Maximum Likelihood Estimation of Latent Affine Processes

David S. Bates

University of Iowa and the
National Bureau of Economic Research

October 30, 2002

Abstract

This article develops a direct 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 jump risk.
“The Lion in Affrik and the Bear in Sarmatia are fierce, ... but Translated into a Contrary Heaven, are of less strength and courage.”

Dorothy Dunnett, *The Ringed Castle*

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. At present, the state of the art is simulation methods such as 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 to 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.

This article derives an alternate recursive maximum likelihood methodology applicable whenever the 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 an equivalent of Bayes’ rule for updating the characteristic function of the latent variable conditional upon observed data. Given this and the affine structure, 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. 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 in the state variable can easily be accommodated, whereas simulation methods such as EMM may 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 variance from options prices.

A further advantage of the approach relative to EMM 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. Smoothed estimates based on the full data set can also be generated recursively. 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 affine stochastic volatility process and simulated data. Section 3 provides estimates of some affine 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 \( y_t \) denote an \((L \times 1)\) vector of variables observed at date \( t \). Let \( x_t \) represent an \((M \times 1)\) vector of latent state variables affecting the dynamics of \( y_t \). The following assumptions are made:

1) \( y_t \) is assumed to be Markov;
2) the latent variables \( x_t \) are assumed stationary; and
3) the characteristic function of \( z_{t+i} \) conditional upon \( z_t \) is exponentially affine in the latent variables \( x_t \):

\[
F_{y,x}(i\Phi, i\psi; z_t) = E\left[e^{i\Phi y_{t+i} + i\psi x_{t+i} | z_t}\right] = \exp[C(\tau; i\Phi, i\psi; y_t) + D(\tau; i\Phi, i\psi)'x_t].
\]

(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 \( p(z_{t+i} | x_t) \) 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 state vector \( x_t \), 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:
\[ G_y(\Phi) = E[e^{i\Phi \beta}] = \int e^{i\Phi \beta} p(y) \, dy \]  
\[ p(y) = \frac{1}{2\pi} \int G(i\Phi) e^{-i\Phi y} \, d\Phi. \]  

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:

\[ p(x, y) = \frac{1}{(2\pi)^2} \int \int F_{x,y}(i\psi, i\Phi) e^{-i\psi x - i\Phi y} \, d\psi \, d\Phi. \]  

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

**Proposition 1.** Let

\[ F_{x,y}(i\psi, i\Phi) = E[e^{i\psi \tilde{x} + i\Phi \tilde{\beta}}] = \int \int e^{i\psi x + i\Phi x} p(x, y) \, dx \, dy \]  

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

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

where

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

is the marginal density of \( y \).
Proof: By Bayes’ law, the conditional characteristic function \( G_{x \mid y} \) can be written as

\[
G_{x \mid y}(i \psi; y) = \int e^{i \psi x} p(x \mid y) \, dx = \frac{1}{p(y)} \int e^{i \psi x} p(x, y) \, dx. \tag{8}
\]

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

\[
F_{x, y}(i \psi; i \Phi) = \int e^{i \Phi y} \left[ G_{x \mid y}(i \psi, y) p(y) \right] \, dy = \int \int e^{i \Phi y + i \psi x} p(x, y) \, dx \, dy. \tag{9}
\]

Consequently, \( G_{x \mid y}(i \psi, y) p(y) \) is the inverse \((\Phi, y)\) Fourier transform of \( F_{x, y}(i \psi, i \Phi) \), yielding (6) above.

Define \( I_t = \{y_0, \ldots, y_t\} \) as the data observed up through date \( t \). Define \( G_{t \mid s}(i \psi) = \mathbb{E}[e^{i \psi x_t} \mid I_s] \) as the characteristic function (CF) of the latent variables \( x_t \) conditional upon observing \( I_s \). Given Proposition 1 and the affine structure, the filtered CF \( G_{t \mid t}(i \psi) \) can be recursively updated as follows:

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

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

**Step 1:** Given \( G_{t \mid t} \), the joint characteristic function of next’s period’s \( z_{t + \tau} = \{y_{t + \tau}, x_{t + \tau}\} \) 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:

\[
F_{y, x \mid t}(i \Phi, i \psi) = \mathbb{E}_t \left[ \mathbb{E}_t \left[ e^{i \Phi y_{t + \tau} + i \psi x_{t + \tau}} \mid x_t \right] \right] = \mathbb{E}_t \left[ e^{C(\tau; \Phi, i \psi) + D(\tau; i \Phi, i \psi) x_t} \mid I_t \right] = e^{C(\tau; i \Phi, i \psi)} G_{t \mid t}[D(\tau; i \Phi, i \psi)]. \tag{11}
\]
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:

$$p(y_{t+1} \mid t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F_{y_t,x_{\tau+t}}[i\Phi, 0] e^{-i\Phi y_{t+1}} d\Phi.$$  \hspace{1cm} (12)

