

Maximum Likelihood Estimation of Latent Affine Processes

David S. Bates

University of Iowa

This article develops a direct filtration-based maximum likelihood methodology for estimating the parameters and realizations of latent affine processes. Filtration is conducted in the transform space of characteristic functions, using a version of Bayes' rule for recursively updating the joint characteristic function of latent variables and the data conditional upon past data. An application to daily stock market returns over 1953–1996 reveals substantial divergences from estimates based on the Efficient Methods of Moments (EMM) methodology; in particular, more substantial and time-varying jump risk. The implications for pricing stock index options are examined.

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

Jacob Ziegler; translated by Richard Eden (1555)

While models proposing time-varying volatility of asset returns have been around for 30 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.

Estimating unobserved and time-varying volatility and jump risk from stock returns is an example of the general state space problem of inferring latent variables from observed data. This problem has two associated subproblems:

I am grateful for comments on earlier versions of this article from Yacine Aït-Sahalia, Luca Benzoni, Michael Chernov, Frank Diebold, Garland Durham, Bjorn Eraker, Michael Florig, George Jiang, Michael Johannes, Ken Singleton, George Tauchen, Chuck Whiteman, and the anonymous referee. Comments were also appreciated from seminar participants at Colorado, ESSEC, Goethe University (Frankfurt), Houston, Iowa, NYU, the 2003 WFA meetings, the 2004 IMA workshop on Risk Management and Model Specification Issues in Finance, and the 2004 Bachelier Finance Society meetings. Address correspondence to David S. Bates, Henry B. Tippie College of Business, University of Iowa, Iowa City, IA 52242-1000, or e-mail: david-bates@uiowa.edu.

1. the *estimation* issue of identifying the parameters of the state space system
2. the *filtration* issue of estimating current values of latent variables from past data, given the parameter estimates.

For instance, the filtration issue of estimating the current level of underlying volatility is the key issue in risk assessment approaches such as Value at Risk. The popular GARCH approach of modeling latent volatility as a deterministic function of past data can be viewed as a simple method of specifying and estimating a volatility filtration algorithm.¹

This article proposes a new recursive maximum likelihood methodology for *semi-affine processes*. The key feature of these processes is that the joint characteristic function describing the joint stochastic evolution of the (discrete-time) data and the latent variables is assumed exponentially affine in the latent variable, but not necessarily in the observed data. Such processes include the general class of affine continuous-time jump-diffusions discussed in Duffie, Pan, and Singleton (2000), the time-changed Lévy processes of Carr, Geman, Madan, and Yor (2003), and various discrete-time stochastic volatility models such as the Gaussian AR(1) log variance process.

The major innovation is to work almost entirely in the transform space of characteristic functions, rather than working with probability densities. While both approaches are, in principle, equivalent, working with characteristic functions has a couple of advantages. First, the approach is more suited to filtration issues, since conditional moments of the latent variable are more directly related to characteristic functions than to probability densities. Second, the set of state space systems with analytic transition densities is limited, whereas there is a broader set of systems for which the corresponding conditional characteristic functions are analytic.

I use a version of Bayes' rule for updating the characteristic function of a latent variable conditional upon observed data. Given this and the semi-affine structure, recursively updating the characteristic functions of observations and latent variables conditional upon past observed data is relatively straightforward. Conditional probability densities of the data needed for maximum likelihood estimation can 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

¹ See Nelson (1992), Nelson and Foster (1994), and Fleming and Kirby (2003) for filtration interpretations of GARCH models.

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.

A caveat is that the updating procedure does involve numerical approximation of $G_{t|t}(\psi)$. The overall estimation procedure is consequently termed *approximate maximum likelihood* (AML), with potentially some loss of estimation efficiency relative to an exact maximum likelihood procedure. However, estimation efficiency could be improved in this approach by using more accurate approximations. Furthermore, the simple moment-based approximation procedure used here is a numerically stable filtration that performs well on simulated data, with regard to both parameter estimation and latent variable filtration.

This approach is, of course, only the latest in a considerable literature concerned with the problem of estimating and filtering state space systems of observed data and stochastic latent variables. Previous approaches include

1. analytically tractable specifications, such as Gaussian and regime-switching specifications
2. GMM approaches based on analytic moment conditions
3. simulation-based approaches, such as Gallant and Tauchen's (2002) Efficient Method of Moments, or Jacquier, Polson, and Rossi's (1994) Monte Carlo Markov Chain approach.

The major advantage of AML is that it provides an integrated framework for parameter estimation *and* latent variable filtration of value for managing risk and pricing derivatives. Moment- and simulation-based approaches, by contrast, focus primarily on parameter estimation, to which must be appended an additional filtration procedure to estimate latent variable realizations. Gallant and Tauchen (1998, 2002), for instance, propose identifying via "reprojection" an appropriate rule for inferring volatility from past returns, given an observable relationship between the two series in the many simulations. Johannes, Polson, and Stroud (2002) append a particle filter to the MCMC parameter estimates of Eraker, Johannes, and Polson (2003). And while filtration is an integral part of Gaussian and regime-switching models, it remains an open question as to whether these are adequate approximations of the asset return/latent volatility state space system.²

² Ruiz (1994) and Harvey, Ruiz, and Shephard (1994) apply the Kalman filtration associated with Gaussian models to latent volatility. 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.

The AML approach has two weaknesses. First, the approach is limited at present to semi-affine processes, whether discrete- or continuous-time, whereas simulation-based methods are more flexible. 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. Furthermore, some recent interesting expanded-data approaches also fit within the affine structure; for instance, the intradaily “realized volatility” used by Anderson, Bollerslev, Diebold, and Ebens (2001). Finally, some discrete-time nonaffine processes become affine after appropriate data transformations [e.g., the stochastic log variance model examined *inter alia* by Harvey, Ruiz, and Shephard (1994) and Jacquier, Polson, and Rossi (1994)].

Second, AML has a “curse of dimensionality” originating in its use of numerical integration. It is best suited for a single data source; two data sources necessitate bivariate integration, while using higher-order data is probably infeasible. However, extensions to multiple latent variables appear possible.

The parameter estimation efficiency of AML appears excellent for two processes for which we have performance benchmarks. For the discrete-time log variance process, AML is more efficient than EMM and almost as efficient as MCMC, while AML and MCMC estimation efficiency are comparable for the continuous-time stochastic volatility/jump model with constant jump intensity. Furthermore, AML’s *filtration* efficiency also appears to be excellent. For continuous-time stochastic volatility processes, AML volatility filtration is substantially more accurate than GARCH when jumps are present, while leaving behind little information to be gleaned by EMM-style reprojection.

Section 1 in this article derives the basic algorithm for arbitrary semi-affine processes, and discusses alternative approaches. Section 2 runs diagnostics, using data simulated from continuous-time affine stochastic volatility models with and without jumps. 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 (2003). For direct comparison with EMM-based estimates, I use the Anderson et al. data set of daily S&P 500 returns over 1953–1996, which were graciously provided by Luca Benzoni. Section 4 discusses option pricing implications, while Section 5 concludes.

1. Recursive Evaluation of Likelihood Functions for Affine Processes

Let $\mathbf{y}_t$ denote an $(L \times 1)$ vector of variables observed at discrete dates indexed by 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 strictly stationary and strongly ergodic.
3. The characteristic function of $\mathbf{z}_{t+1}$ conditional upon observing $\mathbf{z}_t$ is an exponentially affine function of the latent variables $\mathbf{x}_t$:

$$\begin{aligned} F(i\Phi, i\psi | \mathbf{z}_t) &\equiv E[e^{i\Phi' \mathbf{y}_{t+1} + i\psi' \mathbf{x}_{t+1}} | \mathbf{z}_t] \\ &= \exp[C(i\Phi, i\psi; \mathbf{y}_t) + D(i\Phi, i\psi; \mathbf{y}_t)' \mathbf{x}_t] \end{aligned} \quad (1)$$

Equation (1) is the joint conditional characteristic function associated with the joint transition density $p(\mathbf{z}_{t+1} | \mathbf{z}_t)$. As illustrated in the appendices, it can be derived from either discrete- or continuous-time specifications of the stochastic evolution of $\mathbf{z}_t$. The conditional characteristic function in Equation (1) can also be generalized to include dependencies of $\mathbf{C}(\bullet)$ and $\mathbf{D}(\bullet)$ upon other exogenous variables, such as seasonalities or irregular time gaps from weekends and holidays.

Processes satisfying the above conditions will be called *semi-affine processes* (i.e., processes with a conditional characteristic function that is exponentially linear in the latent variables $\mathbf{x}_t$ but not necessarily in the observed data $\mathbf{y}_t$). Processes for which the conditional characteristic function is also exponentially affine in $\mathbf{y}_t$ and, therefore, of the form

$$F(i\Phi, i\psi | \mathbf{z}_t) = \exp[\mathbf{C}^0(i\Phi, i\psi) + \mathbf{C}^y(i\Phi, i\psi)' \mathbf{y}_t + \mathbf{D}(i\Phi, i\psi)' \mathbf{x}_t] \quad (2)$$

will be called *fully affine* processes. Fully affine processes have been extensively used as models of stock and bond price evolution. Examples include

1. The continuous-time affine jump-diffusions summarized in Duffie, Pan, and Singleton (2000)
2. The time-changed Lévy processes considered by Carr, Geman, Madan and Yor (2001) and Huang and Wu (2004), in which the rate of information flow follows a square-root diffusion
3. All discrete-time Gaussian state-space models
4. The discrete-time AR(1) specification for log variance $x_t \equiv \ln V_t$, where y_t is either the demeaned log absolute return [Ruiz (1994)] or the high-low range [Alizadeh, Brandt, and Diebold (2002)].

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 involves multi-dimensional integration for higher-dimensional $\mathbf{y}_t$. It is also assumed that the univariate data y_t observed at discrete time intervals have been made stationary if necessary, as is standard practice in time series analysis. In

financial applications this is generally done via differencing, such as the log-differenced asset prices used below. Alternative techniques can also be used (e.g., detrending if the data are assumed stationary around a trend).

While written in a general form, specification (1) or (2) actually places substantial constraints upon possible stochastic processes. Fully affine processes are closed under time aggregation; by iterated expectations, the multiperiod joint transition densities $p(\mathbf{z}_{t+n}|\mathbf{z}_t, n)$ for $n > 1$ have associated conditional characteristic functions that are also exponentially fully affine in the state variables $\mathbf{z}_t$. By contrast, semi-affine processes do not time-aggregate in general, unless also fully affine. Consequently, while it is possible to derive Equation (2) from continuous-time affine processes such as those in Duffie, Pan, and Singleton (2000), it is not possible in general to derive processes satisfying Equation (1) but not Equation (2) from a continuous-time foundation. However, there do exist discrete-time semi-affine processes that are not fully affine; for instance, $\mathbf{z}_{t+1}$ multivariate normal with conditional moments that depend linearly on x_t but nonlinearly on y_t .

The best-known discrete-time affine process is the Gaussian state-space system discussed in Hamilton (1994, Ch. 13), for which the conditional density function $p(\mathbf{z}_{t+1}|\mathbf{z}_t)$ 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 conditional density functions. Their popularity for bond and option pricing models lies in the ability to compute densities, distributions, and option prices 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. Each fully summarizes what is known about a given random variable.

In order to conduct the equivalent of Kalman filtration using conditional characteristic functions, we need the equivalent of Bayesian updating for characteristic functions. The following describes the procedure for arbitrary random variables.

1.1 Bayesian updating of characteristic functions: the static case

Consider an arbitrary pair (y, x) of continuous random variables, with joint probability density $p(y, x)$. Let

$$\begin{aligned} F(i\Phi, i\Psi) &\equiv E[e^{i\Phi y + i\Psi x}] \\ &= \iint e^{i\Phi y + i\Psi x} p(y, x) dy dx \end{aligned} \quad (3)$$

be their joint characteristic function. The marginal densities $p(y)$ and $p(x)$ have corresponding univariate characteristic functions $F(i\Phi, 0)$ and $F(0, i\psi)$, respectively. It will prove convenient below to label the characteristic function of the latent variable x separately:

$$\begin{aligned} G(i\psi) &\equiv E[e^{i\psi x}] \\ &= \int e^{i\psi x} p(x) dx \\ &= F(0, i\psi). \end{aligned} \quad (4)$$

Similarly, $G(\psi) = E[e^{\psi x}]$ is the moment generating function of x , while $g(\psi) \equiv \ln G(\psi)$ is its cumulant generating function. Derivatives of these evaluated at $\psi = 0$ provide the noncentral moments and cumulants of x , respectively.

The characteristic functions in Equations (3) and (4) are the Fourier transforms of the probability densities. Correspondingly, the probability densities can be evaluated as the inverse Fourier transforms of the characteristic functions; for instance,

$$p(x) = \frac{1}{2\pi} \int G(i\psi) e^{-i\psi x} d\psi \quad (5)$$

and

$$p(y, x) = \frac{1}{(2\pi)^2} \iint F(i\Phi, i\psi) e^{-i\Phi y - i\psi x} d\Phi d\psi. \quad (6)$$

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

Proposition 1. [Bartlett (1938)]. *The characteristic function of x conditional upon observing y is*

$$G_{x|y}(i\psi|y) = \frac{\frac{1}{2\pi} \int F(i\Phi, i\psi) e^{-i\Phi y} d\Phi}{p(y)} \quad (7)$$

where

$$p(y) = \frac{1}{2\pi} \int F(i\Phi, 0) e^{-i\Phi y} d\Phi \quad (8)$$

is the marginal density of y .

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

$$\begin{aligned} G_{x|y}(i\psi|y) &= \int e^{i\psi x} p(x|y) dx \\ &= \frac{1}{p(y)} \int e^{i\psi x} p(y, x) dx. \end{aligned} \quad (9)$$

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

$$\begin{aligned} F(i\Phi, i\psi) &= \int e^{i\Phi y} [G_{x|y}(i\psi|y)p(y)] dy \\ &= \iint e^{i\Phi y + i\psi x} p(y, x) dx dy. \end{aligned} \quad (10)$$

Consequently, $G_{x|y}(i\psi|y)p(y)$ is the *inverse* Fourier transform of $F(i\Phi, \bullet)$, yielding Equation (7) above.

1.2 Dynamic Bayesian updating of characteristic functions

While expressed in a static setting, Proposition 1 also applies to conditional expectations and the sequential updating of the characteristic functions of dynamic latent variables x_t conditional upon the latest datum y_t . The following notation will be used below for such filtration:

$Y_t \equiv \{y_1, \dots, y_t\}$ is the data observed by the econometrician up through date t .

$E[\cdot|Y_t] \equiv E_t(\cdot)$ is the expectational operator conditional upon observing data Y_t .

$F(i\Phi, i\psi|z_t) = E[e^{i\Phi y_{t+1} + i\psi x_{t+1}} | y_t, x_t]$ is the joint characteristic function [Equation (1)] conditional upon observing the Markov state variables z_t (including the latent variable realization x_t).

$F(i\Phi, i\psi|Y_t) = E[e^{i\Phi y_{t+1} + i\psi x_{t+1}} | Y_t]$ is the joint characteristic function of next period's variables conditional upon observing only past data Y_t .

$G_{t|s}(i\psi) \equiv E[e^{i\psi x_t} | Y_s]$ is the characteristic function of the latent variable x_t at time t conditional upon observing Y_s .

Given Proposition 1 and the semi-affine structure, the filtered characteristic function $G_{t|t}(i\psi)$ 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 unconditional characteristic function of the latent variable. This can be computed from the conditional characteristic function associated with the multiperiod transition density $p(x_t|y_0, x_0; t)$, by taking the limit as t goes to infinity. For fully affine processes, the multiperiod conditional characteristic function is of the affine form³

³ Semi-affine processes will also generate multiperiod characteristic functions of the form in Equation (11) if $p(x_{t+1}|x_t)$ does not depend on y_t . Such processes will have $e^{\psi(i\psi, t)} = 0$.

$$E[e^{i\psi x_t}|y_0, x_0] = \exp[c^0(i\psi, t) + c^y(i\psi, t)y_0 + d(i\psi, t)x_0]. \quad (11)$$

Since x_t is assumed strictly stationary and strongly ergodic, the limit as t goes to infinity does not depend on initial conditions (y_0, x_0) and converges to the unconditional characteristic function:⁴

$$\begin{aligned} G_{0|0}(i\psi) &= E[e^{i\psi x_t}] \text{ for all } t \\ &= \lim_{t \rightarrow \infty} E[e^{i\psi x_t}|y_0, x_0] \\ &= \lim_{t \rightarrow \infty} \exp[c^0(i\psi, t)]. \end{aligned} \quad (12)$$

For continuous-time (fully affine) models, the time interval between observations is a free parameter; solving Equation (2) also provides an analytic solution for Equation (11).⁵ For discrete-time models, Equation (11) can be solved from Equation (2) by repeated use of iterated expectations, yielding a series solution for $c^0(i\psi, \infty)$.

Step 1: Given $G_{t|t}(\psi) \equiv E[e^{\psi x_t}|Y_t]$, the joint characteristic function of next period's (y_{t+1}, x_{t+1}) conditional on data observed through date t can be evaluated by iterated expectations, exploiting the Markov assumption and the semi-affine structure of characteristic functions given in Equation (1):

$$\begin{aligned} F(i\Phi, i\psi|Y_t) &= E[E(e^{i\Phi y_{t+1} + i\psi x_{t+1}}|y_t, x_t)|Y_t] \\ &= E[e^{C(i\Phi, i\psi; y_t) + D(i\Phi, i\psi; y_t)x_t}|Y_t] \\ &= e^{C(i\Phi, i\psi; y_t)} G_{t|t}[D(i\Phi, i\psi; y_t)]. \end{aligned} \quad (13)$$

Step 2: The density function of next period's datum y_{t+1} conditional upon data observed through date t can be evaluated by Fourier inversion of its conditional characteristic function:

$$p(y_{t+1}|Y_t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(i\Phi, 0|Y_t) e^{-i\Phi y_{t+1}} d\Phi. \quad (14)$$

Step 3: Using Proposition 1, the conditional characteristic function of next period's x_{t+1} is

⁴ See Karlin and Taylor (1981), pp. 220–221 and 241.

⁵ An illustration is provided in Appendix A.2, Equations (A.19) and (A.20).

$$\begin{aligned}
 G_{t+1|t+1}(i\psi) &= \frac{\frac{1}{2\pi} \int_{-\infty}^{\infty} F(i\Phi, i\psi | \mathbf{Y}_t) e^{-i\Phi y_{t+1}} d\Phi}{p(y_{t+1} | \mathbf{Y}_t)} \\
 &= \frac{\frac{1}{2\pi} \int_{-\infty}^{\infty} G_{t|t}[D(i\Phi, i\psi; y_t)] e^{C(i\Phi, i\psi; y_t) - i\Phi y_{t+1}} d\Phi}{p(y_{t+1} | \mathbf{Y}_t)}
 \end{aligned} \tag{15}$$

Step 4: Repeat steps 1–3 for subsequent values of t . Given underlying parameters θ , the log likelihood function for maximum likelihood estimation is $\ln L(\mathbf{Y}_T | \theta) = \ln p(y_1 | \theta) + \sum_{t=2}^T \ln p(y_t | \mathbf{Y}_{t-1}, \theta)$.

