Likelihood-based estimation and specification analysis of one- and two-factor SV models with leverage effects [*]

Garland B. Durham[†]
University of Iowa

January 29, 2004

Abstract

Techniques for simulated maximum likelihood (SML) estimation, filtering, and assessing the fit of stochastic volatility models are examined. Both one- and two-factor models (with leverage effects) are considered. The techniques are computationally efficient, robust, straightforward to implement, and easy to adapt to new models. Using these techniques, it is possible to estimate single-factor models over data sets of several thousand observations in several seconds. The computational efficiency of the techniques means that Monte Carlo studies assessing both the small sample statistical properties as well as the numerical properties of the estimators are easy to do. Such studies are important for all simulation estimators, including simulation-based Bayesian and method of moments estimators. The application looks at S&P 500 index returns. Even the simple single-factor models adequately capture the dynamics of volatility; the problem is to get the shape of the returns distribution right. Although including a second volatility factor improves the fit over the basic single-factor models, a new formulation of the SV-t model (a single factor model, but with t rather than normal errors in the observation equation) provides the best fit. However, all the models considered fail in a similar manner: they are unable to capture the left tail of the distribution. Fitting this part of the distribution is important for option-pricing and risk management. Although it may be possible to come up with ad hoc parametric models that fit particular data series and sample periods, a promising alternative might be to look at single-factor models with flexible forms for the error distributions.

[†]Department of Economics, W376 Pappajohn Bldg., University of Iowa, Iowa City, IA 52242-1000, USA; Phone: 319 335 0844; email: garland-durham@uiowa.edu; web: www.biz.uiowa.edu/faculty/gdurham.

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 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 greatly complicated since the state is not uniquely determined by the data. An appealing feature of SV models is that they have a natural continuous-time interpretation, whereas ARCH/GARCH models are fundamentally discrete-time in nature. 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, Bollerslev, Diebold and Labys 2002b; Barndorff-Nielsen and Shephard 2002), hi-lo quotes (Alizadeh, Brandt and Diebold 2002) or option prices (Jones 1999; Hol and Koopman 2002).

This paper examines tools for likelihood-based analysis of standard one- and two-factor SV models used in finance, building on work by Shephard and Pitt (1997), Pitt and Shephard (1999), Durbin and Koopman (1997, 2000, 2001), 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 (2002), who also use Monte Carlo methods to integrate out the unobserved states, and Fridman and Harris (1998), who use recursive numerical integration.

The primary contribution of this paper is to demonstrate tools that are robust, computationally efficient, straightforward to implement, and easy to adapt to new 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 idea is that it should be possible for most users to estimate and assess a variety of models of practical interest on multiple data sets without having to expend a great deal of effort.

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. It is typically known that the simulation estimator converges to some standard estimator with well-known properties as the simulation length goes to infinity (with fixed data). Of interest is how close to convergence the estimator is for the simulation lengths actually used in practice.

A common approach is to fix a particular pseudorandom sequence and report the corresponding estimates. The seed might, for example, be a family member's

3

birthdate. The hope is that the estimates corresponding to other seqences would not differ by much. Convincing evidence justifying this hope may be difficult to come by.

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 serves two purposes. The standard deviation gives an idea of the numerical stability of the estimator with respect to different seeds. But also, the mean will display much less variablility than the individual estimates.

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.

One of the advantages of the computational efficiency of the tools used in this paper is that careful studies of their numerical properties are feasible. 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.

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, Gallant, Ghysels and Tauchen (2003)), the results are not directly comparable. 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. This limits the number of model variants and data sets that can be studied in detail (given finite computing resources). In addition, Monte Carlo studies on the numerical properties of the estimator as well as its small sample statistical properties would be costly to perform for many of the models of practical interest.

The empirical work looks at S&P 500 index returns over the period June 23, 1980 through September 2, 2002. The results suggest that even the simple single factor models are able to capture the volatility dynamics (e.g., volatility clustering) exhibited by the data. The problem is capturing the *shape* of the returns distribution

4

(conditional on the volatility factor). While including a second volatility factor helps, all of the models fail in a similar manner. None is able to explain the extreme left tail of the distribution. In addition to formal statistical tests, QQ-plots clearly illustrating the nature of the models' failure are provided.

The best model fit is obtained by a new formulation of the SV-t model. This model differs from models considered previously in the manner in which 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, Johannes and Polson (2001) 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 idea of using single-factor models with more flexible distributional forms for the innovations looks like a promising direction for future work.

The rest of this paper is organized as follows: Section 2 describes the estimation approach used, Section 3 begins the empirical work and fits a variety of models, Section 4 discusses the issues of filtering and smoothing, Section 5 provides model diagnostics, Section 6 looks at the numerical performance of the estimators, and Section 7 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 $p(x, v)$ 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

$$
\begin{aligned}
L(\theta|x) &= \int \frac{p(x, v; \theta)}{q(v)}\, dQ(v) \\
&\approx \frac{1}{S} \sum_{s=1}^{S} \frac{p(x, v^{(s)})}{q(v^{(s)})}.
\end{aligned}
\tag{2}
$$

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 equation (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 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). This issue is addressed in Section 6.

The stochastic volatility models examined in this paper are of form

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

(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 $p(x, v)$. One first computes

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

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 $n \cdot n_v$ 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 very efficient computationally and can be implemented in just a few lines of Fortran code. Pseudocode and detailed implementation notes are available in the appendix. 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 (sample code available upon request).

Jacquier, Polson and Rossi (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

$$X_t = \exp(V_t/2)\epsilon_t$$
$$V_t = \alpha + \phi V_{t-1} + \sigma_V \eta_t$$

(4)

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

Following JPR, I estimate model (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 or multiple relative extrema.

Computational cost for this model is about 0.05 seconds 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 seconds 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.

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

8

returns that remains even after allowing for time-varying volatility. The specification used in this paper is given by

$$X_t = \mu + \sigma_X \exp(V_{t-1}/2)\epsilon_t$$
$$V_t = \phi V_{t-1} + \sigma_V \left(\rho\epsilon_t + \sqrt{1-\rho^2}\eta_t\right) \tag{5}$$

where $\epsilon_t \sim t_\nu$, $\eta_t \sim N(0,1)$, 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, Polson and Rossi (2002) using Bayesian MCMC techniques. The specification they use is given by

$$X_t = \sigma_X \exp(V_t/2)\lambda_t\epsilon_t$$
$$V_t = \phi V_{t-1} + \sigma_V \eta_t \tag{6}$$

where $\nu/\lambda_t \sim \chi^2_\nu$, ϵ_t and η_t are $N(0,1)$, and $\text{corr}(\epsilon_t, \eta_t) = \rho$. Note that the product $\lambda_t\epsilon_t$ has the t_ν distribution.

There are two key differences between these models. First, the timings are different: (5) is a martingale difference sequence, while (6) is not (due to the presence of V_t in the right hand side of the observation equation). This issue is examined in more detail in Section 3. Secondly, (5) 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. (2001) argue that this is an important feature of the data.

SV-t models have also been estimated using likelihood-based techniques by Chib, Nardari and Shephard (2002), Sandmann and Koopman (1998) and Liesenfeld and Richard (2002), but only with uncorrelated errors (note that (5) and (6) are equivalent if the errors are not correlated). As will be shown in Section 3, the models without correlation are not empirically relevant, at least for S&P 500 index returns.