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

$$G_{t+\tau \mid t, \tau}(i\psi) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F_{y_t,x_{\tau+t}}[i\Phi, i\psi] e^{-i\Phi y_{t+\tau}} d\Phi$$  \hspace{1cm} p(y_{t+\tau} \mid t)$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} G_{t \mid t}[D(\tau; i\Phi, i\psi)] e^{C(\tau; i\Phi, i\psi) - i\Phi y_{t+\tau}} d\Phi$$  \hspace{1cm} p(y_{t+\tau} \mid t)$$

Step 4: Repeat steps 1-3 for subsequent values of $t$.

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.

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

$$\hat{x}_{t+1 \mid t, 1} = G_{t+1 \mid t, 1}'(0)$$

$$P_{t+1 \mid t, 1} = Var_{t+1}(x_{t+1})$$

$$= G_{t+1 \mid t, 1}''(0) - \hat{x}_{t+1 \mid t, 1}^2,$$  \hspace{1cm} (14)

where derivatives with respect to $\psi$ are taken inside the integral in (13).

Smoothed estimates can also be computed, using the joint distribution of $(x_t, x_{t+\tau}, y_{t+\tau})$ conditional upon information $I_t$:
\[
F_{y_t, x_t, x_{t-\tau}}(\Phi, \psi_0, \psi_1) = E \left[ E \left[ e^{\Phi y_t x_t + \psi_0 y_t x_{t-\tau} | x_t} \right] \right] \\
= E \left[ e^{C(\tau; \Phi, \psi_1) + [\psi_0 + D(\tau; \Phi, \psi_1)] x_t | I_t} \right] \\
= e^{C(\tau; \Phi, \psi_1) G_{t|t}[\psi_0 + D(\tau; \Phi, \psi_1)]}.
\] (15)

Smoothed characteristic functions \(G_{t|T}(\psi) = E[\exp(i\psi x_t) | I_T]\) are generated by beginning at the penultimate observation \(t = T - \tau\), updating the MGF of \(x_t\) conditional upon \((x_{t-\tau}, y_{t-\tau}, I_T)\) via a double invocation of Proposition 1, and proceeding recursively backwards over time. Smoothed estimates of the latent variable are then computed analogously to (14): \(\hat{x}_{t|T} = G_{t|r} \cdot (0)\).

### I.B Implementation

The recursion in (10) - (13) indicates that for a given conditional characteristic function \(G_{t|t}(i\psi)\) and an observed datum \(y_{t-\tau}\), it is possible to compute an updated CF \(G_{t+1|t+\tau}(i\psi)\) that fully summarizes all that is known about the latent variable \(x_{t-\tau}\). To implement the recursion, it is necessary to 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 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 \(\kappa_4\). The associated density and characteristic functions are

\[
p(x) = 1 + \frac{\kappa_4}{24} (x^4 - 6x^2 + 3) n(x) \]
\[
G(i\psi) = e^{-\psi^2} \left( 1 + \frac{\kappa_4}{24} \psi^4 \right)
\] (16)

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 Table 1 is just Bayesian updating, any legitimate approximate prior characteristic function \( \hat{G}_{t_{\text{p}}}^*(i\psi) \) from a known distribution will generate a legitimate posterior characteristic function \( \hat{G}_{t_{\text{t}}}^*(i\psi) \). 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 deemed necessary.

2. A Monte Carlo examination of approximation error

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 is simulated:

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

with asset returns and variance shocks assumed independent. This process is the affine equivalent of the benchmark stochastic volatility model studied by Harvey, Ruiz, and Shephard (1994), Jacquier, Polson and Rossi (1994), and Shephard and Kim (1993). Rather than the conditionally Gaussian AR(1) process for log volatility of those papers, however, the square-root process for volatility implies a noncentral chi-squared transition density in discrete time:

\[
\begin{align*}
\frac{2V_{t+1}}{\kappa} | \quad & V_t \sim \chi^2 \left( \frac{4\alpha}{\sigma^2}, \frac{2V_t e^{-\beta \Delta t}}{\kappa} \right) \\
\text{for} \quad & \kappa = \frac{\sigma^2}{2\beta} \left[ 1 - e^{-\beta \Delta t} \right].
\end{align*}
\]

Given the assumed conditional independence of volatility increments and asset returns, the joint conditional characteristic function in (1) above takes the separable form.
\[
F_{y,\nu}(i\Phi, i\psi \mid V_t) = E \left[ e^{i\Phi y_{t+1} + i\psi V_{t+1}} \mid V_t \right] = \exp \left[ -\frac{2\alpha}{\sigma^2} \ln(1 - i\kappa \psi) + \left( -i\frac{\Phi^2 \Delta t}{2} + \frac{e^{-\beta \Delta t} i\psi}{1 - i\kappa \psi} \right) V_t \right].
\]

The unconditional density of \( V_0 \) has a gamma distribution, with associated characteristic function
\[
G_{0\mid 0}(\psi) = (1 - i\kappa_0 \psi)^{-\nu_0}
\]
for \( \nu_0 = 2\alpha/\sigma^2 \) and \( \kappa_0 = \sigma^2/2\beta \). The unconditional mean and variance of \( V_0 \) are \( \kappa_0 \nu_0 = \alpha/\beta \) and \( \kappa_0^2 \nu_0 = (\alpha/\beta)(\sigma^2/2\beta) \), respectively.

