Monte Carlo methods for estimating, smoothing, and filtering one- and two-factor stochastic volatility models

Garland B. Durham

Leeds School of Business, University of Colorado, 419 UCB, Boulder, CO 80309-0419, USA

Abstract

One- and two-factor stochastic volatility models are assessed over three sets of stock returns data: S&P 500, DJIA, and Nasdaq. Estimation is done by simulated maximum likelihood using techniques that are computationally efficient, robust, straightforward to implement, and easy to adapt to different models. The models are evaluated using standard, easily interpretable time-series tools. The results are broadly similar across the three data sets. The tests provide no evidence that even the simple single-factor models are unable to capture the dynamics of volatility adequately; the problem is to get the shape of the conditional returns distribution right. None of the models come close to matching the tails of this distribution. Including a second factor provides only a relatively small improvement over the single-factor models. Fitting this aspect of the data is important for option pricing and risk management.

1. Introduction

Equity returns data commonly exhibit volatility clustering and non-Gaussian distributions. A huge literature devoted to analyzing models that try to account for

E-mail address: Garland.durham@colorado.edu.

these characteristics has developed. Such models are important for pricing derivative securities and risk management.

The literature has several dominant strands. The bulk of applied work has used some version of the ARCH/GARCH class of models. Such models can generate data possessing the features mentioned above, and statistical analysis is straightforward since the volatility state is easily deduced from the data. On the other hand, these models imply a deterministic link between the return and volatility processes that may be difficult to justify on either empirical or theoretical grounds. An alternative is the class of stochastic volatility (SV) models. While the additional source of randomness in SV models provides more flexibility in fitting the data, statistical analysis is more complicated since the state is not uniquely determined by the data. Regime-switching models provide a third alternative (e.g., Hamilton, 1990). Recent work by Geweke and Amisano (2001) suggests the idea of using compound Markov mixtures of normals. This approach provides a highly flexible modeling framework, lending a nonparametric flavor to the endeavor. Other approaches to volatility modeling include using the information available in high-frequency data (Andersen et al., 2003; Barndorff-Nielsen and Shephard, 2002), hi-lo quotes (Alizadeh et al., 2002) or option prices (Jones, 2003; Koopman et al., 2005).

This paper demonstrates tools for likelihood-based analysis of standard one- and two-factor SV models used in finance, building on work by Kim et al. (1998), Shephard and Pitt (1997), Durbin and Koopman (1997, 2000), and Sandmann and Koopman (1998). The likelihood function is approximated by using simulation to integrate out an unobserved "auxiliary" variable (the volatility process in this case). Estimation is carried out by maximizing the approximate likelihood. This approach is commonly referred to as simulated maximum likelihood estimation (SMLE). It has a great deal in common with the Bayesian Markov chain Monte Carlo (MCMC) approach that has become popular in recent years. SMLE is loosely related to simulated method of moments (Duffie and Singleton, 1993) and efficient method of moments (Gallant and Tauchen, 1996), which also rely upon simulation as a tool to approximate estimation criteria that are unavailable in closed form.

Other related work on maximum likelihood estimation of SV models includes Danielson and Richard (1993), Danielson (1994), and Liesenfeld and Richard (2003), who also use Monte Carlo methods to integrate out the unobserved states, and Fridman and Harris (1998), who use recursive numerical integration.

The tools used in this paper are robust, computationally efficient, straightforward to implement, and easy to adapt to different models. Log volatility models with one and two factors, affine models, and a new formulation of the SV-t model are examined. All the models allow for the possibility of leverage effects. In addition to estimation, issues related to filtering, smoothing, model diagnostics, and numerical performance are considered.

Using these tools, the likelihood of a standard single factor SV model on a data set of several thousand observations can be approximated in a fraction of a second on a typical PC. The algorithm can be implemented in a few dozen lines of Fortran code. Adapting the code to a new model is simply a matter of providing a few model-specific functions, code for which is easily obtained using symbolic manipulation

software such as Maple. The smoothers, filters, and model diagnostics are also easy to implement in a few lines of Fortran code.

The critical requirement for being able to use the tools described in this paper is that it be possible to specify in closed form the joint density of the observed and unobserved variables conditional on their past. While this is often possible (such as with the SV models considered in this paper), there are many interesting models used in economics where it is not. In such cases, the simulated method of moments approach—which requires only that one be able to generate synthetic data from the model—may still be feasible.

With any simulation-based estimator, a careful examination of numerical issues is essential. Such estimators are based on simulations generated using sequences of pseudorandom numbers. The estimator is a function not only of the data, but also of the seed used to generate the pseudorandom sequence. Given a fixed data set, each choice of seed implies a different parameter estimate. A common approach is to fix a particular pseudorandom sequence and report the corresponding estimates. The hope is that the estimates corresponding to other sequences would not differ by much. Convincing evidence is often not provided, and in some cases may be costly to obtain.

For each of the models considered in this paper, estimates based on many different seeds for the random number generator are computed. The mean and standard deviation of the individual estimates is reported. This provides information as to the numerical stability of the estimator with respect to different seeds. By taking an increasing sequence of simulation lengths, it is possible to observe the rate at which the simulation-induced variance dissipates. By checking for systematic patterns in location shifts for the parameter estimates with increasing simulation length, it is also possible to address the issue of simulation-induced bias.

It should be noted that simulation error is an issue not just with SMLE, but with simulated method of moments and Bayesian MCMC estimators as well. The problem is that assessing simulation error may be difficult given the high computational demands of some of these estimators. One of the advantages of the computational efficiency of the tools used in this paper is that careful studies of their numerical properties (as well as their small sample statistical properties) are feasible.

This paper is the first time that several of the models considered have been estimated using likelihood-based tools. While similar (and yet more sophisticated models) have been estimated using the efficient method of moments (e.g., Chernov et al., 2003), parameter estimates obtained using the two approaches—as well as the precision of those estimates—can differ substantially. The model diagnostics examined in this paper are also substantially different from those provided by EMM. And finally, EMM is more computationally demanding than the techniques considered in this paper. The merits of EMM are well-documented in the literature. This paper is intended to further explore alternatives.

The empirical work looks at three data sets: S&P 500 index returns over the period June 23, 1980 through September 2, 2002; Dow Jones Industrial Average returns over the period January 1953 through July 30, 1999; and Nasdaq returns from October, 1984 through September 15, 2004. A variety of models is estimated over

each data set. The availability of the log likelihood means that models can be evaluated using Kullback–Leibler and related information criteria. The models are also subjected to a battery of diagnostic tests.

The results are qualitatively similar for all the data sets:

- I find no evidence that even the simple single factor models are unable to capture the *dynamics* of the volatility process. In contrast, Gallant et al. (1997) using a longer data set (1928–1987) find evidence of more complex dynamics in the volatility process. Their results suggest modeling volatility as a fractionally differenced AR(2) process. Bollerslev and Mikkelsen (1996), Ding et al. (1993) and others also find evidence of long memory in stock returns. Andersen et al. (2002), Chernov et al. (2003) and others find evidence in favor of a second volatility factor. The failure of the diagnostics used in this paper to find evidence of such behavior may be due to the shorter sample period or to lack of power in the tests.
- The more critical problem is to capture the *shape* of the conditional returns distribution. While including a second volatility factor helps some, all of the models fail in a similar manner: none is able to explain the extreme left tail of the distribution. This is in agreement with earlier findings of Gallant et al. (1997) and others, though some of the diagnostic information presented in this paper is new. In addition to formal statistical tests, QQ-plots clearly illustrating the nature of the models' failure are provided.

Surprisingly good results are obtained using a new formulation of the SV-t model. Performance is comparable to the more commonly used two-factor models. This model differs from those considered previously in the literature in the way that correlation between returns and the volatility process is introduced. The volatility process inherits some of the kurtosis found in returns. In particular, large absolute returns (e.g., "crash" days) are associated with simultaneous "jumps" in volatility, an effect that Eraker et al. (2003) argue is an important feature of the data. Using t rather than normal errors adds kurtosis to the returns distribution. But skewness is needed as well. The SV-t model is unable to account for the long left tail.

The rest of this paper is organized as follows: Section 2 describes the estimation approach used, Section 3 looks at the numerical performance of the estimators, Section 4 discusses the issues of filtering and smoothing, Section 5 is the application, and Section 6 concludes.

2. Estimation methodology

The basic idea underlying SMLE is as follows. Suppose that $x = (x_1, \ldots, x_n)$ is a realization from some random vector $X = (X_1, \ldots, X_n)$ for which direct evaluation of the density function $p(x)$ is infeasible, but that there exists some collection of (unobserved) auxiliary variables $V = (V_1, \ldots, V_n)$ such that the joint density $p(x, v) \leftarrow$ is easy to evaluate. The likelihood of a parameter vector θ can then be obtained by

integrating out the auxiliary variables:

$$L(\theta|x) = \int p(x, v; \theta)\, dv. \tag{1}$$

This is generally a very high-dimensional integral that must be evaluated using Monte Carlo techniques. The idea is to draw samples $v^{(1)}, \ldots, v^{(S)}$ from some density q, referred to as an importance density, and compute

$$L(\theta|x) = \int \left(\frac{p(x, v; \theta)}{q(v)}\right) dQ(v)$$

$$\approx \frac{1}{S} \sum_{s=1}^{S} \frac{p(x, v^{(s)})}{q(v^{(s)})}. \tag{2}$$

Thus the likelihood is approximated by a weighted average across "simulated" draws from q. Although q will usually depend on x and θ, this will be suppressed in the notation. The dependence of p on θ will usually be supressed as well. The estimation step is performed by maximizing the approximate likelihood thus obtained.

The theory of Monte Carlo integration is well understood (e.g., Judd, 1998). Convergence of the sum on the right-hand side of (2) follows from a straightforward application of the strong law of large numbers (treating x and θ as fixed). It is sufficient to verify that the integral in Eq. (2) exists. In practice, one would also like for the variance of $p(x, V)/q(V)$ (with respect to q) to be finite, so that a central limit theorem may be applied to show that convergence is at rate $\sqrt{S}$. The issue is essentially whether the tails of q are sufficiently thick with respect to those of p. This is not just of theoretical concern, but a very practical problem. If the tails of q are too thin, very large values of p/q will be drawn occasionally and the sum will be erratic over repeated trials.

While a theoretical verification that the regularity conditions are satisfied is of course useful, it is easy to design importance samplers that satisfy them yet perform so poorly that they are of little use in practice (e.g., Geweke, 1989). It is also possible to take an importance density which works well over most of its range and fix problems far out in the tails by truncation (e.g., Kloek and van Dijk, 1978). A careful examination of the convergence properties of the sum in (2) for the particular problem at hand is essential (indeed, providing an estimate of the numerical error should be part of the standard operating procedure for *any* simulation-based estimator).

The SV models examined in this paper are of form

$$X_{t+1} = \mu_X(X_t, V_\tau) + \sigma_X(X_t, V_\tau)\varepsilon_{t+1},$$
$$V_{t+1} = \mu_V(V_t) + \sigma_V(V_t)\eta_{t+1}, \tag{3}$$

where X_t is an n_x-dimensional observed component, V_t is an n_v-dimensional latent volatility component, and (ε_t, η_t) is an $(n_x + n_v)$-dimensional random variable with mean zero and variance Ω. By setting τ equal to either t or $t+1$, the above model encompasses the different timings that appear in the literature.

