

Selection of Multivariate Stochastic Volatility Models via Bayesian Stochastic Search

Antonello Loddo, Capital One Financial Corporation

Shawn Ni, Department of Economics, University of Missouri

Dongchu Sun, Department of Statistics, University of Missouri

Abstract

We propose a Bayesian stochastic search approach to selecting restrictions on multivariate regression models where the errors exhibit deterministic or stochastic conditional volatilities. We develop a Markov Chain Monte Carlo (MCMC) algorithm that generates posterior restrictions on the regression coefficients and Cholesky decompositions of the covariance matrix of the errors. Numerical simulations with artificially generated data show that the proposed method is effective in selecting the data-generating model restrictions and improving the forecasting performance of the model. Applying the method to daily foreign exchange rate data, we conduct stochastic search on a VAR model with stochastic conditional volatilities.

KEY WORDS: Bayesian VAR, Particle Filter, Markov Chain Monte Carlo, Model Selection.

JEL CLASSIFICATIONS: C11, C15, C32.

1 Introduction

For every empirical problem in economics, researchers must choose an appropriate model for the data at hand. Selecting among a given class of models, the researcher will contemplate the trade off between the scope of the model and the goodness of fit. If one considers a model space that includes all potential models of interest, then the model choice becomes restricting the model space (i.e., obtaining a submodel). Appropriate restrictions (submodels) often give rise to sharper finite sample inference and better forecasts. In this paper, we consider a multivariate linear regression framework where variables are potentially correlated contemporaneously as well as serially. In addition, some variables may exhibit time-varying, even stochastic, conditional volatilities. The unrestricted model in general consists of over-parameterized regression and stochastic volatility equations. The main objective of the paper is to develop a stochastic search model selection procedure that jointly restricts conditional volatilities and parameters in the regression models.

We consider the following p -vector multivariate regression model:

$$\mathbf{y}_t = \mathbf{b}_0 + \mathbf{B}_1 \mathbf{x}_t + \mathbf{e}_t, \quad (1)$$

for $t = 1, \dots, T$, where $\mathbf{y}_t = (y_{1t}, \dots, y_{pt})'$ is a $p \times 1$ -vector of variables, \mathbf{b}_0 is a $p \times 1$ -vector of unknown parameters, \mathbf{x}_t is a $q \times 1$ -vector of known endogenous or exogenous variables, $\mathbf{B}_1 \equiv (\mathbf{b}_1, \dots, \mathbf{b}_q)$ is a $p \times q$ -matrix of unknown parameters, \mathbf{e}_t are serially independent $N_p(\mathbf{0}, \boldsymbol{\Sigma}_t)$ errors. If \mathbf{x}_t contains lags of \mathbf{y}_t , model (1) becomes a Vector Autoregression (VAR). $\boldsymbol{\Sigma}_t$ is an unknown $p \times p$ positive definite matrix, which permits a Cholesky decomposition,

$$\boldsymbol{\Sigma}_t = \boldsymbol{\Gamma} \boldsymbol{\Lambda}_t \boldsymbol{\Gamma}', \quad (2)$$

where $\boldsymbol{\Gamma}$ is a lower triangular matrix with unit diagonal elements and $\boldsymbol{\Lambda}_t = \text{diag}(\lambda_{1t}, \dots, \lambda_{pt})$. The decomposition (2) implies that time variation in the covariance matrix stems from sources associated with each of the p variables in the model. Let $h_{jt} = \log \lambda_{jt}$ and $\mathbf{h}_t = (h_{1t}, \dots, h_{pt})'$. We

model \mathbf{h}_t as follows:

$$\mathbf{h}_t = \mathbf{a}_0 + \text{diag}(\mathbf{h}_{t-1})\boldsymbol{\beta} + \mathbf{A}_1\mathbf{z}_t + \text{diag}(\boldsymbol{\delta})\mathbf{v}_t, \quad (3)$$

where $\mathbf{z}_t = (z_{t,1}, \dots, z_{t,r})'$ is an r -vector observable exogenous variable uncorrelated with \mathbf{v}_t , \mathbf{a}_0 is an unknown p -vector; $\mathbf{A}_1 = \{a_{jk}\}$, $j = 1, \dots, p$, $k = 1, \dots, r$ is a $p \times r$ matrix of unknown parameters; $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$ is a p -vector of unknown parameters, each bounded by unity; \mathbf{v}_t are iid $N_p(\mathbf{0}, \mathbf{I}_p)$; and $\boldsymbol{\delta} = (\delta_1, \dots, \delta_p)'$ is a p -vector of unknown nonnegative parameters. If $\delta_i = 0$ then the i th variable is driven by a shock with non-stochastic (but possibly time-varying) conditional volatility. h_0 is assumed to follow the stationary distribution implied by (3).

Our interests in stochastic volatility (SV) models stem from their popularity in analysis of macroeconomic and financial market data. Boscher et al. (2000) and Hol & Koopman (2002) show that in some empirical studies SV models make better forecasts than GARCH models do. Bollerslev et al. (1992), Bauens et al. (2004), Taylor (1994), and Shephard (1996), among others, survey the literature on GARCH and SV models. Jacquier et al. (1994) address the advantages of SV models over GARCH models for high kurtosis time series data. Recently, Bayesian SV models have been extensively studied and much progress has been made (see Jacquier et al. (2002), Aguilar & West (2000), Chib et al. (2002) and (2006) among others). However, the existing studies take the regression and the SV models as given. Such an approach would be questionable when there is substantial uncertainty regarding the specification of the model.

In economic analysis of time series data, the nature of time-varying volatility plays a critical role in the construction of the appropriate model. For example, in option-pricing literature, the widely used Black & Scholes (1973) formula is applicable when the conditional variance of underline asset return is known. If the asset return exhibits stochastic volatility, then the Black-Scholes formula leaves the volatility risk unpriced and yields biased results. Empirical studies show that ignoring the volatility risk substantially hampers the performance of the option pricing model (e.g., see Hull & White (1987) for option pricing of stock prices and Melino & Turnbull (1990) for option pricing

of currency exchange rates). The significance of SV in understanding macroeconomic phenomena was rarely considered in earlier business-cycle literature. More recently, several authors argue that SV is an important attribute of macroeconomic time series data and should be introduced in dynamic stochastic general equilibrium (DSGE) models. For example, Christensen & Kiefer (2000) and Tauchen (2005) explore the implications of SV on asset pricing and Justiniano & Primiceri (2008) argue that SV enhances DSGE models' ability to explain the post-war U.S. business cycle fluctuations. However, redundant variables in a model adversely affect its forecasting ability. Unnecessary introduction of SV is especially costly because of the complexity in computation.

In this paper, we consider the problem that a researcher is uncertain about the dynamic relationship among a set of macroeconomic variables and whether some or all of the variables are generated by a SV model. Instead of settling for a specific set of restrictions a priori, we choose one or a few models from a sufficiently large pool of candidate models. The practical challenge is that even with constant volatilities, the number of possible specifications of the regression model is too large for exhaustive model comparison. Variable selection problem in a univariate model with m explanatory variables involves comparing 2^m competing models. The amount of computation is prohibitive for a moderate size m . George & McCulloch (1993) propose a Bayesian MCMC stochastic search algorithm in univariate regression framework that greatly reduces the amount of computation, and George et al. (2008) extend the algorithm to VAR models with stationary errors. Due to its latent variable nature, stochastic volatility model selection is more challenging than regression model selection.

In this study we conduct stochastic model selection for time varying conditional volatility models. We develop an MCMC algorithm for model selection on coefficients $(\mathbf{B}_1, \mathbf{A}_1)$ as well as $\Psi = \Gamma^{-1'}$ and δ . There are 2^{pq} competing models for \mathbf{B}_1 , 2^{pr} models for \mathbf{A}_1 , $2^{p(p-1)/2}$ models for Ψ and 2^p models for δ . The total number of competing models is $2^{p(r+q+1)+p(p-1)/2}$. The selection of restrictions on the regression coefficients must be conducted jointly with the selection

of restrictions on the volatility model. Following George & McCulloch (1993), we index a model by the tightness of prior vector on parameters. To exclude (or include) a variable, we put a tight (or loose) prior with zero mean on the parameter associated with the variable, which corresponds to an index of zero (or one). The prior of the parameter is a mixture of the tight and loose priors weighted by the prior of the indexes. The posterior of the model indexes constitutes the distribution of selected models.

A major technical difficulty concerns the Bayesian simulation of the latent stochastic volatilities. One can cast the multivariate SV model in a space state framework, with (1) as the measurement equation and (3) as the state equation. Because the state variable is the variance of the error term, the widely-used Kalman filter cannot be used to integrate out the unobserved stochastic volatility. For posterior simulation of stochastic volatilities, we use particle filters (based on a sequential updating of volatilities vector-wise). Simulation studies show that the method is effective for inference on SV.

Specification on the nature of time-varying volatility through parameter δ brings a new challenge to stochastic search. In the existing stochastic search literature, the parameters are assumed to have mixed normal priors. Such an approach is not applicable to the selection of δ , which cannot be negative. We find that the posterior of the model index on δ is sensitive to the hyper-parameters of a mixed inverse gamma prior. To mitigate the sensitivity to prior settings, we consider a hierarchical structure for the prior of δ , with a diffuse gamma prior on the hyper-parameter of the mixed inverse gamma prior on δ . Numerical simulations on artificially generated data show that the hierarchical prior is very effective in selecting the true volatility models.

Stochastic search model selection casts a net that covers the model space and obtains submodels that provide a good fit for observed data. These submodels usually produce better forecasts than the unrestricted model because the latter tends to overfit. Numerical simulations demonstrate that restricted models obtained via stochastic search outperforms the unrestricted model in out-

of-sample forecasting. Boscher et al. (2000) and Hol & Koopman (2002) compare several models with time varying variances, but the number of candidate models in this study is much larger.

The rest of the paper is organized as follows. Section 2 defines the notation and discusses priors. Section 3 derives conditional posterior distributions and lays out the algorithms for Bayesian model selection. Section 4 reports simulation results. Section 5 applies the stochastic search method developed in the paper to exchange rate data. Section 6 offers concluding remarks.

2 The Model and Prior

2.1 Likelihood Functions

We give several formulas for the likelihood functions. First we know that the likelihood function of $(\mathbf{b}_0, \mathbf{B}_1, \boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_T)$ based on $\mathbf{y} = \text{vec}(\mathbf{y}_1, \dots, \mathbf{y}_T)$, is

$$[\mathbf{y} \mid \mathbf{b}_0, \mathbf{B}_1, \boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_T] \propto \prod_{t=1}^T |\boldsymbol{\Sigma}_t|^{-\frac{1}{2}} \text{etr} \left\{ -\frac{1}{2} \sum_{t=1}^T (\mathbf{y}_t - \mathbf{b}_0 - \mathbf{B}_1 \mathbf{x}_t)' \boldsymbol{\Sigma}_t^{-1} (\mathbf{y}_t - \mathbf{b}_0 - \mathbf{B}_1 \mathbf{x}_t) \right\}, \quad (4)$$

where $\text{etr}(\cdot)$ stands for $\exp(\text{trace}(\cdot))$. It follows from (2) that

$$\boldsymbol{\Sigma}_t^{-1} = \boldsymbol{\Psi} \boldsymbol{\Lambda}_t^{-1} \boldsymbol{\Psi}', \quad (5)$$

where $\boldsymbol{\Psi} = \boldsymbol{\Gamma}^{-1'}$ is an upper unit triangular matrix.

