Efficient Estimation of Stochastic Diffusion Models with Leverage Effects and Jumps

Juan Carlos Artigas and Ruey S. Tsay
Graduate School of Business
University of Chicago
Preliminary Draft

Abstract

Several Markov Chain Monte Carlo (MCMC) methods have been proposed in the literature to estimate stochastic diffusion models with stochastic volatility. Most of the methods either cannot deal with the leverage effects or require intensive computation. In this paper, we investigate the difficulty of estimating the leverage effect and propose an efficient method to overcome the difficulty. The proposed method is a combination of those of Kim, Shephard, and Chib (1998) and Jacquier, Polson and Rossi (2002), but makes use of the technique of forward filtering and backward sampling to improve computing efficiency and nonlinear filters to estimate leverage effects. Real and simulated data are used to demonstrate the proposed method and to compare with the existing methods.

Key Words: Diffusion process, Gibbs sampler, Kalman filter, Leverage effect, and Markov chain Monte Carlo.

1 Introduction

Volatility has been a key topic in financial research, especially in the areas of asset allocation, derivative pricing, and risk management. For instance, the diffusion term of a stochastic diffusion equation for option pricing is a function of volatility. For many years, volatility was assumed to be constant, a hypothesis that was dropped lately in favor of a more flexible structure. This leads to the development of stochastic volatility process that is believed to be closer to reality.

Many methods have been proposed to estimate stochastic diffusion equations with stochastic volatility processes, such as maximum likelihood, quasi-maximum likelihood, efficient method of moments and Monte Carlo simulations. All these methods have pros and cons, but one method that is commonly used is the Markov Chain Monte Carlo (MCMC) procedure. Examples include Eraker (2001), Jacquier, Polson and Rossi (2002), Kim, Shephard and Chib (2001), among others.

The story does not end with the pure stochastic volatility models. In recent years, researchers have included correlation and jumps in the mean and volatility equations to cope with the leverage effect and the unusual observations found in market data. This is the case in Eraker, Johannes and Polson (2003), and Chib, Nardari and Shephard (2002).

The algorithms to estimate these volatility models via MCMC methods also vary widely in the literature. Some do it recursively, others jointly, some use strong priors and some not so strong. The fact of the matter is that all these algorithms have some good and some not so good properties.
The objective of this study is to examine the challenges that some of these algorithms pose and outline ways to overcome them.

We also derive in detail an algorithm that is a blend of different estimation methods and that we use to estimate the structure in the 3-month Treasury bill (T-bill) rate, after the accuracy of the procedure is verified through a simulation study.

The paper is organized as follows. In Section 2 we briefly review the stochastic volatility model. We then discuss in Section 3 the theoretical implications of different models and the empirical advantages of using one algorithm versus another. Section 4 proposes a new compromised algorithm to estimate the stochastic volatility models. Section 5 shows the empirical results using both simulated and real data. Section 6 concludes.

2 Stochastic Volatility Models

In this section we introduce the stochastic volatility (SV) models, with or without jumps, in a rather general framework. Then we focus on two particular specifications of the models, namely the ones described in Eraker et al. (2003) and Chib et al. (2002). Finally, we define the model we use as a baseline for our investigation.

We start with the description of the process of interest. Let \( X_t := (Y_t, Z_t)' \) be a 2-dimensional diffusion process such that

\[
dX_t = \mu(X_t; \theta)dt + \sigma(X_t; \theta)dW_t + \xi_t dN_t
\]

where \( \theta \) is a set of parameters in the space \( \Theta \subseteq \mathbb{R} \); \( \mu(\cdot) \) and \( \sigma(\cdot) \) are functions from \( \mathbb{R} \times \Theta \rightarrow \mathbb{R}^2 \) and \( \mathbb{R} \times \Theta \rightarrow \mathbb{R}^{2 \times 2} \); \( W_t \) is a 2-dimensional Brownian Motion; \( \xi_t \) is a bivariate jump size process; and \( N_t = (N^y_t, N^z_t) \) with \( N^y_t \) and \( N^z_t \) being Poisson processes with intensities \( \lambda_y \) and \( \lambda_z \), respectively.

This is a very general setup. Different specifications of the mean and variance functions lead to some of the well-known SV models such as the Cox-Ingersoll-Ross model, the Vasicek model, and some other more recent formulations.

The case where \( \lambda_y = \lambda_z = 0 \) is the simple stochastic volatility model. When any one of the intensities is greater than zero, we have a stochastic volatility model with jumps. Often one assumes \( \lambda_z > 0 \) only if \( \lambda_y > 0 \).

To be able to estimate this model using Markov chain Monte Carlo methods, one has to use a discretized version of (1), namely, the Euler approximation. This approximation can be represented (in a loose way) as follows:

\[
\Delta X_t = \mu(X_t, \theta) \Delta t + \sigma(X_t, \theta) \sqrt{\Delta t} \Delta W_t + \xi_t J_t
\]

where \( \Delta W_t \sim N_2(0, \Sigma) \), with unit variance and correlation \( \rho \); and \( J_t = (J^y_t, J^z_t) \) with each component being an independent Bernoulli random variable with probabilities of success \( \lambda_y \) and \( \lambda_z \), respectively.

We study three particular representations of (2) found in the literature. The first one is by Eraker et al. (2003), the second one by Chib et al. (2002), and the last one by Eraker (2001).

Henceforth assume \( \Delta t = 1 \) to simplify the notation. The stochastic diffusion model of Eraker et al. (2003) is

\[
\begin{align*}
\Delta Y_t & = \mu + \sqrt{Z_{t-1}} \varepsilon_t + \xi^y_t J^y_t, \\
\Delta Z_t & = \kappa(\theta - Z_{t-1}) + \sigma_z \sqrt{Z_{t-1}} \eta_t + \xi^z_t J^z_t
\end{align*}
\]

(3.1) (3.2)
where $Y_t$ can be the log price or the yield of an asset; and $(\varepsilon_t, \eta_t)' \sim N_2(0, \Sigma)$. Clearly, different specifications of the jump times and sizes lead to alternative models. Four categories of model are available. The first model has the stochastic volatility model without jumps, the second has jumps in the mean equation (SVJ model) only, the third allows for independent jumps in the mean and the volatility (SVIJ); and the fourth model contains jumps in the mean conditional on jumps in the volatility (SVCJ). Eraker et al. (2003) describe and estimate each of these models.

The next model, which resembles the one described in Chib et al. (2002), is

$$\Delta Y_t = \theta_r + \kappa_r Y_{t-1} + e_t^2 Z_{t-1} \varepsilon_t + \xi_t^y J_t^y, \quad (4.1)$$

$$\Delta Z_{t+1} = \theta_z + \kappa_z Z_t + \sigma_z \eta_t + \xi_t^z J_t^z \quad (4.2)$$

where $\varepsilon_t$ and $\eta_t$ are independent standard Normals. In their paper, Chib et al. assume $\lambda_z = 0$. Hence, they derive algorithms only for the SV and SVJ models.

Finally, the model that we use for this study is a slight modification of the one described in Eq. (3.1)-(3.2). It is also an extension of the model used by Eraker (2001). It can be written as

$$\Delta Y_t = \theta_r + \kappa_r Y_{t-1} + e_t^2 Z_{t-1} \varepsilon_t + \xi_t^y J_t^y, \quad (5.1)$$

$$\Delta Z_t = \theta_z + \kappa_z Z_{t-1} + \sigma_z \eta_t + \xi_t^z J_t^z \quad (5.2)$$

where $(\varepsilon_t, \eta_t)'$ are bivariate Normal with mean zero, unit variance, and correlation $\rho$.

The difference between models in (4.1)-(4.2) and (5.1)-(5.2) lies in the way the errors relate to each other. In the case of independent errors, the index specification has no impact on the derivation of the algorithm. Nevertheless, when the errors are correlated the two models refer to two different, but related, processes.

### 3 Implications of Models and Algorithms

In this section we explain the difficulties and advantages found in the estimation of different stochastic volatility models. We conclude that the model specification (5.1)-(5.2) has certain properties that make it preferable.

Let us start with the model in Eq. (3.1)-(3.2). The procedure that Eraker et al. (2003) propose to estimate the model uses Gibbs sampler for most parameters, and Metropolis-Hastings for the correlation parameter $\rho$ and for the volatility path $Z_1, \ldots, Z_n$. The latter is sampled one at a time. To sample from the posterior distribution of $Z_t$ they use the information at $t-1$ and $t+1$ as described in Eraker (2001). The first clear disadvantage of such a specification is the use of a squared-volatility process. Assuming that the errors are Normal, the squared process is counterintuitive and forces the estimation to use some sort of strong prior to constrain the volatility process to be non-negative. In fact, a condition for the process to have at most a 97.5% chance of being non-negative is that $\kappa \theta > 2 \sigma_z^2$, which seems to be an strong restriction on the parameters. Hence, even when most of the posterior distributions of the parameters in the model have nice and clean results, a misspecification on the priors can lead to a highly inefficient algorithm.

Our limited empirical experience shows that sampling the volatility process can be very sensitive towards modifications on the proposal distribution for the M-H algorithm (the posterior distribution appears to be highly non-Gaussian and thus an efficient proposal is not easily found), and very dependent on the particular processes analyzed. A similar fact was experienced for the correlation

As noted before, $\Sigma := \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$.
Nevertheless, Jacquier et al. (2002) propose a way to sample from the joint distribution of $\rho$ and $\sigma_z$ that we find to be more stable and appealing. In particular, the algorithm is not so dependent on the proposal distribution for $\rho$. In summary, the ideas of this algorithm to use the advantage of the Gibbs sampler to obtain draws from the posterior distributions of the parameters are good and straightforward to follow. Nevertheless, the algorithm seems very slow and case dependent when one tries to sample from the volatility path.

