

Bayesian Estimation of Wishart Autoregressive Stochastic Volatility Model

Ming Lin^a Changjiang Liu^b Linlin Niu^{a*}

^a WISE, Xiamen University

^b Essence Securities Company

November 2012

Abstract

The Wishart autoregressive (WAR) process is a powerful tool to model multivariate stochastic volatility (MSV) with correlation risk and derive closed-form solutions in various asset pricing models. However, making inferences of the WAR stochastic volatility (WAR-SV) model is challenging because the latent volatility series does not have a closed-form transition density. Based on an alternative representation of the WAR process with lag order $p = 1$ and integer degrees of freedom, we develop an effective two-step procedure to estimate parameters and the latent volatility series. The procedure can be applied to study other varying-dimension problems. We show the effectiveness of this procedure with a simulated example. Then this method is used to study the time-varying correlation of US and China stock market returns.

Keywords: Bayesian posterior probability, Markov chain Monte Carlo, Multivariate stochastic volatility, Sequential Monte Carlo, Wishart autoregressive process.

^{*a}: WISE, Economics Building A307, Xiamen University, 361005, Xiamen, China. E-Mail: ming.lin50@gmail.com and linlin.niu@gmail.com; ^b: 19F B Tower, Xin Sheng Plaza, 5 Jin Rong Avenue, West District, Beijing, 100033, China. E-Mail: changjiang.liu@gmail.com.

1 Introduction

Time-varying volatility and changing correlations are seminal phenomena in economic and financial time series. The ability to jointly model these features is important in many economic studies and financial applications, such as understanding the dynamic structure of the macroeconomy, asset pricing, risk management and optimal portfolio allocation (Primiceri, 2005; Harvey, 1989; Christodoulakis and Satchell, 2002; Engle, 2002; Tse and Tsui, 2002; Buraschi et al., 2010). To that end, there have been extensive applications of multivariate GARCH models (Bollerslev et al., 1988; Bollerslev, 1990; Engle and Kroner, 1995; Engle, 2002) and expanding development of multivariate stochastic volatility (MSV) models (Harvey et al., 1994; Jacquier et al., 1994; McAleer, 2005; Yu and Meyer, 2006) in the literature. Compared with GARCH-type models, where the current volatility is deterministic given past information, stochastic volatility models introduce additional innovations and are more flexible. In addition, MSV models are often linked to continuous-time models. However, the latent structure in MSV models also renders complication in estimation.

Gourieroux et al. (2009) propose the Wishart autoregressive (WAR) process to model multivariate stochastic volatilities. The WAR process is regarded as the discretization of a multivariate extension of the Cox-Ingersoll-Ross (CIR) model (Cox et al., 1985; Gourieroux, 2006). The process generates symmetric and positive definite volatility matrices with a relatively small number of parameters. Moreover, this model enables the closed-form forecast of future volatilities given the current volatility.

The theoretically appealing properties of the WAR process were quickly explored in financial modeling. Thanks to its well-defined conditional Laplace transform, the WAR process has been successfully applied to developing theoretical models for the term structure of interest rates (Gourieroux and Sufana, 2003; Buraschi et al., 2008; Cieslak and Povala, 2009; Niu, 2009), and to deriving convenient models with analytical solutions for option pricing and for optimal portfolio allocation under time-varying correlation risks (Gourieroux and Sufana, 2010; Da Fonseca et al., 2007, 2008; Buraschi et al.,

2010).

However, the empirical applications of these models are still limited to cases where the underlying volatility process is assumed to be known *ex ante*. For example, both [Gourieroux et al. \(2009\)](#) and [Buraschi et al. \(2010\)](#) calculate the realized volatility matrices from high-frequency data. Then the method of moments is applied to estimate the parameters. It would be more useful to estimate the models in a general situation where the volatility-covolatility process is latent. In fact, many other existing MSV models, such as those surveyed by [Asai et al. \(2006\)](#) and empirically compared by [Asai and McAleer \(2009\)](#), can be used in the general case where a Markov chain Monte Carlo (MCMC) procedure is employed to estimate the parameters and the underlying volatility process simultaneously.

The major challenge hindering the inference of the latent WAR process is the fact that its conditional distribution follows a noncentral Wishart distribution, whose density function does not have a closed-form expression. Calculating the density involves evaluation of the hypergeometric function, which is a series expansion and needs to be computed recursively ([Gourieroux et al., 2009](#)). The complexity of this conditional density function imposes a serious computational burden in estimations with the MCMC method.

In this paper, we focus on the WAR process with lag order $p = 1$ and integer degrees of freedom, and utilize an alternative representation to circumvent the problem that the WAR process does not have a closed-form transition density. This equivalent representation states the WAR process as a cross-product of an underlying mean-zero vector autoregressive process with normal innovations. We develop a two-step procedure to infer the WAR stochastic volatility model. At first, given the degrees of freedom, other parameters in the model and the latent volatility-covolatility process can be estimated simultaneously using an MCMC procedure. Then the degrees of freedom are determined through Bayesian posterior probabilities, which are calculated using a sequential Monte Carlo (SMC) procedure.

The remainder of this paper is organized as follows. Section 2 briefly introduces the WAR model, its applications in financial modeling and the challenges in its estimation. In Section 3, we describe the two-step estima-

tion procedure. Section 4 demonstrates the effectiveness of the procedure with a simulated example. Section 5 applies this method to study the time-varying correlation of Shanghai and New York stock market returns. Section 6 concludes.

2 The WAR Stochastic Volatility Model

The WAR stochastic volatility (WAR-SV) model we consider is described as

$$\begin{aligned} \text{observation:} \quad & Y_t \sim N(0, \Omega_t), \\ \text{state dynamics:} \quad & \{\Omega_t\} \sim \text{WAR process}, \end{aligned} \tag{1}$$

where Y_t is a vector observation at time t , and the latent state, Ω_t , is its volatility matrix, which follows a WAR process. The estimation method we propose in this paper can be easily extended to models where Y_t has a non-zero conditional mean, for example, $\{Y_t\}$ can either follow a vector autoregressive process or its conditional mean can have a volatility-in-mean term.

2.1 The WAR Process

Gourieroux and Sufana (2003) first propose the WAR process of lag order $p = 1$ in a new MSV model to be applied to the term structure of interest rates, and the generalized model with lag order $p \geq 1$ is discussed in detail in Gourieroux et al. (2009). It is defined through the conditional Laplace transform as follows.

Definition 1 *The Wishart autoregressive process of order p , called $WAR(p)$ and denoted by $W_n(K; M_1, \dots, M_p, \Sigma)$, is a $n \times n$ matrix process $\{\Omega_t\}$ with the conditional Laplace transform:*

$$\begin{aligned} \Psi_{t-1}(\Gamma) &= E_{t-1} [\exp \text{Tr}(\Gamma \Omega_t)] \\ &= \frac{\exp \text{Tr} \left[\Gamma (I_n - 2\Sigma\Gamma)^{-1} \sum_{j=1}^p M_j \Omega_{t-j} M_j' \right]}{[\det(I_n - 2\Sigma\Gamma)]^{K/2}}, \end{aligned}$$

where Γ is an $n \times n$ matrix, E_{t-1} is the conditional expectation given all information up to time $t - 1$, Tr stands for the trace of a matrix, I_n is the

identity matrix of size n , $K \geq n$ is the degrees of freedom, p is the lag order, M_1, \dots, M_p are $n \times n$ matrices of the autoregressive coefficients, M_j' is the transpose of M_j , and Σ is an $n \times n$ symmetric positive definite matrix.

The WAR process is appealing to modeling time-varying volatility-covolatility matrices because it ensures symmetry and positive definiteness of the matrix Ω_t . [Gourieroux et al. \(2009\)](#) show that the $W_n(K; M_1, \dots, M_p, \Sigma)$ process has the dynamic representation

$$\Omega_t = K\Sigma + \sum_{j=1}^p M_j \Omega_{t-j} M_j' + \eta_t, \quad (2)$$

where η_t is an innovation matrix satisfying $E_{t-1}(\eta_t) = \mathbf{0}$. This dynamic enables closed-form forecasts of future states given the current state and past states.

When lag order $p = 1$, the process has a direct continuous-time analog ([Gourieroux, 2006](#)) which is popular in the derivation of finance models. [Gourieroux et al. \(2009\)](#) state that the WAR process of order $p = 1$ is able to generate a large spectrum of persistence patterns in volatility and covolatility. We focus on the WAR(1) process in the following.

2.2 Financial Applications

A prominent feature of the WAR process in financial and economic applications is that its closed-form conditional Laplace transform helps derive closed-form solutions for asset pricing models with stochastic volatility, in particular those with time-varying correlations.

When applied to the term structure of interest rates in discrete time, [Gourieroux and Sufana \(2003\)](#) assume that the stochastic discount factor $D_{t,t+1}$ in period $(t, t + 1)$ is exponential matrix affine in some underlying factor Ω_t , that is,

$$D_{t,t+1} = \exp[d + Tr(C\Omega_{t+1})],$$

where Ω_t follows a WAR process. Under this assumption, the bond price of maturity τ at time t can be derived as

$$P_{t,\tau} = \exp \{A_\tau + Tr[C_\tau \Omega_t]\},$$

where A_τ and C_τ are matrix coefficients with closed-form expression.

Niu (2009) derives an affine term structure model under a more general setting with the stochastic discount factor determined by a vector autoregressive process $\{Y_t\}$. The time-varying volatility-covolatility of the innovation in $\{Y_t\}$ is modeled by a WAR process of Ω_t . The bond price, $P_{t,\tau}$, then has the closed-form solution

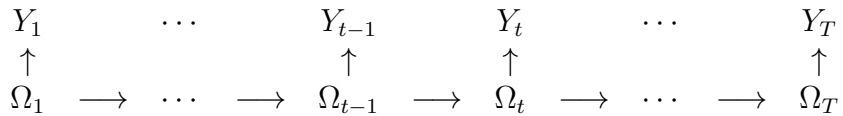
$$P_{t,\tau} = \exp \{A_\tau + B_\tau Y_t + Tr[C_\tau \Omega_t]\}.$$

Buraschi et al. (2008) and Cieslak and Povala (2009) study equilibrium term structures under stochastic volatility using the WAR process in a continuous-time framework. The WAR process is also used to study the implications of correlation risk in derivative pricing (Gourieroux and Sufana, 2010; Buraschi et al., 2010; Da Fonseca et al., 2007, 2008).

2.3 Challenges in Estimating the WAR-SV model

When the volatilities $\{\Omega_t\}$ are observable, Gourieroux et al. (2009) propose a moment method to estimate the parameters in a WAR(1) process. However, in general stochastic volatility models, the volatilities are unobservable. In these circumstances, the inference of the parameters and the latent volatility process relies on exploring information contained in the observations $\{Y_t\}$.