Denote $\mathbf{B} = (\mathbf{b}_0, \mathbf{B}_1)$ and $\mathbf{b} = \text{vec}(\mathbf{B})$. The likelihood function of $(\mathbf{b}, \boldsymbol{\Psi}, \boldsymbol{\Lambda}_1, \dots, \boldsymbol{\Lambda}_T)$ is then

$$[\mathbf{y} \mid \mathbf{b}, \boldsymbol{\Psi}, \boldsymbol{\Lambda}_1, \dots, \boldsymbol{\Lambda}_T] \propto |\boldsymbol{\Psi}|^T \prod_{t=1}^T |\boldsymbol{\Lambda}_t|^{-\frac{1}{2}} \text{etr} \left\{ -\frac{1}{2} \sum_{t=1}^T \mathbf{S}_t(\mathbf{b}) \boldsymbol{\Psi} \boldsymbol{\Lambda}_t^{-1} \boldsymbol{\Psi}' \right\}, \quad (6)$$

where

$$\mathbf{S}_t(\mathbf{b}) = (\mathbf{y}_t - \mathbf{b}_0 - \mathbf{B}_1 \mathbf{x}_t)(\mathbf{y}_t - \mathbf{b}_0 - \mathbf{B}_1 \mathbf{x}_t)'. \quad (7)$$

Denote

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_T \end{pmatrix}, \quad \boldsymbol{\Lambda} = \text{diag}(\boldsymbol{\Lambda}_1, \dots, \boldsymbol{\Lambda}_T) = \text{diag}(\lambda_{11}, \lambda_{21}, \dots, \lambda_{p1}, \lambda_{12}, \dots, \lambda_{pT}).$$

The likelihood of $(\mathbf{b}, \Psi, \Lambda)$ is then

$$\begin{aligned}
L(\mathbf{b}, \Psi, \Lambda) &= [\mathbf{y} \mid \mathbf{b}, \Psi, \Lambda] \\
&\propto |\Psi|^T |\Lambda|^{-\frac{1}{2}} \text{etr} \left\{ -\frac{1}{2} [\mathbf{y} - (\mathbf{X}' \otimes \mathbf{I}_p) \mathbf{b}]' (\mathbf{I}_T \otimes \Psi) \Lambda^{-1} (\mathbf{I}_T \otimes \Psi') [\mathbf{y} - (\mathbf{X}' \otimes \mathbf{I}_p) \mathbf{b}] \right\} \\
&= |\Psi|^T |\Lambda|^{-\frac{1}{2}} \text{etr} \left\{ -\frac{1}{2} [(\mathbf{I}_T \otimes \Psi') \mathbf{y} - (\mathbf{X}' \otimes \Psi') \mathbf{b}]' \Lambda^{-1} [(\mathbf{I}_T \otimes \Psi') \mathbf{y} - (\mathbf{X}' \otimes \Psi') \mathbf{b}] \right\}. \quad (8)
\end{aligned}$$

For a full Bayesian analysis, one needs to assign a prior to parameters. In the likelihood function (8), $\mathbf{b} = \text{vec}(\mathbf{b}_0, \mathbf{B}_1)$ and Ψ are parameters pertaining to the regression model (1), Λ involves parameters $(\mathbf{a}_0, \beta, \mathbf{A}_1, \delta)$ in SV equation (3). We now discuss the prior specification.

2.2 Priors of $(\mathbf{b}_0, \mathbf{B}_1, \Psi)$

We assume the priors of components of (\mathbf{b}, Ψ) are independent. Here are the marginal priors.

(i) *Priors of $\mathbf{b}_0 = (b_{10}, \dots, b_{p0})'$.* We assume that the intercept b_{i0} is always included in the model. We also assume independent priors for b_{i0} :

$$b_{i0} \stackrel{\text{indep}}{\sim} N(b_{i0}^0, \xi_{i0}^2), \quad i = 1, \dots, p. \quad (9)$$

(ii) *Priors of $\mathbf{B}_1 = \{b_{ij}\}_{p \times q}$.* Each element b_{ij} is associated with an indicator $\gamma_{b,ij}$. If $\gamma_{b,ij} = 1$, b_{ij} is included; and if $\gamma_{b,ij} = 0$, b_{ij} is excluded. Then b_{ij} has a two-stage prior: for fixed $p_{b,ij} \in (0, 1)$,

$$P(\gamma_{b,ij} = 1) = 1 - P(\gamma_{b,ij} = 0) = p_{b,ij}, \quad i = 1, \dots, p, \quad j = 1, \dots, q. \quad (10)$$

For given $\gamma_b = (\gamma_{b,11}, \gamma_{b,12}, \dots, \gamma_{b,pq})'$, we let

$$(b_{ij} \mid \gamma_{b,ij}) \stackrel{\text{indep}}{\sim} (1 - \gamma_{b,ij}) N(0, \kappa_{b,ij}^2) + \gamma_{b,ij} N(0, c_{b,ij}^2 \kappa_{b,ij}^2), \quad (11)$$

for $i = 1, \dots, p$ and $j = 1, \dots, q$, where $\kappa_{b,ij}$ are small and $c_{b,ij}$ are large constants. If we write

$$\eta_{b,ij} = c_{b,ij}^{\gamma_{b,ij}} = \begin{cases} 1, & \text{if } \gamma_{b,ij} = 0, \\ c_{b,ij}, & \text{if } \gamma_{b,ij} = 1, \end{cases}$$

and $\mathbf{D}_{b,j} = \text{diag}((\eta_{b,1j}\kappa_{b,1j})^2, \dots, (\eta_{b,pj}\kappa_{b,pj})^2)$, then (11) is equivalent to

$$(\mathbf{b}_j \mid \boldsymbol{\gamma}_{b,j}) \stackrel{\text{indep}}{\sim} N_p(\mathbf{0}, \mathbf{D}_{b,j}), \quad \text{for } j = 1, \dots, q. \quad (12)$$

Combining the priors in (ia) and (ib), we can write the prior for \mathbf{b} as

$$(\mathbf{b} \mid \boldsymbol{\gamma}_b) \sim N(\bar{\mathbf{b}}, \bar{\boldsymbol{\Xi}}), \quad (13)$$

where

$$\begin{aligned} \bar{\mathbf{b}} &= (b_{10}^0, \dots, b_{p0}^0, 0, \dots, 0)', \\ \bar{\boldsymbol{\Xi}} &= \text{diag}(\xi_{10}^2, \dots, \xi_{p0}^2, (\eta_{b,11}\kappa_{b,11})^2, \dots, (\eta_{b,pq}\kappa_{b,pq})^2). \end{aligned}$$

(ii) Priors of $\boldsymbol{\Psi}$.

For $j = 2, \dots, p$, let $\boldsymbol{\psi}_j$ be a vector containing the non-redundant elements of the j^{th} column of $\boldsymbol{\Psi}$, i.e. $\boldsymbol{\psi}_j = (\psi_{1j}, \dots, \psi_{j-1,j})'$. Also, define a vector of indicators of length $j - 1$, $\boldsymbol{\gamma}_{\psi,j} = (\gamma_{\psi,1j}, \dots, \gamma_{\psi,j-1,j})'$. We assume that elements of $\boldsymbol{\psi}_j$ may be included in the model ($\gamma_{\psi,ij} = 1$) or not ($\gamma_{\psi,ij} = 0$). Let the model index for ψ_{ij} , $\gamma_{\psi,ij}$, be independent Bernoulli ($p_{\psi,ij}$) random variables: for fixed $p_{\psi,ij} \in (0, 1)$,

$$P(\gamma_{\psi,ij} = 1) = 1 - P(\gamma_{\psi,ij} = 0) = p_{\psi,ij}, \quad i = 1, \dots, j - 1, \quad j = 1, \dots, p. \quad (14)$$

For given $\boldsymbol{\gamma}_{\psi,j} = (\gamma_{\psi,1j}, \dots, \gamma_{\psi,j-1,j})'$, we assume that

$$(\psi_{ij} \mid \boldsymbol{\gamma}_{\psi,ij}) \stackrel{\text{indep}}{\sim} (1 - \gamma_{\psi,ij})N(0, \kappa_{\psi,ij}^2) + \gamma_{\psi,ij}N(0, c_{\psi,ij}^2 \kappa_{\psi,ij}^2), \quad (15)$$

for $i = 1, \dots, j - 1$ and $j = 2, \dots, p$, where $\kappa_{\psi,ij}$ are small and $c_{\psi,ij}$ are large constants. If we write

$$\eta_{\psi,ij} = c_{\psi,ij}^{\gamma_{\psi,ij}} = \begin{cases} 1, & \text{if } \gamma_{\psi,ij} = 0, \\ c_{\psi,ij}, & \text{if } \gamma_{\psi,ij} = 1, \end{cases}$$

and $\mathbf{D}_{\psi,j} = \text{diag}((\eta_{\psi,1j}\kappa_{\psi,1j})^2, \dots, (\eta_{\psi,j-1,j}\kappa_{\psi,j-1,j})^2)$, then (15) is equivalent to

$$(\boldsymbol{\psi}_j \mid \boldsymbol{\gamma}_{\psi,j}) \stackrel{\text{indep}}{\sim} N_{j-1}(\mathbf{0}, \mathbf{D}_{\psi,j}), \quad (16)$$

for $j = 2, \dots, p$. Note that a slight modification of the setting allows for modelling whether some elements of Ψ are equal. Instead of centering the prior of these elements at 0, we can set them at a common mean in prior (15). When the parameter index $\gamma = 0$ the corresponding element approximately equals to the common mean and when $\gamma = 1$ it is not restricted. But to simplify notation, throughout the paper we only consider priors centered at 0.

2.3 Priors of $(\mathbf{a}_0, \boldsymbol{\beta}, \mathbf{A}_1, \boldsymbol{\delta})$.

We assume that \mathbf{a}_0 , $\boldsymbol{\beta}$, \mathbf{A}_1 , and $\boldsymbol{\delta}$ have mutually independent priors.

(i) *Priors of $\mathbf{a}_0 = (a_{10}, \dots, a_{p0})'$.* For fixed (\bar{a}_{j0}, σ_a) , we assume that

$$(a_{j0}) \stackrel{indep}{\sim} N(\bar{a}_{j0}, \sigma_a^2). \quad (17)$$

(ii) *Priors of $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$.* For fixed $(\bar{\beta}_j, \sigma_\beta)$, we assume that

$$(\beta_j) \stackrel{indep}{\sim} N(\bar{\beta}_j, \sigma_\beta^2). \quad (18)$$

(iii). *Priors of \mathbf{A}_1 .* Let the model index for a_{jk} , $\gamma_{a,jk}$ be independent Bernoulli $(p_{a,jk})$ random variables: For fixed $p_{a,jk} \in (0, 1)$,

$$P(\gamma_{a,jk} = 1) = 1 - P(\gamma_{a,jk} = 0) = p_{a,jk}, \text{ for } j = 1, \dots, p, k = 1, \dots, r. \quad (19)$$

For given $\boldsymbol{\gamma}_{a,j} = (\gamma_{a,j1}, \gamma_{a,j2}, \dots, \gamma_{a,jr})'$, we assume that

$$(a_{jk} | \boldsymbol{\gamma}_{a,jk}) \stackrel{indep}{\sim} (1 - \gamma_{a,jk})N(0, \kappa_{a,jk}^2) + \gamma_{a,jk}N(0, c_{a,jk}^2 \kappa_{a,jk}^2). \quad (20)$$

where $\kappa_{a,jk}$ would be small and $c_{a,jk}$ would be large constants. Later, we also write \mathbf{A}_1 in terms of its row vectors: $\mathbf{A}_1 = (\tilde{\mathbf{a}}_1', \dots, \tilde{\mathbf{a}}_p')$. Here $\tilde{\mathbf{a}}_j = (a_{j1}, \dots, a_{jr})'$, $j = 1, \dots, p$. Denote

$$\eta_{a,jk} = c_{a,jk}^{\gamma_{a,jk}} = \begin{cases} 1, & \text{if } \gamma_{a,jk} = 0, \\ c_{a,jk}, & \text{if } \gamma_{a,jk} = 1. \end{cases}$$

and $\mathbf{D}_{a,j} = \text{diag}((\eta_{a,j1}\kappa_{a,j1})^2, \dots, (\eta_{a,jr}\kappa_{a,jr})^2)$. We know that the prior of $\tilde{\mathbf{a}}_j$ for given $\gamma_{a,j} = (\gamma_{a,j1}, \dots, \gamma_{a,jr})$ is

$$(\tilde{\mathbf{a}}_j | \gamma_{a,j}) \stackrel{\text{indep}}{\sim} N_r(\mathbf{0}, \mathbf{D}_{a,j}). \quad (21)$$