The simplest approximating function in subsequent periods to assume a gamma distribution:
\[
\hat{G}_t(i\psi) = (1 - i\kappa_t \psi)^{-\nu_t}.
\]

This approximation is exact for \( t = 0 \), and imposes appropriate nonnegativity constraints for \( t > 0 \). The density is summarized by the parameter pair \((\kappa_t, \nu_t)\). Furthermore, updated estimates \((\kappa_{t+1}, \nu_{t+1})\) can be generated from equations (13) and (14):
\[
\kappa_{t+1} \nu_{t+1} = \hat{\nu}_{t+1 \mid t+1} = G_{t+1 \mid t+1}'(0)
\]
\[
\kappa_{t+1}^2 \nu_{t+1} = Var_{t+1}(V_{t+1}) = G_{t+1 \mid t+1}''(0) - \hat{\nu}_{t+1 \mid t+1}^2.
\]

The resulting updated approximation \( \hat{G}_{t+1 \mid t+1}(\psi) = (1 - \kappa_{t+1})^{-\nu_{t+1}} \) matches the first and second moments of the true \( G_{t+1 \mid t+1}(\psi) \) computed from (13) for different values of \( \psi \), and is therefore comparable to a second-order Taylor approximation.

It may be useful to compare this approach with 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 conditionally Gaussian distribution of log volatility. The HRS approach suffers from an “inlier” problem resulting from the fact that log absolute returns are
not approximately Gaussian, given occasional near-zero returns. Here, the information inferred from observed data in (13) is exact. The approximation error enters in the specification of the prior density for latent variance, which affects the relative weights of the data and the prior. Furthermore, the updating of the moments of $V_{t+1}$ at each step is optimal from a mean-variance criterion. Consequently, the propositions in Ruiz (1994) can probably be used here as well to show that the parameter estimates from using approximating characteristic functions are consistent.

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. 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 was set to $\sigma = \sqrt{4\alpha/\beta} = .187$, instead of the ABL estimate of .185. This produces an integer number (five) of degrees of freedom for the noncentral $\chi^2$ density in (18), simplifying the Monte Carlo generation of shocks from that distribution. 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 generated for various values of $T$. Normally distributed log-differenced returns were generated given the variances, and the parameters $\hat{\theta} = \{\hat{\alpha}, \hat{\beta}, \hat{\sigma}\}$ were estimated by the maximum likelihood methodology described above, initiated at the true parameter values. For comparison, the parameters $\hat{\theta}_f$ were also estimated by direct maximum likelihood conditional upon observing the $\{V_t\}_{t=0}^{T-1}$ sample path, using the noncentral $\chi^2$ transition densities and initial gamma distribution.

Table 2 summarizes the results. The estimation methodology is consistent, with the estimated bias (average bias rows) decreasing with longer data samples. The fastest convergence is for the average variance $\alpha/\beta$ and the volatility of variance parameter $\sigma$. The slowest is for the parameter $\beta$ 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. Here, of course, the values of that series must be inferred from noisy returns data, which almost doubles the magnitude of the bias. The \( (\hat{\beta}, \hat{\beta}_\nu) \) estimates have a substantial correlation of 60-70%. Other parameter estimates are also substantially correlated with the estimates conditional on directly observing the \( V(t) \) data, except for the \( \hat{\sigma} \) estimates.

Figures 1 and 2 illustrate the accuracy of the \( \hat{V}_{t|t} \) filtration conditional upon using the true parameters, for the first 1000 observations (four years) of a generated sample of 12,000 observations. The estimate begins at the unconditional mean of \( (0.116)^2 \) at time 0, but converges rapidly towards the true value as returns are observed. Changes in the filtered estimate \( \hat{V}_{t|t} \) perforce lag behind changes in the true variance \( V_t \), since the filtered estimate is inferred from past squared returns. The divergence is typically less than .01, but is occasionally larger. To put this error in perspective: a variance estimate (towards the end of the sample) of .01 when the true variance is .02 implies corresponding annualized volatilities of 10% and 14.1%, respectively.

