

# SV mixture models with application to S&P 500 index returns<sup>☆</sup>

Garland B. Durham<sup>\*</sup>

## Abstract

Understanding both the dynamics of volatility and the shape of the distribution of returns conditional on the volatility state is important for many financial applications. A simple single-factor stochastic volatility model appears to be sufficient to capture most of the dynamics. It is the shape of the conditional distribution that is the problem. This paper examines the idea of modeling this distribution as a discrete mixture of normals. The flexibility of this class of distributions provides a transparent look into the tails of the returns distribution. Model diagnostics suggest that the model, SV-mix, does a good job of capturing the salient features of the data. In a direct comparison against several affine-jump models, SV-mix is strongly preferred by Akaike and Schwarz information criteria.

## 1. Introduction

Equity returns are typically highly non-Gaussian. Time-varying volatility accounts for much of this non-Gaussianity. But even after allowing for time-varying volatility, model

---

<sup>☆</sup>I am grateful for the helpful comments and suggestions of David Bates, John Geweke, and Michael Stutzer. I am especially grateful to the anonymous referee, whose many suggestions substantially improved the paper. Any remaining errors are my own.

<sup>\*</sup>Tel.: +1 303 494 4292.

*E-mail address:* [garland.durham@colorado.edu](mailto:garland.durham@colorado.edu).

residuals are likely to be fat-tailed and left-skewed. Because these features of the data are important for option pricing, risk management, and other applications, much work has been done trying to model them. This paper introduces a new model and demonstrates some recently developed techniques for assessing model performance. The model is built on a standard stochastic volatility (SV) framework but takes a semiparametric approach (mixture of normals) toward fitting the distribution of returns conditional on the volatility factor. Results suggest that the model is largely successful in capturing not only the dynamics of volatility but also the shape of the distribution of returns.

Two main classes of models are used to explain time-varying volatility: generalized autoregressive conditional heteroskedasticity (GARCH) and SV. In both, volatility is a random process. In GARCH models, the link between the data and this volatility process is deterministic, while in SV models the volatility process incorporates an additional source of noise. Given a model, Bayes' rule can be used to infer the distribution of the volatility variable conditional on the data. In GARCH models, this distribution is singular (up to an initial condition). The deterministic link between the data and the volatility process posited by GARCH models is difficult to justify, either theoretically or empirically. However, it makes estimation and analysis of such models much simpler, accounting for their widespread use.

This paper restricts attention to SV models. The tools used for estimation and inference are both computationally efficient and straightforward to implement. Standard & Poor's (S&P) 500 index returns have been widely studied and thus provide a useful test case.

The basic SV model is given by

$$\begin{aligned} dS_t &= \mu_S dt + \sigma_S \exp(V_t/2) dW_{1t}, \\ dV_t &= \psi V_t dt + \sigma_V dW_{2t}, \end{aligned} \tag{1} \leftarrow$$

where  $S_t$  represents log price,  $V_t$  is the latent volatility process, and  $W_{1t}$  and  $W_{2t}$  are (possibly correlated) Brownian motions. The Euler scheme approximation of this model is given by

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

where  $X_t = S_t - S_{t-1}$ ,  $\mu_X = \mu_S$ ,  $\sigma_X = \sigma_S$ ,  $\phi = \psi + 1$ , and  $\varepsilon_t$  and  $\eta_t$  are independently and identically distributed (iid) standard normal with  $\text{corr}(\varepsilon_t, \eta_t) = \rho$ .<sup>1</sup> While the continuous-time version of this model is useful for deriving theoretical results, statistical analysis proceeds much more simply with the discrete-time approximation. If  $V_t$  is highly persistent, the two differ by little. This is the case in many applications (including the one considered here). It is Eq. (2) and variations of it that are the subject of this paper.

This simple model appears to do a good job of capturing volatility clustering, i.e., explaining  $\text{Var}[X_{t+1}|X_t, X_{t-1}, \dots]$ . However, it is unable to adequately capture other features of the conditional distribution of returns, such as skewness and kurtosis.<sup>2</sup> The idea

---

<sup>1</sup>For the purposes of this paper, the data are observed daily, time is measured in days, and the Euler scheme is based on a discretization interval equal to one day.

<sup>2</sup>Although the conditional returns distribution implied by the Euler scheme approximation is normal, the continuous-time version allows for some non-normality. In particular, correlation between  $W_1$  and  $W_2$  induces skewness in this conditional distribution, which could be important in fitting the data. This issue is examined in more detail in Appendix A.

of this paper is to use a highly flexible functional form, a mixture of normals, for the distribution of  $\varepsilon_t$ , enabling the model to reflect more accurately the features of the conditional distribution of returns presented by the data. The proposed model also sets  $\eta_t = \rho\varepsilon_t + \sqrt{1 - \rho^2}u_t$ , where  $u_t$  is  $N(0, 1)$  and uncorrelated with  $\varepsilon_t$ . Thus,  $\eta_t$  still has mean zero, variance one, and  $\text{corr}(\eta_t, \varepsilon_t) = \rho$ , but it inherits some non-Gaussianity from  $\varepsilon_t$ . A potentially useful implication of this specification is that a large move in the price of the asset implies a large move in the volatility as well. In particular, volatility jumps following a market crash. Eq. (2) with these specifications for  $\varepsilon_t$  and  $\eta_t$  is referred to as the stochastic volatility mixture (SV-mix) model.

The mixture distribution is constructed as follows. Suppose the distribution is given by a mixture of  $K$  normals with means  $\mu_k$ , standard deviations  $\sigma_k$ , and weights  $p_k$  for  $k = 1, \dots, K$ . A draw from this distribution is obtained by first drawing from the discrete distribution on  $1, \dots, K$  with weights  $p_1, \dots, p_K$ , and then drawing from the corresponding normal distribution. The density of the mixture is given by  $\sum_{k=1}^K p_k \phi(\mu_k, \sigma_k)$ , where  $\phi(\mu, \sigma)$  is the density of a Gaussian random variable with mean  $\mu$  and standard deviation  $\sigma$ .

The appeal of the SV-mix approach is the way in which it combines the structure of the standard SV model with a flexible approach to modeling the conditional distribution of returns. This flexibility gives the approach an almost nonparametric flavor. Given a sufficient number of components, the mixture distribution can approximate any distribution to an arbitrary degree of precision (e.g., [McLachlan and Peel, 2000](#)). The number of components used determines the degree of smoothing, similar to the bandwidth choice in a kernel method. A wide range of density shapes can be obtained using mixtures with three or four components. In practice, the problem is generally to estimate some unknown density based on sample information. The appropriate number of mixture components to use can be determined by some information criterion, e.g., Schwarz criterion or Akaike information criterion. As the amount of data increases, more information regarding the true distribution is available, justifying a mixture with more components.<sup>3</sup> A closely related idea to the modeling framework used in this paper is the compound Markov mixture of normals used by [Geweke and Amisano \(2001\)](#).

The SV-mix model can be thought of as the Euler-scheme approximation of a continuous-time model with regime switching, i.e.,

$$\begin{aligned} dS_t &= \mu_X dt + \sigma_X \exp(V_t/2) dZ_{1t} \\ dV_t &= \psi V_t dt + \sigma_V dZ_{2t}, \end{aligned} \tag{3} \leftarrow$$

where  $dZ_{1t} = \mu_{k_t} dt + \sigma_{k_t} dW_{1t}$  and  $dZ_{2t} = \rho dZ_{1t} + \sqrt{1 - \rho^2} dW_{2t}$  with  $W_{1t}$  and  $W_{2t}$  independent Brownian motions. On each interval  $[t, t + 1)$  for  $t = 1, 2, \dots$ , an iid draw, denoted  $k_t$ , from the discrete distribution on  $1, \dots, K$  with weights  $p_1, \dots, p_K$  determines the particular values  $\mu_{k_t}$  and  $\sigma_{k_t}$  used to generate  $Z_{1t}$  and  $Z_{2t}$  on that interval.

One possible extension of this model would be to draw the  $k_t$  from a stationary Markov chain instead of assuming them to be iid. Another extension might be to treat the timing of the draws as random. These extensions are left for future research.

An alternative would be to think of the SV-mix model simply as descriptive of the dynamics of daily returns and saying nothing regarding the detailed structure of the underlying continuous-time data-generating process.

<sup>3</sup>When the number of components is allowed to increase with sample size, the model is referred to as a Gaussian mixture sieve (see, e.g., [Priebe, 1994](#)).

The SV-mix approach is similar in spirit to the affine-jump models studied by [Bates \(1996\)](#) and many others. A low-probability, high-variance component in the mixture of the SV-mix model plays a similar role to the jumps in the affine-jump models. Because the mixture term is also included in the volatility innovations in the SV-mix model, the approach of this paper is most similar to a model proposed by [Eraker, Johannes, and Polson \(2003\)](#) which has correlated jumps in both the volatility and returns processes. But the jumps in that model are independent of the volatility factor (in both size and intensity), whereas the mixture term of the SV-mix model is scaled by volatility. Affine-jump models where the jump intensity (but not size) depends upon the volatility factor have been studied by, e.g., [Bates \(2005\)](#), [Andersen, Benzoni, and Lund \(2002\)](#), and [Pan \(2002\)](#), but without the inclusion of jumps in the volatility process.

Because the model is estimated over daily data, it is impossible to distinguish whether the large negative returns that appear occasionally in the data are the result of jumps or if the price path is instead continuous. This paper makes no claims about this. Moreover, although one could compute filtered estimates of the probability that a return is a realization from a particular component of the mixture (i.e., whether it includes a jump or not), no attempt to do so is made in this paper. The mixture is regarded simply as a mechanism for generating a flexible family of distributions and the mixture components are not themselves considered to be of interest.

In standard jump-diffusion models, jumps occur in addition to the diffusive part of returns. Hence, unless one were willing to accept that jumps occur nearly every day and explain a larger part of returns than does the diffusion part of the model, it is difficult for these models to capture the full range of shapes for the conditional returns density possible with the SV-mix model.<sup>4</sup>

In a direct comparison with the affine-jump models considered by [Eraker, Johannes, and Polson](#), the SV-mix model provides a large increase in log likelihood and is strongly preferred by standard information-based model choice criteria. The techniques used in this paper corroborate the findings of [Eraker, Johannes, and Polson](#) regarding the importance of including jumps in volatility. The results also support speculation by [Eraker, Johannes, and Polson](#) that allowing for time-varying jump intensity would further improve the performance of their models.

For option-pricing and risk-management applications, the flexibility provided by the mixture of normals approach looks to be especially useful. For example, given a panel of put and call options with varying strike prices, information regarding the predictive density of the underlying security under the risk-neutral measure can be inferred. Out-of-the-money option prices are highly informative about the tails of the risk-neutral measure. The relationship between this risk-neutral measure and the physical measure governing the returns process is of interest. In particular, differences between the physical and risk-neutral measures are associated with risk premia.

But, strong parametric assumptions implied by standard SV models essentially force a certain structure on the tails of the physical measure. If these parametric assumptions do not

---

<sup>4</sup>It is easy to find a jump-diffusion model with the same Euler-scheme approximation as the SV-mix model. The diffusion part of the model would play the role of the mixture component with smallest variance. Adding in jumps would give rise to the larger variance components. Writing the preferred model of this paper in this form would require two jump processes. Jumps would be expected to occur on 86% of days and explain over half of the standard deviation of conditional returns.

accurately reflect the true returns-generating process, difficulties in trying to reconcile the tail-shapes of physical and risk-neutral measures would be expected, potentially resulting in incorrect inference regarding risk premia. If comparisons between these two measures are to be meaningful, models that accurately reflect the distributions seen in the data are needed. The SV-mix modeling approach looks to be a promising way to proceed in this direction.

An extensive literature examining the relationship between physical and risk-neutral measures exists. For example, [Bates \(2000\)](#), using S&P 500 futures and futures option prices to fit a variety of affine and affine-jump models, finds the risk of large negative returns implied by option prices difficult to reconcile with the absence of such events in the returns data. He also finds that a square-root diffusion process driving instantaneous volatility and jump risk is unable to account for the large and typically positive volatility shocks implied by option prices (and speculates that including jumps in volatility could help), that the volatility of volatility needed to match volatility smiles implied by option prices is too high to be consistent with the time-series properties of option prices, and that the models have difficulty simultaneously matching the volatility smiles implied by both short- and long-maturity options. [Pan \(2002\)](#) uses S&P 500 index option prices and returns data to assess some of the same models considered by Bates. She finds evidence of a substantial jump-risk premium and that this jump-risk premium is highly correlated with the level of volatility. In addition to helping explain the volatility smile implied by short-maturity options, she finds this jump-risk premium to explain much of the inconsistency between the level of volatility implied by near-the-money options versus that observed in the spot market. Using S&P 500 index returns alone, [Eraker, Johannes, and Polson \(2003\)](#) find that, once jumps in volatility are included, large risk premia are no longer needed to generate steep IV curves as seen in the data. Building on this work, [Eraker \(2004\)](#), using both returns and option price data, finds little evidence of a significant jump-risk premium. He does find evidence supporting a jump-risk premium in a model with time-varying jump intensity as well as jumps in volatility, but the size of the premium is much smaller than that found by Pan. However, Eraker also finds that none of the models he considers improves much over the basic affine model in fitting option prices. [Andersen, Benzoni, and Lund \(2002\)](#), using S&P index returns data and models that include time-varying jump intensity but not jumps in volatility (similar to Bates and Pan), also find that they are able to generate implied volatility curves similar to those observed in the data without large risk premia. They do not find evidence of state-dependent jump intensity.

There is much variation in both the methodologies and results in this literature. To avoid the analysis being dependent on a specific representation of the risk premia as well, this paper restricts attention to a careful study of the physical measure. Among other things, the paper looks at predictive distributions of cumulative returns over various horizons. Although SV-mix does better than the other models at matching these distributions, some evidence of minor mis-specification is uncovered here that is not apparent using other diagnostics. However, at the 20-day forecast horizon and beyond, there is not much difference among the various models. Analogous forecast distributions under the risk-neutral measure are what determine European option prices with corresponding times to maturity. Thus although the models described in this paper could be helpful in understanding the relationship between physical and risk-neutral measures at short horizons, they are not likely to be of much help at longer horizons.

The SV-mix approach also provides a useful tool for investigating the implications of some interesting hypothetical scenarios. For example, suppose that one wishes to

think about the implications of a peso-type situation involving a low-probability, high-impact event. It is both easy and transparent to add a small lump of mass to one tail of the distribution. One can experiment with the implications of changing the location, scale, and weight of this lump. Conversely, the approach could also be useful in studying what option prices have to say about the market's assessment of the risk of such events. The SV-mix framework provides an alternative to purely nonparametric techniques such as, for example, that proposed by [Jackwerth and Rubinstein \(1996\)](#), imposing some structure while still allowing a great deal of flexibility to match observed option prices.

The rest of this paper is organized as follows. Section 2 gives an overview of the statistical methodology used, Section 3 introduces the models under consideration and fits them to data, Section 4 examines some diagnostics of model fit, Section 5 looks at forecasting performance, Section 6 provides a comparison with some affine/affine-jump models, and Section 7 concludes. An online supplement is available that includes some additional tables and figures as well as appendices examining conditional densities of some continuous-time models (as opposed to their Euler-scheme approximations) and assessing the out-of-sample performance of the SV-mix model.

