

Estimation of Stochastic Volatility Models with Heavy Tails and Serial Dependence*

Joshua C.C. Chan
Research School of Economics
Centre for Applied Macroeconomic Analysis (CAMA)
Australian National University

Cody Y.L. Hsiao
Crawford School of Public Policy
Centre for Applied Macroeconomic Analysis (CAMA)
Australian National University

November 2013

Abstract

Financial time series often exhibit properties that depart from the usual assumptions of serial independence and normality. These include volatility clustering, heavy-tailedness and serial dependence. A voluminous literature on different approaches for modeling these empirical regularities has emerged in the last decade. In this paper we review the estimation of a variety of highly flexible stochastic volatility models, and introduce some efficient algorithms based on recent advances in state space simulation techniques. These estimation methods are illustrated via empirical examples involving precious metal and foreign exchange returns. The corresponding MATLAB code is also provided.

Keywords: stochastic volatility, scale mixture of normal, state space model, Markov chain Monte Carlo, financial data

JEL classification codes: C11, C22, C58

*This article was prepared for *Bayesian Inference in the Social Sciences*, edited by Ivan Jeliazkov and Xin-She Yang and to be published by John Wiley & Sons.

1 Introduction

Financial time series often exhibit properties that depart from the usual assumptions of serial independence and normality. These include volatility clustering, heavy-tailedness and serial dependence. A voluminous literature on different approaches for modeling these empirical regularities has emerged in the last decade. In this paper we review the estimation of a variety of highly flexible stochastic volatility models, and introduce some efficient algorithms based on recent advances in state space simulation techniques. These estimation methods are illustrated via empirical examples involving precious metal and foreign exchange returns. The corresponding MATLAB code is also provided.¹

The remaining of the paper is structured as follows. Section 2 first discusses the basic stochastic volatility model and its estimation. In particular, we provide details of the auxiliary mixture sampler and the precision sampler for linear Gaussian state space models. In Section 3 we extend the basic stochastic volatility model to allow for moving average errors. We then discuss an efficient estimation method based on fast band matrix routines.

Lastly, Section 4 considers another extension—instead of the conventional assumption of a Gaussian error distribution, we discuss some heavy-tailed distributions that can be written as scale mixtures of Gaussian distributions. We demonstrate the relevance of these heavy-tailed stochastic volatility models through an empirical example.

2 Stochastic Volatility Model

In financial time series, it is often observed that large changes in observations tend to be followed by large changes, while small changes are followed by small changes—the phenomenon that is referred to as **volatility clustering**. For instance, during financial crises movements in financial asset returns tend to be large (of either sign), whereas in “normal periods” the same asset returns might exhibit little time variation.

As an example, Figure 1 depicts the AUD/USD daily returns from January 2005 to December 2012. For a long stretch of time from early 2005 to mid-2007, the daily returns mostly fluctuate between $\pm 2\%$. However, during the global financial crisis of 2007-2008, the volatility of the daily returns increases dramatically—often reaching as high as $\pm 4\%$, sometimes even larger.

Models that assume constant variance, by definition, cannot accommodate time-varying volatility, and therefore cannot model volatility clustering that is a prominent feature in a wide variety of financial data. In this section we introduce a class of state space models that allow for this important feature. Extensions of these simple models are discussed in later sections.

¹MATLAB code is available at <http://people.anu.edu.au/joshua.chan>.

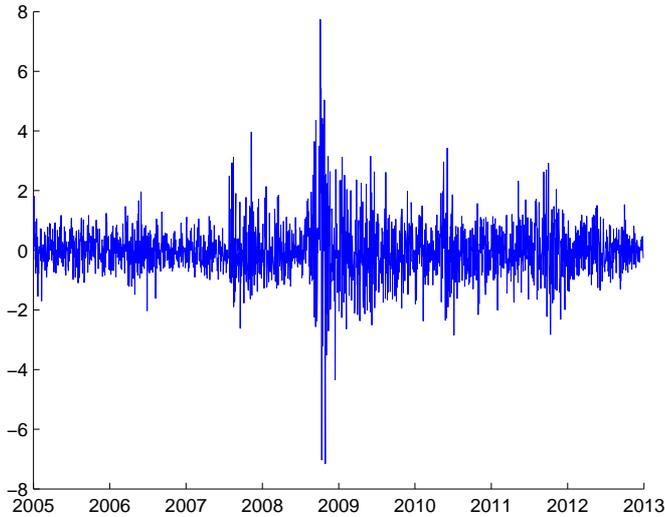


Figure 1: AUD/USD daily returns from January 2005 to December 2012.

We focus our discussion on modeling the variance of the time series, and for the moment assume that the observations have a zero mean; a constant mean or a suitable conditional mean process such as an $AR(p)$ component can be added later on, as is done in the empirical example in Section 2.3. Specifically, under the **stochastic volatility** model, the observation at time t is given by

$$y_t = e^{\frac{1}{2}h_t} \varepsilon_t \quad (1)$$

for $t = 1, \dots, T$, where $\varepsilon_t \sim \mathcal{N}(0, 1)$. Hence, the conditional variance of y_t is $Var(y_t | h_t) = e^{h_t}$, and the state h_t is often called the **log-volatility**. The states are assumed to evolve according to a stationary process

$$h_t = \mu_h + \phi_h(h_{t-1} - \mu_h) + \zeta_t \quad (2)$$

for $t = 2, \dots, T$, where $\zeta_t \sim \mathcal{N}(0, \sigma_h^2)$ and is independent of ε_t at all leads and lags. Here we assume that $|\phi_h| < 1$, and the states are initialized with $h_1 \sim \mathcal{N}(\mu_h, \sigma_h^2 / (1 - \phi_h^2))$, which is the stationary distribution of the process.

To complete the model specification, we assume independent prior distributions for μ_h , ϕ_h and σ_h^2 , i.e., $p(\mu_h, \phi_h, \sigma_h^2) = p(\mu_h)p(\phi_h)p(\sigma_h^2)$. In particular, consider the following independent prior distributions:

$$\mu_h \sim \mathcal{N}(\mu_{h0}, V_{\mu_h}), \quad \phi_h \sim \mathcal{N}(\phi_{h0}, V_{\phi_h})1(|\phi_h| < 1), \quad \sigma_h^2 \sim \mathcal{IG}(\nu_h, S_h), \quad (3)$$

where $1(\cdot)$ denotes the indicator function and $\mathcal{IG}(\cdot, \cdot)$ represents the inverse-gamma distribution. Note that we impose the stationarity condition $|\phi_h| < 1$ through the prior distribution of ϕ_h .

The stochastic volatility model is an example of a *nonlinear* state space model where the measurement equation (1) is nonlinear in the state. One major challenge of estimating this

nonlinear model is that the joint conditional density of the states $\mathbf{h} = (h_1, \dots, h_T)'$ given the model parameters and the data is high-dimensional and nonstandard—in contrast to a linear Gaussian state space model where the conditional density of the states is Gaussian. Consequently, Bayesian estimation using MCMC becomes more difficult.

2.1 Auxiliary Mixture Sampler

In this subsection we discuss a popular approach—the so-called **auxiliary mixture sampler** [13]—for estimating the stochastic volatility model (1)–(2). In a nutshell, the idea is to approximate the nonlinear stochastic volatility model using a mixture of linear Gaussian models, where the estimation of the latter models is standard. Specifically, we first transform the observation y_t so that the measurement equation becomes linear in the log-volatility h_t . More precisely, we square both sides of (1) and take the logarithm:

$$y_t^* = h_t + \varepsilon_t^*, \quad (4)$$

where $y_t^* = \log y_t^2$ and $\varepsilon_t^* = \log \varepsilon_t^2$. In practice, we often set $y_t^* = \log(y_t^2 + c)$ for some small constant c , say, $c = 10^{-4}$, to avoid numerical problems when y_t is close to zero. Then, (2) and (4) define a linear state space model in h_t . Note, however, that the error ε_t^* no longer has a Gaussian distribution—in fact, it follows a $\log\text{-}\chi_1^2$ distribution—and the machinery for fitting linear Gaussian state space models cannot be directly applied.

To tackle this difficulty, the second step of the auxiliary mixture sampling approach is to obtain an appropriate Gaussian mixture that well approximates $f(\varepsilon_t^*)$, the density function of ε_t^* . More precisely,

$$f(\varepsilon_t^*) \approx \sum_{i=1}^n p_i \varphi(\varepsilon_t^*; \mu_i, \sigma_i^2), \quad (5)$$

where $\varphi(x; \mu, \sigma^2)$ denotes the Gaussian density with mean μ and variance σ^2 , p_i is the probability of the i -th mixture component, and n is the number of components.

We can equivalently write the mixture density in (5) in terms of an auxiliary random variable $s_t \in \{1, \dots, n\}$ that serves as the mixture component indicator (hence, the name of the approach):

$$(\varepsilon_t^* | s_t = i) \sim \mathcal{N}(\mu_i, \sigma_i^2), \quad (6)$$

$$P(s_t = i) = p_i. \quad (7)$$

Using this representation, we have a linear Gaussian model conditional on the component indicator s_t and the simulation techniques for estimating such a model can be applied.