A key issue is whether the approximating gamma density for the distribution of the filtered \( \hat{V}_{t|t} \) estimate is reasonably accurate. On simulated data, this can be directly assessed by examining the frequency with which \( \hat{V}_{t|t} \) realizations fall within the quantiles of the gamma distribution of \( \hat{V}_{t|t} \). The realized frequencies over a run of 12,000 observations indicate the gamma approximation is indeed accurate:

<table>
<thead>
<tr>
<th>Quantile p:</th>
<th>.05</th>
<th>.10</th>
<th>.25</th>
<th>.50</th>
<th>.75</th>
<th>.90</th>
<th>.95</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prob[V(t) &lt; V_{p</td>
<td>t}(t)</td>
<td>\mathcal{I}_t]|</td>
<td>.043</td>
<td>.090</td>
<td>.243</td>
<td>.492</td>
<td>.743</td>
</tr>
</tbody>
</table>

where \( V_{p|t}(t) \) is the 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,

---

1See Nankervis and Savin, 1988 and Stambaugh 1999 for a discussion of these biases.
I am indebted to Luca Benzoni for providing the data. The ABL estimates are from Table IV, converted to an annualized basis. Based on ABL’s annualized parameters in footnote 10 on p. 1261, it appears there are decimal-point errors in the daily parameter estimates of their Tables III and IV. On a daily decimal basis, the SV parameter estimates in Table IV should be $\mu_0 = 0.0300 \times 10^{-2}$, $\alpha = 0.0069 \times 10^{-4}$, $\beta = 0.0129$, $\sigma = 0.0734 \times 10^{-2}$.

This article uses 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. I will not replicate 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 the joint likelihood function.

The first three columns of Table 3 present estimates of the stochastic volatility model without jumps from Chernov et al, Andersen et al, and the methodology of this paper. As discussed in CGGT,
estimating the parsimonious stochastic volatility model without jumps creates conflicting demands for the volatility mean reversion parameter $\beta$ and the volatility of volatility parameter $\sigma$. 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 the latter.

My estimates are affected by both phenomena, but matching the volatility persistence clearly dominates. While constraining $\sigma$ 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.

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 $9 \times 10^{-16}$. 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 -0.9% and standard deviation 5.1%. As outliers are now primarily explained by the jump component, the parameters

and $\rho = -.5877$. I have also modified the decimal point of their $c$ estimate ($\mu_1$ in my notation) to make the order of magnitude comparable to mine.
governing volatility dynamics parameters are modified: \( \sigma \) 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 \( 10^{-6} \). The time-invariant jump component \( \lambda_0 \) ceases to be statistically significant when \( \lambda_1 \) 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

\[
z_{t+1} = N^{-1}[CDF(y_{t+1} | I_t, \hat{\theta})]
\]

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 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 4.1%, and has a \( z \)-score of -5.44. 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
\begin{align}
    \text{Prob}[dN = 1] &= 130.61 \, V_t, \\
    \ln (1 + \tilde{k}) &\sim \begin{cases} 
        \mathcal{N}[.002, (.029)^2] & \text{with prob .985} \\
        \mathcal{N}[-.212, (.061)^2] & \text{with prob .015}
    \end{cases}
\end{align}

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 an average arrival rate of 1.85 per year, and an extremely infrequent crash corresponding to the 1987 outlier. Log likelihood rises from 39,306.84 to 39,313.94 -- 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 -3.59. The increase in log likelihood has a marginal significance level of .003 under a likelihood ratio test, given 3 additional parameters.

3.B. Filtration estimates

Figure 4 illustrates the filtered estimates of latent variance from the SV and SVJ1 models, respectively. The variance estimates are generally comparable. However, the outliers in 1955 and 1987 have more of an impact on variance assessments under the SV model than under the jump models.

Figure 5 illustrates the typical impact of asset returns on volatility assessments. For comparability with the “news impact curves” used by Hentschel (1995) in assessing GARCH models, the revisions in expectations are specified with regard to conditional standard deviations, rather than variances, using the Taylor approximation

\[ E[\sqrt{V}] \approx \sqrt{E[V]} \left[ 1 - \frac{1}{8} \frac{\text{Var}[V]}{E[V]^2} \right] \]

The estimates were calibrated from a median volatility day (October 14, 1960) with annualized volatility of 11.0%, and filtered gamma distribution parameters \((\kappa_t, v_t) = (.00229, 5.89)\).

All news impact curves are tilted, with negative returns having a larger impact on volatility assessments than positive returns. 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. All models process the information in small asset returns similarly. Under the SVJ0 model, however, 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.