As before, implementing the estimator for the model shown in (5) 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 essentially 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, Hsu and Tauchen (1999), Alizadeh et al. (2002), and Chernov et al. (2003) among others. Liesenfeld and Richard (2002) estimate and assess a two-factor model (but without leverage effects) using likelihood-based tools.

A standard two-factor model is given by

$$
\begin{aligned}
X_t &= \mu + \sigma_X \exp(U_{t-1}/2 + V_{t-1}/2)\epsilon_{1t} \\
V_t &= \phi_V V_{t-1} + \sigma_V \epsilon_{2t} \\
U_t &= \phi_U U_{t-1} + \sigma_U \epsilon_{3t}
\end{aligned}
\tag{7}
$$

where $\epsilon_{it} \sim N(0,1)$ and $\text{corr}(\epsilon_{it}, \epsilon_{jt}) = \rho_{ij}$. Note that if $\phi_U = 0$, then $\exp(U_{t-1}/2)\epsilon_{1t}$ is an iid mixture of normals. If $\rho_{31} = 0$ as well, then it is a *scale* mixture of normals (i.e., the normals are always centered at zero). Recall that the t distributed innovations which appear in the SV-t model may also be thought of as iid scale mixtures of normals, but the scaling factor has a different distribution.

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 second (on a 2 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 $h = \int \sigma_V^{-1}(v) \, dv$.

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

$$X_t = \mu + h_{t-1}\epsilon_t$$
$$h_t = h_{t-1} + \frac{\phi}{2h_{t-1}} \left(\alpha - \frac{\sigma^2}{4\phi} - h_{t-1}^2 \right) + \frac{\sigma}{2}\eta_t \tag{8}$$

where ϵ_t and η_t are both $N(0,1)$ and $\mathrm{corr}(\epsilon_t, \eta_t) = \rho$. 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 beyond the scope of this paper.

3 Application

In this section, some variants of the models described in the previous section are estimated using daily S&P 500 index returns from June 23, 1980 to September 2, 2002 ($N = 5616$). The data are plotted in Figure 1.

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, Benzoni and Lund (2002a). 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 paper 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 the diagnostics discussed in Section 5. The models under consideration are summarized in Table 2. Parameter estimates and log likelihoods are shown in Table 3.

We first look at the single factor SV model

$$X_t = \mu + \sigma_X \exp(V_\tau/2)\epsilon_t$$
$$V_t = \phi V_{t-1} + \sigma_V \eta_t$$

where ϵ_t and η_t are both iid $N(0,1)$ and $\text{corr}(\epsilon_t, \eta_t) = \rho$. Two different timings appear in the literature, depending upon whether $\tau = t$ or $\tau = t - 1$. If $\tau = t - 1$, the resulting model is a martingale difference sequence (after subtracting off the unconditional mean, μ). It also represents the Euler scheme approximation of the underlying continuous-time model. This model will be referred to as SV1-EUL. If $\tau = t$, 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). This model will be referred to as SV1-JPR. If $\rho = 0$, the timing issue is irrelevant and the model will be referred to simply as SV1.

Whether the asymmetry introduced by the JPR timing is important in practice can only be determined empirically and with respect to particular data sets. 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 it is nice that $X_t - \mu$ is a martingale difference sequence for SV1-EUL, the departure from "martingale-ness" implied by SV1-JPR may be small. Furthermore, it is not obvious that the true data generating process has this characteristic (that the data had to be run through an ARMA prefilter suggests otherwise). Rather than relying on possibly shaky theoretical arguments, it is better to assess the models empirically.

Looking at the estimates, we first note that ρ is highly significant (leverage effect). Potentially more interesting is that SV1-JPR provides a significant improvement over SV1-EUL (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). On the other hand, Yu (2002) finds in favor of the martingale timing (using S&P 500 index data from 1980-87 and a Bayesian estimator), so this result may not be robust. In practice, it may not make much difference which version of the model one uses: parameter estimates as well as the forecasts and diagnostics examined in subsequent sections 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.

The two-factor model is specified as

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

with $\epsilon_{it} \sim N(0,1)$ and $\text{corr}(\epsilon_{it}, \epsilon_{jt}) = \rho_{ij}$. There is a timing issue similar to that with the single-factor model. Three variants are considered. In the following discussion, V will refer to the persistent factor and U to the nonpersistent factor. SV2-EUL uses $\tau = t - 1$ (for both U and V). This model has the same nice properties as the single-factor martingale model. SV2-JPR uses $\tau = t$. While this model is not a martingale, the additional sources of non-Gaussianity resulting from the appearance of the contemporaneous volatility state in the observation equation may be useful, as turned out to be the case in the single-factor models. I also look at a hybrid model, SV2-HYB, which uses the martingale timing for the persistent factor but not the nonpersistent factor.

Based on the Kullback-Leibler information criterion, 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 would probably be considered the preferred model. A version of the two-factor model with all correlations set to zero (i.e., no leverage effect) is also estimated. For this version of the model, the timing issue is irrelevant. As in the single-factor case, this version of the model is of little relevance empirically.

The parameter ϕ_U governs the persistence of the non-persistent factor and ρ_{31} controls the extent 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 3 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).

Despite having two fewer parameters, the SV-t model does even better than the two-factor models in terms of log likelihood. It is useful to compare these models more closely. SV2-EUL with $\phi_U = 0$ may be rewritten as

$$
\begin{aligned}
X_t &= \mu + \sigma_X \exp(V_{t-1}/2)\nu_t \\
V_t &= \phi V_{t-1} + \sigma_V \zeta_t
\end{aligned}
\tag{9}
$$

where $\nu_t = \exp(U_{t-1}/2)\epsilon_{1t}$ is a scale mixture of normals with no autocorrelation. If $\rho_{31} = 0$, these innovations are iid, otherwise there is some "leverage effect" (i.e., $\text{Var}(\nu_t)$ depends upon ν_{t-1}). The volatility innovations, ζ_t, are normal and possibly correlated with ν_t. The SV-t model is also of form (9): 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 the results on ϕ_U and ρ_{31} pointed out in the preceding paragraph, it is not surprising that the performance of these models is similar. The particular forms of the mixing distribution implied by these two alternatives does not appear to make much difference empirically.

It is useful to clarify what is meant by one- versus two-factor models. Models of form (9) will be referred to as single-factor as long as ν_t is iid, even if there exists a possible construction of the distribution as a mixture. The SV-t and SV2-EUL

models (as well as the other SV2 variants) with $\phi_U = \rho_{31} = 0$ are of this form. For the unrestricted SV2 models with the parameter estimates shown in Table 3, on the other hand, the ν_t not quite iid.

Figures 2 and 3 show predictive densities for $X_t|V_{t-1}$ for several models. The densities are computed by fixing $V_{t-1} = 0$ 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 in the single-factor model 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.

The idea of using a single-factor model but with flexible distributional forms for the innovations looks promising. Mixture distributions are easy to work with technically and offer great flexibility. The formulation used for the SV-t model in this paper (wherein large absolute returns are associated with simultaneous "jumps" in volatility) may be worth particular attention.

Table 3 also displays estimates for a single factor affine model. These results confirm the findings of ABL and CGGT 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).