The only missing piece is a suitable Gaussian mixture. By matching the moments of the $\log\text{-}\chi_1^2$ distribution, [13] propose a seven-component Gaussian mixture

$$f(x) = \sum_{i=1}^7 p_i \varphi(x; \mu_i - 1.2704, \sigma_i^2),$$

Table 1: A seven-component Gaussian mixture for approximating the $\log\text{-}\chi_1^2$ distribution.

| component | p_i | μ_i | σ_i^2 |
|-----------|---------|-----------|--------------|
| 1 | 0.00730 | -10.12999 | 5.79596 |
| 2 | 0.10556 | -3.97281 | 2.61369 |
| 3 | 0.00002 | -8.56686 | 5.17950 |
| 4 | 0.04395 | 2.77786 | 0.16735 |
| 5 | 0.34001 | 0.61942 | 0.64009 |
| 6 | 0.24566 | 1.79518 | 0.34023 |
| 7 | 0.25750 | -1.08819 | 1.26261 |

where the values of the parameters are given in Table 1. We emphasize that these values are fixed and do not depend on any unknown parameters. Hence, this approximation does not require any additional computation time in the estimation.

In summary, using the Gaussian mixture approximation in (6) and (7), the model (2) and (4) is now conditionally linear Gaussian given the component indicators $\mathbf{s} = (s_1, \dots, s_T)'$. For later reference, let $\mathbf{y}^* = (y_1^*, \dots, y_T^*)'$, and similarly define \mathbf{h} , $\boldsymbol{\zeta}$ and $\boldsymbol{\varepsilon}^*$. Bayesian analysis can be performed using a sample from the joint posterior distribution $p(\mathbf{h}, \mathbf{s}, \mu_h, \phi_h, \sigma_h^2 | \mathbf{y})$. Posterior draws can be obtained via a Gibbs sampler that cycles through

1. $p(\mathbf{h} | \mathbf{y}^*, \mathbf{s}, \mu_h, \phi_h, \sigma_h^2)$;
2. $p(\mathbf{s} | \mathbf{y}^*, \mathbf{h}, \mu_h, \phi_h, \sigma_h^2) = p(\mathbf{s} | \mathbf{y}^*, \mathbf{h})$;
3. $p(\mu_h | \mathbf{y}, \mathbf{h}, \mathbf{s}, \phi_h, \sigma_h^2) = p(\mu_h | \mathbf{h}, \phi_h, \sigma_h^2)$;
4. $p(\phi_h | \mathbf{y}, \mathbf{h}, \mathbf{s}, \mu_h, \sigma_h^2) = p(\phi_h | \mathbf{h}, \mu_h, \sigma_h^2)$;
5. $p(\sigma_h^2 | \mathbf{y}, \mathbf{h}, \mathbf{s}, \mu_h, \phi_h) = p(\sigma_h^2 | \mathbf{h}, \mu_h, \phi_h)$.

These MCMC draws are from the approximate model, and one can reweight the draws using importance sampling weights to get the exact posterior moments under the original model. However, this step is often skipped in practice as reweighting makes little difference; see also the discussion in [13].

2.2 Precision Sampler for Linear Gaussian State Space Models

Given the prior distributions in (3), the conditionally linear Gaussian model in (2), (4), (6) and (7) can be estimated using standard MCMC techniques. In particular, one can simulate from the joint distribution of the log-volatilities $p(\mathbf{h} | \mathbf{y}^*, \mathbf{s}, \mu_h, \phi_h, \sigma_h^2)$ using Kalman filter-based algorithms such as those discussed in [2, 9, 7, 8].

Instead, here we apply a new algorithm that exploits the special structure of the model. Specifically, we will show below that the precision matrix—inverse of the covariance

matrix—of $p(\mathbf{h} | \mathbf{y}^*, \mathbf{s}, \mu_h, \phi_h, \sigma_h^2)$ is a band matrix. That is, it contains only a small number of nonzero elements along a diagonal band. As such, computations involving band matrices can be done much faster compared to those involving full matrices.

For instance, obtaining the Cholesky decomposition of a band $T \times T$ matrix with fixed bandwidth involves only $\mathcal{O}(T)$ operations (see, e.g., [11], pp. 156) as opposed to $\mathcal{O}(T^3)$ for a full matrix of the same size. A similar reduction in computations holds for other operations such as multiplication and forward-backward substitution for solving linear systems. These band matrix algorithms are implemented in standard packages such as MATLAB, GAUSS and R.

The idea of exploiting band precision matrices to speed up computations can be traced back to [20], who discusses simulation of Gaussian Markov random fields; [4] and [17] propose similar algorithms for linear Gaussian state space models, which are later used in various applications in, e.g., [1, 5, 6, 12].

We first derive the joint distribution of the states $p(\mathbf{h} | \mathbf{y}^*, \mathbf{s}, \mu_h, \phi_h, \sigma_h^2)$. To that end, we rewrite (4) in matrix notation:

$$\mathbf{y}^* = \mathbf{h} + \boldsymbol{\varepsilon}^*,$$

where

$$(\boldsymbol{\varepsilon}^* | \mathbf{s}) \sim \mathcal{N}(\mathbf{d}, \Sigma_{\mathbf{y}^*}),$$

with $\mathbf{d} = (\mu_{s_1} - 1.2704, \dots, \mu_{s_T} - 1.2704)'$, $\Sigma_{\mathbf{y}^*} = \text{diag}(\sigma_{s_1}^2, \dots, \sigma_{s_T}^2)$, and the fixed parameters μ_1, \dots, μ_7 and $\sigma_1^2, \dots, \sigma_7^2$ are given in Table 1. By a simple change of variable, we have $(\mathbf{y}^* | \mathbf{s}, \mathbf{h}) \sim \mathcal{N}(\mathbf{h} + \mathbf{d}, \Sigma_{\mathbf{y}^*})$. It follows that

$$\log p(\mathbf{y}^* | \mathbf{s}, \mathbf{h}) = -\frac{1}{2}(\mathbf{y}^* - \mathbf{h} - \mathbf{d})' \Sigma_{\mathbf{y}^*}^{-1} (\mathbf{y}^* - \mathbf{h} - \mathbf{d}) + c_1, \quad (8)$$

where c_1 is a normalization constant.

Next, rewrite the state equation (2) in matrix form:

$$\mathbf{H}_{\phi_h} \mathbf{h} = \tilde{\boldsymbol{\alpha}} + \boldsymbol{\zeta},$$

where $\boldsymbol{\zeta} \sim \mathcal{N}(\mathbf{0}, \Sigma_{\mathbf{h}})$, $\Sigma_{\mathbf{h}} = \text{diag}(\sigma_h^2/(1 - \phi_h^2), \sigma_h^2, \dots, \sigma_h^2)$, and

$$\tilde{\boldsymbol{\alpha}} = \begin{pmatrix} \mu_h \\ (1 - \phi_h)\mu_h \\ \vdots \\ (1 - \phi_h)\mu_h \end{pmatrix}, \quad \mathbf{H}_{\phi_h} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ -\phi_h & 1 & 0 & \cdots & 0 \\ 0 & -\phi_h & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & -\phi_h & 1 \end{pmatrix}.$$

Here \mathbf{H}_{ϕ_h} is a band matrix with determinant $|\mathbf{H}_{\phi_h}| = 1$ for all values of ϕ_h . By a change of variable, we have

$$(\mathbf{h} | \mu_h, \phi_h, \sigma_h^2) \sim \mathcal{N}(\boldsymbol{\alpha}, (\mathbf{H}'_{\phi_h} \Sigma_{\mathbf{h}}^{-1} \mathbf{H}_{\phi_h})^{-1}),$$

where $\boldsymbol{\alpha} = \mathbf{H}_{\phi_h}^{-1} \tilde{\boldsymbol{\alpha}}$. Noting that $|(\mathbf{H}'_{\phi_h} \Sigma_{\mathbf{h}}^{-1} \mathbf{H}_{\phi_h})^{-1}| = |\Sigma_{\mathbf{h}}| = (\sigma_h^2)^T / (1 - \phi_h^2)$, we have

$$\log p(\mathbf{h} | \mu_h, \phi_h, \sigma_h^2) = -\frac{1}{2} \log \left(\frac{(2\pi\sigma_h^2)^T}{1 - \phi_h^2} \right) - \frac{1}{2} (\mathbf{h} - \boldsymbol{\alpha})' \mathbf{H}'_{\phi_h} \Sigma_{\mathbf{h}}^{-1} \mathbf{H}_{\phi_h} (\mathbf{h} - \boldsymbol{\alpha}). \quad (9)$$

Notice that in this case the precision matrix $\mathbf{H}'_{\phi_h} \Sigma_{\mathbf{h}}^{-1} \mathbf{H}_{\phi_h}$ is also band; in fact, it is tridiagonal.