The importance sampler is based on the Laplace approximation to $p(x,v)$ (see e.g., Gelman et al., 1995). One first computes

$$\hat{v} = \operatorname*{argmax}_{v} \log p(x,v)$$

and

$$H = \frac{\partial^2}{\partial v^2} \log p(x, \hat{v}).$$

The importance density is given by the multivariate normal with mean $\hat{v}$ and variance $-H^{-1}$. The mode, $\hat{v}$, of $p(x,v)$ is obtained using Newton's method. Although this would appear to be costly since it involves solving a nn_v dimensional system of linear equations, the Hessian is positive definite symmetric banded (with n_v off-diagonals). Efficient techniques are available to solve linear systems with this structure. Note that there is never any need to obtain H^{-1} explicitly.

This approach is similar to that used by Durbin and Koopman (1997, 2000), but the implementation is more straightforward, especially for models involving correlation between ε_t and η_t. Durbin and Koopman approximate the model by a linear Gaussian state-space model. The approximating model is computed iteratively using the Kalman filter and smoother.

It is easy to write this importance sampler down. As discussed above, the critical step is to verify its performance in practice for the particular model at hand. This issue is addressed in Section 3.

The estimation procedure is very efficient computationally and can be implemented in just a few lines of Fortran code. The only model dependent code involves the computation of $\log p(x,v)$ and its first and second derivatives with respect to v. In practice, $\log p(x,v)$ is obtained as the sum of terms of form $\log p(x_{t+1}, v_{t+1}|x_t, v_t)$. The derivatives of $\log p(x,v)$ are constructed by stacking up the derivatives of these terms blockwise. The formulae are typically easy to obtain. The simplest way to do so is using Maple or some other symbolic manipulation software. Code and detailed implementation notes are available upon request.

Jacquier et al. (1994) (JPR hereafter) compare the performance of their MCMC estimator for SV models against several other approaches in an extensive simulation study that has become a standard benchmark in this literature. They look at the model

$$\begin{aligned}
X_t &= \exp(V_t/2)\varepsilon_t, \\
V_t &= \alpha + \phi V_{t-1} + \sigma_V \eta_t,
\end{aligned} \qquad (4)$$

where the (ε_t, η_t) are iid standard normal.

Following JPR, I estimate (4) over synthetic data generated using various parameter settings that they argue are representative of much financial data. The distribution of the estimates obtained using the approach described in this paper is essentially the same as found by JPR and others. Details are available upon request. The estimator is very well-behaved with these models. Experiments over thousands of simulated data sets never resulted in any apparent problems with convergence.

Computational cost for this model is about 0.05 s per evaluation of the likelihood function (on a 2 GHz PC) with $n = 1000$ observations and $S = 64$ draws from the importance sampler. Time required to maximize the likelihood is more variable. Using two-sided numerical derivatives, the Broyden–Fletcher–Goldfarb–Shapiro optimizer, and a reasonable start value, times of around 3 s are typical.

It is sometimes found in empirical work that negative returns are associated with a subsequent increase in volatility. This is often referred to as the "leverage effect". No changes in the main body of code are needed to include correlation between returns and the volatility process in (4). The only thing needed is a minor change in the Maple code used to obtain expressions for $\log p(x_t, v_t | x_{t+1}, v_{t+1})$ and its first and second derivatives. The resulting estimator works with essentially the same efficiency and robustness as for the uncorrelated case. Some simulation results are shown in Table 1. Parameter settings for ϕ, σ_X and σ_V are taken from the "middle" case of JPR. Various settings for the correlation parameter ρ are tried.

An alternative formulation of this model is

$$
\begin{aligned}
X_t &= \mu + \sigma_X \exp(V_{t-1}/2)\varepsilon_t, \\
V_t &= \phi V_{t-1} + \sigma_V v_t
\end{aligned}
\tag{5} \leftarrow
$$

with $\operatorname{corr}(\varepsilon_t, v_t) = \rho$. This is a different way of organizing the parameters, but more critically, the timing is different (depending upon whether V_t or V_{t-1} appears in the returns equation). The model in (5) is a martingale difference sequence[1] (after subtracting off the unconditional mean, μ). It also represents the Euler scheme approximation of the underlying continuous-time model.

The model in (4), on the other hand, has the feature that large absolute returns are associated with concurrent shifts in the level of volatility, introducing an additional source of non-Gaussianity into the model. In particular, the distribution of $X_t | V_{t-1}$ is skewed if $\rho \neq 0$ with this timing. Jacquier et al. (2002) argue that this effect may help to explain the extremely large negative returns that are seen occasionally in the data ("crash" days).

Whether the asymmetry introduced by the JPR timing is important in practice can only be determined empirically. That the model with this timing is not the Euler scheme approximation to the underlying continuous-time model does not appear to be critical. The Euler scheme is one approximation; it is neither the only one nor necessarily the best. And while this model is not a martingale difference sequence, the departure from "martingale-ness" may be small. Furthermore, it is typically not obvious that the true data generating process has this characteristic. A careful empirical assessment of the timing issue is undertaken in Section 5. Both timings of

---

[1]To see why the model in (4) is not a martingale difference sequence, rewrite it as

$$
\begin{aligned}
X_t &= \exp(V_t/2)(\sqrt{1 - \rho^2}\eta_t + \rho\varepsilon_t), \\
V_t &= \alpha + \phi V_{t-1} + \sigma_V \eta_t
\end{aligned}
$$

and note that $\mathrm{E}[\exp(V_t/2)\eta_t | V_{t-1}] = \exp(\alpha + \phi V_{t-1}/2)\mathrm{E}[\exp(\sigma_V \eta_t/2)\eta_t] \neq 0$. The same reasoning holds even if the conditioning is with respect to $X_1, \ldots, X_{t-1}$ rather than V_{t-1}.

Table 1
Simulation studies

Single-factor (see Eq. (4))

	σ_X	ϕ	σ_V	ρ
True	0.0252	0.9500	0.260	0.000
	0.0253	0.9476	0.256	0.000
	(0.0015)	(0.0124)	(0.028)	(0.074)
True	0.0252	0.9500	0.260	−0.250
	0.0253	0.9476	0.257	−0.255
	(0.0015)	(0.0120)	(0.027)	(0.071)
True	0.0252	0.9500	0.260	−0.500
	0.0253	0.9479	0.258	−0.507
	(0.0014)	(0.0106)	(0.025)	(0.062)
True	0.0252	0.9500	0.260	−0.750
	0.0253	0.9486	0.259	−0.757
	(0.0011)	(0.0080)	(0.021)	(0.043)

SV-t (see Eq. (6))

	μ	σ_X	ϕ	σ_V	ρ	v
True	0.0000	0.0070	0.980	0.100	−0.500	8.00
	−0.0000	0.0070	0.976	0.103	−0.515	8.45
	(0.0002)	(0.0005)	(0.010)	(0.022)	(0.116)	(1.83)

Two-factor (see Eq. (8))

	μ	σ_X	ϕ_V	σ_V	ρ_{21}	ϕ_U	σ_U	ρ_{31}
True	0.00000	0.00800	0.980	0.090	−0.400	0.20	0.500	0.000
	−0.00002	0.00806	0.970	0.103	−0.390	0.15	0.435	0.005
	(0.00019)	(0.00043)	(0.023)	(0.032)	(0.154)	(0.30)	(0.071)	(0.103)

Affine (see Eq. (9))

	μ	ϕ	α	σ	ρ
True	0.00000	0.030	0.000200	0.0010	−0.500
	0.00000	0.041	0.000200	0.0011	−0.501
	(0.00032)	(0.030)	(0.000012)	(0.0003)	(0.149)

The mean and standard deviation of parameter estimates are shown. For the single-factor model with correlation, the data generating parameters correspond to the middle cell of the simulation study in Jacquier et al. (1994). For the remaining models, the data generating parameters correspond approximately to parameter estimates obtained in the empirical sections of this paper. All results are based on 5000 replications with $N = 2000$ observations each. The single-factor and SV-t models use $S = 64$ draws from the importance sampler; the two-factor and affine models use $S = 256$.

the model can be estimated with just a minor modification in the code. There is essentially no effect on the numerical performance.

Another model of interest is the SV-t model, which uses t in place of normally distributed errors in the returns process. The idea is to model the excess kurtosis in returns that remains even after allowing for time-varying volatility. The specification used in this paper is given by

$$X_t = \mu + \sigma_X \exp(V_{t-1}/2)\varepsilon_t,$$
$$V_t = \phi V_{t-1} + \sigma_V\left(\rho\varepsilon_t + \sqrt{1-\rho^2}\eta_t\right), \qquad (6)$$

where $\varepsilon_t \sim t_v$, $\eta_t \sim N(0,1)$, and ε_t and η_t are independent. Correlation between returns and volatility is introduced by including ε_t in the innovations of both.

An alternative formulation of the model is estimated by Jacquier et al. (2002) using Bayesian MCMC techniques. The specification they use is given by

$$X_t = \sigma_X \exp(V_t/2)\lambda_t\varepsilon_t,$$
$$V_t = \phi V_{t-1} + \sigma_V\eta_t, \qquad (7)\leftarrow$$

where $v/\lambda_t \sim \chi_v^2$, ε_t and η_t are $N(0,1)$, and $\mathrm{corr}(\varepsilon_t, \eta_t) = \rho$. Note that the product $\lambda_t\varepsilon_t$ has the t_v distribution.

There are two key differences between these models. First, there is a timing issue analogous to the one discussed above. But also, the presence of ε_t in the volatility equation of (6) implies that the volatility process inherits some of the kurtosis found in returns. In particular, large absolute returns (e.g., "crash" days) are associated with simultaneous "jumps" in the volatility level. Eraker et al. (2003) argue that this is an important feature of the data.

SV-t models have also been estimated using likelihood-based techniques by Chib et al. (2002), Sandmann and Koopman (1998) and Liesenfeld and Richard (2003), but only with uncorrelated errors (note that (6) and (7) are equivalent if the errors are not correlated). As will be shown in Section 5, the models without correlation are not empirically relevant, at least for the data sets examined in this paper.

Implementing the estimator for the model shown in (6) requires no changes in the main body of code. Expressions for the transition density and its derivatives are easily obtained using Maple. Performance and robustness of the estimator are similar to the case with normally distributed errors. See Table 1 for some results from a small simulation study.

Another approach to capturing the non-Gaussianity of returns is to include a second volatility factor. The first factor is highly persistent and captures volatility clustering. The second factor has little persistence. Its role is primarily to control the shape of the distribution of returns. Engle and Lee (1998) use a two-factor GARCH model. Two-factor SV models have been explored (using different techniques) by Gallant et al. (1999), Alizadeh et al. (2002), and Chernov et al. (2003) among others. Liesenfeld and Richard (2003) estimate and assess a two-factor model (but without leverage effects) using likelihood-based tools.

A standard two-factor model is given by

$$X_t = \mu + \sigma_X \exp(U_{t-1}/2 + V_{t-1}/2)\varepsilon_{1t},$$
$$V_t = \phi_V V_{t-1} + \sigma_V \varepsilon_{2t},$$
$$U_t = \phi_U U_{t-1} + \sigma_U \varepsilon_{3t}, \qquad (8)$$

where $\varepsilon_{it} \sim N(0,1)$ and $\mathrm{corr}(\varepsilon_{it}, \varepsilon_{jt}) = \rho_{ij}$. In addition to (8), two alternative timings are of interest for this model. The first uses the JPR timing for both factors, i.e., U_t and V_t in place of U_{t-1} and V_{t-1} in the returns equation. The other timing of potential interest is a hybrid using the Euler-scheme timing for the persistent factor and the JPR timing for the non-persistent factor.