## 2. Methods

The techniques used for estimation, filtering, and specification analysis in this paper are described in detail in [Durham \(2006\)](#). For completeness, a brief sketch of the techniques is provided here, as well as some details regarding implementation with the SV-mix model. These techniques are closely related to ideas used in [Durbin and Koopman \(1997, 2000\)](#), [Sandmann and Koopman \(1998\)](#), [Shephard and Pitt \(1997\)](#), and [Liesenfeld and Richard \(2003\)](#).

### 2.1. Estimation

Estimation is based on the simulated maximum likelihood approach. The idea is that  $x = (x_1, \dots, x_n)$  is a realization from some random vector  $X = (X_1, \dots, X_n)$  for which direct evaluation of the density function is infeasible, but that there exists an (unobserved) auxiliary random vector  $V = (V_1, \dots, V_n)$  such that the joint density  $p(x, v)$  is easy to evaluate. The likelihood function can then be evaluated at a candidate model parameter vector  $\theta$  by integrating out the auxiliary variable,

$$L(\theta|x) = \int p(x, v; \theta) dv. \quad (4)$$

This is generally a very high-dimensional integral that must be evaluated using Monte Carlo techniques. One draws samples  $v^{(1)}, \dots, v^{(S)}$  from some density  $q$ , referred to as the importance density, and computes

$$\begin{aligned} L(\theta|x) &= \int \frac{p(x, v; \theta)}{q(v)} dQ \\ &\approx \frac{1}{S} \sum_{s=1}^S \frac{p(x, v^{(s)})}{q(v^{(s)})}. \end{aligned} \quad (5)$$

Thus, the likelihood is approximated by a weighted average across “simulated” draws from  $q$ . The estimation step is performed by maximizing the approximate likelihood.

The problem is simply to find a good importance sampler. The importance sampler used in this paper is based on a Laplace approximation to  $p(x, v)$ . One first computes

$$\hat{v} = \underset{v}{\operatorname{argmax}} \log p(x, v) \leftarrow \quad (6)$$

and

$$H = \leftarrow \frac{\partial^2}{\partial v^2} \log p(x, \hat{v}). \quad (7)$$

The importance density is given by the multivariate normal with mean  $\hat{v}$  and variance  $-H^{-1}$ . The mode,  $\hat{v}$ , of  $\log p(x, v)$  is obtained using Newton's method. Although this would appear to be costly because each step involves solving a high-dimensional system of linear equations, the Hessian is positive definite, symmetric, and banded (with the number of sub- and super-diagonals equal to the number of volatility factors). Efficient techniques are available to solve linear systems with this structure. There is never any need to obtain  $H^{-1}$  explicitly. This is important because the inverse does not maintain the banded structure and would require a great deal of effort to compute and an enormous amount of memory to store.

As demonstrated by [Durham \(2006\)](#), this approach is very efficient computationally for standard one- and two-factor SV models. It works equally well with the SV-mix models proposed by this paper. Computational cost for the SV-mix model using a mixture with three components is around 2 second for one evaluation of the likelihood using  $S = 256$  on a data set of 5,615 observations (2 GHz PC). Numerical error can be assessed by repeating the estimation many times using different seeds for the random number generator used in constructing the  $v^{(s)}$ . Given the computational efficiency of the estimator, such studies are relatively painless. Numerical performance of the approach was found to be even better for the SV-mix models than for the models with Gaussian errors and about the same as for the SVt model.

Furthermore, the approach is easy to implement, requiring just a few dozen lines of Fortran code. All that is needed to adapt the estimator to a new model are subroutines providing  $\log p(x, v)$  and its first and second derivatives with respect to  $v$ . For models of form Eq. (2),  $\log p(x, v)$  can be obtained as the sum of terms of form  $\log p(x_{t+1}, v_{t+1}|v_t)$ . The gradient is thus constructed from elements of form

$$\frac{\partial}{\partial v_t} [\log p(x_{t+1}, v_{t+1}|v_t) + \log p(x_t, v_t|v_{t-1})]. \quad (8)$$

The Hessian has diagonal elements of form

$$\frac{\partial^2}{\partial v_t^2} [\log p(x_{t+1}, v_{t+1}|v_t) + \log p(x_t, v_t|v_{t-1})] \leftarrow \quad (9)$$

and off-diagonal elements of form

$$\frac{\partial^2}{\partial v_t \partial v_{t+1}} \log p(x_{t+1}, v_{t+1}|v_t). \quad (10)$$

The formulae for these derivatives can be obtained using Maple or some other symbolic manipulation software. Alternatively, it is tedious though straightforward to obtain them by hand.

## 2.2. Filtering

It is often useful to know the filtered distribution of  $V_t$  conditional on  $x_1, \dots, x_t$ . Among other things, this is needed to compute predictive distributions for returns. An easy way to do this is by means of a particle filter. The particle filter consists of a collection of points  $\{v_t^{(s)}, s = 1, \dots, S\}$  for each  $t = 1, \dots, n$ . The discrete uniform distribution on these points approximates the filter distribution  $V_t | X_1, \dots, X_t$ . The filter is constructed recursively. For simplicity, the discussion below is specialized to models of form Eq. (2).

To initialize the filter, let  $v_0^{(1)}, \dots, v_0^{(S)}$  be draws from the marginal distribution of the volatility factor.<sup>5</sup> To construct the recursion, we need to be able to advance to time  $t + 1$  given the (approximate) filter at time  $t$ . This proceeds as follows. First, compute weights  $w_t^{(s)}$  proportional to  $p(x_{t+1} | v_t^{(s)})$ , normalizing so that they sum to one. Next, resample from the points  $v_t^{(s)}$  using the newly constructed weights. Denote the resampled points  $\tilde{v}_t^{(s)}$ . Finally, draw  $v_{t+1}^{(s)}$  from  $p(v_{t+1} | \tilde{v}_t^{(s)}, x_{t+1})$ . The particle filter has nice convergence properties as  $S \rightarrow \infty$  (e.g., Crisan, 2001).

The particle filter described above is somewhat different from the one used in Durham (2006). The filter in that paper was intended to work with SV models with different timing conventions. The one described here is optimized for the model timing used in this paper. In addition to being simpler to implement and more efficient, using this filter also makes it possible to provide results for the SVt model, which was not feasible using the filter from the other paper.

The particle filter is fast to compute and easy to code. For the SV-mix model with three components and a sample with 5,615 observations, computational cost is about two minutes for a filter with ten thousand particles (2 GHz PC), which is sufficient to provide a very high level of accuracy with this filter. Accuracy is assessed by repeating the computations with different settings for the random number seed and checking that the results do not vary by much. Once again, the computational efficiency of the technique makes such studies relatively painless. Such checks are often omitted when more computationally demanding techniques are used, leading to results that might not be reliable.

## 2.3. Diagnostics

The model diagnostics used in this paper are based on standard time-series residual analysis techniques. The problem is that, due to the presence of the latent state variable, it is not obvious how to obtain the residuals. The idea here is to construct “generalized residuals” using the output of the particle filter. In particular, the density of  $X_{t+1} | x_1, \dots, x_t$  can be estimated by

$$\hat{p}(x_{t+1} | x_1, \dots, x_t) = \frac{1}{S} \sum_{s=1}^S p(x_{t+1} | v_t^{(s)}) \quad (11)$$

Similarly, its cumulative distribution function can be estimated by

$$z_{t+1} = \text{prob}(X_{t+1} \leq x_{t+1} | x_1, \dots, x_t) = \frac{1}{S} \sum_{s=1}^S \text{prob}(X_{t+1} \leq x_{t+1} | v_t^{(s)}) \quad (12)$$

<sup>5</sup>For the standard SV models, the volatility factor is AR(1) with Gaussian innovations, so the marginal distribution is easy to determine. For the SV-mix model, the marginal distribution can be approximated by simulation (because the process is ergodic).

If the model is correctly specified, the  $z_t$  should be iid uniform(0,1). The generalized residuals are obtained by applying the inverse of the normal cumulative distribution function,  $\tilde{z}_t = \Phi^{-1}(z_t)$ . For a correctly specified model, these should be iid  $N(0, 1)$ . The usual kinds of tests can now be done to test the hypothesis that the model is correctly specified, e.g., Jarque-Bera or some other test to assess the unconditional distribution of the  $\tilde{z}_t$  and Box-Pierce tests or standard tests for autoregressive conditional heteroskedasticity (ARCH) behavior to look for dynamic structure.

A great deal of work has appeared recently dealing with more sophisticated approaches to specification testing based on generalized residuals. Papers include [Bontemps and Meddahi \(2005\)](#), [Duan \(2003\)](#), [Bai \(2003\)](#), [Hong and Li \(2002\)](#), and [Diebold, Gunther, and Tay \(1998\)](#). Many of the ideas proposed in these papers could be applied within the context described above. But such work is not undertaken here.

### 3. Data and estimation

This section examines the performance of the SV-mix model over daily S&P 500 index data from June 23, 1980 to September 2, 2002 ( $n = 5,616$ ). The data exhibit a small amount of autocorrelation, possibly stemming from nonsynchronous trading of the individual stocks comprising the index. One way to remove this correlation is by passing the data through an autoregressive moving average (ARMA) filter. This is the approach taken by, for example, [Andersen, Benzoni, and Lund \(2002\)](#). The empirical results reported in this paper are all based on data that have 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 4. [Fig. 1](#) shows some plots of the data.

The models under consideration are described in [Table 1](#). SV1 and SV2 are, respectively, the standard one- and two-factor log volatility models with normal errors. SVt is a one-factor model, but with Student- $t$  instead of normal errors in the returns process. SV0, which treats returns as iid normal, is included for reference. SV-mix models with two and three components are referred to as MIX2 and MIX3, respectively. All of the models include correlation between returns and volatilities. This correlation has been found to be an important feature of the data (commonly referred to as the leverage effect).

Estimates and log likelihoods for the models are shown in [Table 2](#). Because the models are not nested, it is not straightforward to test one model against the other directly using, e.g., likelihood ratio tests. However, the log likelihood still provides a means of assessing model fit in terms of the Kullback-Leibler information of the data relative to the fitted model.

The danger here is over-fitting, that is, favoring models that describe artifacts of the sample that are neither features of the actual data-generating process nor economically useful. One approach to the problem of model choice is to use some information criterion based on the log likelihood plus a penalty function that depends on the number of estimated parameters. Common choices include the Akaike information criterion (AIC) and Schwarz criterion (SC), which call for penalties of one and  $\log \sqrt{n}$  (points per parameter, respectively (the SC penalty is 4.31 points per parameter for the data set under consideration here). The results of this paper are mostly clearcut enough to render the choice of information criterion moot.

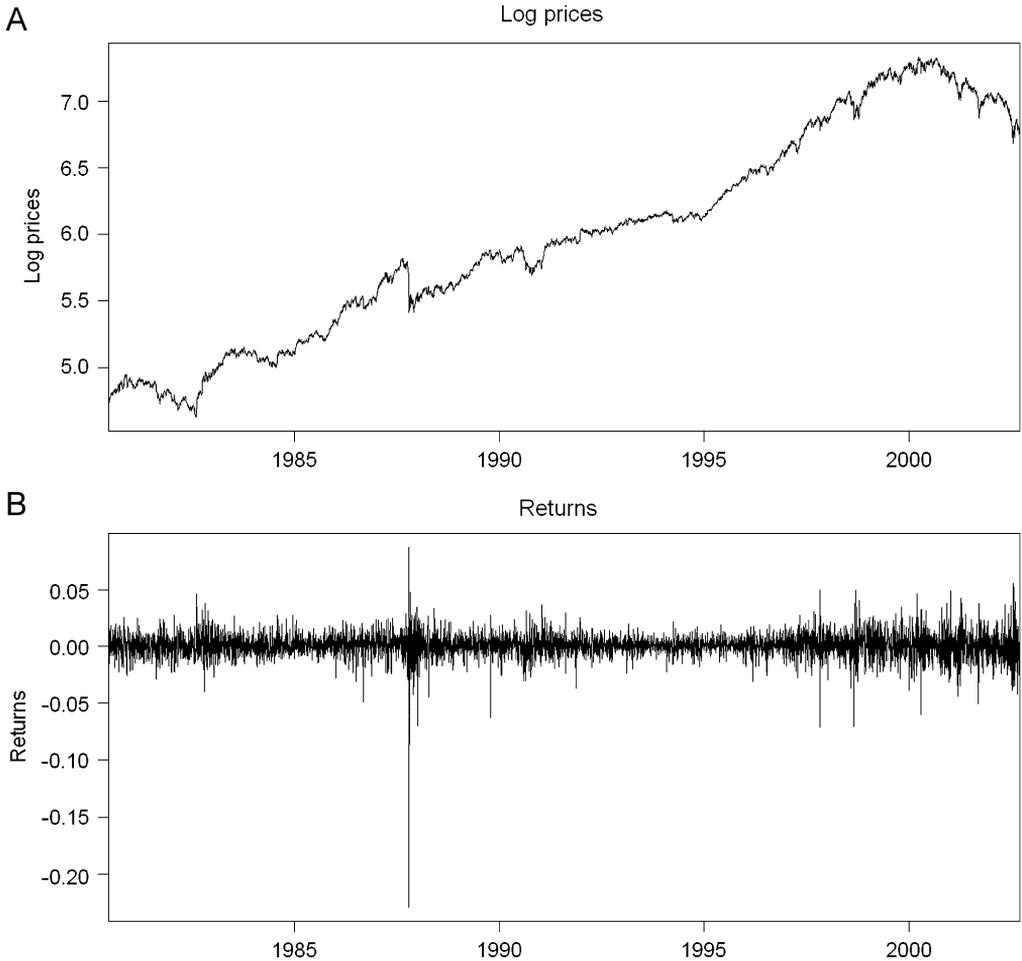


Fig. 1. Standard & Poor's 500 index, June 23, 1980 to September 2, 2002.

In practice, none of these models might be the true data-generating process. Of interest is whether they describe the data in an economically useful manner. This issue is addressed in more detail in subsequent sections.

SV2 provides a substantial increase in log likelihood over SV1. Somewhat surprisingly, SVt does even better than the more complicated (and more commonly used) two-factor model. The log likelihood of MIX2, the simplest of the mixture models, is about the same as that of SVt. Both SC and AIC prefer SVt because it has two fewer parameters. MIX3 provides a large increase in log likelihood over MIX2 and is the preferred model overall. Adding a fourth component to the mixture was found to provide essentially no further benefit.

All of the models show volatility to be highly persistent with autoregressive coefficients around 0.98–0.99. This is robust to model specification. The correlation between returns and innovations to this persistent volatility process is around  $-0.5$ . This is also robust.