$G_{t|t}(i\psi)$ is the time- t prior characteristic function of the latent variable x_t , while $F(i\Phi, i\psi | \mathbf{Y}_t)$ in Equation (13) is the time- t prior joint characteristic function for (y_{t+1}, x_{t+1}) . Step 3 is the equivalent of Bayesian updating and yields the posterior characteristic function of latent x_{t+1} —which is also the time- $(t+1)$ prior characteristic function for the next time step. The equivalent steps for updating moment generating functions and the associated conditional density functions are given in Table 1.

Table 1
Fourier inversion approach to computing likelihood functions

Densities	Associated moment generating functions
Conditional density of x_t :	
$p(x_t \mathbf{Y}_t)$	$G_{t t}(\psi)$
Joint conditional density of (y_{t+1}, x_{t+1}) :	
$p(y_{t+1}, x_{t+1} \mathbf{Y}_t)$	$F(\Phi, \psi \mathbf{Y}_t) = E \left[E \left(e^{\Phi y_{t+1} + \psi x_{t+1}} \mid y_t, x_t \right) \mid \mathbf{Y}_t \right]$
$= \int p(y_{t+1}, x_{t+1} y_t, x_t) p(x_t \mathbf{Y}_t) dx_t$	$= E \left[e^{C(\Phi, \psi; y_t) + D(\Phi, \psi; y_t)x_t} \mid \mathbf{Y}_t \right]$
Conditional density evaluation:	$= e^{C(\Phi, \psi; y_t)} G_{t t}[D(\Phi, \psi; y_t)]$
$p(y_{t+1} \mathbf{Y}_t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(i\Phi, 0 \mathbf{Y}_t) e^{-i\Phi y_{t+1}} d\Phi$	
Updated conditional density of x_{t+1} :	
$p(x_{t+1} \mathbf{Y}_{t+1}) = \frac{p(y_{t+1}, x_{t+1} \mathbf{Y}_t)}{p(y_{t+1} \mathbf{Y}_t)}$	$G_{t+1 t+1}(\psi) = \frac{\frac{1}{2\pi} \int_{-\infty}^{\infty} F(i\Phi, \psi \mathbf{Y}_t) e^{-i\Phi y_{t+1}} d\Phi}{p(y_{t+1} \mathbf{Y}_t)}$

Moment generating functions:

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

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

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

is the moment generating function of x_t conditional on observing data $\mathbf{Y}_t = \{y_1, \dots, y_t\}$. Its initial value $G_{0|0}(\psi) = E[\exp(\psi x_0)]$ is the unconditional moment generating function of x_0 . Subsequent $G_{t|t}$ s and the conditional densities of the data used in the likelihood function can be recursively updated as shown in the table.

Filtered estimates of next period's latent variable realization and the accompanying precision can be computed from derivatives of the moment generating function $G_{t+1|t+1}(\psi)$ in Equation (15):

$$\begin{aligned}\hat{x}_{t+1|t+1} &= G'_{t+1|t+1}(0) \\ &= \frac{1}{2\pi p(y_{t+1}|Y_t)} \int_{-\infty}^{\infty} F_{\psi}(i\Phi, 0|Y_t) e^{-i\Phi y_{t+1}} d\Phi\end{aligned}\quad (16)$$

$$\begin{aligned}P_{t+1|t+1} &\equiv \text{Var}_{t+1}(x_{t+1}) \\ &\equiv G''_{t+1|t+1}(0) - \hat{x}_{t+1|t+1}^2 \\ &= \frac{1}{2\pi p(y_{t+1}|Y_t)} \int_{-\infty}^{\infty} F_{\psi\psi}(i\Phi, 0|Y_t) e^{-i\Phi y_{t+1}} d\Phi - \hat{x}_{t+1|t+1}^2.\end{aligned}\quad (17)$$

1.3 Implementation

The recursion in Equations (13)–(15) indicates that, for a given prior characteristic function $G_{t|t}(i\psi)$ of latent x_t and an observed datum y_{t+1} , it is possible to compute an updated posterior characteristic function $G_{t+1|t+1}(i\psi)$ that fully summarizes the filtered distribution of latent x_{t+1} . 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 Chebyshev polynomials, it is possible to achieve arbitrarily precise approximations to $G_{t|t}(i\psi)$.

However, such atheoretic methods do not necessarily preserve the shape restrictions that make a given atheoretic approximating function $\hat{G}_{t|t}(i\psi)$ 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

$$\begin{aligned}p(x) &= \left[1 + \frac{\kappa_4}{24}(x^4 - 6x^2 + 3)\right] n(x) \\ G(i\psi) &= e^{-1/2\psi^2} \left(1 + \frac{\kappa_4}{24}\psi^4\right)\end{aligned}\quad (18)$$

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 that are generated directly from distributions with known properties. As the recursion in Equations (13)–(15) is just Bayesian updating, any legitimate approximate prior $\hat{G}_{t|t}(i\psi)$ from a known distribution will generate a legitimate posterior characteristic function $\hat{G}_{t+1|t+1}(i\psi)$.

The choice of approximating function also involves a trade-off between accuracy and computational speed. Evaluating $G_{t+1|t+1}(D)$ at each complex-valued point D involves numerical integration. More evaluations map the surface more accurately but are also slower.

The following analysis therefore uses a simple moment-matching approximation similar in spirit to Kalman filtration. An approximate $\hat{G}_{t+1|t+1}(i\psi)$ from a two-parameter distribution is used at each time step. The parameter values at each step are determined by the conditional mean $\hat{x}_{t+1|t+1}$ and variance $P_{t+1|t+1}$ of the latent variable x_{t+1} , which are evaluated as described in Equations (16) and (17). The choice of distribution depends upon the properties of the latent variable. If x_{t+1} is unbounded, as in the discrete-time log variance model in Appendix B, the natural approximating characteristic function is Gaussian:

$$\ln \hat{G}_{t+1|t+1}(i\psi) = \hat{x}_{t+1|t+1}(i\psi) + 1/2 P_{t+1|t+1}(i\psi)^2. \quad (19)$$

If x_{t+1} is nonnegative, as in the continuous time square-root variance process of Appendix A.2, a gamma characteristic function is more appropriate:

$$\begin{aligned} \ln \hat{G}_{t+1|t+1}(i\psi) &= -v_{t+1} \ln(1 - \kappa_{t+1} i\psi) \\ \kappa_{t+1} &\equiv P_{t+1|t+1} / \hat{x}_{t+1|t+1} \\ v_{t+1} &\equiv \hat{x}_{t+1|t+1}^2 / P_{t+1|t+1}. \end{aligned} \quad (20)$$

In both cases, the initial unconditional distribution is a member of the family of conditional distributions. Given the moment matching, these approximations can be viewed as second-order Taylor approximations to the true log characteristic functions. Further details on computing these moments are in the appendices. Three numerical integrations are required at each time step: one for the conditional density [from Equation (14)] and two for the moment conditions in Equations (16) and (17).

The approach is analogous to the Kalman filtration approach used by Ruiz (1994) and Harvey, Ruiz, and Shephard (1994) of updating the conditional mean and variance of the latent state variable based on the latest datum. The major difference is that Kalman filtration uses strictly linear updating for the conditional mean, whereas Equations (16) and (17) give the optimal *nonlinear* moment updating rules, conditional upon

the distributional properties of observed y_{t+1} , and conditional upon a correctly specified prior characteristic function $G_{t|t}(i\psi)$. Approximation error enters in that the prior characteristic function $\hat{G}_{t|t}(i\psi)$ is approximate rather than exact, which can distort the appropriate relative weights of the datum and the prior when updating posterior moments.

There may well be better approximation methodologies that bring the approach closer to true maximum likelihood estimation. However, the above moment-matching approach can be shown to be a *numerically stable* filtration, despite the complexity of the concatenated integrations. Filtration errors die out as more data are observed. This follows from a vector autoregression representation of the data and the revisions in conditional moments derived in Appendix A.1. In the fully affine case from Equation (2), with joint conditional cumulant generating function of the form

$$\ln F(\Phi, \psi | y_t, x_t) = C^0(\Phi, \psi) + C^y(\Phi, \psi)y_t + D(\Phi, \psi)x_t, \quad (21)$$

the inputs to the algorithm follow a stable VAR of the form

$$\begin{pmatrix} y_{t+1} \\ \hat{x}_{t+1|t+1} \\ P_{t+1|t+1} \end{pmatrix} = \begin{pmatrix} C_\Phi^0 \\ C_\psi^0 \\ C_{\psi\psi}^0 \end{pmatrix} + \begin{pmatrix} C_\Phi^y & D_\Phi & 0 \\ C_\psi^y & D_\psi & 0 \\ C_{\psi\psi}^y & D_{\psi\psi} & D_\psi^2 \end{pmatrix} \begin{pmatrix} y_t \\ \hat{x}_{t|t} \\ P_{t|t} \end{pmatrix} + \begin{pmatrix} u_{t+1} \\ v_{t+1} \\ w_{t+1} - v_{t+1}^2 \end{pmatrix} \quad (22)$$

or

$$\mathbf{Z}_{t+1} = \mathbf{A} + \mathbf{B}\mathbf{Z}_t + \boldsymbol{\varepsilon}_{t+1} \quad (23)$$

where

1. $\hat{E}_t$ is the expectational operator using the approximate prior, conditional upon data $\mathbf{Y}_t$;
2. $u_{t+1} \equiv y_{t+1} - \hat{E}_t y_{t+1}$ is the observed prediction error for y_{t+1} of the approximate prior;
3. $v_{t+1} \equiv (\hat{E}_{t+1} - \hat{E}_t)x_{t+1}$ is the revision in the latent variable estimate using the above algorithm and approximate prior, given an additional datum y_{t+1} ;
4. $w_{t+1} \equiv (\hat{E}_{t+1} - \hat{E}_t)x_{t+1}^2 - 2v_{t+1}(C_\psi^0 + C_\psi^y y_t + D_\psi \hat{x}_{t|t})$;

and all partial derivatives (C_ψ^0 , D_ψ , etc.) are evaluated at $\Phi = \psi = 0$. Since x_t and y_t are assumed stationary, the block-triangular matrix $\mathbf{B}$ is well behaved: its eigenvalues have modulus less than 1, and $\mathbf{B}^n \rightarrow \mathbf{0}$ as $n \rightarrow \infty$.

The matrices $\mathbf{A}$ and $\mathbf{B}$ are determined analytically by the specification [from Equation (21)] of the affine process. Approximation issues consequently affect only the signals $\boldsymbol{\varepsilon}_{t+1}$ inferred from observing y_{t+1} . Under an exact prior ($\hat{E}_t = E_t$), the signals $(u_{t+1}, v_{t+1}, w_{t+1})$ would be serially uncorrelated and independent of all lagged data $\mathbf{Y}_t$. But even under an

approximate prior, the impact of inefficiently processing the latest observation y_{t+1} when computing the signals dies out geometrically over time. In the absence of news (e.g., when projecting forward in time), the conditional moments converge to the unconditional moments $(\mathbf{I} - \mathbf{B})^{-1}\mathbf{A}$, where $\mathbf{I}$ is the identity matrix.

Since the posterior variance $P_{t+1|t+1}$ is computed by Bayesian updating, it is strictly positive.⁶ Since the approximate Bayesian updating algorithm depends upon the prior moments $(\hat{x}_{t|t}, P_{t|t})$ and the latest observation y_{t+1} , all of which are stationary, the signals ϵ_{t+1} are also stationary, and the overall system is well behaved. The extent to which the use of an approximate prior leads to inefficient signals is of course an important issue, which will be examined later in this article for specific models. However, the matrices $\mathbf{A}$ and $\mathbf{B}$ that determine the weighting of past signals when computing the mean and variance of x_{t+1} are not affected by approximation issues.

It may be possible to generate better approximations to the conditional characteristic functions $\hat{G}_{t|t}(i\psi)$ by using multiparameter approximating distributions that match higher moments as well. Noncentral moments of all orders also have a stable block-triangular VAR representation similar to Equation (22), so higher-moment filtrations will also be numerically stable.

1.4 Comparison with alternate approaches

There have, of course, been many alternative approaches to the problem of estimating and filtering dynamic state space systems. These approaches fall into two categories: those that assume the Markov system $\mathbf{z}_t$ is fully observable and those for which $\mathbf{z}_t = (y_t, x_t)$ contains latent components x_t that must be estimated from past realizations $\mathbf{Y}_{t-1}$. Included in the first category are those approaches that infer the current values of x_t from other data sources: from bond prices for multifactor term structure models, from option prices for latent volatility, or deterministically from past returns in GARCH models. In the discussion that follows, I will primarily focus on the approaches for estimating affine models of stochastic volatility.

If $\mathbf{z}_t$ is fully observable or inferable, direct maximum likelihood estimation of the parameter vector $\boldsymbol{\theta}$ is, in principle, feasible. In some cases, the log transition densities $\ln p(\mathbf{z}_t|\mathbf{z}_{t-1}, \boldsymbol{\theta})$ that enter into the log likelihood function can be analytic, as in the multifactor CIR model of Pearson and Sun (1994).⁷ Alternatively, the log densities may be

⁶ In the special case of Kalman filtration, the variance revision $w_{t+1} - v_{t+1}^2$ is a nonpositive deterministic nonlinear function of the prior variance $P_{t|t}$, which in turn converges to a steady-state minimal value conditional upon a steady information flow [see Hamilton (1994, Ch. 13) or Appendix B]. For general affine processes, variance revisions are stochastic and can be positive.

⁷ Pearson and Sun note that, if $\mathbf{z}_t$ is inferred from observed data, an additional Jacobian term for the data transformation is required in the log likelihood function.

evaluated numerically by Fourier inversion of the conditional characteristic function, by numerical finite-difference methods [Lo (1988)], by simulation methods [e.g., Durham and Gallant (2002)], or by good approximation techniques [Aït-Sahalia (2002)]. Some [e.g., Pan (2002)] use GMM rather than maximum likelihood, based on conditional moment conditions derived from the conditional characteristic function. Feuerverger and McDunnough (1981a,b) show that a continuum of moment conditions derived directly from characteristic functions achieves the efficiency of maximum likelihood estimation. Singleton (2001) and Carrasco, Chernov, Florens, and Ghysels (2003) explore how to implement this *empirical characteristic function* approach using a finite number of moment conditions.

This article is concerned with the second category of state space systems—those that include latent variables. Current approaches for estimating such systems include

1. analytically tractable filtration approaches such as Gaussian and regime-switching models;
2. GMM approaches that use moment conditions evaluated either analytically or by simulation methods; and
3. the Bayesian Monte Carlo Markov Chain approach of Jacquier, Polson, and Rossi (1994) and Eraker, Johannes, and Polson (2003).

Filtration approaches typically rely on specific state space structures that permit recursive updating of the conditional density functions $p(y_t | Y_{t-1}, \theta)$ of observed data used in maximum likelihood estimation. While there exist some affine processes that fall within these categories,⁸ both Gaussian and regime-switching models can be poor descriptions of the stochastic volatility processes examined below. 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 articles use such models as approximations for volatility evolution and estimation. Harvey, Ruiz, and Shephard (1994) explore the Kalman filtration associated with Gaussian models, while Sandmann and Koopman (1998) add a simulation-based correction for the deviations from Gaussian distributions. 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.

One major strand of the literature on GMM estimation of stochastic volatility processes without or with jumps focuses on models for which moments of the form $E[y_t^m y_{t-L}^n]$ for $m, n \geq 0$ can be evaluated analytically. Examples include Melino and Turnbull (1990), Andersen and Sørensen

⁸ See, for example Jegadeesh and Pennacchi (1996) for a Kalman filtration examination of the multifactor Vasicek bond pricing model.

(1996), Ho, Perraudin, and Sørensen (1996), Jiang and Knight (2002), and Chacko and Viceira (2003). This is relatively straightforward for models with affine conditional characteristic functions. As illustrated in Jiang and Knight (2002), iterated expectations can be used to generate unconditional characteristic functions $F(i\Phi) \equiv E[\exp(i\Phi y_t)]$ or joint characteristic functions $F(i\Phi_0, \dots, i\Phi_L) \equiv E[\exp(i\Phi_0 y_t + \dots + i\Phi_L y_{t-L})]$. Unconditional moments and cross-moments of returns can then be computed by taking derivatives. Alternatively, one can generate moment conditions by directly comparing theoretical and empirical characteristic functions. Feuerverger (1990) shows that a continuum of such moment conditions for different values of Φ s is equivalent to maximum likelihood estimation premised on the limited-information densities $p(y_t|y_{t-1}, \dots, y_{t-L}; \theta)$ — a result cited in Jiang and Knight (2002) and Carrasco et al. (2003).

The second strand of GMM estimation evaluates theoretical moments numerically by Monte Carlo methods, using the simulated method of moments approach of McFadden (1989) and Duffie and Singleton (1993). Eliminating the requirement of analytic tractability greatly increases the range of moment conditions that can be used. The currently popular Efficient Method of Moments (EMM) methodology of Gallant and Tauchen (2002) uses first-order conditions from the estimation of an auxiliary discrete-time semi-nonparametric time series model as the moment conditions to be satisfied by the postulated continuous-time process.

The approximate maximum likelihood (AML) methodology is closest in spirit to the filtration approaches that evaluate $p(y_t|Y_{t-1}, \theta)$ recursively over time. Indeed, the discretization of possible variance realizations within the range of ± 3 unconditional standard deviations [Fridman and Harris (1998)] can be viewed as a particular point-mass approximation methodology for the conditional characteristic function:

$$\hat{G}_{t|t}(i\psi) \approx \sum_j \pi_{t|t}(j) \exp[i\psi x^{(j)}] \quad (24)$$

where $\pi_{t|t}(j) \equiv \text{Prob}[x_t = x^{(j)}|Y_t]$. Fridman and Harris use the Bayesian state probability updating of regime-switching models to recursively update the state probabilities $\pi_{t|t}$ over time.

The AML methodology has strengths and weaknesses relative to Fridman and Harris. AML can be used with a broad class of discrete- or continuous-time affine models, whereas the Fridman and Harris approach requires a discrete-time model with explicit conditional transition densities $p(x_t|x_{t-1}; \theta)$ for the latent variable. Second, AML can accommodate correlations between observed data and the latent variable evolution (e.g., correlated asset returns and volatility shocks), whereas the regime-switching structure used by Fridman and Harris relies on

conditional independence when updating state probabilities. Both approaches generate filtered estimates $\hat{x}_{t|t}$, but the Fridman and Harris approach can also readily generate smoothed estimates $\hat{x}_{t|T}$. Finally, the AML approach, at present, lacks a systematic method of increasing the accuracy of $\hat{G}_{t|t}$ estimates, whereas it is simple to increase the number of grid points in Equation (24).

The major advantage of AML relative to moment- and simulation-based approaches is that it directly provides filtered estimates $\hat{x}_{t|t}$ for use in assessing risk or pricing derivatives. Most other methods do not and must append an additional filtration methodology. Melino and Turnbull (1990), for instance, use an extended Kalman filter calibrated from their GMM parameter estimates. EMM practitioners use reprojection. The MCMC approach of Eraker, Johannes, and Polson (2003) provides smoothed but not filtered latent variable estimates, to which Johannes, Polson, and Stroud (2003) append a particle filter.