The dynamic conditional structure of the WAR(1)-SV model is portrayed as the following diagram:



Let $\Theta = (M, \Sigma, K)$, its likelihood function is

$$\begin{aligned}
 L(\Theta) &= P(Y_{1:T}; \Theta) \\
 &= \int P(Y_{1:T}, \Omega_{1:T}; \Theta) d\Omega_{1:T} \\
 &= \int \prod_{t=1}^T P(\Omega_t | \Omega_{t-1}; \Theta) P(Y_t | \Omega_t) d\Omega_{1:T},
 \end{aligned}$$

where $Y_{1:T} = (Y_1, \dots, Y_T)$ and $\Omega_{1:T} = (\Omega_1, \dots, \Omega_T)$. When evaluating the likelihood function, we need to integrate out the latent volatilities $\Omega_1, \dots, \Omega_T$, which makes the task of obtaining the maximum likelihood estimation of the parameters formidable.

Common methods of inference with stochastic volatility models include the Markov chain Monte Carlo (MCMC) (Philipov and Glickman, 2006), the simulated maximum likelihood estimation (SMLE) (Durbin and Koopman, 1997; Shephard and Pitt, 1997), and nonlinear filtering methods such as the unscented Kalman filter (Cieslak and Povala, 2009), etc. The advantage of MCMC compared to SMLE is that it obtains estimates of both the parameters and the latent volatilities, simultaneously. Also, unlike the unscented Kalman filter, the MCMC method approaches the true model distribution without systematic bias.

However, to the best of our knowledge, none of the existing studies applying the WAR process use the MCMC method to infer the latent volatilities. This is due to the nonexistence of closed-form transition density in the WAR process (Gourieroux et al., 2009).

The complexity of the conditional density function in the Wishart process seems devastating for efficient inference calculation. To circumvent this problem, we focus on the WAR(1) process with integer degrees of freedom and utilize an alternative representation of this special class of the WAR process.

2.4 WAR(1)-SV Model with Integer Degrees of Freedom

As a special case of Model (1), the WAR(1)-SV model can be written as

$$\begin{aligned} \text{observation:} & \quad Y_t \sim N(0, \Omega_t), \\ \text{state equation:} & \quad \{\Omega_t\} \sim \text{WAR(1) process.} \end{aligned}$$

When the degrees of freedom K is an integer, this system has an alternative representation as follows (Gourieroux et al., 2009):

$$\begin{aligned}
\text{observation:} \quad & Y_t \sim N(0, \Omega_t), \\
\text{state equation:} \quad & \Omega_t = Z_t Z_t', \\
& Z_t = M Z_{t-1} + \Xi_t,
\end{aligned} \tag{3}$$

where $Z_t = (z_{1,t}, \dots, z_{K,t})$ and $\Xi_t = (\xi_{1,t}, \dots, \xi_{K,t})$ are $n \times K$ matrices, and $\{\xi_{1,t}\}, \dots, \{\xi_{K,t}\}$ are independent Gaussian white noises with $\xi_{i,t} \sim N(0, \Sigma)$ for $i = 1, \dots, K$. In this new representation, the latent state process $\{Z_t\}$ has a Gaussian transition density. However, the dimension of Z_t depends on the unknown degrees of freedom K .

We develop a two-step procedure to estimate this model. First, we use the MCMC method to make inference of the parameters (M, Σ) and the latent volatilities $\Omega_t = Z_t Z_t'$, $t = 1, \dots, T$, under given degrees of freedom. Then the degrees of freedom are determined through Bayesian posterior probabilities. The results of the MCMC procedure are used to develop an efficient sequential Monte Carlo (SMC) algorithm to calculate the Bayesian posterior probabilities. We will discuss the estimation procedure in detail in the next section.

3 Bayesian Estimation

3.1 Given K , Inference of Other Parameters and Latent States

When the number of degrees of freedom is given, the joint posterior distribution of $(M, \Sigma, Z_{1:T})$ in Model (3) is

$$\begin{aligned}
& P(M, \Sigma, Z_{1:T} \mid Y_{1:T}; K) \\
& \propto P(M, \Sigma, Z_{1:T}, Y_{1:T} \mid K) \\
& = P(M, \Sigma, Z_1 \mid K) P(Y_1 \mid Z_1; K) \\
& \quad \times \prod_{t=2}^T P(Z_t \mid Z_{t-1}, M, \Sigma; K) P(Y_t \mid Z_t; K),
\end{aligned}$$

where $P(M, \Sigma, Z_1 | K)$ are the prior distributions of parameters (M, Σ) and the initial state Z_1 . In the following, we always use $P(\cdot)$ to denote the distribution in Model (3).

We use the MCMC method to draw samples from the joint posterior distribution $P(M, \Sigma, Z_{1:T} | Y_{1:T}; K)$ and make inferences. For simplicity, we suppress the degrees of freedom K and let $(M^{*(s)}, \Sigma^{*(s)}, Z_{1:T}^{*(s)})$, $s = 1, \dots, S$, be the samples generated by MCMC. Then for any function $h(M, \Sigma, Z_{1:T})$ with finite expectation,

$$\frac{1}{S} \sum_{s=1}^S h(M^{*(s)}, \Sigma^{*(s)}, Z_{1:T}^{*(s)})$$

is a consistent estimation of the expectation $E_P [h(M, \Sigma, Z_{1:T}) | Y_{1:T}; K]$. A more detailed discussion of the MCMC method can be found in Robert and Casella (1999) and Liu (2001).

In the following, we describe some implementation details of the MCMC procedure.

3.1.1 Prior Distributions

We assume that the prior distribution

$$P(M, \Sigma, Z_1 | K) = P(M | K) P(\Sigma | K) P(Z_1 | K),$$

that is, M , Σ , and Z_1 are independent given K . A dependent joint prior can also be used.

We use conjugate priors for M and Σ . For ease of presentation, we rearrange the elements of the $n \times n$ matrix M into an n^2 -dimensional vector $\vec{M} = (M_{1.}, \dots, M_{n.})'$, where $M_{i.}$ is the i -th row of matrix M . The prior $P(\vec{M} | K)$ is set as the normal distribution $N(\mu_{\vec{M}, \text{prior}}, \Sigma_{\vec{M}, \text{prior}})$. The prior $P(\Sigma | K)$ follows an inverse Wishart distribution $W^{-1}(V_{\text{prior}}, d_{\text{prior}})$ with its probability density function denoted as

$$\frac{|V_{\text{prior}}|^{d_{\text{prior}}/2}}{2^n d_{\text{prior}}^{n/2} \Gamma_n(d_{\text{prior}}/2)} |\Sigma|^{-(d_{\text{prior}}+n+1)/2} \exp \left\{ -Tr(V_{\text{prior}} \Sigma^{-1}) / 2 \right\},$$

where $|\cdot|$ stands for the determinant of the matrix. The hyperparameters $(\mu_{\vec{M},prior}, \Sigma_{\vec{M},prior})$ and (V_{prior}, d_{prior}) in the prior may affect the results, especially when the sample size T is small. In practice, we can use a flat prior with large $\Sigma_{\vec{M},prior}$ and small d_{prior} .

For the prior of $Z_1 = (z_{1,1}, \dots, z_{1,K})$, we let $z_{1,1}, \dots, z_{1,K}$ be independent and identically distributed (i.i.d.) following a normal distribution $N(\mu_{z,prior}, \Sigma_{z,prior})$, where $\mu_{z,prior}$ is a zero vector and

$$\Sigma_{z,prior} = \frac{1}{KT} \sum_{t=1}^T Y_t Y_t'.$$

Under this prior distribution, we have $E(Z_1 Z_1') = \frac{1}{T} \sum_{t=1}^T Y_t Y_t'$.

3.1.2 Iteratively Updating with Hybrid MH-Gibbs Sampler

We draw samples $(\vec{M}^{*(s)}, \Sigma^{*(s)}, Z_{1:T}^{*(s)})$ from the posterior distribution $P(M, \Sigma, Z_{1:T} | Y_{1:T}; K)$ by iteratively updating each component with the Gibbs sampling or the Metropolis-Hasting (MH) sampling.

Suppose at $s = 0$, we have the initial sample $(M^{*(0)}, \Sigma^{*(0)}, Z_{1:T}^{*(0)})$, then the steps to generate samples $(M^{*(s)}, \Sigma^{*(s)}, Z_{1:T}^{*(s)})$, $s = 1, 2, \dots, S$, are illustrated as follows:

1. Draw $M^{*(s)}$ from distribution $P(M | \Sigma^{*(s-1)}, Z_{1:T}^{*(s-1)}, Y_{1:T}; K)$.
2. Draw $\Sigma^{*(s)}$ from distribution $P(\Sigma | M^{*(s)}, Z_{1:T}^{*(s-1)}, Y_{1:T}; K)$.
3. For $t = 1, \dots, T$,
 - (a) Draw $Z_t^{*(s)}$ from distribution $G(Z_t | M^{*(s)}, \Sigma^{*(s)}, Z_{1:t-1}^{*(s)}, Z_{t+1:T}^{*(s-1)}, Y_{1:T}; K)$.
 - (b) Accept the new sample $Z_t^{*(s)}$ with probability

$$p = \min \left\{ 1, \frac{P(Z_t^{*(s)} | rest; K) G(Z_t^{*(s-1)} | rest; K)}{P(Z_t^{*(s-1)} | rest; K) G(Z_t^{*(s)} | rest; K)} \right\}, \quad (4)$$

where $rest = (M^{*(s)}, \Sigma^{*(s)}, Z_{1:t-1}^{*(s)}, Z_{t+1:T}^{*(s-1)}, Y_{1:T})$. Otherwise, let $Z_t^{*(s)} = Z_t^{*(s-1)}$.

Detailed calculations of the sampling distributions and the acceptance rates are presented in Appendix A.

3.2 Selection of K

In Model (3), the dimension of state variable Z_t depends on the degrees of freedom K . Green (1995) proposes the reversible jump MCMC method for Bayesian model determination when the dimension of the model space is not fixed. However, how to choose an efficient jump between state spaces of differing dimensionality, in general cases, remains a challenging problem. We use the Bayesian posterior probability $P(K | Y_{1:T})$ to select the proper model (Kass and Raftery, 1995; Green, 2001). More specifically, suppose $\mathbb{K} = \{n, n+1, \dots, n+r\}$ is the set of all possible values of K . We calculate $P(K | Y_{1:T})$ for every $K \in \mathbb{K}$, then choose the K with the largest posterior probability as the degrees of freedom of the model.

The posterior probability of K , given observations Y_1, \dots, Y_T , can be written as

$$\begin{aligned}
& P(K | Y_{1:T}) \\
& \propto P(K)P(Y_{1:T} | K) \\
& \propto P(K) \int P(M, \Sigma, Z_{1:T}, Y_{1:T} | K) dZ_{1:T} dM d\Sigma \\
& = P(K) \int P(M, \Sigma | K)P(Z_1 | K)P(Y_1 | Z_1; K) \\
& \quad \times \prod_{t=2}^T P(Z_t | Z_{t-1}, M, \Sigma; K)P(Y_t | Z_t; K) dZ_{1:T} dM d\Sigma.
\end{aligned} \tag{5}$$

Usually, the prior distribution $P(K)$ is assumed to be uniformly distributed in \mathbb{K} , but the calculation of $P(Y_{1:T} | K)$ involves high dimensional integration over the parameters (M, Σ) and the latent states (Z_1, \dots, Z_T) as in Equation (5).