Even the SV filtration is not necessarily compatible with GARCH filtrations. In the stochastic volatility model, the precision $P_{t|t}$ of the variance prior varies stochastically over time, in a fashion that is correlated with but not fully summarized by the variance estimate $\hat{V}_{t|t}$. The relative weighting of the information in asset returns and the variance prior therefore changes over time in a fashion that cannot be captured by GARCH specifications. It is an open question as to whether this variable weighting improves 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. The joint conditional distribution of log-differenced asset prices and the number of jumps $\tilde{n} \equiv N(t+1) - N(t)$ has an affine specification. Consequently, the characteristic function $G_n(i\xi \mid y_{t-1}^\frac{1}{2}, I_t) = E[e^{i\xi\tilde{n}^\frac{1}{2}} \mid y_{t-1}^\frac{1}{2}, I_t]$ can also be evaluated by Proposition 1.

While it is possible to evaluate the daily probability that $n$ jumps occurred by Fourier inversion, it is simpler to compute the conditional mean $E[n_{t-1} \mid I_{t-1}, I_t] = G_n'(0 \mid y_{t-1}^\frac{1}{2}, I_t)$. At the daily horizon, $n$ is essentially binomial, and the conditional mean 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 quite 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 would appear, however, that the SNP/EMM methodology may be quite sensitive to infrequent outliers. Chernov et al (2002), find very different estimates from Andersen et al for the stochastic volatility model, despite using similar methodologies and similar data sets. My 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

\[
F_{y_t, v_t, \tau}(i\Phi, i\psi, i\xi) = E\left[ \exp\left( i\Phi \tilde{y}_{t+1} + i\psi \tilde{V}_{t+1} + i\xi \int_{\tau-t}^{\tau+1} \tilde{V}_t \, dt \right) \mid V_t \right]
\]

\[
= \exp[C(i\Phi, i\psi, i\xi) + D(i\Phi, i\psi, i\xi) V_t]
\]

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 $(y_{t+1}, \overline{V}_{t+1})$ 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 transition densities

By Itô’s lemma, the stochastic volatility/jump diffusion in equation (20) implies a log asset price evolution of the form

\[ ds = [\mu_0 + (\mu_1 - \frac{1}{2}) V - (\lambda_0 + \lambda_1 V) \bar{k}] dt + \sqrt{V} dW + \gamma dN \]  

(A.1)

\[ dV = (\alpha - \beta V) dt + \sigma \sqrt{V} dW_v \]  

(A.2)

where \( s \) is the log of the asset price, \( \text{Cov}(dW, dW_v) = \rho \ dt \), \( \gamma \) is normally distributed \( \mathcal{N}[\gamma, \delta^2] \), and \( \bar{k} = e^{\gamma + \frac{\gamma^2}{2}} - 1 \). The joint moment generating function underlying the transition density \( p(s_T, V_T \mid s_0, V_0) \) is

\[ F(\Phi, \psi; T) = E \left[ e^{\Phi s_T + \psi V_T} \mid s_0, V_0 \right] , \]  

(A.3)

which solves the backward Kolmogorov partial differential equation

\[ F_T = [\mu_0 + (\mu_1 - \frac{1}{2}) V - (\lambda_0 + \lambda_1 V) \bar{k}] F_s + (\alpha - \beta V) F_V + \frac{1}{2} V \left( F_{ss} + 2 \rho \sigma_v F_{sv} + \sigma_v^2 F_{VV} \right) + \lambda E [F(s + \gamma, V) - F] \]  

(A.4)

subject to the boundary condition

\[ F(\Phi, \psi; 0) = e^{\Phi s_0 + \psi v_0} . \]  

(A.5)

The solution is

\[ F(\Phi, \psi; T) = \exp[\Phi s_T + C(T; \Phi, \psi) + D(T; \Phi, \psi) V_0 ] \]  

(A.6)

where

\[ E(\Phi) = \exp[\gamma \Phi + \frac{1}{2} \delta^2 \Phi^2] - (1 + \bar{k}) \Phi \]  

(A.7)

\[ C(T; \Phi, \psi) = [\mu - \lambda \bar{k}] \Phi T - \frac{\alpha T}{\sigma^2} (\rho \sigma \Phi - \beta - \gamma) + \lambda_0 T E(\Phi) \]

\[ - \frac{2 \alpha}{\sigma^2} \ln \left[ 1 + \frac{1}{2} (\rho \sigma_v \Phi - \beta - \gamma) \frac{1 - e^{-\gamma T}}{\gamma} \right] - \frac{2 \alpha}{\sigma^2} \ln \left[ 1 - \kappa(\Phi) \psi \right] \]  

(A.8)
\[ D(T; \Phi, \psi) = \frac{-2(\mu_1 - \frac{1}{2})\Phi - \Phi^2}{\rho\sigma\Phi - \beta + \gamma \frac{1 + e^{\gamma T}}{1 - e^{\gamma T}}} + \frac{\Lambda(\Phi)\psi}{1 - \kappa(\Phi)\psi} , \quad (A.9) \]