The other issue is, of course, how the various methods compare with regard to parameter estimation efficiency. AML presumably suffers some loss of efficiency for poor-quality $\hat{G}_{t|t}$ approximations. The performance of moment-based approaches depends upon moment selection. Ad hoc moment selection can reduce the performance of GMM approaches, while EMM's moment selection procedure asymptotically approaches the efficiency of maximum likelihood estimation. The latent-variable empirical characteristic function approaches of Jiang and Knight (2002) and Carrasco et al. (2003) can, at best, achieve the efficiency of a maximum likelihood procedure that uses the limited-information densities $p(y_t|y_{t-1}, \dots, y_{t-L}; \theta)$. Given that L is, in practice, relatively small, there may be substantial efficiency losses relative to maximum likelihood estimation that uses the densities $p(y_t|Y_{t-1}; \theta)$. Conditional moment estimates in Equation (22) place considerable weight on the signals from longer-lagged observations when the latent variable is persistent (i.e., when D_ψ is near 1).

It is not possible to state in general which approaches will work best for which models. However, some guidance is provided by Andersen, Chung, and Sørensen's (1999) summary of the relative estimation efficiency of various approaches for the benchmark log variance process. In Appendix B, I develop the corresponding AML methodology and append the results to those of Andersen et al. The Jacquier, Polson, and Rossi (1994) MCMC method competes with Sandmann and Koopman's (1998) approach for most efficient. The AML and Fridman and Harris (1998) approaches are almost as efficient, followed by EMM, GMM, and the inefficient Kalman filtration approach of Harvey, Ruiz, and Shephard (1994). And although Andersen et al. do not examine the latent-variable empirical characteristic function approach, Jiang and Knight's (2002) moment conditions resemble those of the GMM procedure and would probably perform comparably.

2. A Monte Carlo Examination of Parameter Estimation and Volatility Filtration

The above algorithm can be used with any discrete- or continuous-time model that generates a discrete-time exponentially semi-affine conditional characteristic function of the form of Equation (1) above. One such process is the continuous-time fully affine stochastic volatility/jump process

$$\begin{aligned} dS_t/S_t &= [\mu_0 + \mu_1 V_t - (\lambda_0 + \lambda_1 V_t)\bar{k}]dt + \sqrt{V_t}(\rho dW_{1t} + \sqrt{1 - \rho^2} dW_{2t}) \\ &\quad + (e^{\gamma_s} - 1)dN_t \\ dV_t &= (\alpha - \beta V_t)dt + \sigma\sqrt{V_t}dW_{1t} \end{aligned} \quad (25)$$

where

- dS_t/S_t is the instantaneous asset return;
- V_t is its instantaneous variance conditional upon no jumps;
- W_{1t} and W_{2t} are independent Wiener processes;
- N_t is a Poisson counter with intensity $\lambda_0 + \lambda_1 V_t$ for the incidence of jumps;
- $\gamma_s \sim N(\bar{\gamma}, \delta^2)$ is the random Gaussian jump in the log asset price conditional upon a jump occurring;
- $\bar{k}$ is the expected percentage jump size: $\bar{k} \equiv E(e^{\gamma_s} - 1) = e^{\bar{\gamma} + \frac{1}{2}\delta^2} - 1$.

This model generates an analytic, exponentially affine conditional characteristic function $F_{y,V}$ for observed discrete-time log-differenced asset prices $y_{t+1} \equiv \ln(S_{t+1}/S_t)$ and variance $x_{t+1} \equiv V_{t+1}$ that is given in Appendix A.2.

Variants of the model have been estimated on stock index returns by Andersen, Benzoni, and Lund (2002), Chernov, Gallant, Ghysels, and Tauchen (2003), and Eraker, Johannes, and Polson (2003). All use simulation-based methods. The first two articles (henceforth ABL and CGGT, respectively) use the SNP/EMM methodology of Gallant and Tauchen for daily stock index returns over 1953–1996 and 1953–1999, respectively. The third article (henceforth EJP) uses Bayesian MCMC methods for daily S&P returns over 1980–1999, as well as NASDAQ returns over 1985–1999. The latter two articles also examine the interesting affine specification in which there are jumps in latent variance that may be correlated with stock market jumps.

2.1 Parameter estimates on simulated data

To test the accuracy of parameter estimation from the AML algorithm in Section 1, 100 independent sample paths of daily returns and latent

variance were generated over horizons of 1,000–12,000 days (roughly 4–48 years) using a Monte Carlo procedure described in Appendix A.6. Three models were examined: a stochastic volatility (SV) process, a stochastic volatility/jump process (SVJ0) with constant jump intensity λ_0 , and a stochastic volatility/jump process (SVJ1) with jump intensity $\lambda_1 V_t$. Parameter values for the SV and SVJ1 models were based upon those estimated in Section 3. By contrast, the SVJ0 parameter values were taken from Eraker et al. (2003), in order to replicate their study of MCMC parameter estimation efficiency. The EJP parameter values are generally close to those estimated in Section 3. However, the annualized unconditional nonjump variance α/β is somewhat higher: $(14.1\%)^2$, as opposed to the AML estimate of $(12.0\%)^2$.

Parameters were estimated for each simulated set of returns by using the Davidon-Fletcher-Powell optimization routine in GQOPT to maximize the log-likelihood function computed using the algorithm described in Section 1. The true parameter values were used as starting values, with parameter transformations used to impose sign constraints. In addition, the parameter space was loosely constrained to rule out the sometimes extreme parameter values tested at the early stages of quadratic hill-climbing: $\sigma \in [.03, .50]$, $|\rho| < .95$, etc. None of these constraints was binding at the final optimized parameter estimates, with one exception: the $|\bar{\gamma}| < .15$ constraint was binding for 7 of the 100 SVJ1 runs on the shortest 1000-observation data samples. These runs had very low estimated jump intensities and were from simulations in which only a few small jumps occurred. As all seven estimates were observationally identical to the corresponding no-jump estimates from the SV model, parameter estimates $\lambda_1 = \bar{\gamma} = \delta = 0$ were used for these runs when computing the summary statistics reported in Table 4. The problem reflects the difficulty of estimating jump parameters on small data sets and is not an issue for longer data samples.

Various authors argue that maximum likelihood is ill suited for estimating mixtures of distributions. Arbitrarily high likelihood can be achieved if the conditional mean equals some daily return and the specification permits some probability of extremely low daily variance on that day.⁹ To preclude this, all models were estimated subject to the additional parameter constraint $2\alpha > \sigma^2$ —a constraint that was never binding at the final estimates. This constraint implies that latent variance cannot attain its lower barrier of zero and can be viewed as imposing a strong prior belief that there is not near-zero daily stock market variance when markets are open.

⁹ Hamilton (1994, p.689) discusses the issue in the context of regime-switching models. Honoré (1998) raises the issue for jump-diffusion processes.

The optimizations involved, on average, 6–10 steps for the SV model and 9–15 steps for the SVJ1 model, with on average about 14(18) log likelihood function evaluations per step for the SV (SVJ1) model given linear stretching and numerical gradient computation. The optimizations converged in fewer steps for the longer data sets. Actual computer time required for each optimization averaged between 4 minutes (1000 observations) and 30 minutes (12,000 observations) for the SV model, and between 10 and 80 minutes for the SVJ1 model, on a 3.2 GHz Pentium 4 PC. Estimation of the SVJ0 model on 4000 observations averaged about 29 minutes per optimization.

Table 2 summarizes the results of replicating the EJP Monte Carlo examination of MCMC parameter estimation efficiency for the SVJ0 model, on 4000-day samples with two jump intensities of 1.51 and 3.78 jumps per year, respectively. Overall, it would appear that AML parameter estimation efficiency is broadly comparable to that of MCMC. Estimation biases are smaller for the AML procedure, especially for the variance mean reversion parameter β . The root mean squared errors of

Table 2
Parameter estimates on simulated daily data: SVJ0 model

	Mean	SV parameters				Jump parameters		
	μ_0^*	α/β	β	σ	ρ	λ_0	$\bar{\gamma}$	δ
True values:	.126	.020	3.78	.252	-.400	1.51	-.030	.035
Bias								
MCMC	.007	.001	1.49	.029	-.023	.20	-.007	-.005
AML	.000	.000	.35	.001	.004	.05	-.001	-.004
RMSE								
MCMC	.033	.003	2.12	.037	.065	0.61	.018	.007
AML	.030	.002	1.01	.027	.079	0.55	.014	.009
True values:	.126	.020	3.78	.252	-.400	3.78	-.030	.035
Bias								
MCMC	.009	.002	1.61	.043	-.013	.10	-.004	.001
AML	-.002	.000	.38	.001	-.008	-.05	-.002	-.002
RMSE								
MCMC	.035	.010	2.21	.075	.067	.79	.009	.005
AML	.034	.002	1.00	.025	.087	.86	.009	.005

MCMC: Results from Eraker, Polson, and Rossi (2003, Table VII), in annualized units
AML: Approximate maximum likelihood estimates

$$d \ln S_t = \mu_0^* dt + \sqrt{V_t} \left(\rho dW_{1t} + \sqrt{1 - \rho^2} dW_{2t} \right) + \gamma_s dN_t$$
$$dV_t = (\alpha - \beta V_t) dt + \sigma \sqrt{V_t} dW_{1t}$$
$$Prob(dN_t = 1) = \lambda_0 dt, \gamma_s \sim N(\bar{\gamma}, \delta^2)$$

100 sample paths were simulated for data sets of 4000 days (roughly 16 years). All parameter estimates are in annualized units.

parameter estimates are generally comparable, with neither MCMC nor AML clearly dominating.¹⁰

Tables 3 and 4 summarize the results for the SV and SVJ1 models, respectively. The tables also report results for parameters (α, β, σ) 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. These latter estimates provide an unattainable bound on asymptotic parameter estimation efficiency for those parameters, given that latent variance is not in fact directly observed.

The AML estimation methodology appears broadly consistent, with the RMSE of parameter estimates roughly declining at rate $1/\sqrt{T}$.

Table 3
Parameter estimates on simulated daily data: SV model

Returns-based estimates										
	<i>T</i> (days)	mean		SV parameters				<i>V_t</i> -based estimates		
		μ ₀	μ ₁	√α/β	β	σ	ρ	√α/β	β	σ
True values:		.026	3.68	.126	5.94	.306	−.576	.126	5.94	.306
Average bias	1000	−.021	5.08	−.002	1.85	.013	−.033	−.002	.93	.000
$\hat{\theta} - \theta$	2000	−.025	5.43	.000	.67	.000	−.037	.000	.46	.000
	4000	.001	3.82	−.001	.63	.003	−.002	−.001	.37	−.001
	8000	.013	2.50	.001	.09	.000	.000	.000	.05	.000
	12000	.013	2.55	.001	−.04	−.005	−.001	.000	.12	.000
standard error	1000	.010	.80	.001	.31	.006	.011	.001	.17	.001
	2000	.007	.52	.001	.20	.004	.008	.001	.14	.001
	4000	.005	.39	.001	.13	.003	.006	.001	.09	.000
	8000	.003	.29	.000	.08	.002	.004	.000	.07	.000
	12000	.003	.28	.000	.06	.001	.003	.000	.05	.000
RMSE:	1000	.101	9.47	.011	3.57	.057	.114	.007	1.95	.007
	2000	.077	7.52	.010	2.10	.035	.087	.005	1.46	.005
	4000	.045	5.47	.006	1.47	.026	.056	.003	.97	.003
	8000	.036	3.79	.005	.85	.016	.040	.002	.66	.002
	12000	.034	3.75	.004	.61	.014	.033	.002	.53	.002
Correlation between	1000							.95	.66	.12
returns- and <i>V_t</i>	2000							.97	.76	.11
based estimates	4000							.95	.67	.03
	8000							.95	.72	.17
	12000							.93	.62	.35

$$dS_t/S_t = (\mu_0 + \mu_1 V_t)dt + \sqrt{V_t}(\rho dW_{1t} + \sqrt{1 - \rho^2}dW_{2t})$$

$$dV_t = (\alpha - \beta V_t)dt + \sigma\sqrt{V_t}dW_{1t}$$

100 sample paths were simulated for data sets of 1,000–12,000 days (roughly 4–48 years).

¹⁰ The results in EJP’s Table VII were converted into annualized units. In addition, it seems likely that EJP are reporting standard deviations rather than root mean squared errors of parameter estimates, since their reported RMSEs are occasionally less than the absolute bias. Consequently, the EJP numbers were also adjusted using $RMSE^2 = SD^2 + bias^2$. As estimated biases were generally small, only the $RMSE(\hat{\beta})$ numbers were significantly affected by this adjustment.

Table 4
Parameter estimates on simulated daily data: SVJ1 model

		Returns-based estimates										
		mean		SV parameters			jump parameters			V_T -based estimates		
		T (days)	μ_0	μ_1	$\sqrt{\alpha/\beta}$	β	σ	ρ	λ_1	$\bar{\gamma}$	δ	$\sqrt{\alpha/\beta}$ β σ
True values:			.040	3.09	.119	4.25	.246	-.611	93.4	-.024	.039	.119 4.25 .246
Average bias $\hat{\theta} - \theta$	1000		-.025	5.33	.000	1.58	.015	-.033	2.0 ^a	-.007 ^a	-.017 ^a	.000 .87 .001
	2000		-.022	5.32	-.001	1.16	.009	-.006	0.4	-.009	-.010	-.002 .67 .001
	4000		-.005	3.64	-.001	.53	.002	-.004	0.1	-.003	-.005	-.001 .34 -.001
	8000		.001	2.70	.000	.08	-.002	-.005	1.3	-.001	-.001	.000 .06 .000
	12000		.009	2.15	.001	.00	-.002	-.003	-2.7	-.002	-.001	.000 .07 .000
standard error	1000		.010	.90	.002	.27	.005	.013	8.3 ^a	.003 ^a	.004 ^a	.001 .16 .001
	2000		.007	.63	.001	.18	.003	.008	6.1	.002	.002	.001 .11 .000
	4000		.005	.43	.001	.12	.002	.006	3.9	.002	.001	.001 .08 .000
	8000		.003	.26	.001	.07	.001	.004	2.5	.001	.000	.001 .05 .000
	12000		.003	.24	.000	.05	.001	.004	2.1	.001	.000	.000 .04 .000
RMSE:	1000		.105	10.44	.016	3.17	.056	.134	83.1 ^a	.035 ^a	.043 ^a	.015 1.83 .005
	2000		.074	8.24	.010	2.12	.034	.084	61.1	.027	.018	.010 1.31 .004
	4000		.048	5.62	.007	1.29	.025	.056	39.2	.019	.011	.007 .83 .003
	8000		.032	3.72	.005	.70	.015	.038	24.8	.009	.005	.005 .53 .002
	12000		.027	3.23	.004	.50	.011	.038	21.0	.007	.004	.004 .41 .002
Correlation between returns- and V_T -based estimates	1000											.92 .68 .17
	2000											.71 .44 .07
	4000											.96 .66 .30
	8000											.95 .76 .33
	12000											.89 .63 .13

$$dS_t/S_t = (\mu_0 + \mu_1 V_t - \lambda_1 V_t \bar{k})dt + \sqrt{V_t}(\rho dW_{1t} + \sqrt{1 - \rho^2} dW_{2t}) + (e^{\gamma_t} - 1)dN_t$$
$$dV_t = (\alpha - \beta V_t) dt + \sigma \sqrt{V_t} dW_{1t}$$
$$Prob(dN_t = 1) = \lambda_1 V_t dt, \bar{\gamma}_s \sim N(\bar{\gamma}, \delta^2)$$

100 sample paths were simulated for data sets of 1,000–12,000 days (roughly 4–48 years).
^aJump parameter estimates for 7 runs (out of 100) were set to the observationally identical values of zero.

The estimated bias (average bias rows) for all parameters and parameter transformations also generally decreased with longer data samples. There do not appear to be significant biases in the estimates of jump parameters: the sensitivity λ_1 of jump intensities to changes in latent variance, or the mean and standard deviation of jumps conditional upon jumps occurring. However, the λ_1 estimates are quite noisy even with 48 years of daily data. The RMSE of 21.0 is a substantial fraction of the true value of 93.4. There are substantial biases for two parameter estimates: the sensitivity μ_1 of expected stock returns to the current level of variance and the parameter β that determines the serial correlation and the associated half-life of the variance process. The $\hat{\mu}_1$ bias remains even at 12000-day

(48-year) samples; the $\hat{\beta}$ bias is still substantial at 16-year samples but disappears for longer samples. The latter bias reflects in magnified fashion the small-sample biases for a persistent series that would exist even if the V_t series were directly observed, as is illustrated in the second sets of $\hat{\beta}$ estimates conditioned upon observing V_t .¹¹ Here, of course, the values of latent variance must be inferred from noisy returns data, which almost double the magnitude of the bias relative to the V_t -dependent estimates.

The estimates of some of the parameters of the latent stochastic variance process perform surprisingly well. Daily returns are very noisy signals of daily latent variance, and yet the RMSE of returns-based parameter estimates of $\sqrt{\alpha/\beta}$ and β is typically less than double that of estimates conditioned on actually observing the underlying variance. Furthermore, the parameter estimates 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 observed, its volatility parameter σ would be pinned down quite precisely even for relatively small data sets: a RMSE of .005–.007 on data sets 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. Similar results are reported for the discrete-time log variance process in Appendix B. The mean reversion parameter estimate for latent log variance has two to three times the RMSE of estimates conditioned on actually observing log variance, while the RMSE of the estimate of the volatility of log variance rises seven- to eightfold.

2.2 Filtration

A major advantage of the filtration algorithm is that it provides estimates of latent variable realizations conditional upon past data. Figure 1 illustrates how volatility assessments are updated conditional upon the last observation for three models estimated below: the stochastic volatility model (SV), the stochastic volatility/jump model with constant jump intensities (SVJ0), and the stochastic volatility/jump model with variance-dependent jump intensities (SVJ1). For comparability with Hentschel's (1995) study of GARCH models, the figure illustrates *volatility* revisions $(E_{t+1} - E_t)\sqrt{V_{t+1}}$, using the conditional moments¹² of V_{t+1} and the Taylor approximation

¹¹ See Nankervis and Savin (1988) and Stambaugh (1999) for a discussion of these biases.

¹² The mean and variance of $V_{t|t}$ can be derived from those of $V_{t|t}(\kappa_t v_t$ and $\kappa_t^2 v_t$, respectively) by using Equation (A.2). The mean and variance of $V_{t+1|t+1}$ are updated from (κ_t, v_t) conditional on the observed asset return by using the algorithm in Equations (13)–(17).

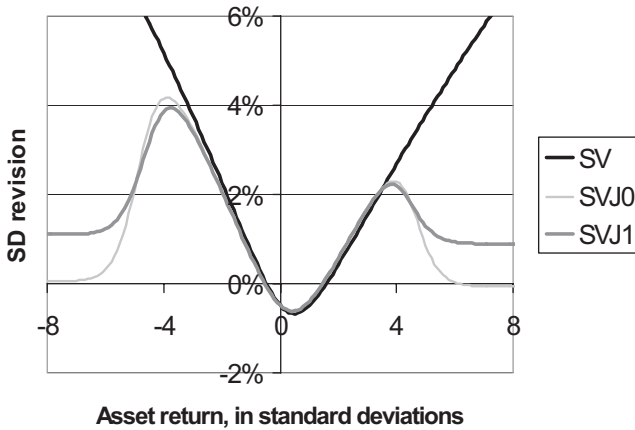


Figure 1

News impact curves for various models

The graph shows the revision in estimated annualized standard deviation $(E_{t+1} - E_t)\sqrt{V_{t+1}}$, conditional upon observing a standardized return of magnitude $y_{t+1}/\sqrt{\hat{V}_{t+1}/252}$.