15

4 Smoothing and filtering

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

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

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. Computational cost is about two hours of CPU time on a 750 MHz PC. For the single-factor SV models, the sampler is more efficient and 10,000 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, Shephard and Chib 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, $E(V_t | \mathcal{F}_t)$, where $\mathcal{F}_t$ denotes the information set generated by $X_1, \ldots, X_t$. One way to obtain these is by means of a particle filter. 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 $\hat{F}(v_t|\mathcal{F}_t)$ that approximate the exact densities $F(v_t|\mathcal{F}_t)$. 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 $\hat{F}(v_t|\mathcal{F}_t)$, advance the particles by drawing from $F(v_{t+1}|v_t, \mathcal{F}_t)$, and then weight to adjust for the new information implied by X_{t+1}.

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

$$
\begin{aligned}
p(v_{t+1}|\mathcal{F}_{t+1}) &= \int p(v_{t+1}, v_t|\mathcal{F}_{t+1})\, dv_t \\
&= \int \frac{p(v_{t+1}, v_t|\mathcal{F}_{t+1})}{p(v_t|\mathcal{F}_t)}\, dP(v_t|\mathcal{F}_t).
\end{aligned}
\tag{10}
$$

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

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

Plugging (11) into (10) gives

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

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

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

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.

5   Model diagnostics

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 $X_{t+1}|\mathcal{F}_t$ can be estimated by

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

Similarly, its cdf can be estimated by

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

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 "pseudo-residuals" should be iid $N(0,1)$.

The Jarque-Bera test is used 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. 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 Table 4 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 $z_t = \text{prob}(X_{t+1}^2 \leq x_{t+1}^2|\mathcal{F}_t)$. This formulation makes it impossible to disentangle the right and left tails of the distribution, diminishing its usefulness.

The Box-Pierce and ARCH tests suggest that even the simple single-factor models are able to capture the volatility dynamics adequately. The issue of timing (SV1-JPR vs. SV1-EUL) makes little difference here. The two-factor models (including

18

SV2-EUL with the restriction $\phi_U = \rho_{31} = 0$) do slightly worse on these tests but also fail to be rejected. The correlograms in Figures 4 and 5 tell much the same story.

In contrast, the Jarque-Bera test results imply that none of the models is able to capture the shape of the returns distribution. The situation is more clearly illustrated by the QQ-plots in Figures 6 and 7. 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.

The affine model does poorly in both dimensions, as shown by the test statistics in Table 4 and more clearly by the graphics in Figures 4 and 6.

Unfortunately, there does not appear to be an easy way to carry out a similar analysis for the SV-t model. Computing 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 this problem numerically, but doing so would be computationally costly and is not addressed in this paper. In any event, it seems likely that the results would be similar to those for the SV2 models.

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

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

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).

The key point to take from the foregoing specification analysis is that the simple single-factor log volatility model is sufficient to capture the volatility dynamics of returns; the problem is to get the right shape for the returns distribution conditional on volatility. Including a second factor or using t innovations fattens the tails somewhat, but neither offers sufficient flexibility to capture the left tail. Understanding this part of the distribution is important for hedging, option pricing,

19

and risk management. Chernov et al. (2003) try a more flexible specification for the volatility factors which appears to do better on the particular data set they consider. Given sufficient patience, it should always be possible to come up with an *ad hoc* specification that is acceptable for a particular data set. Unfortunately, this process would have to be repeated for every different data set examined, a prospect with little appeal (particularly if one is forced to use a more computationally costly estimator than the one used here). Better would be to come up with a non-parametric approach to fitting the returns distribution. One way forward might be the SV-mix class of models examined in Durham (2003), which uses a discrete mixture of normals to fit the conditional distribution of returns, $X_t|V_{t-1}$.

6   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.

It is useful to look first at the convergence properties of the Monte Carlo approximation to the likelihood. Recall that

$$
\begin{aligned}
L(\theta|x) &= \int p(x,v)\,dv \\
&= \int \frac{p(x,v)}{q(v)}\,dQ(v) \\
&\approx \frac{1}{S}\sum_{s=1}^{S} \frac{p(x,v^{(s)})}{q(v^{(s)})}
\end{aligned}
\tag{12}
$$

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

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

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 obtain a large number of draws from $p(x,V)/q(V)$, where V has density q. Suppose that the right tail of $\log\left(p(x,V)/q(V)\right)$ is thinner than that of Z, the Gaussian distribution with the same mean and variance. Since $\exp(Z)$ has a finite second moment, we 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 $\log\left(p(x,V)/q(V)\right)$ might be approximately normally distributed. Figure 8 shows

histograms and QQ-plots of 100,000 draws from $\log\big(p(x, V)/q(V)\big)$ using the data and two of the models considered in Section 3. These figures are supportive of the argument outlined above. It seems reasonable to expect that the sum in (12) converges at rate $\sqrt{S}$, at least eventually.

On the other hand, if the variance of $\log\big(p(x, V)/q(V)\big)$ 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 information is possibly of interest but 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 take the mean of the entire collection of estimates computed. The standard deviation of the individual estimates serves as an indication of the simulation error. Of course, 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 for the data set under consideration.

Table 5 shows the results of trying this idea using the data and some of the models considered in Section 3. The mean and standard deviation of the estimates obtained using each of 500 different sequences of random numbers 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 models, 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 dissipates as S gets large.

For the two-factor model, there is a significant amount of variation across simulations in the estimates for several of the parameters. However, the variance of

the mean of the estimates — which is what is actually reported — should be much smaller. There are also significant shifts 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.

Note that the issue of numerical error is a problem 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 careful investigations of numerical issues are possible.

7 Conclusion

This paper examines 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. Model diagnostics are based on standard time-series techniques.

The application looks at S&P 500 index returns. The results suggest that rather than adding a second volatility factor, better results may be obtained by staying within the single factor framework but using more flexible distributions for the innovations.

References

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

Andersen, T. G., Benzoni, L. and Lund, J. (2002a). An empirical investigation of continuous-time equity retrun models, *Journal of Finance* **57**: 1239–1284.

Andersen, T. G., Bollerslev, T., Diebold, F. X. and Labys, P. (2002b). Modeling and forecasting realized volatility, Working paper, Kellogg School of Management, Nortwestern University.

Barndorff-Nielsen, O. E. and 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.

Barndorff-Nielsen, O. E., Nicolato, E. and Shephard, N. (2002). Some recent developments in stochastic volatility modelling, *Quantitative Finance* **2**: 11–23.

Bates, D. S. (1996). Jumps and stochastic volatility: Exchange rate processes implicit in deutsche mark options, *Review of Financial Studies* **9**: 69–107.

Brandt, M. W. and Santa-Clara, P. (2002). Simulated likelihood estimation of diffusions with an application to exchange rate dynamics in incomplete markets, *Journal of Financial Economics* **63**: 161–210.

Carr, P., Geman, H., Madan, D. B. and Yor, M. (2001). Stochastic volatility for Lévy processes, Unpublished paper: Robert H Smith School of Business, Maryland University.

Chernov, M., Gallant, A. R., Ghysels, E. and Tauchen, G. (2003). Alternative models for stock price dynamics, forthcoming in *Journal of Econometrics*.

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

Dai, Q. and Singleton, K. J. (2000). Specification analysis of affine term structure models, *Journal of Finance* **55**: 1943–1978.

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

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

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

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

Durbin, J. and 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.

Durbin, J. and Koopman, S. (2001). *Time Series Analysis by State Space Methods*, Oxford University Press, Oxford, U.K.

Durham, G. B. (2003). SV mixture models with application to s&p 500 index returns, Working paper, University of Iowa.

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

Elerian, O., Chib, S. and Shephard, N. (2001). Likelihood inference for discretely observed non-linear diffusions, *Econometrica* **69**: 959–993.