Table 1  
Model specifications

Model	Specification	Error terms
SV0	$X_t = \mu_X + \sigma_X \varepsilon_t$	$\varepsilon_t \sim \mathcal{N}(0, 1) \leftarrow$
SV1	$X_t = \mu_X + \sigma_X \exp(V_{t-1}/2) \varepsilon_t$ $V_t = \phi V_{t-1} + \sigma_V \eta_t$	$\varepsilon_t, \eta_t \sim \mathcal{N}(0, 1)$ $\text{corr}(\varepsilon_t, \eta_t) = \rho$
SV2	$X_t = \mu_X + \sigma_X \exp(U_{t-1}/2 + V_{t-1}/2) \varepsilon_t$ $V_t = \phi_V V_{t-1} + \sigma_V \eta_t$ $U_t = \phi_U U_{t-1} + \sigma_U \zeta_t$	$\varepsilon_t, \eta_t, \zeta_t \sim \mathcal{N}(0, 1)$ $\text{corr}(\varepsilon_t, \eta_t) = \rho_V$ $\text{corr}(\varepsilon_t, \zeta_t) = \rho_U$
SVt	$X_t = \mu_X + \sigma_X \exp(V_{t-1}/2) \varepsilon_t$ $V_t = \phi V_{t-1} + \sigma_V \eta_t$	$\varepsilon_t \sim t_\nu, \zeta_t \sim \mathcal{N}(0, 1) \leftarrow$ $\text{corr}(\varepsilon_t, \zeta_t) = 0$
SV-mix	$X_t = \mu_X + \sigma_X \exp(V_{t-1}/2) \varepsilon_t$ $V_t = \phi V_{t-1} + \sigma_V \eta_t$	$\eta_t = \rho \varepsilon_t + \sqrt{1 - \rho^2} \zeta_t \leftarrow$ $\varepsilon_t \sim \text{Normal mixture}(0, 1) \leftarrow$ $\zeta_t \sim \mathcal{N}(0, 1)$ $\text{corr}(\varepsilon_t, \zeta_t) = 0$ $\eta_t = \rho \varepsilon_t + \sqrt{1 - \rho^2} \zeta_t \leftarrow$

Fig. 2 shows the log forecast density of  $X_{t+1}$  that is implied by each of the fitted models conditional on  $V_t = 0$ . Conditioning on a different value for  $V_t$  would change the scale of the densities but have little effect on the shape. This conditional distribution is of considerable interest for option-pricing and risk-management applications. The mixture models are particularly interesting because they provide something close to a nonparametric look at this distribution. The long left tail of the mixture distributions captures the occasional crash days. Over the rest of their support, the mixture distributions are remarkably close to normal. As shown by the estimates in Table 2, the mixture distributions exhibit skewness around  $-0.5$  and kurtosis around 6. These estimates are reasonably robust to the number of mixture coefficients.

In MIX3, the crash state ( $k = 2$ ) has a mean of  $-3.6$ , standard deviation of 2.7, and probability of 0.004. Thus this state would be expected to occur about once per year on average. In particular, this state captures not just the well-known crashes such as October 19, 1987 and October 13, 1989, but many less extreme events as well.

For comparison, Fig. 2 also shows the conditional log densities corresponding to the SV1, SV2, and SVt models. The SV1 conditional density is Gaussian. The conditional density of SV2 is obtained by integrating across the marginal density of  $U_t$ . These densities are all symmetric. They lack the long left tail of the mixture densities. The densities implied by SVt and SV2 are similar in shape. Both are substantially fatter than the mixture models in the right tail.

Panel C of Fig. 2 zooms in on the region around the mode of the conditional densities. MIX3 includes a component that is concentrated around zero ( $k = 3$ ). This makes the density slightly more peaked than the normal. In fact, it is very close to the SVt model ( $t$  with about 8 degrees of freedom) in this region. In standard jump-diffusion models, jumps occur in addition to the diffusive part of returns. Thus, although it would be possible to construct a jump-diffusion model with the same Euler-scheme approximation as MIX3, it would require two jump components, with jumps expected to occur on 86% of days and explaining over half of the standard deviation of conditional returns.

Table 2

Parameter estimates, Standard & Poor's 500 index returns, June 23, 1980 to September 2, 2002

Data are prefiltered using an ARMA(2,1) model. The skewness and kurtosis columns for the mixture models (Panel C) correspond to the densities implied by the estimated mixture parameters. The parameter estimates for the mixture components are shown in Panel D. No standard errors are associated with the first mixture component. These parameters are implied by the constraints that the mixture probabilities sum to one and that the mixture have mean zero and variance one.

Panel A									
Model	Log $L$	$\mu$	$\sigma_X$	$\phi_V$	$\sigma_V$	$\rho_{21}$	$\phi_U$	$\sigma_U$	$\rho_{31}$
SV0	17569.66	0.00000 (0.00000)	0.01059 (0.00000)						
SV1	18533.34	0.00007 (0.00011)	0.00855 (0.00041)	0.9762 (0.0045)	0.170 (0.015)	-0.432 (0.047)			
SV2	18575.86	0.00008 (0.00010)	0.00830 (0.00063)	0.9905 (0.0027)	0.101 (0.014)	-0.459 (0.088)	0.15 (0.27)	0.468 (0.066)	-0.215 (0.112)
Panel B									
Model	Log $L$	$\mu$	$\sigma_X$	$\phi_V$	$\sigma_V$	$\rho_{21}$	$\nu$		
SVt	18582.24	0.00008 (0.00010)	0.00755 (0.00052)	0.9895 (0.0026)	0.103 (0.011)	-0.488 (0.052)	8.09 (0.89)		
Panel C									
Model	Log $L$	$\mu$	$\sigma_X$	$\phi_V$	$\sigma_V$	$\rho_{21}$	skew	kurt	
MIX2	18583.99	-0.00003 (0.00011)	0.00893 (0.00055)	0.98551 (0.00311)	0.129 (0.012)	-0.579 (0.045)	-0.46	5.55	
MIX3	18604.57	-0.00004 (0.00011)	0.00902 (0.00058)	0.9872 (0.0029)	0.118 (0.010)	-0.577 (0.047)	-0.46	6.37	
Panel D									
Model	$\log p_1$	$\mu_1$	$\sigma_1$	$\log p_2$	$\mu_2$	$\sigma_2$	$\log p_3$	$\mu_3$	$\sigma_3$
MIX2	-0.011	0.023	0.947	-4.558 (0.572)	-2.195 (1.062)	2.440 (0.381)			
MIX3	-0.176	0.015	1.029	-5.504 (0.411)	-3.611 (1.247)	2.651 (0.595)	-1.848 (0.242)	0.015 (0.048)	0.444 (0.058)

One problem with the mixture of normals approach is how to assess the precision of the estimated mixture density in an intuitively meaningful way. Although the usual standard errors are available for the parameter estimates, the complex interactions between the model parameters make it difficult to gauge the extent to which parameter uncertainty translates into uncertainty regarding the shape of the density itself.

Bayesian techniques provide a useful approach to addressing this issue (see, e.g., Escobar and West, 1995). Let  $\psi$  denote the vector of parameters that determines the mixture density, and let  $\vartheta$  denote the remaining model parameters. Let  $f(\cdot; \psi)$  be the mixture density implied by  $\psi$ . The goal is to assess the range of shapes for  $f$  corresponding to choices for  $\psi$  that are consistent with the data. Fixing  $\vartheta$  at the maximum likelihood estimate, and given a prior for  $\psi$ , it is possible to draw a sample,  $\{\psi^{(s)}, \psi^s = \psi, s = 1, \dots, S\}$ , from the posterior distribution of  $\psi$  conditional on the data using, e.g., the Metropolis-Hastings algorithm. Examination of the collection of densities implied by this sample,

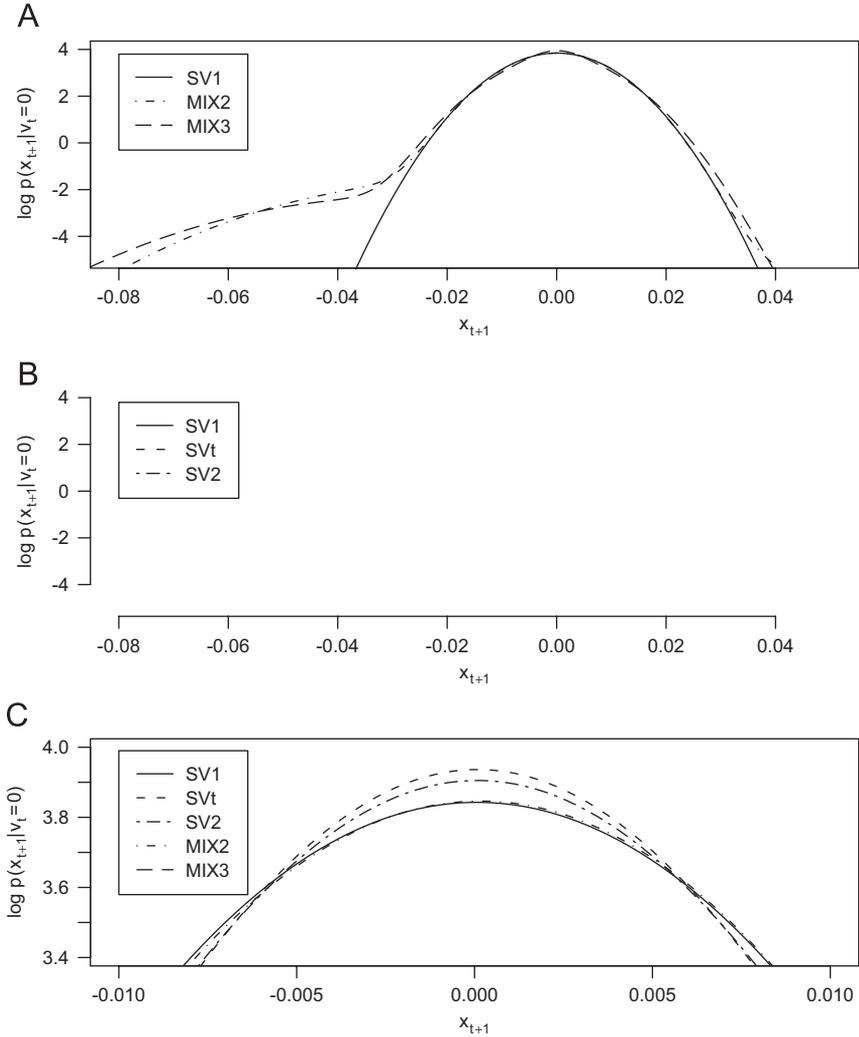


Fig. 2. Log density of  $X_{t+1}$  conditional on  $V_t = 0$  for various models. Models are defined in Table 1. MIX2 and MIX3 are SV-mix models with 2 and 3 components, respectively. Panel C zooms in on the region near the mode.

$\{f^{(s)} \leftarrow f(\cdot; \psi^{(s)}), \psi^{(s)} = 1, \dots, S\}$ , provides information regarding the range of plausible shapes for  $f$ .

A convenient way to summarize this information is as follows. Given a fixed point,  $u_0$ , and  $\alpha \in (0, 1)$ , let  $f_{\alpha}(u_0)$  be the  $\alpha$  quantile of  $\{f^{(s)}(u_0), \psi^{(s)} = 1, \dots, S\}$ . Then, for example, the interval  $(f_{0.025}(u_0), f_{0.975}(u_0))$  can be interpreted as a 95% Bayesian confidence interval for the value of the mixture density evaluated at  $u_0$ .<sup>6</sup> Repeating this process across some range

<sup>6</sup>Bayesian confidence intervals (also referred to as credible intervals or credible sets) are closely related to classical confidence intervals. The advantages in the present context are that the Bayesian confidence intervals are more straightforward to compute, and they are (in some sense) exact, whereas the classical analogs are based on asymptotics that are not likely to be good in this case because the log likelihood surface is highly nonquadratic in

of values for  $u_0$ , it is possible to construct a pointwise confidence interval for the entire density.

The procedure described above was implemented for the MIX3 model. A simple random walk Metropolis-Hastings algorithm was used to obtain draws from  $\psi$  conditional on the data, with  $\vartheta$  fixed at its maximum likelihood estimate, and a flat prior on  $\psi$ .<sup>7</sup> The algorithm was run for ten thousand iterations and the output was checked to confirm that the parameter space was adequately covered.

The upper panel of Fig. 3 shows a selection of some of the densities obtained in this manner. The (pointwise) interquartile range and 95% confidence band for the mixture density corresponding to the MIX3 model is plotted in the lower panel of Fig. 3. The density is estimated precisely in the center of the distribution, where there are many observations. Not surprisingly, the tails are estimated less precisely, because the data are much sparser. Nonetheless, the basic shape of the density is well defined. Most of the uncertainty occurs beyond 4 standard deviations from the mean.

The same draws for the mixture parameter can also be used to compute confidence intervals for other quantities of interest. For example, the posterior density for the probability of the event ( $\varepsilon < -4$ ) has mean 0.0017 and interquartile range [0.0012, 0.0021]. The 95% confidence interval is [0.0006, 0.0032]. The point estimate of the probability of this event (based on the maximum likelihood parameter estimates) is 0.0021.

The event ( $\varepsilon < -4$ ) corresponds loosely to crash days (for MIX3, this is equivalent to a return of less than  $-0.0362$  conditional on  $V_t = \Theta$ ). For comparison, the maximum likelihood estimates for the probability of this event are 0.000011 for SV1, 0.00029 for SV2, and 0.00065 for SVt. Thus, a practitioner whose analysis was based on the SV1 model, for example, would believe the risk of such crash days to be more than one hundred times less likely than the practitioner using the MIX3 model.

#### 4. Diagnostics

This section looks at some diagnostics for model fit based on ideas discussed in Section 2. Fig. 4 plots the generalized residuals for several models. Correlograms for the squared generalized residuals are shown in Fig. 5.  $QQ$ -plots of the generalized residuals against the standard normal distribution are shown in Fig. 6. Table 3 shows results of Jarque-Bera tests for normality of the generalized residuals, Box-Pierce tests for autocorrelation in the squared generalized residuals, and LM tests for the presence of ARCH effects in the generalized residuals.

---

*(footnote continued)*

the mixture parameters (McLachlan and Peel, 2000). Wald confidence intervals computed using the delta method are likely to be particularly poor. Carlin and Louis (2000, Section 4.3) show that in some applications Bayesian confidence intervals can have better frequentist properties than the usual frequentist confidence intervals. Furthermore, given some regularity conditions, the posterior distribution for large  $n$  is approximately normal with mean equal to the posterior mode and covariance equal to the negative inverse Hessian, thus the Bayesian confidence interval is asymptotically equivalent to commonly used frequentist confidence intervals.

<sup>7</sup>Fixing  $\vartheta$  at its maximum likelihood estimate focuses attention on uncertainty in the mixture density itself, isolating it from interactions with the remaining model parameters. Alternatively, it would be possible to integrate across the posterior of  $\vartheta$  (given some prior). Although this does not pose any technical difficulties, it is less clear how to interpret the results. At any rate, for the application examined here, it does not make much difference because  $\vartheta$  is precisely estimated.