The log-volatility model described in Eq. (4.1)-(4.2) seems more attractive from the theoretical point of view. It is defined in such a way that imposes the non-negativity constraint of the volatility. Nevertheless, it creates a more complex posterior distribution for some parameters. However, all distributions involved are tractable or can be easily approximated.

To overcome the difficulties of sampling from the non-Gaussian posterior distribution of the volatility, Kim et al. (1998) propose a way to draw directly from the more stable multivariate distribution of the whole volatility path $\{Z_1, \ldots, Z_n\}$ at once. To achieve this, they use ideas from linear state-space models. The main result comes from the linearization of (4.1) to transform the non-linear system into a linear one. Then, a mixture of Normals is used to approximate the distribution of the transformed error. Following this line of work, Chib et al. (2002) extend the algorithm to a more general framework. The other parameters in the model are sampled in blocks either using a Gibbs sampler or Metropolis-Hastings. The performance of this methodology is better, in particular, in our experience, the algorithm is even ten times faster than the previous one, and also more stable. It seems not to be case dependent and the selection of priors has little impact on the final results. A drawback emerges with this simplicity: when we linearize (4.1) we break the correlation structure of the errors and, thus, the estimation of $\rho$ is no longer straightforward.

For this study, we adopt the model specification described in Eq. (5.1)-(5.2) for simplicity and because we believe that the volatility has an infinitesimally delayed impact on the return (or yield.) It also has the good properties of a log-volatility model, as well as some simple economical interpretations. In particular, as discussed in Yu (2004), this specification ensures the correct effect of the leverage effect. Namely, if the correlation among the errors is negative, a fall in the return (or yield) translates into an increase in the volatility. This corresponds to the so-called leverage effect.

We aim to formulate an algorithm that uses the positive features of the previous two algorithms. The next section describes in length the proposed methodology.

4 Algorithm to estimate Stochastic Volatility Models

In what follows, we describe the proposed algorithm to estimate the stochastic diffusion model. We state the algorithm in a general way at the beginning, and then more extensively for each step. In particular, we present the algorithm to estimate the SVJ model, from which the SV model is a special case.

Thus, the model that we use in this section is a special case of (5.1)-(5.2), namely

$$\Delta Y_t = \theta_t + \kappa_t Y_{t-1} + \epsilon_t^2 Z_{t-1} \gamma_{t-1} \epsilon_t + \xi_{t} J_t, \quad (6.1)$$
$$\Delta Z_t = \theta_z + \kappa_z Z_{t-1} + \sigma_z \eta_t \quad (6.2)$$

where $(\epsilon_t, \eta_t)'$ are bivariate Normal with mean zero, unit variance and correlation $\rho$; $J_t \sim \text{Ber}(\lambda)$; and $\xi_t \sim N(\mu_J, \sigma_J^2)$. Let $\theta$ represent the set of parameters in the simple stochastic volatility model, and $\theta_J$ the parameters of the jump process (both times and sizes.)
In section 4.6 we extend this model to allow for contemporaneous jumps in the mean and the volatility.

**Proposed Algorithm**

0. Initialize any needed variables.
1. Sample each element of \( \theta \) from its posterior distribution.
2. Sample from the multivariate posterior of the log-volatility path \( Z_1, \ldots, Z_n \).
   - If no jumps are present, go to 1, otherwise
3. Sample the jump times, and the intensity parameter.
4. Sample from the posterior of the remaining elements of \( \theta_j \).
5. Sample from the posterior distribution of the jump sizes.
6. Go to 1.

Before we start with the specifics of each step, let us define some notation.

Let \( Y^t := \{Y_1, \ldots, Y_t\} \), i.e. the collection of observed data up to time \( t \); and \( Y_s^t := \{Y_s, Y_{s+1}, \ldots, Y_t\} \), in words, the observed data between time \( s \) and time \( t \). Similarly, define \( Z^t, Z_s^t, J^t \) and \( \xi^t \). Now, let \( \theta_i \) represent the \( i \)th subset of parameters \( \theta \); and \( \theta^- \) be the subset of remaining parameters, in other words, \( \theta^- := \theta^i \). Finally, let \( V^t := \{Y^t, Z^t, J^t, \xi^t\} \), or the collection of variable information up to time \( t \).

Let \( L(\cdot) \) be the likelihood function of some subset of \( \theta \) and notice that, given the full collection of data, we can write down the likelihood in three equivalent ways. They are

\[
L(\theta_i) = \prod_{t=1}^{n} p(X_t|X_{t-1}, J_t, \xi^y_t, \theta, \theta_j) \quad (7.1)
\]

\[
= \prod_{t=1}^{n} p(Z_t|X_{t-1}, Y_t, J_t, \xi^y_t, \theta, \theta_j)p(Y_t|X_{t-1}, J_t, \xi^y_t, \theta, \theta_j) \quad (7.2)
\]

\[
= \prod_{t=1}^{n} p(Y_t|X_{t-1}, Z_t, J_t, \xi^y_t, \theta, \theta_j)p(Z_t|Z_{t-1}, \theta) \quad (7.3)
\]

Firstly, we have that \( X_t|X_{t-1}, J_t, \xi^y_t, \theta, \theta_j \sim N_2(\mu, D \Sigma D') \); where

\[
\mu = \begin{bmatrix}
\mu_y \\
\mu_z
\end{bmatrix} = \begin{bmatrix}
\theta_r + (1 + \kappa_r)Y_{t-1} + \xi^y_t J_t \\
\theta_z + (1 + \kappa_z)Z_{t-1}
\end{bmatrix},
\]

\[
D = \begin{bmatrix}
\sigma_y & 0 \\
0 & \sigma_z
\end{bmatrix},
\]

\[
\Sigma = \begin{bmatrix}
1 & \rho \\
\rho & 1
\end{bmatrix}.
\]

Secondly, \( Z_t|X_{t-1}, Y_t, \theta, \theta_j \sim N(\mu_{z|y}, \sigma_{z|y}^2) \), with

\[
\mu_{z|y} = \theta_z + (1 + \kappa_z)Z_{t-1} + \rho_{z|y}^2(Y_t - \mu_y),
\]

\[
\sigma_{z|y}^2 = \sigma_z^2(1 - \rho^2);
\]

Finally, \( Y_t|X_{t-1}, Z_t, \theta, \theta_j \sim N(\mu_{y|z}, \sigma_{y|z}^2) \), with

\[
\mu_{y|z} = \theta_r + (1 + \kappa_r)Y_{t-1} + \xi^y_t J_t + \rho_{z|y}^2(Z_t - \mu_z),
\]

\[
\sigma_{y|z}^2 = e^{2\kappa_{z|y}}Y_{t-1}^2(1 - \rho^2);
\]
4.1 Gibbs sampler and M-H for the non-jump parameters

We present here the conditional posterior distributions of the parameters $\theta$. A detailed derivation of these and all other distributions can be found in Appendix A.

4.1.1 Posterior of $(\theta_r, \kappa_r)$

The posterior distribution of these two parameters is straightforward, and represents the least squares solution with heteroscedastic errors. For simplicity in notation, assume the priors of both parameters to be non-informative. Thus, we have that

$$p(\theta_r, \kappa_r | Y^n, Z^n, \xi^n, J^n, \theta_J) \sim N_2(\hat{\mu}_r, S_r)$$

with $\hat{\mu}_r = (X'X)^{-1}X'y$, and $S_r = (1 - \rho^2)(X'X)^{-1}$, where the $k$th element of the $(n \times 1)$ vector $y$ is $[(\Delta Y_k - \xi_k^0 J_k - \rho \frac{\sigma_y}{\sigma} (Z_k - \mu_z))e^{-\frac{1}{2}Z_k^{-1}Y_k^{-\beta}}]$, and the $k$th row of the $(n \times 2)$ matrix $X$ is $[e^{-\frac{1}{2}Z_k^{-1}Y_k^{-\beta}}, e^{-\frac{1}{2}Z_k^{-1}Y_k^{1-\beta}}]$.

4.1.2 Posterior of $(\theta_z, \kappa_z)$

Similarly, we have that

$$p(\theta_z, \kappa_z | Y^n, Z^n, \xi^n, J^n, \theta_J) \sim N_2(\hat{\mu}_z, S_z)$$

with $\hat{\mu}_z = (X'X)^{-1}X'z$, and $S_z = \sigma_z^2(1 - \rho^2)(X'X)^{-1}$, where the $k$th element of the $(n \times 1)$ vector $z$ is $[(\Delta Z_k + \rho \frac{\sigma_y}{\sigma} (Y_k - \mu_y))$, and the $k$th row of the $(n \times 2)$ matrix $X_z$ is $[1, Z_{k-1}]$.

4.1.3 Posterior of $(\sigma_z, \rho)$

To sample from this joint distribution, we use the result of Jacquier et al. (2002) and rewrite the likelihood as follows.

Let

$$\Omega = \begin{bmatrix} 1 & \rho \sigma_z \\ \rho \sigma_z & \sigma_z^2 \end{bmatrix} \text{ and } B = \begin{bmatrix} \sigma_y & 0 \\ 0 & 1 \end{bmatrix},$$

and notice that $X_t | X_{t-1}, J_t, \xi_t^y, \theta, \theta_J \sim N_2(\mu, B\Omega B')$. Re-parametrize $\Omega$ as

$$\Omega = \begin{bmatrix} 1 & \psi \\ \psi & \phi + \psi^2 \end{bmatrix}.$$

Let $u$ and $v$ be $(n \times 1)$ error vectors from (6.1) and (6.2), respectively. In other words, the $k$th element of $u$ is $[(Y_k - \mu_y)/\sigma_y]$, and the $k$th element of the vector $v$ is $[Z_k - \mu_z]$. Note that $u$ corresponds to a standardized error, but $v$ is not standardized. Therefore, using non-informative priors, we have

$$p(\psi | X^n, J^n, \xi^n, \theta_J, \theta_J) \sim N(\tilde{\psi}, \phi/(u'u)), $$

where $\tilde{\psi} = u'v/u'u$. And

$$p(\phi | X^n, J^n, \xi^n, \theta_J, \theta_J) \sim IG((n - 1)/2, v'v - (u'v)^2/(u'u)).$$

Finally, transform back to the original parametrization using $\sigma_z^2 = \phi + \psi^2$ and $\rho = \psi/\sigma_z$. 