According to the principle of importance sampling (Robert and Casella, 1999), if we generate samples $(M^{(j)}, \Sigma^{(j)}, Z_{1:T}^{(j)})$, $j = 1, \dots, m$, from a *trial distribution* $Q(M, \Sigma, Z_{1:T})$ whose support covers the support of $P(M, \Sigma, Z_{1:T} |$

$Y_{1:T}; K$), and let the weight

$$w^{(j)} = \frac{P(M^{(j)}, \Sigma^{(j)}, Z_{1:T}^{(j)}, Y_{1:T} | K)}{Q(M^{(j)}, \Sigma^{(j)}, Z_{1:T}^{(j)})},$$

then

$$\frac{1}{m} \sum_{j=1}^m w^{(j)} \xrightarrow{a.s.} E_Q(w) = P(Y_{1:T} | K). \quad (6)$$

The choice of the trial distribution Q will greatly affect the performance of the estimator (6), especially when calculating high dimensional integrations. In theory, the “perfect” trial distribution is

$$Q(M, \Sigma, Z_{1:T}) = P(M, \Sigma, Z_{1:T} | Y_{1:T}; K). \quad (7)$$

In this case, the variance of weight w becomes zero and the estimator (6) gives the exact value of $P(Y_{1:T} | K)$. However, to use the “perfect” trial distribution (7), we need to know the value of

$$P(M, \Sigma, Z_{1:T} | Y_{1:T}; K) = \frac{P(M, \Sigma, Z_{1:T}, Y_{1:T} | K)}{P(Y_{1:T} | K)}$$

for the weight calculation, which involves the unknown value $P(Y_{1:T} | K)$ we want to obtain. Although the “perfect” trial distribution is infeasible in practice, we should choose a trial distribution close to $P(M, \Sigma, Z_{1:T} | Y_{1:T}; K)$ and easy to draw samples from.

In general, it is difficult to directly find an efficient trial distribution in such a high dimensional space as $(M, \Sigma, Z_{1:T})$. Liu and Chen (1998) propose a sequential Monte Carlo (SMC) method, called *sequential importance sampling with resampling* (SISR), to sequentially build up high-dimensional random samples.

3.2.1 Sequential Importance Sampling with Resampling

Suppose $f_0(M, \Sigma), f_1(M, \Sigma, Z_1), \dots, f_t(M, \Sigma, Z_{1:t}), \dots, f_T(M, \Sigma, Z_{1:T})$ is a sequence of functions of $(M, \Sigma, Z_{1:t})$ with increasing dimensionality. The SISR method is described as follows.

1. At time $t = 0$, draw $(M^{(j)}, \Sigma^{(j)})$ from the trial distribution $Q(M, \Sigma)$.
Let $w_0^{(j)} = \frac{f_0(M^{(j)}, \Sigma^{(j)})}{Q(M^{(j)}, \Sigma^{(j)})}$.
2. At times $t = 1, 2, \dots, T$,
 - (a) Sampling: Draw $Z_t^{(j)}$ from the trial distribution $Q(Z_t | M^{(j)}, \Sigma^{(j)}, Z_{1:t-1}^{(j)})$.
 - (b) Updating Weights: Update the weights by letting

$$w_t^{(j)} = w_{t-1}^{(j)} \frac{f_t(M^{(j)}, \Sigma^{(j)}, Z_{1:t}^{(j)})}{f_{t-1}(M^{(j)}, \Sigma^{(j)}, Z_{1:t-1}^{(j)}) Q(Z_t^{(j)} | M^{(j)}, \Sigma^{(j)}, Z_{1:t-1}^{(j)})}.$$

- (c) Resampling: Denote $(M, \Sigma, Z_{1:t})$ by $Z_{0:t}$. When $t < T$,
 - i. draw $Z_{0:t}^{new(j)}$, $j = 1, \dots, m$, from the distribution

$$\sum_{j=1}^m \frac{w_t^{(j)}}{\sum_{k=1}^m w_t^{(k)}} \delta(Z_{0:t} - Z_{0:t}^{(j)}),$$

where $\delta(\cdot)$ is the Dirichlet function, and set $w_t^{new(j)} = \frac{1}{m} \sum_{j=1}^m w_t^{(j)}$;

- ii. let $Z_{0:t}^{(j)} = Z_{0:t}^{new(j)}$ and $w_t^{(j)} = w_t^{new(j)}$ for $j = 1, \dots, m$.

The resampling step is used to rejuvenate the skewed samples and plays a key role in the SMC method. Many other resampling schemes are outlined in Kitagawa (1996), Liu and Chen (1998), Carpenter et al. (1999), Crisan and Lyons (2002) and Pitt (2002).

It is easy to show that at time t ,

$$\frac{1}{m} \sum_{j=1}^m w_t^{(j)} \rightarrow \int f_t(M, \Sigma, Z_{1:t}) dZ_{1:t} dM d\Sigma.$$

If we let

$$f_T(M, \Sigma, Z_{1:T}) = P(M, \Sigma, Z_{1:T}, Y_{1:T} | K), \quad (8)$$

then $\frac{1}{m} \sum_{j=1}^m w_T^{(j)}$ is a consistent estimator of $P(Y_{1:T} | K)$.

3.2.2 Implementation of the SMC Procedure

We now describe the implementation of the SMC procedure in detail. The calculations and algorithmic steps are presented in Appendix B.

1. *Choice of the intermediate target function $f_t(M, \Sigma, Z_{1:t})$.* To obtain consistent estimation of $P(Y_{1:T} | K)$ with SISR, it only requires that Equation (8) holds. However, choice of $f_t(M, \Sigma, Z_{1:t})$, $0 \leq t < T$, will affect the efficiency of the SISR method through the resampling step. A natural choice of the intermediate target function $f_t(M, \Sigma, Z_{1:t})$ is letting

$$f_t(M, \Sigma, Z_{1:t}) = P(M, \Sigma, Z_{1:t}, Y_{1:t} | K), \quad t = 0, 1, \dots, T. \quad (9)$$

This choice does not consider using future observations $Y_{t+1:T}$. After the resampling step at time t , the random samples $(M^{(j)}, \Sigma^{(j)}, Z_{1:t}^{(j)})$, $j = 1, \dots, m$, will approximately follow the distribution

$$P(M, \Sigma, Z_{1:t} | Y_{1:t}; K) \propto f_t(M, \Sigma, Z_{1:t}),$$

which is different from the “perfect” marginal distribution $P(M, \Sigma, Z_{1:t} | Y_{1:T}; K)$ as in Equation (7), especially when t is small. It may result in many unrepresentative samples and diminish the efficiency.

Instead of using (9) as the intermediate target function, we let

$$f_t(M, \Sigma, Z_{1:t}) = \left[\frac{P(M, \Sigma | Y_{1:T}; K)}{P(M, \Sigma | K)} \right]^{(T-t)/T} P(M, \Sigma, Z_{1:t}, Y_{1:t} | K), \quad (10)$$

so that $f_0(M, \Sigma) = P(M, \Sigma | Y_{1:T}; K)$ and Equation (8) holds. Then the distribution of random samples after the resampling step at time t will approximately follow the distribution

$$\begin{aligned} \pi_t(M, \Sigma, Z_{1:t}) &\propto f_t(M, \Sigma, Z_{1:t}) \\ &= \left[\frac{P(M, \Sigma | Y_{1:T}; K)}{P(M, \Sigma | K)} \right]^{(T-t)/T} P(M, \Sigma, Z_{1:t}, Y_{1:t} | K) \\ &\propto [P(Y_{1:T} | M, \Sigma; K)]^{(T-t)/T} P(M, \Sigma, Z_{1:t}, Y_{1:t} | K). \end{aligned} \quad (11)$$

Consider the approximation

$$\begin{aligned} [P(Y_{1:T} | M, \Sigma; K)]^{(T-t)/T} &\approx P(Y_{t+1:T} | M, \Sigma; K) \\ &\approx P(Y_{t+1:T} | M, \Sigma, Z_t; K), \end{aligned} \quad (12)$$

where the first approximation is based on the number of observations that can provide information about (M, Σ) in $P(Y_{t+1:T} | M, \Sigma; K)$ and $P(Y_{1:T} | M, \Sigma; K)$. Combining equations (11) and (12), we then have

$$\begin{aligned}\pi_t(M, \Sigma, Z_{1:t}) &\propto P(Y_{t+1:T} | M, \Sigma, Z_t; K) P(M, \Sigma, Z_{1:t}, Y_{1:t} | K) \\ &= P(Y_{t+1:T} | M, \Sigma, Z_{1:t}, Y_{1:t}; K) P(M, \Sigma, Z_{1:t}, Y_{1:t} | K) \\ &= P(M, \Sigma, Z_{1:t}, Y_{1:T} | K) \\ &\propto P(M, \Sigma, Z_{1:t} | Y_{1:T}; K),\end{aligned}$$

which is the “perfect” marginal distribution.

2. *Approximation of $P(M, \Sigma | Y_{1:T}; K)$.* In the intermediate target function (10), $P(M, \Sigma | Y_{1:T}; K)$ cannot be directly obtained. However, with the random samples generated by the MCMC procedure described in Section 3.1, we can obtain an approximation of $P(M, \Sigma | Y_{1:T}; K)$.

Suppose $(M^{*(s)}, \Sigma^{*(s)}, Z_{1:T}^{*(s)})$, $s = 1, \dots, S$, are the random samples generated by the MCMC procedure following the posterior distribution $P(M, \Sigma, Z_{1:T} | Y_{1:T}; K)$. Then we can approximate $P(M, \Sigma | Y_{1:T}; K)$ by

$$\hat{P}(\vec{M}, \Sigma | Y_{1:T}; K) = \hat{P}(\vec{M} | Y_{1:T}; K) \hat{P}(\Sigma | Y_{1:T}; K),$$

where $\hat{P}(\vec{M} | Y_{1:T}; K)$ follows a $N(\mu_{\vec{M}, post}, \Sigma_{\vec{M}, post})$ distribution and $\hat{P}(\Sigma | Y_{1:T}; K)$ is an inverse Wishart distribution $W^{-1}(V_{post}, d_{post})$. We let

$$\begin{aligned}\mu_{\vec{M}, post} &= \frac{1}{S} \sum_{s=1}^S \vec{M}^{*(s)}, \\ \Sigma_{\vec{M}, post} &= \frac{1}{S} \sum_{s=1}^S (\vec{M}^{*(s)} - \mu_{\vec{M}, post})' (\vec{M}^{*(s)} - \mu_{\vec{M}, post}), \\ d_{post} &= n + 3 + \frac{2}{n} \sum_{i=1}^n \hat{E}^2(\Sigma_{i,i}) / \widehat{\text{Var}}(\Sigma_{i,i}), \\ V_{post} &= (d_{post} - n - 1) \frac{1}{S} \sum_{s=1}^S \Sigma^{*(s)},\end{aligned}$$

where $\hat{E}^2(\Sigma_{i,i})$ and $\widehat{\text{Var}}(\Sigma_{i,i})$ are the sample mean and the sample variance of $\Sigma_{i,i}^{(s)}$, $s = 1, \dots, S$, respectively. These parameters are obtained by matching

the mean and the variance of the normal distribution and the inverse Wishart distribution with the sample mean and the sample variance of $M^{*(s)}$ and $\Sigma^{*(s)}$, $s = 1, \dots, S$ (Press, 1982).