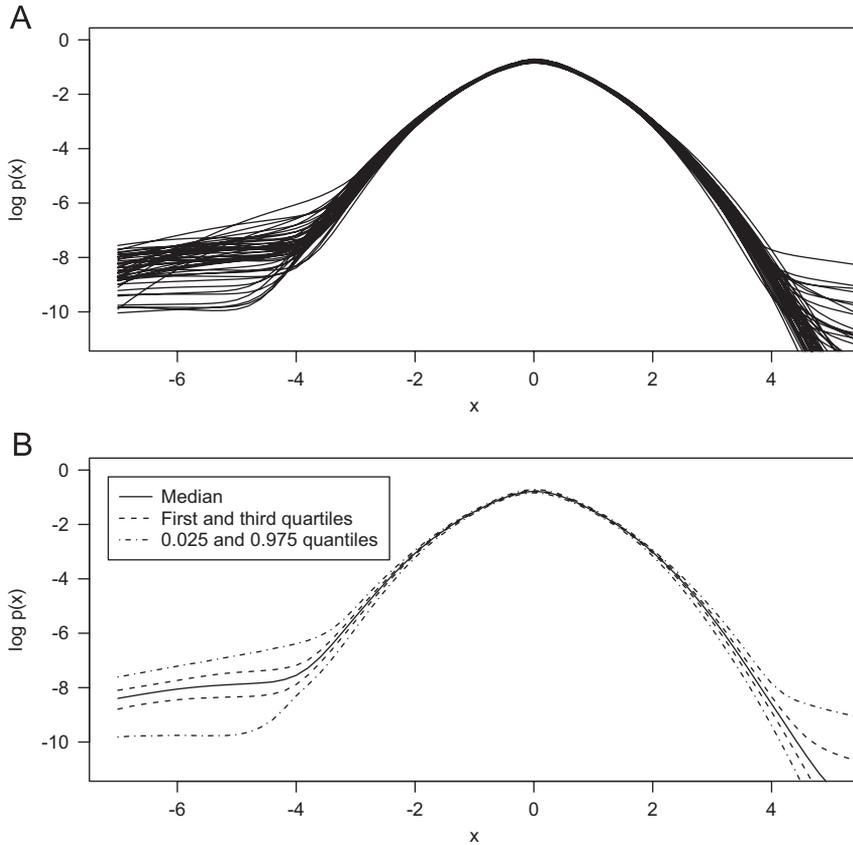


Fig. 3. Panel A shows a sample of MIX3 log mixture densities corresponding to draws from the posterior distribution of the mixture parameters. Panel B shows Bayesian confidence bands for the log mixture density corresponding to MIX3.

As expected, the correlogram corresponding to the SV0 model indicates a great deal of persistence in volatility. In contrast, the correlograms corresponding to the other models suggest that they are all largely successful in filtering out this persistence. The Box-Pierce test on 20 lags rejects none of the models at conventional significance levels. However, if more lags are considered, a small amount of long-term persistence in the squared residuals can be detected. All of the models have  $p$ -values around 0.01 for the Box-Pierce test on 250 lags. This is consistent with long memory in volatility, as suggested by [Ding and Granger \(1996\)](#), [Bollerslev and Mikkelsen \(1996\)](#), [Gallant, Hsieh, and Tauchen \(1997\)](#), and others. None of the models successfully captures this feature of the data.

I now turn to the marginal distribution of the generalized residuals. Looking at the  $qq$ -plots in [Fig. 6](#), the results are much in line with what one might expect from the density plots in [Fig. 2](#). Although SV1, SVt, and SV2 improve enormously upon SV0, they fail to capture the fat left tail of the conditional returns distribution adequately. SVt does slightly better than SV2 in the left tail but is too fat in the right tail. It is easy to see from these plots where the models and the data begin to diverge. It is even possible to determine exactly

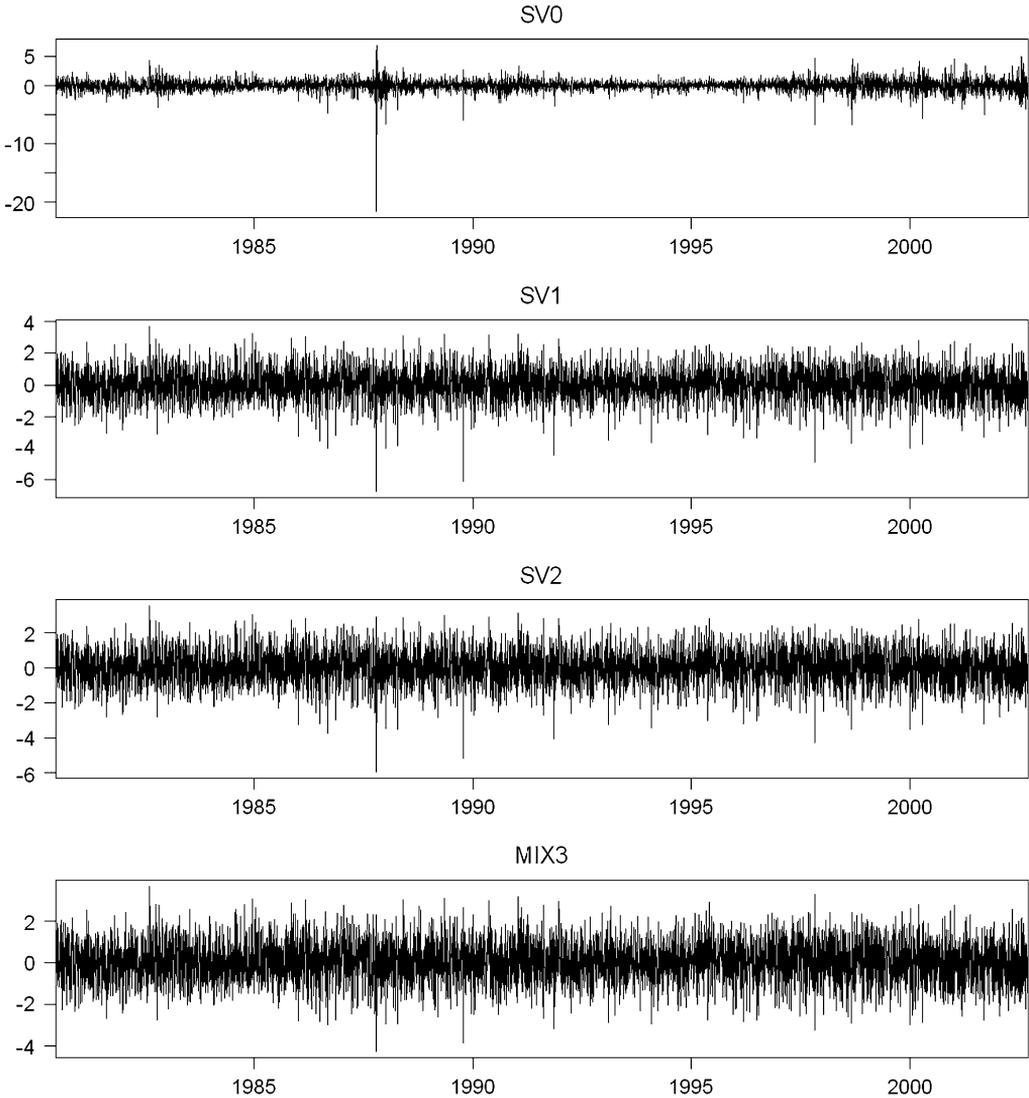


Fig. 4. Generalized residuals, Standard & Poor's 500 index returns, June 23, 1980 to September 2, 2002.

which observations are involved. The lack of fit affects considerably more than just the several well-known extreme events that appear in the sample.

The SV-mixture models do a much better job of matching the data. Although neither MIX2 nor MIX3 is able to explain the crashes of October 19, 1987 or October 13, 1989 completely, the MIX3 model is almost perfect over the rest of the distribution. The MIX2 model does slightly worse at explaining the two crash days and is also too thin in the right tail.

Looking at the Jarque-Bera tests in Table 3, SV1 and SV2 are overwhelmingly rejected. SVt is barely rejected at the 99% level. MIX2 is rejected at the 95% but not the 99% level, while MIX3 is not in danger of rejection at any conventional level.

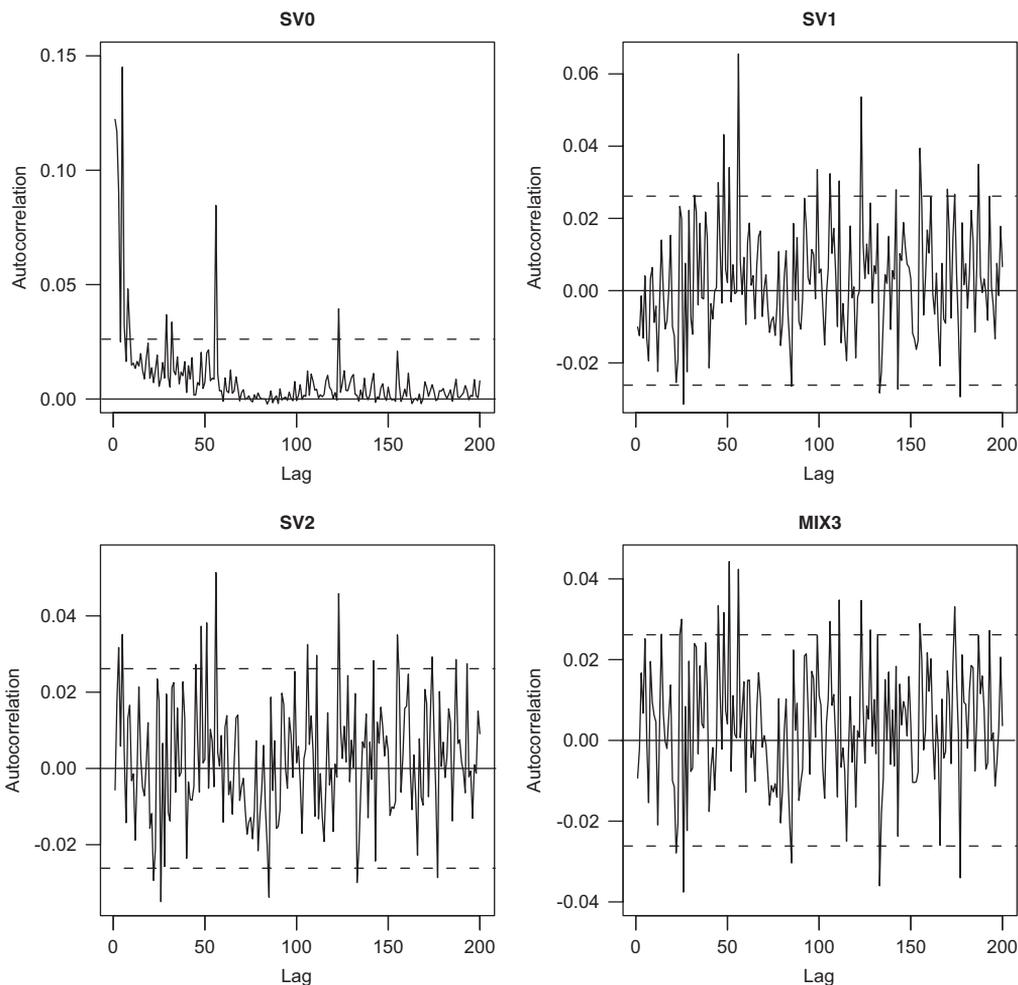


Fig. 5. Correlograms for squared generalized residuals. Two hundred lags are shown. The dashed lines mark the rejection region (5% significance level) for the individual correlation coefficients.

The *qq*-plot for the SV0 model in Fig. 6 shows that the frequency of large positive returns is similar to that of large negative returns (see also Schwert, 1990). Thus it is surprising that the mixture densities fatten the left tail but not the right. The explanation for this appears to be that large positive returns tend to occur when volatility is already relatively high. Once the volatility state is accounted for, it looks as though the innovations driving the price process,  $\varepsilon_t = (X_t - \mu_X) / [\sigma_X \exp(V_{t-1}/2)]$ , are about what would be predicted by a normal distribution in the right tail (but not the left). This explanation is supported by the *qq*-plot for SV1 in Fig. 6, which shows little evidence of misspecification in the right tail. A close examination of the time series of returns in the bottom panel of Fig. 1 lends additional support. Thus, although the right and left tails of the unconditional returns distribution could be similar, the tails of the conditional distribution are not.

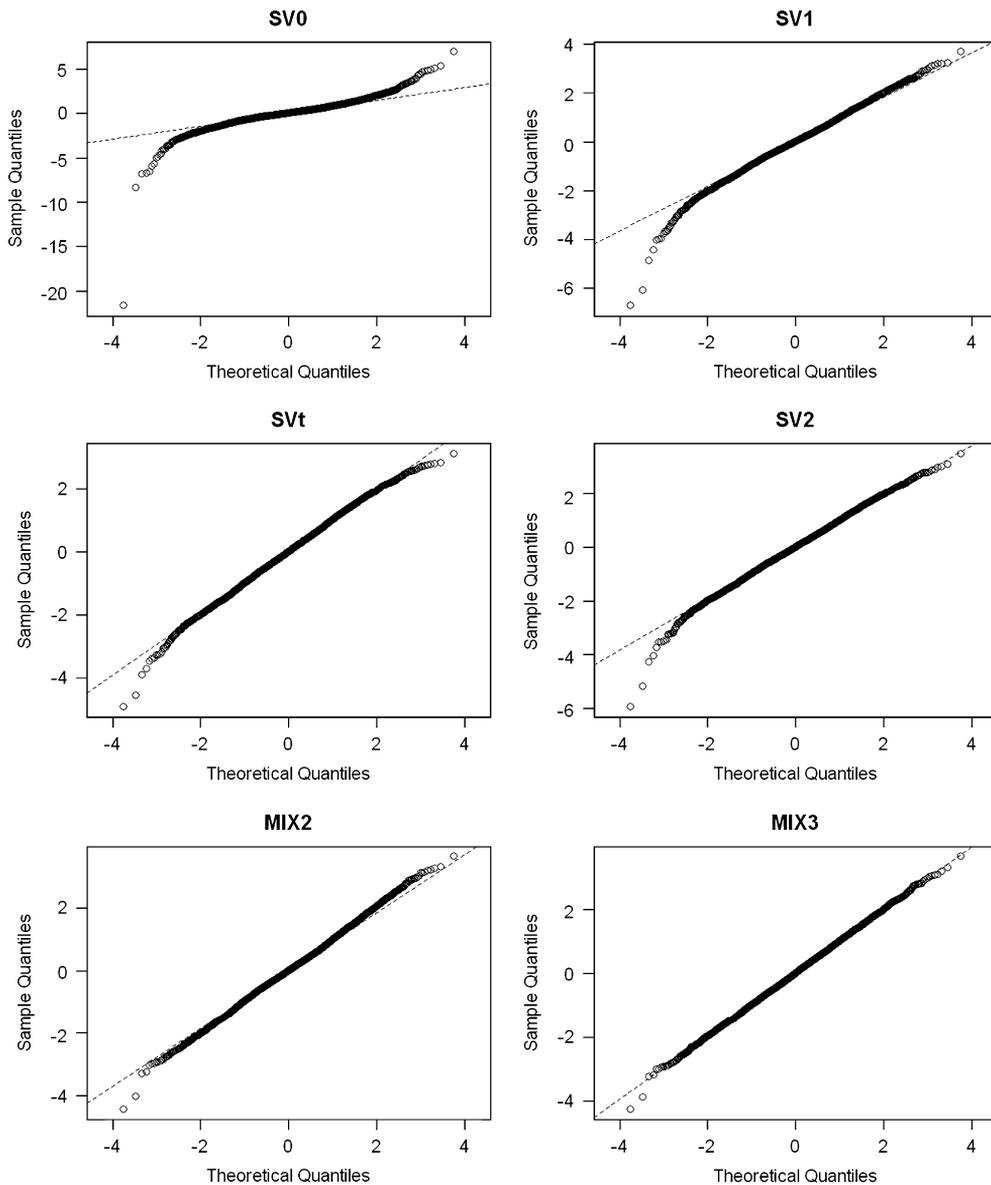


Fig. 6.  $QQ$ -plots of generalized residuals against the standard normal, Standard & Poor's 500 index returns, June 23, 1980 to September 2, 2002.