6
4.2 Posterior of $\beta$

To draw from the non-standard distribution of $\beta$, we follow the Metropolis-Hastings algorithm proposed by Eraker (2001). Basically one draws from the following proposal distribution

$$N \left( \beta - \frac{\ell'(\beta)}{\ell''(\beta)}, -\left(\ell''(\beta)\right)^{-1} \right)$$

evaluated at $\beta = \beta^{(h-1)}$ (the previous draw in the chain), where $l(\beta)$ is the log-likelihood function of $\beta$.

4.3 Sampling the Volatility Path

To sample the posterior of the volatility, we base our methodology on the one described by Chib, Nardari and Shephard (2002), but use the Forward-Filtering Backward Sampling (FFBS) method instead of the backward recursions they use.

We will distinguish between two cases, one when there is correlation between the errors, and the other one where there is none.

4.3.1 No correlation in the errors

As mentioned earlier, the idea is to sample directly from the multivariate density of the volatility path $\{Z_1, \ldots, Z_n\}$, using the Markov property of the model. To simplify notation, we only condition on the sequence $\{Y_1, \ldots, Y_n\}$ implying that we are also using the information of jumps and parameters. Thus,

$$p(Z_1, \ldots, Z_n|Y^n) = p(Z_n|Y^n)p(Z_{n-1}|Z_n, Y^n) \cdots p(Z_1|Z_2, Y^n).$$

When one has a linear system, the solution to this problem is to use the Kalman filter, and then use some sort of backward sampler. Obviously, system (6.1)-(6.2) is not linear in the observation equation. To overcome this difficulty, Chib et al. (2002) propose to transform (6.1) to make it linear and then simply use known techniques to obtain the joint distribution.

Let $Y_t^* := \log(Y_t - \mu_y)^2 - \beta \log(Y_{t-1})^2$. Hence, (6.1) can be re-expressed as

$$Y_t^* = Z_{t-1} + \varepsilon_t^*$$

where $\varepsilon_t^* = \log(\varepsilon^2)$. They approximate the distribution of $\varepsilon^*$ with a mixture of Normals with seven components. This distribution is given by

$$p(\varepsilon_t^*) = \sum_{i=1}^{7} q_i \phi \left( \frac{\varepsilon_t^* - m_i}{v_i} \right),$$

where $\phi(\cdot)$ represents the standard Normal density. The values for $q_i$, $m_i$ and $v_i$ are presented in Table 1, which was obtained directly from Chib et al. (2002). This representation can be also written in terms of an auxiliary indicator variable $s_t$, such that $\varepsilon_t^*|s_t = i \sim N(m_i, v_i^2)$ and $P(s_t = i) = q_i$.

Therefore, to sample from the posterior of the volatility path, one first samples from the posterior of $s_t$ which is given by

$$p(s_t|Y_t^*, Z_{t-1}) \propto q_i \phi \left( \frac{Y_t^* - Z_{t-1} - m_i}{v_i} \right).$$
Table 1: Parameter components for the approximation to the distribution of $\varepsilon_t^*$

<table>
<thead>
<tr>
<th>$i$</th>
<th>$q_i$</th>
<th>$m_i$</th>
<th>$v_i^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.00730</td>
<td>-11.4004</td>
<td>5.7960</td>
</tr>
<tr>
<td>2</td>
<td>0.10556</td>
<td>-5.2432</td>
<td>2.6137</td>
</tr>
<tr>
<td>3</td>
<td>0.00002</td>
<td>-9.8373</td>
<td>5.1795</td>
</tr>
<tr>
<td>4</td>
<td>0.04395</td>
<td>1.5075</td>
<td>0.1674</td>
</tr>
<tr>
<td>5</td>
<td>0.34001</td>
<td>-0.6510</td>
<td>0.6401</td>
</tr>
<tr>
<td>6</td>
<td>0.24566</td>
<td>0.5248</td>
<td>0.3402</td>
</tr>
<tr>
<td>7</td>
<td>0.25750</td>
<td>-2.3586</td>
<td>1.2626</td>
</tr>
</tbody>
</table>

Then, from the linear system

$$Y_t^* = m_i + Z_{t-1} + w_t$$  \hfill (8.1)
$$Z_t = \theta_z + (1 + \kappa_z) Z_{t-1} + \sigma_z \eta_t$$  \hfill (8.2)

where $w_t \sim N(0, v_i^2)$ and $\text{Cov}(w_t, \eta_t) = 0$, one uses the forward and backward recursions proposed by Artigas (2003). The actual Kalman Filter and Backward Sampling recursions are shown in Appendix B.

4.3.2 Non-zero correlation in the errors

For the case $\rho \neq 0$, things are not as straightforward. To see this, recall that $\rho$ is the correlation between $\varepsilon_t$ and $\eta_t$. When we linearize (6.1) we transform the error $\varepsilon_t$, and the correlation between the errors $\varepsilon_t^*$ and $\eta_t$ is no longer $\rho$.

Let $\rho^* := \text{Corr}(\varepsilon^*, \eta)$. Through an empirical study, one can check that $\rho^* \approx 0$, regardless of the value of $\rho$. This happens when we square $\varepsilon_t$. Hence, the linear filter alone is unable to recover that dependence structure.

A way to include this information is to use conditional expectations. Thus, instead of directly calculating the correlation between $\varepsilon^*$, and $\eta$, we do it conditional on the sign that $\varepsilon$ takes.

Not surprisingly, once we use conditional expectation we can see a clear relationship between $\varepsilon^*$ and $\eta$, as a function of $\rho$. To understand this, a simulation study was conducted using 1,000 samples of 100,000 observations for values of $\rho$ from $-1$ to $1$ in increments of 0.01. Figure 1 shows the unconditional correlation and the conditional correlation given $\varepsilon_t > 0$. The case when $\varepsilon_t < 0$ is a reflection through the origin of the former case.

Let $\rho^*_{\varepsilon > 0} := E[\varepsilon^*_t \eta | \varepsilon > 0]$. This relation can be approximated quite accurately by the following equation

$$\hat{\rho}^*_{\varepsilon > 0} = 0.25 \log \left( \frac{1 + \rho}{1 - \rho} \right).$$  \hfill (9)

Figure 2 shows the fitted values and residuals of model (9), as well as a more simple linear fit to $\rho^*_{\varepsilon > 0}$ and its corresponding residuals.

It can be shown that $\rho^*_{\varepsilon < 0} = -\rho^*_{\varepsilon > 0}$. Thus, once the relationship has been established, we run the filter for the linear system in (8.1) and (8.2), the only difference being that $E[w_t \eta_t] = \hat{\rho}^*_{\varepsilon > 0}$ in the case that $u_t > 0$ (the observed residuals of the mean equation), and $E[w_t \eta_t] = -\hat{\rho}^*_{\varepsilon > 0}$ when $u_t < 0$. 

8
Figure 1: Correlation between $\varepsilon^* = \log(\varepsilon^2)$ and $\eta$.

Unfortunately, this simple procedure only recovers the sign but not the magnitude of the correlation successfully. Nonetheless, it helps to understand the implications of the linearization of system (6.1)-(6.2).

We want to find the joint distribution of $\varepsilon^*_t$ and $\eta_t$. Note that the joint distribution of $\varepsilon_t$ and $\eta_t$ is given by

$$f_{\varepsilon_t, \eta_t}(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{ -\frac{1}{2(1-\rho^2)} (x^2 - 2\rho xy + y^2) \right\};$$

therefore,

$$f_{\varepsilon^*_t, \eta_t}(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{ -\frac{1}{2(1-\rho^2)} (e^x - (1-\rho^2)x + y^2) \right\} \cosh\left( \frac{\rho}{1-\rho^2} e^{\frac{1}{2}x} y \right),$$

where $\cosh(z) = \sum_{k=0}^{\infty} \frac{z^{2k}}{(2k)!}$. The complete derivation is presented in Appendix B.

Not surprisingly the marginal distribution of $\varepsilon^*_t$ does not depend on $\rho$. This result was verified through numerical integration and simulations. Table 2 presents the results of a simulation study used to obtain the marginal distribution of $\varepsilon^*_t$.

Indeed, we have that

$$f_{\varepsilon^*_t}(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{ -\frac{1}{2} (e^x - x) \right\},$$

which can be approximated by the mixture of Normals described in the previous section.