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

The estimates were calibrated from a median volatility day with a prior volatility estimate of 11.4%, and initial 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. 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 *nonmonotonic* 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).

Figure 2 illustrates 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 100,000-observation sample generated from the SVJ1 model. The filtered estimate tracks latent volatility quite well, with an overall R^2 of 70% over the full sample. Changes in filtered volatility

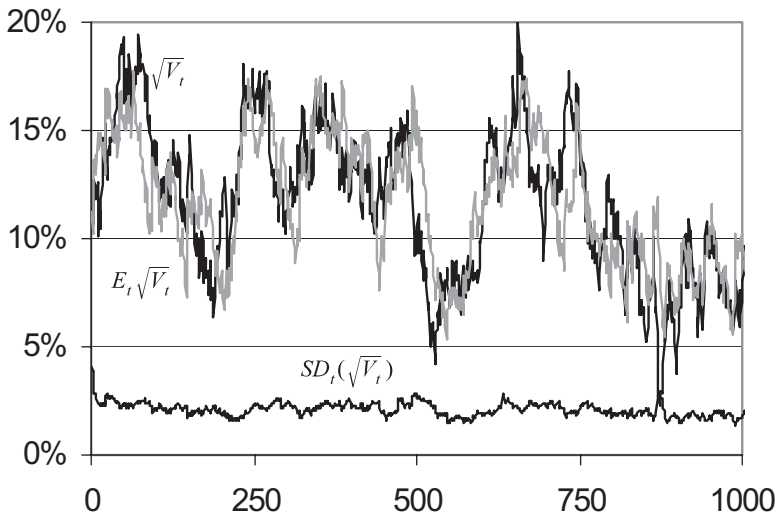


Figure 2
Annualized latent volatility, and its filtered estimate and standard deviation: SVJ1 model

performance lag behind changes in the true volatility, since the filtered estimate must be inferred from past returns. The absolute divergence was usually less than 5% (roughly two standard deviations) but was occasionally larger. To put this error in perspective, the volatility estimate in mid-sample of 15% when the true volatility was 10% represents a substantial error when pricing short-maturity options. The magnitude of this error reflects the low informational content of daily returns for estimating latent volatility and variance.

2.2.1 Filtration diagnostics. A key issue is whether force-fitting posterior distributions into a gamma distribution each period reduces the quality of the filtrations. Two sets of diagnostics were used on simulated data. First, the accuracy and informational efficiency of the variance filtration was compared with various GARCH approaches and with Gallant and Tauchen's (2002) reprojection technique. Second, the extent of specification error in conditional distributions was assessed using higher-moment and quantile diagnostics. The diagnostics were run on two 200,000-day (794-year) data sets simulated from the SV and SVJ1 processes, respectively. The first 100,000 days were used for in-sample estimation and testing, while the subsequent 100,000 days were used for out-of-sample tests.

Three GARCH models were estimated on the SV simulated data: the standard GARCH(1,1) and EGARCH specifications and Hentschel's

(1995) generalization (labeled HGARCH) that nests these and other ARCH specifications. Filtration performance was assessed based on how well the filtered volatility and variance estimates tracked the true latent values, as measured by overall R^2 .

As shown in Table 5, the approximate maximum likelihood SV filtration outperforms all three ARCH models, when either the true SV parameters or the parameters estimated from the full sample are used. However, the EGARCH and HGARCH specifications that take into account the correlations between asset returns and volatility shocks perform almost as well as the SV model when estimating latent volatility and variance from past returns.

For the SVJ1 data with jumps, the ARCH models were modified to t -ARCH specifications to capture the conditionally fat-tailed property of returns. Despite the modification, the t -GARCH and t -EGARCH filtrations performed abysmally, while even the t -HGARCH specification substantially underperformed the SVJ1 filtration. The problem is, of course, the jumps. As illustrated in Figure 1, the optimal filtration under the postulated stochastic volatility/jump process is essentially to ignore large outliers. The GARCH failure to do this can generate extremely high volatility estimates severely at odds with true latent volatility. For instance, the simulated data happened to include a crash-like -21% return. The (in-sample) annualized volatility estimates of 108%, 159%, and 32% from the t -GARCH, t -EGARCH and t -HGARCH models for the day immediately after the outlier greatly exceeded the true value of 10%.

Table 5
Volatility and variance filtration performance under various approaches

	R^2 on SV data			R^2 on SVJ1 data	
Filtration method	Volatility	Variance	Filtration method	Volatility	Variance
In-sample fits					
GARCH	.598	.553	t -GARCH	−.008	−2.222
EGARCH	.674	.649	t -EGARCH	.356	−2.184
HGARCH	.678	.649	t -HGARCH	.585	.453
SV (θ)	.703	.690	SVJ1 (θ)	.723	.700
SV ($\hat{\theta}$)	.704	.692	SVJ1 ($\hat{\theta}$)	.726	.702
SV ($\hat{\theta}$) + reprojection		.695	SVJ1($\hat{\theta}$) + reprojection		.705
Out-of-sample fits					
HGARCH	.678	.650	t -HGARCH	.603	.430
SV (θ)	.699	.687	SVJ1 (θ)	.747	.727
SV ($\hat{\theta}$) + reprojection		.689	SVJ1($\hat{\theta}$) + reprojection		.727

Criterion : $R^2 \equiv 1 - \frac{\sum (x_t - \hat{x}_{t|t-1})^2}{\sum (x_t - \bar{x})^2}$, for $x_t = \sqrt{V_t}$ (volatility) or $x_t = V_t$ (variance).

Two samples of 200,000 daily observations (794 years) were generated from the SV and SVJ1 processes, respectively. In-sample fits use the first 100,000 observations for parameter estimation and filtration. Out-of-sample fits use the subsequent 100,000 observations for filtration.

The informational efficiency of the algorithm's filtrations was assessed using a reprojection technique based on that in Gallant and Tauchen (2002). The true variance V_t was regressed on current and lagged values of the filtered variance estimates $\hat{V}_{t|t}$, as well as on lags of returns and absolute returns and a constant.¹³ Comparing the resulting R^2 s with those of the AML algorithm tests jointly for forecasting biases and for any information not picked up by the contemporaneous filtered estimate $\hat{V}_{t|t}$.¹⁴ The lag length was set equal to $4.32 \ln 2/\beta$ (4.32 half-lives of variance shocks), which implies from Equation (22) that less than 5% of the potential information in the omitted data from longer lags is still relevant for variance forecasting. Given β estimates, this generated 128- and 178-day lag lengths of the three dependent variables for the SV and SVJ1 models respectively. All regressions were run in RATS.

The reprojection results in Table 5 indicate very little information is picked up by the additional regressors. In-sample, the R^2 s of the variance estimates increase only from 69.0% to 69.5% for the SV variances generated from the SV process, and from 70.0% to 70.6% for the SVJ1 variances. In the out-of-sample tests, there is virtually no improvement in forecasting ability. The in-sample increase in R^2 are of course statistically significant, given almost 100,000 observations. Nevertheless, the virtually identical R^2 performance suggests little deterioration in latent variance estimation from approximating prior distributions by an gamma distribution.

2.2.2 Diagnostics of conditional distributions. Two additional diagnostics were used to assess how well gamma approximations captured the overall conditional distributions of latent variance. The first diagnostic was based upon computing higher conditional moments of posterior distributions. The algorithm uses only the posterior mean and variance of the latent variable as inputs to next period's prior distribution. However, the posterior skewness and excess kurtosis can be computed by similar numerical integration methods and compared with the gamma-based values of $2/\sqrt{\nu_{t+1}}$ and $6/\nu_{t+1}$, respectively.

The results reported in Table 6 indicate the posterior distributions of latent variance from the SV model are invariably positively skewed and leptokurtic. The divergences in moments from the gamma moments are positive but near zero on average, and with a small standard deviation. Overall, it would appear that the gamma posterior approximation generally does a good job. However, the min/max ranges for moment

¹³ Gallant and Tauchen (2002) use lagged variance estimates from their SNP-GARCH approach as regressors.

¹⁴ Testing the improvement in R^2 via an F -test is statistically equivalent to examining whether the regressors have any explanatory power for the residuals $V_t - \hat{V}_{t|t}$.

Table 6
Higher Conditional Moments of Latent Variance

	Average	Standard deviation	Min	Max	Percentage less than 0
Conditional moment estimates					
SV: $SK\acute{E}W$	.87	.25	.34	2.75	0%
$XK\grave{U}RT$	1.23	.80	.18	11.62	0%
SVJ1: $SK\acute{E}W$	.80	.23	-.19	2.56	0.02%
$XK\grave{U}RT$	1.05	.68	-.79	10.08	0.06%
Divergence from gamma-based estimates					
SV: $SK\acute{E}W - SK\acute{E}W_{\text{gamma}}$	.03	.08	-.11	.71	48.9%
$XK\grave{U}RT - XK\grave{U}RT_{\text{gamma}}$	.10	.27	-.36	5.06	48.7%
SVJ1: $SK\acute{E}W - SK\acute{E}W_{\text{gamma}}$	.02	.07	-.91	.64	50.5%
$XK\grave{U}RT - XK\grave{U}RT_{\text{gamma}}$	.07	.23	-2.58	3.99	50.0%

Summary statistics of the conditional moment estimates, and of the divergence from gamma-based estimates, based on 100,000 observations of simulated data from the SV and SVJ1 models, respectively.

divergences indicate there are specific days on which a more flexible specification might be preferable.

The posterior moments for the SVJ1 model are also generally close to the gamma moments. Again, however, there are specific days in which a more flexible specification is desirable. In particular, the posterior distribution can occasionally be negatively skewed and platykurtic. This reflects the fact that the posterior distributions for latent variance from the SVJ1 model are mixtures of distributions, the posterior distributions conditional upon n jumps occurring weighted by the posterior probabilities of n jumps. For days with substantial ambiguity ex post regarding whether a jump has or has not occurred, the posterior distribution for latent variance can be multimodal and platykurtic.

It should be emphasized that all of the above posterior moment computations involve updating conditional upon a gamma prior distribution. As such, they provide a strictly *local* diagnostic of whether a more flexible class of distributional approximations would better capture posterior distributions at any single point in time.

A second diagnostic was used to assess the overall performance of the approximate conditional distributions: the frequency with which simulated V_t realizations fell within the quantiles of the conditional $V_{t|t}$ gamma distributions. The realized frequencies over runs of 100,000 observations indicate the gamma conditional distributions do, on average, capture the conditional distribution of V_t realizations quite accurately:

Quantile p :	.010	.050	.100	.250	.500	.750	.900	.950	.990
SV	.008	.042	.089	.232	.478	.734	.891	.944	.987
SVJ1	.008	.042	.086	.227	.469	.722	.879	.935	.983

The above results were for filtrations using the true parameter vector θ . The results using in-sample estimated $\hat{\theta}$ were identical.

In summary, the approximate maximum likelihood methodology performs well on simulated data. Parameter estimation is about as efficient as the MCMC approach for two processes for which we have benchmarks: the discrete-time log variance process, and the continuous time stochastic volatility/jump process SVJ0. Volatility and variance filtrations are more accurate than GARCH approaches, especially for processes with jumps, while the filtration error is virtually unpredictable by EMM-style reprojection. Finally, the mean- and variance-matching gamma conditional distributions assess the quantiles of variance realizations quite well on average. However, there are individual days for which matching higher moments better would be desirable.

3. Estimates from Stock Index Returns

For estimates on observed stock returns, I use 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. I will not repeat the data description in that article, but two comments are in order. First, Andersen et al. prefilter the data to remove an MA(1) component that may be attributable to nonsynchronous trading in the underlying stocks. Second, there were three substantial negative outliers: 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% minicrash on October 13, 1989.

The first three columns of Table 7 present estimates of the stochastic volatility model without jumps (SV) from Chernov et al., Andersen et al., and the AML methodology of this article.¹⁵ 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 suggests lower volatility of volatility and slower mean reversion. In CGGT's estimates, the former effect dominates; in ABL's estimates, the latter dominates.

AML 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

¹⁵ The ABL estimates are from their Table IV, converted to an annualized basis.

Table 7
Model estimates and comparison with EMM-based results

	SV			SVJ0, $\lambda_1 = 0$			SVJ1, $\lambda_1 \neq 0$	
	CGGT	ABL	AML	CGGT	ABL	AML	ABL	AML
μ_0		.051 (.032)	.026 (.025)		.037 (.045)	.028 (.027)	.037 (.095)	.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)
α	1.283	.051 (.010)	.093 (.011)	.044	.047 (.013)	.063 (.009)	.047 (.017)	.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)
σ	1.024 (.030)	.197 (.018)	.315 (.018)	.207 (.02)	.184 (.019)	.244 (.016)	.184 (.019)	.237 (.015)
ρ	-.199 (.000)	-.597 (.045)	-.579 (.031)	-.483 (.10)	-.620 (.067)	-.612 (.031)	-.620 (.086)	-.611 (.031)
$\sqrt{\alpha/\beta}$	.096	.114	.125 (.004)	.125	.113	.120 (.004)	.113	.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)
λ_0				1.70	5.09 (.43)	.744 (.217)	5.09 (7.18)	.000 (.000)
λ_1							.70 (488.0)	93.4 (33.4)
$\bar{\gamma}$				-.030 (.002)		-.010 (.010)		-.002 (.006)
δ				.008 (.001)	.012 (.001)	.052 (.009)	.012 (.001)	.039 (.008)
ln L		39,192.45 ^a	39,233.87		39,238.03 ^a	39,294.79	39,238.03 ^a	39,309.51

$$dS_t/S_t = [\mu_0 + \mu_1 V_t - (\lambda_0 + \lambda_1 V_t)\bar{k}]dt + \sqrt{V_t}(\rho dW_{1t} + \sqrt{1 - \rho^2}dW_{2t}) + (e^{\gamma_t} - 1)dN_t$$
$$dV_t = (\alpha - \beta V_t)dt + \sigma\sqrt{V_t}dW_{1t}$$
$$Prob[dN_t = 1] = (\lambda_0 + \lambda_1 V_t)dt, \bar{\gamma}_s \sim N[\bar{\gamma}, \delta^2]$$

CGGT: Chernov et al. (2003) EMM-based estimates on daily DJIA returns, 1953–July 16, 1999 (11,717 obs.)
ABL: Andersen et al. (2002) EMM-based estimates on daily S&P 500 returns, 1953–1996 (11,076 obs.)
AML: Approximate maximum likelihood estimates of this article, using the ABL data. All parameters are in annualized units except the variance shock half-life $HL \equiv 12 \ln 2/\beta$, which is in months.
^aABL log likelihoods were evaluated at the ABL parameter estimates using the AML methodology.

39,049—a strong rejection of the constraint with an associated P -value less than 10^{-16} . Although the CGGT data include a few outliers in 1997–1999 that are not in the ABL data used here, the likelihood impact per outlier of a larger $\hat{\sigma}$ seems insufficient to explain the difference in results.¹⁶

¹⁶ It 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.

Although the AML stochastic volatility parameter estimates are qualitatively similar to those of ABL on the same data set, there are statistically significant 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 former divergence is especially significant statistically, given a standard error of only .018.¹⁷ 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 10^{-15} . 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 approximate 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 7, 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 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) shows that a volatility-dependent jump intensity component $\lambda_1 V_t$ helps explain the cross section of stock index option prices. Some weak time series evidence for the specification is provided in Bates and Craine (1999), while Eraker et al. (2003) find stronger empirical support. In contrast to the results in ABL, the final column of Table 7 indicates 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. The Monte Carlo simulations in Table 4 establish that the standard error estimates are reasonable, and that the AML estimation methodology can identify the presence of time-varying jump intensities.

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

$$y_{t+1}^* = N^{-1}[CDF(y_{t+1} | \mathbf{Y}_t, \hat{\theta})] \quad (27)$$

¹⁷ The Monte Carlo RMSE results in Tables 2 and 3 for 12,000-observation data samples indicate that the estimated asymptotic standard errors in Table 7 are reliable in general.

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 y^* s should be independent and identical draws from a normal $N(0, 1)$ distribution.

Figure 3 examines specification accuracy using normal probability plots generated by Matlab, which plot the theoretical quantiles (line) and empirical quantiles (+) against the ordered normalized data y^* . Unsurprisingly, the stochastic volatility model (SV) is unable to match the tail properties of the data; there are far too many extreme outliers. The models with jumps (SVJ0, SVJ1) do substantially better. However, both have problems with the 1987 crash, which is equivalent in probability to a negative draw of more than five standard deviations from a Gaussian distribution. As such moves should be observed only once every 14,000 years, the single 1987 outlier constitutes substantial evidence against both models.

To address this issue, an additional stochastic volatility/jump model SVJ2 was estimated with two separate jump processes, each with V_t -dependent jump intensity.¹⁸ The resulting estimates for the stochastic volatility component are roughly unchanged; the jump parameters become

$$\begin{aligned} \text{Prob}[dN_{1t} = 1] &= 131.1 V_t dt, & \ln(1 + k_1) &\sim N[.001, (.029)^2] \\ \text{Prob}[dN_{2t} = 1] &= 2.4 V_t dt, & \ln(1 + k_2) &\sim N[-.222, (.007)^2]. \end{aligned} \quad (28)$$

The conditional distribution of daily returns is approximately a mixture of three normals: diffusion-based daily volatility that varies stochastically over a range of .2%–1.8%, infrequent symmetric jumps with a larger standard deviation of 2.9% and a 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. The increase in log likelihood has a marginal significance level of .0008 under a likelihood ratio test, given three additional parameters.

The y^* -score for the 1987 crash drops in magnitude to a more plausible value of -3.50 . As shown in Figure 3, the resulting empirical and theoretical quantiles for the SVJ2 model are much more closely aligned. However, there remain small deviations from perfect alignment that indicate some scope for further model improvement.

¹⁸ Multiple-jump processes are equivalent to a single-jump process in which the jump is drawn from a mixture of distributions.