Define $\tilde{\mathbf{a}}_j^* = (a_{j0}, \beta_j, \tilde{\mathbf{a}}_j)'$: combining (17) and (21) we can write

$$(\tilde{\mathbf{a}}_j^* | \gamma_{a,j}) \stackrel{\text{indep}}{\sim} N_{r+2}(\bar{\mathbf{a}}_j, \mathbf{\Omega}_j), \quad (22)$$

where

$$\begin{aligned} \bar{\mathbf{a}}_j &= (\bar{a}_{j0}, \bar{\beta}_j, 0, \dots, 0)' \\ \mathbf{\Omega}_j &= \text{diag}(\sigma_a^2, \sigma_\beta^2, (\eta_{a,j1}\kappa_{a,j1})^2, \dots, (\eta_{a,jr}\kappa_{a,jr})^2). \end{aligned}$$

(iv) *Prior for δ* . The elements in $\boldsymbol{\delta} = (\delta_1, \dots, \delta_p)'$ are objects of model selection. A positive δ_j corresponds to stochastic volatility in the innovation of variable j . In multivariate time series literature, a random variable is usually assumed to be generated from a model with either deterministic or stochastic conditional volatilities. Using our stochastic search algorithm for $\boldsymbol{\delta}$, we can estimate a very general model for heteroscedastic time series and make a data driven selection between a stochastic and a deterministic process for each variable. Let the model index for δ_j , γ_{δ_j} be independent Bernoulli (p_{δ_j}) random variables: for fixed $p_{\delta_j} \in (0, 1)$,

$$P(\gamma_{\delta_j} = 1) = 1 - P(\gamma_{\delta_j} = 0) = p_{\delta_j}, \quad j = 1, \dots, p. \quad (23)$$

For given γ_{δ_j} ($j = 1, \dots, p$), we assume that δ_j is a mixture of independent inverse gamma, $(\delta_j^2 | \gamma_{\delta_j}, q_j) \sim IG(v_{j0}; q_j s_{j0})$ with probability γ_{δ_j} and $(\delta_j^2 | \gamma_{\delta_j}, q_j) \sim IG(v_{j0}; s_{j0})$ with probability $1 - \gamma_{\delta_j}$.

In other words, the density of δ_j^2 has the form

$$[\delta_j^2 | \gamma_{\delta_j}, q_j] \propto (\eta_{\delta_j} s_{j0})^{v_{j0}} (\delta_j^2)^{-(v_{j0}+1)} \exp\left(-\frac{\eta_{\delta_j} s_{j0}}{\delta_j^2}\right), \quad (24)$$

where

$$\eta_{\delta_j} = q_j^{\gamma_{\delta_j}} = \begin{cases} 1, & \text{if } \gamma_{\delta_j} = 0, \\ q_j, & \text{if } \gamma_{\delta_j} = 1. \end{cases}$$

Here the scale parameter v_{j0} is a given positive constant greater than 2; the shape parameter s_{j0} is a small positive constant so that with an appropriately chosen q_j the mean and the variance of the prior with $\gamma_{\delta_j} = 1$, $q_j s_{j0}/(v_{j0} - 1)$ and $(q_j s_{j0})^2/\{(v_{j0} - 1)^2(v_{j0} - 2)\}$, are large while prior mean and variance corresponding to $\gamma_{\delta_j} = 0$, $s_{j0}/(v_{j0} - 1)$ and $s_{j0}^2/\{(v_{j0} - 1)^2(v_{j0} - 2)\}$, are close to zero. The choice of q_j needs some further work. The model selection on the other parameters of the model is a selection of linear regression coefficients with normal distribution, like in George & McCulloch (1993) and George et al. (2008), but $\boldsymbol{\delta}$ follows an inverse gamma distribution, so even small changes in the values of its hyper-parameters can have a large impact (see Kim et al. (2002) for a discussion on normal gamma priors).

Experiments show that fixing hyper parameter $\mathbf{q} = (q_1, \dots, q_p)'$ leads to unsatisfactory results. Simulation studies show that no arbitrary value of \mathbf{q} is universally effective. When q_j is set as a large (small) number, the stochastic volatility parameter δ_j is often mistakenly excluded when the true value δ_j is small (large). Fixing q_j in accordance with the magnitude of δ_j is infeasible because δ_j is unknown. We solve this problem by assigning a hierarchal prior structure to the hyper-parameter \mathbf{q} . Specifically, we set the prior on \mathbf{q} as:

$$q_j \sim Ga(\alpha_q, \beta_q), \text{ for } j = 1, \dots, p. \quad (25)$$

In this way each single δ_j will have a different data-driven posterior value of q_j .

2.4 Choice of Hyper-parameters

The prior setting for the numerical simulation discussed later in the paper is as follows. The hyper-parameter is .5 for all the Bernoulli priors in the elements of $(\mathbf{B}_1, \boldsymbol{\Psi}, \mathbf{A}_1, \boldsymbol{\delta})$, making it equally likely to include or exclude each variable. The prior on each element of \mathbf{b}_0 is $N(0, 50)$. For hyper-parameters on stochastic search of $\boldsymbol{\Psi}$, \mathbf{B}_1 , and \mathbf{A}_1 we set $\kappa_{ij} = 0.1$, $c_{ij} = d_{ij} = 50.0$. The hyper-parameters for $\boldsymbol{\delta}$ are $v_{j0} = 6.0$ and $s_{j0}^2 = .001$, and the hyper-parameters for q are $\alpha_q = 5$ and $\beta_q = 1$. We set the prior mean α_{j0} at 0 and standard deviation σ_α at 10 (consequently the

priors are not centered at the true values and are quite flat).

For β_j ($j = 1, \dots, p$), Kim et al. (1998) employ a $Beta(20, 1.5)$ prior on the interval $(-1, 1)$, Their prior has the mean .860, standard deviation .107, and is quite informative. We employ a truncated normal prior with the same mean .860 but with a larger variance of unity. To preserve the stationarity of the SV process, we truncate the prior of β outside the interval $(-1, 1)$. This prior is quite flat on the interval $(-1, 1)$. In the web appendix, we discuss how sensitive the simulation results are to hyper-parameters.

3 Posterior Computation

The joint posterior of $(\mathbf{b}, \Psi, \Lambda, \mathbf{a}_0, \beta, \mathbf{A}_1, \delta, \gamma_b, \gamma_\psi, \gamma_a, \gamma_\delta, \mathbf{q})$ has the form

$$\begin{aligned} [\mathbf{b}, \Psi, \Lambda, \mathbf{a}_0, \beta, \mathbf{A}_1, \delta, \gamma_b, \gamma_\psi, \gamma_a, \gamma_\delta, \mathbf{q} \mid \mathbf{y}] &\propto [\mathbf{y} \mid \mathbf{b}, \Psi, \Lambda] [\Lambda \mid \mathbf{a}_0, \beta, \mathbf{A}_1, \delta] \\ &\times [\mathbf{b}_0] [\mathbf{B}_1 \mid \gamma_b] [\gamma_b] [\Psi \mid \gamma_\psi] [\gamma_\psi] [\mathbf{a}_0, \beta] [\mathbf{A}_1 \mid \gamma_a] [\gamma_a] [\delta \mid \gamma_\delta, \mathbf{q}] [\gamma_\delta] [\mathbf{q}], \end{aligned} \quad (26)$$

where $[\mathbf{y} \mid \mathbf{b}, \Psi, \Lambda]$ is the likelihood function, given by (8), $[\Lambda \mid \mathbf{a}_0, \beta, \mathbf{A}_1, \delta]$ is given through (3), $[\mathbf{b}_0]$ is given by (9), $[\mathbf{B}_1 \mid \gamma_b]$ is given by (12), $[\gamma_b]$ is given by (10), $[\Psi \mid \gamma_\psi]$ is given by (16), $[\gamma_\psi]$ is given by (14), $[\mathbf{a}_0]$ is given by (17), $[\beta]$ is given by (18), $[\mathbf{A}_1 \mid \gamma_a]$ is given by (21), $[\gamma_a]$ is given by (19), $[\delta \mid \gamma_\delta, \mathbf{q}]$ is given by (24), $[\gamma_\delta]$ is given by (23) and $[\mathbf{q}]$ is given by (25). To use an MCMC algorithm, we now derive the full conditional posteriors for $(\mathbf{b}, \Psi, \Lambda, \mathbf{a}_0, \beta, \mathbf{A}_1, \delta, \gamma_b, \gamma_\psi, \gamma_a, \gamma_\delta, \mathbf{q})$.

3.1 Conditional Posteriors for \mathbf{b} and γ_b

Fact 1 (a) *The conditional posterior distribution of \mathbf{b} given $(\gamma_b, \Psi, \gamma_\psi, \mathbf{a}_0, \beta, \mathbf{A}_1, \delta, \gamma_a, \Lambda, \mathbf{q}; \mathbf{y})$ depends only on $(\Psi, \Lambda, \gamma_b; \mathbf{y})$ and has the form*

$$(\mathbf{b} \mid \Psi, \Lambda, \gamma_b; \mathbf{y}) \sim N_m(\widehat{\mathbf{b}}, \widehat{\Xi}), \quad (27)$$

where

$$\widehat{\Xi} = \left\{ (\mathbf{X} \otimes \Psi) \Lambda^{-1} (\mathbf{X}' \otimes \Psi') + \bar{\Xi}^{-1} \right\}^{-1}, \quad (28)$$

$$\hat{\mathbf{b}} = \hat{\Xi} \left\{ (\mathbf{X} \otimes \Psi) \Lambda^{-1} (\mathbf{I}_T \otimes \Psi') \mathbf{y} + \hat{\Xi}^{-1} \bar{\mathbf{b}} \right\}. \quad (29)$$

(b) Denote $\gamma_{b,(-ij)} = (\gamma_{b,kl} : (k,l) \neq (i,j))$. Given prior independence for b_{ij} , the conditional posterior distributions of γ_b for given $(\mathbf{b}, \Psi, \Lambda, \mathbf{a}_0, \beta, \mathbf{A}_1, \delta, \gamma_{b,(-ij)}, \gamma_\psi, \gamma_a, \mathbf{q})$ depend only on \mathbf{B}_1 ,

$$(\gamma_{b,ij} | \mathbf{B}_1) = (\gamma_{b,ij} | b_{ij}) \stackrel{indep}{\sim} \text{Bernoulli} \left(\frac{u_{b,ij1}}{u_{b,ij1} + u_{b,ij2}} \right), \quad (30)$$

where

$$\begin{aligned} u_{b,ij1} &= \frac{1}{c_{b,ij}} \exp \left(-\frac{b_{ij}^2}{2c_{b,ij}^2 \kappa_{b,ij}^2} \right) p_{b,ij}, \\ u_{b,ij2} &= \exp \left(-\frac{b_{ij}^2}{2\kappa_{b,ij}^2} \right) (1 - p_{b,ij}). \end{aligned} \quad (31)$$

PROOF. Using the likelihood (8) part (a) is obvious. For part (b), recall that γ_b depends on data indirectly, then, $u_{b,ij1} \propto [\mathbf{b} | \gamma_{b,(-ij)}, \gamma_{b,ij} = 1] p_{b,ij}$, and $u_{b,ij2} \propto [\mathbf{b} | \gamma_{b,(-ij)}, \gamma_{b,ij} = 0] (1 - p_{b,ij})$. The formula (31) follows from the prior independence of b_{ij} . \square

3.2 Conditional Posteriors for Ψ and γ_ψ

To derive conditional distributions of Ψ , we use the likelihood function (6) of $(\mathbf{b}, \Psi, \Lambda_1, \dots, \Lambda_T)$. Note that $\mathbf{S}_t = \mathbf{S}_t(\mathbf{b})$ represents the covariance of residuals \mathbf{e}_t . Let $\mathbf{S}_{t,j}$ be the upper-left $j \times j$ submatrix of $\mathbf{S}_t(\mathbf{b})$. So $\mathbf{S}_t = \mathbf{S}_{t,p}$. We denote the $(i,j)^{th}$ component of $\mathbf{S}_t(\mathbf{b})$ by $s_{t,ij}$. For $j = 2, \dots, p$, define $\mathbf{s}_{t,j} = (s_{t,1j}, \dots, s_{t,j-1,j})'$. Define

$$v_{t,1} = s_{t,11}, \quad v_{t,j} = \frac{|\mathbf{S}_{t,j}|}{|\mathbf{S}_{t,j-1}|}, \quad \text{for } j = 2, \dots, p. \quad (32)$$