Hence, the conditional distribution of $\eta_t | \varepsilon^*_t = x$ is given by

$$f_{\eta_t | \varepsilon^*_t}(y|x) = \frac{1}{\sqrt{2\pi(1-\rho^2)}} \exp\left\{ -\frac{1}{2(1-\rho^2)} (y^2 + \rho^2 e^x) \right\} \cosh\left( \frac{\rho}{1-\rho^2} e^{\frac{1}{2}x} y \right).$$

(10)
This result has a strong implication. It is easy to see that when \( \rho = 0 \), the conditional distribution in (10) is Normal and the Kalman filter remains unaltered. Nevertheless, when \( \rho \neq 0 \) the conditional distribution is no longer Normal and the filter needs to be modified.\(^2\)

Let \( \epsilon_t := \rho \epsilon_t + \sqrt{1 - \rho^2} \eta^*_t \), where \( \eta^*_t \) is a standard normal independent of \( \epsilon_t \). Note that the joint distribution of \( \epsilon_t \) and \( \eta_t \) is equivalent to the joint distribution of \( \epsilon_t \) and \( \epsilon_t \). The filter can be re-written as

\[
Y^*_t = Z_{t-1} + \epsilon^*_t
\]

\[
Z_t = \theta + (1 + \kappa_z)Z_{t-1} + \rho \sigma_z(Y_t - \mu_y)Y_{t-1}^{-\beta}e^{-\frac{1}{2}Z_{t-1}} + \sigma_z\sqrt{1 - \rho^2} \eta^*_t.
\]

This is clearly a non-linear filter. Nevertheless, it is a more tractable filter and one can use standard methodologies to sample from the state variable. For example, one can use a first order

\(^2\)Empirical evidence shows that when the magnitude of \( \rho \) is large, or when \( \epsilon_t \) is 1.5 standard deviations away from the mean, the distribution is bimodal.
Table 2: Results from a simulation study using 10,000 samples of 100,000 observations from the joint distribution of $(\varepsilon_t, \eta_t)$, from which the distribution of $\varepsilon_t^* = \log(\varepsilon_t^2)$ was obtained.

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>0.75</th>
<th>0.5</th>
<th>0.25</th>
<th>0</th>
<th>-0.25</th>
<th>-0.5</th>
<th>-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>-1.26</td>
<td>-1.28</td>
<td>-1.24</td>
<td>-1.26</td>
<td>-1.27</td>
<td>-1.24</td>
<td>-1.28</td>
</tr>
<tr>
<td>Std.Dev.</td>
<td>2.18</td>
<td>2.23</td>
<td>2.18</td>
<td>2.20</td>
<td>2.22</td>
<td>2.19</td>
<td>2.25</td>
</tr>
<tr>
<td>Skewness</td>
<td>-1.57</td>
<td>-1.58</td>
<td>-1.56</td>
<td>-1.61</td>
<td>-1.61</td>
<td>-1.57</td>
<td>-1.60</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>4.42</td>
<td>4.52</td>
<td>4.45</td>
<td>4.62</td>
<td>4.66</td>
<td>4.67</td>
<td>4.36</td>
</tr>
</tbody>
</table>

Taylor approximation around the conditional mean, or compute the time and measurement equations directly from the conditional distribution. The Kalman filter equations for the first order approximation can be found on Appendix B.

Furthermore, we can see from (11.2) that $\partial Z_t/\partial Y_t = \rho \sigma_y Y_{t-1}^{-\beta} e^{-\frac{1}{2}Z_{t-1}}$. Thus, if $\rho < 0$ (ceteris paribus) a drop in the yield translates into an increment in the volatility. Notice that when we are working with stock (log)prices we set $\beta = 0$. In that case, letting $r_t := \Delta Y_t$ (i.e. the log-return) gives us the same result, since $\partial Z_t/\partial r_t = \rho \sigma_y e^{-\frac{1}{2}Z_{t-1}}$. This is consistent with the discussion raised by Yu (2004).

4.4 Posterior of the Jump Times and the Intensity Parameter

To calculate the posterior of the jump times, we consider two cases, one when there is a jump: $J_t = 1$; and the other one when there is no jump: $J_t = 0$.

Directly from the conditional formulation of the likelihood, we have that the posterior probability of a jump at time $t$ is given by

$$P(J_t = 1|X^n, \xi^n, \theta, \theta_J) \propto \phi \left( \frac{\Delta Y_t - \theta - \kappa Y_{t-1} - \xi^0 - \rho \sigma_y (Z_t - \mu_z)}{e^{\frac{1}{2}Z_{t-1}Y_{t-1}^{1/2}(1-\rho^2)}} \right) P(J_t = 1)$$

$$= \lambda \phi \left( \frac{\Delta Y_t - \theta - \kappa Y_{t-1} - \xi^0 - \rho \sigma_y (Z_t - \mu_z)}{e^{\frac{1}{2}Z_{t-1}Y_{t-1}^{1/2}(1-\rho^2)}} \right).$$

Similarly, the posterior probability of no jump at time $t$ is

$$P(J_t = 0|X^n, \xi^n, \theta, \theta_J) \propto (1 - \lambda) \phi \left( \frac{\Delta Y_t - \theta - \kappa Y_{t-1} - \rho \sigma_y (Z_t - \mu_z)}{e^{\frac{1}{2}Z_{t-1}Y_{t-1}^{1/2}(1-\rho^2)}} \right).$$

Note that we have the probabilities up to a proportionality constant, thus we can obtain the probability by a ratio of these quantities.

The posterior distribution of the intensity parameter is straightforward. Assume that $\pi(\lambda) \sim \text{Beta}(\alpha_0, \beta_0)$. Then,

$$p(\lambda|J^n) \sim \text{Beta}(\alpha_0 + n - \sum_{t=1}^n J_t, \beta_0 + \sum_{t=1}^n J_t).$$

4.5 Posterior of the Jump Sizes and $\mu_J$

To start with a simpler model, assume that $\sigma_J$ is known. Thus, we only need to compute the posterior distribution of $\mu_J$, and then the posterior distribution of the jump sizes.
Following the work by Chib, et al., we also find that the estimation of $\mu_J$ is more reliable if we use $\xi^y_t$ and $\mu_J$ as a block and sample from the marginalized posterior of their joint distribution. In other words, if we integrate out the jump sizes.

It can be shown that the posterior distribution of $Y_t$ marginalized over $\xi^y_t$ is

$$p(Y_t | X_{t-1}, Z_t, J_t, \theta, \theta_J) \propto \exp \left\{ -\frac{1}{2} \frac{(Y_t - \theta_J - (1 + \kappa_J)Y_{t-1} - \mu_J J_t - \rho \sigma_J^2 (Z_t - \mu_J))^2}{\sigma^2_J + \sigma^2_J J_t} \right\}.$$ 

Thus,

$$p(\mu_J | X^n, J^n, \theta, \theta_J) \sim N(\hat{\mu}_J, S_J),$$

where

$$S_J = \left( \sum_{t=1}^n \frac{J_t}{\sigma^2_{y|z} + \sigma^2_J} \right)^{-1}$$

and $\hat{\mu}_J = S_J \sum_{t=1}^n \frac{(\Delta Y_t - \kappa_J Y_{t-1} - \rho \sigma_J^2 (Z_t - \mu_J)) J_t}{\sigma^2_{y|z} + \sigma^2_J J_t}$.

Finally, for the jump sizes, we condition on the value of $J_t$. If there is no jump, then we do not have new information about the size and we simply draw from the prior distribution. Otherwise, the posterior distribution is

$$p(\xi^y_t | X_t, X_{t-1}, J_t = 1, \theta, \theta_J) \sim N\left( \tau_t^{-1} \alpha_t, \tau_t \right),$$

where

$$\alpha_t = \frac{(\Delta Y_t - \kappa_J Y_{t-1} - \rho \sigma_J^2 (Z_t - \mu_J))}{\sigma^2_{y|z} + \sigma^2_J J_t},$$

$$\tau_t^{-1} = \frac{1}{\sigma^2_{y|z} + 1/\sigma^2_J J_t}.$$

### 4.6 Contemporaneous jumps in the mean and the volatility

A logical next step is to consider the case when the jumps occur in both the mean and the volatility equation. This is the so-called Stochastic Volatility with Conditional Jumps (SVCJ) model. In that case, consider the model

$$\Delta Y_t = \theta_J + \kappa_J Y_{t-1} + \epsilon^2_t \zeta_{1-t-1} \xi_t^y + \xi^y_t J_t, \quad (12.1)$$

$$\Delta Z_t = \theta_J + \kappa_J Z_{t-1} + \sigma_J \eta_t + \xi^z_t J_t, \quad (12.2)$$

where $(\xi_t, \eta_t)$ are bivariate Normal with mean zero, unit variance and correlation $\rho$; $J_t \sim \text{Ber}(\lambda)$; $\xi^z_t = \text{sign}_t \zeta_t^z$, with $\zeta^z_t \sim N(\mu^z_J, \sigma^z_J)$; and $\xi^y_t | \xi_t^z \sim N(\mu^y_J + \rho J_t \xi^z_t, \sigma^y_J)$.

To sample from the posterior distribution of the parameters, do exactly as before except for the jump part.

In summary, the posterior of $\lambda$, simply use the bivariate density instead of the conditional density (since the jumps affect the volatility as well.) The posterior of $\mu^z_J \sim N(\hat{\mu}_J, \hat{S}_J)$; the posterior of $\xi^z_t \sim N(\hat{\mu}^z_J, \hat{S}_J)$ if $J_t = 1$; and if $J_t = 0$, $\xi^z_t$ comes from prior. The posterior of $\mu^y_J, \rho_J$ is obtained by regression of $\xi^y_t$ on $\xi^z_t$. The posterior of $\xi^y_t \sim N(\hat{\mu}^y_J, \hat{S}_J^2)$ if $J_t = 1$; and if $J_t = 0$, $\xi^y_t$ comes from
5 Results

In this section we present the results from the estimation of the SV, SVJ, and SVCJ models for both simulated data and real data. Let us start with the description of the data we use to fit the models. It corresponds to the weekly (Wednesday) yield of the 3-month Treasury bill from 1/1/1954 to 10/5/1997. This particular sample period was chosen to be able to compare the results obtained by Eraker (2001), where a more detailed description of the data can be found.

The next subsection presents the simulation study that was performed to assess the accuracy and efficiency of the algorithm. It also compares the results with those obtained using Eraker (2001) methodology to sample from the volatility path, i.e. for each $t$, one samples from the posterior distribution $p(Z_t|X_{t-1}, X_{t+1}, \xi_{t+1}, J_{t+1}, \theta, \theta_J)$. Similarly, subsection 5.2 shows the results using the Treasury bill data.