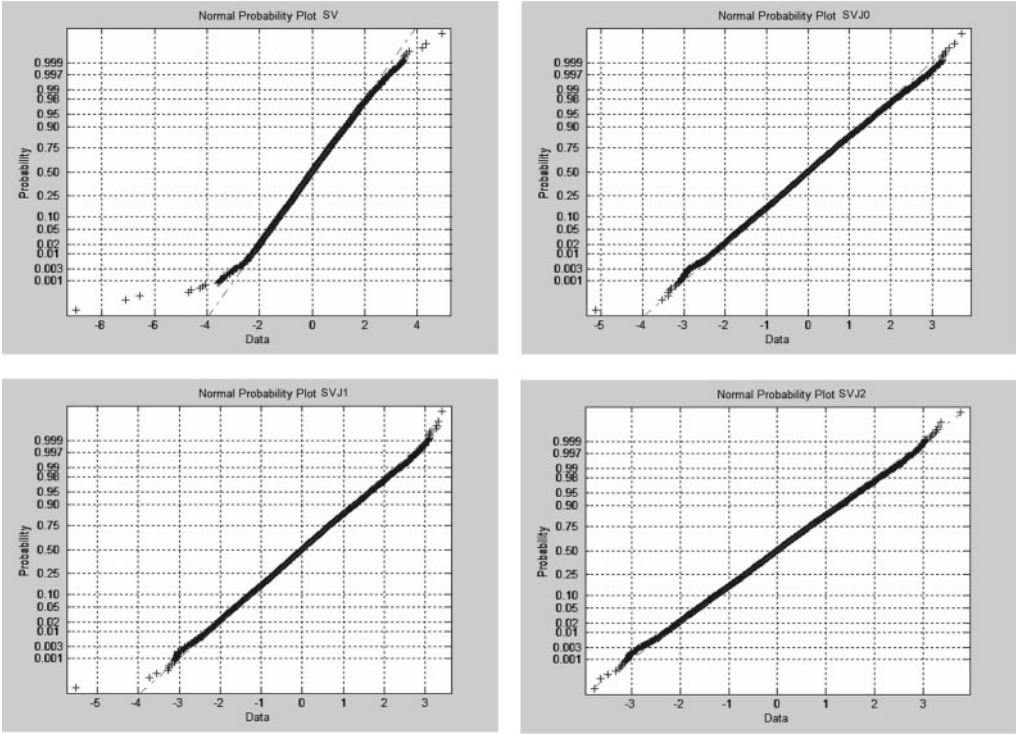


Figure 3
Normal probability plots for the normalized returns $y_{t+1}^* \equiv N^{-1}[\text{CDF}(y_{t+1}|Y_t, \hat{\theta})]$, for different models
 The diagonal line gives the theoretical quantiles conditional upon correct specification; + gives the empirical quantiles.

3.1 Filtration

Figure 4 shows the filtered estimates of the latent state variable $\sqrt{V_t}$ from the SVJ1 model, and the divergence between SVJ1 and SV estimates of diffusive volatility $\sqrt{V_t}$. 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—as one would expect from the news impact curves in Figure 1.

The number of jumps on any given day is also a latent variable that can be inferred from observed returns. It is shown in Appendix A that the joint conditional distribution of log-differenced asset prices and the number of jumps $\Delta N_{t+1} \equiv N_{t+1} - N_t$ has an affine specification. Consequently, the characteristic function $G_{\Delta N}(i\xi | Y_{t+1}) \equiv E[e^{i\xi \Delta N_{t+1}} | y_{t+1}, Y_t]$ can 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: $E[\Delta N_{t+1} | Y_{t+1}] = G'_{\Delta N}(0 | Y_{t+1})$. At the daily horizon, ΔN_{t+1} is essentially binomial, and the estimated number of jumps is approximately the probability that a jump occurred. Unsurprisingly, in Figure 5 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

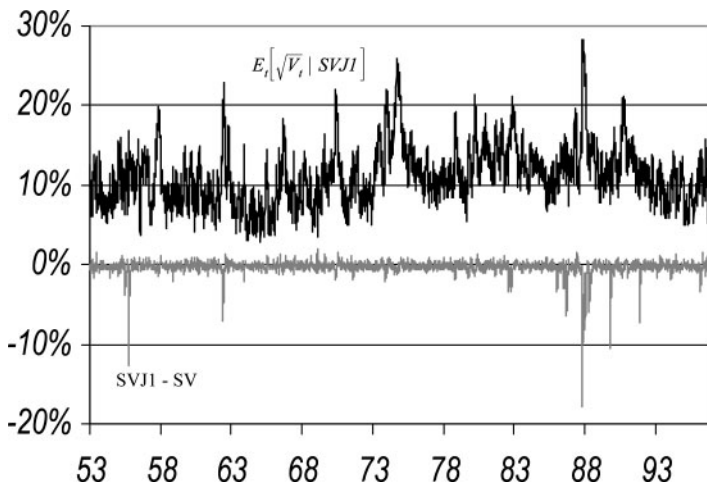


Figure 4
Annualized filtered volatility estimate $E_t\sqrt{V_t}$ from the stochastic volatility/jump model SVJ1, and its divergence from SV volatility estimates.

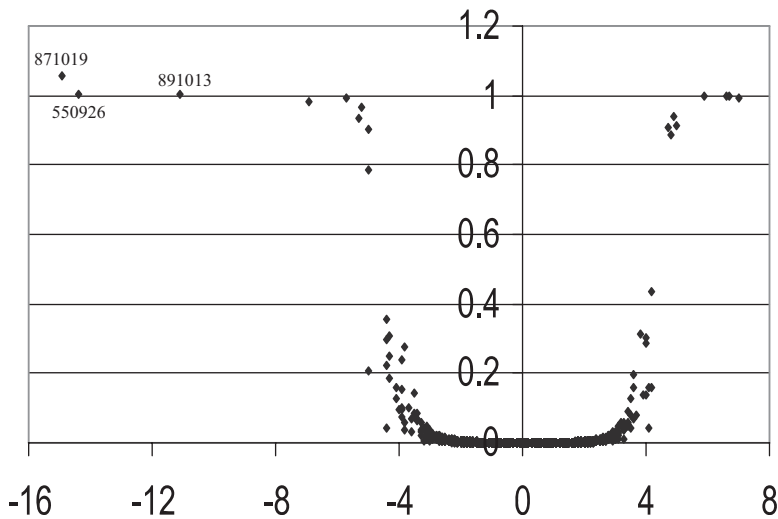


Figure 5
Estimated number of jumps, versus standardized returns $y_{t+1}/\sqrt{\hat{V}_{it}/252}$, SVJ0 model
 The values are approximately the probability for each day that a jump occurred.

frequent intermediate-sized moves that underpin the overall estimate of jump intensities (e.g., .744 jumps per year in the SVJ0 model).

4. Implications for Option Prices

One of the major advantages of using affine processes is their convenience for pricing options. If the log of the pricing kernel is also affine in the state variables, the “risk-neutral” asset price process used when pricing options is also affine, and options can be rapidly priced by Fourier inversion of the associated exponentially affine characteristic function. Under the risk-neutral parameters θ^* , the upper tail probability conditional upon knowing latent variance V_t is

$$P_2 \equiv \text{Prob}^*[\ln(S_{t+T}/S_t) > \ln(X/S_t) | V_t] \\ = \frac{1}{2} + \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\exp[C^*(i\Phi, 0; T) + D^*(i\Phi, 0; T) V_t - i\Phi \ln(X/S_t)]}{i\Phi} d\Phi \quad (29)$$

where $C^*(\bullet)$ and $D^*(\bullet)$ are variants of $C(\bullet)$ and $D(\bullet)$ evaluated at the risk-neutral parameters θ^* and time interval T . The value for a European call option $c(S_t, V_t, T; X)$ with maturity T and strike price X can be evaluated by using $\partial c / \partial X = -e^{-rT} P_2$, substituting Equation (29) for P_2 , and integrating with respect to X :

$$c(S_t, V_t, T; X) = S_t e^{-d_t T} - e^{-rT} X \left[\frac{1}{2} + \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{C^*(i\Phi, 0; T) + D^*(i\Phi, 0; T) V_t - i\Phi \ln(X/S_t)}}{i\Phi(1 - i\Phi)} d\Phi \right] \quad (30)$$

where d_t is the current dividend yield, and the ex-dividend spot price $S_t e^{-d_t T}$ is a constant of integration determined by the value of a call option with zero strike price.¹⁹ Given the affine structure inside the integrand in Equation (30), an econometrician's valuation of a European call option conditional upon observing past returns $\mathbf{Y}_t$ and the current asset price S_t is

$$c(S_t, T; X | \mathbf{Y}_t) = E[c(S_t, V_t, T; X) | \mathbf{Y}_t, S_t] = S_t e^{-d_t T} - e^{-rT} X \left[\frac{1}{2} + \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{C^*(i\Phi, 0; T)} G_{t|t}[D^*(i\Phi, 0; T)] e^{-i\Phi \ln(X/S_t)}}{i\Phi(1 - i\Phi)} d\Phi \right]. \quad (31)$$

The filtration used in estimating the conditional characteristic function $G_{t|t}(\bullet)$ of current latent variance is based on the objective parameter values, not the risk-neutral parameters. The option valuation in Equation (31) is roughly equivalent to using a filtered variance estimate $\hat{V}_{t|t}$ in (30) and adding a Jensen's inequality correction for state uncertainty $P_{t|t}$.

The risk adjustments incorporated into the divergence between the actual parameters θ and the risk-neutral parameters θ^* depend upon the pricing kernel M_t . The general affine pricing kernel specification commonly used in the affine literature takes the form

$$d \ln M_t = \mu_m dt - R d \ln S_t - R_V dV_t - R_J \gamma_s dN_t, \quad (32)$$

where μ_m is such that $E_t(dM_t/M_t) = -r_t dt$. This reduced-form specification nests various approaches. For instance, the submodel with $R_V = R_J = 0$ is the myopic power utility pricing kernel used in the implicit pricing kernel literature and in Coval and Shumway's (2001) empirical examination of the profits from straddle positions. If S_t is viewed as a good proxy for overall wealth, R measures relative risk aversion. More generally, R reflects the projection of pricing kernel innovations upon asset returns and can be greater or less than the coefficient of relative risk aversion.

R_V determines a volatility risk premium *in addition to* the wealth-related effect that volatility innovations tend to move inversely with

¹⁹ This approach is equivalent to but more efficient than the approaches in Bates (1996, 2000) and Bakshi and Madan (2000). Those approaches require two univariate integrations, whereas Equation (30) has only one, with an integrand that falls off more rapidly in Φ . Carr and Madan (2000), Lewis (2001), and Attari (2004) have derived similar formulas by alternate methods.

stock market returns. While theoretically zero under myopic preferences such as log utility, R_V will diverge from zero for representative agents concerned with volatility-related shifts in the investment opportunity set. Bakshi and Kapadia (2003) and Coval and Shumway attribute the overpricing of stock index options to a significantly negative R_V . Similarly, R_J determines the additional risk premium on market stock jumps *beyond* the direct wealth-related effects on marginal utility captured by $Rd\ln S_t$. Nonzero R_J can arise if there are jump-related changes in investment opportunities (volatility-jump models) or in preferences (e.g., habit formation models), or if investors are directly averse to jumps [Bates (2001)] or to jump estimation uncertainty [Liu, Pan, and Wang (2004)]. R_J will also differ from zero if the jump-contingent projection of pricing kernel innovations on asset returns diverges from the diffusion-based projection.

Using this pricing kernel, the objective and risk-neutral processes for *excess* returns for the SVJ2 model will be of the form

Objective measure:

$$\begin{aligned} dS_t/S_t + (d_t - r_t)dt &= \mu_1 V_t dt + \sqrt{V_t} \left(\rho dW_{1t} + \sqrt{1 - \rho^2} dW_{2t} \right) \\ &\quad + \sum_{i=1}^2 [(e^{\gamma_i} - 1) dN_{it} - \lambda_i V_t \bar{k}_i dt] \\ dV_t &= (\alpha - \beta V_t) dt + \sigma \sqrt{V_t} dW_{1t} \end{aligned} \quad (33)$$

Risk-neutral measure:

$$\begin{aligned} dS_t/S_t + (d_t - r_t)dt &= \sqrt{V_t} \left(\rho dW_{1t}^* + \sqrt{1 - \rho^2} dW_{2t}^* \right) \\ &\quad + \sum_{i=1}^2 [(e^{\gamma_i^*} - 1) dN_{it}^* - \lambda_i^* V_t \bar{k}_i^* dt] \\ dV_t &= (\alpha - \beta^* V_t) dt + \sigma \sqrt{V_t} dW_{1t}^* \end{aligned} \quad (34)$$

where

W_{1t} and W_{2t} are independent Wiener processes;

N_{it} are Poisson counters with instantaneous intensities $\lambda_i V_t$ for $i = 1, 2$;

$\gamma_i \sim N(\bar{\gamma}_i, \delta_i^2)$ is the jump size of log asset prices conditional upon a jump occurring;

$\bar{k}_i = \exp(\bar{\gamma}_i + 1/2\delta_i^2) - 1$ is the expected percentage jump size conditional upon a jump;

$(\lambda_i^*, \bar{\gamma}_i^*, \bar{k}_i^*)$ are the corresponding values for the risk-neutral process (34);

W_{it}^* and N_{it}^* are the corresponding risk-neutral Wiener processes and Poisson counters.

If dividend yields and interest rates are nonstochastic, Equation (34) is also the risk-neutral process for futures returns used in pricing futures options.

The affine pricing kernel constrains both objective and risk-neutral parameter values. As discussed in Bates (1991, Appendix 1), the instantaneous equity premium estimated under the objective probability measure is

$$(\mu_0 + \mu_1 V_t + d_t - r_t)dt = -E_t \left[\left(\frac{dS_t}{S_t} \right) \left(\frac{dM_t}{M_t} \right) \right]. \quad (35)$$

For the SVJ2 model, this implies

$$\begin{aligned} \mu_0 &= 0 \\ \mu_1 &= R + R_V \rho \sigma + \sum_{i=1}^2 (\lambda_i^* \bar{k}_i^* - \lambda_i \bar{k}_i), \end{aligned} \quad (36)$$

while the risk-neutral parameter values used in pricing derivatives are

$$\begin{aligned} \lambda_i^* &= \lambda_i E \left[1 + \frac{dM_t}{M_t} | dN_i = 1 \right] \\ &= \lambda_i \exp[-(R + R_J) \bar{\gamma}_i + 1/2 (R + R_J)^2 \delta_i^2] \\ \bar{\gamma}_i^* &= \bar{\gamma}_i + \text{Cov}(\gamma_i, d \ln M_t | dN_i = 1) \\ &= \bar{\gamma}_i - (R + R_J) \delta_i^2 \\ \beta^* &= \beta - E \left[\frac{dV_t}{V_t} \frac{dM_t}{M_t} \right] \\ &= \beta + R \rho \sigma + R_V \sigma^2. \end{aligned} \quad (37)$$

4.1 Estimated option prices

To illustrate the option pricing implications of the AML estimation methodology, I will focus upon the myopic power utility pricing kernel, with $R_V = R_J = 0$. This is done for two reasons. First, the appropriate risk adjustments for pricing options can be estimated solely from past stock index excess returns under the AML methodology presented above, whereas pricing volatility or jump risk premia under other specifications requires incorporating additional information from option prices or option returns.

Second, the power utility pricing kernel is a useful performance measure that has been previously employed for assessing returns on options

[Coval and Shumway (2001)] and on mutual funds [e.g., Cumby and Glen (1990)]. It is also the standard benchmark against which implicit pricing kernels are compared [Rosenberg and Engle (2002) and Bliss and Panagirtzoglou (2004)]. While equivalent to a conditional CAPM criterion when stock market and option returns are approximately conditionally normal, using a power utility pricing kernel is more robust to the substantial departures from normality observed with returns on out-of-the-money options.²⁰ Large deviations of observed option prices from the power utility valuations indicate investment opportunities with an excessively favorable conditional return/risk tradeoff. Of course, it is possible that such investment opportunities can be rationally explained by investors' aversions to time-varying volatility or to jump risk, which would show up in nonzero R_V or R_J .

Since Equation (33) requires excess returns, the ABL data were adjusted by monthly dividend yields and 1-month Treasury bill rates obtained from the Web sites of Robert Shiller and Ken French, respectively. In addition, the ABL data set was extended through 2001 using CRSP data, for use in out-of-sample tests. Since the substantial autocorrelations in daily S&P 500 index returns estimated by Andersen et al. (2002) over 1953–1996 were not apparent over the post-1996 period, post-1996 excess returns were used directly, without prefiltration.

Table 8 contains estimates of the SVJ2 model on returns and excess returns and the corresponding estimates of the risk-neutral parameters under the myopic power utility pricing kernel. Unsurprising, the unconstrained SVJ2 estimates on raw and excess returns in the first two rows of each panel are virtually identical, except for a difference in the mean parameter μ_1 . The dividend and interest rate series are smooth, inducing little change in the estimates of volatility dynamics and jump risk.

Affine pricing kernel models imply that $\mu_0 = 0$ for the SVJ2 model. The estimated sensitivity μ_1 of the equity premium to the current level of variance determines the risk aversion parameter R for the power utility pricing kernel and constrains the risk aversion parameters for other models. In Table 8, the $\mu_0 = 0$ constraint is borderline insignificant, with a marginal significance level of 6.8% under a likelihood ratio test. The risk aversion estimate is approximately 4. The implications for the risk-neutral parameters used in option pricing are twofold. First, the risk-neutral frequency of 1987-like crashes is more than double that of the objective frequency. Second, the risk-neutral volatility dynamics implicit in the term structure of implied volatilities involve somewhat slower mean reversion to a somewhat higher level: a half-life of 2.0 months

²⁰ Goetzmann, Ingersoll, Spiegel, and Welch (2002) discuss problems with the traditional Sharpe ratio performance measure when funds can take positions in options.

Table 8
Estimates of the multi-jump model SVJ2 and associated risk-neutral parameter estimates

	mean parameters			Stochastic volatility parameters				
	<i>R</i>	μ_0	μ_1	α	β	β^*	σ	ρ
Returns		.041 (.024)	2.8 (2.1)	.059 (.007)	4.15 (.57)		.233 (.012)	-.614 (.033)
Excess returns		.045 (.024)	1.7 (2.1)	.060 (.007)	4.22 (.59)		.234 (.012)	-.610 (.033)
Excess returns	3.94 (1.26)	0	4.8 (1.4)	.067 (.004)	4.72 (.56)	4.12 (.53)	.240 (.012)	-.627 (.029)
Jump parameters (<i>i</i> = 1,2)								
	λ_i	λ_i^*	$\tilde{\gamma}_i$	$\tilde{\gamma}_i^*$	δ_i	$\ln L$		
Returns	131.1 (38.8) 2.4 (1.6)		.001 (.004) -.222 (.092)		.029 (.004) .007 (.036)	39,317.81		
Excess Returns	121.1 (36.0) 1.5 (2.8)		.002 (.004) -.217 (.043)		.030 (.004) .005 (.062)	39,314.69		
Excess Returns	121.1 (36.2) 1.6 (2.1)	121.3 (36.4) 3.7 (4.3)	.001 (.004) -.216 (.024)	-.002 (.004) -.216 (.024)	.030 (.004) .003 (.012)	39,313.03		

Data: daily S&P 500 returns and excess returns over 1953–1996. Standard errors are in parentheses.

rather than 1.8 months, and a long-run level of $(12.8\%)^2$ instead of $(11.9\%)^2$.

For comparison, daily settlement prices for the Chicago Mercantile Exchange’s American options on S&P 500 futures and the underlying futures contracts were obtained from the Institute for Financial Markets for the period from the option contracts’ inception on January 28, 1983, through June 29, 2001. The use of end-of-day option prices synchronizes nicely with the AML variance filtration, which uses closing S&P 500 index returns. The shortest-maturity out-of-the-money call and put options with a maturity of at least 14 days were selected. The corresponding implicit standard deviations (ISDs) were computed using the Barone-Adesi and Whaley (1987) American option pricing formula, for comparison with those computed from AML European option price estimates. In addition, a benchmark at-the-money ISD was constructed for each day by interpolation.²¹

Figure 6 shows the observed and estimated volatility smirks for out-of-the-money put and call options on December 31, 1996, which was the last day of the ABL data set. Several results are evident. First, the ISDs

²¹ A minor issue is the difference between American and European options, and the computation of associated ISDs. The price of an American futures option is bounded below by the European price and bounded above by the future value of the European price. The resulting error in annualized European at-the-money ISD estimates is bounded by $\pm ISD_t \times rT$ ($\pm 0.04\%$ on December 31, 1996), while the errors for out-of-the-money ISD estimates are even smaller.