Engle, R. F. and Lee, G. G. (1998). A permanent and transitory component model of stock return volatility, *in* R. F. Engle and H. White (eds), *Cointegration, Causality, and Forecasting: A Festschrift in Honor of Clive W.J. Granger*, Oxford University Press, 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. and Polson, N. G. (2001). The impact of jumps in returns and volatility, Working paper, Duke University.

Fridman, M. and 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. and Tauchen, G. E. (1996). Which moments to match?, *Econometric Theory* **12**: 657–681.

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

Gallant, A. R., Hsu, C.-T. and Tauchen, G. E. (1999). Using daily range data to calibrate volatility diffusions and extract the forward integrated variance, *Review of Economics and Statistics*, (forthcoming).

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

Geweke, J. and Amisano, G. (2001). Compound Markov mixture models with applications in finance, Working paper, University of Iowa.

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

Greene, W. H. (2002). *Econometric Analysis*, 5 edn, 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.

Harvey, A., Ruiz, E. and Shephard, N. (1994). Multivariate stochastic variance models, *Review of Economic Studies* **61**: 247–264.

Hol, E. and Koopman, S. J. (2002). Forecasting the variability of stock index returns with stochastic volatility models and implied volatility, Working paper, Free University of Amsterdam.

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

Jacquier, E., Polson, N. G. and Rossi, P. E. (2002). Bayesian analysis of stochastic volatility models with fat-tails and correlated errors, forthcoming in *Journal of Econometrics*.

Jones, C. S. (1999). The dynamics of stochastic volatility: Evidence from underlying and options markets, Working Paper, Simon School of Business, University of Rochester.

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

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

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

Liesenfeld, R. and Richard, J.-F. (2002). Univariate and multivariate stochastic volatility models: Estimation and diagnostics, Forthcoming in *Journal of Empirical Finance*.

Pan, J. (2001). Integrated time-series analysis of spot and option prices, Working paper, MIT Sloan School of Management.

Pedersen, A. R. (1995). A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations, *Scandinavian Journal of Statistics* **22**: 55–71.

Piazzesi, M. (2001). An econometric model of the yield curve with macroeconomic jump effects, Working paper, Anderson School of Business, UCLA.

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

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

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

Taylor, S. (1986). *Modelling Financial Time Series*, John Wiley, New York.

Yu, J. (2002). MCMC methods for estimating stochastic volatility models with leverage effects: Comments on Jacquier, Polson and Rossi (2002), Working paper (University of Aukland, Department of Economics).

Table 1. Simulation studies.

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 the 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 $S = 256$.

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. (5))

	μ	σ_X	ϕ	σ_V	ρ	ν
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. (7))

	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. (8))

	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)

Table 2. Models

SV1-EUL:

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

SV1-JPR:

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

SV-t:

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

SV2-EUL:

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

SV2-JPR:

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

SV2-HYB:

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

AFF:

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

Table 3. Parameter estimates, S&P 500 index returns.

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 $S = 1024$. Each model is estimated 100 times using different seeds for the random number generator use to construct the simulations (as described in Section 6). The mean of these estimates is reported. Standard errors are immediately below each set of estimates, followed by the standard deviation of the 100 individual estimates.

Model	Log L	μ	σ_X	ϕ	σ_V	ρ	ν
SV1-EUL	18533.34	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	18538.84	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	18503.12	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	18582.24	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	18553.50	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)

Model	Log L	μ	σ_X	ϕ_V	σ_V	ρ_{21}	ϕ_U	σ_U	ρ_{31}
SV2-EUL	18575.86	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	18569.78	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	18575.21	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	18572.85	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	18543.92	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)	

Model	Log L	μ	α	ϕ	σ	ρ
AFF	18473.44	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)

Table 4. Model diagnostics

	Jarque-Bera	ARCH(20)	Box-Pierce(20)
Raw data	439771.42	272.70	368.60
	(0.0000)	(0.0000)	(0.0000)
Affine	2722.82	59.79	63.74
	(0.0000)	(0.0000)	(0.0000)
SV1-EUL	288.93	12.35	12.70
	(0.0000)	(0.9036)	(0.8898)
SV1-JPR	168.42	12.66	13.08
	(0.0000)	(0.8915)	(0.8740)
SV2-EUL	42.56	20.50	21.04
	(0.0000)	(0.4269)	(0.3948)
SV2-EUL ($\phi_U = \rho_{31} = 0$)	53.05	20.91	21.54
	(0.0000)	(0.4023)	(0.3662)
SV2-JPR	34.89	20.46	20.91
	(0.0000)	(0.4298)	(0.4022)
SV2-HYB	30.26	17.86	18.27
	(0.0000)	(0.5963)	(0.5695)

Table 5. Numerical error

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

SV1-EUL

S	μ	σ_X	ϕ	σ_V	ρ	$\log L$
32	0.0000693	0.0085533	0.97663	0.1685	-0.4331	18532.62
	(0.0001059)	(0.0004175)	(0.00442)	(0.0144)	(0.0476)	
	(0.0000018)	(0.0000040)	(0.00048)	(0.0021)	(0.0030)	(1.15)
128	0.0000688	0.0085516	0.97633	0.1699	-0.4323	18533.21
	(0.0001059)	(0.0004182)	(0.00439)	(0.0143)	(0.0475)	
	(0.0000015)	(0.0000033)	(0.00042)	(0.0019)	(0.0026)	(0.94)
512	0.0000685	0.0085508	0.97616	0.1707	-0.4320	18533.41
	(0.0001059)	(0.0004088)	(0.00460)	(0.0149)	(0.0468)	
	(0.0000013)	(0.0000030)	(0.00039)	(0.0017)	(0.0021)	(0.64)
2048	0.0000684	0.0085496	0.97601	0.1714	-0.4314	18533.58
	(0.0001059)	(0.0004145)	(0.00445)	(0.0144)	(0.0475)	
	(0.0000010)	(0.0000022)	(0.00030)	(0.0014)	(0.0017)	(0.45)
8192	0.0000682	0.0085487	0.97585	0.1722	-0.4310	18533.73
	(0.0001059)	(0.0004140)	(0.00449)	(0.0146)	(0.0476)	
	(0.0000008)	(0.0000021)	(0.00029)	(0.0013)	(0.0014)	(0.37)
32768	0.0000681	0.0085482	0.97578	0.1725	-0.4307	18533.78
	(0.0001059)	(0.0004116)	(0.00454)	(0.0147)	(0.0474)	
	(0.0000006)	(0.0000018)	(0.00024)	(0.0011)	(0.0012)	(0.24)