The approach used to estimate this model is basically the same as for the one-factor model, but the implementation is more involved. It is primarily a matter of "bookkeeping". The Hessian required in the construction of the importance sampler is constructed blockwise from the second derivatives of $p(x_t, u_t, v_t | x_{t-1}, u_{t-1}, v_{t-1})$ with respect to $(u_t, v_t, u_{t-1}, v_{t-1})$. To maintain the banded form for the Hessian of the transition matrix, the latent factors must be interleaved, $(u_1, v_1, u_2, v_2, \ldots, u_n, v_n)$. Also, the blocks of the second derivative matrix are 4×4 (versus 2×2 for the single-factor case), which means that there are ten cells to fill in for each block (rather than three for the single-factor case; recall that the Hessian is symmetric). Otherwise, the algorithm differs little from the single-factor case.

Computational cost is greater than for the single-factor model but remains modest. More data is needed to obtain reasonably precise estimates for some of the parameters (the rate of mean reversion of the non-persistent factor is difficult to estimate precisely even with several thousand observations). Also, more draws from the importance sampler are needed to obtain acceptable levels of numerical precision. The simulation studies reported in Table 1 are based on $n = 2000$ observations and $S = 256$ draws from the importance sampler. Computational cost is less than $1\,\mathrm{s}$ (on a $2\,\mathrm{GHz}$ PC) per evaluation of the likelihood function.

The estimation approach described above can also be used with the affine class of SV models, but slightly more work is required. First, the model should be transformed to one where the coefficient of the Brownian motion in the latent process is constant. Given the continuous-time version of the model

$$dY_t = \mu\, dt + \sqrt{V_t}\, dW_{1t},$$
$$dV_t = \phi(\alpha - V_t)\, dt + \sigma\sqrt{V_t}\, dW_{2t},$$

where W_1 and W_2 are (possibly correlated) Brownian motions, use Ito's rule with the transformation $h_t = \sqrt{V_t}$ to get

$$dY_t = \mu\, dt + h_t\, dW_{1t}$$
$$dh_t = \frac{\phi}{2h_t}\left(\alpha - \frac{\sigma^2}{4\phi} - h_t^2\right) dt + \frac{\sigma}{2}\, dW_{2t}.$$

This is the same "volatility-stabilizing transformation" used in Durham and Gallant (2002). The idea is the same in either context. The transformed model is closer to

Gaussian and the performance of the importance sampler is dramatically improved. For the generic case where the coefficient of W_{2t} is $\sigma_V(v)$, the desired transformation is given by $h = \int \sigma_V^{-1}(v)\,\mathrm{d}v$.

The form of the model that is actually estimated is the Euler approximation

$$X_t = \mu + h_{t-1}\varepsilon_t,$$

$$h_t = h_{t-1} + \frac{\phi}{2h_{t-1}}\left(q - \frac{\sigma^2}{4\phi} - h_{t-1}^2\right) + \frac{\sigma}{2}\eta_t, \tag{9}$$

where ε_t and η_t are both $N(0,1)$ and $\mathrm{corr}(\varepsilon_t, \eta_t) = \rho$. Once the model is in this form, adapting the estimator to work with it again requires some modification of the Maple code used to obtain $p(x_t, v_t | x_{t-1}, v_{t-1})$ and its derivatives, but no changes to the main body of code. Results of a simulation study are shown in Table 1.

It seems plausible that the estimation approach used in this paper can be extended to work with affine models with jumps and/or more than a single volatility factor. However, such work is not undertaken here.

A summary of the models is shown in Table 2.

3. Numerical performance

While it is easy to demonstrate that the estimators used in this paper are asymptotically equivalent to the corresponding maximum likelihood estimators as S goes to infinity, (e.g., Gouriéroux and Monfort, 1996, Proposition 3.2), what is really needed is some understanding of the nature of the approximation error given the simulation lengths used in practice.

Recall that

$$\begin{aligned}
L(\theta|x) &= \int p(x,v)\,\mathrm{d}v \\
&= \int \frac{p(x,v)}{q(v)}\,\mathrm{d}Q(v) \\
&\quad \frac{1}{S}\sum_{s=1}^{S} \frac{p(x,v^{(s)})}{q(v^{(s)})},
\end{aligned} \tag{10}$$

where $v^{(1)}, \ldots, v^{(S)}$ are iid samples from the importance density q and the dependence of both p and q on θ is suppressed in the notation. The almost sure convergence of the sum follows from the strong law of large numbers since $L(\theta|x) < \infty$ by definition. Note that the choice of q does not come into play here.

A central limit theorem can be applied to get $\sqrt{S}$ convergence if additionally

$$E_Q\left[\left(\frac{p(x,V)}{q(V)}\right)^2\right] = \int \left(\frac{p(x,v)}{q(v)}\right)^2 \mathrm{d}Q(v) < \infty.$$

It is generally difficult to assess whether this integral is finite analytically, but a numerical investigation into the issue is straightforward. In particular, it is easy to

Table 2
Models

SV1-EUL

$$X_t = \mu + \sigma_X \exp(V_{t-1}/2)\varepsilon_t \qquad\qquad \varepsilon_t, \eta_t \quad N(0,1)$$
$$V_t = \phi V_{t-1} + \sigma_V \eta_t \qquad\qquad\qquad \text{corr}(\varepsilon_t, \eta_t) = \rho$$

SV1-JPR

$$X_t = \mu + \sigma_X \exp(V_t/2)\varepsilon_t \qquad\qquad \varepsilon_t, \eta_t \quad N(0,1)$$
$$V_t = \phi V_{t-1} + \sigma_V \eta_t \qquad\qquad\qquad \text{corr}(\varepsilon_t, \eta_t) = \rho$$

SV-t

$$X_t = \mu + \sigma_X \exp(V_{t-1}/2)\varepsilon_t \qquad\qquad \varepsilon_t \quad t_v, \psi \quad \eta_t \quad N(0,1)$$
$$V_t = \phi V_{t-1} + \sigma_V\left(\rho\varepsilon_t + \sqrt{1-\rho^2}\eta_t\right) \qquad \text{corr}(\varepsilon_t, \eta_t) = 0$$

SV2-EUL

$$X_t = \mu + \sigma_X \exp(U_{t-1}/2 + V_{t-1}/2)\varepsilon_{1t} \qquad \varepsilon_{it} \quad N(0,1)$$
$$V_t = \phi_V V_{t-1} + \sigma_V \varepsilon_{2t} \qquad\qquad\qquad \text{corr}(\varepsilon_{it}, \varepsilon_{jt}) = \rho_{ij}$$
$$U_t = \phi_U U_{t-1} + \sigma_U \varepsilon_{3t}$$

SV2-JPR

$$X_t = \mu + \sigma_X \exp(U_t/2 + V_t/2)\varepsilon_{1t} \qquad\quad \varepsilon_{it} \quad N(0,1)$$
$$V_t = \phi_V V_{t-1} + \sigma_V \varepsilon_{2t} \qquad\qquad\qquad \text{corr}(\varepsilon_{it}, \varepsilon_{jt}) = \rho_{ij}$$
$$U_t = \phi_U U_{t-1} + \sigma_U \varepsilon_{3t}$$

SV2-HYB

$$X_t = \mu + \sigma_X \exp(U_t/2 + V_{t-1}/2)\varepsilon_{1t} \qquad \varepsilon_{it} \quad N(0,1)$$
$$V_t = \phi_V V_{t-1} + \sigma_V \varepsilon_{2t} \qquad\qquad\qquad \text{corr}(\varepsilon_{it}, \varepsilon_{jt}) = \rho_{ij}$$
$$U_t = \phi_U U_{t-1} + \sigma_U \varepsilon_{3t}$$

AFF

$$X_t = \mu + V_{t-1}\left(\sqrt{1-\rho^2}\varepsilon_t + \rho\eta_t\right) \qquad \varepsilon_t, \eta_t \quad N(0,1)$$
$$V_t = V_{t-1} + \frac{\phi}{2V_{t-1}}\left(\phi - \frac{\sigma^2}{4\phi} - V_{t-1}^2\right) + \frac{\sigma}{2}\eta_t \qquad \text{corr}(\varepsilon_t, \eta_t) = 0$$

---

obtain a large number of draws from $p(x, V)/q(V)$, where V has density q. Suppose that the right tail of $\log(p(x, V)/q(V))$ is thinner than that of Z, a Gaussian random variable with the same mean and variance. Since $\exp(Z)$ has a finite second moment, one could conclude that $p(x, V)$ does also.

Note that we may write

$$\log \frac{p(x, V)}{q(V)} = \log p(x) + \log\left(\frac{p(V_1|x)}{q(V_1)}\right) + \sum_{i=1}^{n-1} \log\left(\frac{p(V_{i+1}|V_i, x)}{q(V_{i+1}|V_i)}\right).$$

Since this is a sum of many random variables, it seems plausible to hope that $\log(p(x, V)/q(V))$ might be approximately normally distributed. Fig. 1 shows
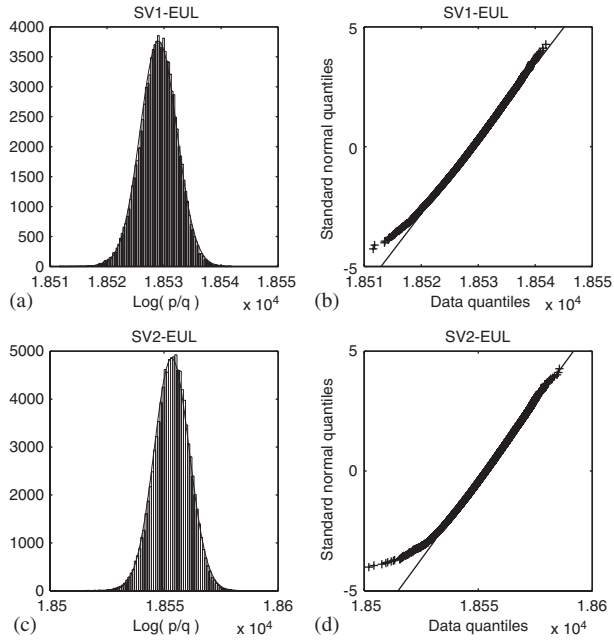
Fig. 1. Histograms and QQ-plots for 100,000 draws from $\log(p(x, V)/q(V))$, where V has the importance density, q.

histograms and QQ-plots of 100,000 draws from $\log(p(x, V)/q(V))$ using the data and two of the models considered in Section 5. These figures are supportive of the argument outlined above. It seems reasonable to expect that the sum in (10) converges at rate $\sqrt{S}$, at least eventually.

On the other hand, if the variance of $\log(p(x, V)/q(V))$ is large, then $p(x, V)/q(V)$ will be severely skewed, and "eventually" could be a long time coming. Furthermore, between the computation of the likelihood and the evaluation of the estimator lies an optimization step. Therefore, the preceding argument, while possibly of interest, is less than conclusive.

The simplest way to address the issue of simulation error is possibly the best: re-estimate the model many times using different seeds for the random number generator. This provides direct and unambiguous evidence on the variability of the estimator as a function of the sequence of random numbers used.

Ultimately, what is needed is a single parameter estimate for a given data set. Rather than arbitrarily choosing the estimate corresponding to one particular seed, it is better to report the mean of the entire collection of estimates computed. The standard deviation of the individual estimates serves as an indication of the simulation error. Since the variance of the mean will be much less than the variance of the individual estimates, this will be a conservative estimate of the numerical error associated with the parameter estimate actually reported.

The simulation estimator may also be biased with respect to the true maximum likelihood estimator (this is a different issue from possible bias in the MLE itself). A

careful analysis requires repeating the procedure described above for each of several different settings for S. The results can be used to see if there is any drift in the estimates as S increases.

Table 3 shows the results from trying this idea using the S&P 500 data examined in Section 5. The mean and standard deviation of the estimates obtained using each of 500 different seeds are shown. Several different values for S are tried. Histograms of the estimates provide further insight and are available upon request.