Since $p(\mathbf{h} | \mathbf{y}^*, \mathbf{s}, \mu_h, \phi_h, \sigma_h^2) \propto p(\mathbf{y}^* | \mathbf{s}, \mathbf{h}) p(\mathbf{h} | \mu_h, \phi_h, \sigma_h^2)$, using the expressions in (8) and (9), we have

$$\begin{aligned} & \log p(\mathbf{h} | \mathbf{y}^*, \mathbf{s}, \mu_h, \phi_h, \sigma_h^2) \\ &= -\frac{1}{2} (\mathbf{y}^* - \mathbf{h} - \mathbf{d})' \Sigma_{\mathbf{y}^*}^{-1} (\mathbf{y}^* - \mathbf{h} - \mathbf{d}) - \frac{1}{2} (\mathbf{h} - \boldsymbol{\alpha})' \mathbf{H}'_{\phi_h} \Sigma_{\mathbf{h}}^{-1} \mathbf{H}_{\phi_h} (\mathbf{h} - \boldsymbol{\alpha}) + c_2 \\ &= -\frac{1}{2} [\mathbf{h}' (\mathbf{H}'_{\phi_h} \Sigma_{\mathbf{h}}^{-1} \mathbf{H}_{\phi_h} + \Sigma_{\mathbf{y}^*}^{-1}) \mathbf{h} \\ & \quad - 2\mathbf{h}' (\mathbf{H}'_{\phi_h} \Sigma_{\mathbf{h}}^{-1} \mathbf{H}_{\phi_h} \boldsymbol{\alpha} + \Sigma_{\mathbf{y}^*}^{-1} (\mathbf{y}^* - \mathbf{d}))] + c_3, \end{aligned} \quad (10)$$

where c_2 and c_3 are constants not depending on \mathbf{h} . Since this log-density is quadratic in \mathbf{h} , it is Gaussian and we have

$$(\mathbf{h} | \mathbf{y}^*, \mathbf{s}, \mu_h, \phi_h, \sigma_h^2) \sim \mathcal{N}(\hat{\mathbf{h}}, \mathbf{K}_{\mathbf{h}}^{-1})$$

for some mean vector $\hat{\mathbf{h}}$ and precision matrix $\mathbf{K}_{\mathbf{h}}$. To determine $\hat{\mathbf{h}}$ and $\mathbf{K}_{\mathbf{h}}$, compare (10) with the log-density

$$\log p(\mathbf{h} | \mathbf{y}^*, \mathbf{s}, \mu_h, \phi_h, \sigma_h^2) = -\frac{1}{2} (\mathbf{h}' \mathbf{K}_{\mathbf{h}} \mathbf{h} - 2\mathbf{h}' \mathbf{K}_{\mathbf{h}} \hat{\mathbf{h}}) + c_4$$

for some normalization constant c_4 . Hence, we have

$$\mathbf{K}_{\mathbf{h}} = \mathbf{H}'_{\phi_h} \Sigma_{\mathbf{h}}^{-1} \mathbf{H}_{\phi_h} + \Sigma_{\mathbf{y}^*}^{-1}, \quad \hat{\mathbf{h}} = \mathbf{K}_{\mathbf{h}}^{-1} (\mathbf{H}'_{\phi_h} \Sigma_{\mathbf{h}}^{-1} \mathbf{H}_{\phi_h} \boldsymbol{\alpha} + \Sigma_{\mathbf{y}^*}^{-1} (\mathbf{y}^* - \mathbf{d})).$$

Next, we discuss an efficient method to sample from the high-dimensional Gaussian distribution $\mathcal{N}(\hat{\mathbf{h}}, \mathbf{K}_{\mathbf{h}}^{-1})$ by exploiting the special structure of the precision matrix $\mathbf{K}_{\mathbf{h}}$. More specifically, since $\mathbf{K}_{\mathbf{h}}$ is a band matrix, $\hat{\mathbf{h}}$ can be obtained quickly by solving the linear system

$$\mathbf{K}_{\mathbf{h}} \mathbf{x} = \mathbf{H}'_{\phi_h} \Sigma_{\mathbf{h}}^{-1} \mathbf{H}_{\phi_h} \boldsymbol{\alpha} + \Sigma_{\mathbf{y}^*}^{-1} (\mathbf{y}^* - \mathbf{d})$$

for \mathbf{x} , which avoids computing the inverse $\mathbf{K}_{\mathbf{h}}^{-1}$.

To get a draw from $\mathcal{N}(\hat{\mathbf{h}}, \mathbf{K}_{\mathbf{h}}^{-1})$, we first obtain the Cholesky factor $\mathbf{C}_{\mathbf{h}}$ of $\mathbf{K}_{\mathbf{h}}$, i.e. $\mathbf{K}_{\mathbf{h}} = \mathbf{C}_{\mathbf{h}} \mathbf{C}'_{\mathbf{h}}$. Again, this operation can be done quickly as $\mathbf{K}_{\mathbf{h}}$ is a band matrix. Next, we obtain T independent $\mathcal{N}(0, 1)$ draws $\mathbf{U} = (U_1, \dots, U_T)'$, and return $\mathbf{V} = \hat{\mathbf{h}} + (\mathbf{C}'_{\mathbf{h}})^{-1} \mathbf{U}$. It is easy to check that the mean vector of \mathbf{V} is $\hat{\mathbf{h}}$ and its covariance matrix is

$$(\mathbf{C}'_{\mathbf{h}})^{-1} ((\mathbf{C}'_{\mathbf{h}})^{-1})' = (\mathbf{C}'_{\mathbf{h}})^{-1} (\mathbf{C}_{\mathbf{h}})^{-1} = (\mathbf{C}_{\mathbf{h}} \mathbf{C}'_{\mathbf{h}})^{-1} = \mathbf{K}_{\mathbf{h}}^{-1}.$$

We note that once again $(\mathbf{C}'_{\mathbf{h}})^{-1}\mathbf{U}$ can be computed quickly by solving a suitable linear system without computing the inverse $(\mathbf{C}'_{\mathbf{h}})^{-1}$. This completes Step 1 of the Gibbs sampler for the stochastic volatility model.

To implement Step 2, note that $p(\mathbf{s} | \mathbf{y}^*, \mathbf{h}) = \prod_{t=1}^T p(s_t | y_t^*, h_t)$, and therefore we can sample each s_t independently conditional on \mathbf{y}^* and \mathbf{h} . Since s_t is a discrete random variable that follows a seven-point distribution, it can be easily sampled via the inverse-transform method (see, e.g., Algorithm 3.2 in [16]), provided we can compute $P(s_t = i | y_t^*, h_t)$ for $i = 1, \dots, 7$. In fact, we have

$$P(s_t = i | y_t^*, h_t) = \frac{1}{c_t} p_i \varphi(y_t^*; h_t + \mu_i - 1.2704, \sigma_i^2),$$

where $c_t = \sum_{j=1}^7 p_j \varphi(y_t^*; h_t + \mu_j - 1.2704, \sigma_j^2)$ is a normalization constant.

Steps 3 and 5 can be done easily, as both conditional distributions are standard. In particular,

$$(\mu_h | \mathbf{h}, \phi_h, \sigma_h^2) \sim \mathcal{N}(\hat{\mu}_h, D_{\mu_h}),$$

where $D_{\mu_h} = (V_{\mu_h}^{-1} + \mathbf{X}'_{\mu_h} \Sigma_{\mathbf{h}}^{-1} \mathbf{X}_{\mu_h})^{-1}$ and $\hat{\mu}_h = D_{\mu_h} (V_{\mu_h}^{-1} \mu_{h0} + \mathbf{X}'_{\mu_h} \Sigma_{\mathbf{h}}^{-1} \mathbf{z}_{\mu_h})$ with $\mathbf{X}_{\mu_h} = (1, 1 - \phi_h, \dots, 1 - \phi_h)'$, $\mathbf{z}_{\mu_h} = (h_1, h_2 - \phi_h h_1, \dots, h_T - \phi_h h_{T-1})'$ and $\Sigma_{\mathbf{h}} = \text{diag}(\sigma_h^2 / (1 - \phi_h^2), \sigma_h^2, \dots, \sigma_h^2)$. In addition, we have

$$(\sigma_h^2 | \mathbf{h}, \mu_h, \phi_h) \sim \mathcal{IG}(\nu_h + T/2, \tilde{S}_h),$$

where $\tilde{S}_h = S_h + [(1 - \phi_h^2)(h_1 - \mu_h)^2 + \sum_{t=2}^T (h_t - \mu_h - \phi_h(h_{t-1} - \mu_h))^2] / 2$.