It follows from the formula of determinant of a partitioned matrix and the fact that covariance matrices are positive definite that $v_{t,j} = s_{t,jj} - \mathbf{s}'_{t,j} \mathbf{S}_{t,j-1}^{-1} \mathbf{s}_{t,j} > 0$ for $j = 2, \dots, p$. We define ψ_j as the $j-1$ elements of j th column of Ψ above the diagonal element. So for $j = 2, \dots, p$, $\psi_j = (\psi_{1j}, \dots, \psi_{j-1,j})'$. This recursive formula is immediate,

$$\text{tr}(\Psi'_p \mathbf{S}_{t,p} \Psi_p \Lambda_{t,p}^{-1}) = \text{tr}(\Psi'_{p-1} \mathbf{S}_{t,p-1} \Psi_{p-1} \Lambda_{t,p-1}^{-1}) + \lambda_{t,p}^{-1} v_{t,p} + \lambda_{t,p}^{-1} \mathbf{g}'_t \mathbf{S}_{t,p-1} \mathbf{g}_t,$$

where $\mathbf{g}_t = \boldsymbol{\psi}_{p-1} + \mathbf{S}_{t,p-1}^{-1} \mathbf{s}_{t,p}$. The likelihood function $[\mathbf{y} \mid \mathbf{b}, \boldsymbol{\Psi}, \boldsymbol{\Lambda}_1, \dots, \boldsymbol{\Lambda}_T]$ of (6) has the form,

$$\left(\prod_{t=1}^T \prod_{j=1}^p \frac{1}{\sqrt{\lambda_{jt}}} \right) \exp \left[-\frac{1}{2} \left\{ \sum_{t=1}^T \sum_{j=1}^p \frac{v_{t,j}}{\lambda_{jt}} + \sum_{t=1}^T \sum_{j=2}^p \frac{1}{\lambda_{jt}} (\boldsymbol{\psi}_j + \mathbf{S}_{t,j-1}^{-1} \mathbf{s}_{t,j})' \mathbf{S}_{t,j-1} (\boldsymbol{\psi}_j + \mathbf{S}_{t,j-1}^{-1} \mathbf{s}_{t,j}) \right\} \right].$$

This expression allows us to derive the conditional posterior of $\boldsymbol{\Psi}$.

Fact 2 (a) *The conditional posteriors of $\boldsymbol{\psi}_2, \dots, \boldsymbol{\psi}_p$ given $(\gamma_b, \mathbf{b}, \boldsymbol{\gamma}_\psi, \mathbf{a}_0, \boldsymbol{\beta}, \mathbf{A}_1, \boldsymbol{\delta}, \boldsymbol{\gamma}_a, \boldsymbol{\Lambda}, \mathbf{q}; \mathbf{y})$ are independent and depend only on $(\mathbf{b}, \boldsymbol{\gamma}_\psi, \boldsymbol{\Lambda}; \mathbf{y})$. We have the form,*

$$(\boldsymbol{\psi}_j \mid \mathbf{b}, \boldsymbol{\gamma}_\psi, \boldsymbol{\Lambda}; \mathbf{y}) \stackrel{indep}{\sim} N_{j-1}(\boldsymbol{\mu}_j, \boldsymbol{\Delta}_j), \quad (33)$$

where

$$\boldsymbol{\Delta}_j = \left(\sum_{t=1}^T \lambda_{jt}^{-1} \mathbf{S}_{t,j-1} + \mathbf{D}_{\boldsymbol{\psi},j}^{-1} \right)^{-1} \text{ and } \boldsymbol{\mu}_j = -\boldsymbol{\Delta}_j \sum_{t=1}^T \lambda_{jt}^{-1} \mathbf{s}_{t,j}. \quad (34)$$

(b) *For $j = 2, \dots, p$ and $i = 1, \dots, j-1$, denote $\boldsymbol{\gamma}_{\psi,(-ij)} = (\gamma_{\psi,1j}, \dots, \gamma_{\psi,i-1,j}, \gamma_{\psi,i+1,j}, \dots, \gamma_{\psi,j-1,j})'$.*

Given prior independence for $(\boldsymbol{\psi}_{1j}, \dots, \boldsymbol{\psi}_{j-1,j})$, the conditional posterior of $\gamma_{\psi,ij}$ given $(\gamma_b, \mathbf{b}, \boldsymbol{\Psi}, \boldsymbol{\gamma}_{\psi,(-ij)}, \mathbf{a}_0, \boldsymbol{\beta}, \mathbf{A}_1, \boldsymbol{\delta}, \boldsymbol{\gamma}_a, \boldsymbol{\Lambda}, \mathbf{q}; \mathbf{y})$ depends only on $\boldsymbol{\Psi}$ and has the form,

$$(\gamma_{\psi,ij} \mid \boldsymbol{\Psi}) = (\gamma_{\psi,ij} \mid \boldsymbol{\psi}_{ij}) \stackrel{indep}{\sim} \text{Bernoulli} \left(\frac{u_{\psi,ij1}}{u_{\psi,ij1} + u_{\psi,ij2}} \right), \quad (35)$$

where

$$\begin{aligned} u_{\psi,ij1} &= \frac{1}{c_{\psi,ij}} \exp \left(-\frac{\psi_{ij}^2}{2c_{\psi,ij}^2 \kappa_{\psi,ij}^2} \right) p_{\psi,ij}, \\ u_{\psi,ij2} &= \exp \left(-\frac{\psi_{ij}^2}{2\kappa_{\psi,ij}^2} \right) (1 - p_{\psi,ij}). \end{aligned} \quad (36)$$

PROOF. The conditional posterior density of $(\boldsymbol{\psi}_2, \dots, \boldsymbol{\psi}_p)$, given $(\mathbf{b}, \boldsymbol{\Lambda}_1, \dots, \boldsymbol{\Lambda}_T, \boldsymbol{\gamma}_\psi; \mathbf{y})$ is

$$\begin{aligned} & [\boldsymbol{\psi}_2, \dots, \boldsymbol{\psi}_p \mid \mathbf{b}, \boldsymbol{\gamma}_\psi, \boldsymbol{\Lambda}_1, \dots, \boldsymbol{\Lambda}_T; \mathbf{y}] \\ & \propto \exp \left\{ -\frac{1}{2} \sum_{j=2}^p \sum_{t=1}^T \frac{1}{\lambda_{jt}} (\boldsymbol{\psi}_j + \mathbf{S}_{t,j-1}^{-1} \mathbf{s}_{t,j})' \mathbf{S}_{t,j-1} (\boldsymbol{\psi}_j + \mathbf{S}_{t,j-1}^{-1} \mathbf{s}_{t,j}) \right\} \exp \left\{ -\frac{1}{2} \sum_{j=2}^p \boldsymbol{\psi}_j' \mathbf{D}_{\boldsymbol{\psi},j}^{-1} \boldsymbol{\psi}_j \right\} \\ & \propto \exp \left\{ -\frac{1}{2} \sum_{j=2}^p (\boldsymbol{\psi}_j - \boldsymbol{\mu}_j)' \boldsymbol{\Delta}_j^{-1} (\boldsymbol{\psi}_j - \boldsymbol{\mu}_j) \right\}, \end{aligned}$$

where $\boldsymbol{\mu}_j$ and $\boldsymbol{\Delta}_j$ are defined in (34). Part (a) follows from direct computation. For part (b), recall the fact that γ_ψ depends on data indirectly, then, $u_{\psi,ij1} \propto [\boldsymbol{\psi}_j \mid \gamma_{\psi,(-ij)}, \gamma_{\psi,ij} = 1] p_{\psi,ij}$ and $u_{\psi,ij2} \propto [\boldsymbol{\psi}_j \mid \gamma_{\psi,(-ij)}, \gamma_{\psi,ij} = 0](1 - p_{\psi,ij})$. The expression above, under prior independence of $\psi_{1j}, \dots, \psi_{j-1,j}$, gives the formula (36). \square

3.3 Conditional Posteriors of $(\mathbf{a}_0, \boldsymbol{\beta})$, (\mathbf{A}_1, γ_a) and $(\boldsymbol{\delta}, \gamma_\delta, \mathbf{q})$

Recall $h_{jt} = \log(\lambda_{jt})$. Define $\mathbf{H} = (h_{jt})_{p \times T}$ and the column vector $\tilde{\mathbf{h}}_j = (h_{j1}, \dots, h_{jT})'$. Recall that $\tilde{\mathbf{a}}_j'$ is the j^{th} row of \mathbf{A}_1 . We define

$$\mathbf{W}_j = \begin{pmatrix} 1 & h_{0j} & z_{11} & \cdots & z_{1r} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & h_{T-1,j} & z_{T1} & \cdots & z_{Tr} \end{pmatrix}.$$

The conditional posterior of $\mathbf{a}_0, \boldsymbol{\beta}, \mathbf{A}_1, \gamma_a$ and $\boldsymbol{\delta}$ are as follows:

Fact 3 (a) *The conditional posterior distributions of $\tilde{\mathbf{a}}_j^* = (a_{j0}, \beta_j, \tilde{\mathbf{a}}_j)'$, $j = 1, \dots, p$ given $(\gamma_b, \mathbf{b}, \gamma_\psi, \boldsymbol{\Psi}, \boldsymbol{\delta}, \gamma_a, \mathbf{H}, \mathbf{q}; \mathbf{y})$ are independent and depend only on $(\mathbf{H}, \gamma_a, \boldsymbol{\delta})$.*

$$(\tilde{\mathbf{a}}_j^* \mid \mathbf{H}, \gamma_a, \boldsymbol{\delta}) = (\tilde{\mathbf{a}}_j^* \mid \tilde{\mathbf{h}}_j, \gamma_{a,j}, \delta_j) \stackrel{\text{indep}}{\sim} N_{r+2}(\boldsymbol{\nu}_j, \boldsymbol{\Upsilon}_j), \quad (37)$$

where

$$\boldsymbol{\nu}_j = \boldsymbol{\Upsilon}_j (\delta_j^{-2} \mathbf{W}_j' \tilde{\mathbf{h}}_j + \boldsymbol{\Omega}_j^{-1} \tilde{\mathbf{a}}_j) \text{ and } \boldsymbol{\Upsilon}_j = (\delta_j^{-2} \mathbf{W}_j' \mathbf{W}_j + \boldsymbol{\Omega}_j^{-1})^{-1}.$$

(b) *The conditional posterior of $(\delta_1, \dots, \delta_p)$ given $(\gamma_\delta, \gamma_b, \mathbf{b}, \gamma_\psi, \boldsymbol{\Psi}, \mathbf{a}_0, \boldsymbol{\beta}, \mathbf{A}_1, \gamma_a, \mathbf{H}, \mathbf{q}; \mathbf{y})$ are mutually independent and depend only on $(\gamma_\delta, \mathbf{H}, \mathbf{a}_0, \boldsymbol{\beta}, \mathbf{A}_1, \mathbf{q})$,*

$$\begin{aligned} (\delta_j^2 \mid \gamma_\delta, \mathbf{H}, \mathbf{a}_0, \boldsymbol{\beta}, \mathbf{A}_1) &= (\delta_j^2 \mid \gamma_{\delta j}, \tilde{\mathbf{h}}_j, a_{j0}, \beta_j, \tilde{\mathbf{a}}_j) \\ &\stackrel{\text{indep}}{\sim} \mathbf{IG}\left(v_{j0} + \frac{1}{2}(T-2), \eta_{\delta j}^2 s_{j0}^2 + \frac{1}{2}(\tilde{\mathbf{h}}_j - \mathbf{W}_j \tilde{\mathbf{a}}_j)'(\tilde{\mathbf{h}}_j - \mathbf{W}_j \tilde{\mathbf{a}}_j)\right); \end{aligned} \quad (38)$$