SV2-EUL

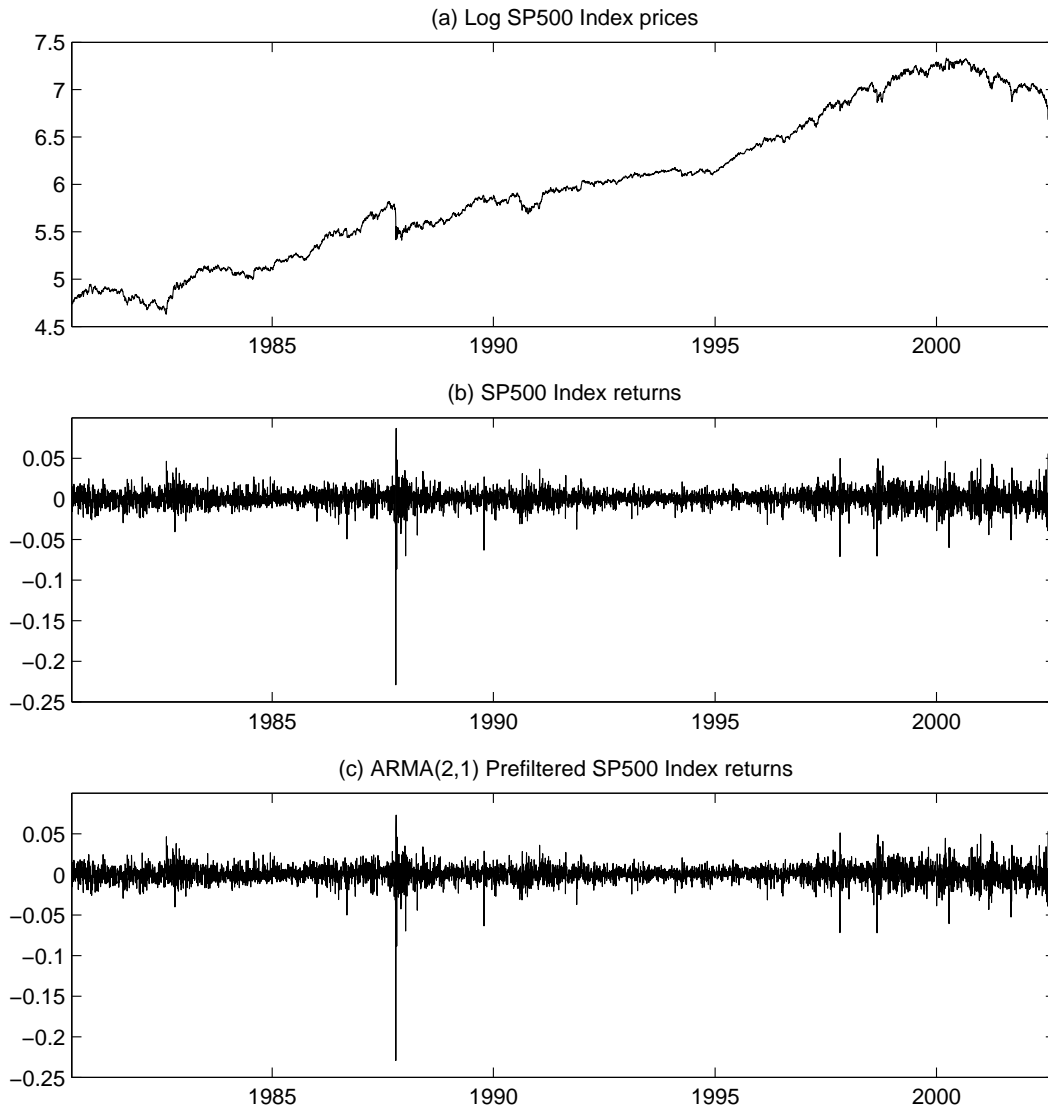| S	μ	σ_X	ϕ_V	σ_V	ρ_{21}	ϕ_U	σ_U	ρ_{31}	$\log L$
32	0.000087	0.00836	0.9913	0.0957	-0.429	0.311	0.424	-0.279	18572.09
	(0.000104)	(0.00055)	(0.0029)	(0.0121)	(0.061)	(0.199)	(0.033)	(0.069)	
	(0.000006)	(0.00006)	(0.0012)	(0.0075)	(0.042)	(0.204)	(0.036)	(0.073)	(2.62)
128	0.000085	0.00833	0.9909	0.0984	-0.445	0.236	0.445	-0.249	18573.86
	(0.000104)	(0.00071)	(0.0025)	(0.0139)	(0.105)	(0.137)	(0.052)	(0.077)	
	(0.000007)	(0.00005)	(0.0010)	(0.0065)	(0.035)	(0.186)	(0.030)	(0.065)	(2.34)
512	0.000083	0.00831	0.9906	0.1005	-0.457	0.167	0.462	-0.223	18575.34
	(0.000103)	(0.00063)	(0.0025)	(0.0120)	(0.074)	(0.152)	(0.042)	(0.072)	
	(0.000007)	(0.00004)	(0.0009)	(0.0057)	(0.027)	(0.169)	(0.022)	(0.056)	(2.10)
2048	0.000081	0.00829	0.9904	0.1016	-0.465	0.132	0.473	-0.208	18576.59
	(0.000103)	(0.00062)	(0.0025)	(0.0115)	(0.068)	(0.168)	(0.040)	(0.074)	
	(0.000006)	(0.00003)	(0.0008)	(0.0051)	(0.023)	(0.146)	(0.016)	(0.047)	(1.89)

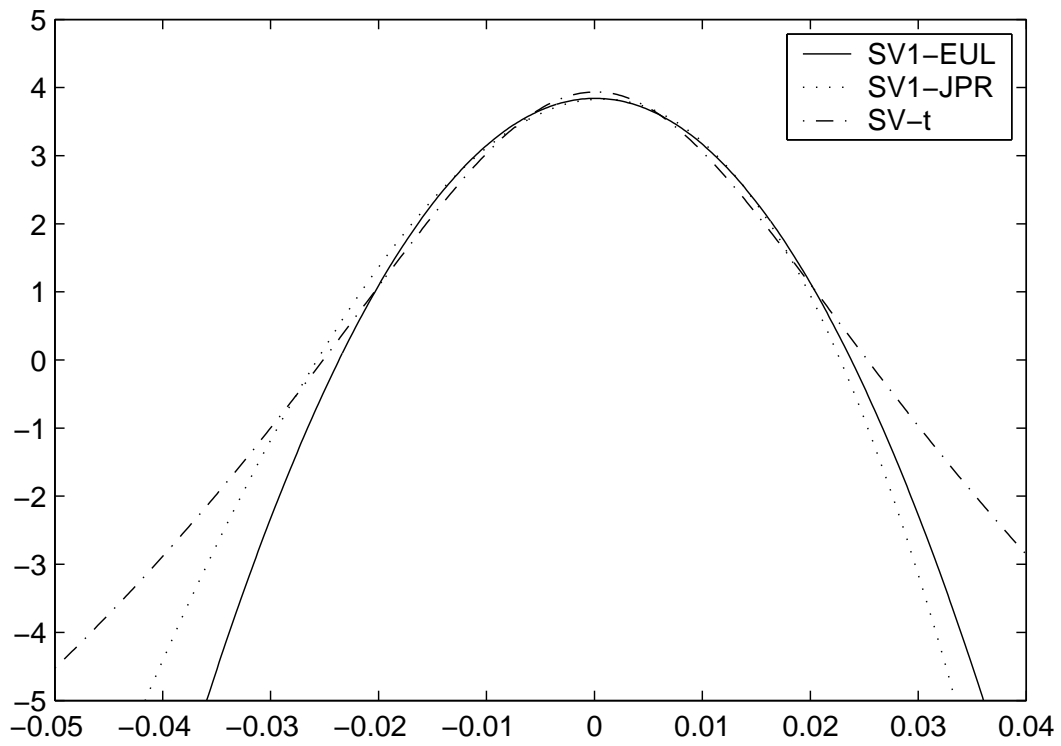Figure 1. S&P 500 index returns, 6/23/1980 - 9/2/2002.