5.1 Simulation Study

We divide this section in three: first we talk about the SV model without correlation, then we include correlation in the process, and finally we simulate and estimate a process with jumps in the mean.

5.1.1 SV model with no correlation

According to the results found in Eraker (2001), ten simulated data sets were created. To simulate the data the Euler discretization was used with $\Delta = 20$, and then every 20th observation was chosen. The parametrization that he uses is slightly different from the one described in (6.1) and (6.2). In particular, he uses $\sigma_r$ to account for the mean of the log-volatility process instead of $\theta_z$. Nevertheless, there is a 1-to-1 relation between them.

The algorithm described in Section 4 was run for 10,000 iterations and the first 3,000 were used as burn-in. The algorithm for the volatility described in Eraker (2001) was run for 30,000 iterations (10,000 burn-in) because it seems to take longer to converge. Furthermore, for a fixed number of iterations the second algorithm takes ten times as long. The values of the parameters used to simulate the process, along with the posterior means and posterior standard deviations with the last 7,000 and 20,000 draws, respectively, are given in Table 3.

From Part I of Table 3 we can see that the results seem reasonable. The standard errors of $\theta_r$ and $\kappa_r$ are a bit large. The same happens with the standard error of $\beta$. Part II shows that the results are consistent using either methodology.

5.1.2 SV model with negative correlation

Ten more datasets were simulated including a negative correlation in the process. In particular, we used $\rho = -0.4$. The choice of the value relates to findings in the literature, especially to that of Eraker et al. (2003), although this negative correlation refers to the returns of the S&P 500 and Nasdaq index. To make sure the small sample size was not an issue in the estimation of the correlation, a data set with 10,000 observations was used.

In the comparative study, ten similar simulated data sets with 2,500 observations were created. The smaller sample size responds to the fact that the second algorithm is much slower and such a large data set becomes almost intractable.

The performance of the algorithm described in Section 4 is good. It was run for 10,000 iterations, and the first 3,000 were used as burn-ins. The correlation is recovered (other simulated processes
Table 3: Simulation results for the SV model
This table presents the posterior means and standard errors (in parenthesis) of the SV model from 10 simulated data sets with 2,000 observations. Part I corresponds to the results produced by the algorithm described in Section 4 using 10,000 iterations from which the first 3,000 were discarded. Part II corresponds to the results produced by the algorithm with Eraker (2001) methodology to sample from the volatility path, using 30,000 iterations from which the first 10,000 were discarded.

<table>
<thead>
<tr>
<th></th>
<th>$\theta_r$</th>
<th>$\kappa_r$</th>
<th>$\theta_z$</th>
<th>$\kappa_z$</th>
<th>$\sigma_z$</th>
<th>$\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>0.0123</td>
<td>-0.0020</td>
<td>-0.1302</td>
<td>-0.0201</td>
<td>0.2457</td>
<td>0.7813</td>
</tr>
<tr>
<td>Mean</td>
<td>0.0109</td>
<td>-0.0018</td>
<td>-0.1251</td>
<td>-0.0217</td>
<td>0.2294</td>
<td>0.7752</td>
</tr>
<tr>
<td>Std. Dev</td>
<td>(0.0054)</td>
<td>(0.0008)</td>
<td>(0.0232)</td>
<td>(0.0071)</td>
<td>(0.0246)</td>
<td>(0.0201)</td>
</tr>
</tbody>
</table>

With positive correlation were used, although they are not discussed here), both in sign and in magnitude. This can be seen in Table 4. Nonetheless, the algorithm needs to have some prior specification about the sign of $\rho$ and its magnitude. This is not crucial, but speeds the convergence of the algorithm.

Figure 3 represents the plot of the draw from each posterior versus the iteration. There we can see that the parameters seem to have stabilized. For $\sigma_v$, $\beta$, and $\rho$ we can observe some persistence, but it does not seem to be high.

5.1.3 SVJ model
The algorithm for the SVJ model was tested using a simulated process, consisting of 2,000 observations. The values of the jump parameters where chosen as follows: $\lambda$ corresponds to an average of two jumps per year (using weekly data); $\mu_J$ and $\sigma_J$ reflect large positive jumps in the yield.

First, to study the properties of the jump, part of the algorithm was run using the volatility path $Z_1, \ldots, Z_n$ as given. This way we can be sure we are getting the right jump times, intensity and sizes when the right path is chosen. The summary results are presented in Table 5.

From Table 5 we can see that all the parameters can be reliably estimated when the log-volatility is given. The important study here is whether the jump times and jump sizes can be accurately reproduced. Figure 4 presents the mean probability of a jump (calculated averaging the probabilities of jump over the 7,000 iterations) when the true jump is one, and the estimated mean probability of a jump when there is no true jump. We can see that the algorithm was able to pick most jumps correctly (just underestimating a couple) and was producing few false positives (setting $J_t^{(h)} = 1$, when there is no jump).

Similarly, the sizes seem correct. Figure 5 presents the true size when the true jump is one, and the estimated size. The graph when the true jump is zero is not shown, but it also shows the capabilities of the algorithm to reproduce sizes correctly.
Table 4: Simulation results for the SV model with negative correlation

This table presents the posterior means and standard errors (in parenthesis) of the SV model with correlation. Part I shows the results produced by the algorithm described in section 4, from 10 simulated data sets containing 10,000 observations. The algorithm was run for 10,000 iterations from which the first 3,000 were discarded. Part II corresponds to the results using the alternative sampling procedure from the volatility path as described in Eraker (2001). The data sets contained 2,500 observations and were run for 30,000 iterations from which the first 10,000 were discarded.

<table>
<thead>
<tr>
<th></th>
<th>$\theta_r$</th>
<th>$\kappa_r$</th>
<th>$\theta_z$</th>
<th>$\kappa_z$</th>
<th>$\sigma_z$</th>
<th>$\beta$</th>
<th>$\rho$</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>0.0123</td>
<td>-0.0020</td>
<td>-0.1302</td>
<td>-0.0201</td>
<td>0.2457</td>
<td>0.7813</td>
<td>-0.4000</td>
</tr>
<tr>
<td>Mean</td>
<td>0.0141</td>
<td>-0.0024</td>
<td>-0.1378</td>
<td>-0.0205</td>
<td>0.2302</td>
<td>0.7804</td>
<td>-0.3814</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>(0.0061)</td>
<td>(0.0019)</td>
<td>(0.02980)</td>
<td>(0.0062)</td>
<td>(0.0161)</td>
<td>(0.0206)</td>
<td>(0.0417)</td>
</tr>
</tbody>
</table>

Finally, the algorithm was run to estimate all parameters, volatility, jumps times, and sizes. The results are shown in Table 6. From there we can see that the estimate of $\lambda$ is rather large, meaning that we are overestimating the number of jumps. This gets reflected as well in the estimate of $\mu_J$ which is smaller than the true value. The other estimates are not so different from what we had before. For the case of the jump times, we have more false positives than before, as can be seen in Figure 6. This is simply a reflection of the estimated value of $\lambda$.

5.1.4 SVCJ model

Now the algorithm for the SVCJ model was tested using a simulated process, consisting of 2,000 observations. Similarly, the jump parameters where chosen so that $\lambda$ corresponds to an average of two jumps per year (using weekly data); $\mu_J^x$, $\mu_J^y$ and $\sigma_J^x$, and $\sigma_J^y$ reflect positive jumps in the yield three standard deviations away from the mean. Notice that for the yield case, we do not have to deal with the sign, since we are restricting the jumps to be positive. To study the properties of the jump, the algorithm was run using the volatility path $Z_1, \ldots, Z_n$ as given. The summary results are presented in Table 7.

Finally, the algorithm was run to estimate all parameters, volatility, jumps times and sizes. The results are shown in Table 8. The new paramtrization makes the estimates more reliable. Indeed, the estimate of $\lambda$ is large but not as much as it was for the SVJ model. We obtain similar results for the other parameters as well.

5.2 Estimation of the SV model for the Treasury bill

Using the t-bill data we estimated the SV model with correlation. The estimation of the SVJ model will be reported later.

Figure 7 presents the time series plot of the Treasury bill yield rate used to fit the SV model.
Table 9 shows the posterior means and standard errors of the parameters using the t-bill data. It also presents the results obtained by Eraker (2001) when $\Delta t = 1$, and when $\Delta t = 1/8$. Finally, it shows the results using the alternative algorithm to sample from the volatility path. From there we can see that the results of the algorithm described in section 4 seem consistent, and they actually are closer to the computations when $\Delta = 1/8$. Notice that the latter corresponds to a case when Eraker uses data augmentation for the estimation. With respect to the alternative methodology, the results are also very similar except for the correlation parameter for which the filter methodology produces a smaller magnitude estimate.

Finally, Figure 8 shows plots of the standardized residuals of the SV model for the t-bill. The errors of the return show heavier tails than normal. This result is of common knowledge, and the use of a t-distribution instead of normal may alleviate the unusual observations.

The last table, presents the estimated parameters for the SVCJ model for the T-bill data. As we can see from 10, the results seem consistent with the previous findings, although the leverage effect is smaller. The estimate of $\lambda$ points towards an approximate 2 jumps per year, and the correlation between jumps is negligible.
Table 5: Simulation results for the SVJ model with $Z_1, \ldots, Z_n$ given.
This table presents the posterior means and standard errors (in parenthesis) of the SVJ model given the volatility produced by the algorithm using 10,000 iterations from which the first 3,000 were discarded.