(c) *For $j = 1, \dots, p$, the conditional posterior of $\gamma_{\delta j}$ given $(\gamma_b, \mathbf{b}, \boldsymbol{\Psi}, \gamma_a, \mathbf{a}_0, \boldsymbol{\beta}, \mathbf{A}_1, \boldsymbol{\delta}, \gamma_\psi, \mathbf{H}, \mathbf{q}; \mathbf{y})$*

depends only on (δ_j, q_j) , and has the form

$$(\gamma_{\delta_j} \mid \delta_j, q_j) \stackrel{\text{indep}}{\sim} \text{Bernoulli}\left(\frac{u_{\delta_j1}}{u_{\delta_j1} + u_{\delta_j2}}\right), \quad (39)$$

where

$$u_{\delta_j1} = \exp\left(-\frac{q_j s_{j0}}{\delta_j^2}\right) p_{\delta_j} q_j^{v_{j0}} \quad \text{and} \quad u_{\delta_j2} = \exp\left(-\frac{s_{j0}}{\delta_j^2}\right) (1 - p_{\delta_j}). \quad (40)$$

(d) For $j = 1, \dots, p$, denote $\gamma_{a,(-jk)} = (\gamma_{a,j1}, \dots, \gamma_{a,jk-1}, \gamma_{a,jk+1}, \dots, \gamma_{a,jr})'$. Under the assumption of prior independence for the elements of $\tilde{\mathbf{a}}_j$, the conditional posterior of $\gamma_{a,jk}$ given $(\gamma_b, \mathbf{b}, \Psi, \gamma_{a,(-jk)}, \mathbf{a}_0, \beta, \mathbf{A}_1, \delta, \gamma_\psi, \mathbf{H}, \mathbf{q}; \mathbf{y})$ depends only on \mathbf{A}_1 , and has the form

$$(\gamma_{a,jk} \mid \mathbf{A}_1) = (\gamma_{a,jk} \mid a_{jk}) \stackrel{\text{indep}}{\sim} \text{Bernoulli}\left(\frac{u_{a,jk1}}{u_{a,jk1} + u_{a,jk2}}\right), \quad (41)$$

where

$$u_{a,jk1} = \frac{p_{a,jk}}{c_{a,jk}} \exp\left(-\frac{a_{jk}^2}{c_{a,jk}^2 2\kappa_{a,jk}^2}\right) \quad \text{and} \quad u_{a,jk2} = (1 - p_{a,jk}) \exp\left(-\frac{a_{jk}^2}{2\kappa_{a,jk}^2}\right). \quad (42)$$

(e) For $j = 1, \dots, p$, the conditional posterior of q_j given $(\gamma_b, \mathbf{b}, \Psi, \gamma_a, \mathbf{a}_0, \beta, \mathbf{A}_1, \delta, \gamma_\psi, \mathbf{H}, \gamma_\delta; \mathbf{y})$ depends only on $(\delta_j, \gamma_{\delta_j})$, and has the form

$$(q_j \mid \delta_j, \gamma_{\delta_j}) \sim \text{Ga}(v_{j0}\gamma_{\delta_j} + \alpha_q, \frac{s_{j0}}{\delta_j^2}\gamma_{\delta_j} + \beta_q). \quad (43)$$

PROOF. Parts (a) and (b) can easily be proved using regression theory results. For part (c), recall that γ_δ depends on data indirectly, then, $u_{\delta_j1} \propto [\tilde{\delta}_j \mid \gamma_{\delta_j} = 1] p_{\delta_j}$ and $u_{\delta_j2} \propto [\tilde{\delta}_j \mid \gamma_{a,jk} = 0](1 - p_{\delta_j})$. Substituting the density of Inverse Gamma to the expressions above gives the formula (40). Note that the scale parameter cancels out and does not affect the conditional posterior of the model index γ_{δ_j} . For part (d), recall that γ_a depends on data indirectly. Then, $u_{a,jk1} \propto [\tilde{\mathbf{a}}_j \mid \gamma_{a,(-jk)}, \gamma_{a,jk} = 1] p_{a,jk}$, and $u_{a,jk2} \propto [\tilde{\mathbf{a}}_j \mid \gamma_{a,(-jk)}, \gamma_{a,jk} = 0](1 - p_{a,jk})$. The two expressions, together with prior independence of (a_{j1}, \dots, a_{jr}) , give the formula (42). Part (e) comes from direct computation. \square

3.4 Conditional Posterior of Λ

Updating the stochastic volatility λ is more complicated, because the latent volatility λ_{jt} is serially correlated and the full conditional of λ is not in closed-form. Uhlig (1997) using a Beta distribution for the ratio of the volatilities finds a closed-form for the conditional posterior. But his Gibbs sampler requires numerical integration. We simulate Λ through particle filters, using the following notation of the likelihood function of $(\mathbf{b}, \Psi, \Lambda_1, \dots, \Lambda_T)$:

$$L(\mathbf{b}, \Psi, \Lambda_1, \dots, \Lambda_T) = [\mathbf{y} \mid \mathbf{b}, \Psi, \Lambda_1, \dots, \Lambda_T] \propto \prod_{t=1}^T \prod_{j=1}^p \lambda_{jt}^{-\frac{1}{2}} \text{etr} \left\{ -\frac{1}{2} \sum_{t=1}^T \mathbf{Q}_t(\mathbf{b}, \Psi) \Lambda_t^{-1} \right\}, \quad (44)$$

where

$$\mathbf{Q}_t(\mathbf{b}, \Psi) \equiv \Psi \mathbf{S}_t(\mathbf{b}) \Psi' = (q_{t,ij})_{p \times p}, \quad (45)$$

and $\mathbf{S}_t(\mathbf{b})$ is given by (7). In SV literature, researchers often find it easier to simulate $h_{jt} = \log(\lambda_{jt})$ instead of λ_{jt} . We now turn to the particle filter approach to updating λ_{jt} .

3.4.1 Filtering and Smoothing of Conditional Posterior of Λ

Several authors apply particle filter on nonlinear state space models (e.g., Carter & Kohn, 1994, 1996), Chib et al. (2002) and (2006)). We apply the particle filter to stochastic search. Note that conditional on data $\mathbf{Y}_t = (\mathbf{y}_1, \dots, \mathbf{y}_t)$ and parameters $\boldsymbol{\theta}$, the density of Λ is

$$[\Lambda \mid \mathbf{y}, \boldsymbol{\theta}] = \prod_{t=1}^T [\Lambda_t \mid \mathbf{Y}_t, \boldsymbol{\theta}],$$

where

$$[\Lambda_t \mid \mathbf{Y}_t, \boldsymbol{\theta}] = \prod_{j=1}^p \lambda_{jt}^{-\frac{1}{2}} \text{etr} \left\{ -\frac{1}{2} \mathbf{Q}_t(\mathbf{b}, \Psi) \Lambda_t^{-1} \right\} = \prod_{j=1}^p \lambda_{jt}^{-\frac{1}{2}} \exp \left\{ \sum_{j=1}^p -\frac{1}{2} q_{tjj}(\mathbf{b}, \Psi) \lambda_{jt}^{-1} \right\}.$$

Recall that $h_{jt} = \log(\lambda_{jt})$. The particle filter is an algorithm based on the model and prediction and draws \mathbf{h}_t given $(\mathbf{y}_t, \mathbf{Y}_{t-1}, \boldsymbol{\theta})$,

$$[\mathbf{h}_t \mid \mathbf{y}_t, \mathbf{Y}_{t-1}, \boldsymbol{\theta}] \propto [\mathbf{y}_t \mid \mathbf{h}_t, \boldsymbol{\theta}] [\mathbf{h}_t \mid \mathbf{Y}_{t-1}, \boldsymbol{\theta}], \text{ for } t = 1, \dots, T.$$

An importance sampling particle filter is as follows.

Algorithm F. To initialize the filter, draw M particles of \mathbf{h}_0 from the stationary distribution implied by (3). Suppose at *Stage* t , we have $(\mathbf{h}_{t-1}^1, \dots, \mathbf{h}_{t-1}^M)$ drawn from $(\mathbf{h}_{t-1} | \mathbf{Y}_{t-1}, \boldsymbol{\theta})$.

Step 1. For $l = 1, \dots, M$, draw \mathbf{h}_t^l from

$$(\mathbf{h}_t^l | \mathbf{h}_{t-1}^l, \boldsymbol{\theta}) \sim N_p(\mathbf{a}_0 + \text{diag}(h_{1,t-1}^l, \dots, h_{p,t-1}^l)\boldsymbol{\beta} + \mathbf{A}_1 \mathbf{z}_t, \text{diag}(\delta_1^2, \dots, \delta_p^2)).$$

Step 2. For $l = 1, \dots, M$, compute $\boldsymbol{\Sigma}_t^l = \boldsymbol{\Gamma} \text{diag}(e^{h_{1t}^l}, \dots, e^{h_{pt}^l}) \boldsymbol{\Gamma}'$ and the importance weight

$$w_t^l = \frac{[\mathbf{y}_t | \boldsymbol{\Sigma}_t^l, \boldsymbol{\theta}]}{\sum_{l=1}^M [\mathbf{y}_t | \boldsymbol{\Sigma}_t^l, \boldsymbol{\theta}]}.$$

Step 3. Sample M draws of \mathbf{h}_t from $(\mathbf{h}_t^1, \dots, \mathbf{h}_t^M)$ with replacement and relative weight $\mathbf{w}_t = (w_t^1, \dots, w_t^M)$.

This completes *Stage* t . Continue with *Stage* $t + 1$ until *Stage* T .

The resampling step 3 over-samples particles with high importance weights. There will be more discussion on refinement of the algorithm later in this section.

For $t = 1, \dots, T$, filtering yields \mathbf{h}_t conditioning on the current observation \mathbf{Y}_t and the parameters $\boldsymbol{\theta}$. To update $\boldsymbol{\theta}$ given (\mathbf{h}, \mathbf{y}) in an MCMC algorithm, we need the entire series of stochastic volatility $\mathbf{h} = (\mathbf{h}_1, \dots, \mathbf{h}_T)'$ conditioning on the entire data set \mathbf{y} and parameter $\boldsymbol{\theta}$. The conditional posterior $[\mathbf{h} | \mathbf{y}, \boldsymbol{\theta}]$ is obtained through smoothing, using the Markovian structure of the model

$$[\mathbf{h}_T, \mathbf{h}_{T-1}, \dots, \mathbf{h}_1 | \mathbf{y}_T, \mathbf{y}_{T-1}, \dots, \mathbf{y}_1, \boldsymbol{\theta}] = [\mathbf{h}_T | \mathbf{Y}_T, \boldsymbol{\theta}] \prod_{t=1}^{T-1} [\mathbf{h}_t | \mathbf{h}_{t+1}, \mathbf{Y}_t, \boldsymbol{\theta}]. \quad (46)$$

Note that treating \mathbf{h}_{t+1} as observation and applying the Bayes rule, we have

$$[\mathbf{h}_t | \mathbf{h}_{t+1}, \mathbf{y}_t, \boldsymbol{\theta}] \propto [\mathbf{h}_t | \mathbf{Y}_t, \boldsymbol{\theta}] [\mathbf{h}_{t+1} | \mathbf{h}_t, \boldsymbol{\theta}],$$

where $[\mathbf{h}_{t+1} | \mathbf{h}_t, \boldsymbol{\theta}]$ is the Gaussian model by assumption of the stochastic volatility equation (3), and a numerical draw of $[\mathbf{h}_t | \mathbf{Y}_t, \boldsymbol{\theta}]$ is the result of filtering. Smoothing for $(\mathbf{h} | \mathbf{y}, \boldsymbol{\theta})$ is achieved by utilizing the recursive structure of (46) and implementing the following algorithm.

Algorithm S

Step 1. From the numerical result of filtering, draw $\mathbf{h}_T \sim (\mathbf{h}_T \mid \mathbf{Y}_T, \boldsymbol{\theta})$.