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

\[ \Lambda(\Phi) = \frac{\left( \frac{e^{\gamma T} + 1}{e^{\gamma T} - 1} \right)^2 - 1}{\left( \frac{e^{\gamma T} + 1}{e^{\gamma T} - 1} + \frac{\beta - \rho\sigma\Phi}{\gamma} \right)^2} \quad (A.11) \]

\[ \kappa(\Phi) = \frac{\sigma^2}{\gamma \frac{e^{\gamma T} + 1}{e^{\gamma T} - 1} + \beta - \rho\sigma\Phi} \quad (A.12) \]

The moment generating functions underlying the marginal transition densities of \( s_T \) and \( V_T \) are of course given by \( F(\Phi, 0, T) \) and \( F(0, \psi, T) \), respectively. Given (A.6), the joint moment generating function of log-differenced asset prices \( y = \Delta s = \ln(S_t \Delta_t / S_t) \) conditional upon observing \( V_t \) is

\[ F_{y,v}(\Phi, \psi; \Delta t) = \exp \left[ C(\Delta t; \Phi, \psi) + D(\Delta t; \Phi, \psi) V_t \right] . \quad (A.13) \]

**A.2 Numerical integration, and outliers**

Given the gamma MGF approximation \( G_{t+1}(\psi) = (1 - \kappa_t\psi)^{-\psi_t} \) for the latent variance \( V_t \), the evaluation of the conditional density of the datum \( y_{t+1} \) involves integrals of the form

\[ p(y_{t+1} \mid I_t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{C(\psi, 0) - \psi \ln(1 - \kappa_t D(\psi, 0)] - \psi y_{t+1}} d\Phi \quad (A.14) \]

An upper limit \( \Phi_{\max} \) was computed for which truncation error would be less than \( \varepsilon = 10^{-10} \), and

\[ ^1 \text{For } \Phi = 0, \Lambda(0) = e^{-\beta T} \text{ and } \kappa(0) = \frac{1}{2}\sigma^2 (1 - e^{-\beta T})/\beta . \]
the integral was computed numerically using IMSL’s adaptive DQDAG integration routine over 
\((0, \Phi_{\text{max}})\). The updating of mean and variance in (14) involved similar integrals.

For extreme outliers, \(p(y)\) takes on near-zero values that are difficult for the numerical integration 
routine to evaluate. For such values, a scaled density function can be evaluated. The Fourier 
transform of the scaled density function \(e^{ay'}p(y^*)\) of the standardized return

\[
y^* = y_{t+1}/\sqrt{V_{t+1}}\Delta t
\]

is

\[
\int [e^{ay'}p(y^*)] e^{\Phi y'} dy' = E\left[ e^{(a+\Phi)y} \right] = E\left[ \exp \left( \frac{(a+\Phi)y}{\sqrt{V\Delta t}} \right) \right] = \exp \left[ C \left( \frac{a + \Phi}{\sqrt{V\Delta t}}, 0 \right) \right] G_{t+1} \left[ D \left( \frac{a + \Phi}{\sqrt{V\Delta t}}, 0 \right) \right]
\]

so the scaled density can be computed by Fourier inversion:

\[
e^{ay'}p(y^*) = \frac{1}{2\pi} \int \exp \left[ C \left( \frac{a + i\Phi}{\sqrt{V\Delta t}}, 0 \right) \right] G_{t+1} \left[ D \left( \frac{a + i\Phi}{\sqrt{V\Delta t}}, 0 \right) \right] e^{-i\Phi y'} d\Phi.
\]

The density of observed data is then

\[
p(y) = e^{-ay'} \left[ e^{ay'}p(y^*) \right] / \sqrt{V_{t+1}}\Delta t.
\]

The transformation is the equivalent of importance sampling for low-probability events in Monte Carlo integration. Similar transformations can be used in conjunction with the updating of the moments of latent variance from the joint transformation of \(e^{ay'}p(y^*)\) and \(V_{t+1}\).

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:** \(a = 2\) for the 1987 crash, 0 otherwise;
- **SVJ2:** \(a = 0\) always.
References


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.