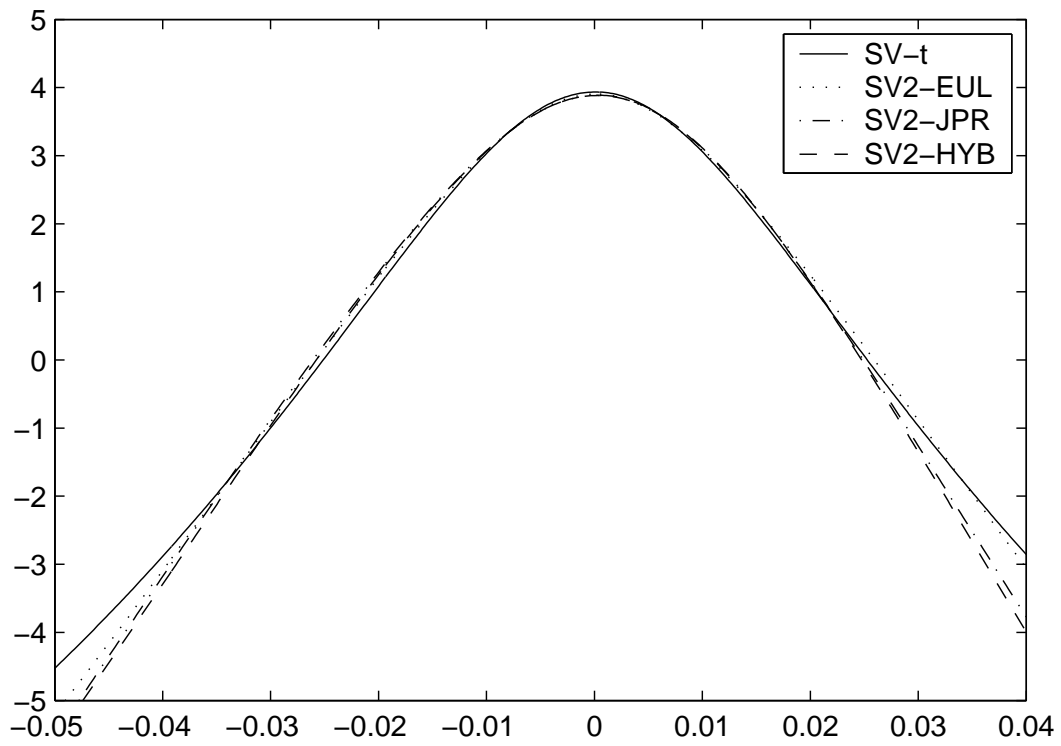Figure 2. Predictive densities for $X_t|V_{t-1}$.