The SV-mix model follows the work of EJP in that large negative returns (jumps) are associated with large increases (jumps) in volatility. The flip side of this is that large positive returns are associated with decreases in volatility. It is unclear if this is a feature of the data or just an artifact of the linear correlation relationship between  $\varepsilon_t$  and  $\eta_t$ , the innovations driving the price and volatility processes, respectively. Although some evidence regarding the possibility of nonlinearity in this relationship is provided in Section 5, a more detailed investigation is left for future work. At any rate, because the

Table 3

Tests for autoregressive conditional heteroskedasticity (ARCH) effects and normality of generalized residuals, Standard & Poor's 500 index returns, June 23, 1980 to September 2, 2002

The ARCH test is based on 20 lags. Box-Pierce tests are based on 20 and 250 lags of the squared generalized residuals. *P*-values are shown in parentheses.

Model	Jarque-Bera	ARCH(20)	Box-Pierce(20)	Box-Pierce(250)
SV0	439,869.1 (0.0000)	272.70 (0.0000)	368.10 (0.0000)	479.4943 (0.0000)
SV1	248.19 (0.0000)	13.09 (0.8735)	13.28 (0.8651)	316.44 (0.0028)
SV2	51.94 (0.0000)	24.92 (0.2046)	25.60 (0.1794)	307.48 (0.0076)
SVt	10.96 (0.0042)	18.65 (0.5446)	19.36 (0.4983)	305.43 (0.0095)
MIX2	8.13 (0.0171)	15.48 (0.7485)	16.22 (0.7030)	301.33 (0.0145)
MIX3	0.19 (0.9079)	17.92 (0.5926)	18.70 (0.5412)	305.96 (0.0090)

right tail of the mixture density is close to Gaussian, the link between positive returns and volatility decreases in the SV-mix model is about the same as that implied by the usual leverage effect in, say, the standard SV1 model.

## 5. Forecasting

The conditional distributions for returns described in Section 4 are all based on the idea that  $V_t$  is known (i.e.,  $V_t = 0$  for the plots in Fig. 2). In practice, this information is not available and must be inferred from past returns. The appropriate forecast density operationally is  $p(x_{t+1}|x_t, x_{t-1}, \dots, x_1)$ . Fortunately, the output of the particle filter provides an easy way to estimate this. For each  $t$ , the particle filter provides draws  $\{v_t^{(1)}, \dots, v_t^{(S)}\}$  from  $V_t|X_t, X_{t-1}, \dots, X_1$ . The desired forecast density can thus be computed as

$$p(x_{t+1}|x_t, x_{t-1}, \dots, x_1) = \int p(x_{t+1}|v_t) dp(v_t|x_t, x_{t-1}, \dots, x_1) \leftarrow \frac{1}{S} \sum_{s=1}^S p(x_{t+1}|v_t^{(s)}) \leftarrow \tag{13}$$

As these expressions show, the forecast density is a mixture of the conditional densities  $p(x_{t+1}|v_t)$  across possible realizations of  $V_t$ , weighted according to their likelihood based on past returns.

Forecast densities for August 1, 1991 computed using this approach with various models are shown in Fig. 7 (this arbitrarily chosen date is at the midpoint of the sample period). As compared with the densities shown in Fig. 2 (which treat  $V_t$  as known with certainty), accounting for uncertainty in the level of volatility fattens the tails of all the models. Relative differences between the various models, however, remain qualitatively similar.

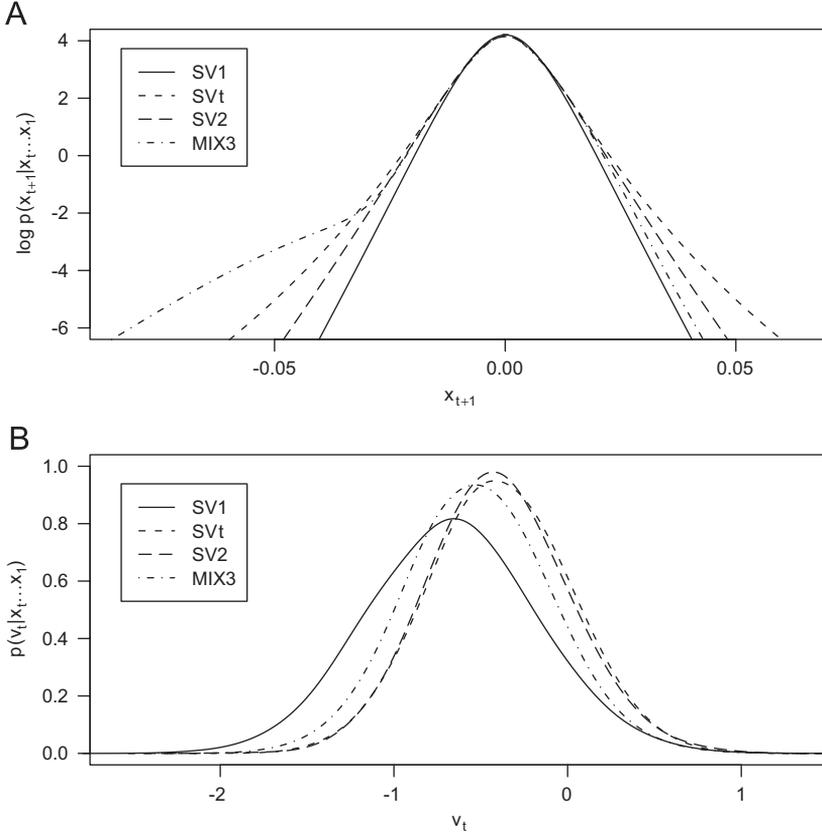


Fig. 7. Forecast densities. Panel A shows forecast density for  $X_{t+1}|X_t, \dots, X_1$  for August 1, 1991. Panel B shows the filtered density for  $V_t|X_t, \dots, X_1$  for the same date.

The filtered densities of  $V_t|X_t, X_{t-1}, \dots, X_1$  implied by various models for the date August 1, 1991 are shown in the lower panel of Fig. 7. Time-series plots of filtered volatility estimates are shown in Fig. 8 (see also Fig. 20 in Appendix A). The models differ mostly in how they react to extreme returns. In particular, because it has a thinner-tailed conditional returns distribution, SV1 requires higher levels of volatility to accommodate such observations. But, overall, volatility estimates are reasonably robust to model specification.

A similar approach can be used to compute predictive densities over longer time periods. Let  $X_t^{t+k}$  denote the cumulative log return over a period of  $k$  days, beginning at time  $t$ , and suppose the predictive density of  $X_t^{t+k}|X_1, \dots, X_t$  is desired. For simplicity, restrict attention to the SV1 model for now. As before, let  $\{v_t^{(s)}, s = 1, \dots, S\}$  be draws from the particle filter. For each particle, simulate a volatility path,  $\{v_t^{(s)}, \dots, v_{t+k}^{(s)}\}$  (do not condition on values of  $X_\tau$  for  $\tau > t$ ). Now, the problem is to compute

$$p(x_t^{t+k}|x_1, \dots, x_t) = \int p(x_{t+k}|v_t, \dots, v_{t+k}) dp(v_t, \dots, v_{t+k}|x_1, \dots, x_t) \\ = \frac{1}{S} \sum_{s=1}^S p(x_t^{t+k}|v_t^{(s)}, \dots, v_{t+k}^{(s)}). \quad (14)$$

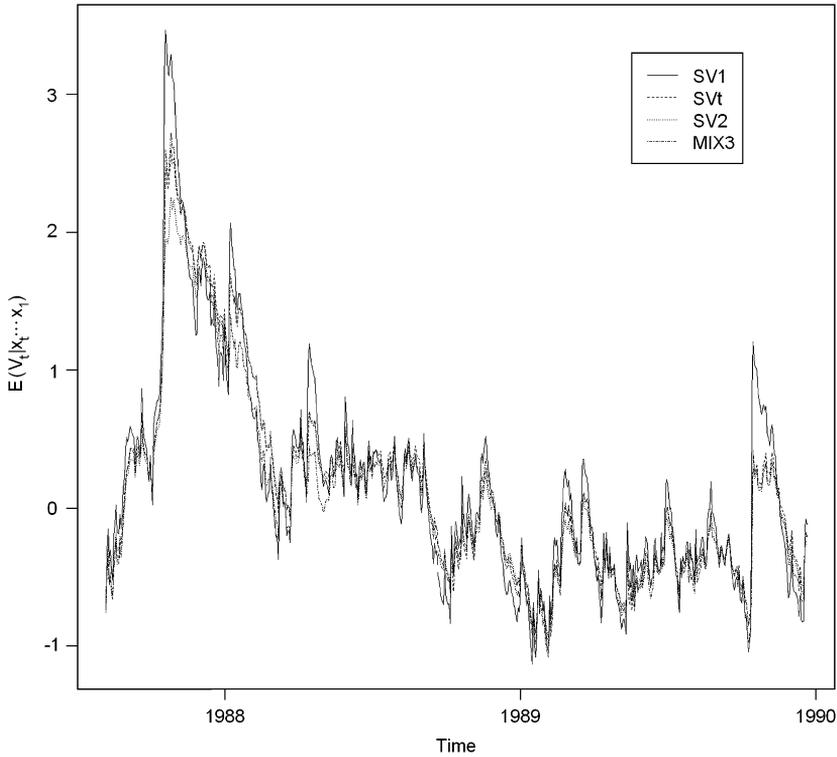


Fig. 8. Filtered volatility estimates for various models.

But this is straightforward to do because the conditional distributions are Gaussian. Fixing  $s$  for the moment, let  $\eta_\tau^{(s)\leftarrow}$  be the innovation used to generate  $v_\tau^{(s)}$  for  $t < \tau \leq t+k$ . Then the predictive distribution for the daily log return  $X_\tau$  conditional on the simulated path  $v_t^{(s)\leftarrow}, \dots, v_{t+k}^{(s)\leftarrow}$  is normal with mean  $\mu_X + \sigma_X \exp(v_{\tau-1}^{(s)\leftarrow}/2) \rho \eta_\tau$  and variance  $\sigma_X^2 \exp(v_{\tau-1}^{(s)\leftarrow}) (1 - \rho^2)$ . The cumulative log return  $X_t^{t+k}$  conditional on the simulated volatility path is normal with mean and variance equal to the sum of the daily means and variances, respectively.

Predictive densities for the other models can be constructed in a similar manner but with a little more work. For example, for the SV-mix model, the mixture states need to be simulated in addition to the volatility. The same idea works for the SVt model, because a  $t$  distribution is just an inverse gamma mixture of normals. For the SV2 model, paths for both volatility factors must be simulated.

This idea of computing an unconditional distribution by averaging over conditional distributions is commonly referred to as Rao-Blackwellization (e.g., Robert and Casella, 2004, Section 4.2).

Predictive distributions for five- and ten-day cumulative returns are shown for the date August 1, 1991 in Fig. 9 (see Fig. 21 in Appendix A for 15- and 20-day forecast distributions). Compared with the one-day predictive densities shown in Fig. 7, the

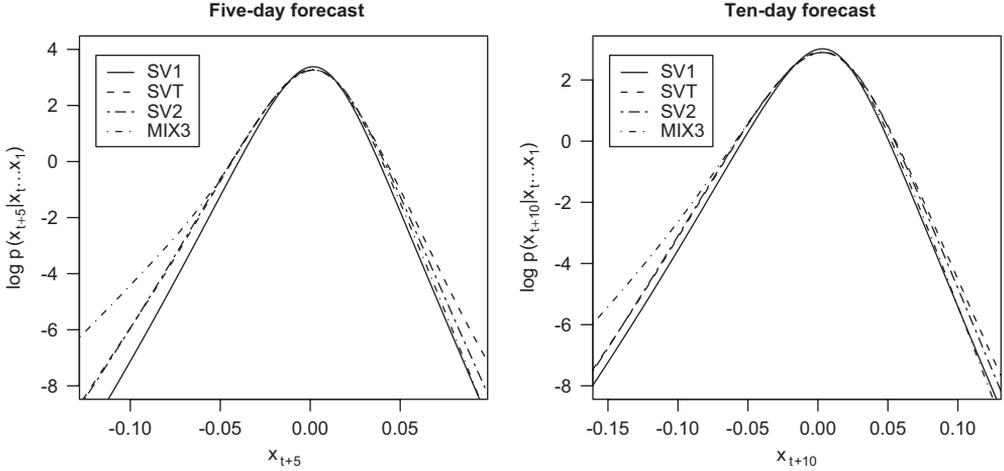


Fig. 9. Forecast densities of cumulative returns at five- and ten-day horizons for August 1, 1991.

predictive densities for all of the models become closer to normal as the prediction interval grows. Nonetheless, the left tail of the SV-mix predictive density remains somewhat fatter than the other models out to at least ten days (two weeks). By 20 days out, not much difference between the various models remains.

It is possible to assess how well these predictive densities fit the data using an approach along the lines of the diagnostics described in Section 4. In particular, because the predictive densities can be computed, it is also straightforward to compute the quantities

$$z_t^{t+k} = \text{prob}(X_t^{t+k} \leq x_t^{t+k} | x_1, \dots, x_t), \quad (15)$$

where  $x_t^{t+k}$  is the observed cumulative log return. As before, the generalized residuals,  $\tilde{z}_t^{t+k} = \Phi^{-1}(z_t^{t+k})$ , should be normally distributed with mean zero and variance one if the model is correctly specified. But in contrast to the previously discussed situation, they would be autocorrelated (because the periods over which the cumulative returns are computed overlap).<sup>8</sup>

Figs. 10 and 11 show *qq*-plots for the generalized residuals of five- and ten-day cumulative returns, respectively. At the five-day horizon, the MIX3 model fits the left tail of the data better than the other models, but, somewhat surprisingly, all of the models have some trouble in the right tail. At the ten-day horizon, all of the models miss the left tail and continue to have a small amount of difficulty in the right tail. At the 20-day horizon (see Fig. 22 in Appendix A), all of the models have about the same difficulties fitting the right tails as at the five- and ten-day horizons. The lack of fit in the left tail is still apparent, though substantially less than at ten days (MIX3 does slightly better than the other

<sup>8</sup>A nonautocorrelated subsample could be obtained by taking only every  $k$ th generalized residual. But this has its own problems. In particular,  $k$  different subsamples are possible, depending on whether the subsampling was begun on day  $1, 2, \dots, \text{or } k - 1$ . It turns out that the diagnostics can vary considerably between subsamples. A better idea might be to use the specification test of Bontemps and Meddahi (2005), which accounts for possible autocorrelation.