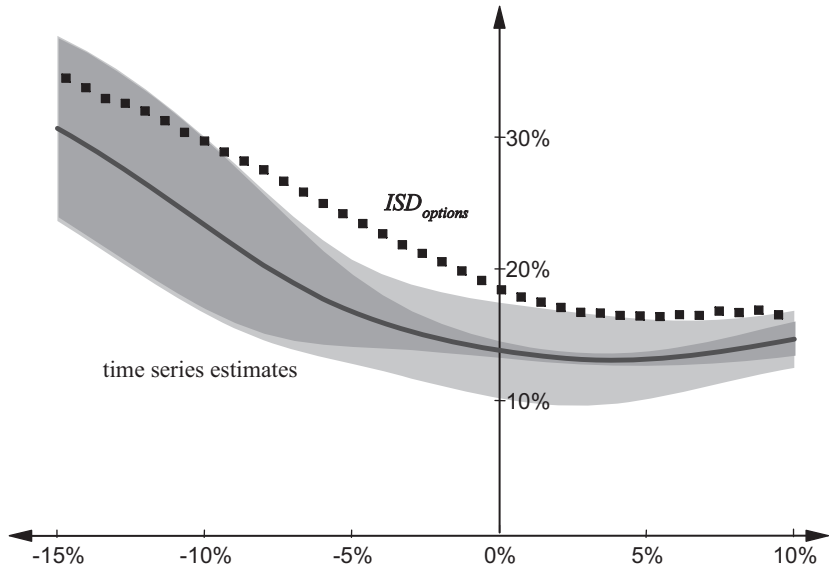


Figure 6

Observed and estimated ISDs for 17-day Jan.1997 S&P 500 futures options on Dec.31, 1996

The dark grey area is the 95% confidence interval given only parameter estimation uncertainty. The light grey area is the 95% confidence interval given both parameter and state uncertainty.

estimated from stock index returns exhibit a substantial volatility smirk. This smirk partly reflects the negative correlation between asset returns and volatility shocks but also reflects the substantial and time-varying crash risk—in particular, the risk-adjusted estimate of infrequent 1987-like crashes.

Second, estimates are imprecise. Option price estimates and the associated ISDs contain two types of error:

parameter estimation error, with variance $\frac{\partial c}{\partial \theta'} \text{Var}(\hat{\theta}) \frac{\partial c}{\partial \theta}$, and

filtration error (or state uncertainty), with variance $\left(\frac{\partial c}{\partial \hat{V}_{t|t}} \right)^2 P_{t|t}$.

The at-the-money estimated ISD contains little parameter uncertainty, reflecting the relative robustness of volatility estimates to model specification. However, there is considerable state uncertainty regarding the current level of V_t and the at-the-money ISD. This state uncertainty reflects the relatively low informational content of daily returns and could be reduced by using more informative data sources. Out-of-the-money put

and call option prices are less affected by state uncertainty but are more affected by parameter uncertainty. This reflects the difficulties in estimating the jump parameters, which are of key importance in determining the probability of these options paying off.

Third, the deviations between estimated and observed ISDs were sometimes large enough to be statistically significant. S&P 500 futures options were mostly overpriced on December 31, 1996, from the perspective of a myopic power utility investor with a risk aversion of 4 who is 100% invested in S&P 500 stocks. This sort of overpricing is responsible for the high Sharpe ratios reported by various authors for option-selling strategies.

Figure 7 shows at-the-money ISDs for short-maturity options on Wednesdays over 1983–2001, the corresponding AML filtered ISD estimates, and the divergence between the two. The estimates prior to 1997 are “in-sample,” in that the parameter estimates underlying the filtration are from the full 1953–1996 data set. The post-1996 filtered estimates are out-of-sample.

Overall, the filtered estimates track the at-the-money option ISDs reasonably well. There are, however, some interesting and persistent

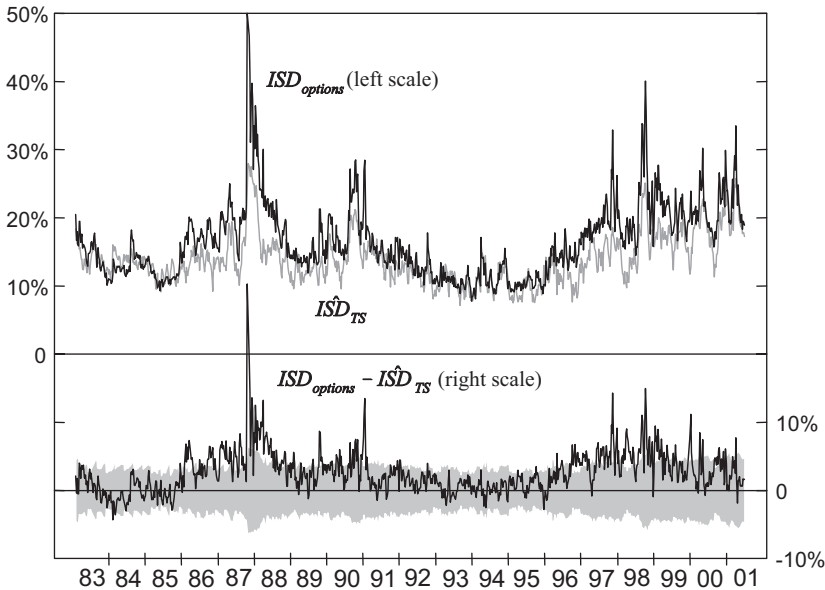


Figure 7
Observed and estimated ISD's for at-the-money S&P 500 futures options; Wednesday data
 Estimated ISD's are based on SVJ2 parameter estimates from 1953–1996, and on filtered variance estimates from daily returns. The grey area is the 95% confidence interval given both parameter and state uncertainty.

deviations between the two series. For instance, at-the-money options were substantially overpriced during the two years *preceding* the 1987 crash, when judged against the time series estimates of appropriate risk-adjusted value. The extent of the overpricing soared after the crash but gradually dwindled over 1988–1991, only to reemerge in late 1996 and in subsequent years. The mispricing was especially pronounced following the minicrash of October 1997, and in the fall of 1998. Overall, option overpricing was most pronounced when observed and estimated ISDs were both relatively high—a result mirroring standard results from regressing realized volatility on ISDs.

Over 1988–1996, ISDs averaged 2.2% higher than the time series valuations, while the average gap was 4.6% over 1997–2001. It is possible that this greater gap reflects out-of-sample instabilities in the time series model. However, it seems more plausible to attribute it to structural shifts in the stock index options market—in particular, to the bankruptcy of Long Term Capital Management in August 1998. By 2000, estimated ISDs were again tracking observed ISDs fairly closely.

5. Conclusions and Extensions

This article has presented a new approximate maximum likelihood methodology for estimating continuous-time affine processes on discrete-time data: both parameter values and latent variable realizations. Results on simulated data indicate the parameter estimation efficiency is excellent for those processes for which we have performance benchmarks. Furthermore, the approach directly generates a filtration algorithm for estimating latent variable realizations, of value for risk management and derivatives pricing. Most other approaches must append an additional filtration procedure to the parameter estimation methodology.

The AML approach was used to estimate the parameters of an affine stochastic volatility/jump process, using the data set of Andersen, Benzoni, and Lund (2002). Parameter estimates were similar to the EMM-based estimates of Andersen et al., but differ in statistically significant fashions. In particular, I find a generally higher volatility of volatility, a more substantial jump component, and strong support for the hypothesis that jumps are more likely when volatility is high. Furthermore, Monte Carlo simulations establish that the AML estimation methodology can reliably identify these phenomena.

The AML methodology differs substantially from the SNP/EMM methodology used by Andersen et al., so the source of divergence is not immediately apparent. It does appear that the EMM methodology may be 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. (2003) find very different estimates from Andersen et al. for the stochastic volatility model, for a similar data set but a different auxiliary model. A Monte Carlo examination of whether the EMM estimation procedure is indeed robust to stochastic volatility/jump processes would appear desirable.

This article has focused on classical maximum likelihood estimation. However, the recursive likelihood evaluation methodology presented here can equally be used in Bayesian estimation, 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, for example, high-low ranges, or “realized” intradaily variance. 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 alternative data sources in conjunction with multifactor models of latent variance will be explored in future research.

Appendix A

A.1 Conditional moment dynamics

For fully affine stochastic processes, the conditional joint cumulant generating function is

$$\begin{aligned}\ln F(\Phi, \psi | Y_t) &\equiv \ln \hat{E}[e^{\Phi y_{t+1} + \psi x_{t+1}} | Y_t] \\ &= C^0(\Phi, \psi) + C^y(\Phi, \psi) y_t + g_{t|t}[D(\Phi, \psi)]\end{aligned}\quad (\text{A.1})$$

where $g_{t|t}(\psi) \equiv \ln \hat{E}[e^{\psi x_t} | Y_t]$ is the cumulant generating function of latent x_t conditional upon using the approximate maximum likelihood filtration and observing data Y_t . Conditional means and variances can be evaluated by taking derivatives of the joint cumulant generating function:

$$\begin{aligned}\hat{E}_t y_{t+1} &= C_\Phi^0 + C_\Phi^y y_t + g'_{t|t}(0) D_\Phi \\ &= C_\Phi^0 + C_\Phi^y y_t + \hat{x}_{t|t} D_\Phi \\ \hat{x}_{t+1|t} &= C_\psi^0 + C_\psi^y y_t + \hat{x}_{t|t} D_\psi \\ P_{t+1|t} &= C_{\psi\psi}^0 + C_{\psi\psi}^y y_t + g'_{t|t} D_{\psi\psi} + g''_{t|t} D_\psi^2 \\ &= C_{\psi\psi}^0 + C_{\psi\psi}^y y_t + \hat{x}_{t|t} D_{\psi\psi} + P_{t|t} D_\psi^2\end{aligned}\quad (\text{A.2})$$

where all derivatives of C^0 , C^y and D are evaluated at $\Phi = \psi = 0$. The first and second derivatives of $g_{t|t}$, evaluated at $D(0,0) = 0$, give the time- t conditional mean and variance of x_t .

Conditional moment dynamics can be evaluated by writing each noncentral moment as the sum of the prior expectation and the revision in expectations in light of new data. For conditional means this takes the form

$$\begin{pmatrix} y_{t+1} \\ \hat{x}_{t+1|t+1} \end{pmatrix} = \begin{pmatrix} C_\Phi^0 \\ C_\psi^0 \end{pmatrix} + \begin{pmatrix} C_\Phi^y & D_\Phi \\ C_\psi^y & D_\psi \end{pmatrix} \begin{pmatrix} y_t \\ \hat{x}_{t|t} \end{pmatrix} + \begin{pmatrix} u_{t+1} \\ v_{t+1} \end{pmatrix}\quad (\text{A.3})$$

where $u_{t+1} \equiv y_{t+1} - \hat{E}_t y_{t+1}$ and $v_{t+1} \equiv \hat{x}_{t+1|t+1} - \hat{x}_{t+1|t}$.

Using the same approach for revising the conditional second noncentral moment of the latent state variable yields

$$(P_{t+1|t+1} + \hat{x}_{t+1|t+1}^2) = (P_{t+1|t} + \hat{x}_{t+1|t}^2) + (\hat{E}_{t+1} - \hat{E}_t)x_{t+1}^2. \quad (\text{A.4})$$

Substituting in the expressions for $P_{t+1|t}$ and $\hat{x}_{t+1|t}$ from Equation (A.2) yields the dynamics of conditional variance:

$$P_{t+1|t+1} = C_{\psi\psi}^0 + C_{\psi\psi}^y y_t + D_{\psi\psi} \hat{x}_{t|t} + D_{\psi}^2 P_{t|t} + (w_{t+1} - u_{t+1}^2), \quad (\text{A.5})$$

where $w_{t+1} \equiv (\hat{E}_{t+1} - \hat{E}_t)x_{t+1}^2 - 2v_{t+1}(C_{\psi}^0 + C_{\psi}^y y_t + D_{\psi} \hat{x}_{t|t})$. This, along with Equation (A.3), yields the vector autoregression for conditional moments that is given in Equation (22).

The dynamics of the higher noncentral conditional moments of x_{t+1} and y_{t+1} can be evaluated similarly, using derivatives of the moment generating function

$$F(\Phi, \psi | Y_t) = e^{C^0(\Phi, \psi) + C^y(\Phi, \psi)y_t} G_{t|t}[D(\Phi, \psi)] \quad (\text{A.6})$$

to evaluate $\hat{E}_t x_{t+1}^m$ and $\hat{E}_t y_{t+1}^m$ for arbitrary integer m . The m th derivative of $G_{t|t}$, evaluated at $D(0,0) = 0$, is the noncentral conditional moment $\hat{E}_t x_{t+1}^m$.

A.2 Joint moment generating functions from continuous-time affine processes

Let

$$F(\Phi, \psi, \xi | s_t, V_t, N_t, t, T) \equiv E_t[e^{\Phi s_T + \psi V_T + \xi N_T} | s_t, V_t, N_t] \quad (\text{A.7})$$

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 a Poisson counter. Since F is a conditional expectation, it is a martingale:

$$E_t[dF(\bullet | s_t, V_t, N_t, t, T)] = 0. \quad (\text{A.8})$$

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, and depends on the time gap $\tau \equiv T - t$ between observations:

$$F(\Phi, \psi, \xi | s_t, V_t, N_t; \tau) = \exp[\Phi s_t + \xi N_t + C(\Phi, \psi, \xi; \tau) + D(\Phi, \psi, \xi; \tau) V_t]. \quad (\text{A.9})$$

By Itô's lemma, the stochastic volatility/jump-diffusion in Equation (25) implies a log asset price evolution of the form

$$\begin{aligned} ds_t &= [\mu_0 + (\mu_1 - 1/2)V_t - (\lambda_0 + \lambda_1 V_t)\bar{k}]dt + \sqrt{V_t}(\rho dW_{1t} + \sqrt{1 - \rho^2}dW_{2t}) + \gamma_s dN_t \\ dV_t &= (\alpha - \beta V_t)dt + \sigma\sqrt{V_t}dW_{1t} \end{aligned} \quad (\text{A.10})$$

where (W_{1t}, W_{2t}) are independent Brownian motions, N_t is a Poisson counter with intensity $\lambda_0 + \lambda_1 V_t$, γ_s 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

$$\begin{aligned}
 F_\tau = & [\mu_0 + (\mu_1 - 1/2)V_t - (\lambda_0 + \lambda_1 V_t)\bar{k}]F_s + (\alpha - \beta V_t)F_V \\
 & + 1/2 V_t (F_{ss} + 2\rho\sigma F_{sV} + \sigma^2 F_{VV}) \\
 & + (\lambda_0 + \lambda_1 V_t)E[F(\bullet|s_t + \gamma_s, N_t + 1, V_t; \tau) - F]
 \end{aligned} \tag{A.11}$$

which is solved subject to the boundary condition

$$F(\Phi, \psi, \xi|s_t, V_t, N_t; \tau = 0) = e^{\Phi s_t + \psi V_t + \xi N_t}. \tag{A.12}$$

Plugging Equation (A.9) into Equation (A.11) yields a recursive system of ordinary differential equations in τ that $C(\bullet; \tau)$ and $D(\bullet; \tau)$ must solve, subject to the boundary conditions $C(\bullet; 0) = 0$ and $D(\bullet; 0) = \psi$. The resulting solutions are:

$$\begin{aligned}
 C(\Phi, \psi, \xi; \tau) = & \mu_0 \Phi \tau - \frac{\alpha \tau}{\sigma^2} [\rho \sigma \Phi - \beta - \gamma(\Phi, \xi)] + \lambda_0 \tau E(\Phi, \xi) \\
 & - \frac{2\alpha}{\sigma^2} \ln \left[1 + [\rho \sigma \Phi - \beta - \gamma(\Phi, \xi)] \frac{1 - e^{\gamma(\Phi, \xi)\tau}}{2\gamma(\Phi, \xi)} \right] - \frac{2\alpha}{\sigma^2} \ln[1 - K(\Phi, \xi; \tau)\psi]
 \end{aligned} \tag{A.13}$$

$$\begin{aligned}
 D(\Phi, \psi, \xi; \tau) = & 2 \frac{1/2 \Phi^2 + (\mu_1 - 1/2)\Phi + \lambda_1 E(\Phi, \xi)}{\gamma(\Phi, \xi) \frac{e^{\gamma(\Phi, \xi)\tau} + 1}{e^{\gamma(\Phi, \xi)\tau} - 1} + \beta - \rho \sigma \Phi \\
 & + \frac{\Lambda(\Phi, \xi; \tau)\psi}{1 - K(\Phi, \xi; \tau)\psi}
 \end{aligned} \tag{A.14}$$

where

$$E(\Phi, \xi) = e^{\xi + \bar{\gamma}\Phi + 1/2\delta^2\Phi^2} - (1 + \bar{k}\Phi) \tag{A.15}$$

$$\gamma(\Phi, \xi) = \sqrt{(\rho \sigma \Phi - \beta)^2 - 2\sigma^2[1/2\Phi^2 + (\mu_1 - 1/2)\Phi + \lambda_1 \gamma(\Phi, \xi)]} \tag{A.16}$$

$$K(\Phi, \xi; \tau) = \frac{\sigma^2}{\gamma(\Phi, \xi) \frac{e^{\gamma(\Phi, \xi)\tau} + 1}{e^{\gamma(\Phi, \xi)\tau} - 1} + \beta - \rho \sigma \Phi} \tag{A.17}$$

$$\Lambda(\Phi, \xi; \tau) = \frac{\left(\frac{e^{\gamma(\Phi, \xi)\tau} + 1}{e^{\gamma(\Phi, \xi)\tau} - 1} \right)^2 - 1}{\left(\frac{e^{\gamma(\Phi, \xi)\tau} + 1}{e^{\gamma(\Phi, \xi)\tau} - 1} + \frac{\beta - \rho \sigma \Phi}{\gamma(\Phi, \xi)} \right)^2}. \tag{A.18}$$

Equations (A.13)–(A.18) are identical to those in Appendix D of Pan (2002) when $\xi = 0$, but are written in a form that makes it easy to take analytic derivatives with respect to ψ . In the two-jump SVJ2 model of Equation (28), $\lambda_0 = 0$ and the $\lambda_1 E(\Phi, \xi)$ terms are replaced by $\sum_{i=1}^2 \lambda_i E_i(\Phi, \xi)$, where $E_i(\Phi, \xi) = \exp(\xi + \bar{\gamma}_i \Phi + 1/2\Phi^2 \delta_i^2) - 1 - \Phi[\exp(\bar{\gamma}_i + 1/2\delta_i^2) - 1]$.