<table>
<thead>
<tr>
<th></th>
<th>$\theta_r$</th>
<th>$\kappa_r$</th>
<th>$\theta_z$</th>
<th>$\kappa_z$</th>
<th>$\sigma_z$</th>
<th>$\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>0.0123</td>
<td>−0.0020</td>
<td>−0.1302</td>
<td>−0.0201</td>
<td>0.2457</td>
<td>0.7813</td>
</tr>
<tr>
<td>Mean</td>
<td>0.0176</td>
<td>−0.0021</td>
<td>−0.1084</td>
<td>−0.0164</td>
<td>0.2417</td>
<td>0.7742</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>(0.0097)</td>
<td>(0.0005)</td>
<td>(0.0272)</td>
<td>(0.0040)</td>
<td>(0.0039)</td>
<td>(0.0053)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\lambda$</th>
<th>$\mu_j$</th>
<th>$\sigma_j$</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>0.0400</td>
<td>4.0000</td>
</tr>
<tr>
<td>Mean</td>
<td>0.0344</td>
<td>3.9616</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>(0.0021)</td>
<td>(0.0901)</td>
</tr>
</tbody>
</table>

Table 6: Simulation results for the SVJ model.
This table presents the posterior means and standard errors (in parenthesis) of the SVJ model by the algorithm using 10,000 iterations from which the first 3,000 were discarded.

<table>
<thead>
<tr>
<th></th>
<th>$\theta_r$</th>
<th>$\kappa_r$</th>
<th>$\theta_z$</th>
<th>$\kappa_z$</th>
<th>$\sigma_z$</th>
<th>$\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>0.0123</td>
<td>−0.0020</td>
<td>−0.1302</td>
<td>−0.0201</td>
<td>0.2457</td>
<td>0.7813</td>
</tr>
<tr>
<td>Mean</td>
<td>0.0097</td>
<td>−0.0014</td>
<td>−0.1676</td>
<td>−0.0116</td>
<td>0.2017</td>
<td>0.8942</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>(0.0057)</td>
<td>(0.0006)</td>
<td>(0.0272)</td>
<td>(0.0040)</td>
<td>(0.0139)</td>
<td>(0.1253)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\lambda$</th>
<th>$\mu_j$</th>
<th>$\sigma_j$</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>0.0400</td>
<td>4.0000</td>
</tr>
<tr>
<td>Mean</td>
<td>0.0724</td>
<td>2.8460</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>(0.0121)</td>
<td>(0.5801)</td>
</tr>
</tbody>
</table>

6 Conclusion

Some concluding remarks are in order. First, the algorithm for the stochastic diffusion model without jumps and correlation seems stable and reliable. For both the simulated and real data, the results are satisfactory. At first, the estimation of the parameter $\beta$ took a long time to become stable. Using a more sophisticated Metropolis-Hastings the results improved.

When correlation is added, the original Kalman filter methodology is not able to recover this structure, since it is destroyed when we linearize the mean equation. To solve this issue the conditional distribution as a function of $\rho$ was found. From there the inclusion of the information in the linear filter through Taylor approximation was used a first step, but a more accurate algorithm is achieved using the non-gaussian conditional distribution.

For the jumps, the belief of the authors is that the failure of the algorithm to recover the jump parameters more accurately is related to the fact that, if the prior of $\lambda$ is not set correctly, every observation could be regarded as a jump. The algorithm works fine when the true path is given. Perhaps the linearization of the mean equation has some relevance here as well.

Second, there is evidence that stochastic volatility models with jumps not only in the mean reflect better the behaviour of the market. Such models can be found in Eraker et al., among
Figure 4: Mean probabilities of jump of the SVJ model given volatility.

others. This issue will be investigated further.

Appendix A

Here we present the derivation for the posterior distributions of the parameters discussed in Section 4.1,

(a) Posterior distribution of \((\theta_r, \kappa_r)\)

Using the notation used in subsection 4.1.1, and letting \(\mu_r = (\theta_r, \kappa_r)\)' , we have that

\[
p(\theta_r, \kappa_r| X^n, J^n, \xi^n, \theta^-, \theta_f) \propto \prod_{t=1}^n p(Y_t| Y_{t-1}, Z^t, \theta, \theta_f) \pi(\theta_r) \pi(\kappa_r)
\]

\[
\propto \prod_{t=1}^n \exp \left\{ -\frac{1}{2} \frac{(Y_t - \mu_{\theta|z})^2}{\sigma_{\theta|z}^2} \right\} = \exp \left\{ -\frac{1}{2} (\mathbf{y} - \mathbf{X}\mu_r)'(\mathbf{y} - \mathbf{X}\mu_r) \right\} \\
\propto \exp \left\{ -\frac{1}{2} (\mu_r - \hat{\mu}_r)'(\mathbf{X}'\mathbf{X}(\mu_r - \hat{\mu}_r)) \right\}.
\]
This implies that
\[ p(\theta_r, \kappa_r|X^n, J^n, \xi^n, \theta^-, \theta_f) \sim N_2(\tilde{\mu}_r, (X'X)^{-1}), \]
where \( \tilde{\mu}_r = (X'X)^{-1}X'y \)

Note that the inclusion of more informative prior distributions is straightforward and follows the classic Bayesian updating.

(b) Posterior distribution of \((\theta_z, \kappa_z)\)

Similarly, let \( \mu_z = (\theta_z, \kappa_z)' \). Then,
\[
p(\theta_z, \kappa_z|X^n, J^n, \xi^n, \theta^-, \theta_f) \propto \prod_{t=1}^n p(Z_t|Z_{t-1}, Y^t, \theta, \theta_f)\pi(\theta_z)\pi(\kappa_z) \\
\propto \prod_{t=1}^n \exp \left\{ -\frac{1}{2} \frac{(Z_t-\mu_z)^2}{\sigma_{z_t}^2} \right\} = \exp \left\{ -\frac{1}{2} (y_z - X_z\mu_r)'(y_z - X_z\mu_z) \right\} \\
\propto \exp \left\{ -\frac{1}{2} (\mu_z - \tilde{\mu}_z)'(\sigma_z^{-2}X_z'X_z)(\mu_z - \tilde{\mu}_z) \right\}
\]

This implies that
\[ p(\theta_z, \kappa_z|X^n, J^n, \xi^n, \theta^-, \theta_f) \sim N_2(\mu_z, \sigma_z^{-2}(X_z'X_z)^{-1}), \]
where \( \hat{\mu}_z = (X'_z X_z)^{-1} X'_z y_z \).

(c) Posterior of \((\sigma_z, \rho)\)

For the joint posterior of \(\sigma_z\) and \(\rho\), recall the parametrization defined in section 4.1.3. Now, let \(R\) be the variance-covariance matrix of the residuals. Then

\[
R = \begin{bmatrix} u'u & u'v \\ v'u & v'v \end{bmatrix}.
\]

It is easy to re-write (7.1) with the new parametrization as

\[
L(\theta) \propto |\Omega|^{-\frac{3}{2}} \exp \left\{ -\frac{1}{2} \text{tr}(\Omega^{-1} R) \right\}
\]

Furthermore, note that \(|\Omega| = \phi\), and that

\[
\Omega^{-1} = \frac{1}{\phi} \begin{pmatrix} \psi^2 & -\psi \\ -\psi & 1 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} =: \frac{1}{\phi} S + \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.
\]
Table 7: Simulation results for the SVCJ model with $Z_1, \ldots, Z_n$ given.
This table presents the posterior means and standard errors (in parenthesis) of the SVCJ model
given the volatility produced by the algorithm using 10,000 iterations from which the first 3,000
were discarded.

<table>
<thead>
<tr>
<th></th>
<th>$\theta_r$</th>
<th>$\kappa_r$</th>
<th>$\theta_z$</th>
<th>$\kappa_z$</th>
<th>$\sigma_z$</th>
<th>$\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>0.0123</td>
<td>-0.0020</td>
<td>-0.1302</td>
<td>-0.0201</td>
<td>0.2457</td>
<td>0.7813</td>
</tr>
<tr>
<td>Mean</td>
<td>0.0167</td>
<td>-0.0019</td>
<td>-0.1101</td>
<td>-0.0171</td>
<td>0.2359</td>
<td>0.7699</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>(0.0081)</td>
<td>(0.0006)</td>
<td>(0.0254)</td>
<td>(0.0051)</td>
<td>(0.0042)</td>
<td>(0.0121)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$\lambda$</th>
<th>$\mu_j^y$</th>
<th>$\sigma_{jy}$</th>
<th>$\mu_j^z$</th>
<th>$\sigma_{jz}$</th>
<th>$\rho_j$</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>0.0100</td>
<td>0.6000</td>
<td>1.0000</td>
<td>0.6000</td>
<td>1.0000</td>
<td>0.5000</td>
</tr>
<tr>
<td>Mean</td>
<td>0.0093</td>
<td>0.6142</td>
<td>1.0000</td>
<td>0.5724</td>
<td>1.0000</td>
<td>0.4300</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>(0.0032)</td>
<td>(0.0571)</td>
<td>(0.0721)</td>
<td>(0.0081)</td>
<td>(0.0121)</td>
<td>(0.0571)</td>
</tr>
</tbody>
</table>

Table 8: Simulation results for the SVCJ model.
This table presents the posterior means and standard errors (in parenthesis) of the SVCJ model
by the algorithm using 10,000 iterations from which the first 3,000 were discarded.