For the single-factor model, the estimator is reasonably stable even with a small number of draws. This is fortunate, because if $\sqrt{S}$ convergence is setting in, it is doing so slowly. Notice that there is some movement in the mean of the estimates with increasing S. This suggests the presence of a small amount of simulation-induced bias that disappears as S gets large.

For the two-factor model, there is a moderate amount of variation across simulations in the estimates for several of the parameters. There is also a shift in location for some of the estimates with increasing S, especially for ϕ_U and ρ_{31}. It seems likely that larger settings for S would yield estimates closer to zero for both of these parameters.

The issue of numerical error should be addressed with any simulation-based estimator, including simulated method of moments and Bayesian MCMC estimators. The advantage of the techniques used in this paper is that they are efficient enough that a careful investigation of numerical issues is possible.

4. Smoothing, filtering and diagnostics

It is straightforward to obtain estimates of the smoothed volatilities, $\mathrm{E}(V_1, \ldots, V_n | x_1, \ldots, x_n)$. One again simulates many draws from the importance sampler q. For each sample path, $v^{(s)}$, one computes the weight

$$w_s = \frac{p(x, v^{(s)})/q(v^{(s)})}{\sum_{i=1}^{S} p(x, v^{(i)})/q(v^{(i)})}, \quad s = 1, \ldots, S.$$

The collection of sample paths and weights may be thought of as defining a discrete probability measure which approximates that of $V_1, \ldots, V_n | x_1, \ldots, x_n$. The smoothed volatilities are estimated by the mean of the approximating distribution,

$$\bar{V} = \sum_{s=1}^{S} v^{(s)} w_s.$$

Other expectations, such as the variance, can be estimated in a similar manner.

While this approach to smoothing is simple, it is not very efficient. For the two-factor SV models, numerical error in the smoothed volatility is still readily apparent over test runs using 10,000 sample paths each (one can check the numerical error by repeating the exercise several times with different seeds for the random number generator). Using one million paths seems to be sufficient to reduce the numerical error to insignificance. On a data set with 5000 observations, computational cost is

Table 3
Numerical error

SV1-EUL

S	$\log L$	μ	σ_X	ϕ	σ_V	ρ
32	18,532.62	0.0000693	0.0085533	0.97663	0.1685	−0.4331
		(0.0001059)	(0.0004175)	(0.00442)	(0.0144)	(0.0476)
	(1.15)	(0.0000018)	(0.0000040)	(0.00048)	(0.0021)	(0.0030)
128	18,533.21	0.0000688	0.0085516	0.97633	0.1699	−0.4323
		(0.0001059)	(0.0004182)	(0.00439)	(0.0143)	(0.0475)
	(0.94)	(0.0000015)	(0.0000033)	(0.00042)	(0.0019)	(0.0026)
512	18,533.41	0.0000685	0.0085508	0.97616	0.1707	−0.4320
		(0.0001059)	(0.0004088)	(0.00460)	(0.0149)	(0.0468)
	(0.64)	(0.0000013)	(0.0000030)	(0.00039)	(0.0017)	(0.0021)
2048	18,533.58	0.0000684	0.0085496	0.97601	0.1714	−0.4314
		(0.0001059)	(0.0004145)	(0.00445)	(0.0144)	(0.0475)
	(0.45)	(0.0000010)	(0.0000022)	(0.00030)	(0.0014)	(0.0017)
8192	18,533.73	0.0000682	0.0085487	0.97585	0.1722	−0.4310
		(0.0001059)	(0.0004140)	(0.00449)	(0.0146)	(0.0476)
	(0.37)	(0.0000008)	(0.0000021)	(0.00029)	(0.0013)	(0.0014)
32,768	18,533.78	0.0000681	0.0085482	0.97578	0.1725	−0.4307
		(0.0001059)	(0.0004116)	(0.00454)	(0.0147)	(0.0474)
	(0.24)	(0.0000006)	(0.0000018)	(0.00024)	(0.0011)	(0.0012)

SV2-EUL

S	$\log L$	μ	σ_X	ϕ_V	σ_V	ρ_{21}	ϕ_U	σ_U	ρ_{31}
32	18,572.09	0.000087	0.00836	0.9913	0.0957	−0.429	0.311	0.424	−0.279
		(0.000104)	(0.00055)	(0.0029)	(0.0121)	(0.061)	(0.199)	(0.033)	(0.069)
	(2.62)	(0.000006)	(0.00006)	(0.0012)	(0.0075)	(0.042)	(0.204)	(0.036)	(0.073)
128	18,573.86	0.000085	0.00833	0.9909	0.0984	−0.445	0.236	0.445	−0.249
		(0.000104)	(0.00071)	(0.0025)	(0.0139)	(0.105)	(0.137)	(0.052)	(0.077)
	(2.34)	(0.000007)	(0.00005)	(0.0010)	(0.0065)	(0.035)	(0.186)	(0.030)	(0.065)
512	18,575.34	0.000083	0.00831	0.9906	0.1005	−0.457	0.167	0.462	−0.223
		(0.000103)	(0.00063)	(0.0025)	(0.0120)	(0.074)	(0.152)	(0.042)	(0.072)
	(2.10)	(0.000007)	(0.00004)	(0.0009)	(0.0057)	(0.027)	(0.169)	(0.022)	(0.056)
2048	18,576.59	0.000081	0.00829	0.9904	0.1016	−0.465	0.132	0.473	−0.208
		(0.000103)	(0.00062)	(0.0025)	(0.0115)	(0.068)	(0.168)	(0.040)	(0.074)
	(1.89)	(0.000006)	(0.00003)	(0.0008)	(0.0051)	(0.023)	(0.146)	(0.016)	(0.047)

Each model is estimated 500 times for each setting of S, each time using a different seed for the random number generator used to construct the simulations. The S&P 500 data is used. The mean of the estimates, the standard error corresponding to one particular set of estimates, and the standard deviation of the 500 individual estimates are reported.

about 1 h (2 GHz PC). For the single-factor SV models, the sampler is more efficient and 100,000 sample paths appear to be enough. This runs in a matter of several minutes.

Using the MCMC approach to sample directly from the density of $V|X$ may be a more efficient solution to the smoothing problem (e.g., Kim et al., 1998; Shephard and Pitt, 1997; Eraker, 2001; Jacquier et al., 1994). On the other hand, the approach described above may be less costly in terms of programming effort, which could ultimately be more important.

In practice, one may be more interested in the filtered volatilities, $E(V_t|\mathscr{F}_t)$, where $\mathscr{F}_t$ denotes the information set generated by $X_1, \ldots, X_t$. One way to obtain these is by means of a particle filter (e.g., Gordon et al., 1993 or Pitt and Shephard, 1999). Kim et al. (1998) use a particle filter for a single factor model without leverage effect. The version described below works for the one- and two-factor models with leverage effects and various timings considered in this paper.

A particle filter is comprised of a collection of discrete probability distributions $\hat{F}(v_t|\mathscr{F}_t)$ that approximate the exact densities $F(v_t|\mathscr{F}_t)$. For each t, the approximating density is defined by a collection of points $v_t^{(s)}$ and probability weights $w_t^{(s)}$, $s = 1, \ldots, S$. These are constructed recursively.

Heuristically, the idea at each step is to draw "particles" from the time t filter $\hat{F}(v_t|\mathscr{F}_t)$, advance the particles by drawing from $F(v_{t+1}|v_t, \mathscr{F}_t)$, and then weight to adjust for the new information implied by X_{t+1}.

More formally, suppose that $\hat{p}(v_t|\mathscr{F}_t)$ is known and the goal is to obtain $\hat{p}(v_{t+1}|\mathscr{F}_{t+1})$. First, notice that

$$
\begin{aligned}
p(v_{t+1}|\mathscr{F}_{t+1}) &= \int p(v_{t+1}, v_t|\mathscr{F}_{t+1})\,dv_t \\
&= \int \frac{p(v_{t+1}, v_t|\mathscr{F}_{t+1})}{p(v_t|\mathscr{F}_t)}\,dP(v_t|\mathscr{F}_t).
\end{aligned}
\tag{11}
$$

Also, from $p(v_{t+1}, v_t, x_{t+1}|\mathscr{F}_t) = p(v_{t+1}, v_t|\mathscr{F}_{t+1})p(x_{t+1}|\mathscr{F}_t)$, we get

$$
\begin{aligned}
p(v_{t+1}, v_t|\mathscr{F}_{t+1}) &= \frac{p(v_{t+1}, v_t, x_{t+1}|\mathscr{F}_t)}{p(x_{t+1}|\mathscr{F}_t)} \\
&= \frac{p(x_{t+1}|v_{t+1}, v_t, \mathscr{F}_t)p(v_{t+1}|v_t, \mathscr{F}_t)p(v_t|\mathscr{F}_t)}{p(x_{t+1}|\mathscr{F}_t)}.
\end{aligned}
\tag{12}
$$

Plugging (12) into (11) gives

$$
p(v_{t+1}|\mathscr{F}_{t+1}) = \int \frac{p(x_{t+1}|v_{t+1}, v_t, \mathscr{F}_t)p(v_{t+1}|v_t, \mathscr{F}_t)}{p(x_{t+1}|\mathscr{F}_t)}\,dP(v_t|\mathscr{F}_t).
$$

Thus, to advance the filter, first draw a point $v_t^{(s)}$ from $\hat{p}(v_t|\mathscr{F}_t)$, then draw $v_{t+1}^{(s)}$ from $p(v_{t+1}|v_t, \mathscr{F}_t)$. Repeat for $s = 1, \ldots, S$. These are the new particles. The weights are given by

$$
w^{(s)} = p(x_{t+1}|v_{t+1}^{(s)}, v_t^{(s)}) \Big/ \sum_{s=1}^{S} p(x_{t+1}|v_{t+1}^{(s)}, v_t^{(s)}).
$$

Note that the conditioning must be on both v_{t+1} and v_t since the innovations in the observed and unobserved components of the model may be correlated. Without correlation, the preceding argument would be much simpler.

The algorithm described above is reasonably efficient, can be written in just a few lines of Fortran code (available on request), and works for all of the models considered in this paper. More efficient implementations are possible.

A standard approach to specification analysis of time series models is to look at the residuals. Of interest is their unconditional distribution and dynamic structure. Due to the presence of the latent factor, it is not obvious how to go about doing this in the current setting. However, the construction of the particle filter suggests the following idea. The density of $X_{t+1}|\mathscr{F}_t$ can be estimated by

$$\hat{p}(x_{t+1}|\mathscr{F}_t) = \frac{1}{S}\sum_{s=1}^{S} p(x_{t+1}|v_{t+1}^{(s)\leftarrow}, v_t^{(s)}).$$

Similarly, its cdf can be estimated by

$$z_t = \text{prob}(X_{t+1} \leqslant x_{t+1}|\mathscr{F}_t) = \frac{1}{S}\sum_{s=1}^{S} \text{prob}(X_{t+1} \leqslant x_{t+1}|v_{t+1}^{(s)\leftarrow}, v_t^{(s)}).$$

If the model is correctly specified, these quantities should be iid uniform$(0, 1)$. While it would be possible to base analysis directly upon these, it is useful first to transform them by the inverse of the normal cumulative distribution function, $\tilde{z}_t = \Phi^{-1}(z_t)$. For a correctly specified model, these generalized residuals should be iid $N(0, 1)$.