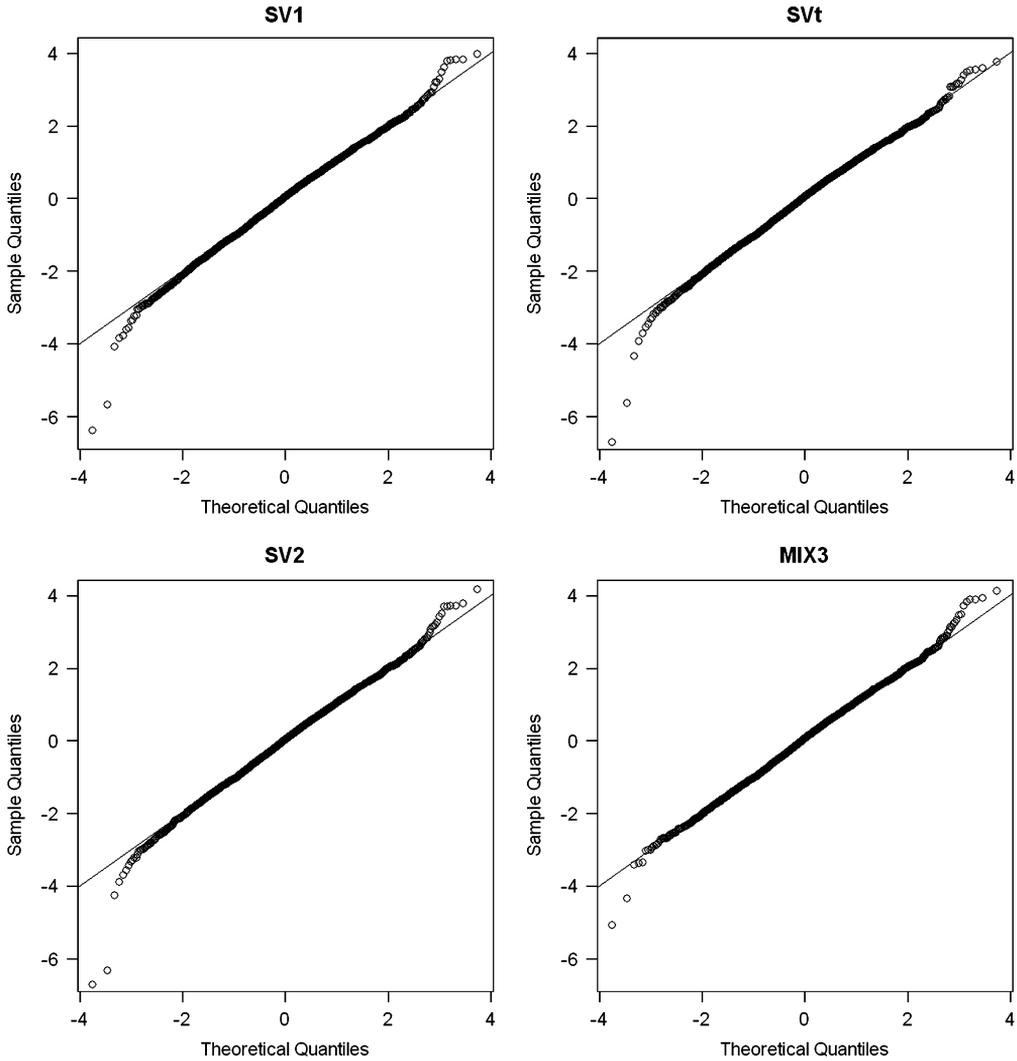


Fig. 10. *QQ*-plots for generalized residuals of five-day cumulative returns.

models; SV2 is worst). Because of the presence of autocorrelation, the Jarque-Bera test is no longer valid, so it is difficult to judge the extent to which these problems represent model mis-specification as opposed to sample variation.

Figs. 10 and 11 are mildly supportive of the idea of a nonlinear relationship between returns and changes in volatility. The evidence suggests that estimated volatility increases too little in response to large negative returns and that it decreases too much in response to large positive returns (that is, the model-implied leverage effect is too weak for large negative returns and too strong for large positive returns). In both cases, the result is that subsequent generalized residuals are slightly exaggerated. This is possibly related to results found by Ghysels, Santa-Clara, and Valkanov (2004). Using their mixed data sampling

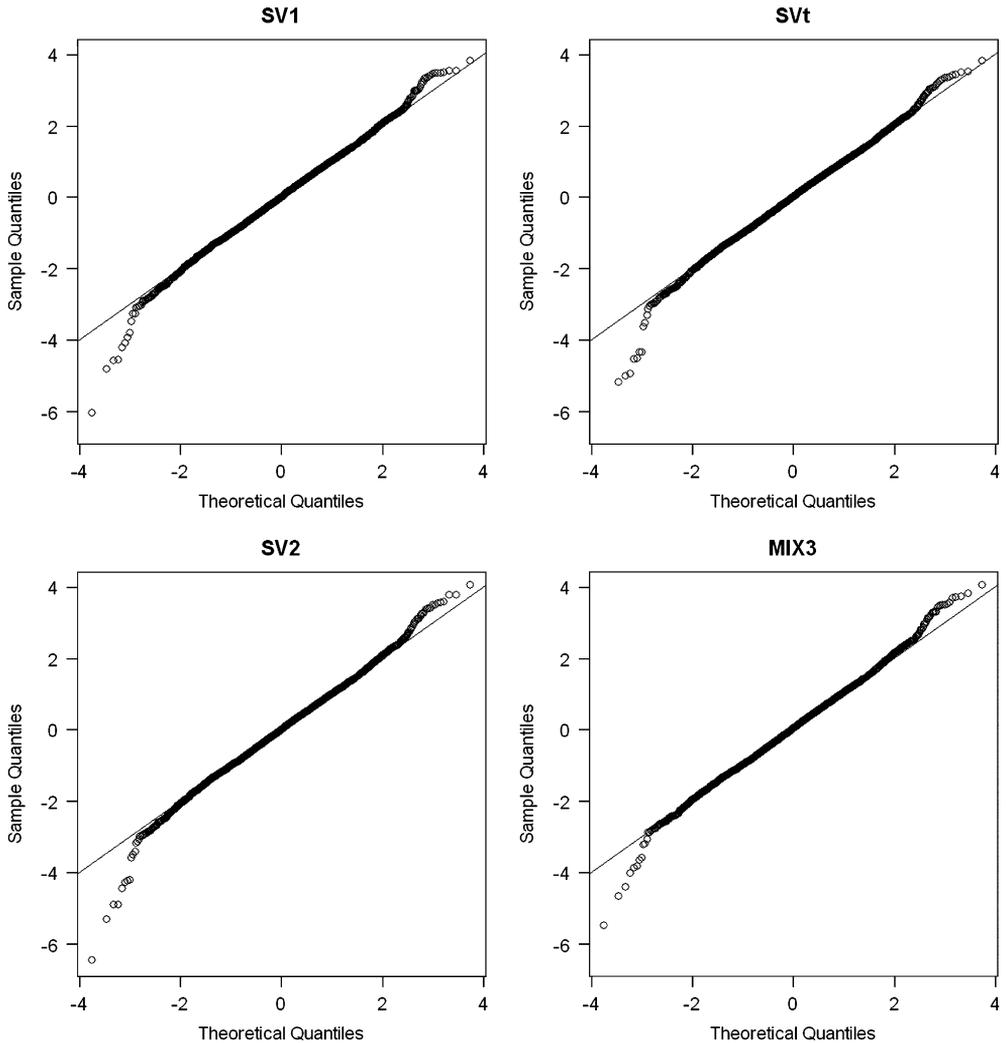


Fig. 11. *QQ*-plots for generalized residuals of ten-day cumulative returns.

(MIDAS) approach with different weights for the effects of positive and negative returns on volatility, they find that negative shocks have a larger immediate impact on volatility than do positive shocks, but the impact of negative shocks is short-lived while that of positive shocks is extremely persistent.

The approach described above, whereby forecast densities are computed by integrating across the filtered distribution of the latent volatility factor, is also important for correctly computing option prices implied by a risk-neutral model (see also Bates, 2005). For example, a risk-neutral forecast density generated by initializing the volatility factor at its expected value conditional on past returns (instead of integrating over the full distribution) has tails that are too thin. Option prices are biased downward and volatility smiles are less

pronounced relative to those based on the correct forecast density. Furthermore, all of the models provide less than perfect fits to the data at the horizons examined. Imputing risk premia based on differences between these models and the risk-neutral models implied by option prices is thus an exercise that should be undertaken with a great deal of caution (see also Bates, 2003).

## 6. Comparison with affine models

Affine and affine-jump models are commonly used in the literature. This section takes some of the tools applied elsewhere in this paper and applies them to several members of the affine and affine-jump class. For concreteness, the same models and data as Eraker, Johannes, and Polson (2003) (EJP hereafter) are used. The data are made up of S&P 500 index returns from January 2, 1980 to December 31, 1999. This sample period differs only slightly from the data set used throughout the rest of this paper. The general form of the model is

$$\begin{aligned} dY_t &= \alpha dt + \sqrt{V_t} dW_{1t} + \xi_Y dN_{1t}, \\ dV_t &= \kappa(\alpha - V_t) dt + \sigma_V \sqrt{V_t} dW_{2t} + \xi_V dN_{2t}, \end{aligned} \tag{16}$$

where  $W_1$  and  $W_2$  are Brownian motions with correlation  $\rho$ , and  $N_1$  and  $N_2$  are Poisson processes (possibly identical) with constant arrival intensities  $\lambda_Y$  and  $\lambda_V$  and jump sizes  $\xi_Y$  and  $\xi_V$ , respectively. To distinguish them, the SV models considered in the preceding sections of this paper are sometimes referred to as log volatility models.

EJP looked at the following special cases:

**AFF:** No jumps, i.e.,  $\lambda_Y = \lambda_V = 0$ .

**AFF-J:** Jumps in returns with size  $\xi_Y \sim N(\mu_Y, \sigma_Y)$ ; no jumps in volatility.

**AFF-CJ:** Contemporaneous jumps in both returns and volatility (i.e.,  $N_1 = N_2$ ) with correlated sizes,  $\xi_V \sim \exp(\mu_V)$  and  $\xi_Y | \xi_V \sim N(\mu_Y + \rho_J \xi_V, \sigma_Y)$ .

**AFF-IJ:** Jumps in returns and volatility driven by independent Poisson processes and with sizes  $\xi_V \sim \exp(\mu_V)$  and  $\xi_Y \sim N(\mu_Y, \sigma_Y)$ .

Following EJP (and the approach used elsewhere in this paper), the Euler scheme approximation to the models is used throughout. Although the estimation technique described in Section 2 works for the AFF and AFF-J models, it works poorly for the models with jumps in volatility. This is because the importance sampler relies on a normal approximation to the true distribution of the volatility factor conditional on the data. The presence of jumps in volatility throws the approximation off by enough that the Monte Carlo integration is no longer practically feasible (though it remains theoretically valid). However, an alternative estimation strategy based on the particle filter can be used. The particle filter is easily implemented for all of these models and allows the likelihood to be approximated for any candidate parameter vector. The only drawback is that the resulting criterion function is not smooth (because of the nature of the particle filter). However, optimization can still proceed using, e.g., simulated annealing. Standard errors can be obtained by fitting a quadratic surface to the log likelihood function in a neighborhood of the mode. While this approach is easy to implement and works with a great deal of generality, it is computationally intensive. Nonetheless, letting things run for a day or so,

it is possible to get accurate estimates. The optimization is made easier because the EJP estimates provide good start values. In fact, EJP's estimates differ only slightly from the final estimates obtained using the approach outlined above. This is not surprising because the EJP estimates are given by the mean of the posterior with an uninformative prior while the maximum likelihood estimates are given by the mode of the likelihood, which is equivalent to the mode of the posterior with a flat prior. The closeness of the estimates provides corroboration of EJP's work. The log likelihoods corresponding to the two sets of estimates (MLE versus EJP) differ by only around one point at most, and the diagnostics obtained using either set of estimates are similar. For ease of comparison, the EJP estimates are used throughout the rest of this section (see Table 4).

As found by EJP (and others), including jumps in returns provides a large improvement in model fit over the basic affine model. Further corroborating EJP's findings, including jumps in volatility provides additional improvement. Both the AFF-IJ and AFF-CJ are preferred over AFF-J by both the AIC and SC (the SC imposes a penalty of 4.26 points per parameter with this data set). However, in contrast to EJP, who find evidence in favor of independent jumps, I do not find much difference between the AFF-IJ and AFF-CJ models (both have the same number of parameters).

To compare the affine and affine-jump models with the models used throughout the rest of this paper, those models were reestimated over the EJP data. The parameter estimates differed only slightly from those displayed in Table 2 and so are not reported (but are available upon request). The log likelihoods are displayed in Table 4.<sup>9</sup> The log likelihood of the MIX3 model exceeds that of the best of the affine-jump models by 46 points (this model has one additional parameter versus either AFF-IJ or AFF-CJ) and is clearly preferred by both the SC and AIC. SVt and SV2 are also preferred over all of the affine-jump models by safe margins. The affine-jump models are all preferred over the basic SV1 model.

Correlograms for the squared generalized residuals are shown in Fig. 12. The affine models without jumps in volatility show significant autocorrelation through the first several lags. Box-Pierce tests on 20 lags reject AFF and AFF-J (see Table 5). Introducing jumps in volatility takes care of this problem, providing additional evidence in favor of the findings of EJP. None of the log volatility models was rejected by this test. All of the affine and affine-jump models are rejected on Box-Pierce tests with 250 lags (as were the log volatility models).

$QQ$ -plots for the generalized residuals are shown in Fig. 13.<sup>10</sup> The model without jumps is too thin in the left tail, as expected, but all of the models with jumps do much better, missing only the crash of October 1987. Nonetheless, all of the models are rejected by the Jarque-Bera test (Table 5).

Forecast densities for  $X_{t+1}|V_t$  are shown in Fig. 14. For the log volatility models,  $V_t$  affects only the scaling but not the shape of this density. However, for the affine-jump models, the diffusion term but not the jump size is scaled by  $V_t$ , so the shape of the density varies depending on the level of volatility. Forecast densities are

<sup>9</sup>To assure the results were comparable, log likelihoods were computed using both the Monte Carlo integration described in Section 4 and the particle filter (100,000 particles). The results were nearly identical.

<sup>10</sup>Similar  $qq$ -plots are shown in Eraker, Johannes, and Polson (2003), but they are different from those shown in this paper in an important way: The EJP  $qq$ -plots use smoothed residuals, whereas the plots in this paper use filtered residuals. That is, the EJP residual  $\tilde{z}_{t+1}$  is computed by integrating over the distribution of  $V_t$  conditional on the full data set, while this paper integrates over the distribution of  $V_t$  conditional on data observed up to time  $t$ .

Table 4

Parameter estimates for affine and affine-jump models, Standard & Poor's 500 index returns, January 2, 1980 to December 31, 1999

The data set is the same as used by Eraker, Johannes, and Polson (2003) (EJP hereafter). EJP use percentage returns, whereas the results reported elsewhere in this paper are based on decimal returns. Although the estimates reported below are reproduced from EJP, the log likelihoods are based on transforming the returns to decimal form to be comparable with results elsewhere in this paper.