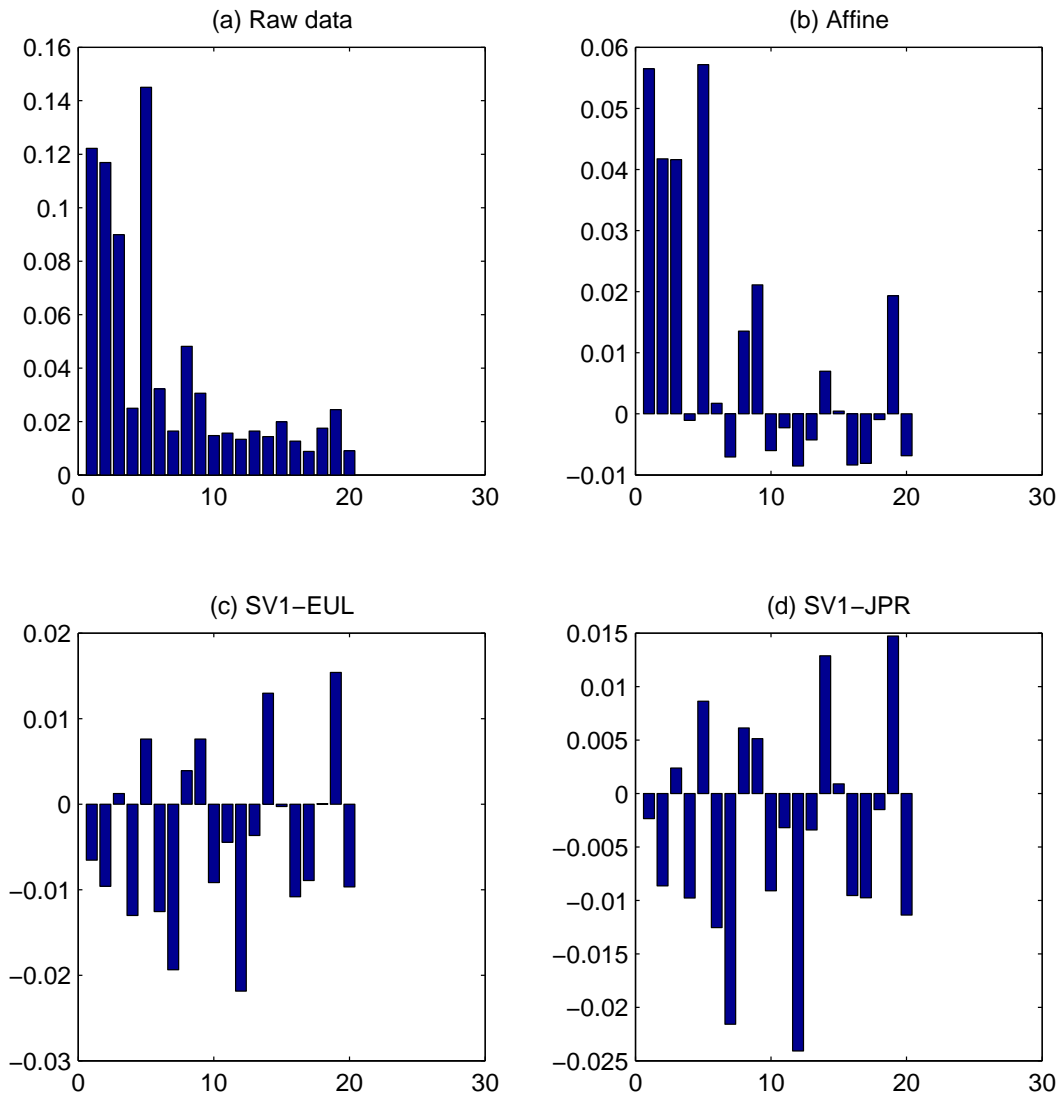Figure 3. Predictive densities for $X_t|V_{t-1}$.
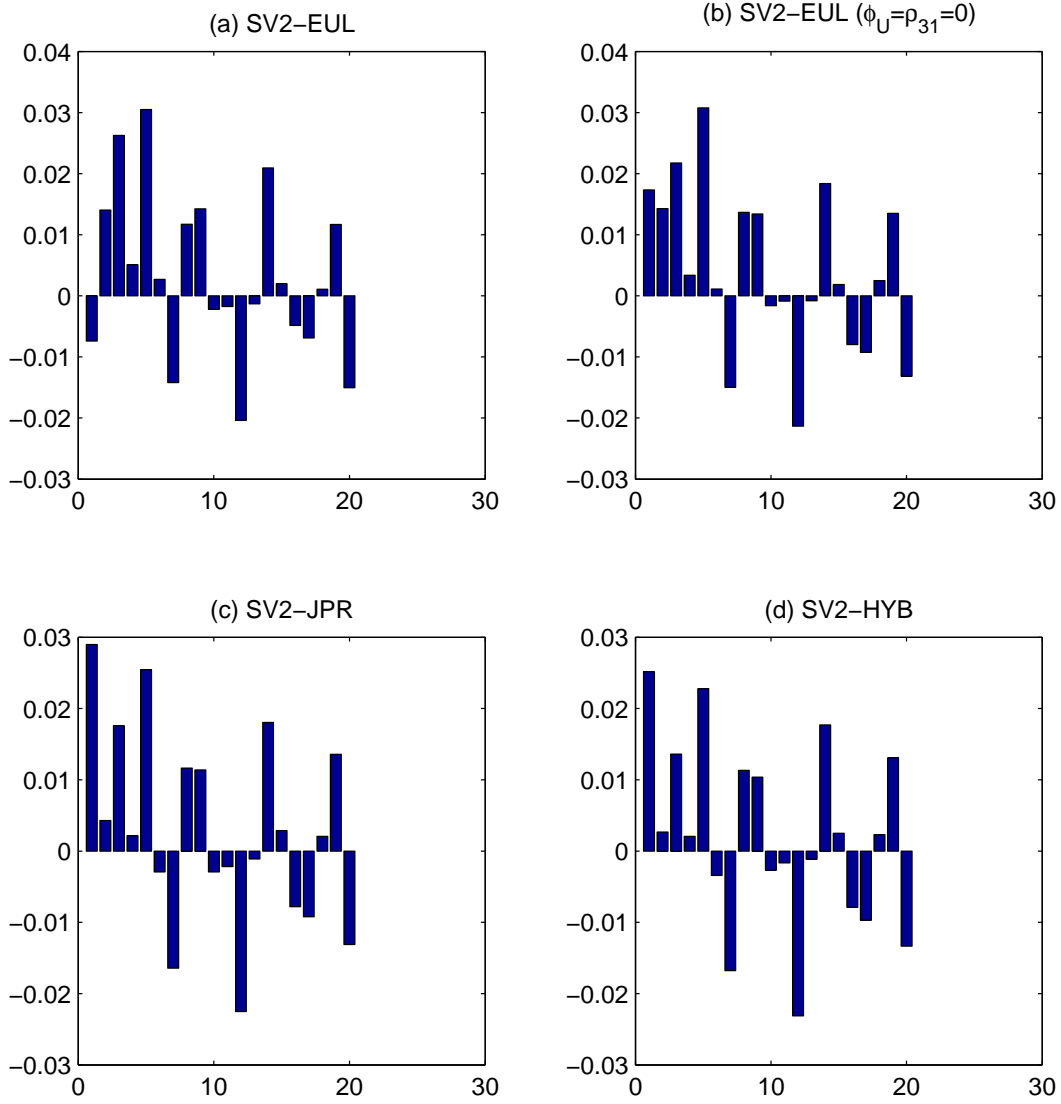
Figure 4. corroleograms for squared "pseudo-residuals".
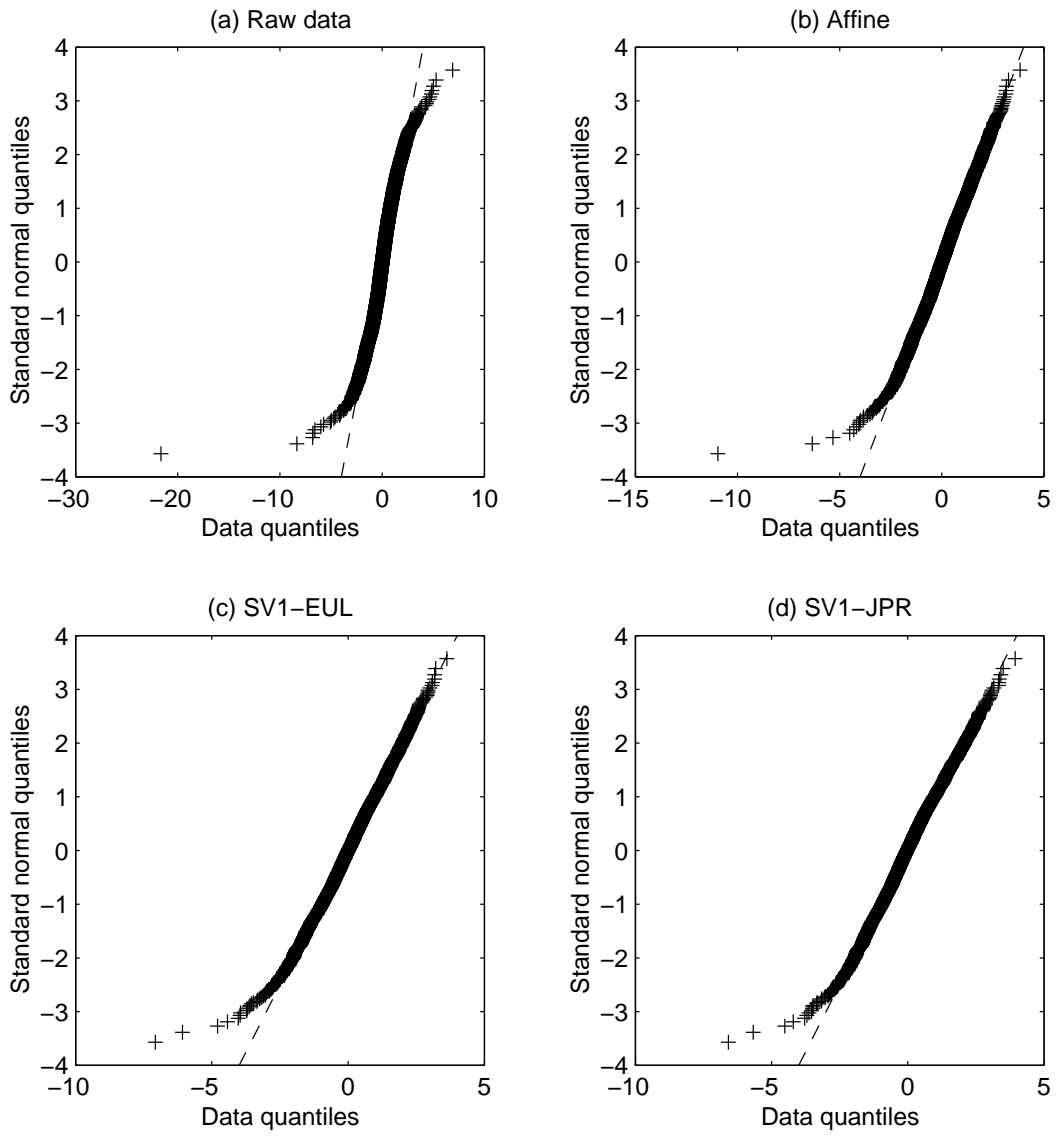
Figure 5. Corroleograms for squared "pseudo-residuals".
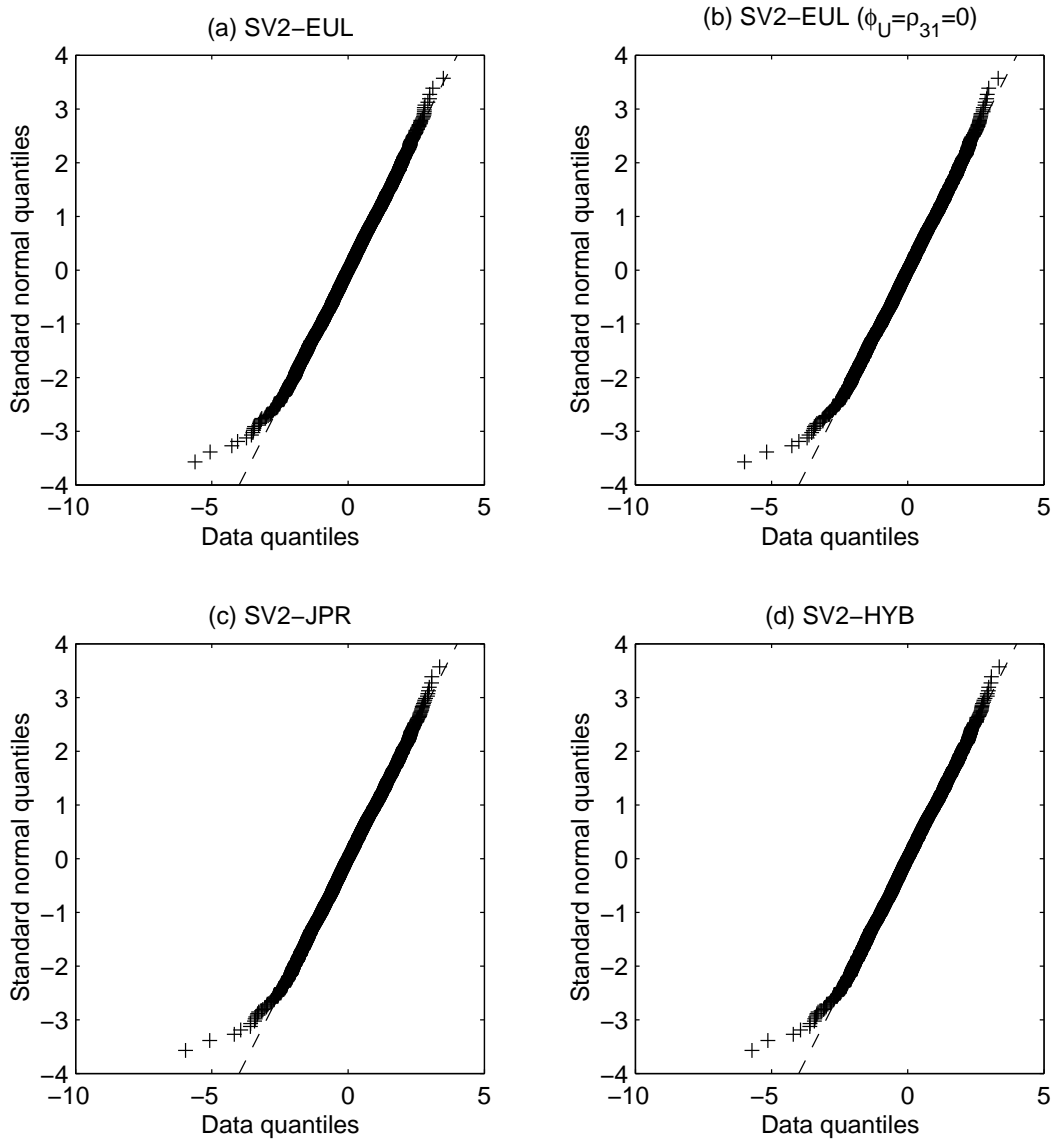
Figure 6. QQ-plots for "pseudo-residuals".