Table 1: Fourier inversion approach to computing likelihood functions

Let

\[ F(\Phi, \psi; x_t) = E[e^{\Phi y_{t+1} + \psi x_{t+1}} | x_t] = \exp[C(\Phi, \psi) + D(\Phi, \psi) x_t] \]

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

\[ G_{t|s}(\psi) = E\left[ e^{\psi x_t} | I_s \right] \]

be the moment generating function (MGF) of \(x_t\) conditional on observing data \(I_s = \{y_1, \ldots, y_s\}\). 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 \(\theta\), subsequent MGF’s \(G_{t|t}\) and the likelihood function can be updated via the following recursion:

<table>
<thead>
<tr>
<th>Densities</th>
<th>Associated moment generating functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conditional density (p(x_t</td>
<td>t))</td>
</tr>
<tr>
<td>Joint likelihood of data and latent variable (p(y_{t+1}, x_{t+1}</td>
<td>t))</td>
</tr>
<tr>
<td>Evaluation of likelihood function</td>
<td>(p(y_{t+1}</td>
</tr>
<tr>
<td>Updated conditional density of (x_{t+1})</td>
<td>(G_{t+1</td>
</tr>
<tr>
<td>Noncentral moments of (x_{t+1}): (E[x_{t+1}^n</td>
<td>t+1] = \frac{\partial^n G_{t+1</td>
</tr>
</tbody>
</table>
Table 2. Estimates on simulated data.

**log-differenced returns:** \( y_{t+1} - N[0, V_t \Delta t] \), 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 based on observed returns only (first columns), and based on also observing the latent variance realizations (last columns). All parameters are in annualized units **except** the half-life of variance shocks \( HL = 12 \ln 2/\beta \), which is in months.

Average: \( \text{Avg}(\hat{\theta}) \); avg. bias: \( \text{Avg}(\hat{\theta} - \theta) \); std. error: \( \sqrt{\text{Var}(\hat{\theta})}/1000 \); **RMSE:** \( \sqrt{\text{Avg}((\hat{\theta} - \theta)^2)} \).

Corr is the correlation \( \text{Corr}(\hat{\theta}, \hat{\theta}_V) \) between the two sets of parameter estimates.

<table>
<thead>
<tr>
<th>T (days)</th>
<th>Estimates ( \hat{\theta} ) from returns only</th>
<th>Estimates ( \hat{\theta}_V ) if latent variance ( V(t) ) were observed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \alpha )</td>
<td>( \beta )</td>
</tr>
<tr>
<td>True value:</td>
<td>0.044</td>
<td>3.25</td>
</tr>
<tr>
<td>Average: 1000</td>
<td>0.065</td>
<td>4.98</td>
</tr>
<tr>
<td>Average: 2000</td>
<td>0.053</td>
<td>4.06</td>
</tr>
<tr>
<td>Average: 4000</td>
<td>0.049</td>
<td>3.65</td>
</tr>
<tr>
<td>Average: 8000</td>
<td>0.047</td>
<td>3.48</td>
</tr>
<tr>
<td>Average: 12000</td>
<td>0.046</td>
<td>3.41</td>
</tr>
<tr>
<td>Avg bias: 1000</td>
<td>0.021</td>
<td>1.73</td>
</tr>
<tr>
<td>Avg bias: 2000</td>
<td>0.010</td>
<td>0.81</td>
</tr>
<tr>
<td>Avg bias: 4000</td>
<td>0.005</td>
<td>0.40</td>
</tr>
<tr>
<td>Avg bias: 8000</td>
<td>0.003</td>
<td>0.23</td>
</tr>
<tr>
<td>Avg bias: 12000</td>
<td>0.002</td>
<td>0.16</td>
</tr>
<tr>
<td>std.error: 1000</td>
<td>0.001</td>
<td>0.11</td>
</tr>
<tr>
<td>std.error: 2000</td>
<td>0.001</td>
<td>0.06</td>
</tr>
<tr>
<td>std.error: 4000</td>
<td>0.000</td>
<td>0.03</td>
</tr>
<tr>
<td>std.error: 8000</td>
<td>0.000</td>
<td>0.02</td>
</tr>
<tr>
<td>std.error: 12000</td>
<td>0.000</td>
<td>0.02</td>
</tr>
<tr>
<td>RMSE: 1000</td>
<td>0.043</td>
<td>3.41</td>
</tr>
<tr>
<td>RMSE: 2000</td>
<td>0.023</td>
<td>1.85</td>
</tr>
<tr>
<td>RMSE: 4000</td>
<td>0.013</td>
<td>1.07</td>
</tr>
<tr>
<td>RMSE: 8000</td>
<td>0.008</td>
<td>0.70</td>
</tr>
<tr>
<td>RMSE: 12000</td>
<td>0.007</td>
<td>0.56</td>
</tr>
<tr>
<td>Corr: 1000</td>
<td>0.62</td>
<td>0.63</td>
</tr>
<tr>
<td>Corr: 2000</td>
<td>0.64</td>
<td>0.68</td>
</tr>
<tr>
<td>Corr: 4000</td>
<td>0.62</td>
<td>0.71</td>
</tr>
<tr>
<td>Corr: 8000</td>
<td>0.60</td>
<td>0.71</td>
</tr>
<tr>
<td>Corr: 12000</td>
<td>0.61</td>
<td>0.71</td>
</tr>
</tbody>
</table>
Table 3: Model estimates, and comparison with results from Chernov et al (2002) and Andersen et al (2002). New estimates in bold face. Model:

\[
\frac{dS}{S} = (\mu_0 + \mu_1 V - \lambda_i \bar{k}) dt + \sqrt{V} dW + (e^\gamma - 1) dN
\]

\[
dV = (\alpha - \beta V) dt + \sigma \sqrt{V} dW \]

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

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


All parameters in annualized units except the volatility shock half-life \( HL = 12 \ln 2 / \beta \), which is in months.