This paper uses the Jarque–Bera test to assess the unconditional distribution of $\tilde{z}_t$ for the various models under consideration. The Box–Pierce test and the standard LM test for ARCH behavior (e.g., Greene, 2002) are used to look for dynamic structure. The Box–Pierce test is done on the squared residuals, since this is the feature of the data that is of interest. It is a good idea to try computing these statistics several times using different sequences of random numbers to construct the particle filter. If the statistics differ significantly across replications, more precision will be needed in the filter. Although 10,000 particles were enough to obtain good estimates of the filtered volatilities, the test statistics were not sufficiently stable across replications. The results reported in Section 5 were constructed using 100,000 particles for the single-factor models and one million particles for the two-factor models.

Kim et al. (1998) use a similar approach to assess a single-factor model without leverage effect. However, they look at $z_t = \text{prob}(X_{t+1}^2 \leqslant x_{t+1}^2|\mathscr{F}_t)$. This formulation, which makes it impossible to disentangle the right and left tails of the distribution, seems to be less useful. Similar diagnostics are also used by Liesenfeld and Richard (2003), but again only for models with uncorrelated errors. They use the Kolmogorov–Smirnoff test rather than Jarque–Bera to assess the unconditional density. But that test turns out to have little power in this context. They also use a different approach for filtering. This approach to model diagnostics is also discussed by Berkowitz (2001) and Bontemps and Meddahi (2005) among others, but not applied in the context of models with latent state variables.

Unfortunately, there does not appear to be an easy way to compute these diagnostics for the SV-t model. Obtaining z_t requires that one first evaluate the density $p(x_{t+1}|v_{t+1}, v_t)$. But for the SV-t model, this involves the density of a linear combination of normal and t random variables, which is not available analytically (note that this problem does not come up if $\rho = 0$). It should be possible to overcome the numerical problems, but doing so would be computationally costly and is not addressed in this paper.

It would be possible to undertake a similar analysis using smoothed rather than filtered residuals, i.e., with

$$w_t = \text{prob}(X_t \leqslant x_t | x_1, \ldots, x_{t-1}, x_{t+1}, \ldots, x_n).$$

However, this approach seems to be less useful. These is no reason to expect these w_t to be independent. Also, the smoothing effect causes the Jarque–Bera test to be badly sized (the test almost never rejects).

5. Application

5.1. S&P 500

In this subsection, some of the models described in Section 2 are estimated over daily S&P 500 index returns from June 23, 1980 to September 2, 2002 ($N = 5616$). The data are plotted in Figs. 2 and 3.

The data exhibit a small amount of autocorrelation, possibly due to non-synchronous trading of the individual stocks comprising the index. One way to remove this correlation is by passing the data through an ARMA filter. This is the approach taken by, for example, Andersen et al. (2002). An alternative approach would be to include an additional factor to capture mean dynamics (e.g., Chernov et al., 2003).

The empirical results reported in this subsection are all based on data that has been prefiltered using an ARMA(2, 1) model. Whether filtered or unfiltered data are used makes little difference in either the parameter estimates or diagnostics. The models under consideration are summarized in Table 2. Parameter estimates are shown in Table 4.

The first thing to notice is that in all of the models ρ is highly significant (leverage effect). This is important because this parameter is often set to zero for reasons of computational simplicity. The results of such studies are likely to be of limited practical use.

For the SV1 models, the JPR timing provides a significant improvement over the Euler scheme timing (model comparisons are based on Kullback–Leibler information, i.e., the difference in log likelihood; note that both models have the same number of free parameters). This is supportive of the Jacquier et al. (1994) argument that the non-Gaussianity introduced by this timing is empirically useful. On the other hand, Yu (2004) finds in favor of the Euler-scheme timing (using S&P 500 index data from 1980–87 and a Bayesian estimator), so this result may not be robust.
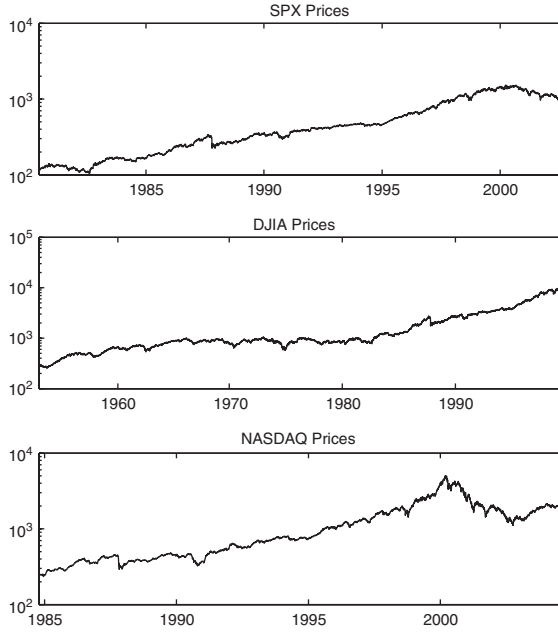
Fig. 2. (a) S&P 500 index prices, 6/24/1980–9/2/2002; (b) DJIA prices, 1/2/1953–7/30/1999; and (c) Nasdaq prices, 10/11/1984–9/15/2004.

In practice, it may not make much difference which version of the model one uses: parameter estimates as well as the associated forecasts and diagnostics are similar either way. Given that the persistence of the volatility factor is over 0.98, it is not surprising that the timing issue makes little difference.

For the two-factor models, the Kullback–Leibler information criterion suggests that there is little reason to prefer one of these models over the others, at least on this data set. Nonetheless, SV2-EUL does slightly better than the alternatives; given its appealing theoretical properties, it should probably be considered the preferred model.

The parameter ϕ_U governs the persistence of the less persistent factor and ρ_{31} controls the size of the leverage effect with respect to this factor. For all of the two-factor models, these parameters are marginally significant at best. For SV2-EUL, ϕ_U is not significantly different from zero at any conventional significance level, while ρ_{31} has a p-value of 0.048. Table 4 also includes estimates for SV2-EUL with both of these parameters pinned to zero. Although the likelihood ratio test rejects the restriction, the Bayesian information criterion prefers the smaller model (the reduction in log likelihood is around 6 points on over 5000 observations).

Note that SV2-EUL with $\phi_U = \rho_{31} = 0$ may be rewritten as

$$X_t = \mu + \sigma_X \exp(V_{t-1}/2)v_t,$$
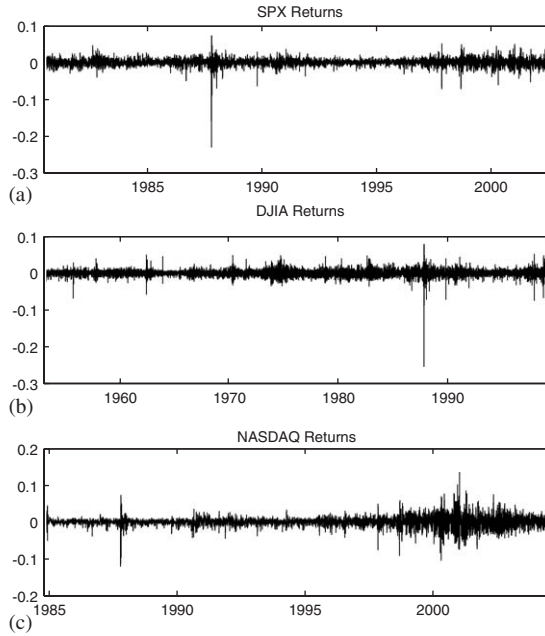$$V_t = \phi V_{t-1} + \sigma_V \zeta_t, \tag{13} \leftarrow$$

Fig. 3. (a) S&P 500 index returns, 6/24/1980–9/2/2002; (b) DJIA returns, 1/2/1953–7/30/1999; and (c) Nasdaq returns, 10/11/1984–9/15/2004.

Table 4
Parameter estimates, S&P 500

Model	Log L	μ	σ_X	ϕ	σ_V	ρ	v
SV1-EUL	18,533.34	0.00007	0.00855	0.9762	0.170	−0.432	
		(0.00011)	(0.00041)	(0.0045)	(0.015)	(0.047)	
	(0.87)	(0.00000)	(0.00000)	(0.0004)	(0.002)	(0.002)	
SV1-JPR	18,538.84	0.00024	0.00872	0.9828	0.148	−0.462	
		(0.00010)	(0.00050)	(0.0035)	(0.012)	(0.049)	
	(0.62)	(0.00000)	(0.00001)	(0.0002)	(0.001)	(0.003)	
SV1	18,503.12	0.00025	0.00862	0.9816	0.151	*	
		(0.00010)	(0.00048)	(0.0040)	(0.013)		
	(0.77)	(0.00000)	(0.00000)	(0.0003)	(0.002)		
SVt	18,582.24	0.00008	0.00755	0.9895	0.103	−0.488	8.09
		(0.00010)	(0.00052)	(0.0026)	(0.011)	(0.052)	(0.89)
	(0.11)	(0.00000)	(0.00000)	(0.0000)	(0.000)	(0.001)	(0.01)
SVt	18,553.50	0.00020	0.00773	0.9934	0.085	*	7.68
		(0.00010)	(0.00068)	(0.0022)	(0.010)		(0.83)
	(0.07)	(0.00000)	(0.00000)	(0.0000)	(0.000)		(0.01)

Table 4 (*continued*)

	Log L	μ	σ_X	ϕ_V	σ_V	ρ_{21}	ϕ_U	σ_U	ρ_{31}
SV2-EUL	18,575.86	0.00008	0.00830	0.9905	0.101	−0.459	0.15	0.468	−0.215
		(0.00010)	(0.00063)	(0.0027)	(0.014)	(0.088)	(0.27)	(0.066)	(0.112)
	(1.81)	(0.00001)	(0.00003)	(0.0008)	(0.005)	(0.020)	(0.14)	(0.017)	(0.048)
SV2-EUL	18,569.78	0.00008	0.00827	0.9885	0.113	−0.506	*	0.468	*
		(0.00010)	(0.00055)	(0.0027)	(0.011)	(0.053)		(0.032)	
	(1.82)	(0.00001)	(0.00002)	(0.0004)	(0.002)	(0.011)		(0.009)	
SV2-JPR	18,575.21	0.00011	0.00838	0.9878	0.119	−0.536	−0.25	0.439	0.076
		(0.00011)	(0.00055)	(0.0030)	(0.012)	(0.051)	(0.11)	(0.037)	(0.036)
	(2.06)	(0.00001)	(0.00002)	(0.0005)	(0.003)	(0.011)	(0.05)	(0.014)	(0.008)
SV2-HYB	18,572.85	0.00011	0.00837	0.9880	0.118	−0.527	−0.19	0.449	−0.066
		(0.00011)	(0.00055)	(0.0030)	(0.012)	(0.051)	(0.15)	(0.036)	(0.033)
	(1.97)	(0.00001)	(0.00002)	(0.0005)	(0.003)	(0.012)	(0.06)	(0.013)	(0.005)
SV2	18,543.92	0.00020	0.00833	0.9921	0.093	*	−0.13	0.493	*
		(0.00010)	(0.00067)	(0.0024)	(0.010)		(0.12)	(0.034)	
	(2.08)	(0.00000)	(0.00001)	(0.0003)	(0.002)		(0.04)	(0.011)	

	Log L	μ	α	ϕ	σ	ρ
AFF	18,473.44	0.00003	0.000100	0.0186	0.00129	−0.464
		(0.00011)	(0.000009)	(0.0039)	(0.00013)	(0.051)
	(2.49)	(0.00001)	(0.000000)	(0.0010)	(0.00004)	(0.010)

The data are prefiltered using an ARMA(2,1) model. The sample period is 6/23/1980–9/2/2002 ($N = 5616$ observations). A '*' indicates that the parameter has been fixed at zero. The SV1 and SV-t models are estimated using $S = 256$ draws from the importance sampler; the SV2 and affine models use $S = 4024$. Each model is estimated 100 times using different seeds for the random number generator used to construct the simulations (as described in Section 3). The mean of these estimates is reported. Standard errors are immediately below each set of estimates. Finally, the standard deviation of the 100 individual estimates is reported (this provides information as to the numerical uncertainty of the estimates).

where $v_t = \exp(U_{t-1}/2)\varepsilon_{1t}$ is an iid scale mixture of normals. The volatility innovations, ζ_t, are normal. The SV-t model is also of form (13). The return innovations are iid scale mixtures of normals (but with a different mixing distribution) and each volatility innovation is a linear combination of a normal and a t. Given that neither ϕ_U nor ρ_{31} differ from zero by much in the two-factor model, it is not surprising that the performance of the SV-t and SV2 models is similar. The particular form of the mixing distributions implied by these two alternatives does not make a great deal of difference empirically. Nonetheless, the SV-t model dominates. This is not meant to imply that the SV-t model is the

"correct" model; only that it may have empirical advantages (as well as being much easier to estimate) compared to the commonly-used SV2 model, at least on this data set.