Step 2. For $t = T - 1, \dots, 1$, draw $\mathbf{h}_t \mid \mathbf{h}_{t+1}$ by reweighing the filtered $[\mathbf{h}_t \mid \mathbf{Y}_t, \boldsymbol{\theta}]$ by $[\mathbf{h}_{t+1} \mid \mathbf{h}_t, \boldsymbol{\theta}]$.

3.5 An MCMC Algorithm of Filtering and Smoothing of Vector-wise SV

From the result of the previous subsections, we develop the following algorithm for simulating posterior quantities $(\mathbf{b}, \gamma_b, \boldsymbol{\Psi}, \gamma_\psi, \mathbf{a}_0, \boldsymbol{\beta}, \mathbf{A}_1, \gamma_a, \boldsymbol{\delta}, \gamma_\delta, \mathbf{q}, \boldsymbol{\Lambda})$: Suppose at the end of cycle $k - 1$, we have $(\boldsymbol{\Psi}^{(k-1)}, \mathbf{b}^{(k-1)}, \mathbf{a}_0^{(k-1)}, \boldsymbol{\beta}^{(k-1)}, \mathbf{A}_1^{(k-1)}, \boldsymbol{\delta}^{(k-1)}, \gamma_b^{(k-1)}, \gamma_\psi^{(k-1)}, \gamma_a^{(k-1)}, \gamma_\delta^{(k-1)}, \mathbf{q}^{(k-1)}, \boldsymbol{\Lambda}^{(k-1)})$.

Step 1: Draw $(\gamma_\delta^{(k)} \mid \boldsymbol{\delta}^{(k-1)}, \mathbf{q}^{(k-1)})$ from (40);

Step 2: Draw $(\mathbf{q}^{(k)} \mid \boldsymbol{\delta}^{(k-1)}, \gamma_\delta^{(k)})$ from (43);

Step 3: Draw $(\mathbf{a}_0^{(k)}, \boldsymbol{\beta}^{(k)}, \mathbf{A}_1^{(k)})$: for $j = 1, \dots, p$, draw $(\tilde{\mathbf{a}}_j^{(k)} \mid \boldsymbol{\Lambda}^{(k-1)}, \gamma_a^{(k-1)}, \boldsymbol{\delta}^{(k-1)})$ from (37).

Step 4: Draw $(\boldsymbol{\delta}^{(k)} \mid \gamma_\delta^{(k)}, \gamma_a^{(k-1)}, \gamma_b^{(k-1)}, \gamma_\psi^{(k-1)}, \boldsymbol{\Lambda}^{(k-1)}, \mathbf{a}_0^{(k)}, \boldsymbol{\beta}^{(k)}, \mathbf{A}_1^{(k)}, \mathbf{q}^{(k)})$ from (38).

Step 5: For $j = 1, \dots, p$, draw $(\gamma_{a,j}^{(k)} \mid \mathbf{A}_1^{(k)})$ from the Bernoulli distribution (42).

Step 6: Draw $(\boldsymbol{\Lambda}^{(k)} \mid \mathbf{a}_0^{(k)}, \boldsymbol{\beta}^{(k)}, \mathbf{A}_1^{(k)}, \mathbf{b}^{(k-1)}, \boldsymbol{\delta}^{(k)}, \boldsymbol{\Psi}^{(k-1)}; \mathbf{y})$ based on the Algorithm F (in combination with Algorithm S).

Step 7: For $\boldsymbol{\Psi}^{(k)}$, draw $(\psi_j^{(k)} \mid \boldsymbol{\Lambda}^{(k)}, \mathbf{b}^{(k-1)}, \gamma_\psi^{(k-1)}; \mathbf{y})$ from (33), $j = 2, \dots, p$.

Step 8: Draw $(\gamma_\psi^{(k)} \mid \boldsymbol{\Psi}^{(k)})$ from the Bernoulli distribution (35).

Step 9: Draw $(\mathbf{b}^{(k)} \mid \boldsymbol{\Psi}^{(k)}, \boldsymbol{\Lambda}^{(k)}, \gamma_b^{(k-1)}; \mathbf{y})$ from the normal distribution (27).

Step 10: Draw $(\gamma_b^{(k)} \mid \mathbf{b}^{(k)})$ from the Bernoulli distribution (30).

3.6 Model Selection Based on MCMC Simulations

The MCMC simulation yields numerical draws of the posterior of model index $\boldsymbol{\gamma} = (\gamma_\delta, \gamma_a, \gamma_\psi, \gamma_b)$. Each draw of $\boldsymbol{\gamma}$ represents a selected submodel. To find the best submodel, George & McCulloch (1993) compute the sample posterior mode of $\boldsymbol{\gamma}$, corresponding to the one most visited in MCMC simulations. There are two difficulties to implementing this method in our study: first, the number

of models here is much larger and it is computationally expensive to store all the posterior draws of model index γ and calculate their frequencies. We solve this problem by observing that each draw of γ is a string of zeroes and ones; hence, it can be read as a binary number. It can be univocally recoded to decimal or hexadecimal notation with a much more compact form.

The second practical difficulty of selecting a submodel by the posterior mode of the model index is if the researcher wants to use stochastic search for model calibration, it is difficult to gauge the importance of each parameter. Our solution is to compute the posterior mean of each parameter in γ . The resultant value (between zero and one) reflects the importance of the single parameter. Of course, the model chosen based on the posterior mean of the model index may not be the one most frequently visited in MCMC (which is the true data generating model in most cases in our simulation study). Recognizing these practical challenges, we report the posterior mean of the model index as well as the frequency with which the true model is visited through stochastic search.

4 Simulation Studies

In this section we report the results of simulation studies. We simulate 100 samples from a SV model and evaluate the posterior of the parameters as well as model indexes obtained through Bayesian stochastic search. For each data sample we draw 10,000 MCMC cycles after 1,000 burn-in runs. The MCMC chains converge quite fast. The results are virtually the same if we change the burn-in runs to either 500 or 2,000.

4.1 A Benchmark Model with Sample Size $T=100$

We start with a small benchmark model of four variables with a moderate sample size ($T = 100$). In this case, stochastic search is shown effectively recovering the data generating parameters and the model restrictions. For each set of simulated observations, the running time for a MCMC chain with 11,000 cycles is roughly 5 minutes on an Intel Pentium 4 PC. Variations to the model

specifications are also considered, including the sample size and prior setting, as we explore the robustness and the performance of the methodology.

We simulate a VAR model with $p = 4$ variables and one lag giving the true parameters

$$\mathbf{\Psi} = \begin{pmatrix} 1 & .5 & .5 & .6 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \text{and } \mathbf{B} = \begin{pmatrix} .5 & .9 & .5 & 0 & 0 & 0 \\ .5 & 0 & 0 & .5 & 0 & 0 \\ .5 & 0 & 0 & 0 & .5 & 0 \\ .5 & .9 & 0 & 0 & 0 & .5 \end{pmatrix}.$$

The exogenous variables in the regression \mathbf{x} and the SV equation \mathbf{z} are both scalars, with $x_{it} = \cos(t/2)$, $z_{it} = \sin(t^2)$, for $i = 1, \dots, 4$ and $t = 1, \dots, T$. The true SV equation parameters are $\mathbf{A}_1 = (0, 0, 0, .4)$, $\boldsymbol{\beta} = (.9, .9, .8, .7)'$, $\boldsymbol{\alpha} = (.1, .1, .1, .1)'$ and $\boldsymbol{\delta} = (.1, .1, 1, 10^{-6})'$. The true parameter setting implies that the fourth component of the conditional variance is effectively non-stochastic. We start from the stationary distribution for fixed data generating parameters and take the eleventh observation as the initial observation.

Model selection is made on $\mathbf{\Psi}$ and \mathbf{b} in the regression and \mathbf{A}_1 and $\boldsymbol{\delta}$ in the stochastic volatility equation. We now report the posterior mean of model indexes and parameters. The posterior mean of model indexes (γ_Ψ, γ_b) averaged over 100 data samples are

$$\widehat{\gamma}_\Psi = \begin{pmatrix} * & \mathbf{.798} & \mathbf{.891} & \mathbf{.974} \\ * & * & .034 & .030 \\ * & * & * & .037 \\ * & * & * & * \end{pmatrix}, \quad \widehat{\gamma}_b = \begin{pmatrix} * & \mathbf{.911} & \mathbf{.860} & .032 & .038 & .039 \\ * & .110 & .047 & \mathbf{.785} & .044 & .043 \\ * & .096 & .037 & .030 & \mathbf{.783} & .033 \\ * & \mathbf{.865} & .045 & .029 & .033 & \mathbf{.881} \end{pmatrix}.$$

The boldfaced elements in the estimated model index matrices correspond to unity and the rest elements correspond to zero in the true data generating model. From the estimates of γ_Ψ and γ_b , it appears that the stochastic search approach recovers the true model reasonably well. The averages of the 100 posterior means of parameters are

$$\widehat{\mathbf{\Psi}} = \begin{pmatrix} 1 & .463 & .478 & .599 \\ 0 & 1 & .002 & .002 \\ 0 & 0 & 1 & -.002 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \widehat{\mathbf{B}} = \begin{pmatrix} .542 & .961 & .442 & -.004 & .005 & .012 \\ .579 & -.005 & -.012 & .424 & -.004 & .010 \\ .583 & .008 & -.002 & .006 & .422 & .006 \\ .512 & .911 & .020 & .004 & .003 & .459 \end{pmatrix}.$$

Again, these estimates are close to the true data generating parameters specified in the beginning of the example, with the exception of the well known downward bias in the AR coefficients of regression model.

Now we turn to the parameters of the SV equation. The averages of the posterior mean of intercepts and AR parameters of the SV equations are $\boldsymbol{\alpha} = (.135, .136, .118, .098)'$, and $\boldsymbol{\beta} = (.865, .864, .766, .703)'$; close to the true parameters $\boldsymbol{\alpha} = (.1, .1, .1, .1)'$ and $\boldsymbol{\beta} = (.9, .9, .8, .7)'$.

(Insert Table 1 around here.)

The model restriction and the magnitude of AR coefficient in the SV equation, \mathbf{A}_1 , are estimated with a high precision. The model selection of the SV parameter $\boldsymbol{\delta}$ is also effective. All stochastic components are correctly selected. The non-stochastic component is incorrectly selected only 2.5 percent of the time. The selection of the SV parameter $\boldsymbol{\delta}$ extends the stochastic search approach originated by George & McCulloch (1993). The innovation lies in specification of a hierarchical prior on hyper-parameter \mathbf{q} . The simulation results in Table 1 are based on parameter \mathbf{q} , whose prior specification and posterior computation were discussed in Sections 2.3 and 3.3, respectively.

4.2 The Same Model with Sample Size $T=1000$.

Now we explore the performance of the stochastic search method with a larger sample size T . For the same model and parameters, the frequency of visiting the true data generating model increases with T . For $T = 1,000$, the parameters of the SV equation are very precisely estimated. The averages of the posterior means of intercepts and AR parameters of the SV equations are $\boldsymbol{\alpha} = (.106, .105, .104, .100)'$, and $\boldsymbol{\beta} = (.895, .895, .792, .701)'$; much closer to the true parameters $(.1, .1, .1, .1)'$ and $(.9, .9, .8, .7)'$ than those for $T = 100$. The posterior means (averaged over 100 samples) of the model index and the regression parameters are also closer to the true values as the sample size increases to 1,000.

$$\hat{\gamma}_\psi = \begin{pmatrix} * & \mathbf{.999} & \mathbf{.999} & \mathbf{1.00} \\ * & * & .021 & .020 \\ * & * & * & .021 \\ * & * & * & * \end{pmatrix}, \quad \gamma_b = \begin{pmatrix} * & \mathbf{1.00} & \mathbf{.999} & .021 & .022 & .021 \\ * & .037 & .022 & \mathbf{.999} & .022 & .022 \\ * & .030 & .021 & .021 & \mathbf{.999} & .021 \\ * & \mathbf{1.00} & .022 & .021 & .022 & \mathbf{.999} \end{pmatrix}.$$

$$\hat{\Psi} = \begin{pmatrix} 1 & .503 & .496 & .600 \\ 0 & 1 & -.002 & .000 \\ 0 & 0 & 1 & .002 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \hat{B} = \begin{pmatrix} .507 & .907 & .498 & .006 & -.002 & -.004 \\ .516 & -.003 & -.003 & .494 & -.005 & -.003 \\ .509 & -.003 & .001 & -.004 & .495 & -.002 \\ .497 & .897 & .004 & -.005 & .006 & .498 \end{pmatrix}.$$

To visualize the accuracy of the posterior of SV, in Figure 1 we plot the posterior means of SV components h_{1t} and h_{2t} against the generated values (averaged over t) for 100 simulated data samples, with $T = 100$ and 1,000, respectively. The posterior means of h_{1t} and h_{2t} are the average over MCMC cycles and over time t . The figure shows that when the sample size $T = 100$, the posterior mean of SV is a fair estimate of the simulated SV, with a moderate upward bias. For a larger sample size $T = 1,000$, the average of the posterior mean becomes very close to the true SV.