Finally, it follows from (2) and (3) that

$$p(\phi_h | \mathbf{h}, \mu_h, \sigma_h^2) \propto p(\phi_h) g(\phi_h) e^{-\frac{1}{2\sigma_h^2} \sum_{t=2}^T (h_t - \mu_h - \phi_h(h_{t-1} - \mu_h))^2},$$

where $g(\phi_h) = (1 - \phi_h^2)^{1/2} \exp(-\frac{1}{2\sigma_h^2} (1 - \phi_h^2)(h_1 - \mu_h)^2)$ and $p(\phi_h)$ is the truncated normal prior given in (3). The conditional density $p(\phi_h | \mathbf{h}, \mu_h, \sigma_h^2)$ is nonstandard, and we implement an independence-chain Metropolis-Hastings step with proposal distribution $\mathcal{N}(\hat{\phi}_h, D_{\phi_h}) 1(|\phi_h| < 1)$, where $D_{\phi_h} = (V_{\phi_h}^{-1} + \mathbf{X}'_{\phi_h} \mathbf{X}_{\phi_h} / \sigma_h^2)^{-1}$ and $\hat{\phi}_h = D_{\phi_h} (V_{\phi_h}^{-1} \phi_{h0} + \mathbf{X}'_{\phi_h} \mathbf{z}_{\phi_h} / \sigma_h^2)$, with $\mathbf{X}_{\phi_h} = (h_1 - \mu_h, \dots, h_{T-1} - \mu_h)'$ and $\mathbf{z}_{\phi_h} = (h_2 - \mu_h, \dots, h_T - \mu_h)'$. Then, given the current draw ϕ_h , a proposal ϕ_h^* is accepted with probability $\min(1, g(\phi_h^*) / g(\phi_h))$; otherwise the Markov chain stays at the current state ϕ_h .

2.3 Empirical Example: Modeling AUD/USD Returns

We illustrate the estimation of the stochastic volatility model using the data depicted in Figure 1—AUD/USD daily returns from January 2005 to December 2012. More specifically, we allow for a constant conditional mean in the measurement equation

$$y_t = \mu + e^{\frac{1}{2} h_t} \varepsilon_t$$

for $t = 1, \dots, T$, where $\varepsilon_t \sim \mathcal{N}(0, 1)$. The state equation is again given in (2).

We assume the same independent prior distributions for μ_h, ϕ_h and σ_h^2 as in (3). In particular, we set $\mu_{h0} = 0, V_{\mu_h} = 5, \phi_{h0} = 0.95, V_{\phi_h} = 1, \nu_h = 10$ and $S_h = 0.19$. These values imply relatively diffuse priors with prior means $E(\mu_h) = 0, E(\phi_h) = 0.95$ and $E(\sigma_h^2) = 0.02$. For μ we assume the normal prior distribution $\mathcal{N}(\mu_0, V_\mu)$, where $\mu_0 = 0$ and $V_\mu = 5$. Again this prior is relatively diffuse with prior mean $E(\mu) = 0$.

To estimate the model with the constant conditional mean μ , we simply need an extra block to sample from the conditional distribution $p(\mu | \mathbf{y}, \mathbf{h})$, and modify the Gibbs sampler discussed in Section 2.2 by replacing y_t with $y_t - \mu$. In particular, it can be checked that

$$(\mu | \mathbf{y}, \mathbf{h}) \sim \mathcal{N}(\hat{\mu}, D_\mu),$$

where $D_\mu = (V_\mu^{-1} + \sum_{t=1}^T e^{-h_t})^{-1}$ and $\hat{\mu} = D_\mu(V_\mu^{-1}\mu_0 + \sum_{t=1}^T e^{-h_t}y_t)$.

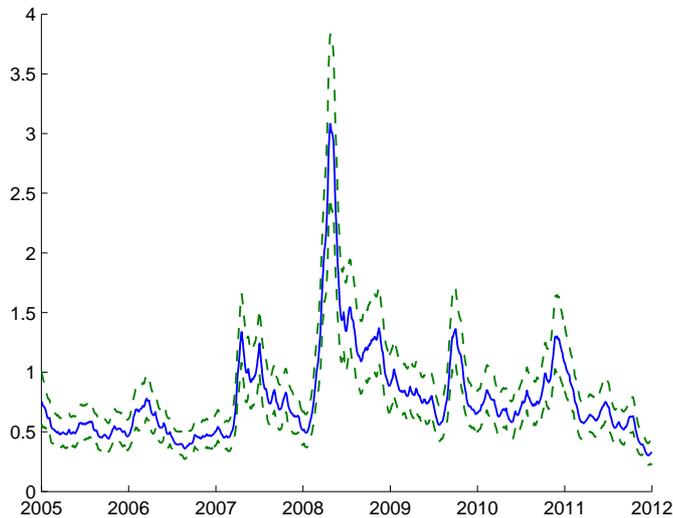


Figure 2: Posterior means (solid line) and 90% credible intervals (dash lines) of the time-varying standard deviation $\exp(h_t/2)$; AUD/USD daily returns data.

We obtain 20000 draws from the posterior distribution using the Gibbs sampler outlined above, after a burn-in period of 1000. Figure 2 depicts the posterior means and quantiles of the time-varying standard deviation $\exp(h_t/2)$. As the figure shows, there is substantial time-variation in the volatility. In particular, between 2005 and early 2007, the estimated standard deviation mostly fluctuates around 0.5%. It increases to about 1% in mid-2007 and peaks at 3% during the global financial crisis. Although it goes down substantially after 2009, it is still much higher than the pre-crisis level.

The posterior means, standard deviations and quantiles of the model parameters are reported in Table 2. For instance, over the period from January 2005 to December 2012,

Table 2: Posterior means, standard deviations and quantiles of model parameters; AUD/USD daily returns data.

| parameter | posterior mean | posterior std. dev. | 5%-tile | 95%-tile |
|--------------|----------------|---------------------|---------|----------|
| μ | -0.029 | 0.013 | -0.051 | -0.006 |
| μ_h | -0.748 | 0.351 | -1.275 | -0.229 |
| ϕ_h | 0.989 | 0.004 | 0.982 | 0.995 |
| σ_h^2 | 0.017 | 0.003 | 0.012 | 0.023 |

the average daily return is estimated to be -0.029% . The posterior mean of the AR(1) coefficient of the state equation is 0.989, indicating a very high level of persistence.

3 Moving Average Stochastic Volatility Model

We now discuss an extension of the plain vanilla stochastic volatility model introduced in Section 2. Specifically, the model (1)–(2) assumes the errors in the measurement equation are serially independent given the log-volatilities. This is often an appropriate assumption for modeling financial data if we expect that the market is in some sense efficient. Of course, this assumption can be tested, for example, by building a model that allows for persistence via moving average errors.

To that end, we consider a variant of the moving average stochastic volatility models proposed in [3]. More precisely, we expand the stochastic volatility model (1)–(2) by allowing the errors in the measurement equation to follow an MA(q) process:

$$y_t = \mu + u_t, \quad (11)$$

$$u_t = \varepsilon_t + \psi_1 \varepsilon_{t-1} + \dots + \psi_q \varepsilon_{t-q}, \quad (12)$$

$$h_t = \mu_h + \phi_h (h_{t-1} - \mu_h) + \zeta_t, \quad (13)$$

where $\varepsilon_t \sim \mathcal{N}(0, e^{h_t})$ and $\zeta_t \sim \mathcal{N}(0, \sigma_h^2)$ are independent of each other, and $\varepsilon_0 = \varepsilon_{-1} = \dots = \varepsilon_{-q+1} = 0$. For identification, we impose the invertibility conditions—the roots of the characteristic polynomial associated with the MA coefficients $\boldsymbol{\psi} = (\psi_1, \dots, \psi_q)'$ are all outside the unit circle. As before, we assume that $|\phi_h| < 1$ and the states are initialized with $h_1 \sim \mathcal{N}(\mu_h, \sigma_h^2 / (1 - \phi_h^2))$. It is obvious that if $\psi_1 = \dots = \psi_q = 0$, the model (11)–(13) reduces to the standard stochastic volatility model discussed in Section 2.

Recall that under the standard stochastic volatility model (1)–(2), the conditional variance of the observation y_t is simply e^{h_t} . Here, under the moving average variant, the conditional variance of y_t is given by

$$\text{Var}(y_t | \mu, \boldsymbol{\psi}, \mathbf{h}) = e^{h_t} + \psi_1^2 e^{h_{t-1}} + \dots + \psi_q^2 e^{h_{t-q}}.$$

That is, the conditional variance is time-varying through two channels: first, it is a moving average of the $q + 1$ most recent variances $e^{h_t}, \dots, e^{h_{t-q}}$; and second, the log-volatilities also evolve according to the stationary AR(1) process in (13).

Compared to the standard stochastic volatility model, another difference is that y_t is no longer serially independent (even after conditioning on the log-volatilities). In fact, its conditional autocovariances are given by

$$\text{Cov}(y_t, y_{t-j} \mid \mu, \boldsymbol{\psi}, \mathbf{h}) = \sum_{i=0}^{q-j} \psi_{i+j} \psi_i e^{h_{t-i}}$$

for $j = 1, \dots, q$ and 0 for $j > q$, where $\psi_0 = 1$. Due to the presence of the time-varying log-volatilities, the autocovariances of y_t are also time-varying.