Fig. 4 shows predictive densities for $X_t|V_{t-1}$ for several models. The densities are computed by fixing $V_{t-1} = \bar{\theta}$ and integrating across $V_t, U_t, U_{t-1}|V_{t-1}$. For SV1-EUL, this predictive density is exactly Gaussian. Using the JPR timing adds a small amount of skewness. The SV-t model fattens the tails. There is not much difference in the predictive densities implied by the SV-t and various SV2 models. Note that all of these densities are close to symmetric and thus offer little in the way of explaining the occasional large negative returns present in the data.

Table 4 also displays estimates for a single-factor affine model. These results confirm the findings of Andersen et al. (2002) and Chernov et al. (2003) that this model is of little empirical relevance, at least for S&P index returns (while both of those papers find that including jumps in the affine model can greatly improve matters, such models pose additional difficulties that are beyond the scope of this paper).

The Box–Pierce and ARCH tests (see Table 5) suggest that even the simple single-factor models may be able to capture the volatility dynamics adequately. The issue of timing (SV1-JPR vs. SV1-EUL) makes little difference here. The correlograms in Figs. 5 and 6 tell much the same story.

Turning to the Jarque–Bera test, the results imply that none of the models is able to capture the shape of the returns distribution. The situation is more clearly
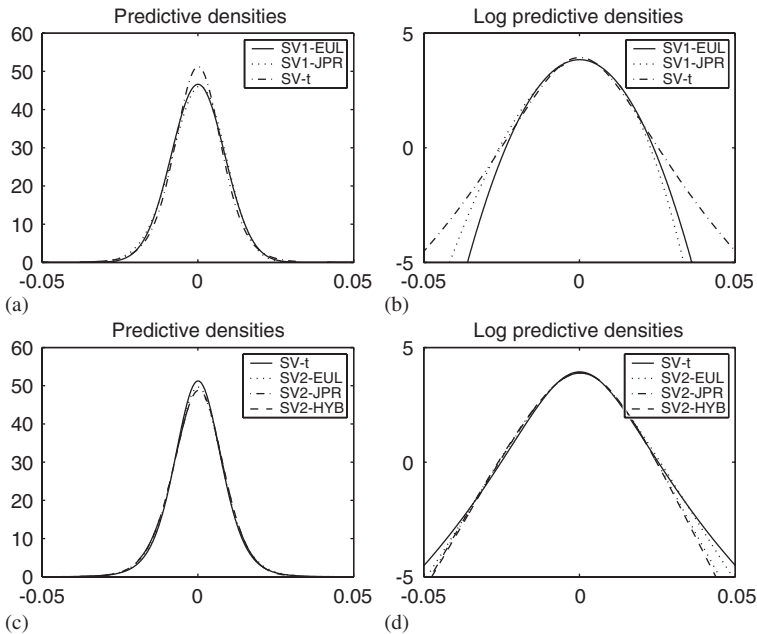

Fig. 4. Predictive densities for $X_t|V_{t-1}$, based on estimates for S&P 500 data.

Table 5
Model diagnostics, S&P 500

	Jarque–Bera	Kolmogorov–Smirnoff	Box–Pierce(20)	ARCH(20)
Raw data	439,771.42	0.071	272.70	368.60
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Affine	2722.82	0.028	59.79	63.74
	(0.0000)	(0.0002)	(0.0000)	(0.0000)
SV1-EUL	288.93	0.025	12.35	12.70
	(0.0000)	(0.0021)	(0.9036)	(0.8898)
SV1-JPR	168.42	0.023	12.66	13.08
	(0.0000)	(0.0050)	(0.8915)	(0.8740)
SV2-EUL	42.56	0.015	20.50	21.04
	(0.0000)	(0.1637)	(0.4269)	(0.3948)
SV2-EUL ($\phi_U = \varphi_{31} = 0$)	53.05	0.014	20.91	21.54
	(0.0000)	(0.1917)	(0.4023)	(0.3662)
SV2-JPR	34.89	0.012	20.46	20.91
	(0.0000)	(0.4048)	(0.4298)	(0.4022)
SV2-HYB	30.26	0.012	17.86	18.27
	(0.0000)	(0.3743)	(0.5963)	(0.5695)

Test statistics and p-values are reported. The Jarque–Bera and Kolmogorov–Smirnoff tests assess the unconditional distribution of the generalized residuals. The Box–Pierce and ARCH tests check for dynamic structure. Both use 20 lags. The Box–Pierce test is based on the squared residuals.

illustrated by the QQ-plots in Figs. 7 and 8. Although including the second volatility factor helps somewhat, all of the models under consideration fail in a similar manner. None is able to capture the extreme left tail of the distribution. The two-factor models are all slightly too thick in the right tail; SV1-EUL gets the right tail almost perfectly. These findings are in accord with earlier work (e.g., Gallant et al., 1997).

The affine model does poorly in both dimensions (dynamic structure and distribution), as shown by the test statistics in Table 5 and more clearly by the graphics in Figs. 5 and 7.

5.2. DJIA

This subsection examines returns on the Dow Jones Industrial Average (DJIA) index from January 2, 1953–July 30, 1999 (see Figs. 2 and 3). This is the same data as
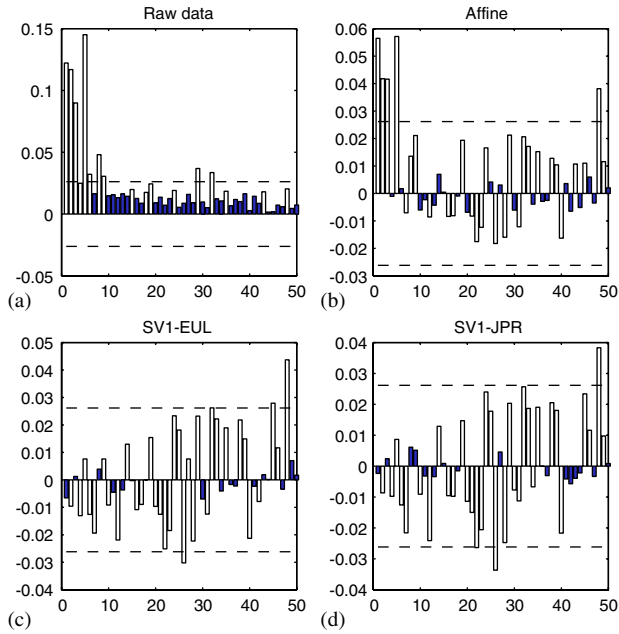
Fig. 5. Correlograms for squared generalized residuals, S&P 500 data. The dashed lines mark the rejection region (5% significance level) for the individual correlation coefficients.
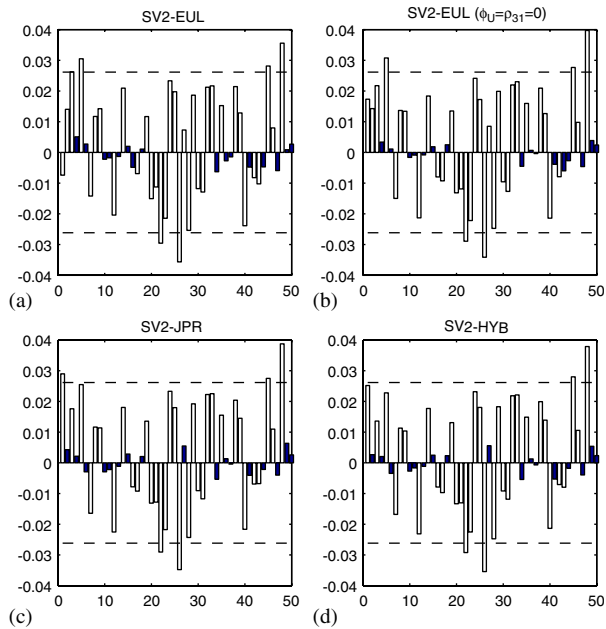

Fig. 6. Correlograms for squared generalized residuals, S&P 500 data. The dashed lines mark the rejection region (5% significance level) for the individual correlation coeffecients.
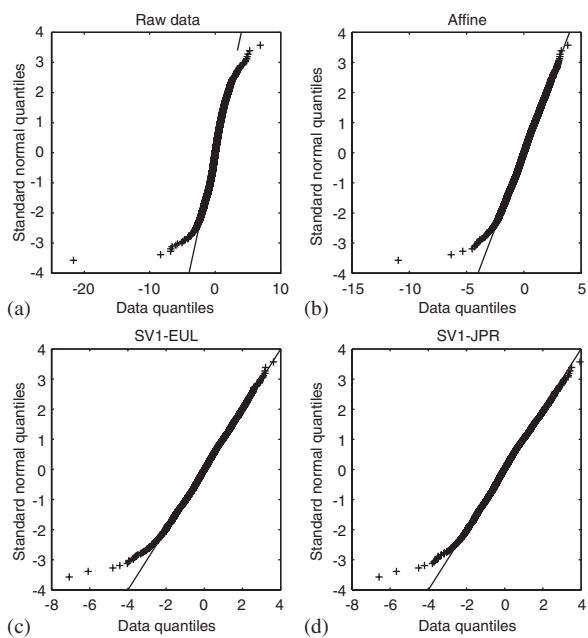
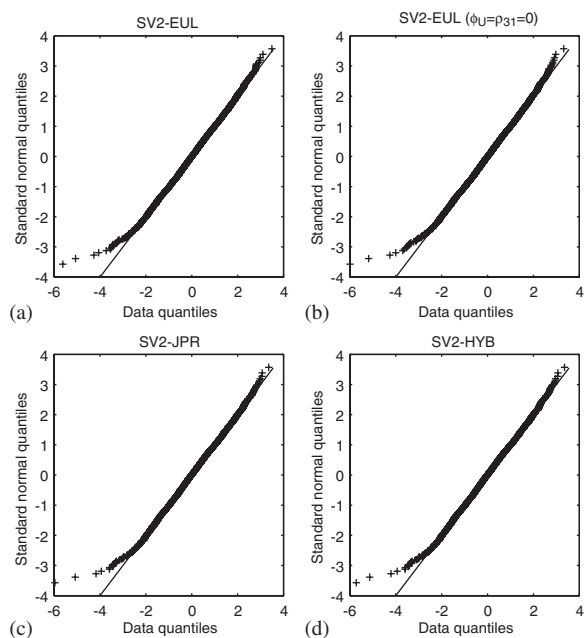Fig. 7. QQ-plots for generalized residuals, S&P 500 data.

Fig. 8. QQ-plots for generalized residuals, S&P 500 data.

used by Chernov et al. (2003). Andersen et al. (2002) use DJIA returns from January 1953 to December 1996. The data is first passed through an AR(2) pre-filter.