3. *Trial Distribution.* At time $t = 0$, we draw samples $(M^{(j)}, \Sigma^{(j)})$ from the trial distribution

$$Q(\vec{M}, \Sigma) = \hat{P}(\vec{M} | Y_{1:T}; K) \hat{P}(\Sigma | Y_{1:T}; K).$$

For ease of computation, we use the state dynamics to generate Z_t , that is, given $(M^{(j)}, \Sigma^{(j)}, Z_{1:t-1}^{(j)})$, $Z_t^{(j)}$ is generated from the distribution

$$Q(Z_t | M^{(j)}, \Sigma^{(j)}, Z_{1:t-1}^{(j)}) = P(Z_t | M^{(j)}, \Sigma^{(j)}, Z_{t-1}^{(j)}; K).$$

4. *MCMC Updating of $(M^{(j)}, \Sigma^{(j)})$.* As noted by Liu and West (2001), Storvik (2002) and many others, the samples of parameters (M, Σ) will degenerate as t increases. As parameters (M, Σ) appear in the state dynamics $P(Z_t | Z_{t-1}, M, \Sigma; K)$ for every t , taking the degeneracy problem into consideration is critical for estimating integration (5).

To circumvent the degeneracy problem, we follow Storvik (2002) and apply a move step after the resampling step to the samples of $(M^{(j)}, \Sigma^{(j)})$. The move step is as follows.

1. Move Step: Update $(M^{(j)}, \Sigma^{(j)})$ with Gibbs sampling following these steps.

- (a) Draw sample $\vec{M}^{(j)}$ from the distribution

$$\begin{aligned} G(\vec{M} | \Sigma^{(j)}, Z_{1:t}^{(j)}) &\propto f_t(\vec{M}, \Sigma^{(j)}, Z_{1:t}^{(j)}) \\ &= \left[\frac{\hat{P}(\vec{M}, \Sigma^{(j)} | Y_{1:T}; K)}{P(\vec{M}, \Sigma^{(j)} | K)} \right]^{(T-t)/T} P(\vec{M}, \Sigma^{(j)}, Z_{1:t}^{(j)}, Y_{1:t} | K). \end{aligned}$$

- (b) Draw sample $\Sigma^{(j)}$ from the distribution

$$\begin{aligned} G(\Sigma | M^{(j)}, Z_{1:t}^{(j)}) &\propto f_t(M^{(j)}, \Sigma, Z_{1:t}^{(j)}) \\ &= \left[\frac{\hat{P}(M^{(j)}, \Sigma | Y_{1:T}; K)}{P(M^{(j)}, \Sigma | K)} \right]^{(T-t)/T} P(M^{(j)}, \Sigma, Z_{1:t}^{(j)}, Y_{1:t} | K). \end{aligned}$$

To save the computing memory, we do not have to record the whole path of the state sample $Z_{1:t}^{(j)}$. As noted by Carvalho et al. (2010), we only need to keep track of the sufficient statistics of $(M^{(j)}, \Sigma^{(j)})$, which is $(\sum_{s=2}^t Z_{s-1}Z'_{s-1}, \sum_{s=2}^t Z_{s-1}Z'_s, \sum_{s=2}^t Z_sZ'_s)$, for the move step.

4 A Simulated Example

We simulate the data using Model (3) with $n = 2$. The true parameters are set as $K = 3$,

$$M = \begin{pmatrix} 0.9 & 0 \\ 0.4 & 0.8 \end{pmatrix}, \quad \text{and} \quad \Sigma = \begin{pmatrix} 1.0 & 0.5 \\ 0.5 & 1.0 \end{pmatrix}.$$

In the simulation, the initial state variables $z_{k,1}$, $k = 1, \dots, K$, are i.i.d. following a normal distribution

$$N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 5.26 & 8.55 \\ 8.55 & 20.32 \end{pmatrix} \right),$$

which is the stationary distribution of the process $z_{k,t} = Mz_{k,t-1} + \xi_{k,t}$. The length of the simulated data is $T = 1000$.

Figure 1 plots one realization of $\{Y_t = (Y_{1,t}, Y_{2,t})', t = 1, \dots, T\}$. The data exhibit heterogeneity due to the persistence parameter in M . Figure 2 plots the underlying volatility matrix series $\Omega_{1:T}$. It is evident that there is a large variation in the variance and covariance.

Given that $K = 3$, we estimate the parameters (M, Σ) and the latent volatility series with the MCMC procedure. In the MCMC procedure, the prior distribution of the initial state variable $z_{k,1}$ is $N(\mathbf{0}, \Sigma_{z,prior})$, where $\Sigma_{z,prior}$ is estimated by $\frac{1}{KT} \sum_{t=1}^T Y_t Y_t'$. The prior distribution of M is set as

$$\begin{pmatrix} M_{11} \\ M_{12} \\ M_{21} \\ M_{22} \end{pmatrix} \sim N \left(\begin{pmatrix} 0.8 \\ 0 \\ 0 \\ 0.8 \end{pmatrix}, \begin{pmatrix} 0.1^2 & 0 & 0 & 0 \\ 0 & 0.5^2 & 0 & 0 \\ 0 & 0 & 0.5^2 & 0 \\ 0 & 0 & 0 & 0.1^2 \end{pmatrix} \right),$$

and the prior distribution of Σ is $W^{-1}(V_{prior}, d_{prior})$, where $d_{prior} = 4$ and

$$V = \Sigma_{z,prior} - M_0 \Sigma_{z,prior} M_0' \quad \text{with} \quad M_0 = \begin{pmatrix} 0.8 & 0 \\ 0 & 0.8 \end{pmatrix}.$$

In this setting, the process $z_{k,t} = M_0 z_{k,t-1} + \xi_{k,t}$ with innovation $\xi_{k,t} \sim N(\mathbf{0}, E(\Sigma))$ will have the stationary distribution $N(\mathbf{0}, \Sigma_{z,prior})$.

We draw 25,000 samples with the MCMC procedure and the first 5,000 samples are thrown away as the burn-in period. The prior distributions and the posterior distributions of the parameters are presented in Figure 3 and Figure 4, respectively. The marginal prior distribution of Σ is obtained using random samples drawn from the joint prior distribution. We find that the posterior distributions are more concentrated around the true values of the parameters than those of the prior distributions. The time-varying correlation of $Y_{1,t}$ and $Y_{2,t}$ is defined as

$$\rho_t = \frac{\omega_{1,2,t}}{\sqrt{\omega_{1,1,t}\omega_{2,2,t}}}, \quad \text{where} \quad \Omega_t = \begin{bmatrix} \omega_{1,1,t} & \omega_{1,2,t} \\ \omega_{2,1,t} & \omega_{2,2,t} \end{bmatrix}.$$

Figure 5 shows the simulated and estimated correlation series. The estimated correlation series is smoother than the simulated correlation series, but still captures the time-varying structure of the simulated correlation series.

We repeat the experiment 100 times and perform the MCMC estimate procedure when $K = 2, 3, 4, 5$ for each data set. We use the mean of the random samples generated by the MCMC procedure as the estimation of the parameter. Figure 6 and Figure 7 report the estimate results for the 100 data sets in boxplots. The estimations of M are less biased under $K = 3$ and $K = 4$ than when $K = 2$ and $K = 5$. The estimated Σ gets smaller as K increases. This is because the WAR(1) process satisfies

$$E(\Omega_t | \Omega_{t-1}) = K\Sigma + M\Omega_{t-1}M'$$

as shown in Equation (2).

We then use the SMC procedure to calculate the Bayesian posterior probabilities. The number of random samples we use is $m = 100,000$. Figure 8 compares the logarithm of $P(K | Y_{1:T})$ under $K = 2, 4, 5$ with the logarithm of $P(K | Y_{1:T})$ under the true model $K = 3$. The zero lines indicate equal likelihoods of two models. Negative values indicate that alternative models have lower likelihoods than the true model, such that the true model with $K = 3$ is selected. It shows that the SMC procedure selects the true model

95 times out of 100 data sets. For the other 5 data sets, it selects the model with $K = 4$.

5 Empirical Study of Time-varying Correlation Between Shanghai and New York Stock Markets

We apply our method to analyze the correlation between Shanghai and New York stock returns. The time variation of stochastic volatility in both stock markets has been documented in numerous studies. Less studied is the time-varying correlation between the two markets. Recently, Chow et al. (2011) model the comovement of Shanghai and New York stock returns with time-varying regressive coefficients. They find that integration between the two markets strengthened considerably after 2002 when China joined WTO. The effect of the current return of one market on the other also became significantly positive and increased after 2002. The upward trend was interrupted during the recent global financial crisis, but recovered after 2008. Since China is opening up its capital market gradually which is rapidly gaining ground in the international capital market, optimal portfolio holdings of Chinese stocks in an international portfolio is important for investors all over the world. Time-varying correlation between assets under management is an important factor in determining optimal holdings in a dynamic setting.

Using a VAR process in continuous time for the volatility-covolatility matrix of asset returns, Buraschi et al. (2010) find that there are distinct hedging components against both stochastic volatility and correlation risk for optimal portfolio construction. They find that under time-varying correlation risk, the hedging demand is typically substantially larger than in univariate models or models with constant correlation. Their paper illustrates the theoretical results by estimating a two-dimensional model with future returns for the S&P500 Index and 30-year Treasury bonds. They use high frequency data to approximate the volatility-covolatility matrix and use the generalized method of moments (GMM) to infer parameter values. As

discussed before, our MCMC procedure is more efficient in the joint inference of parameters and the latent volatility-covolatility process which does not rely on approximation through realized or implied volatility. Moreover, it is essential in a real-time dynamic setting of portfolio rebalancing in the framework of Buraschi et al. (2010).

We model the time-varying volatility and covolatility with a WAR(1) process for Shanghai and New York stock returns. We take the log difference of weekly data from the Shanghai Stock Exchange Composite Index and the New York Stock Composite Index, respectively. We examine weekly returns from 1992-01-27 to 2010-12-27, a total of 956 observations after excluding holidays and missing data in either market. We multiply the log difference with 100 and de-mean the return series, then we apply our method to estimate Model (3).