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. In particular, the marginal transition density of V_T has the cumulant generating function of a noncentral chi-squared random variable:

$$\ln E[e^{\psi V_T} | V_t] = -\frac{2\alpha}{\sigma^2} \ln[1 - K(0, 0; \tau)\psi] + \frac{\Lambda(0, 0; \tau)\psi}{1 - K(0, 0; \tau)\psi} V_t \tag{A.19}$$

where $\Lambda(0, 0; \tau) = e^{-\beta\tau}$, $K(0, 0; \tau) = \sigma^2(1 - e^{-\beta\tau})/2\beta$, and $\tau = T - t$. The unconditional cumulant generating function is the limit as $\tau \rightarrow \infty$,

$$\ln G_{0|0}(\psi) \equiv \ln E[e^{\psi V_T}] = -\frac{2\alpha}{\sigma^2} \ln \left(1 - \frac{\sigma^2}{2\beta} \psi \right) \quad (\text{A.20})$$

which is that of a gamma random variable.

The empirical applications assume the data are spaced at a regular daily time interval. Using t (with a slight abuse of notation) to index observations, the joint moment generating function of discrete-time log-differenced prices $y_{t+1} \equiv \Delta s_{t+1} = \ln(S_{t+1}/S_t)$, number of jumps $\Delta N_{t+1} \equiv N_{t+1} - N_t$, and future variance V_{t+1} conditional upon observing V_t can be computed from Equation (A.9):

$$F_{y,V,\Delta N}(\Phi, \psi, \xi | V_t, \tau) = \exp[C(\Phi, \psi, \xi; \tau) + D(\Phi, \psi, \xi; \tau) V_t], \quad (\text{A.21})$$

where $\tau = 1/252$ is the fixed time interval between observations, in years. 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 transform of the joint transition density in Equation (1) is

$$F_{y,V}(i\Phi, i\psi | V_t) = F_{y,V,\Delta N}(i\Phi, i\psi, 0 | V_t, \tau). \quad (\text{A.22})$$

As discussed above in Equation (13), iterated expectations can be used to compute the joint characteristic function of (y_{t+1}, V_{t+1}) conditional upon data Y_t observed through time t :

$$\begin{aligned} F_{y,V}(i\Phi, i\psi | Y_t) &= E[F_{y,V}(i\Phi, i\psi | V_t) | Y_t] \\ &= e^{C(i\Phi, i\psi, 0; \tau)} G_{t|t}[D(i\Phi, i\psi, 0; \tau)]. \end{aligned} \quad (\text{A.23})$$

$G_{t|t}(\psi) \equiv E[e^{\psi V_t} | Y_t]$ is the conditional moment generating function of V_t . It is approximated by the gamma moment generating function $\hat{G}_{t|t}(\psi) = (1 - \kappa_t \psi)^{-v_t}$, by appropriate choice of (κ_t, v_t) .

A.3 Filtration

Density evaluation and variance updating involves three numerical integrations at each date $t+1$, to evaluate the prior density of the asset return y_{t+1} and the posterior mean and variance of the latent variable V_{t+1} . The density is evaluated by Fourier inversion of Equation (A.23):

$$p(y_{t+1} | Y_t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F_{y,V}(i\Phi, 0 | Y_t) e^{-i\Phi y_{t+1}} d\Phi. \quad (\text{A.24})$$

The noncentral moments $E_{t+1}[V_{t+1}^n]$ are evaluated as in Equations (16) and (17) by taking analytic derivatives of Equation (15) with respect to ψ .²² Defining

$$f(i\Phi, \psi) \equiv \ln F_{y,V}(i\Phi, \psi | Y_t) = C(i\Phi, \psi, 0; \tau) + \ln G_{t|t}[D(i\Phi, \psi, 0; \tau)] \quad (\text{A.25})$$

for $\ln G_{t|t}(D) = -v_t \ln(1 - \kappa_t D)$, the first two posterior noncentral moments are

$$\begin{aligned} E_{t+1}(V_{t+1}) &= \left. \frac{\partial G_{t+1|t+1}(\psi)}{\partial \psi} \right|_{\psi=0} \\ &= \frac{1}{2\pi p(y_{t+1} | Y_t)} \int_{-\infty}^{\infty} f_{\psi}(i\Phi, 0) e^{f(i\Phi, 0) - i\Phi y_{t+1}} d\Phi \end{aligned} \quad (\text{A.26})$$

²² Numerical derivatives are also feasible but reduce the accuracy of the numerically computed log likelihood gradient used in maximum likelihood estimation.

$$\begin{aligned}
 E_{t+1}(V_{t+1}^2) &= \frac{\partial^2 G_{t+1|t+1}(\psi)}{\partial \psi^2} \Big|_{\psi=0} \\
 &= \frac{1}{2\pi p(y_{t+1}|\mathbf{Y}_t)} \int_{-\infty}^{\infty} [f_{\psi\psi}(i\Phi, 0) + f_{\psi}(i\Phi, 0)^2] e^{f(i\Phi, 0) - i\Phi y_{t+1}} d\Phi.
 \end{aligned}
 \tag{A.27}$$

Higher-order posterior moments can be computed similarly, by taking higher-order derivatives. In all cases, the integrand for negative Φ is the complex conjugate of the positive- Φ values. Moments can therefore be evaluated by integrating the real component of the integrand over $[0, \infty)$ and doubling the result.

The first two posterior moments were then used to update the parameters (κ_{t+1}, v_{t+1}) of next period's conditional moment generating function $G_{t+1|t+1}(\psi)$. These were based on the posterior moments

$$\begin{aligned}
 \kappa_{t+1} v_{t+1} &= E_{t+1}(V_{t+1}) \\
 \kappa_{t+1}^2 v_{t+1} &= E_{t+1}(V_{t+1}^2) - [E_{t+1}(V_{t+1})]^2.
 \end{aligned}
 \tag{A.28}$$

The algorithm is initiated at the unconditional gamma characteristic function of the initial V_0 , which is given above in Equation (A.20). 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.

A.4 Inferring jumps

A similar procedure was used to infer whether jumps occurred on a given day. Using Equation (A.9), the prior joint characteristic function of $(y_{t+1}, \Delta N_{t+1})$ conditional on data through date t is

$$F_{y, \Delta N}(i\Phi, i\xi|\mathbf{Y}_t) = e^{C(i\Phi, 0, i\xi; \tau)} G_{t|t}[D(i\Phi, 0, i\xi; \tau)]. \tag{A.29}$$

Proposition 1 can then be invoked to compute the posterior characteristic function

$$G_{\Delta N}(i\xi|\mathbf{Y}_{t+1}) \equiv E[e^{i\xi \Delta N_{t+1}} | y_{t+1}, \mathbf{Y}_t]. \tag{A.30}$$

The posterior moments of the number of jumps $E[(\Delta N_{t+1})^n | \mathbf{Y}_{t+1}]$ can be computed analogously to the approach in Equations (A.24)–(A.27). In particular, since the number of jumps on a given day is approximately a binomial variable, the estimated number of jumps $E[\Delta N_{t+1} | \mathbf{Y}_{t+1}]$ is approximately the probability that a jump occurred on date $t+1$. Since analytic derivatives with respect to ξ are messy, numerical derivatives were used instead.

A.5 Outliers and numerical efficiency

The integrations in Equations (A.24)–(A.27) are valid but numerically inefficient methods of evaluating densities and moments. For extreme return outliers, $p(y)$ takes on near-zero values (e.g., 2×10^{-22} for the 1987 crash under the SV model) that can create numerical problems for the integrations. The *scaled* density function can be evaluated more robustly and more efficiently. The Fourier transform of the scaled density function $e^{ay}p(y)$ is

$$\begin{aligned}
 \int [e^{ay}p(y)]e^{i\Phi y} dy &= E[e^{(a+i\Phi)y}] \\
 &= \exp[f(a + i\Phi, 0)]
 \end{aligned}
 \tag{A.31}$$

for f defined in Equation (A.25). This transformation is also known as *exponential tilting*, or the Esscher transform. It is the Fourier transform equivalent of Monte Carlo importance

sampling, and is the basis of saddlepoint approximations. Fourier inversion of this yields a density evaluation

$$\begin{aligned} p(y) &= \frac{e^{-ay}}{2\pi} \int_{-\infty}^{\infty} e^{f(a+i\Phi, 0) - i\Phi y} d\Phi \\ &= \frac{e^{f(a, 0) - ay}}{\sqrt{2\pi f_{\Phi\Phi}(a, 0)}} \left[\sqrt{\frac{f_{\Phi\Phi}(a, 0)}{2\pi}} \int_{-\infty}^{\infty} e^{f(a+i\Phi, 0) - f(a, 0) - i\Phi y} d\Phi \right]. \end{aligned} \quad (\text{A.32})$$

From saddlepoint approximation theory²³ the optimal y -dependent value of a (labeled a_y) is given implicitly by the minimum of $f(a, 0) - ay$:

$$f_{\Phi}(a_y, 0) = y. \quad (\text{A.33})$$

The scale factor a_y equals 0 when y equals its conditional mean $\bar{y} = f_{\Phi}(0, 0)$, and a_y is of the same sign as $y - \bar{y}$.

Using this scaling factor has several consequences. First, the term inside the brackets in Equation (A.32) is approximately 1, while the term preceding the brackets is the basic saddlepoint approximation for the density: $p(y) \approx \exp[f(a_y, 0) - a_y y] / \sqrt{2\pi f_{\Phi\Phi}(a_y, 0)}$. This follows from a second-order Taylor expansion of the exponent in Equation (A.32):

$$\begin{aligned} \int_{-\infty}^{\infty} e^{f(a_y+i\Phi, 0) - f(a_y, 0) - i\Phi y} d\Phi &\approx \int_{-\infty}^{\infty} \exp[f_{\Phi}(a_y, 0) - y](i\Phi) + 1/2 f_{\Phi\Phi}(a_y, 0)(i\Phi)^2] d\Phi \\ &= \int_{-\infty}^{\infty} \exp[-1/2 f_{\Phi\Phi}(a_y, 0)\Phi^2] d\Phi \\ &= \sqrt{\frac{2\pi}{f_{\Phi\Phi}(a_y, 0)\Phi^2}} \end{aligned} \quad (\text{A.34})$$

given that the first-order term in the Taylor expansion cancels by choice of a_y from (A.33).

Second, using $a = a_y$ makes the numerical integration in Equation (A.32) better behaved. The cancellation of the imaginary component removes an oscillatory component in the integrand in the neighborhood of $\Phi = 0$, and reduces it elsewhere. Furthermore, the integration is using locally the complex-valued path of steepest descent, for which the integrand falls off most rapidly in magnitude near $\Phi = 0$. Finally, evaluating the term in brackets in Equation (A.32) to a given absolute accuracy implies comparable accuracy for the log densities used in maximum likelihood estimation.

Similar rescalings were used to improve the efficiency and robustness of the integrals in Equations (A.26) and (A.27). For each integral, an upper limit $\Phi_{\max}$ was computed analytically for which estimated absolute truncation error would be less than 10^{-9} . The integral was then computed numerically to 10^{-8} accuracy using IMSL's adaptive Gauss-Kronrod DQDAG integration routine over $(0, \Phi_{\max})$, exploiting the fact that the integrands for negative Φ are complex conjugates of those for positive Φ . In the Monte Carlo runs, each integration required, on average, about 136 and 150 evaluations of the integrand for the SV and SVJ models, respectively.

A.6 Monte Carlo data generation

By Itô calculus and Equation (A.10), the orthogonalized state variable $q_t \equiv \ln S_t - (\rho/\sigma)V_t$ follows a jump-diffusion, with innovations uncorrelated with variance shocks:

²³ See Kolassa (1997, Ch. 4) or Barndorff-Nielsen and Cox (1989, Ch. 4).

$$dq_t = (\mu_{q0} + \mu_{q1} V_t) dt + \sqrt{(1 - \rho^2) V_t} dW_{2t} + \gamma_s dN_t \quad (\text{A.35})$$

for $\mu_{q0} \equiv \mu_0 - \lambda_0 \bar{k} - \alpha \rho / \sigma$ and $\mu_{q1} \equiv \mu_1 - 1/2 - \lambda_1 \bar{k} + \beta \rho / \sigma$. Given this orthogonality and the linearity of instantaneous mean, variance, and jump intensity in V_t , the daily time-aggregated innovation $\Delta q \equiv q_{t+\tau} - q_t$ conditional upon the intradaily variance sample path $\{V_s\}_t^{t+\tau}$ is a mixture of normals, with parameters that depend only upon the average intradaily variance:

$$\begin{aligned} \Delta q[\{V_s\}_t^{t+\tau}, n \text{ jumps}] &\sim N[(\mu_{q0} + \mu_{q1} \bar{V}_t) \tau + n \bar{\gamma}, (1 - \rho^2) \bar{V}_t \tau + n \delta^2] \\ n &\sim \text{Poisson}[(\lambda_0 + \lambda_1 \bar{V}_t) \tau]. \end{aligned} \quad (\text{A.36})$$

where τ is the daily time interval (1/252 years) between observations and $\bar{V}_t \tau \equiv \int_t^{t+\tau} V_s ds$.

Daily asset returns were therefore generated by

1. generating intradaily variance sample paths and computing average intradaily variance $\bar{V}_t$ and the daily variance shock $V_{t+\tau} - V_t$;
2. randomly generating the daily number of jumps n given daily average $\bar{V}_t$;
3. randomly generating daily Δq given n and sample $\bar{V}_t$; and
4. computing daily log asset returns $\Delta \ln S \equiv \Delta q + (\rho / \sigma)(V_{t+\tau} - V_t)$.

Intradaily variance sample paths were generated by dividing days into 50 subperiods, and using the exact discrete-time noncentral chi-squared transition density

$$\frac{2V_{t+h}}{K} | V_t \sim \chi^2 \left(\frac{4\alpha}{\sigma^2}, \frac{2V_t}{K} e^{-\beta h} \right) \text{ for } K \equiv \frac{\sigma^2}{2\beta} (1 - e^{-\beta h}) \text{ and } h = \tau/50. \quad (\text{A.37})$$

Random numbers were generated in two fashions. For the SV and SVJ1 simulations, the volatility of volatility parameter σ was rounded to make $m = 4\alpha/\sigma^2$ integer. This permitted exact Monte Carlo generation of noncentral $\chi^2(m, \Lambda_t)$ shocks from the sum of m independent squared normal shocks with unitary variance and mean Λ_t/m .²⁴ For the SVJ0 model, noncentral chi-squared random draws were generated using a substantially slower inverse CDF method, in order to duplicate exactly the parameter values used by Eraker, Johannes, and Polson (2003). The initial variance V_0 was independently drawn from its unconditional gamma density for each sample path.

The latent variance sample paths are therefore exact discrete-time draws from their postulated process, while log asset returns are drawn from the correct family of distributions with the appropriate sensitivity to variance shocks. Discretization error enters only in the evaluation of variance at 50 intradaily points rather than continuously, when computing the intradaily average variance used in generating returns via Equation (A.36).

²⁴ The default random number generator in IMSL (a multiplicative congruential generator with a multiplier of 16807) was found to be insufficiently random for generating intradaily data. A comparison of the statistical properties of daily variances generated from intradaily data with those generated directly at daily frequencies revealed low-order serial correlations in the former that biased the estimates of β by 5–10%. Using IMSL's most powerful random number generator (a higher multiplier, combined with shuffling) eliminated the biases.

Appendix B: The Log Variance Process

The benchmark discrete-time log variance process is

$$\begin{aligned} z_{t+1} &= \sqrt{V_t} \tilde{\varepsilon}_{t+1} \\ \ln V_{t+1} &= \omega + \phi \ln V_t + \sigma_\nu \tilde{\eta}_{t+1} \end{aligned} \quad (\text{B.1})$$

where $(\tilde{\varepsilon}_t, \tilde{\eta}_t)$ are i.i.d. $N(0, 1)$ shocks. By taking logs, the process has an affine state space representation of the form

$$\begin{aligned} y_{t+1} &\equiv \ln z_{t+1}^2 = x_t + \ln \tilde{\varepsilon}_{t+1}^2 \\ x_{t+1} &= \omega + \phi x_t + \sigma_\nu \tilde{\eta}_{t+1} \end{aligned} \quad (\text{B.2})$$

where $x_t \equiv \ln V_t$. The joint characteristic function in Equation (1) conditional upon knowing x_t takes the exponentially affine form

$$\begin{aligned} F(i\Phi, i\psi | x_t) &\equiv E[e^{i\Phi y_{t+1} + i\psi x_{t+1}} | x_t] \\ &= [e^{i\Phi x_t} E(|\tilde{\varepsilon}_{t+1}|^{2i\Phi})] [e^{i\psi(\omega + \phi x_t)} E(e^{i\psi \sigma_\nu \tilde{\eta}_{t+1}})] \\ &= \exp[C(i\Phi, i\psi) + D(i\Phi, i\psi) x_t] \end{aligned} \quad (\text{B.3})$$

where

$$\begin{aligned} C(i\Phi, i\psi) &\equiv [i\Phi \ln 2 + \ln \Gamma(1/2 + i\Phi) - \ln \Gamma(1/2)] + [\omega(i\psi) + 1/2 \sigma_\nu^2 (i\psi)^2] \\ D(i\Phi, i\psi) &= i\Phi + \phi(i\psi) \end{aligned} \quad (\text{B.4})$$

and $\ln \Gamma(\bullet)$ is the natural logarithm of the gamma function.²⁵

B.1 Approximate maximum likelihood

Latent log variance $x_t \equiv \ln V_t$ has an unbounded domain, and an unconditional normal distribution with unconditional moments

$$\begin{aligned} E[x] &\equiv \hat{x}_{0|0} = \frac{\omega}{1 - \phi} \\ Var[x] &\equiv P_{0|0} = \frac{\sigma_\nu^2}{1 - \phi^2} \end{aligned} \quad (\text{B.5})$$

and associated unconditional characteristic function

$$G_{0|0}(i\psi) = \exp[\hat{x}_{0|0} i\psi - 1/2 P_{0|0} \psi^2]. \quad (\text{B.6})$$

Since the domain of the latent variable is unbounded, the natural approximate prior to use in the AML filtration is Gaussian, with associated characteristic function

$$G_{t|t}(i\psi) = \exp[\hat{x}_{t|t} i\psi - 1/2 P_{t|t} \psi^2]. \quad (\text{B.7})$$

Conditional upon observing the datum $y_{t+1} \equiv \ln z_{t+1}^2$, the posterior moments $(\hat{x}_{t+1|t+1}, P_{t+1|t+1})$ can be updated using the algorithm in Equations (13)–(15), with C , D , and $G_{t|t}$ defined by Equations (B.4) and (B.7). Those posterior moments then determine the approximate normal posterior characteristic function $G_{t+1|t+1}(i\psi)$ for the next step.

²⁵ The complex-valued log gamma function was evaluated using the Lanczos approach described in Press et al. (1992, Ch. 6), which is accurate to 2×10^{-10} . IMSL lacks a double-precision complex log gamma function, while its single-precision CLNGAM function was found to be insufficiently smooth for use with filtration and parameter estimation.