<table>
<thead>
<tr>
<th></th>
<th>SV</th>
<th>SV0, ( \lambda_1 = 0 )</th>
<th>SVJ1, ( \lambda_1 \neq 0 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CGGT</td>
<td>ABL</td>
<td>Bates</td>
</tr>
<tr>
<td>( \mu_0 )</td>
<td>0.051</td>
<td>0.026</td>
<td>.037</td>
</tr>
<tr>
<td></td>
<td>(0.032)</td>
<td>(0.025)</td>
<td>(0.045)</td>
</tr>
<tr>
<td>( \mu_1 )</td>
<td>2.58</td>
<td>3.70</td>
<td>4.02</td>
</tr>
<tr>
<td></td>
<td>(2.82)</td>
<td>(1.98)</td>
<td>(3.89)</td>
</tr>
<tr>
<td>( \alpha )</td>
<td>1.283</td>
<td>0.093</td>
<td>0.044</td>
</tr>
<tr>
<td></td>
<td>(0.010)</td>
<td>(0.011)</td>
<td>(0.013)</td>
</tr>
<tr>
<td>( \beta )</td>
<td>137.87</td>
<td>3.93</td>
<td>5.94</td>
</tr>
<tr>
<td></td>
<td>(0.17)</td>
<td>(0.81)</td>
<td>(0.54)</td>
</tr>
<tr>
<td>( \sigma )</td>
<td>1.024</td>
<td>0.197</td>
<td>0.315</td>
</tr>
<tr>
<td></td>
<td>(0.030)</td>
<td>(0.018)</td>
<td>(0.018)</td>
</tr>
<tr>
<td>( \lambda_0 )</td>
<td>-1.99</td>
<td>-0.579</td>
<td>-0.579</td>
</tr>
<tr>
<td></td>
<td>(0.000)</td>
<td>(0.045)</td>
<td>(0.031)</td>
</tr>
<tr>
<td>( \sqrt{\alpha / \beta} )</td>
<td>0.096</td>
<td>0.114</td>
<td>0.125</td>
</tr>
<tr>
<td></td>
<td>(0.004)</td>
<td>(0.004)</td>
<td>(0.004)</td>
</tr>
<tr>
<td>( HL )</td>
<td>0.06</td>
<td>2.12</td>
<td>1.40</td>
</tr>
<tr>
<td></td>
<td>(0.19)</td>
<td>(0.31)</td>
<td>(0.27)</td>
</tr>
<tr>
<td>( \lambda_0 )</td>
<td>1.70</td>
<td>5.09</td>
<td>0.744</td>
</tr>
<tr>
<td></td>
<td>(0.43)</td>
<td>(0.217)</td>
<td>(0.17)</td>
</tr>
<tr>
<td>( \gamma )</td>
<td>-0.030</td>
<td>-0.009</td>
<td>-0.009</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.010)</td>
<td>(0.007)</td>
</tr>
<tr>
<td>( \delta )</td>
<td>0.008</td>
<td>0.012</td>
<td>0.051</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.001)</td>
<td>(0.009)</td>
</tr>
<tr>
<td>( \ln L )</td>
<td>39,192.45</td>
<td>39,233.87</td>
<td>39,294.76</td>
</tr>
</tbody>
</table>
Figure 1: Actual and filtered variance estimates on simulated data.

Figure 2: Filtration error $\hat{V}_{t|t} - V_t$, and associated [5%, 25%, 75%, 95%] quantiles. 50% of the errors should fall in the gray area; 90% of the observations should fall in the grey and yellow areas.

Quantile $p$: .05  .10  .25  .50  .75  .90  .95

$\text{Prob}_t[V(t) < V_p(t)]:$ .043  .090  .243  .492  .743  .897  .951

(Sample path of 12,000 observations)
Figure 3: Distribution of normalized asset returns $N^{-1}(CDF(y_{t+1} \mid I_t))$ 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 -5.44.
Figure 4: Filtered estimates $\hat{\gamma}_t$ from stochastic volatility and stochastic volatility/jump models SV and SVJ1, respectively.
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 an standardized asset return of magnitude $z = y_{t+1} / \sqrt{V_{t+1} \Delta t}$. 
Figure 6. Estimated number of jumps, versus standardized returns $y_{t+1}/\sqrt{V_{t+1}^\Delta t}$. The values are approximately the probability that a jump occurred.