Figure 9 shows the logarithm of the Bayesian posterior probability $P(K | Y_{1:T})$ estimated by the SMC procedure with sample size $m = 200,000$. The result suggests that the degrees of freedom of the model should be around 4.

When $K = 4$, the prior distribution and the posterior distributions of the parameters estimated with the MCMC procedure are presented in Figure 10 and Figure 11. The posterior means of M and Σ are

$$\widehat{M} = \begin{pmatrix} 0.9691 & 0.0074 \\ -0.0004 & 0.9811 \end{pmatrix} \quad \text{and} \quad \widehat{\Sigma} = \begin{pmatrix} 0.4071 & -0.0003 \\ -0.0003 & 0.0529 \end{pmatrix}.$$

Figure 12 plots the estimated correlation between Shanghai and New York stock returns. In the first half of the sample, up to 2001, the correlation between Shanghai and New York stock returns is relatively low, mostly within the range of $[-0.4, 0.4]$ and is concentrated within the range of $[-0.2, 0.2]$ with a slightly negative mean. This is consistent with the findings of Chow and Lawler (2003). However, after 2002 when China joined WTO, the correlation stays more positive and trends upward, reflecting the gradual integration of China's capital market into the world with the United States as the leading market. During the recent financial crisis, starting from the end of 2007, the correlation began to fluctuate wildly towards negative. After the crisis, it quickly resumes the previous positive stance and trends upward again.

Our result captures the main features of the comovement between the two markets as shown in Chow et al. (2011). With the inference of volatility and covolatility modeled by the WAR(1) process, dynamic portfolio allocation can be implemented with hedging both for volatility and covolatility in the framework of Buraschi et al. (2010).

6 Conclusion

The WAR-SV model is analytically convenient for studying asset pricing and optimal portfolio allocation problems with multivariate stochastic volatility and covolatility. But due to the lack of a closed-form transition density, the model is difficult to estimate in applications with latent volatility process. Based on an alternative representation of the WAR process with lag order $p = 1$ and integer degrees of freedom, we develop an efficient two-step procedure to jointly estimate parameters and the latent volatility-covolatility series. With both simulated and real data, we show the effectiveness of this procedure.

The method can be adapted to higher dimension problems or problems with more complex structures of return dynamics. Existing theoretical models using the WAR process can thus be effectively applied to economic and financial problems under time-varying volatilities and changing correlations.

Appendix A: MCMC Procedure

Here we provide the details on each updating step in the MCMC procedure.

1. Update $\vec{M}^{*(s)}$: Draw $\vec{M}^{*(s)}$ from the distribution

$$\begin{aligned}
& P\left(\vec{M} \mid \Sigma^{*(s-1)}, Z_{1:T}^{*(s-1)}, Y_{1:T}; K\right) \\
& \propto P\left(\vec{M} \mid K\right) \prod_{t=2}^T P\left(Z_t^{*(s-1)} \mid Z_{t-1}^{*(s-1)}, \vec{M}; K\right) \\
& \propto \exp\left\{-\frac{1}{2}\left(\vec{M} - \mu_{\vec{M}, \text{prior}}\right)' \Sigma_{\vec{M}, \text{prior}}^{-1}\left(\vec{M} - \mu_{\vec{M}, \text{prior}}\right)\right\} \\
& \quad \times \exp\left\{-\frac{1}{2} \sum_{t=2}^T \sum_{k=1}^K \left(z_{k,t}^{*(s-1)} - \{I_n \otimes z_{k,t-1}^{*(s-1)}\}' \vec{M}\right)' \right. \\
& \quad \left. \left(\Sigma^{*(s-1)}\right)^{-1} \left(z_{k,t}^{*(s-1)} - \{I_n \otimes z_{k,t-1}^{*(s-1)}\}' \vec{M}\right)\right\} \\
& \sim N\left(A_M^{-1} B_M, A_M^{-1}\right),
\end{aligned}$$

where $I_n \otimes z_{k,t-1}^{*(s-1)}$ is the Kronecker product of the identity matrix I_n and vector $z_{k,t-1}^{*(s-1)}$, and

$$\begin{aligned}
A_M &= \Sigma_{\vec{M}, \text{prior}}^{-1} + \sum_{t=2}^T \sum_{k=1}^K \{I_n \otimes z_{k,t-1}^{*(s-1)}\} \left(\Sigma^{*(s-1)}\right)^{-1} \{I_n \otimes z_{k,t-1}^{*(s-1)}\}', \\
B_M &= \Sigma_{\vec{M}, \text{prior}}^{-1} \mu_{\vec{M}, \text{prior}} + \sum_{t=2}^T \sum_{k=1}^K \{I_n \otimes z_{k,t-1}^{*(s-1)}\} \left(\Sigma^{*(s-1)}\right)^{-1} z_{k,t}.
\end{aligned}$$

2. Update $\Sigma^{*(s)}$: Draw $\Sigma^{*(s)}$ from the distribution

$$\begin{aligned}
& P\left(\Sigma \mid M^{*(s)}, Z_{1:T}^{*(s-1)}, Y_{1:T}; K\right) \\
& \propto P\left(\Sigma \mid K\right) \prod_{t=2}^T \prod_{k=1}^K P\left(Z_{k,t}^{*(s-1)} \mid Z_{k,t-1}^{*(s-1)}, \Sigma, M^{*(s)}; K\right) \\
& \propto |\Sigma|^{-(d_{\text{prior}}+n+1)/2} \exp\left\{-\text{Tr}\left(V_{\text{prior}} \Sigma^{-1}\right) / 2\right\} |\Sigma|^{-(T-1)K/2} \\
& \quad \times \exp\left\{-\sum_{t=2}^T \sum_{k=1}^K \left(z_{k,t}^{*(s-1)} - M^{*(s)} z_{k,t-1}^{*(s-1)}\right)' \Sigma^{-1} \left(z_{k,t}^{*(s-1)} - M^{*(s)} z_{k,t-1}^{*(s-1)}\right) / 2\right\} \\
& \sim W^{-1}\left\{V_\Sigma, d_\Sigma\right\},
\end{aligned}$$

where

$$\begin{aligned}
V_\Sigma &= V_{\text{prior}} + \sum_{t=2}^T \sum_{k=1}^K \left(z_{k,t}^{*(s-1)} - M^{*(s)} z_{k,t-1}^{*(s-1)}\right) \left(z_{k,t}^{*(s-1)} - M^{*(s)} z_{k,t-1}^{*(s-1)}\right)' \\
d_\Sigma &= d_{\text{prior}} + (T-1)K.
\end{aligned}$$

3. Update $Z_t^{*(s)}$: For $t = 1, \dots, T$, draw $Z_t^{*(s)}$ from the distribution

$$\begin{aligned} & G\left(Z_t \mid \vec{M}^{*(s)}, \Sigma^{*(s)}, Z_{1:t-1}^{*(s)}, Z_{t+1:T}^{*(s-1)}, Y_{1:T}; K\right) \\ &= \prod_{k=1}^K G\left(z_{k,t} \mid \vec{M}^{*(s)}, \Sigma^{*(s)}, Z_{1:t-1}^{*(s)}, Z_{t+1:T}^{*(s-1)}, Y_{1:T}; K\right), \end{aligned}$$

where for $t = 2, \dots, T-1$,

$$\begin{aligned} & G\left(z_{k,t} \mid \vec{M}^{*(s)}, \Sigma^{*(s)}, Z_{1:t-1}^{*(s)}, Z_{t+1:T}^{*(s-1)}, Y_{1:T}; K\right) \\ & \propto P(z_{t+1,k}^{*(s-1)} \mid M^{*(s)}, \Sigma^{*(s)}, z_{k,t}; K) P(z_{k,t} \mid M^{*(s)}, \Sigma^{*(s)}, z_{k,t-1}^{*(s)}; K) \\ & \propto \exp\left\{-\frac{1}{2}\left(z_{t+1,k}^{*(s-1)} - M^{*(s)} z_{k,t}\right)' \left(\Sigma^{*(s)}\right)^{-1} \left(z_{t+1,k}^{*(s-1)} - M^{*(s)} z_{k,t}\right)\right\} \\ & \quad \times \exp\left\{-\frac{1}{2}\left(z_{k,t} - M^{*(s)} z_{k,t-1}^{*(s)}\right)' \left(\Sigma^{*(s)}\right)^{-1} \left(z_{k,t} - M^{*(s)} z_{k,t-1}^{*(s)}\right)\right\} \\ & \sim N\left(A_Z^{-1} B_Z, A_Z^{-1}\right) \end{aligned}$$

with

$$\begin{aligned} A_Z &= \left(M^{*(s)}\right)' \left(\Sigma^{*(s)}\right)^{-1} M^{*(s)} + \left(\Sigma^{*(s)}\right)^{-1}, \\ B_Z &= \left(M^{*(s)}\right)' \left(\Sigma^{*(s)}\right)^{-1} z_{t+1,k}^{*(s-1)} + \left(\Sigma^{*(s)}\right)^{-1} \left(M^{*(s)} z_{k,t-1}^{*(s)}\right). \end{aligned}$$

When $t = 1$, we have

$$\begin{aligned} A_Z &= \left(M^{*(s)}\right)' \left(\Sigma^{*(s)}\right)^{-1} M^{*(s)} + \Sigma_{z,prior}^{-1}, \\ B_Z &= \left(M^{*(s)}\right)' \left(\Sigma^{*(s)}\right)^{-1} z_{t+1,k}^{*(s-1)} + \Sigma_{z,prior}^{-1} \mu_{z,prior}, \end{aligned}$$

and when $t = T$,

$$\begin{aligned} A_Z &= \left(\Sigma^{*(s)}\right)^{-1}, \\ B_Z &= \left(\Sigma^{*(s)}\right)^{-1} \left(M^{*(s)} z_{k,t-1}^{*(s)}\right). \end{aligned}$$

Then the generated $Z_t^{*(s)}$ is accepted as the new sample with probability

$$p = \min\left\{1, \frac{P\left(Y_t \mid Z_t^{*(s)}; K\right)}{P\left(Y_t \mid Z_t^{*(s-1)}; K\right)}\right\}.$$

Otherwise, we let $Z_t^{*(s)} = Z_t^{*(s-1)}$.

Appendix B: Algorithmic Steps of the SMC Procedure

In the SMC procedure, we choose the intermediate target function as

$$f_t(M, \Sigma, Z_{1:t}) = \left[\frac{\widehat{P}(M, \Sigma | Y_{1:T}; K)}{P(M, \Sigma | K)} \right]^{(T-t)/T} P(M, \Sigma, Z_{1:t}, Y_{1:t} | K)$$

for $t = 0, 1, \dots, T$. The algorithmic steps are as follows.