Overall, the findings are similar to those of the previous subsection. Since DJIA and S&P 500 returns are similar, the differences should be primarily due to the longer sample period of the DJIA data. Parameter estimates are shown in Table 6.

For the single-factor models, the JPR timing is again slightly better than the Euler-scheme timing, and both are soundly trounced by the SV-t model. This is the same relative ranking as with the S&P 500 data. The parameter estimates also differ by little between the two data sets.

Table 6
Parameter estimates, DJIA

Model	Log L	μ	σ_X	ϕ	σ_V	ρ	v		
SV1-EUL	40,693.19	0.00004	0.00723	0.9803	0.149	−0.441			
		(0.00006)	(0.00026)	(0.0027)	(0.009)	(0.033)			
	(0.87)	(0.00000)	(0.00000)	(0.0002)	(0.001)	(0.002)			
SV1-JPR	40,696.47	0.00018	0.00734	0.9847	0.136	−0.457			
		(0.00006)	(0.00030)	(0.0022)	(0.008)	(0.035)			
	(0.77)	(0.00000)	(0.00000)	(0.0001)	(0.001)	(0.002)			
SVt	40,741.26	0.00005	0.00663	0.9883	0.106	−0.491	11.84		
		(0.00006)	(0.00029)	(0.0019)	(0.008)	(0.036)	(1.26)		
	(0.27)	(0.00000)	(0.00000)	(0.0001)	(0.000)	(0.001)	(0.04)		

	Log L	μ	σ_X	ϕ_V	σ_V	ρ_{21}	ϕ_U	σ_U	ρ_{31}
SV2-EUL	40,756.43	0.00005	0.00724	0.9960	0.057	−0.283	0.84	0.238	−0.447
		(0.00006)	(0.00049)	(0.0013)	(0.009)	(0.088)	(0.03)	(0.020)	(0.047)
	(1.18)	(0.00000)	(0.00001)	(0.0003)	(0.002)	(0.022)	(0.01)	(0.007)	(0.011)
SV2-JPR	40,732.74	0.00023	0.00758	0.9978	0.042	−0.245	0.94	0.166	−0.411
		(0.00006)	(0.00069)	(0.0009)	(0.008)	(0.118)	(0.01)	(0.013)	(0.056)
	(1.01)	(0.00000)	(0.00002)	(0.0001)	(0.002)	(0.028)	(0.00)	(0.003)	(0.013)
SV2-HYB	40,744.91	0.00019	0.00743	0.9958	0.066	−0.463	0.89	0.187	−0.352
		(0.00006)	(0.00048)	(0.0022)	(0.015)	(0.053)	(0.05)	(0.023)	(0.061)
	(1.19)	(0.00000)	(0.00002)	(0.0004)	(0.003)	(0.011)	(0.01)	(0.005)	(0.012)

The data are prefiltered using an AR(2) model. The sample period is 1/2/1953–7/30/1999 ($N = 11,726$ observations). The SV1 and SV-t models are estimated using $S = 256$ draws from the importance sampler; the SV2 models use $S = 1024$. Each model is estimated 100 times using different seeds for the random number generator used to construct the simulations (as described in Section 3). The mean of these estimates is reported. Standard errors are immediately below each set of estimates. Finally, the standard deviation of the 100 individual estimates is reported (this provides information as to the numerical uncertainty of the estimates).

For the two-factor models, the Euler-scheme timing does much better than the others (in contrast to the S&P data, the differences between timings are substantial). For the model with this timing, the persistent volatility factor is about twice as persistent as was the case for the S&P 500 data (autoregressive coefficient of 0.996 versus 0.991). The volatility of this factor (σ_V) is about half as large as for the S&P 500 data (implying that the unconditional variance of this factor is about the same for both data sets). Whereas the "nonpersistent" factor had an autoregressive coefficient which was not significantly different from zero with the S&P data, it is highly significant (0.84 with a standard error of 0.03) for the DJIA data.

In contrast to the S&P data, SV2 beats SV-t on this dataset, though not by a huge amount (15 points in the log likelihood on over 10,000 observations). The problem appears to be the persistence of the second volatility factor with the DJIA data. The SV-t model has no way to capture this feature of the data.

Results of the diagnostic tests are similar to those for the S&P 500 data (see Table 7 and Figs. 9 and 10). For the ARCH test and the Box–Pierce test on the squared generalized residuals, SV1-JPR is rejected at the 5% significance level (p-values of 0.04 on both), but none of the other models are rejected at conventional levels. Examination of the correlograms in Fig. 9 provides supporting evidence that all of the models are again doing a good job at capturing the dynamics of volatility. The Jarque–Bera test and QQ-plots also tell much the same story as for the S&P 500 data: including the second volatility factor helps a small amount, but all of the models fail badly at matching the conditional distribution of the data.

Chernov et al. (2003) estimate continuous-time versions of the SV1 and SV2 models, but parameterized differently and with the addition of a stochastic mean factor (obviating the ARMA prefilter). They also find these models inadequate to fit the data and go on to fit a more complicated SV2 model with

Table 7
Model diagnostics, DJIA

	Jarque–Bera	Kolmogorov–Smirnoff	Box–Pierce(20)	ARCH(20)
SV1-EUL	660.309	0.013	24.517	25.006
	(0.0000)	(0.0427)	(0.2205)	(0.2012)
SV1-JPR	443.150	0.015	32.056	32.651
	(0.0000)	(0.0119)	(0.0427)	(0.0368)
SV2-EUL	266.277	0.009	16.129	16.672
	(0.0000)	(0.3607)	(0.7086)	(0.6742)
SV2-JPR	287.855	0.016	22.688	23.112
	(0.0000)	(0.0043)	(0.3044)	(0.2833)

Test statistics and p-values are reported. The Jarque–Bera and Kolmogorov–Smirnoff tests assess the unconditional distribution of the generalized residuals. The Box–Pierce and ARCH tests check for dynamic structure. Both use 20 lags. The Box–Pierce test is based on the squared residuals.
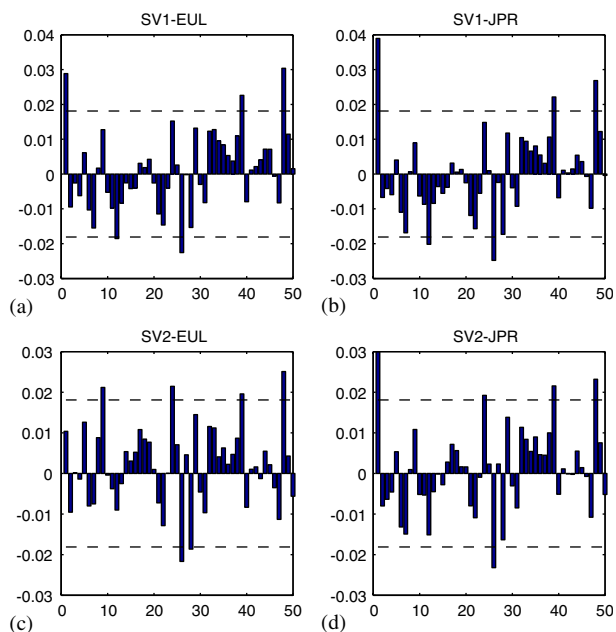
Fig. 9. Correlograms for squared generalized residuals, DJIA data. The dashed lines mark the rejection region (5% significance level) for the individual correlation coeffecients.
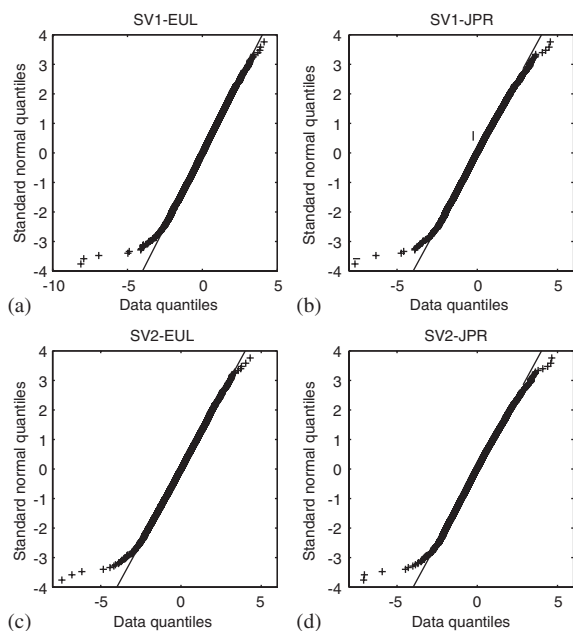

Fig. 10. QQ-plots for generalized residuals, DJIA data.

"volatility-in-volatility". They argue that this extension takes care of the conditional distribution problem.

5.3. Nasdaq

This subsection looks at Nasdaq returns from October 11, 1984 to September 15, 2004 (see Figs. 2 and 3). For both the one- and two-factor models, the JPR timing dominates on this data. SV1-JPR even beats SV-t (by almost 30 points in the log likelihood). In comparison to the S&P 500 and DJIA data, volatility is both slightly higher on average (σ_X) and more variable (σ_V and σ_U) (Table 8).

Table 8
Parameter estimates, NASDAQ

Model	Log L	μ	σ_X	ϕ	σ_V	ρ	v		
SV1-EUL	15,958.75	0.00047	0.00926	0.9851	0.197	−0.299			
		(0.00011)	(0.00087)	(0.0029)	(0.014)	(0.046)			
	(1.01)	(0.00000)	(0.00000)	(0.0002)	(0.002)	(0.003)			
SV1-JPR	16,001.09	0.00057	0.00956	0.9858	0.207	−0.480			
		(0.00011)	(0.00098)	(0.0027)	(0.013)	(0.036)			
	(0.91)	(0.00000)	(0.00001)	(0.0001)	(0.001)	(0.003)			
SVt	15,971.48	0.00053	0.0082	0.99115	0.150	−0.323	11.82		
		(0.00011)	(0.0010)	(0.00235)	(0.015)	(0.049)	(2.24)		
	(0.35)	(0.00000)	(0.0000)	(0.00017)	(0.002)	(0.002)	(0.21)		
	Log L	μ	σ_X	ϕ_V	σ_V	ρ_{21}	ϕ_U	σ_U	ρ_{31}
SV2-EUL	16,018.20	0.00054	0.00943	0.9982	0.064	−0.082	0.82	0.313	−0.473
	(0.00)	(0.00011)	(0.00263)	(0.0010)	(0.012)	(0.110)	(0.03)	(0.024)	(0.050)
	(1.33)	(0.00001)	(0.00010)	(0.0001)	(0.003)	(0.022)	(0.01)	(0.007)	(0.008)
SV2-JPR	16,058.90	0.00076	0.01004	0.9987	0.061	−0.175	0.89	0.266	−0.519
	(0.00)	(0.00011)	(0.00369)	(0.0008)	(0.011)	(0.111)	(0.02)	(0.019)	(0.044)
	(0.85)	(0.00000)	(0.00006)	(0.0001)	(0.002)	(0.023)	(0.00)	(0.004)	(0.008)
SV2-HYB	16,058.90	0.00076	0.01004	0.9987	0.061	−0.175	0.89	0.266	−0.519
	(0.00)	(0.00011)	(0.00369)	(0.0008)	(0.011)	(0.111)	(0.02)	(0.019)	(0.044)
	(0.85)	(0.00000)	(0.00006)	(0.0001)	(0.002)	(0.023)	(0.00)	(0.004)	(0.008)

The data are prefiltered using an AR(2) model. The sample period is 10/11/1984–9/15/2004 ($N = 5027$ observations). The SV1 and SV-t models are estimated using $S = 256$ draws from the importance sampler; the SV2 models use $S = 1024$. Each model is estimated 100 times using different seeds for the random number generator used to construct the simulations (as described in Section 3). The mean of these estimates is reported. Standard errors are immediately below each set of estimates. Finally, the standard deviation of the 100 individual estimates is reported (this provides information as to the numerical uncertainty of the estimates).