<table>
<thead>
<tr>
<th></th>
<th>$\theta_r$</th>
<th>$\kappa_r$</th>
<th>$\theta_z$</th>
<th>$\kappa_z$</th>
<th>$\sigma_z$</th>
<th>$\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>0.0123</td>
<td>-0.0020</td>
<td>-0.1302</td>
<td>-0.0201</td>
<td>0.2457</td>
<td>0.7813</td>
</tr>
<tr>
<td>Mean</td>
<td>0.0101</td>
<td>-0.0016</td>
<td>-0.1251</td>
<td>-0.0174</td>
<td>0.2561</td>
<td>0.8062</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>(0.0061)</td>
<td>(0.0005)</td>
<td>(0.0283)</td>
<td>(0.0058)</td>
<td>(0.0201)</td>
<td>(0.0571)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$\lambda$</th>
<th>$\mu_j^y$</th>
<th>$\sigma_{jy}$</th>
<th>$\mu_j^z$</th>
<th>$\sigma_{jz}$</th>
<th>$\rho_j$</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>0.0100</td>
<td>0.6000</td>
<td>1.0000</td>
<td>0.6000</td>
<td>1.0000</td>
<td>0.5000</td>
</tr>
<tr>
<td>Mean</td>
<td>0.0141</td>
<td>0.5642</td>
<td>1.0000</td>
<td>0.5431</td>
<td>1.0000</td>
<td>0.5417</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>(0.0058)</td>
<td>(0.0810)</td>
<td>(0.0721)</td>
<td>(0.0081)</td>
<td>(0.0121)</td>
<td>(0.0571)</td>
</tr>
</tbody>
</table>

Hence, $\text{tr}(\Omega^{-1} \mathbf{R}) = \text{tr}(\mathbf{SR})/\phi + \mathbf{u}'\mathbf{u}$.

Let $\phi \sim IG(\alpha_0, \beta_0)$, and $\psi|\phi \sim N(0, \phi/2)$. Therefore,

$$
\pi(\psi|\phi, X^n, J^n, \xi^n, \theta^r, \theta^z) \propto \frac{1}{\phi^{n/2}} \exp \left\{ -\frac{1}{2\phi} \text{tr}(\mathbf{SR}) \right\} \pi(\psi, \phi).
$$

After some manipulation, this can be decomposed into a Normal and an Inverse Gamma distributions, namely

$$
p(\psi|X^n, J^n, \xi^n, \theta^r, \theta^z) \sim N\left( \bar{\psi}, \phi/(\beta_0 + \mathbf{u}'\mathbf{u}) \right),
$$

where $\bar{\psi} = \mathbf{u}'\mathbf{v}/(\beta_0 + \mathbf{u}'\mathbf{u})$. And

$$
p(\phi|X^n, J^n, \xi^n, \theta^r, \theta^z) \sim IG\left( (n - 1 + \alpha_0)/2, \beta_0 + \mathbf{v}'\mathbf{v} - (\mathbf{u}'\mathbf{v})^2/\mathbf{u}'\mathbf{u} \right).
$$

(d) Posterior distribution of $\beta$

Let $c_t^2 := (Y_t - \mu_0)^2 e^{-Z_{t-1}}$. The log-likelihood function of beta can be written as

$$
l(\beta) = -\beta \sum_{t=2}^n \log(Y_{t-1}) - \frac{1}{2} \sum_{t=2}^n c_t^2 Y_{t-1}^{1-2\beta} + C
$$

$$
= -\beta \sum_{t=2}^n \log(Y_{t-1}) - \frac{1}{2} \sum_{t=2}^n c_t^2 e^{-2\beta Y_{t-1}} + C
$$
where $C$ is a constant that does not depend of $\beta$. Differenciating with respect to $\beta$ gives us

$$l'(\beta) = -\sum_{t=2}^{n} \log(Y_{t-1}) - \frac{1}{2} \sum_{t=2}^{n} c_t^2 e^{-2\beta Y_{t-1}} (-2 \log(Y_{t-1}))$$

$$= -\sum_{t=2}^{n} \log(Y_{t-1}) + \sum_{t=2}^{n} c_t^2 Y_{t-1}^{-2\beta} (\log(Y_{t-1})).$$

Finally, differenciating again with respect to $\beta$ gives

$$l''(\beta) = -2 \sum_{t=2}^{n} c_t^2 e^{-2\beta Y_{t-1}} (\log(Y_{t-1}))^2 = -2 \sum_{t=2}^{n} c_t^2 Y_{t-1}^{-2\beta} (\log(Y_{t-1}))^2,$$  

which is positive. Hence, the log-posterior of $\beta$ is log-concave. Using a second-order Taylor expansion of the log-posterior, we get the proposal distribution

$$N\left(\beta - \frac{l'(\beta)}{l''(\beta)}, -(l''(\beta))^{-1}\right).$$

Figure 7: Weekly (Wednesday) yield of the 3-month Treasury bill from 1/1/1954 to 10/5/1997.
Table 9: Estimation of the SV for the Treasury bill data
Part I presents the results in Eraker (2001) using $\Delta t = 1$ and $\Delta t = 1/8$. Part I also shows the posterior means and standard errors (in parenthesis) for the SV model fitted to the weekly (Wednesday) three-month t-bill rate from 1/1/1954 to 10/5/1997. The summary statistics were calculated using 7,000 iterations, after a 3,000 iteration burn-in. Part II displays the results obtained by using the alternative methodology to sample from the volatility path, as described in Eraker (2001). The latter algorithm was ran for 25,000 iterations with 5,000 iterations used as burn-in.

<table>
<thead>
<tr>
<th></th>
<th>$\theta_r$</th>
<th>$\kappa_r$</th>
<th>$\theta_z$</th>
<th>$\kappa_z$</th>
<th>$\sigma_z$</th>
<th>$\beta$</th>
<th>$\rho$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta t = 1$</td>
<td>0.0123</td>
<td>-0.0020</td>
<td>-0.1302</td>
<td>-0.0201</td>
<td>0.2457</td>
<td>0.7813</td>
<td>-</td>
</tr>
<tr>
<td>$\Delta t = \frac{1}{8}$</td>
<td>0.0114</td>
<td>-0.0018</td>
<td>-0.1451</td>
<td>-0.0240</td>
<td>0.2770</td>
<td>0.7027</td>
<td>-</td>
</tr>
<tr>
<td>Mean</td>
<td>0.0111</td>
<td>-0.0019</td>
<td>-0.1625</td>
<td>-0.0278</td>
<td>0.2213</td>
<td>0.7729</td>
<td>-0.1518</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>(0.0050)</td>
<td>(0.0012)</td>
<td>(0.0382)</td>
<td>(0.0064)</td>
<td>(0.0201)</td>
<td>(0.0183)</td>
<td>(0.0298)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$\theta_r$</th>
<th>$\kappa_r$</th>
<th>$\theta_z$</th>
<th>$\kappa_z$</th>
<th>$\sigma_z$</th>
<th>$\beta$</th>
<th>$\rho$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.0124</td>
<td>-0.0024</td>
<td>-0.1470</td>
<td>-0.0215</td>
<td>0.2307</td>
<td>0.8106</td>
<td>-0.1719</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>(0.0047)</td>
<td>(0.0012)</td>
<td>(0.0319)</td>
<td>(0.0046)</td>
<td>(0.0108)</td>
<td>(0.0129)</td>
<td>(0.0487)</td>
</tr>
</tbody>
</table>

Appendix B

Filter for volatility path

Forward Filtering

Consider a system in which both shocks are contemporaneous in each of the state-space equations, but the state variable is observed with a delay. Hence, we have

$$
\begin{align*}
z_{k+1} &= F_k z_k + G_k w_{k+1} \\
y_{k+1} &= H_k' z_k + v_{k+1}.
\end{align*}
$$

Table 10: Estimation of the SVCJ model for the Treasury bill data.
This table shows the posterior means and standard errors (in parenthesis) for the SVCJ model fitted to the weekly (Wednesday) three-month t-bill rate from 1/1/1954 to 10/5/1997. The summary statistics were calculated using 7,000 iterations, after a 3,000 iteration burn-in.

<table>
<thead>
<tr>
<th></th>
<th>$\theta_r$</th>
<th>$\kappa_r$</th>
<th>$\theta_z$</th>
<th>$\kappa_z$</th>
<th>$\sigma_z$</th>
<th>$\beta$</th>
<th>$\rho$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.017</td>
<td>-0.0018</td>
<td>-0.1873</td>
<td>-0.0285</td>
<td>0.2258</td>
<td>0.7921</td>
<td>-0.1282</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>(0.0052)</td>
<td>(0.0012)</td>
<td>(0.0258)</td>
<td>(0.0069)</td>
<td>(0.0218)</td>
<td>(0.0243)</td>
<td>(0.0276)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$\lambda$</th>
<th>$\mu_{\lambda}$</th>
<th>$\sigma_{\lambda}$</th>
<th>$\mu_{j}$</th>
<th>$\sigma_{j}$</th>
<th>$\rho_{j}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.0415</td>
<td>-0.0941</td>
<td>0.5000</td>
<td>2.0381</td>
<td>0.5000</td>
<td>0.0638</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>(0.0106)</td>
<td>(0.0571)</td>
<td>(0.2573)</td>
<td>(0.067)</td>
<td>(0.0607)</td>
<td></td>
</tr>
</tbody>
</table>
where \((w_k, v_k)\) is bivariate normal, such that:

(i) \(E[w_k] = E[v_k] = 0\);

(ii) \(\text{Cov}(w_k, w_\ell) = Q_k I\{k=\ell\}\), \(\text{Cov}(v_k, v_\ell) = R_k I\{k=\ell\}\); and

(iii) \(\text{Cov}(w_k, v_\ell) = S_k I\{k=\ell\}\).

Also, assume that \(z_0 \sim N(\mu_0, P_0)\) and that it is independent of \((w_k, v_k)\) for all \(k\).

We want to obtain information about \(z_k\), using measurements up to time \(k\). Nevertheless, it is also convenient to include the information that can be obtained given measurements up to time \(k-1\) (i.e. one-step ahead prediction.)