To complete the model specification, we assume independent prior distributions for μ , $\boldsymbol{\psi}$, μ_h , ϕ_h and σ_h^2 , i.e., $p(\mu, \boldsymbol{\psi}, \mu_h, \phi_h, \sigma_h^2) = p(\mu) p(\boldsymbol{\psi}) p(\mu_h) p(\phi_h) p(\sigma_h^2)$. For $\boldsymbol{\psi}$, we assume a multivariate normal prior with support in the region where the invertibility conditions on $\boldsymbol{\psi}$ hold. For other model parameters, we assume the following independent prior distributions: $\mu \sim \mathcal{N}(\mu_0, V_\mu)$, $\mu_h \sim \mathcal{N}(\mu_{h0}, V_{\mu_h})$, $\phi_h \sim \mathcal{N}(\phi_{h0}, V_{\phi_h}) 1(|\phi_h| < 1)$ and $\sigma_h^2 \sim \mathcal{IG}(\nu_h, S_h)$.

The next section discusses an efficient algorithm to estimate this moving average model that exploits the band structure of covariance matrix of \mathbf{y} . Then in Section 3.2 we illustrate the algorithm and the relevance of the extension through an empirical example that involves Philippine peso returns during the global financial crisis.

3.1 Estimation

We consider an efficient sampler—that builds upon band matrix algorithms—for estimating the moving average stochastic volatility model in (11)–(13). A key ingredient of this sampler is a quick way to evaluate the likelihood function that exploits the band structure of the covariance matrix of \mathbf{y} —instead of using conventional methods based on the Kalman filter (see, e.g., [10]).

To obtain the likelihood function, we derive the joint distribution of the observations \mathbf{y} . First rewrite (12) as:

$$\mathbf{u} = \mathbf{H}_\psi \boldsymbol{\varepsilon}, \tag{14}$$

where \mathbf{H}_ψ is a $T \times T$ lower triangular matrix with ones on the main diagonal, ψ_1 on first lower diagonal, ψ_2 on second lower diagonal, and so forth. For example, for $q = 2$, we have

$$\mathbf{H}_\psi = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ \psi_1 & 1 & 0 & 0 & \cdots & 0 \\ \psi_2 & \psi_1 & 1 & 0 & \cdots & 0 \\ 0 & \psi_2 & \psi_1 & 1 & \cdots & 0 \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \psi_2 & \psi_1 & 1 \end{pmatrix}.$$

Since $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{S}_y)$ with $\mathbf{S}_y = \text{diag}(e^{h_1}, \dots, e^{h_T})$, we have $\mathbf{u} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_y)$, where $\boldsymbol{\Sigma}_y =$

$\mathbf{H}_\psi \mathbf{S}_y \mathbf{H}'_\psi$. It follows from (11) and (14) that

$$(\mathbf{y} \mid \mu, \boldsymbol{\psi}, \mathbf{h}) \sim \mathcal{N}(\mu \mathbf{1}, \boldsymbol{\Sigma}_y),$$

where $\mathbf{1}$ is a $T \times 1$ column of ones.

Note that even though \mathbf{H}_ψ is a $T \times T$ matrix that contains T^2 elements, only $(T - q/2)(q + 1) < T(q + 1)$ of those elements are nonzero. In addition, since \mathbf{S}_y is diagonal, the covariance matrix $\boldsymbol{\Sigma}_y$ is a band matrix. This band structure is again exploited to speed up computations by using band matrix routines.

Furthermore, noting that $|\mathbf{H}_\psi| = 1$ for any $\boldsymbol{\psi}$, we have $|\boldsymbol{\Sigma}_y| = |\mathbf{S}_y| = \exp\left(\sum_{t=1}^T h_t\right)$. The log-density of \mathbf{y} is therefore given by

$$\log p(\mathbf{y} \mid \mu, \boldsymbol{\psi}, \mathbf{h}) = -\frac{T}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^T h_t - \frac{1}{2} (\mathbf{y} - \mu \mathbf{1})' \boldsymbol{\Sigma}_y^{-1} (\mathbf{y} - \mu \mathbf{1}). \quad (15)$$

Once again, the product $\boldsymbol{\Sigma}_y^{-1} (\mathbf{y} - \mu \mathbf{1})$ can be obtained by solving a linear system of equations without computing the inverse of $\boldsymbol{\Sigma}_y^{-1}$. Therefore, the log-likelihood function (15) can be evaluated quickly.

Next, posterior draws can be obtained by sequentially sampling from:

1. $p(\mu \mid \mathbf{y}, \mathbf{h}, \boldsymbol{\psi}, \mu_h, \phi_h, \sigma_h^2) = p(\mu \mid \mathbf{y}, \mathbf{h}, \boldsymbol{\psi});$
2. $p(\mathbf{h} \mid \mathbf{y}, \mu, \boldsymbol{\psi}, \mu_h, \phi_h, \sigma_h^2);$
3. $p(\boldsymbol{\psi} \mid \mathbf{y}, \mathbf{h}, \mu, \mu_h, \phi_h, \sigma_h^2) = p(\boldsymbol{\psi} \mid \mathbf{y}, \mathbf{h}, \mu);$
4. $p(\mu_h \mid \mathbf{y}, \mathbf{h}, \mu, \boldsymbol{\psi}, \phi_h, \sigma_h^2) = p(\mu_h \mid \mathbf{h}, \phi_h, \sigma_h^2);$
5. $p(\phi_h \mid \mathbf{y}, \mathbf{h}, \mu, \boldsymbol{\psi}, \mu_h, \sigma_h^2) = p(\phi_h \mid \mathbf{h}, \mu_h, \sigma_h^2);$
6. $p(\sigma_h^2 \mid \mathbf{y}, \mathbf{h}, \mu, \boldsymbol{\psi}, \mu_h, \phi_h) = p(\sigma_h^2 \mid \mathbf{h}, \mu_h, \phi_h).$

Step 1 can be easily done by drawing from

$$(\mu \mid \mathbf{y}, \mathbf{h}, \boldsymbol{\psi}) \sim \mathcal{N}(\hat{\mu}, D_\mu),$$

where $D_\mu = (V_\mu^{-1} + \mathbf{1}' \boldsymbol{\Sigma}_y^{-1} \mathbf{1})$ and $\hat{\mu} = D_\mu (V_\mu^{-1} \mu_0 + \mathbf{1}' \boldsymbol{\Sigma}_y^{-1} \mathbf{y})$. For Step 2, first note that if we define $\tilde{\mathbf{y}} = \mathbf{H}_\psi^{-1} (\mathbf{y} - \mu \mathbf{1})$, then

$$(\tilde{\mathbf{y}} \mid \mu, \boldsymbol{\psi}, \mathbf{h}) \sim \mathcal{N}(\mathbf{0}, \mathbf{S}_y)$$

with $\mathbf{S}_y = \text{diag}(e^{h_1}, \dots, e^{h_T})$. Hence, the auxiliary mixture sampler discussed in Section 2.1 can be directly applied to sample from $p(\mathbf{h} \mid \mathbf{y}, \mu, \boldsymbol{\psi}, \mu_h, \phi_h, \sigma_h^2)$.

To implement Step 3, observe that given the prior $p(\boldsymbol{\psi})$, the conditional posterior density is given by $p(\boldsymbol{\psi} | \mathbf{y}, \boldsymbol{\mu}, \mathbf{h}) \propto p(\boldsymbol{\psi}) p(\mathbf{y} | \boldsymbol{\mu}, \boldsymbol{\psi}, \mathbf{h})$, where both densities on the right-hand side can be evaluated quickly at any $\boldsymbol{\psi}$. Here we maximize $\log p(\boldsymbol{\psi} | \mathbf{y}, \boldsymbol{\mu}, \mathbf{h})$ numerically and obtain the mode and the negative Hessian evaluated at the mode, denoted as $\hat{\boldsymbol{\psi}}$ and $\mathbf{K}_{\boldsymbol{\psi}}$, respectively. Then, draws from $p(\boldsymbol{\psi} | \mathbf{y}, \boldsymbol{\mu}, \mathbf{h})$ can be obtained using an independence-chain Metropolis-Hastings step with proposal density $\mathcal{N}(\hat{\boldsymbol{\psi}}, \mathbf{K}_{\boldsymbol{\psi}}^{-1})$.

Finally, we can sample from the densities $p(\mu_h | \mathbf{h}, \sigma_h^2, \phi_h)$, $p(\phi_h | \mathbf{h}, \mu_h, \sigma_h^2)$ and $p(\sigma_h^2 | \mathbf{h}, \mu_h, \phi_h)$ in the same way as discussed in Section 2.2.

3.2 Empirical Example: Modeling PHP/USD Returns during Crisis

Figure 3 depicts the daily returns of the Philippine peso against the US dollar from the beginning of the global financial crisis, specifically from July 2007 to December 2012. We fit the data using the moving average stochastic volatility model in (11)–(13). In particular, we consider the MA(1) process $u_t = \varepsilon_t + \psi\varepsilon_{t-1}$. Recall that this moving average variant reduces to the standard stochastic volatility model when $\psi = 0$. Hence, one main focus is to investigate if the posterior density of ψ has mass concentrated around zero.