To further examine the property of simulated SV, we plot the time series of true SV against the posterior of SV. We select a random sample (by using the 100th sample) of generated data with $T = 1,000$, and record the posterior of h_{jt} for $j = 1, \dots, 4, t = 1, \dots, 1,000$. To gain a better view, we plot in Figure 2 the middle portion of the time series ($t = 400$ to 600) of the true SV (h_{1t} in panel (a) and h_{4t} in panel(b)), the posterior mean, and the posterior mean plus and minus 2 times simulated posterior standard deviation bands (roughly 5 percentile to 95 percentile confidence bands for normally distributed random variables). The figure shows that the posterior of SV bands tracks the true SV quite well over time. In the figure, the standard deviation of the posterior of h_{1t} does not shrink over time, which suggests that the re-sampling step of the particle filter successfully prevented degeneration. The deterministic conditional variance h_{4t} is estimated with high precision, the posterior standard deviation is visually negligible compared to the variation of the conditional variance over time.

Next, to evaluate the performance of model selection, we further examine the posterior of model indices. Earlier we reported the posterior mean of the model index for selecting $(\mathbf{B}_1, \Psi, \mathbf{A}_1, \delta)$. One approach to selecting the best model is to include a variable when the posterior mean of the corresponding index is above a threshold. If we set such a threshold (say .9) in the above example for a large sample size ($T = 1,000$), then the true data generating model is selected. If the sample size T is much smaller then setting a high threshold may exclude too many variables. The threshold should be chosen judiciously, depending on the sample size of the data.

An alternative criterion to the posterior mean of model index is the posterior mode, which corresponds to the most frequently visited model. If stochastic search is successful the most visited model is the data generating model. The number of parameters subject to stochastic search is 34, for a total number of 2^{34} competing models. For 10,000 MCMC runs, the algorithm visits several hundred different submodels. For all data samples the following table summarizes the average frequencies that the true model is visited. As the sample size T increases, the frequency of visiting the true data generating model for Ψ and \mathbf{B}_1 increases. But the frequency of visiting the true SV model is not monotonically increasing in the sample size in our experiment.

(Insert Table 2 around here.)

Note that the frequencies are averaged over 100 samples (the cross-sample standard deviations of the frequencies are reported in the parentheses.) So the averaged frequency of matching the whole data-generating model may be larger or smaller than the product of the averaged frequencies of matching each part of the model. Nonetheless, these numbers indicate that the stochastic search algorithm is highly effective in selecting the data generating model even with a modest sample size of $T = 100$. When $T = 100$, the true model is visited a remarkable 11.2% of the time. With a large sample size $T = 1,000$, the average odds of visiting the true model are overwhelming.

It is well known that an unrestricted model (setting all model indexes to unity) is often overfitted and performs poorly in forecasting. We now compare the forecasting performance between an unrestricted model and restricted models obtained from stochastic search. We obtain Bayesian

estimates under the unrestricted model and that through Bayesian stochastic search. We use the estimated parameters to make one-period-ahead forecast of $y_{j,T+1}$, for $j = 1, \dots, 4$. We define $\hat{y}_{j,T+1}^{(true)} = b_{j0} + \mathbf{B}_{j1} \mathbf{x}_{T+1}$, the prediction based on \mathbf{x}_{T+1} and the j th row of the true parameters of $(\mathbf{b}_0, \mathbf{B}_1)$. Let $\hat{y}_{j,T+1}^{(k)}$ be the prediction based on the Gibbs output of the parameters at cycle k , $(\Psi^{(k)}, \mathbf{b}^{(k)}, \mathbf{a}_0^{(k)}, \beta^{(k)}, \mathbf{A}_1^{(k)}, \delta^{(k)}, \gamma_b^{(k)}, \gamma_\psi^{(k)}, \gamma_a^{(k)}, \gamma_\delta^{(k)}, \mathbf{q}^{(k)}, \Lambda^{(k)})$. In fact, $\hat{y}_{j,T+1}^{(k)}$ depends explicitly only on $\mathbf{b}^{(k)}$ and is given by $\hat{y}_{j,T+1}^{(k)} = b_{j0}^{(k)} + \mathbf{B}_{j1}^{(k)} \mathbf{x}_{T+1}$. We define the prediction squared errors to be

$$\frac{1}{K} \sum_{k=1}^K \left[\hat{y}_{j,T+1}^{(true)} - \hat{y}_{j,T+1}^{(k)} \right]^2, \quad (47)$$

which depends only on the estimation errors of the parameters. We average over all MCMC cycles through stochastic search. This is an alternative to a typical Bayesian model averaging (BMA) (see for example Hoeting et al. (1999)). In BMA of N competing models, posteriors of parameters would have to be simulated for each of the N models and the predicted value would then be weighted by the posterior of N model probabilities. For the example above when $p = 4$, $N = 2^{34}$. Clearly, it is impossible to compute all these competing models. An advantage of stochastic search in the article is to draw parameters conditioning on visited models. The frequently visited models are weighted heavily automatically. In contrast to the multiple models visited in stochastic search, the unrestricted model space consists of a single fixed model.

There are other methods besides BMA in the literature of forecasting. For example, Breiman (1996) has proposed an approach named bagging by averaging predictions based on estimates derived from bootstrapped data. It has been shown to improve forecasting in macroeconomics. For example, Inoue & Kilian (2008) show that for forecasting consumer price inflation using 30 macroeconomic indicators, bagging performs better than equal weighted averaging of all prediction models based on each indicator and does equally well as BMA. Although both stochastic search and bagging can be used for model selection, there are several important differences. One advantage of stochastic search is that the resultant model restrictions may have economic interpretations while the averaged prediction from bagging is often *ad hoc* and does not correspond to an economic

model. In addition, bagging is suitable for forecasting a scalar variable with a large number of exogenous predictors. Inoue & Kilian (2008) include lags of inflation in their model and limited bagging to exogenous predictors. The presence of SV further complicates bootstrapping of serially correlated data. Hence, for restricting a VAR with SV, stochastic search is much more convenient. Of course, it is interesting to compare the performance of alternative approaches to forecasting besides stochastic search. We will explore this in the future.

(Insert Table 3 around here.)

Table 3 compares the prediction errors (47) averaged over 100 data samples for the restricted (stochastic search) model with those for the unrestricted models. The average prediction errors of the restricted model are about thirty percent smaller than the unrestricted model when $T = 100$ and around twenty percent smaller when $T = 1,000$. As the sample size increases, the prediction error and the improvement by stochastic search decrease. Maheu & Liu (2009) show that the BMA improves the prediction performance of GARCH models. We find this to be true for SV models. In Figure 3, we plot the frequentist densities of prediction errors of the posterior mean of y_{4T+1} , defined for a given sample as $\frac{1}{K} \sum_{k=1}^K \hat{y}_{4,T+1}^{(k)} - \hat{y}_{4,T+1}^{(true)}$, for $T = 100$, for restricted and unrestricted models. The density of unrestricted model in Figure 3(a) shows greater dispersion than that with restricted model in Figure 3(b). The figure agrees with the results reported in Table 3.

As in Maheu & Liu (2009), we also compute the predictive likelihood of one-step-ahead out-of-sample prediction $L(\mathbf{y}_{T+1} | \mathbf{y}_1, \dots, \mathbf{y}_T)$, after integrating out parameters (including SV) and models. This quantity measures overall predictive fit of the models to the data set. The stochastic search also dominates in terms of predictive likelihood: When $T = 100$ the predictive likelihood for one-step-ahead forecast averaged over all data samples is -5.371 for the unrestricted model and is -5.342 for stochastic search.

The web appendix contains a brief discussion on the robustness of the algorithm to the prior hyper-parameters and the number of observations.

4.3 Discussion on Alternative Approaches to Inference of SV

In the particle filter algorithm, a multinomial resampling adds noise that may distort the drawn filter density from the true density, especially at the tails. The auxiliary sampling importance resampling filter of Pitt & Shephard (1999b) refines the particle filter to produce a more accurate approximation of the true filter density. For the example at hand, we find that the refined resampling produces similar results as Algorithm F.

We now briefly discuss three approaches as alternatives to the one proposed in this study for inference of SV: direct sampling using the Gilks adaptive rejection sampler, Gaussian approximation of SV model by Kim et al. (1998) (who also use particle filter for evaluation of likelihood), and rejection sampling of element-wise volatilities through the filtered volatilities.

The direct sampling approach through *Adaptive Rejection Sampler* by Gilks & Wild (1992) is an alternative to the filter approach when the posterior of each h_{jt} conditional on other parameters is log-concave. In our example, the running time for the adaptive rejection method based on log-concavity is on average about half that of the particle filter algorithm. However, the generated samples of stochastic volatility \mathbf{A} from the the rejection method exhibit a strong serial correlation over the MCMC cycles. It takes more MCMC cycles to achieve a result similar to the particle filter algorithm. Pitt & Shephard (1999a) show that the convergence of a single-move sampling algorithm for h_{jt} conditional on $(\mathbf{b}, \Psi, \mathbf{a}_0, \boldsymbol{\beta}, \mathbf{A}_1, \boldsymbol{\delta}, h_{j,t-1}, h_{j,t+1}; \mathbf{y})$ is very slow, especially when the SV has a strong correlation and the variance of the SV is approximately zero. We find that the particle-filter-based block-move algorithm is quite efficient for such cases. Figure 2 (b) shows that the estimated conditional variance is nearly identical to the ones that generated data, even when the number of burn-in cycles is small, the SV has a large AR coefficient and a near zero variance.

To increase the efficiency of the computation over the single-move rejection-based method, Kim et al. (1998) and Chib et al. (2002) approximate the SV model by a Gaussian-linear model and evaluate the marginal likelihood by using a particle filter. As pointed out by a referee, particle

filters play different roles in our study from that in Kim et al. (1998) and Chib et al. (2002). In the latter studies, particle filters are used to compute predictive likelihood $f(y_{t+1} | h_t, \mathbf{Y}_t, \boldsymbol{\theta})$ given the model parameters $\boldsymbol{\theta}$, where the posterior of $\boldsymbol{\theta}$ is simulated from MCMC in a mixture Gaussian state-space model that approximates the original SV model. In our study, particle filters are used to simulate SV jointly with model parameters $\boldsymbol{\theta}$. This is made possible by doing backward-smoothing following forward-filtering.

The importance sampling particle filter described above involves sampling and weighted re-sampling of vector-wise particles for every Gibbs cycle. Using the fact that the conditional distributions of $(h_{jt} | h_{j,t-1}, \boldsymbol{\theta})$ are independent across $j = 1, \dots, p$, we modify the rejection sampling algorithm specified in Kim et al. (1998) for our multivariate model. Sampling the vector \mathbf{h}_t at once is equivalent to sample from the posterior of h_{jt} element-wise. We find that the element-wise rejection method takes longer time to complete a given number of cycles.

5 An Application to Daily Foreign Exchange Rate Data

Kim et al. (1998) conduct Bayesian estimation on an univariate model of stochastic volatilities of the exchange rates. We extend their analysis by exploring how the exchange rates are correlated and whether they are driven by stochastic volatility in a SV VAR model (with one lag).