Table 9
Model diagnostics, NASDAQ

	Jarque–Bera	Kolmogorov–Smirnoff	Box–Pierce(20)	ARCH(20)
SV1-EUL	108.134	0.029	24.445	24.464
	(0.0000)	(0.0004)	(0.2235)	(0.2227)
SV1-JPR	14.598	0.013	21.798	21.985
	(0.0007)	(0.3230)	(0.3516)	(0.3413)
SV2-EUL	90.824	0.027	23.558	24.946
	(0.0000)	(0.0011)	(0.2623)	(0.2035)
SV2-JPR	7.429	0.010	25.416	25.708
	(0.0244)	(0.7359)	(0.1860)	(0.1757)

Test statistics and p-values are reported. The Jarque–Bera and Kolmogorov–Smirnoff tests assess the unconditional distribution of the generalized residuals. The Box–Pierce and ARCH tests check for dynamic structure. Both use 20 lags. The Box–Pierce test is based on the squared residuals.
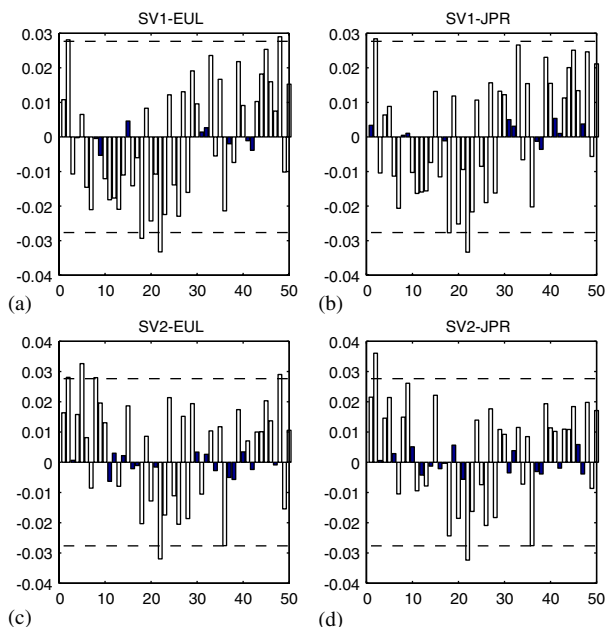

Fig. 11. Correlograms for squared generalized residuals, NASDAQ data. The dashed lines mark the rejection region (5% significance level) for the individual correlation coeffecients.
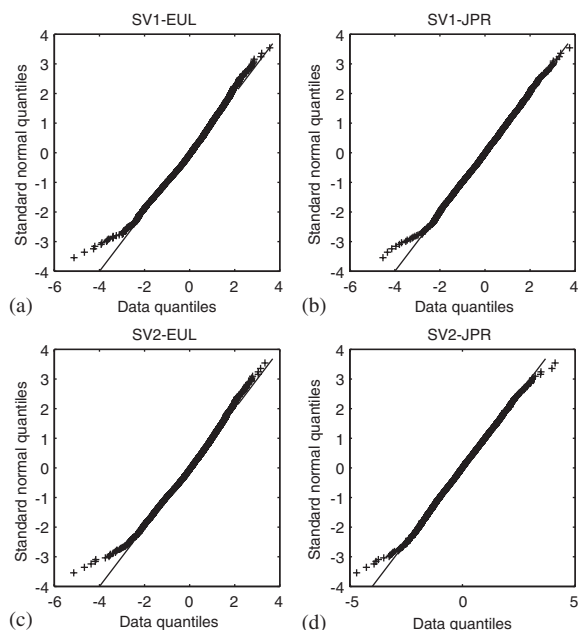
Fig. 12. QQ-plots for generalized residuals, NASDAQ data.

The first factor is yet more persistent than was the case with the DJIA data. Because this factor is so persistent, there is virtually no difference between the JPR and hybrid timings. Looking at the plot of returns in Fig. 3, there is a clear long-term trend in volatility: low in the late-1980s, increasing slowly up to around 2001, and then falling again. It is this pattern that the first factor captures. What one typically thinks of as volatility clustering is captured largely by the second factor, which has an autoregressive coefficient of 0.89 (SV2-JPR).

As with both the S&P 500 and DJIA data, the tests shown in Table 9 and QQ-plots in Fig. 11 suggest that all of the models perform well in capturing the dynamic structure of volatility. The models with the JPR timing do much better at fitting the conditional distribution of returns than was the case with either the S&P 500 or DJIA data (Jarque–Bera test in Table 9 and QQ-plots in Fig. 12). Indeed, SV2-JPR, while rejected at the 5% significance level, is not rejected at the 1% level.

6. Conclusion

This paper demonstrates some easily implemented tools for estimating and assessing one- and two-factor SV models. Monte Carlo studies demonstrating their small sample statistical properties and numerical properties are provided. Such studies are made feasible by the computational efficiency of the tools considered. The model diagnostics are based on standard time-series techniques.

The application looks at returns for the S&P 500, DJIA, and Nasdaq. In contrast to several preceding studies, I find no evidence that even the simple single-factor models are unable to capture the dynamics of the volatility process. The more critical problem is to capture the shape of the conditional returns distribution. Here, all of the models fall badly short.

Of course the usual caveats apply: failure to find evidence regarding the dynamics of the volatility process may be simply due to the inadequacy of the sample or lack of power of the tests.

Acknowledgements

References

Alizadeh, S., Brandt, M.W., Diebold, F.X., 2002. Range-based estimation of stochastic volatility models. Journal of Finance 57, 1047–1091.

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

Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P., 2003. Modeling and forecasting realized volatility. Econometrica 71, 579–625.

Barndorff-Nielsen, O.E., Shephard, N., 2002. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. Journal of the Royal Statistical Society, Series B 64, 253–280.

Berkowitz, J., 2001. Testing density forecasts, with applications to risk management. Journal of Business and Economic Statistics 19, 465–474.

Bollerslev, T., Mikkelsen, H., 1996. Modelling and pricing long-memory in stock market volatility. Journal of Econometrics 73, 151–184.

Bontemps, C., Meddahi, N., 2005. Testing normality: a GMM approach. Journal of Econometrics 124, 149–186.

Chernov, M., Gallant, A.R., Ghysels, E., Tauchen, G., 2003. Alternative models for stock price dynamics. Journal of Econometrics 116, 225–257.

Chib, S., Nardari, F., Shephard, N., 2002. Markov chain Monte Carlo methods for stochastic volatility methods. Journal of Econometrics 108, 281–316.

Danielson, J., 1994. Stochastic volatility in asset prices: estimation with simulated maximum likelihood. Journal of Econometrics 64, 374–400.

Danielson, J., Richard, J.-F., 1993. Accelerated Gaussian importance sampler with application to dynamic latent variable models. Journal of Applied Econometrics 8, 153–173.

Ding, Z., Granger, C., Engle, R., 1993. A long memory property of stock market returns and a new model. Journal of Empirical Finance 1, 83–108.

Duffie, D., Singleton, K.J., 1993. Simulated moments estimation of Markov models of asset prices. Econometrica 61, 929–952.

Durbin, J., Koopman, S., 1997. Monte Carlo maximum likelihood estimation for non-Gaussian state space models. Biometrika 84, 669–684.

Durbin, J., Koopman, S., 2000. Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives. Journal of the Royal Statistical Society, Series B 62, 3–56.

Durham, G.B., Gallant, A.R., 2002. Numerical techniques for simulated maximum likelihood estimation of stochastic differential equations. Journal of Business and Economic Statistics 20, 297–316.

Engle, R.F., Lee, G.G., 1998. A permanent and transitory component model of stock return volatility. In: Engle, R.F., White, H. (Eds.), Cointegration, Causality, and Forecasting: A Festschrift in Honor of Clive W.J. Granger. Oxford University Press, Oxford, pp. 475–497.

Eraker, B., 2001. MCMC analysis of diffusion models with application to finance. Journal of Business and Economic Statistics 19, 177–191.

Eraker, B., Johannes, M., Polson, N.G., 2003. The impact of jumps in volatility and returns. Journal of Finance 58, 1269–1300.

Fridman, M., Harris, L., 1998. A maximum likelihood approach for non-Gaussian stochastic volatility models. Journal of Business and Economic Statistics 16, 284–291.

Gallant, A.R., Tauchen, G.E., 1996. Which moments to match? Econometric Theory 12, 657–681.

Gallant, A.R., Hsieh, D., Tauchen, G.E., 1997. Estimation of stochastic volatility models with diagnostics. Journal of Econometrics 81, 159–192.

Gallant, A.R., Hsu, C.-T., Tauchen, G.E., 1999. Using daily range data to calibrate volatility diffusions and extract the forward integrated variance. Review of Economics and Statistics 81, 617–631.

Gelman, A., Carlin, J., Stern, H., Rubin, D., 1995. Bayesian Data Analysis. Chapman & Hall, London.

Geweke, J., 1989. Bayesian inference in econometric models using Monte Carlo integration. Econometrica 57, 1317–1339.

Geweke, J., Amisano, G., 2001. Compound Markov mixture models with applications in finance. Manuscript, University of Iowa.

Gordon, N., Salmond, D., Smith, A., 1993. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. IEEE Proceedings-F 140, 107–113.

Gouriéroux, C., Monfort, A., 1996. Simulation-Based Econometric Methods. Oxford University Press, Oxford, UK.

Greene, W.H., 2002. Econometric analysis, fifth ed. Prentice-Hall, Upper Saddle River, New Jersey.

Hamilton, J., 1990. Analysis of time series subject to changes in regime. Journal of Econometrics 45, 39–70.

Jacquier, E., Polson, N.G., Rossi, P.E., 1994. Bayesian analysis of stochastic volatility models. Journal of Business and Economic Statistics 12, 371–389.

Jacquier, E., Polson, N.G., Rossi, P.E., 2002. Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. Journal of Econometrics 122, 185–212.

Jones, C.S., 2003. The dynamics of stochastic volatility: evidence from underlying and options markets. Journal of Econometrics 116, 181–224.

Judd, K.L., 1998. Numerical Methods in Economics. MIT Press, Cambridge, MA.

Kim, S., Shephard, N., Chib, S., 1998. Stochastic volatility: likelihood inference and comparison with ARCH models. Review of Economic Studies 65, 361–393.

Kloek, T., van Dijk, H., 1978. Bayesian estimates of equation system parameters: an application of integration by Monte Carlo. Econometrica 46, 1–19.

Koopman, S.J., Jungbacker, B., Hol, E., 2005. Forecasting daily variability of the S&P 100 stock index using historical, realised, and implied volatility measurements. Journal of Empirical Finance, forthcoming.

Liesenfeld, R., Richard, J.-F., 2003. Univariate and multivariate stochastic volatility models: estimation and diagnostics. Journal of Empirical Finance 10, 505–531.

Pitt, M.K., Shephard, N., 1999. Filtering via simulation: auxiliary particle filter. Journal of the American Statistical Association 446, 590–599.

Sandmann, G., Koopman, S.J., 1998. Estimation of stochastic volatility models via Monte Carlo maximum likelihood. Journal of Econometrics 87, 271–301.

Shephard, N., Pitt, M., 1997. Likelihood analysis of non-Gaussian measurement time series. Biometrika 84, 653–667.

Yu, J., 2004. On leverage in a stochastic volatility model. Journal of Econometrics, forthcoming.