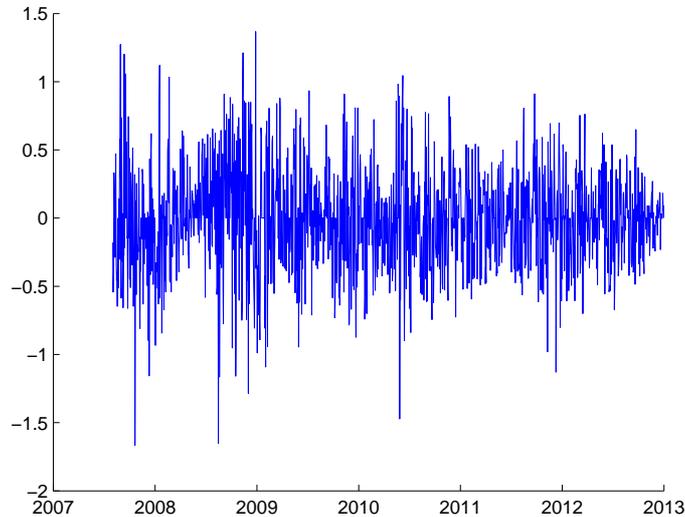


Figure 3: PHP/USD daily returns from July 2007 to December 2012.

For the prior hyperparameters, we set $\mu_0 = 0$, $V_\mu = 5$, $\mu_{h0} = 0$, $V_{\mu_h} = 5$, $\phi_{h0} = 0.95$, $V_{\phi_h} = 1$, $\nu_h = 10$ and $S_h = 0.19$. These values imply relatively diffuse priors with prior means $E(\mu) = 0$, $E(\mu_h) = 0$, $E(\phi_h) = 0.95$ and $E(\sigma_h^2) = 0.02$. For ψ we assume the truncated normal prior distribution $\mathcal{N}(\psi_0, V_\psi)1(|\psi| < 1)$ with $\psi_0 = 0$ and $V_\psi = 1$.

Posterior analysis is based on 20000 draws from the posterior distribution obtained via the sampler described in Section 3.1, after a burn-in period of 1000. We report the posterior means, standard deviations and quantiles of the model parameters in Table 3. The average daily return over the sample period is estimated to be -0.009% . Compared to the results for AUD/USD daily returns data (see Table 2), PHP/USD daily returns exhibit much lower volatility, although ϕ_h is estimated to be about the same magnitude.

Table 3: Posterior means, standard deviations and quantiles of model parameters; PHP/USD daily returns data.

| parameter | posterior mean | posterior std. dev. | 5%-tile | 95%-tile |
|--------------|----------------|---------------------|---------|----------|
| μ | -0.009 | 0.010 | -0.024 | 0.007 |
| μ_h | -2.187 | 0.290 | -2.589 | -1.793 |
| ϕ_h | 0.983 | 0.008 | 0.970 | 0.994 |
| σ_h^2 | 0.018 | 0.004 | 0.012 | 0.025 |
| ψ | 0.141 | 0.026 | 0.098 | 0.184 |

The MA(1) coefficient ψ is quite precisely estimated with posterior mean around 0.141. In addition, a 90% credible interval is (0.098, 0.184)—which excludes zero—indicating that ψ is highly unlikely to be around zero. To investigate this issue further, we compute the posterior density $p(\psi | \mathbf{y})$ as follows. Since this density has support in $(-1, 1)$ and is known up to a normalization constant, it can be evaluated on a grid and renormalized such that the area under the curve is one. We can then estimate $p(\psi | \mathbf{y})$ using the Monte Carlo average

$$\widehat{p(\psi | \mathbf{y})} = \frac{1}{R} \sum_{i=1}^R p(\psi | \mathbf{y}, \mu^{(i)}, \mathbf{h}^{(i)})$$

by summing over the posterior draws $\mu^{(i)}$ and $\mathbf{h}^{(i)}$ for $i = 1, \dots, R$.

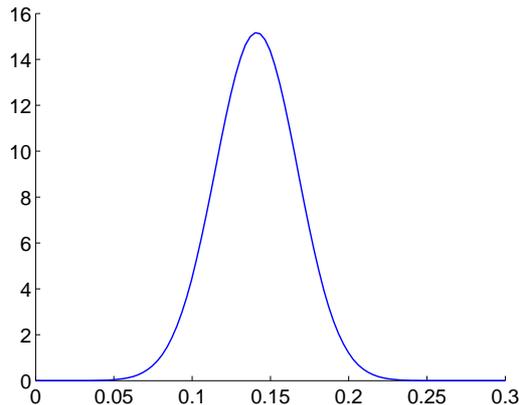


Figure 4: Estimate of $p(\psi | \mathbf{y})$; PHP/USD daily returns data.

The result is reported in Figure 4. The plot shows that most of the mass for ψ is concentrated between 0.05 and 0.25, with virtually no mass around zero. This highlights

the relevance of extending the standard stochastic volatility model to allow for moving average errors. To formally compare the moving average model with the standard variant, we perform a Bayesian model comparison exercise by computing the Bayes factor (see, e.g., [14] pp. 3–4) of the two models.

Since the models are nested—the standard stochastic volatility model can be recovered by setting $\psi = 0$ —the Bayes factor in favor of the moving average model against the standard variant can be obtained using the Savage-Dickey density ratio (see, e.g., [21]): $BF = p(\psi = 0)/p(\psi = 0 | \mathbf{y})$. That is, we evaluate the marginal prior and posterior densities for ψ at zero, and their ratio gives the relevant Bayes factor. The numerator density is a univariate truncated normal density that can be easily evaluated. The denominator density is nonstandard form, but it can be estimated using $p(\widehat{\psi = 0} | \mathbf{y})$. Using this approach, the Bayes factor is estimated to be 49600, indicating overwhelming evidence in favor of the moving average stochastic volatility model.

4 Stochastic Volatility Models with Heavy-Tailed Error Distributions

In addition to volatility clustering, another prominent feature in typical financial data is the presence of outliers. The conventional assumption of a Gaussian error distribution might be inappropriate in this context as Gaussian distributions have exponentially decaying tails. Consequently they have little mass for more extreme values. In this section, we discuss some heavy-tailed distributions that can be used to address this issue. In particular, these distributions can be written as scale mixtures of Gaussian distributions, which facilitates estimation via data augmentation. Specifically, consider the model

$$y_t = \mu + e^{\frac{1}{2}h_t} \lambda_t^{\frac{1}{2}} \varepsilon_t, \quad (16)$$

$$h_t = \mu_h + \phi_h(h_{t-1} - \mu_h) + \zeta_t, \quad (17)$$

where $\varepsilon_t \sim \mathcal{N}(0, 1)$, $\zeta_t \sim \mathcal{N}(0, \sigma_h^2)$ and λ_t are independent of each other. As before, we assume $|\phi_h| < 1$ and $h_1 \sim \mathcal{N}(\mu_h, \sigma_h^2/(1 - \phi_h^2))$.

Different distributional assumptions of the scale mixture variable λ_t leads to different error distributions. For example, if we assume that λ_t has an exponential distribution with mean parameter 2, i.e., $\lambda_t \sim \mathcal{E}(1/2)$, then the error $\tilde{\varepsilon}_t = \lambda_t^{1/2} \varepsilon_t$ follows the double-exponential distribution with density function (see [16] pp. 118-119)

$$f(x) = \frac{1}{2} e^{-|x|}.$$

Note that even though the tails of this density function are exponentially decaying, they decay at a rate that is slower than those of Gaussian distributions.

Another example is the normal log-normal distribution. More precisely, if

$$(\log \lambda_t | \tau) \sim \mathcal{N}(-0.5\tau^2, \tau^2),$$

then $\tilde{\varepsilon}_t = \lambda_t^{1/2} \varepsilon_t$ follows the normal log-normal distribution; for more details see [19].

A third example is the well-known Student-t distribution: if

$$(\lambda_t | \nu) \sim \mathcal{IG}(\nu/2, \nu/2), \quad (18)$$

then $\tilde{\varepsilon}_t = \lambda_t^{1/2} \varepsilon_t$ has a standard Student-t distribution with degree of freedom parameter ν . For a textbook treatment, see [15] pp. 254–255; for applications involving stochastic volatility models, see [18] and [22].

Hence, the stochastic volatility model with Student-t errors are defined by (16), (17) and (18). To complete the model specification, we assume independent prior distributions for μ , ν , μ_h , ϕ_h and σ_h^2 . More specifically, consider: $\mu \sim \mathcal{N}(\mu_0, V_\mu)$, $\nu \sim \mathcal{U}(0, \bar{\nu})$, $\mu_h \sim \mathcal{N}(\mu_{h0}, V_{\mu_h})$, $\phi_h \sim \mathcal{N}(\phi_{h0}, V_{\phi_h})1(|\phi_h| < 1)$ and $\sigma_h^2 \sim \mathcal{IG}(\nu_h, S_h)$. In particular, ν follows a uniform distribution with mean $\bar{\nu}/2$ and support $(0, \bar{\nu})$.