Figure 7. QQ-plots for "pseudo-residuals.

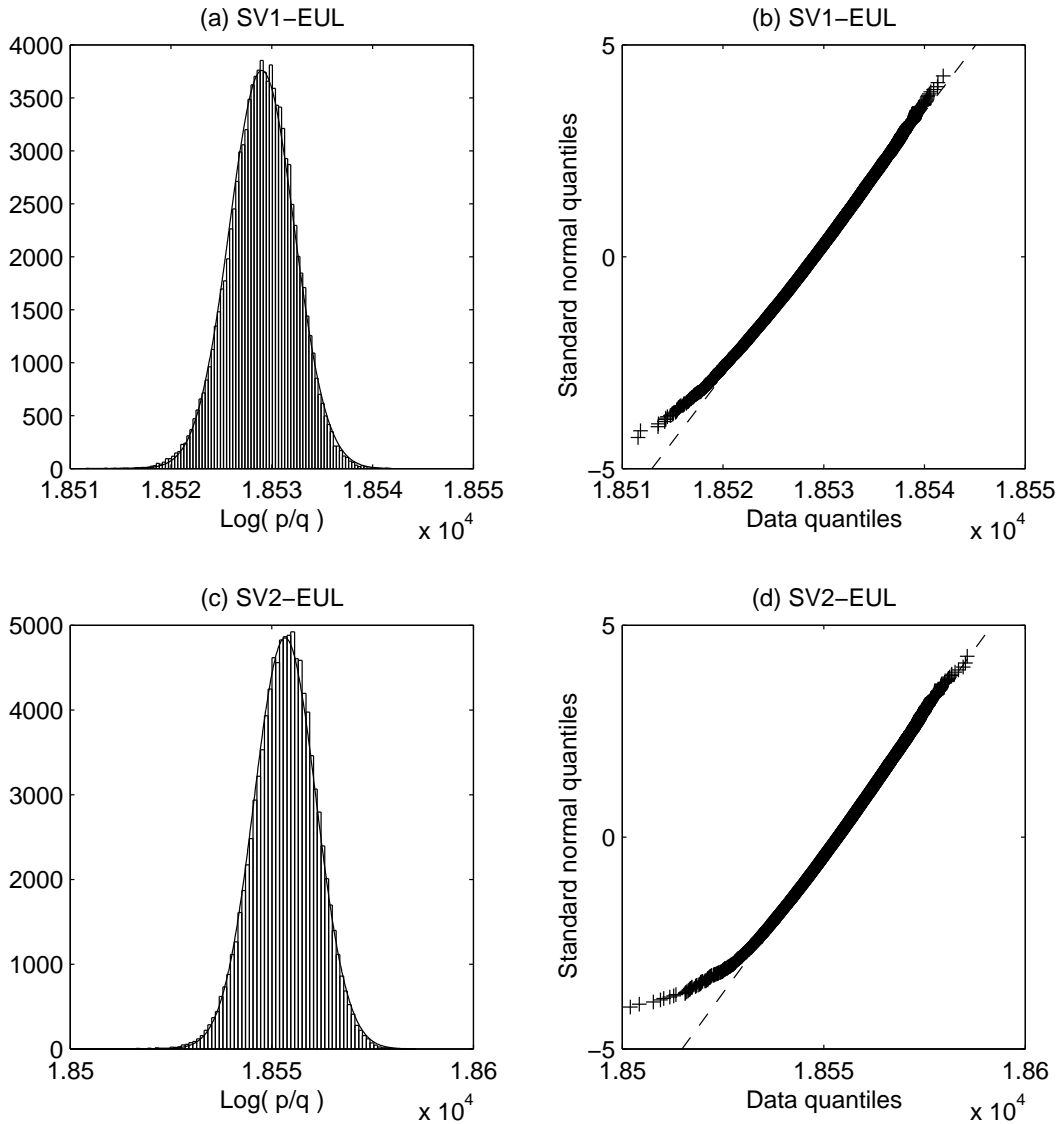Figure 8. Histograms and QQ-plots for 100,000 draws from $\log(p(x, V)/q(V))$, where V is drawn from the importance density, q.

40