For the log volatility models (Panel B), log likelihoods are obtained using the particle filter after reestimating the models over the EJP data set. Estimates are similar to those in Table 2 and are available upon request.

Panel A				
	AFF	AFF-J	AFF-CJ	AFF-IJ
$\mu$	0.04443 (0.0110)	0.0496 (0.0109)	0.0554 (0.0112)	0.0506 (0.0111)
$\alpha$	0.9052 (0.1077)	0.8136 (0.1244)	0.5376 (0.0539)	0.5585 (0.0811)
$\kappa$	0.0231 (0.0068)	0.0128 (0.0039)	0.0260 (0.0041)	0.0250 (0.0057)
$\sigma_V$	0.1434 (0.0128)	0.0954 (0.0104)	0.0790 (0.0074)	0.0896 (0.0115)
$\mu_Y$		-2.5862 (1.3034)	-1.7533 (1.5566)	-3.0851 (3.2485)
$\rho_J$			-0.6008 (0.9918)	
$\sigma_Y$		4.0720 (1.7210)	2.8864 (0.5679)	2.9890 (0.7486)
$\mu_V$			1.4832 (0.3404)	1.7980 (0.5737)
$\rho$	-0.3974 (0.0516)	-0.4668 (0.0579)	-0.4838 (0.0623)	-0.5040 (0.0661)
$\lambda_Y$		0.0060 (0.0021)	0.0066 (0.0020)	0.0046 (0.0020)
$\lambda_V$				0.0055 (0.0032)
Log $L$	16894.01	16957.41	16968.20	16968.95
Panel B				
Model	SV1	SVt	SV2	MIX3
Log $L$	16945.08	16996.01	16992.67	17014.53

shown for three different levels of volatility, corresponding to the tenth, 50th, and 90th percentiles of volatility.<sup>11</sup>

On low volatility days, the jumps are very large relative to the diffusive part of the distribution. The left tail is much fatter than MIX3. At higher levels of volatility, the jumps are much smaller relative to the diffusive part of the distribution. At high levels of

<sup>11</sup>Volatility quantiles were obtained by taking the volatility estimates from the particle filter and sorting.

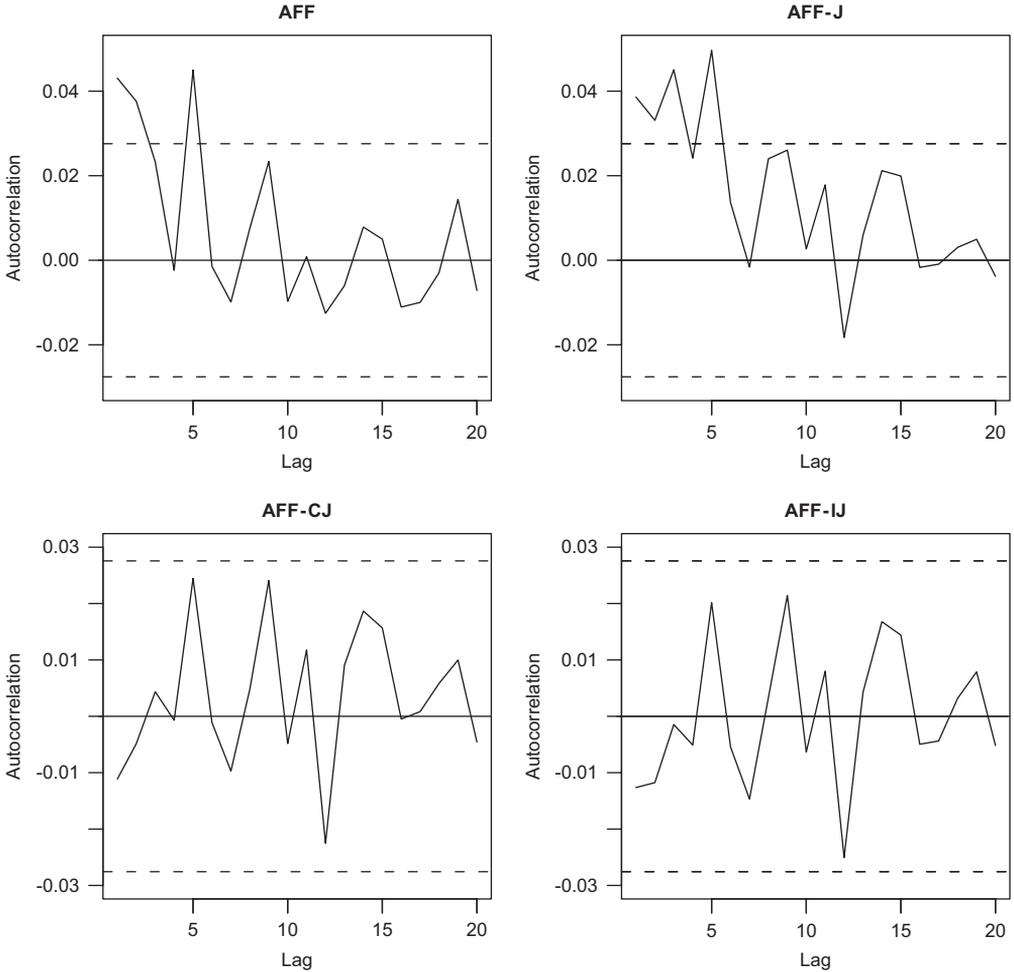


Fig. 12. Correlograms for squared generalized residuals, Eraker, Johannes, and Polson (2003) data. The dashed lines mark the rejection region (5% significance level) for the individual correlation coefficients.

volatility, the left tails of the AFF-CJ and AFF-IJ predictive densities are comparable to (but slightly thinner than) that of MIX3.

Because this feature of the affine-jump models is an important difference relative to MIX3, some additional experiments were performed to see if it is supported by the data. Figs. 15 and 16 show  $qq$ -plots of the generalized residuals corresponding to the observations in the highest and lowest decile with respect to volatility (the plots for AFF-IJ are almost identical to those for AFF-CJ and are thus omitted from the figure). Because the full set of generalized residuals should be iid  $N(0, 1)$  if the model is correctly specified, any subset should be as well. In both the high and low volatility cases, the  $qq$ -plots suggest that MIX3 matches the data well, whereas the affine and affine-jump models show evidence of mis-specification. Furthermore, this mis-specification corresponds to

Table 5

Tests for autoregressive conditional heteroskedasticity (ARCH) effects and normality of generalized residuals, Standard & Poor's 500 index returns, January 2, 1980 to December 31, 1999

The data set is the same as used by [Eraker, Johannes, and Polson \(2003\)](#). The ARCH test is based on 20 lags. Box-Pierce tests are based on 20 and 250 lags of the squared generalized residuals. *P*-values are shown in parentheses.

Model	Jarque-Bera	ARCH(20)	Box-Pierce(20)	Box-Pierce(250)
MIX3	0.22 (0.8978)	16.26 (0.7002)	16.98 (0.6543)	307.78 (0.0074)
AFF	1455.28 (0.0000)	36.73 (0.0126)	37.37 (0.0106)	300.59 (0.0156)
AFF-J	22.12 (0.0000)	46.50 (0.0007)	53.98 (0.0001)	350.23 (0.0000)
AFF-CJ	23.21 (0.0000)	14.16 (0.8222)	14.92 (0.7809)	338.99 (0.0002)
AFF-IJ	19.59 (0.0001)	13.32 (0.8634)	14.24 (0.8179)	335.14 (0.0003)

about what one might expect from looking at the density plots in [Fig. 14](#) under the belief that MIX3 was the true data generating model.<sup>12</sup> The Jarque-Bera test rejects the affine-jump models on the high volatility subset but not the low volatility subset (the Jarque-Bera test only looks at third and fourth moments so failure to reject on the low volatility subset could reflect lack of power in the relevant direction; the *qq*-plots seem fairly suggestive).

Models with time-varying jump intensity (but not size) have been studied by [Bates \(2005\)](#), [Andersen, Benzoni, and Lund \(2002\)](#), and [Pan \(2002\)](#), but without jumps in volatility. Bates fits the model to S&P 500 index returns over 1953–1996 using an approximate maximum likelihood technique. He finds significant evidence in favor of time-varying jump intensity (an improvement of about 15 points in log likelihood). Andersen et al., using the same data set but a simulated method of moments estimator, find the time-dependence parameter for jump intensity to differ negligibly from zero. Pan uses both returns and option prices and finds the state-dependence of jump intensity to be important. This is consistent with EJP's finding that the constant intensity model cannot explain differences in the shape of the option price smirk on low versus high volatility days.

To summarize, the log volatility models significantly outperform the affine models according to two commonly used information criteria. Among the affine models, models with jumps in volatility as well as returns are preferred. Additional evidence supporting jumps in volatility is provided by the diagnostic tests. But there appears to be a need for time-variation in either the size or intensity of jumps in returns, a feature possessed by none of the affine models under consideration in this paper.

<sup>12</sup>Among the low-volatility days, there are no observations corresponding to the extreme left tail in the region where it is thicker under the affine and affine-jump models than under MIX3. Given that such events would be expected to be rare, this does not necessarily represent evidence against this tail shape, but only a lack of evidence in support of it.

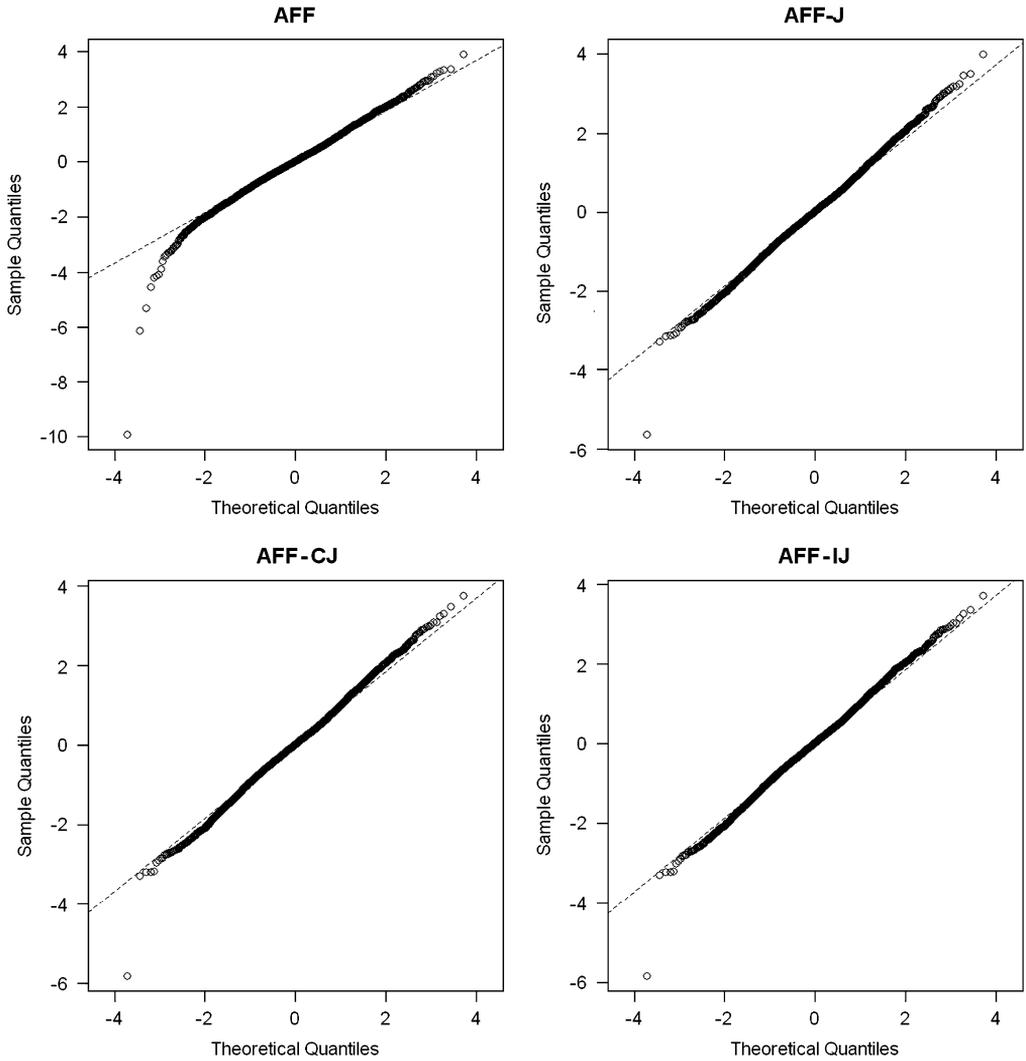


Fig. 13. *QQ*-plots of generalized residuals against the standard normal, Eraker, Johannes, and Polson (2003) data.

## 7. Conclusions

Understanding both the dynamics of volatility and the shape of the distribution of returns conditional on the volatility state is important for many financial applications. A simple single-factor SV model is largely sufficient to capture the dynamics. It is the shape of the conditional distribution that is the problem. Commonly used models lack sufficient flexibility to capture important features of this distribution. Although the SVt and SV2 models capture some of the kurtosis exhibited by returns, they are unable to catch the asymmetry in tail thickness and miss much of the mass in the extreme left tail. The mixture

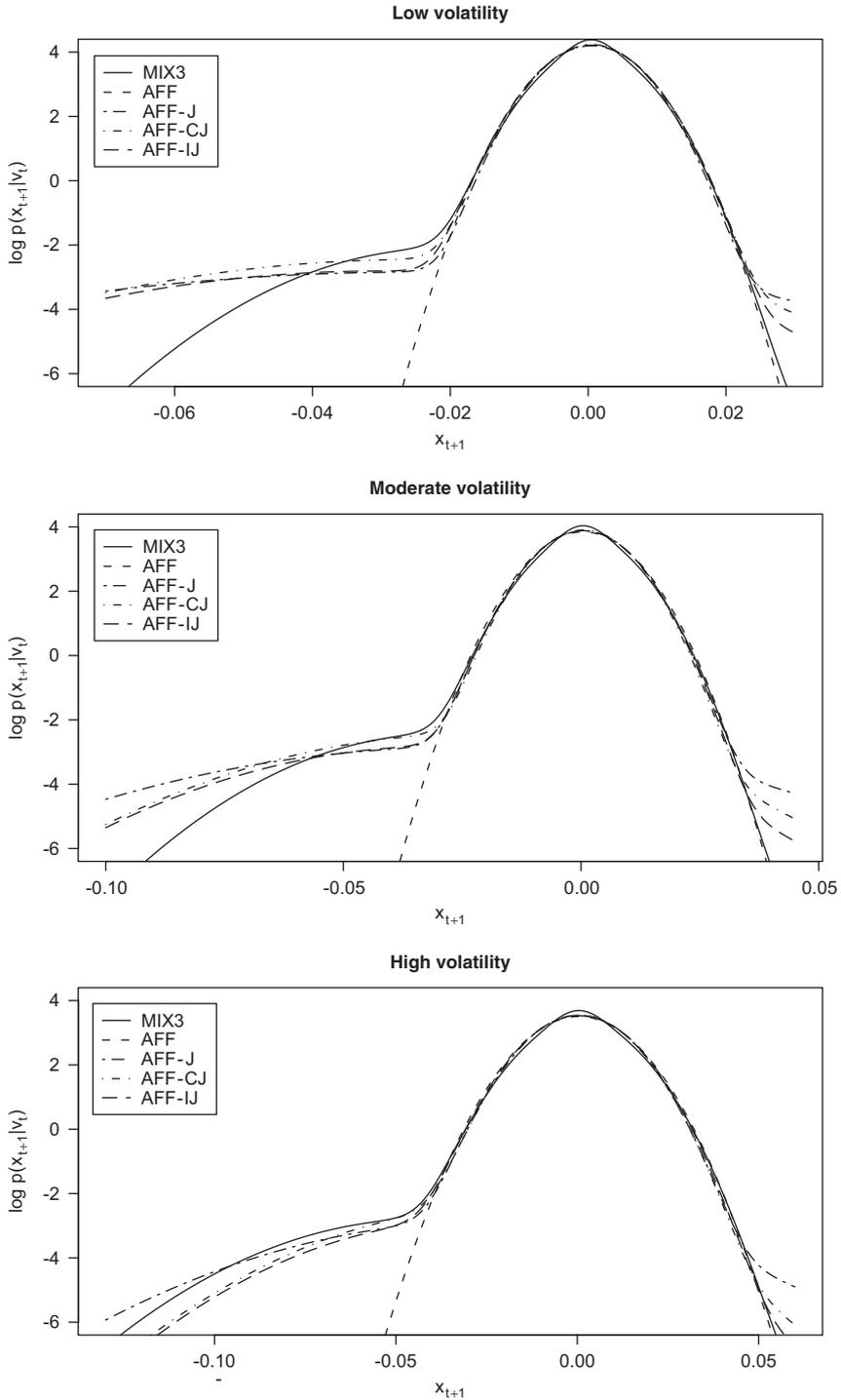


Fig. 14. Log density of  $X_{t+1}$  conditional on  $V_t$  for various models estimated over Eraker, Johannes, and Polson (2003) data. Volatility levels are equal to the tenth, 50th and 90th percentiles of volatility.