Let \(Y_k := \{y_0, y_1, \ldots, y_k\}\), and define

\[
\hat{z}_{k|k-1} := E[z_k|Y_{k-1}], \quad \text{and} \quad P_{k|k-1} := E[(z_k - \hat{z}_{k|k-1})(z_k - \hat{z}_{k|k-1})'].
\]

Similarly, set

\[
\hat{z}_{k|k} := E[z_k|Y_k], \quad \text{and} \quad P_{k|k} := E[(z_k - \hat{z}_{k|k})(z_k - \hat{z}_{k|k})'].
\]
The time-update recursions, for \( k \geq 0 \), are
\[
\begin{align*}
\hat{z}_{k+1|k} &= F_k \hat{z}_{k|k}, \\
P_{k+1|k} &= F_k P_{k|k} F_k^T + G_k Q_{k+1} G_k^T,
\end{align*}
\]
where \( \hat{z}_{0|0} = \mu_0 \), and \( P_{0|0} = P_0 \).

The measurement update recursions, for \( k \geq 0 \), are
\[
\begin{align*}
\hat{z}_{k+1|k+1} &= \hat{z}_{k+1|k} + \hat{K}_k (y_{k+1} - H_k \hat{z}_{k|k}), \\
P_{k+1|k+1} &= P_{k+1|k} - \hat{K}_k (P_k F_k H_k + G_k S_{k+1}), \\
\hat{K}_k &= \left( F_k P_k H_k + G_k S_{k+1} \right) \left( H_k^T P_k H_k + R_{k+1} \right)^{-1}.
\end{align*}
\]

**Backward Sampling**

The ultimate goal is to be able to sample from \( z_1, \ldots, z_M \), using their joint distribution with the information up to time \( M \), i.e. from the full information set \( Y_M \).

First, using the property that the state variable is Markov, we have that
\[
P(z_1, \ldots, z_M | Y_M) = P(z_M | Y_M) \prod_{k=1}^{M-1} P(z_k | z_{k+1}, Y_M).
\]

Let \( \hat{y}_k = E[y_k | Y_k - 1] \), and \( \hat{y}_k = y_k - \hat{y}_k | k - 1 \). Now, let \( Y_{M}^k := \{y_k, \ldots, y_M\} \), and notice that \( \hat{Y}_{M}^k := \{\hat{y}_k, \ldots, \hat{y}_M\} \) contains the same information. Now, note that \( Y_{k+1}^M \) (or equivalently \( \hat{Y}_{k+1}^M \)) is independent of \( Y_k \). Thus
\[
P(z_k | z_{k+1}, Y_M) = P(z_k | z_{k+1}, Y_k, Y_{k+1}^M) = P(z_k | z_{k+1}, Y_k).
\]

Due to the Gaussian structure of the model, the distribution of \( (z_k, z_{k+1}) | Y_k \) is Gaussian, as well. We know the marginal distributions of \( z_k | Y_k \) and \( z_{k+1} | Y_k \) from the different derivations of the filters. The only part that remains is to derive the covariance structure.

Let \( C_{k,k+1} := E[(z_k - \hat{z}_{k|k})(z_{k+1} - \hat{z}_{k+1|k})^T | Y_k] \) and note that, by the orthogonality principle, \( C_{k,k+1} \) is independent of \( Y_k \). In general, we have that
\[
\begin{align*}
\left( \begin{array}{c} z_k \\ z_{k+1} \end{array} \right) | Y_k & \sim N \left( \left( \begin{array}{c} \hat{z}_{k|k} \\ \hat{z}_{k+1|k} \end{array} \right), \left[ \begin{array}{cc} P_k | k \\ C_{k,k+1} & P_{k+1|k} \end{array} \right] \right),
\end{align*}
\]

Therefore,
\[
z_k | z_{k+1}, Y_k \sim N \left( \hat{z}_{k|k} + C_{k,k+1} P_{k+1|k}^{-1} (z_{k+1} - \hat{z}_{k+1|k}), P_k | k - C_{k,k+1} P_{k+1|k}^{-1} C_{k,k+1}^T \right),
\]
where \( C_{k,k+1} = P_k | k F_k^T \).

**Filter for log-volatility when correlation is present**

Here we show the first-order Taylor approximation filter equations when the diffusion process experiences correlation.

Let \( \varepsilon_t := \rho \varepsilon_t + \sqrt{1 - \rho^2} \eta^*_t \), where \( \eta^*_t \) is a standard normal independent of \( \varepsilon_t \). Note that the joint distribution of \( \varepsilon_t \) and \( \eta^*_t \) is equivalent to the joint distribution of \( \varepsilon_t \) and \( \varepsilon_t \).
Recall that and set

\[ Y_t^* = \mu_y + e^{x_t Y_{t-1}} \varepsilon_t \]
\[ Z_t = \theta_z + (1 + \kappa_z) Z_{t-1} + \sigma_z \left( \rho \varepsilon_t + \sqrt{1 - \rho^2} \eta_t^* \right) \]

Thus, after we linearize the first equation, we can re-write the system as

\[ Y_t^* = Z_{t-1} + \varepsilon_t^* \]
\[ Z_t = \theta_z + (1 + \kappa_z) Z_{t-1} + \rho \sigma_z (Y_t - \mu_y) Y_{t-1}^{-\beta} e^{-z_{t-1}} + \sigma_z \sqrt{1 - \rho^2} \eta_t^* \]

where \( Y_t^* = \log(Y_t - \mu_y)^2 - \beta \log(Y_{t-1})^2 \), and \( \varepsilon_t^* = \log(\varepsilon_t^2) \).

To be consistent with the notation of the Kalman filter section, let

\[ f_k(x_k) := (1 + \kappa_z)x_k + \rho \sigma_y (y_{k+1} - \mu_y) y_k^{-\beta} e^{-x_k/2} \]

and set \( F_k = \frac{\partial}{\partial x_k} f_k \bigg|_{x_k = \hat{x}_{k|k}} \).

The only modification of the filter occurs at the time-update recursions which, for \( k \geq 0 \), are given by

\[ \hat{x}_{k+1|k} = f_k(\hat{x}_{k|k}) \]
\[ P_{k+1|k} = F_k P_{k|k} F_k^T + G_k Q_{k+1} G_k' \]

where \( \hat{x}_{0|0} = \mu_0 \), and \( P_{0|0} = P_0 \).

The measurement update recursions remain the same as before. Thus, for \( k \geq 0 \), we have

\[ \hat{x}_{k+1|k+1} = \hat{x}_{k+1|k} + K_k (z_{k+1} - H_k \hat{x}_{k|k}) \]
\[ P_{k+1|k+1} = P_{k+1|k} - K_k F_k P_{k|k} H_k + G_k S_{k+1} \]
\[ K_k = (F_k P_{k|k} H_k + G_k S_{k+1})(H_k P_{k|k} H_k + R_{k+1})^{-1} \]

Joint and conditional distributions of \( \varepsilon_t^* \) and \( \eta_t \)

Recall that

\[ f_{\varepsilon_t, \eta_t}(x, y) = \frac{1}{2\pi \sqrt{1 - \rho^2}} \exp \left\{ -\frac{1}{2(1 - \rho^2)} \left( x^2 - 2\rho xy + y^2 \right) \right\} \]

We want to obtain the joint distribution of \( \varepsilon_t^* = \log(\varepsilon_t^2) \) and \( \eta_t \). Using the transformation theorem we have that

\[ f_{\varepsilon_t^*, \eta_t}(x, y) = \frac{1}{2\pi \sqrt{1 - \rho^2}} \exp \left\{ -\frac{1}{2(1 - \rho^2)} \left( e^{x} - 2\rho(e^{x/2}) y + y^2 \right) \right\} \left| \frac{1}{2} e^{x/2} \right| \]

\[ + \frac{1}{2\pi \sqrt{1 - \rho^2}} \exp \left\{ -\frac{1}{2(1 - \rho^2)} \left( e^{x} - 2\rho(-e^{x/2}) y + y^2 \right) \right\} \left| -\frac{1}{2} e^{x/2} \right| \]

\[ = \frac{1}{2\pi \sqrt{1 - \rho^2}} \exp \left\{ -\frac{1}{2(1 - \rho^2)} \left( e^{x} - (1 - \rho^2) x + y^2 \right) \right\} \]

\[ \times \frac{1}{2} \left( \exp \left\{ \frac{\rho}{1 - \rho^2} e^{x/2} y \right\} + \exp \left\{ -\frac{\rho}{1 - \rho^2} e^{x/2} y \right\} \right) \]

\[ = \frac{1}{2\pi \sqrt{1 - \rho^2}} \exp \left\{ -\frac{1}{2(1 - \rho^2)} \left( e^{x} - (1 - \rho^2) x + y^2 \right) \right\} \cosh \left( \frac{\rho}{1 - \rho^2} e^{x/2} y \right) \]
where $\cosh(z) = \sum_{k=0}^{\infty} \frac{z^{2k}}{(2k)!}$.

For the conditional distribution we simply use the fact that $f(y|x) = f(x,y)/f(y)$. Thus,

$$f_{\eta_\tau t}(y|x) = \frac{1}{\sqrt{2\pi(1-\rho^2)}} \exp\left\{ -\frac{1}{2(1-\rho^2)} \left( e^x - (1-\rho^2)x + y^2 \right) \right\}$$

$$\times \cosh\left( \frac{\rho}{1-\rho^2} e^{x/2} y \right) \exp\left\{ \frac{1}{2} (e^x - x) \right\}$$

$$= \frac{1}{\sqrt{2\pi(1-\rho^2)}} \exp\left\{ -\frac{1}{2(1-\rho^2)} \left( e^x - (1-\rho^2)x + y^2 - (1-\rho^2)e^x + (1-\rho^2)x \right) \right\}$$

$$\times \cosh\left( \frac{\rho}{1-\rho^2} e^{x/2} y \right)$$

$$= \frac{1}{\sqrt{2\pi(1-\rho^2)}} \exp\left\{ -\frac{1}{2(1-\rho^2)} \left( y^2 + \rho^2 e^x \right) \right\} \cosh\left( \frac{\rho}{1-\rho^2} e^{\frac{1}{2} x} y \right).$$

References