4.1 Estimation

We now discuss the estimation of the stochastic volatility model with a Student-t error distribution in (16), (17) and (18). The algorithm is illustrated in Section 4.2 via an empirical example involving daily returns on the silver spot price.

Let $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_T)'$. Posterior draws can be obtained via the following sampler that cycles through:

1. $p(\mu | \mathbf{y}, \mathbf{h}, \boldsymbol{\lambda}, \nu, \mu_h, \phi_h, \sigma_h^2) = p(\mu | \mathbf{y}, \mathbf{h}, \boldsymbol{\lambda});$
2. $p(\mathbf{h} | \mathbf{y}, \boldsymbol{\lambda}, \mu, \nu, \mu_h, \phi_h, \sigma_h^2) = p(\mathbf{h} | \mathbf{y}, \boldsymbol{\lambda}, \mu, \mu_h, \phi_h, \sigma_h^2);$
3. $p(\boldsymbol{\lambda} | \mathbf{y}, \mathbf{h}, \mu, \nu, \mu_h, \phi_h, \sigma_h^2) = \prod_{t=1}^T p(\lambda_t | y_t, h_t, \mu, \nu);$
4. $p(\nu | \mathbf{y}, \mathbf{h}, \boldsymbol{\lambda}, \mu, \mu_h, \phi_h, \sigma_h^2) = p(\nu | \boldsymbol{\lambda});$
5. $p(\sigma_h^2 | \mathbf{y}, \mathbf{h}, \boldsymbol{\lambda}, \mu, \nu, \mu_h, \phi_h) = p(\sigma_h^2 | \mathbf{h}, \mu_h, \phi_h);$
6. $p(\mu_h | \mathbf{y}, \mathbf{h}, \boldsymbol{\lambda}, \mu, \nu, \phi_h, \sigma_h^2) = p(\mu_h | \mathbf{h}, \phi_h, \sigma_h^2);$
7. $p(\phi_h | \mathbf{y}, \mathbf{h}, \boldsymbol{\lambda}, \mu, \nu, \mu_h, \sigma_h^2) = p(\phi_h | \mathbf{h}, \mu_h, \sigma_h^2).$

For Step 1, we only need to modify the corresponding step in the standard stochastic volatility model. More precisely, we have

$$(\mu | \mathbf{y}, \mathbf{h}, \boldsymbol{\lambda}) \sim \mathcal{N}(\hat{\mu}, D_\mu),$$

where $D_\mu = (V_\mu^{-1} + \sum_{t=1}^T e^{-h_t} \lambda_t^{-1})^{-1}$ and $\hat{\mu} = D_\mu (V_\mu^{-1} \mu_0 + \sum_{t=1}^T e^{-h_t} \lambda_t^{-1} y_t)$.

To implement Step 2, we first transform the data $\tilde{\mathbf{y}} = \mathbf{\Lambda}^{-\frac{1}{2}}(\mathbf{y} - \mu\mathbf{1})$, where $\mathbf{\Lambda}^{-\frac{1}{2}} = \text{diag}(\lambda_1^{-\frac{1}{2}}, \dots, \lambda_T^{-\frac{1}{2}})$. Then,

$$(\tilde{\mathbf{y}} | \mu, \mathbf{h}, \boldsymbol{\lambda}) \sim \mathcal{N}(\mathbf{0}, \mathbf{S}_y)$$

with $\mathbf{S}_y = \text{diag}(e^{h_1}, \dots, e^{h_T})$. Therefore, we can directly apply the auxiliary mixture sampler in Section 2.1 to simulate $(\mathbf{h} | \mathbf{y}, \boldsymbol{\lambda}, \mu, \mu_h, \phi_h, \sigma_h^2)$.

Next, since $\lambda_1, \dots, \lambda_T$ are conditionally independent given the model parameters and the data, we can sample each of them sequentially. In fact, we have

$$(\lambda_t | y_t, h_t, \mu, \nu) \sim \mathcal{IG}\left(\frac{\nu + 1}{2}, \frac{\nu + e^{-h_t}(y_t - \mu)^2}{2}\right).$$

To implement Step 4, we first derive the log-density $\log p(\nu | \boldsymbol{\lambda})$: it follows from (18) and the prior assumption $\nu \sim \mathcal{U}(0, \bar{\nu})$ that

$$\log p(\nu | \boldsymbol{\lambda}) = \frac{T\nu}{2} \log(\nu/2) - T \log \Gamma(\nu/2) - \left(\frac{\nu}{2} + 1\right) \sum_{t=1}^T \log \lambda_t - \frac{\nu}{2} \sum_{t=1}^T \lambda_t^{-1} + c_5$$

for $0 < \nu < \bar{\nu}$, where c_5 is a normalization constant. It is easy to check that the first and second derivatives of the log-density with respect to ν are given by

$$\begin{aligned} \frac{d \log p(\nu | \boldsymbol{\lambda})}{d\nu} &= \frac{T}{2} \log(\nu/2) + \frac{T}{2} - \frac{T}{2} \Psi(\nu/2) - \frac{1}{2} \sum_{t=1}^T \log \lambda_t - \frac{1}{2} \sum_{t=1}^T \lambda_t^{-1} \\ \frac{d^2 \log p(\nu | \boldsymbol{\lambda})}{d\nu^2} &= \frac{T}{2\nu} - \frac{T}{4} \Psi'(\nu/2), \end{aligned}$$

where $\Psi(x) = \frac{d}{dx} \log \Gamma(x)$ and $\Psi'(x) = \frac{d}{dx} \Psi(x)$ are respectively the digamma and trigamma functions.

Since the first and second derivatives can be evaluated quickly, we can maximize $\log p(\nu | \boldsymbol{\lambda})$ using, e.g., Newton-Raphson method and obtain the mode and the negative Hessian evaluated at the mode, denoted as $\hat{\nu}$ and K_ν , respectively. Then, we implement an independence-chain Metropolis-Hastings step with proposal distribution $\mathcal{N}(\hat{\nu}, K_\nu^{-1})$.

Finally, sampling from the densities $p(\mu_h | \mathbf{h}, \phi_h, \sigma_h^2)$, $p(\phi_h | \mathbf{h}, \mu_h, \sigma_h^2)$ and $p(\sigma_h^2 | \mathbf{h}, \mu_h, \phi_h)$ can be done as discussed in Section 2.2.

4.2 Empirical Example: Modeling Daily Returns on the Silver Spot Price

In Figure 5 we plot the daily returns on the silver spot price from January 2005 to December 2012. The data exhibit volatility clustering as expected from financial data. In addition, there are numerous more extreme returns—as large as $\pm 15\%$ —indicating

that even after allowing for stochastic volatility, the assumption of a Gaussian error distribution might not be appropriate.

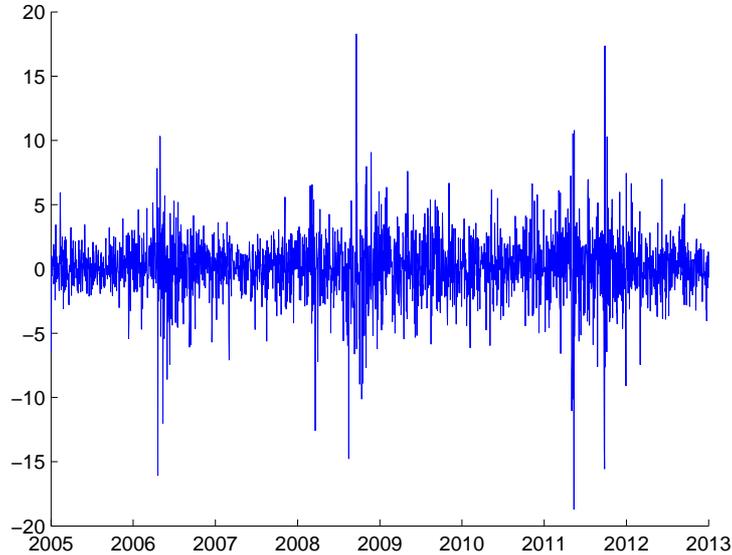


Figure 5: Daily returns on the silver spot price from January 2005 to December 2012.

To accommodate the presence of extreme returns, we fit the data using a stochastic volatility model with Student-t errors. In addition, we also allow for persistence through an MA(1) error process:

$$\begin{aligned} y_t &= \mu + u_t, \\ u_t &= \varepsilon_t + \psi\varepsilon_{t-1}, \end{aligned}$$

where $\varepsilon_t \sim \mathcal{N}(0, \lambda_t e^{h_t})$, $\varepsilon_0 = 0$ and $|\psi| < 1$. The log-volatilities evolve according to the state equation (17), and the distribution for the scale mixture variables is given in (18).