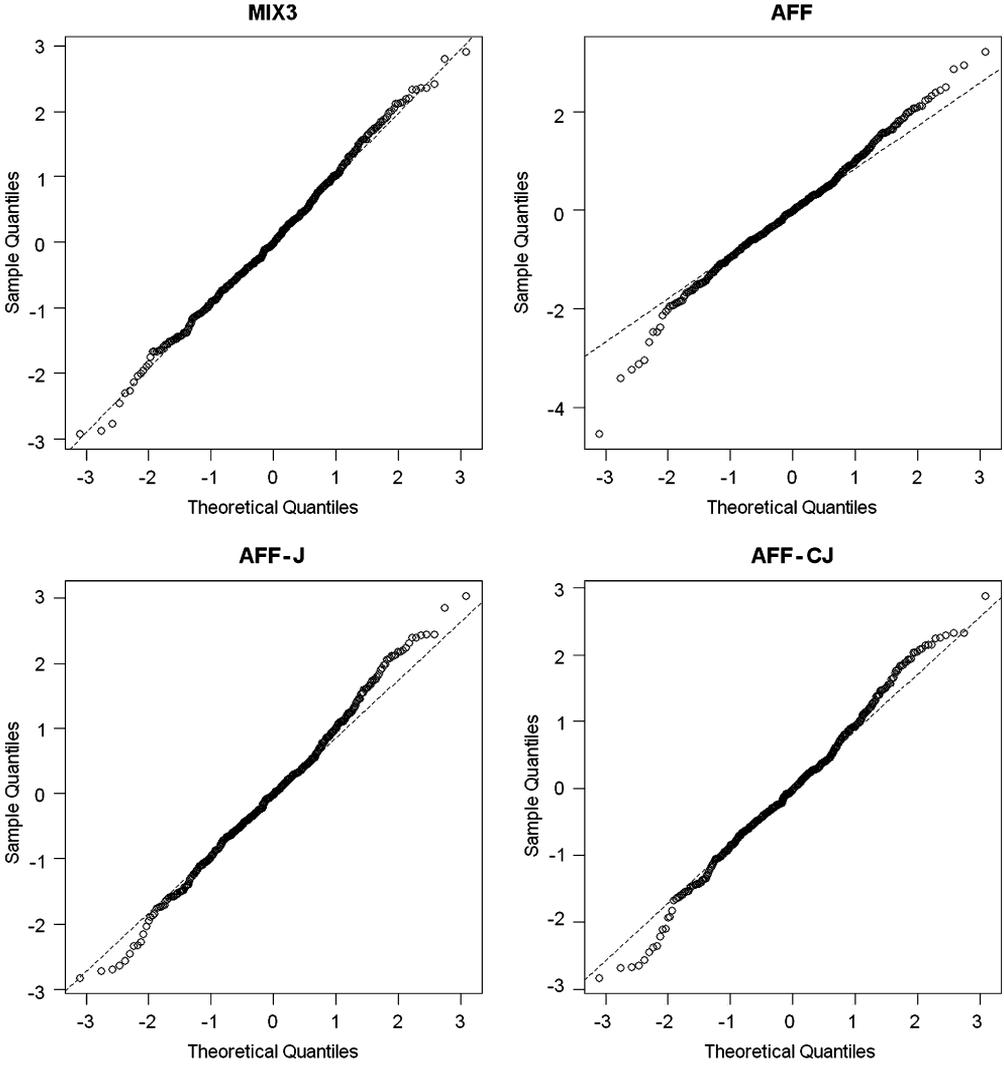


Fig. 15. *QQ*-plots for the observations with volatility in the lower decile.

distributions used in the SV-mix models provide something close to a nonparametric look at the conditional distribution of returns. Model diagnostics suggest that it is successful in reflecting key features of the data.

The SV-mix model also suffers to some extent from the drawbacks of nonparametric estimation: It is difficult to obtain precise estimates in regions of the state space when there are few observations. However, the model does impose some structure compared with a truly nonparametric estimator such as, say, a kernel estimator. This balance between structure and “letting the data speak for itself” is an appealing feature of the model.

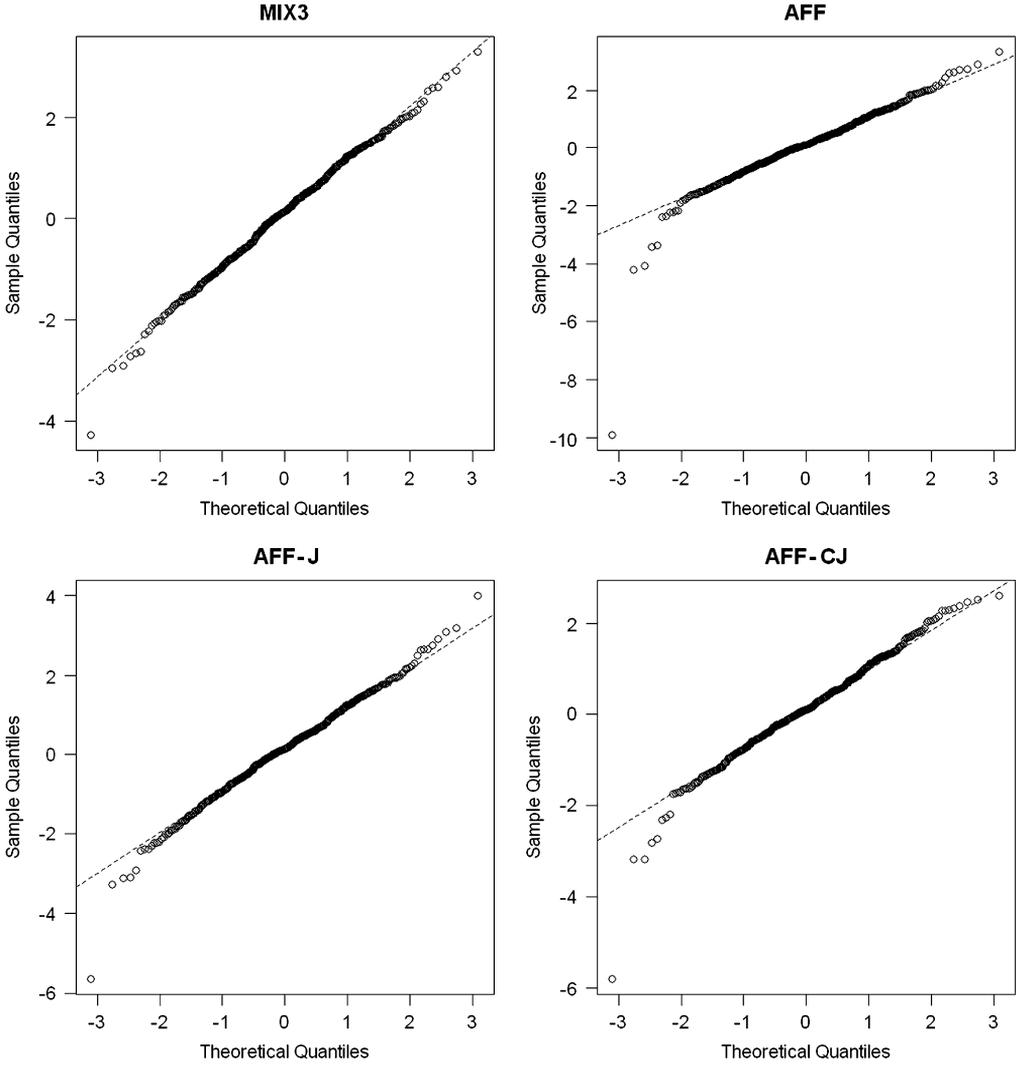


Fig. 16.  $QQ$ -plots for the observations with volatility in the top decile.

Jump-diffusion models represent another path toward the same goal. The jump process plays much the same role as a low probability, high variance component in the mixture distribution. It is easy to find a jump-diffusion model with the same Euler-scheme approximation as the SV-mix model. But, unless one is willing to accept multiple jump processes, with jumps occurring nearly every day and explaining most of the distribution of returns, such models have trouble achieving the flexibility of the SV-mix models. In any event, this paper takes no position on the issue as to whether the returns process is continuous or includes jumps (or on any other intra-daily feature of the returns process). The mixture distribution is regarded simply as a mechanism to generate a flexible family of distributions for daily returns.

In a direct comparison with several affine-jump models studied by Eraker, Johannes, and Polson (2003), an SV-mix model with three mixture components was strongly preferred by Akaike and Schwarz information criteria. Diagnostics support the evidence provided by Eraker, Johannes, and Polson in favor of jumps in volatility as well as returns, but there is also evidence in favor of time-varying jump intensity, a feature included in models examined by several other authors, but not those of Eraker, Johannes, and Polson.

There is a great deal of interest in understanding the relationship between the physical measure, which governs returns, and the risk-neutral measure, from which options are priced. But a crucial step in such explorations must be an adequate description of the physical measure, because errors here lead to faulty inference regarding risk premia. The modeling framework proposed by this paper looks to be useful in this direction.

## Appendix A. Supplementary material

Supplementary material associated with this article can be found in the online version at doi:[10.1016/j.jfineco.2006.06.005](https://doi.org/10.1016/j.jfineco.2006.06.005).

## References

- Andersen, T.G., Benzoni, L., Lund, J., 2002. An empirical investigation of continuous-time equity return models. *Journal of Finance* 57, 1239–1284.
- Bai, J., 2003. Testing parametric conditional distributions of dynamic models. *Review of Economics and Statistics* 85, 531–549.
- Bates, D.S., 1996. Jumps and stochastic volatility: exchange rate processes implicit in deutsche mark options. *Review of Financial Studies* 9, 69–107.
- Bates, D.S., 2000. Post-'87 crash fears in the S&P 500 futures option market. *Journal of Econometrics* 94, 181–238.
- Bates, D.S., 2003. Empirical option pricing: a retrospection. *Journal of Econometrics* 116, 387–404.
- Bates, D.S., 2005. Maximum likelihood estimation of latent affine processes. *Review of Financial Studies* 19, 909–965.
- 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.
- Carlin, B.P., Louis, T.A., 2000. *Bayes and Empirical Bayes Methods for Data Analysis*, second ed. Chapman & Hall, Boca Raton.
- Crisan, D., 2001. Particle filters—a theoretical perspective. In: Doucet, A., de Freitas, N., Gordon, N. (Eds.), *Sequential Monte Carlo Methods in Practice*. Springer, New York, pp. 17–41.
- Diebold, F., Gunther, T., Tay, A., 1998. Evaluating density forecasts with applications to financial risk management. *International Economic Review* 39, 863–883.
- Ding, Z., Granger, C.W., 1996. Modeling volatility persistence of speculative returns: a new approach. *Journal of Econometrics* 73, 185–215.
- Duan, J.-C., 2003. A specification test for time-series models by a normality transformation. Unpublished working paper, Rotman School of Management, University of Toronto, Toronto, Canada.
- 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., 2006. Monte Carlo methods for estimating, smoothing, and filtering one- and two-factor stochastic volatility models. *Journal of Econometrics* 133, 273–305.
- Eraker, B., 2004. Do stock prices and volatility jump? Reconciling evidence from spot and option prices. *Journal of Finance* 59, 1367–1403.

- Eraker, B., Johannes, M., Polson, N.G., 2003. The impact of jumps in returns and volatility. *Journal of Finance* 58, 1269–1300.
- Escobar, M.D., West, M., 1995. Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association* 90, 577–588.
- Gallant, A.R., Hsieh, D., Tauchen, G.E., 1997. Estimation of stochastic volatility models with diagnostics. *Journal of Econometrics* 81, 159–192.
- Geweke, J., Amisano, G., 2001. Compound Markov mixture models with applications in finance. Unpublished working paper, University of Iowa, Iowa City, Iowa.
- Ghysels, E., Santa-Clara, P., Valkanov, R., 2004. There is a risk-return trade-off after all. Unpublished working paper, University of North Carolina, Chapel Hill, NC.
- Hong, Y., Li, H., 2002. Nonparametric specification testing for continuous-time models with applications to spot interest rates. Unpublished working paper, Cornell University, Ithaca, NY.
- Jackwerth, J., Rubinstein, M., 1996. Recovering probability distributions from option prices. *Journal of Finance* 51, 1611–1631.
- Liesenfeld, R., Richard, J.-F., 2003. Univariate and multivariate stochastic volatility models: estimation and diagnostics. *Journal of Empirical Finance* 10, 505–531.
- McLachlan, G., Peel, D., 2000. *Finite Mixture Models*. Wiley, New York.
- Pan, J., 2002. The jump-risk premia implicit in options: evidence from an integrated time-series study. *Journal of Financial Economics* 63, 3–50.
- Priebe, C.E., 1994. Adaptive mixtures. *Journal of the American Statistical Association* 89, 796–806.
- Robert, C.P., Casella, G., 2004. *Monte Carlo Statistical Methods*, second ed. Springer, New York.
- Sandmann, G., Koopman, S.J., 1998. Estimation of stochastic volatility models via Monte Carlo maximum likelihood. *Journal of Econometrics* 87, 271–301.
- Schwert, G.W., 1990. Stock market volatility. *Financial Analysts Journal* 46, 23–34.
- Shephard, N., Pitt, M., 1997. Likelihood analysis of non-Gaussian measurement time series. *Biometrika* 84, 653–667.

## Further Reading

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