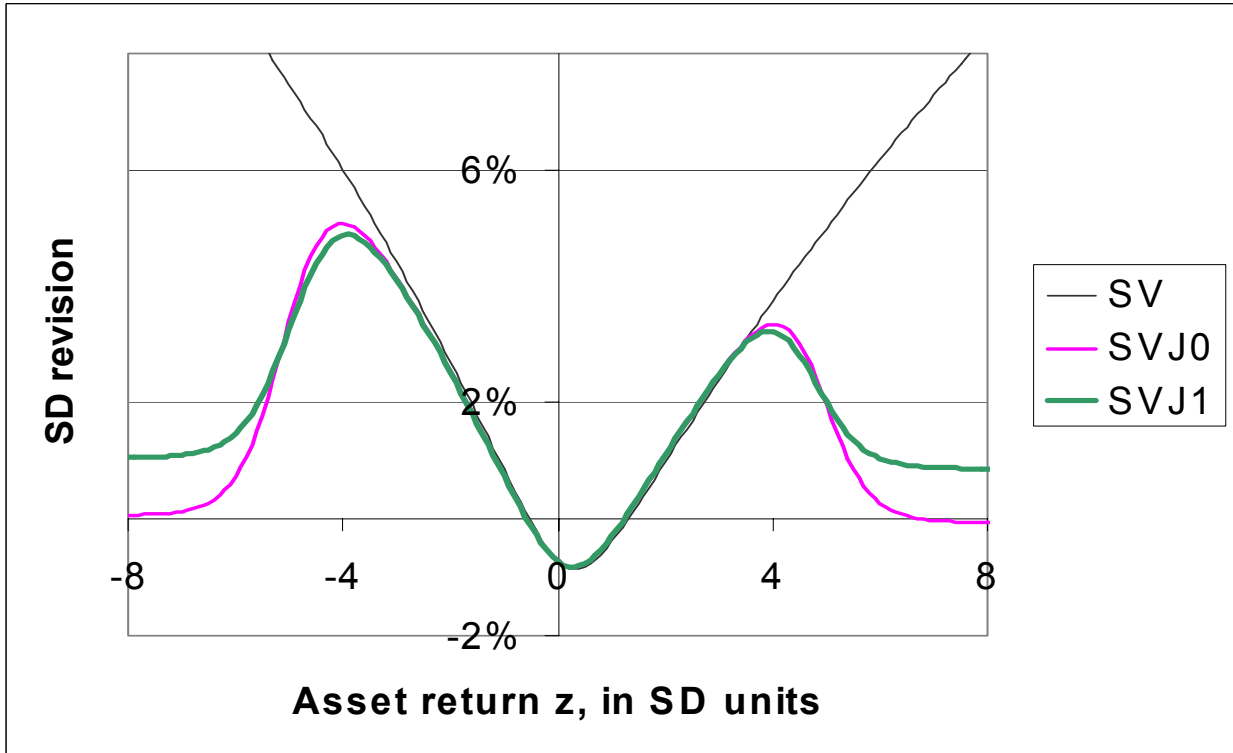


Figure 5: News impact curves for various models. The graph shows the revision in assessed conditional standard deviation, $(E_{t+1} - E_t) \sqrt{V_{t+1}}$, conditional upon observing a standardized asset return of magnitude $z = y_{t+1} / \sqrt{V_{t|t}} \Delta t$.

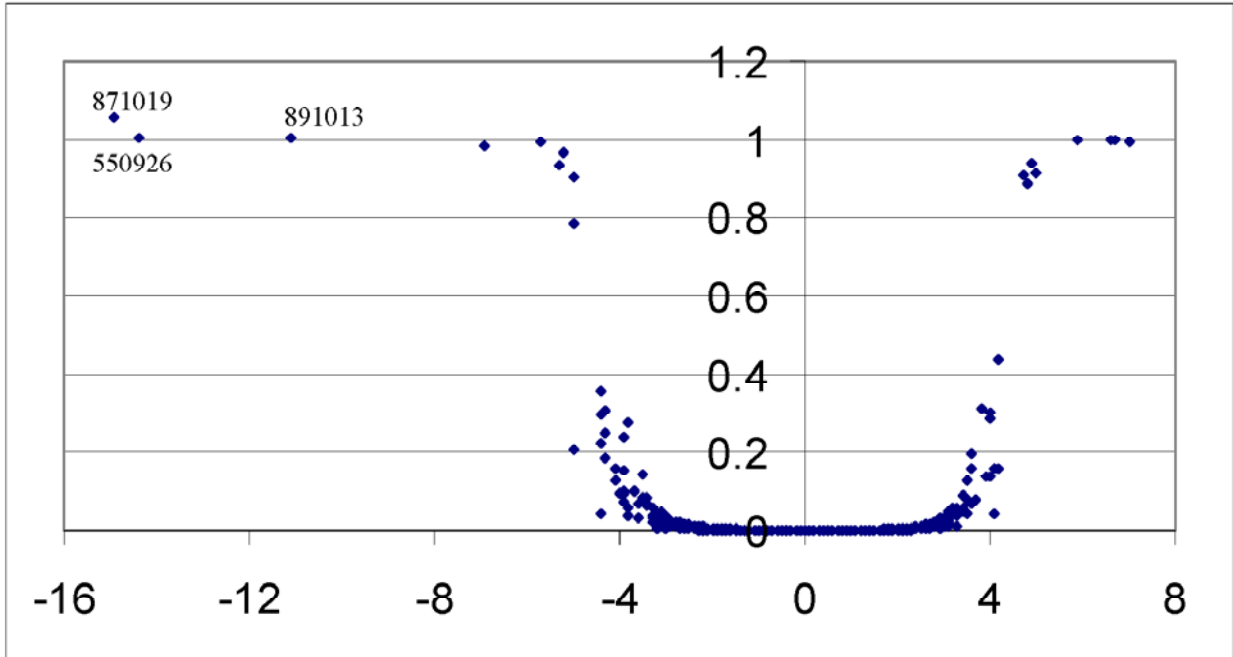


Figure 6: Estimated number of jumps, versus standardized returns $y_{t+1} / \sqrt{\hat{V}_{t|t} \Delta t}$, SVJ0 model. The values are approximately the probability for each day that a jump occurred.