The procedure is similar to the Kalman filtration used by Ruiz (1994) and Harvey, Ruiz, and Shephard (1994). Kalman filtration uses a strictly linear updating of the conditional mean of the latent log variance:

$$\hat{x}_{t+1|t+1} = (\omega + \phi \hat{x}_{t|t}) + \frac{\phi P_{t|t}}{P_{t|t} + 1/2\pi^2} [y_{t+1} - (x_{t|t} - \gamma - \ln 2)] \quad (\text{B.8})$$

where $\gamma \approx .577$ is Euler's constant, and $(-\gamma - \ln 2, 1/2\pi^2)$ are the mean and variance of ε_{t+1}^2 . The Kalman variance updating is nonlinear and deterministic:

$$P_{t+1|t+1} = \frac{\phi^2}{P_{t|t}^{-1} + 2/\pi^2} + \sigma_v^2, \quad (\text{B.9})$$

and converges to a minimal steady-state variance P_{ss} . The key difference is that AML permits nonlinear functions of the latest datum when updating the mean and variance of latent log variance. The AML filtration is the optimal Bayesian updating conditional upon a Gaussian prior. Kalman filtration is suboptimal because ε_{t+1}^2 is not Gaussian.

Figure (8) compares the Kalman and AML volatility filtrations, using the weekly parameter values $(\omega, \phi, \sigma_v) = (-.736, .90, .363)$ from Andersen, Chung, and Sørensen (1999). A Gaussian prior in Equation (B.7) was used for latent log variance in both cases, evaluated at the unconditional mean $\hat{x}_{t|t} = \hat{x}_{0|0} = -7.36$ and the steady-state variance $P_{ss} = .4985$ achieved under Kalman filtration. These values imply an initial volatility estimate of $E_t\sqrt{V_t} = \exp[1/2\hat{x}_{t|t} + 1/8P_{t|t}] = 2.68\%$ weekly, or about 19.3% annualized.

The "inlier" problem noted by Sandmann and Koopman (1998) is apparent: small absolute returns $|z_{t+1}|$ generate large negative values for $y_{t+1} = \ln z_{t+1}^2$, and undesirably

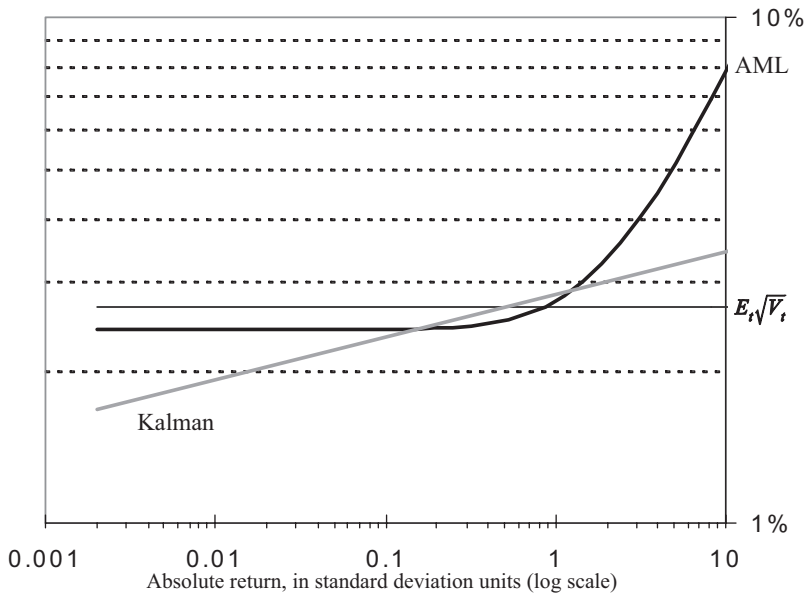


Figure 8
Weekly volatility estimates $E_{t+1}\sqrt{V_{t+1}}$ given a return of absolute size $|z_{t+1}|/E_t\sqrt{V_t}$
Initial volatility estimate: $E_t\sqrt{V_t} = 2.68\%$ weekly. Log scales for both axes.

large downward revisions in Kalman volatility and log variance estimates. The graph illustrates that the Kalman filter also suffers from an “outlier” problem: it substantially underresponds to returns larger than about two standard deviations. The result is inefficient filtration. On a generated sample of 20,000 weekly observations, Kalman filtration had an R^2 of 27% when estimating true volatility $\sqrt{V_t}$ or log variance $\ln V_t$. By contrast, the AML volatility and log variance filtrations had substantially higher R^2 s of 37–38%.

B.2 Parameter estimation efficiency

The AML approach potentially suffers some loss of estimation efficiency from its use of normal distributions to summarize what is known about the latent variable at each point in time. For the log variance process in Equation (B.1), the estimation efficiency can be directly compared with EMM, MCMC, and other approaches. The efficiency of those approaches for the log variance process is summarized in Andersen, Chung, and Sørensen (1999).

Table 9 appends to Andersen et al.’s Table 5 the results from simulating 500 independent sample paths from Equation (B.1) and estimating the parameters via AML, for sample

Table 9
Comparison of estimation methodologies for the discrete-time log variance process

	$T = 500$			$T = 2000$		
	ω	ϕ	σ_v	ω	ϕ	σ_v
True values:	−.736	.90	.363	−.736	.90	.363
Bias						
ML V	−.05	−.01	.00	−.015	−.002	.000
QML	−.7	−.09	.09	−.117	−.02	.020
GMM	.12 ^a	.02 ^a	−.12 ^a	.15	.02	−.08
EMM	−.17	−.02	.02	−.057	−.007	−.004
AML	−.15	−.02	.02	−.039	.005	.005
MCMC	−.13	−.02	−.01	−.026	−.004	−.004
ML	−.13	−.02	.01	NA	NA	NA
MCL	.14	.00	.01	NA	NA	NA
Root mean squared error						
ML V	.17	.02	.01	.076	.010	.006
QML	1.60	.22	.27	.46	.06	.11
GMM	.59 ^a	.08 ^a	.17 ^a	.31	.04	.12
EMM	.60	.08	.20	.224	.030	.049
AML	.42	.06	.08	.173	.023	.043
MCMC	.34	.05	.07	.15	.02	.034
ML	.43	.05	.08	NA	NA	NA
MCL	.27	.04	.08	NA	NA	NA

$$z_{t+1} = \sqrt{V_t} \tilde{\varepsilon}_{t+1}$$
$$\ln V_{t+1} = \omega + \phi \ln V_t + \sigma_v \tilde{\eta}_{t+1}.$$

ML|V and AML estimates are from 500 Monte Carlo sample paths of 500 and 2000 observations, respectively. Results for all other approaches are from comparable runs summarized in Table 5 of Andersen, Chung, and Sørensen (1999). Models:

- ML|V: ML *conditional* upon observing $\{V_t\}$
QML: Harvey, Ruiz and Shephard (1994)
GMM: Andersen and Sørensen (1996)
EMM: Andersen, Chung and Sørensen (1999)
- AML: approximate ML of this article
MCMC: Jacquier, Polson and Rossi (1994)
ML: Fridman and Harris (1998)
MCL: Sandmann and Koopman (1996)

^aAndersen et al. note that the GMM results for $T = 500$ are from runs that did not crash, and are therefore not comparable to results from other methods.

sizes of $T = 500$ and $T = 2000$ weeks, respectively. In addition, parameters were estimated *conditional* upon observing the log variance draws (denoted ML|V), as an unattainable bound on parameter estimation efficiency. The AML and ML|V approaches were estimated subject to the constraints $|\beta| < 1$ and $\sigma > 0$, to ensure stationarity and the existence of an unconditional distribution for the initial observation.

The AML approach is definitely one of the more efficient latent-variable methodologies reported in the table. While Jacquier, Polson, and Rossi's (1994) Monte Carlo Markov Chain approach has the lowest RMSEs for 2000-observation samples and is close to lowest for 500-observation samples, the RMSEs from AML estimation are almost as small. The AML approach outperforms Gallant and Tauchen's (2002) EMM approach, which in turn outperforms the GMM approach of Melino and Turnbull (1990) and the very inefficient Kalman filtrations (QML). AML performs about as well on 500-observation samples as Fridman and Harris's (1998) approach (ML), which also provides filtered estimates of latent volatility.

It may be possible to improve the AML performance further by better approximation methodologies for $G_{t|t}$. However, Table 9 suggests that even the simple two-moment Gaussian approximation is close to achieving the Cramér-Rao limit.

References

- Aït-Sahalia, Y., 2002, "Closed-Form Likelihood Expansions for Multivariate Diffusions," Working Paper 8956, National Bureau of Economic Research, May.
- Alizadeh, S., M. W. Brandt, and F. X. Diebold, 2002, "Range-Based Estimation of Stochastic Volatility Models," *Journal of Finance*, 57, 1047–1091.
- Andersen, T. G., L. Benzoni, and J. Lund, 2002, "An Empirical Investigation of Continuous-Time Equity Return Models," *Journal of Finance*, 57, 1239–1284.
- Andersen, T. G., T. Bollerslev, F. X. Diebold, and H. Ebens, 2001, "The Distribution of Stock Return Volatility," *Journal of Financial Economics*, 61, 43–76.
- Andersen, T. G., H. Chung, and B. E. Sørensen, 1999, "Efficient Method of Moments Estimation of a Stochastic Volatility Model: A Monte Carlo Study," *Journal of Econometrics*, 91, 61–87.
- Andersen, T. G., and B. E. Sørensen, 1996, "GMM Estimation of a Stochastic Volatility Model: A Monte Carlo Study," *Journal of Business and Economic Statistics*, 14, 328–352.
- Attari, M., 2004, "Option Pricing Using Fourier Transforms: A Numerically Efficient Simplification," Charles River Associates; available at www.ssrn.com.
- Bakshi, G., and N. Kapadia, 2003, "Delta-Hedged Gains and the Negative Market Volatility Risk Premium," *Review of Financial Studies*, 16, 527–566.
- Bakshi, G., and D. B. Madan, 2000, "Spanning and Derivative-Security Valuation," *Journal of Financial Economics*, 55, 205–238.
- Barndorff-Nielsen, O. E., and D. R. Cox, 1989, *Asymptotic Techniques for Use in Statistics*, Chapman and Hall, New York.
- Barone-Adesi, G., and R. E. Whaley, 1987, "Efficient Analytic Approximation of American Option Values," *Journal of Finance*, 42, 301–320.
- Bartlett, M. S., 1938, "The Characteristic Function of a Conditional Statistic," *The Journal of the London Mathematical Society*, 13, 63–67.
- Bates, D. S., 1991, "The Crash of '87: Was It Expected? The Evidence from Options Markets," *Journal of Finance*, 46, 1009–1044.
- Bates, D. S., 1996, "Jumps and Stochastic Volatility: Exchange Rate Processes Implicit in PHLX Deutschmark Options," *Review of Financial Studies*, 9, 69–107.

- Bates, D. S., 2000, "Post-'87 Crash Fears in the S&P 500 Futures Option Market," *Journal of Econometrics*, 94, 181–238.
- Bates, D. S., 2001, "The Market for Crash Risk," working Paper 8557, National Bureau of Economic Research, October.
- Bates, D. S., and R. Craine, 1999, "Valuing the Futures Market Clearinghouse's Default Exposure during the 1987 Crash," *Journal of Money, Credit, and Banking*, 31, 248–272.
- Bliss, R. R., and N. Panigirtzoglou, 2004, "Option-Implied Risk Aversion Estimates," *Journal of Finance*, 59, 407–446.
- Carr, P., H. Geman, D. B. Madan, and M. Yor, 2003, "Stochastic Volatility for Lévy Processes," *Mathematical Finance*, 13, 345–382.
- Carr, P., and D. B. Madan, 2000, "Option Valuation Using the Fast Fourier Transform," *Journal of Computational Finance*, 2, 61–73.
- Carrasco, M. et al., 2002, "Efficient Estimation of Jump Diffusions and General Dynamic Models with a Continuum of Moment Conditions," University of North Carolina.
- Chacko, G., and L. M. Viceira, 2003, "Spectral GMM Estimation of Continuous-time Processes," *Journal of Econometrics*, 116, 259–292.
- Chernov, M., et al., 2003, "Alternative Models for Stock Price Dynamics," *Journal of Econometrics*, 116, 225–257.
- Coval, J. D., and T. Shumway, 2001, "Expected Option Returns," *Journal of Finance*, 56, 983–1009.
- Cumby, R. E., and J. D. Glen, 1990, "Evaluating the Performance of International Mutual Funds," *Journal of Finance*, 45, 497–521.
- Duffie, D., J. Pan, and K. J. Singleton, 2000, "Transform Analysis and Asset Pricing for Affine Jump-Diffusions," *Econometrica*, 68, 1343–1376.
- Duffie, D., and K. J. Singleton, 1993, "Simulated Moments Estimation of Markov Models of Asset Prices," *Econometrica*, 61, 929–952.
- Durham, G. B., and A. R. Gallant, 2002, "Numerical Techniques for Maximum Likelihood Estimation of Continuous-Time Diffusion Processes," *Journal of Business and Economic Statistics*, 20, 297–316.
- Eden, R., 1555, *The Decades of the neue worlde or west India, conteynnyng the nauigations and conquestes of the Spanyardes, with the particular description of the moste ryche and large landes and Ilandes lately founde in the west Ocean, perteynyng to the inheritaunce of the kinges of Spayne*, Rycharde Jug, London.
- Eraker, B., M. Johannes, and N. G. Polson, 2003, "The Impact of Jumps in Volatility and Returns," *Journal of Finance*, 58, 1269–1300.
- Feuerverger, A., 1990, "An Efficiency Result for the Empirical Characteristic Function in Stationary Time-Series Models," *The Canadian Journal of Statistics*, 18, 155–161.
- Feuerverger, A., and P. McDunnough, 1981a, "On Some Fourier Methods for Inference," *Journal of the American Statistical Association*, 76, 379–387.
- Feuerverger, A., and P. McDunnough, 1981b, "On the Efficiency of Empirical Characteristic Function Procedures," *Journal of the Royal Statistics Society, Series B*, 43, 20–27.
- Fleming, J., and C. Kirby, 2003, "A Closer Look at the Relation between GARCH and Stochastic Autoregressive Volatility," *Journal of Financial Econometrics*, 1, 365–419.
- Fridman, M., and L. Harris, 1998, "A Maximum Likelihood Approach for Non-Gaussian Stochastic Volatility Models," *Journal of Business and Economic Statistics*, 16, 284–291.
- Gallant, A. R., and G. Tauchen, 1998, "Reprojecting Partially Observed Systems With Application to Interest Rate Diffusions," *Journal of the American Statistical Association*, 93, 10–24.

- Gallant, A. R., and G. Tauchen, 2002, "Simulated Score Methods and Indirect Inference for Continuous-Time Models," University of North Carolina.
- Goetzmann, W. N., J. E. Ingersoll, Jr., M. Spiegel, and I. Welch, 2002, "Sharpening Sharpe Ratios," Working Paper 9116, National Bureau of Economic Research, August.
- Hamilton, J. D., 1994, *Time Series Analysis*, Princeton University Press, Princeton, NJ.
- Harvey, A., E. Ruiz, and N. Shephard, 1994, "Multivariate Stochastic Variance Models," *Review of Economic Studies*, 61, 247–264.
- Hentschel, L., 1995, "All in the Family: Nesting Symmetric and Asymmetric GARCH Models," *Journal of Financial Economics*, 39, 71–104.
- Ho, M. S., W. R. M. Perraudin, and B. E. Sørensen, 1996, "A Continuous Time Arbitrage Pricing Model with Stochastic Volatility and Jumps," *Journal of Business and Economic Statistics*, 14, 31–43.
- Honoré, P., 1998, "Pitfalls in Estimating Jump-Diffusion Models," Aarhus School of Business.
- Huang, J. and L. Wu, 2004, "Specification Analysis of Option Pricing Models Based on Time-Changed Lévy Processes," *Journal of Finance*, 59, 1405–1439.
- Jacquier, E., N. G. Polson, and P. E. Rossi, 1994, "Bayesian Analysis of Stochastic Volatility Models," *Journal of Business and Economic Statistics*, 12, 1–19.
- Jegadeesh, N., and G. Pennacchi, 1996, "The Behavior of Interest Rates Implied by the Term Structure of Eurodollar Futures," *Journal of Money, Credit, and Banking*, 28, 426–446.
- Jiang, G. J., and J. L. Knight, 2002, "Estimation of Continuous Time Processes Via the Empirical Characteristic Function," *Journal of Business and Economic Statistics*, 20, 198–212.
- Johannes, M., N. G. Polson, and J. Stroud, 2002, "Nonlinear Filtering of Stochastic Differential Equations with Jumps," Columbia University.
- Judd, K. L., 1998, *Numerical Methods in Economics*, MIT Press, Cambridge (MA).
- Karlin, S. and H. M. Taylor, 1981, *A Second Course in Stochastic Processes*, Academic Press, Boston.
- Kolassa, J. E., 1997, *Series Approximation Methods in Statistics*, Springer-Verlag, New York.
- Lewis, A. L., 2001, "A Simple Option Formula for General Jump-Diffusion and Other Exponential Lévy Processes," Environ Financial Systems.
- Liu, J., J. Pan, and T. Wang, 2004, "An Equilibrium Model of Rare Event Premia," *Review of Financial Studies*, 18, 131–164.
- Lo, A. W., 1988, "Maximum Likelihood Estimation of Generalized Itô Processes with Discretely Sampled Data," *Econometric Theory*, 4, 231–247.
- McFadden, D., 1989, "A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration," *Econometrica*, 57, 995–1026.
- Melino, A., and S. M. Turnbull, 1990, "Pricing Foreign Currency Options with Stochastic Volatility," *Journal of Econometrics*, 45, 239–265.
- 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.
- Nelson, D. B., 1992, "Filtering and Forecasting with Misspecified ARCH Models I: Getting the Right Variance with the Wrong Model," *Journal of Econometrics*, 52, 61–90.
- Nelson, D. B., and D. P. Foster, 1994, "Asymptotic Filtering Theory for Univariate ARCH Models," *Econometrica*, 62, 1–41.

- Pan, J., 2002, "The Jump-Risk Premia Implicit in Options: Evidence from an Integrated Time-Series Study," *Journal of Financial Economics*, 63, 3–50.
- 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.
- Pearson, N. D., and T. Sun, 1994, "Exploiting the Conditional Density in Estimating the Term Structure: An Application to the Cox, Ingersoll, and Ross Model," *Journal of Finance*, 49, 1279–1304.
- Press, W. H. et al, 1992, *Numerical Recipes in FORTRAN: The Art of Scientific Computing* (2nd ed.), Cambridge University Press, Cambridge.
- Rosenberg, J., and R. F. Engle, 2002, "Empirical Pricing Kernels," *Journal of Financial Economics*, 64, 341–372.
- Ruiz, E., 1994, "Quasi-Maximum Likelihood Estimation of Stochastic Volatility Models," *Journal of Econometrics*, 63, 289–306.
- Sandmann, G., and S. J. Koopman, 1998, "Estimation of Stochastic Volatility Models via Monte Carlo Maximum Likelihood," *Journal of Econometrics*, 87, 271–301.
- Singleton, K. J., 2001, "Estimation of Affine Asset Pricing Models Using the Empirical Characteristic Function," *Journal of Econometrics*, 102, 111–141.
- Stambaugh, R. F., 1999, "Predictive Regressions," *Journal of Financial Economics*, 54, 375–421.

Copyright of Review of Financial Studies is the property of Society for Financial Studies and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.