1. At time $t = 0$, draw $(M^{(j)}, \Sigma^{(j)})$ from the trial distribution

$$Q(M, \Sigma) = \widehat{P}(M, \Sigma | Y_{1:T}; K)$$

and let $w_0^{(j)} = 1$.

2. At times $t = 1, 2, \dots, T$,

- (a) Sampling: Draw $Z_t^{(j)}$ from the trial distribution

$$Q\left(Z_t | M^{(j)}, \Sigma^{(j)}, Z_{1:t-1}^{(j)}\right) = P\left(Z_t | M^{(j)}, \Sigma^{(j)}, Z_{1:t-1}^{(j)}\right).$$

- (b) Updating Weights: Update the weights by letting

$$\begin{aligned} w_t^{(j)} &= w_{t-1}^{(j)} \frac{f_t\left(M^{(j)}, \Sigma^{(j)}, Z_{1:t}^{(j)}\right)}{f_{t-1}\left(M^{(j)}, \Sigma^{(j)}, Z_{1:t-1}^{(j)}\right) Q\left(Z_t^{(j)} | M^{(j)}, \Sigma^{(j)}, Z_{1:t-1}^{(j)}\right)} \\ &= \left[\frac{\widehat{P}(M^{(j)}, \Sigma^{(j)} | Y_{1:T}; K)}{P(M^{(j)}, \Sigma^{(j)} | K)} \right]^{-1/T} P\left(Y_t | Z_t^{(j)}, M^{(j)}, \Sigma^{(j)}; K\right). \end{aligned}$$

- (c) Resampling: Denote $(M, \Sigma, Z_{1:t})$ by $Z_{0:t}$. When $t < T$,

- i. draw $Z_{0:t}^{new(j)}$, $j = 1, \dots, m$, from the distribution

$$\sum_{j=1}^m \frac{w_t^{(j)}}{\sum_{k=1}^m w_t^{(k)}} \delta\left(Z_{0:t} - Z_{0:t}^{(j)}\right),$$

and set $w_t^{new(j)} = \frac{1}{m} \sum_{j=1}^m w_t^{(j)}$;

- ii. let $Z_{0:t}^{(j)} = Z_{0:t}^{new(j)}$ and $w_t^{(j)} = w_t^{new(j)}$ for $j = 1, \dots, m$.

(d) Move Step: When $1 < t < T$, update $(M^{(j)}, \Sigma^{(j)})$ as follows.

i. Draw sample $\vec{M}^{(j)}$ from the distribution

$$\begin{aligned}
& G(\vec{M} \mid \Sigma^{(j)}, Z_{1:t}^{(j)}) \\
& \propto f_t(\vec{M}, \Sigma^{(j)}, Z_{1:t}^{(j)}) \\
& = \left[\frac{\hat{P}(\vec{M}, \Sigma^{(j)} \mid Y_{1:T}; K)}{P(\vec{M}, \Sigma^{(j)} \mid K)} \right]^{(T-t)/T} P(\vec{M}, \Sigma^{(j)}, Z_{1:t}^{(j)}, Y_{1:t} \mid K) \\
& \propto \left[\frac{\hat{P}(\vec{M}, \Sigma^{(j)} \mid Y_{1:T}; K)}{P(\vec{M}, \Sigma^{(j)} \mid K)} \right]^{(T-t)/T} P(\vec{M} \mid K) \prod_{s=2}^t P(Z_s^{(j)} \mid Z_{s-1}^{(j)}, \vec{M}, \Sigma^{(j)}; K) \\
& \sim N(A^{-1}B, A^{-1}),
\end{aligned}$$

where

$$\begin{aligned}
A &= (t/T) \left[\Sigma_{\vec{M}, \text{prior}} \right]^{-1} + (1 - t/T) \left[\Sigma_{\vec{M}, \text{post}} \right]^{-1} + \left[\Sigma^{(j)} \right]^{-1} \otimes \sum_{s=2}^t Z_{s-1} Z'_{s-1}, \\
B &= (t/T) \left[\Sigma_{\vec{M}, \text{prior}} \right]^{-1} \mu_{\vec{M}, \text{prior}} + (1 - t/T) \left[\Sigma_{\vec{M}, \text{post}} \right]^{-1} \mu_{\vec{M}, \text{post}} \\
&\quad + \begin{pmatrix} \sum_{s=2}^t Z_{s-1} Z'_s \left[\Sigma^{(j)} \right]^{-1}_{.1} \\ \vdots \\ \sum_{s=2}^t Z_{s-1} Z'_s \left[\Sigma^{(j)} \right]^{-1}_{.n} \end{pmatrix}.
\end{aligned}$$

ii. Draw sample $\Sigma^{(j)}$ from the distribution

$$\begin{aligned}
& G(\Sigma \mid M^{(j)}, Z_{1:t}^{(j)}) \\
& \propto f_t(M^{(j)}, \Sigma, Z_{1:t}^{(j)}) \\
& = \left[\frac{\hat{P}(M^{(j)}, \Sigma \mid Y_{1:T}; K)}{P(M^{(j)}, \Sigma \mid K)} \right]^{(T-t)/T} P(M^{(j)}, \Sigma, Z_{1:t}^{(j)}, Y_{1:t} \mid K) \\
& \propto \left[\frac{\hat{P}(M^{(j)}, \Sigma \mid Y_{1:T}; K)}{P(M^{(j)}, \Sigma \mid K)} \right]^{(T-t)/T} P(\Sigma \mid K) \prod_{s=2}^t P(Z_s^{(j)} \mid Z_{s-1}^{(j)}, M^{(j)}, \Sigma; K) \\
& \sim W^{-1} \{V, d\},
\end{aligned}$$

where

$$d = (t/T)d_{\text{prior}} + (1 - t/T)d_{\text{post}} + (t - 1)K$$

and

$$\begin{aligned}
V &= (t/T)V_{prior} + (1 - t/T)V_{post} \\
&+ \sum_{s=2}^t Z_s Z'_s - M \sum_{s=2}^t Z_{s-1} Z'_s - \sum_{s=2}^t Z_s Z'_{s-1} M' + M \sum_{s=2}^t Z_{s-1} Z'_{s-1} M'.
\end{aligned}$$

3. The Bayesian posterior probability is estimated by

$$P(K | Y_{1:T}) \propto P(K) \frac{1}{m} \sum_{j=1}^m w_T^{(j)}.$$

Acknowledgements

We thank seminar participants at WISE, Xiamen University for useful comments, the WISE Information Center for computational services, and acknowledge financial support from National Science Foundation of China (70903053 and 11101341) and Central Universities Research Funding of China (2010221093).

References

- Asai, M., and M. McAleer, 2009. The structure of dynamic correlation in multivariate stochastic volatility models, *Journal of Econometrics* 150, 182–192.
- Asai, M., M. McAleer, and J. Yu, 2006. Multivariate stochastic volatility: A review, *Econometric Reviews* 25, 145–175.
- Bollerslev, T., 1990. Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH approach, *Review of Economics and Statistics* 72, 498–505.
- Bollerslev, T., R.F. Engle, and J. Woodridge, 1988. A capital asset pricing model with time varying covariances, *Journal of Political Economy* 96, 116–131.

- Buraschi, A., A. Cieslak, and F. Trojani, 2008. Correlation risk and term structure of interest rates, Imperial College London and University of St. Gallen.
- Buraschi, A., P. Porchia, and F. Trojani, 2010. Correlation risk and optimal portfolio choice, *Journal of Finance* 65, 393–420.
- Carpenter, J., P. Clifford, and P. Fearnhead, 1999. An improved particle for non-linear problems, *IEE Proceedings on Radar, Sonar, and Navigation*, 146, 2–7.
- Carvalho, C.M., M. Johannes, H.F. Lopes, and N. Polson, 2010. Particle learning and smoothing, *Statistical Science* 25, 88–106.
- Chow, G., and C. Lawler, 2003. A time series analysis of Shanghai and New York stock price indices, *Annals of Economics and Finance* 4, 17–35.
- Chow, G., C. Liu, and L. Niu, 2011. Co-movements of Shanghai and New York stock prices by time-varying regressions, *Journal of Comparative Economics* 39, 577–583.
- Christodoulakis, G., and S. Satchell, 2002. Correlated ARCH: modelling the time-varying correlation between financial asset returns, *European Journal of Operations Research* 139, 351–370.
- Cieslak, A., and P. Povala, 2009. Understanding the term structure of yield curve volatility, University of Lugano.
- Cox, J., J. Ingersoll, and S. Ross, 1985. A theory of the term structure of interest rates, *Econometrica* 53, 385–407.
- Crisan, D., and T. Lyons, 2002. Minimal entropy approximations and optimal algorithms, *Monte Carlo Methods and Applications* 8, 343–355.
- Da Fonseca, J., M. Grasselli, and C. Tebaldi, 2007. Option pricing when correlations are stochastic: an analytical framework, *Review of Derivatives Research* 10, 151–180.

- 2008. A multi-factor volatility Heston model, *Quantitative Finance* 8, 591–604.
- Durbin, J., and S. Koopman, 1997. Monte Carlo maximum likelihood estimation for non-Gaussian state space models, *Biometrika* 84, 669–684.
- Engle, R., 2002. Dynamic conditional correlation – a simple class of multivariate GARCH models, *Journal of Business & Economic Statistics* 20, 339–350.
- Engle, R.F., and K.F. Kroner, 1995. Multivariate simultaneous generalized ARCH, *Econometric Theory* 11, 122–150.
- Gourieroux, C., 2006. Continuous time Wishart process for stochastic risk, *Econometric Reviews* 25, 177–217.
- Gourieroux, C., and R. Sufana, 2003. Wishart quadratic term structure models, CREST, CEPREMAP and University of Toronto.
- 2010. Derivative pricing with Wishart multivariate stochastic volatility, *Journal of Business & Economic Statistics* 28, 438–451.
- Gourieroux, C., J. Jasiak, and R. Sufana, 2009. The Wishart autoregressive process of multivariate stochastic volatility, *Journal of Econometrics* 150, 167–181.
- Green, P. J., 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination, *Biometrika* 82, 711–732.
- 2001. Markov chain Monte Carlo methods for computing Bayes factors: A comparative review, *Journal of the American Statistical Association* 96, 1122–1132.
- Harvey, A.C., 1989. Time-varying conditional covariances in tests of asset pricing models, *Journal of Financial Economics* 24, 289–317.
- Harvey, A.C., E. Ruiz, and N. Shephard, 1994. Multivariate stochastic variance models, *Review of Economic Studies* 61, 247–264.