For the prior hyperparameters, we set $\mu_0 = 0$, $V_\mu = 5$, $\psi_0 = 0$ and $V_\psi = 1$, $\mu_{h0} = 0$, $V_{\mu_h} = 5$, $\phi_{h0} = 0.95$, $V_{\phi_h} = 1$, $\nu_h = 10$ and $S_h = 0.19$. These values imply relatively diffuse priors with prior means $E(\mu) = 0$, $E(\psi) = 0$, $E(\mu_h) = 0$, $E(\phi_h) = 0.95$ and $E(\sigma_h^2) = 0.02$. The prior distribution for ν is $\mathcal{U}(0, \bar{\nu})$ with $\bar{\nu} = 50$, implying a prior mean of 25.

Estimation of the model can proceed by combining the samplers in Sections 3.1 and 4.1. More specifically, if we define $\mathbf{z} = \mathbf{H}_\psi^{-1} \mathbf{y}$, then

$$(\mathbf{z} \mid \mathbf{h}, \boldsymbol{\lambda}, \boldsymbol{\psi}, \mu) \sim \mathcal{N}(\mu \mathbf{H}_\psi^{-1} \mathbf{1}, \boldsymbol{\Sigma}_z),$$

where $\boldsymbol{\Sigma}_z = \text{diag}(\lambda_1 e^{h_1}, \dots, \lambda_T e^{h_T})$. Hence, the sampler in Section 4.1 can be applied to \mathbf{z} to sample from the full conditional distributions of \mathbf{h} , $\boldsymbol{\lambda}$, μ , ν , μ_h , ϕ_h and σ_h^2 . Moreover, we can quickly evaluate the density of \mathbf{z} , which can be used for a Metropolis-Hastings step for simulating from the full conditional distribution of $\boldsymbol{\psi}$.

Table 4: Posterior means, standard deviations and quantiles of model parameters; silver spot price daily returns data.

| parameter | posterior mean | posterior std. dev. | 5%-tile | 95%-tile |
|--------------|----------------|---------------------|---------|----------|
| μ | 0.125 | 0.037 | 0.064 | 0.186 |
| μ_h | 1.102 | 0.232 | 0.744 | 1.436 |
| ϕ_h | 0.984 | 0.006 | 0.974 | 0.993 |
| σ_h^2 | 0.017 | 0.005 | 0.011 | 0.024 |
| ψ | -0.088 | 0.022 | -0.120 | -0.050 |
| ν | 6.675 | 1.235 | 5.071 | 8.894 |

We obtain 20000 draws from the posterior distribution, after a burn-in period of 1000. Table 4 reports the posterior means, standard deviations and quantiles of the model parameters. For example, the average daily return over the sample period is estimated to be 0.125%. The estimates for μ_h , ϕ_h and σ_h^2 are similar to those obtained in the previous two empirical examples.

The posterior mean of the MA(1) coefficient ψ is -0.088 with a 90% credible interval $(-0.12, -0.05)$ —which excludes 0—indicating some persistence in the errors. Moreover, the degree of freedom parameter ν is estimated to be about 6.7. This shows that the error distribution seems to have heavier tails than those of a Gaussian distribution.

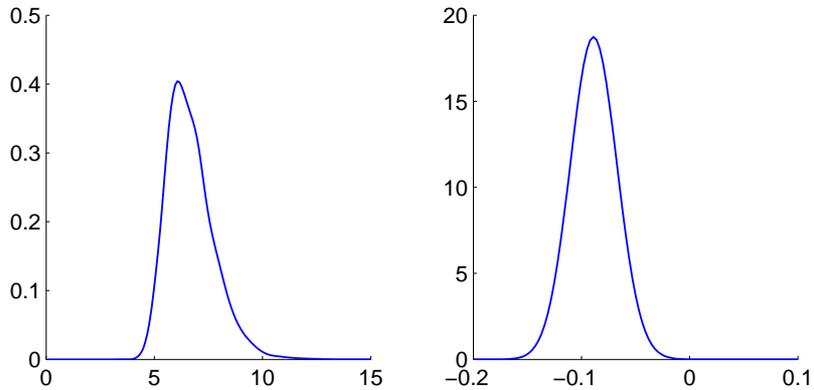


Figure 6: Estimates of $p(\nu | \mathbf{y})$ (left panel) and $p(\psi | \mathbf{y})$ (right panel); silver spot price daily returns data.

The right panel of Figure 6 depicts the estimate of the marginal density $p(\psi | \mathbf{y})$. The density plot indicates that the value 0 is highly unlikely given the data. The left panel shows the density plot of $p(\nu | \mathbf{y})$, in which most of its mass is concentrated between 5 and 10, highlighting the relevance of the extension to a Student-t error distribution.

References

- [1] Belmonte M. and Koop G., “Model Switching and Model Averaging in Time-Varying Parameter Regression Models,” *Strathclyde Discussion Papers in Economics*, No. 13-02, 2013.
- [2] Carter C.K. and Kohn R., “On Gibbs Sampling for State Space Models,” *Biometrika*, Vol. 81, No. 3, pp. 541–553, 1994.
- [3] Chan J.C.C., “Moving Average Stochastic Volatility Models with Application to Inflation Forecast,” *Journal of Econometrics*, Vol. 176, No. 2, pp. 162–172, 2013.
- [4] Chan J.C.C. and Jeliaskov I., “Efficient Simulation and Integrated Likelihood Estimation in State Space Models,” *International Journal of Mathematical Modelling and Numerical Optimisation*, Vol. 1, No. 1/2, pp. 101–120, 2009.
- [5] Chan J.C.C., Koop G., Leon-Gonzalez R. and Strachan R., “Time Varying Dimension Models,” *Journal of Business and Economic Statistics*, Vol. 30, No. 3, pp. 358–367, 2012.
- [6] Chan J.C.C., Koop G. and Potter S.M., “A New Model of Trend Inflation,” *Journal of Business and Economic Statistics*, Vol. 31, No. 1, pp. 94–106, 2013.
- [7] de Jong P. and Shephard N., “The Simulation Smoother for Time Series Models,” *Biometrika*, Vol. 82, No. 2, pp. 339–350, 1995.
- [8] Durbin J. and Koopman S.J., “A Simple and Efficient Simulation Smoother for State Space Time Series Analysis,” Vol. 89, No. 3, pp. 603–615, 2002.
- [9] Frühwirth-Schnatter S., “Data Augmentation and Dynamic Linear Models,” *Journal of Time Series Analysis*, Vol. 15, No. 2, pp. 183–202, 1994.
- [10] Gardner G., Harvey A.C. and Phillips G.D.A., “Algorithm AS 154: An Algorithm for Exact Maximum Likelihood Estimation of Autoregressive-Moving Average Models by Means of Kalman Filtering,” *Journal of the Royal Statistical Society Series C*, Vol. 29, No. 3, pp. 311–322, 1980.
- [11] Golub G.H. and van Loan C.F., *Matrix Computations*, Johns Hopkins University Press, Baltimore, 1983.
- [12] Jeliaskov I. and Liu R., “A Model-Based Ranking of U.S. Recessions,” *Economics Bulletin*, Vol. 30, No. 3, pp. 2289–2296, 2010.
- [13] Kim S., Shepherd N. and Chib S., “Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models,” *Review of Economic Studies*, Vol. 65, No. 3, pp. 361–393, 1998.
- [14] Koop G., *Bayesian Econometrics*, John Wiley & Sons, New York, 2003.

- [15] Koop G., Poirier D.J. and Tobias J.L. *Bayesian Econometric Methods*, Cambridge University Press, New York, 2007.
- [16] Kroese D.P., Taimre T. and Botev Z.I., *Handbook of Monte Carlo Methods*, John Wiley & Sons, New York, 2011.
- [17] McCausland W.J., Miller S. and Pelletier D., “Simulation Smoothing for State-space Models: A Computational Efficiency Analysis,” *Computational Statistics and Data Analysis*, Vol. 55, No. 1, pp 199–212, 2011.
- [18] Nakajima J. and Omori Y., “Leverage, Heavy-tails and Correlated Jumps in Stochastic Volatility Models,” *Computational Statistics and Data Analysis*, Vol. 53, No. 6, pp. 2335–2353, 2009.
- [19] Omori Y., Chib S., Shephard N. and Nakajima J. “Stochastic Volatility with Leverage: Fast and Efficient Likelihood Inference,” *Journal of Econometrics*, Vol. 140, No. 2, pp. 425–449, 2007.
- [20] Rue H., “Fast Sampling of Gaussian Markov Random Fields with Applications,” *Journal of the Royal Statistical Society Series B*, Vol. 63, No. 2, pp. 325–338, 2001.
- [21] Verdinelli I. and Wasserman L., “Computing Bayes Factors Using a Generalization of the Savage-Dickey Density Ratio,” *Journal of the American Statistical Association*, Vol. 90, No. 430, pp. 614–618, 1995.
- [22] Wang J.J.J., Chan J.S.K. and Choy S.T. Boris, “Stochastic Volatility Models with Leverage and Heavy-tailed Distributions: A Bayesian Approach Using Scale Mixtures,” *Computational Statistics and Data Analysis*, Vol. 55, No. 1, pp. 852–862, 2011.