- Jacquier, E., N. G. Polson, and P. E. Rossi, 1994. Bayesian analysis of stochastic volatility models, *Journal of Business & Economic Statistics* 12, 371–417.
- Kass, R. E., and A. E. Raftery, 1995. Bayes factors, *Journal of the American Statistical Association* 90, 773–795.
- Kitagawa, G., 1996. Monte Carlo filter and smoother for non-Gaussian non-linear state space models, *Journal of Computational and Graphical Statistics* 5, 1–25.
- Liu, J., and M. West, 2001. Combined parameter and state estimation in simulation-based filtering, in: A. Doucet J. F. G. de Freitas and N. J. Gordon, eds, *Sequential Monte Carlo in Practice*(Springer-Verlag, New York).
- Liu, J. S., 2001. *Monte Carlo strategies in scientific computing* (Springer-Verlag, New York).
- Liu, J.S., and R. Chen, 1998. Sequential Monte Carlo methods for dynamic systems, *Journal of the American Statistical Association* 93, 1032–1044.
- McAleer, M., 2005. Automated inference and learning in modeling financial volatility, *Econometric Theory* 21, 232–261.
- Niu, L., 2009. An affine term structure model with auxiliary stochastic volatility-covolatility, Xiamen University.
- Philipov, A., and M.E. Glickman, 2006. Multivariate stochastic volatility via Wishart processes, *Journal of Business & Economic Statistics* 24, 313–328.
- Pitt, M. K., 2002. Smooth particle filters for likelihood and maximisation, University of Warwick.
- Press, S. J., 1982. *Applied multivariate analysis* (Dover Publications, New York).

- Primiceri, G., 2005. Time varying structural vector autoregressions and monetary policy, *The Review of Economic Studies* 72, 821–852.
- Robert, R., and G. Casella, 1999. Monte Carlo statistical methods (Springer, New York).
- Shephard, N., and M. Pitt, 1997. Likelihood analysis of non-Gaussian measurement time series, *Biometrika* 84, 653–667.
- Storvik, G., 2002. Particle filters in state space models with the presence of unknown static parameters, *IEEE Transaction of Signal Processing* 50, 281–289.
- Tse, Y., and A. Tsui, 2002. A multivariate GARCH model with time-varying correlations, *Journal of Business & Economic Statistics* 20, 351–362.
- Yu, J., and R. Meyer, 2006. Multivariate stochastic volatility models: Bayesian estimation and model comparison, *Econometric Reviews* 25, 361–384.

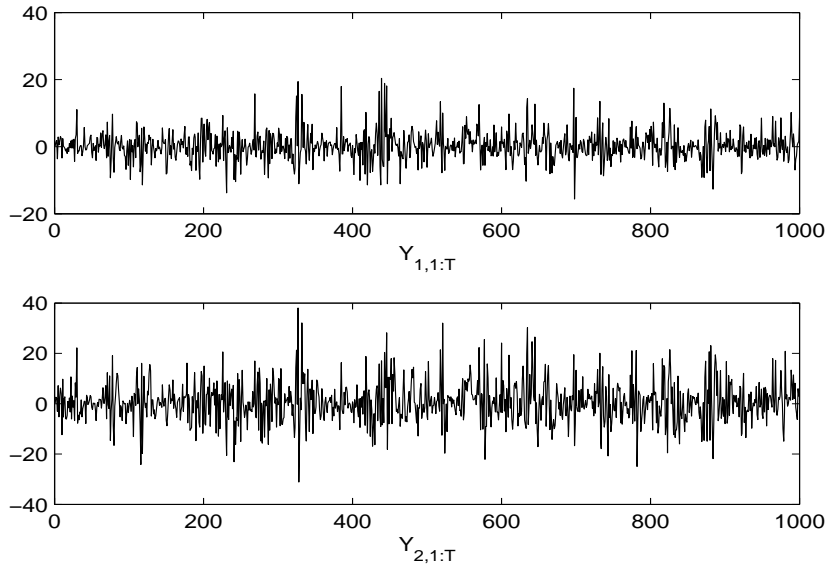


Figure 1: Simulated $Y_{1,1:T}$ and $Y_{2,1:T}$.

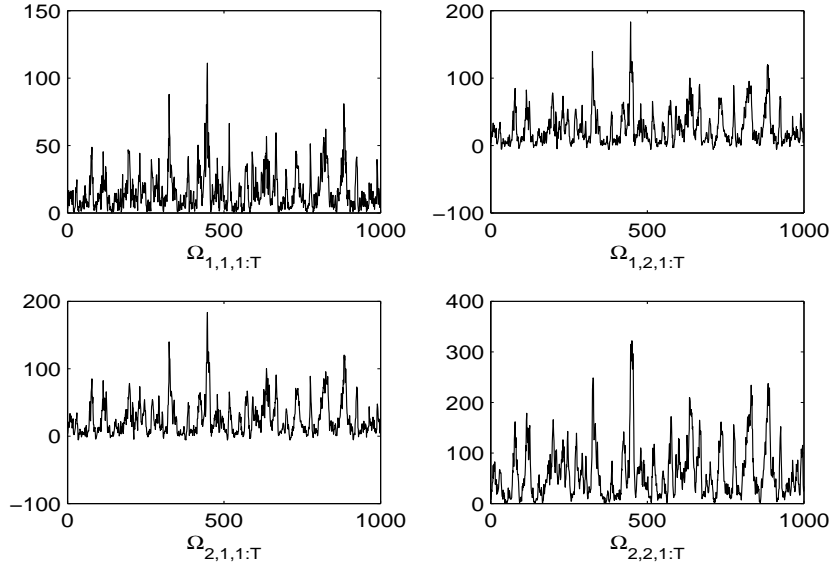


Figure 2: Simulated variance matrix $\Omega_{1:T}$.

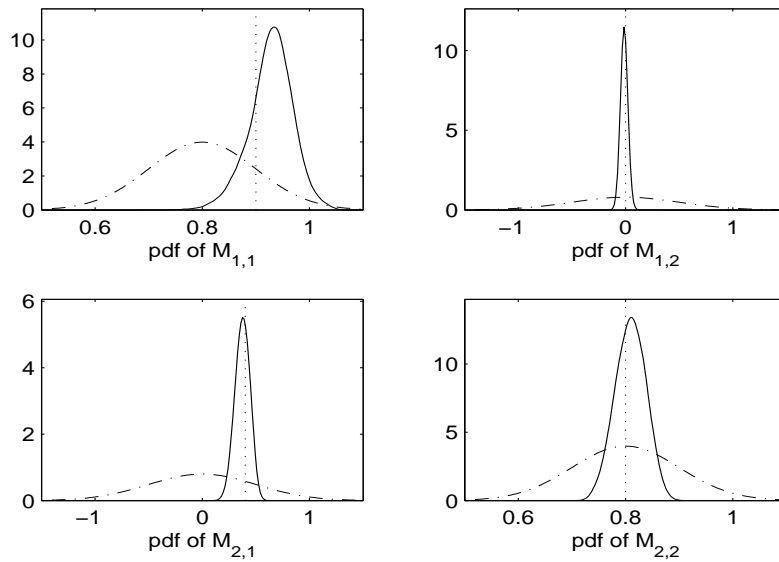


Figure 3: Prior distribution (dashed line) and posterior distribution (solid line) of M . The vertical dotted line represents the true value of the parameter.

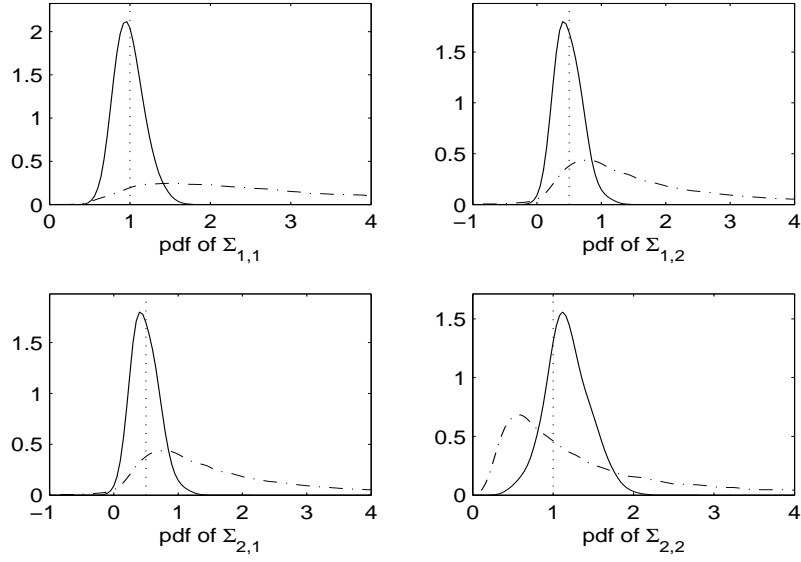


Figure 4: Prior distribution (dashed line) and posterior distribution (solid line) of Σ . The vertical dotted line represents the true value of the parameter.

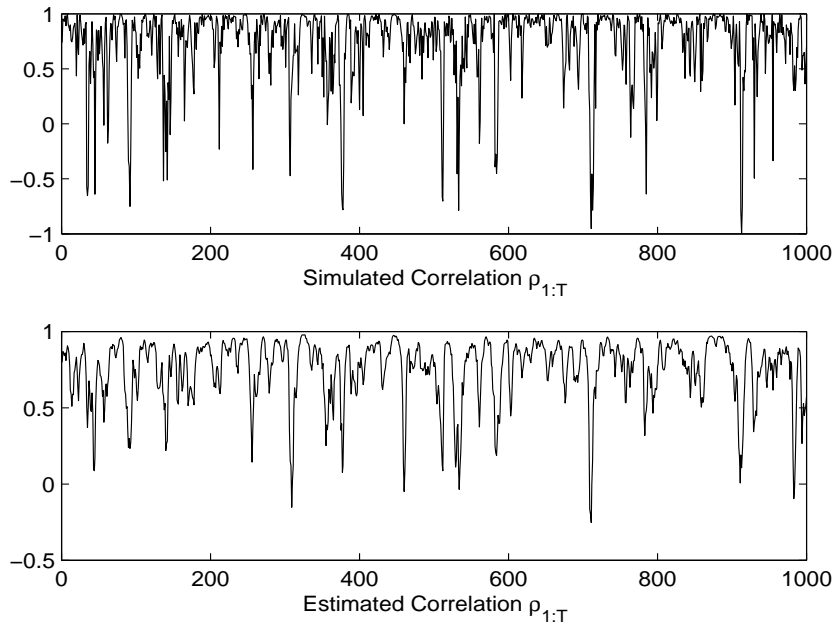


Figure 5: Simulated and estimated correlation series.

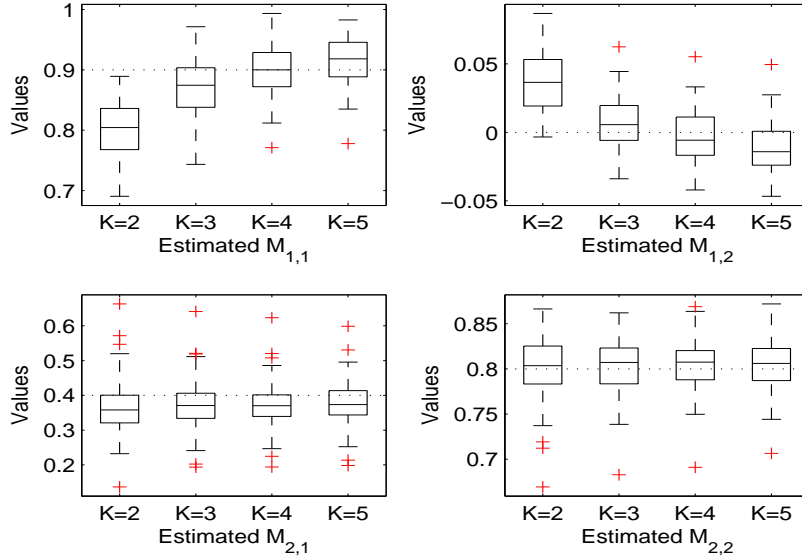


Figure 6: Boxplots of estimated M for 100 data sets. The dotted line represents the true value of the parameter.

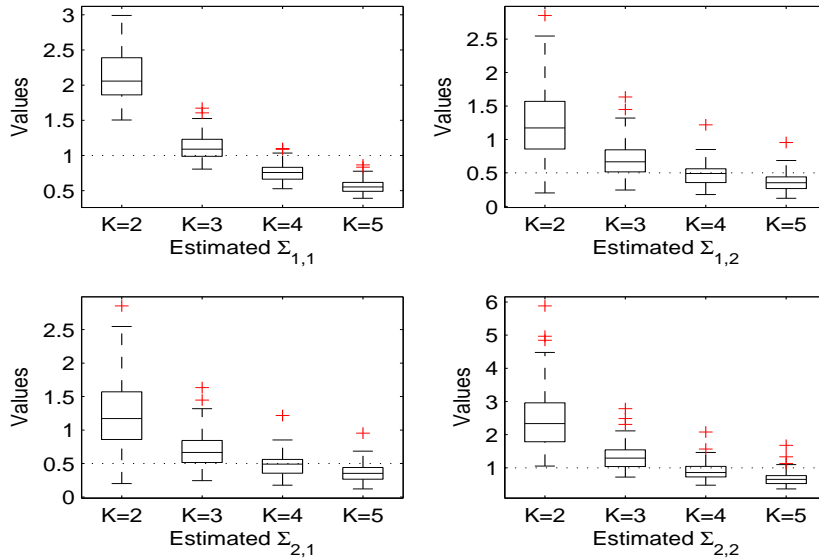


Figure 7: Boxplots of estimated Σ for 100 data sets. The dotted line represents the true value of the parameter.

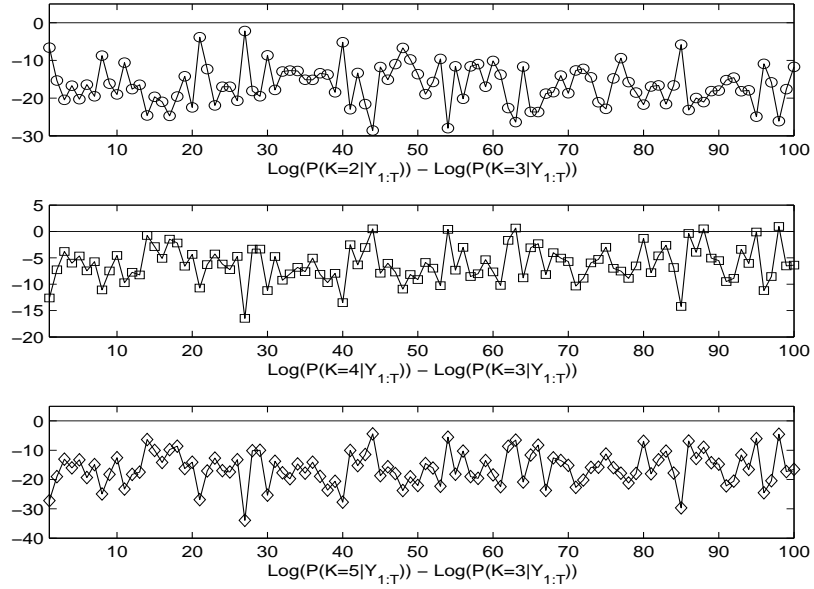


Figure 8: The difference between $\log(P(K | Y_{1:T}))$ under $K = 2, 4, 5$ and $\log(P(K | Y_{1:T}))$ under the true model $K = 3$.

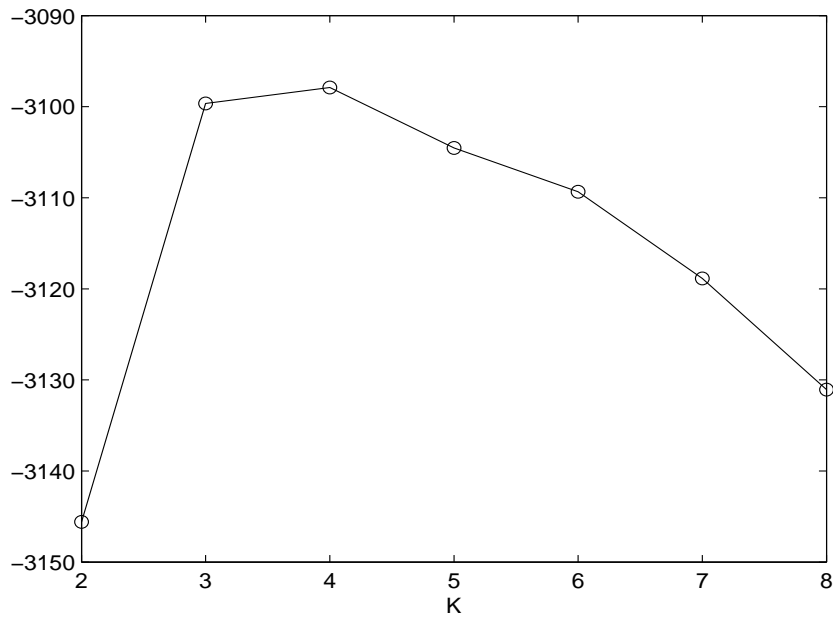


Figure 9: Logarithm of the Bayesian posterior probabilities for different K .

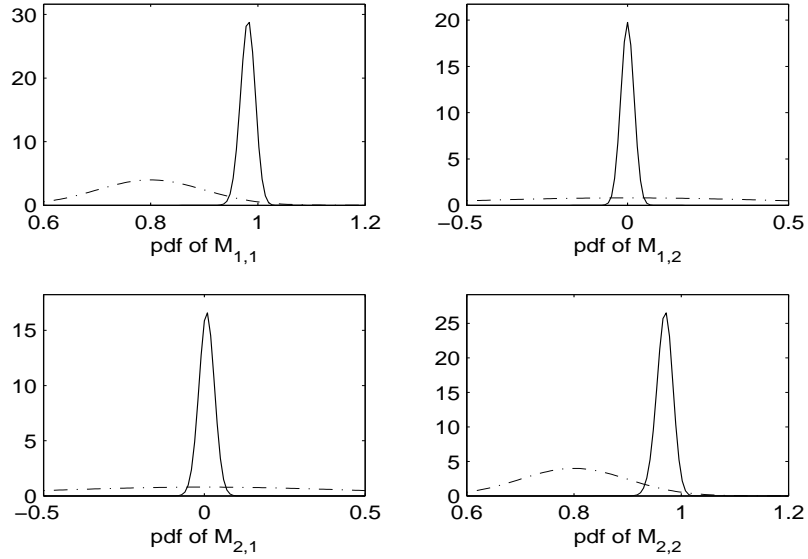


Figure 10: Prior distribution (dashed line) and posterior distribution (solid line) of M .

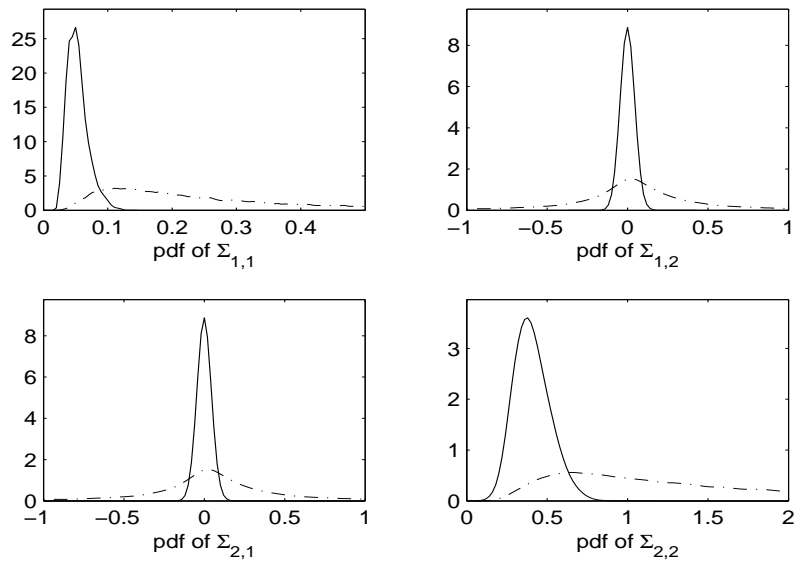


Figure 11: Prior distribution (dashed line) and posterior distribution (solid line) of Σ .

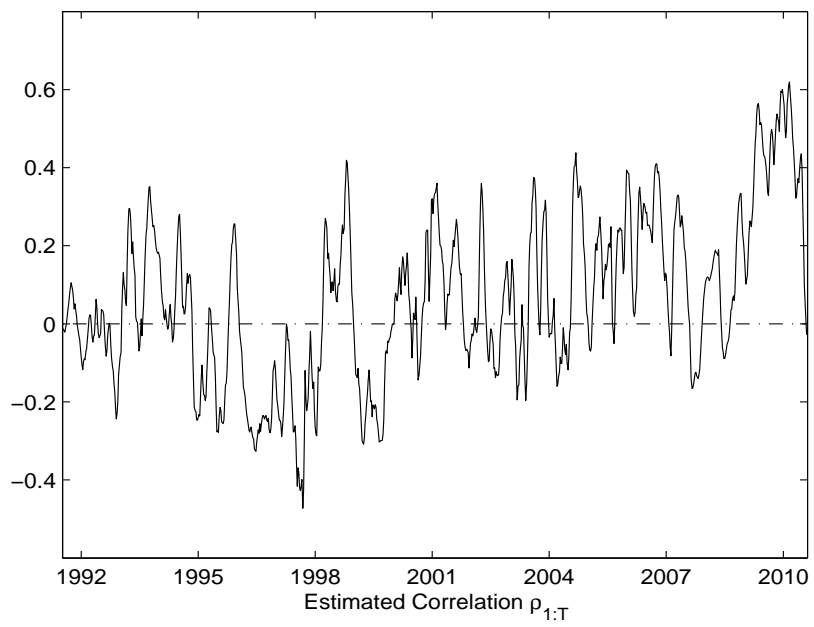


Figure 12: Time-varying correlation between Shanghai and New York stock returns .