We examine the daily exchange rates of the U.S. Dollar against the Euro, British Pound, Japanese Yen, Canadian Dollar, Mexican Peso, Brazilian Real, Switzerland Franc, Norwegian Krone, Singapore Dollar and Indian Rupee (with this order in the VAR) between January 2001 and November 2005. The data set are obtained from the Federal Reserve Bank of St Louis. To make our result comparable to that by Kim et al. (1998), we focus on the demeaned log differences of exchange rate as they do. Prior hyperparameters are the same as those in the numerical example, except the volatility prior parameters are set as $s_0 = 10^{-8}$, $\alpha_q = 0.1$, and $\beta_q = 0.05$. We draw 50,000 MCMC cycles with 10,000 burn-in runs. Figure 4 is a scatter plot of the posterior mean

of the non-redundant elements of Ψ (defined in (5)) and the corresponding model index γ_ψ . The plot of Ψ and γ_ψ suggests a strong correlation among the exchange rates. Figure 4 also plots the posterior mean of regression coefficient matrix $\mathbf{B} = (\mathbf{b}_0, \mathbf{B}_1)$ and the model index of \mathbf{B}_1 , γ_b . All the elements of \mathbf{B} and γ_b are close to zero. The figure supports the restriction imposed by Kim et al. (1998) that daily exchange rate returns are serially uncorrelated. Without imposing the restriction a priori as they do, we start from a more general VAR model and arrive at the same conclusion via model selection. The posterior mean of the stochastic volatility and that of the corresponding model selection index are $\hat{\boldsymbol{\delta}} = (.0002, .0010, .0147, .0018, .0129, .1693, .0158, .0068, .0141, .0089)'$, $\hat{\boldsymbol{\gamma}}_\delta = (.8930, 1.0000, 1.0000, 1.0000, 1.0000, 1.0000, 1.0000, 1.0000, .9780, 1.0000)'$. Although the posterior mean of the SV parameter vector $\boldsymbol{\delta}$ is small in magnitude, all elements of $\boldsymbol{\delta}$ are selected at high frequencies. We conclude that the daily returns of exchange rates of dollar against major foreign currencies are serially uncorrelated and contemporaneously correlated, and that their conditional variances are stochastic.

6 Discussion

In this study we develop and implement an algorithm for Bayesian model selection of multivariate stochastic volatility models. We extend the stochastic search method of George & McCulloch (1993) for linear regression, and George et al. (2008) for multivariate time series with fixed covariances. Our method applies to time varying covariances and provides a computationally feasible algorithm to select among a large number of competing models. Our simulation studies show that the particle filter-smoothing algorithm is effective for a small scale model with a moderately large sample size. We find that restricted models obtained through stochastic search outperform unrestricted models in out-of-sample forecasts.

A number of generalizations are possible. First, a nonindependent setup of the Bernoulli hyper priors would allow great flexibility. Second, the model can be extended to allow jumps and/or

non-Gaussian errors with heavy tails (see Chib et al. (2006)). Finally, the modified Cholesky decomposition used in the paper enjoy advantages such as computational feasibility and easy statistical interpretability of the parameters, but other parameterizations could be considered (e.g., those by Chiu et al. (1996) or Chen & Dunson (2003)). A comparison of different decompositions in multivariate stochastic volatility models may lead to interesting future developments.

Acknowledgement: Ni's research was supported by a grant from the MU Research Council and by Hong Kong Research Grant Council grant CUHK4128/03H. Sun's research was supported by the National Science Foundation grant SES-0720229, and NIH grants R01-MH071418 and R01-CA112159.

References

- Aguilar, O. & West, M. (2000), 'Bayesian dynamic factor models and portfolio allocation', *Journal of Business and Economic Statistics* **18**(3), 338–357.
- Bauens, L., Laurent, S. & Rombouts, J. V. K. (2004), Multivariate GARCH models: a survey. CORE Discussion Paper.
- Black, F. & Scholes, M. (1973), 'The pricing of option and corporate liabilities', *Journal of Political Economy* **81**, 637–659.
- Bollerslev, T., Chou, R. Y. & Kroner, K. F. (1992), 'ARCH modeling in finance. A review of the theory and empirical evidence', *Journal of Econometrics* **52**, 5–59.
- Boscher, H., Fronk, E.-M. & Pigeot, I. (2000), 'Forecasting interest rates volatilities by GARCH (1, 1) and stochastic volatility models', *Statistical Papers* **41**(4), 409–422.
- Breiman, L. (1996), 'Bagging predictors', *Machine Learning* **24**, 123–140.
- Carter, C. K. & Kohn, R. (1994), 'On Gibbs sampling for state space models', *Biometrika* **81**, 541–553.
- Carter, C. K. & Kohn, R. (1996), 'Markov chain Monte Carlo in conditionally Gaussian state space models', *Biometrika* **83**, 589–601.

- Chen, Z. & Dunson, D. B. (2003), ‘Random effects selection in linear mixed models’, *Biometrics* **59**(4), 762–769.
- Chib, S., Nardari, F. & Shephard, N. (2002), ‘Markov chain Monte Carlo methods for stochastic volatility models’, *Journal of Econometrics* **108**, 281–316.
- Chib, S., Nardari, F. & Shephard, N. (2006), ‘Analysis of high dimensional multivariate stochastic volatility models’, *Journal of Econometrics* **134**, 341–371.
- Chiu, T. Y. M., Leonard, T. & Tsui, K.-W. (1996), ‘The matrix-logarithmic covariance model’, *Journal of the American Statistical Association* **91**, 198–210.
- Christensen, B. & Kiefer, N. M. (2000), Anticipated and unanticipated stochastic volatility in empirical option pricing. Working Paper.
- George, E. I. & McCulloch, R. E. (1993), ‘Variable selection via Gibbs sampling’, *Journal of the American Statistical Association* **88**, 881–889.
- George, E., Sun, D. & Ni, S. (2008), ‘Stochastic search model selection for restricted VAR models’, *Journal of Econometrics* **142**, 553–580.
- Gilks, W. R. & Wild, P. (1992), ‘Adaptive rejection sampling for Gibbs sampling’, *Applied Statistics* **41**, 337–348.
- Hoeting, J. A., Madigan, D., Raftery, A. E. & Volinsky, C. T. (1999), ‘Bayesian model averaging: a tutorial (Pkg: p382-417)’, *Statistical Science* **14**(4), 382–401.
- Hol, E. & Koopman, S. J. (2002), Stock index volatility forecasting with high frequency data. Tinbergen Institute Discussion Paper.
- Hull, J. & White, A. (1987), ‘The pricing of option on assets with stochastic volatilities’, *Journal of Finance* **42**, 281–300.
- Inoue, A. & Kilian, L. (2008), ‘How useful is bagging in forecasting economic time series? A case study of U.S. consumer price inflation’, *Journal of the American Statistical Association* **103**, 511–522.

- Jacquier, E., Polson, N. G. & Rossi, P. E. (1994), ‘Bayesian analysis of stochastic volatility models (Disc: pp389-417)’, *Journal of Business and Economic Statistics* **12**, 371–389.
- Jacquier, E., Polson, N. G. & Rossi, P. E. (2002), ‘Bayesian analysis of stochastic volatility models’, *Journal of Business and Economic Statistics* **20**, 69–87.
- Justiniano, A. & Primiceri, G. E. (2008), ‘The time varying volatility of macroeconomic fluctuations’, *American Economic Review* **98**, 604–641.
- Kim, H., Sun, D. & Tsutakawa, R. K. (2002), ‘Lognormal vs. gamma: Extra variations’, *Biometrical Journal* **44**(3), 305–323.
- Kim, S., Shephard, N. & Chib, S. (1998), ‘Stochastic volatility: likelihood inference and comparison with arch models’, *Review of Economic Studies* **65**, 361–394.
- Maheu, J. M. & Liu, C. (2009), ‘Forecasting realized volatility: A bayesian model averaging approach’, *Journal of Applied Econometrics* **24**, 709–733.
- Melino, A. & Turnbull, S. M. (1990), ‘Pricing foreign currency options with stochastic volatility’, *Journal of Econometrics* **45**, 239–265.
- Pitt, M. K. & Shephard, N. (1999a), ‘Analytic convergence rates and parameterization issues for the Gibbs sampler applied to state space models’, *Journal of Time Series Analysis* **20**, 63–85.
- Pitt, M. K. & Shephard, N. (1999b), ‘Filtering via simulation: Auxiliary particle filters’, *Journal of the American Statistical Association* **94**, 590–599.
- Shephard, N. G. (1996), ‘Statistical aspects of ARCH and stochastic volatility’, *Time Series Models in Econometrics, Finance and Other Fields, Monographs on Statistics and Applied Probability* **65**, 1–67.
- Tauchen, G. (2005), Stochastic volatility in general equilibrium. Working Paper.
- Taylor, S. (1994), ‘Modeling stochastic volatility: A review and comparative study’, *Mathematical Finance* **4**, 183–204.
- Uhlig, H. (1997), ‘Bayesian vector autoregressions with stochastic volatility’, *Econometrica* **65**, 59–73.

Table 1: Performance of Selection of $(\mathbf{A}_1, \boldsymbol{\delta})$ for $T = 100$, Averaged Over 100 Data Samples.

true \mathbf{A}_1	0	0	0	.4
average of posterior means \mathbf{A}_1	-.017	.011	.027	.401
true γ_{A_1}	0	0	0	1
average of posterior means γ_{A_1}	.021	.023	.023	.983
true $\boldsymbol{\delta}$.1	.1	.1	10^{-6}
average of posterior means $\boldsymbol{\delta}$.095	.096	.095	.012
true γ_δ	1	1	1	0
average of posterior means γ_δ	1.000	1.000	1.000	.025

Table 2: Average Frequency of Visiting the True Model ($p = 4$) (cross-sample standard deviations in parentheses).

parameter	entire model $(\mathbf{B}_1, \boldsymbol{\Psi}, \mathbf{A}_1, \boldsymbol{\delta})$	\mathbf{B}_1	$\boldsymbol{\Psi}$	$(\mathbf{A}_1, \boldsymbol{\delta})$
# of parameters	34	20	6	8
$T = 100$.112 (.089)	.197 (.136)	.623 (.244)	.901 (.003)
$T = 1,000$.598 (.016)	.718 (.018)	.937 (.003)	.889 (.007)

Table 3: Comparison of Squared Prediction Errors (47) Based on Stochastic Search (SS) and Those Based on Unrestricted Model, Averaged over 100 Data Samples.

Model	pred. error y_1	pred. error y_2	pred. error y_3	pred. error y_4
SS, $T = 100$.1358	.1419	.1160	.1160
No Restriction, $T = 100$.1814	.2330	.1803	.1714
% improvement with SS	25.14%	39.10%	35.50%	32.3%
SS, $T = 1,000$.0166	.0136	.0107	.0158
No Restriction, $T = 1,000$.0201	.0182	.0131	.0181
% improvement with SS	17.41%	25.27%	18.32%	12.71%

Web Appendix: Robustness to Model Specification and Hyper-parameters

When sample size T is smaller for a given model or when the dimension of the model is larger (e.g., $p \geq 6$) for a given sample size, the frequency of visiting the true model is lower. From Table 3, when the sample size is small, the estimates from the unrestricted model are more erratic and the percentage of improvement of using the stochastic search is larger. Our experiments also show that when the sample size T is small, in general the posterior mean of the model index is more reliable than the posterior mode, and BMA is more advisable than relying on a single selected model.

Of course, the reported results depend on the hyper-parameters. For a small sample size T , although the posterior of δ may not be very sensitive to the IG prior on δ , the model selection index for δ can vary with the prior. For most of the parameters, the reported results hold qualitatively when we alter the priors though. For instance, when we reduce the prior variance for the AR coefficient of the SV equation from unity to that implied by the Kim et al. (1998), the simulation results are almost identical to those reported. Some hyper-parameters play a critical role in shaping the results for model selection. For example, the ratio of the variance of selected and excluded parameters in Ψ and B_1 is 50. If the ratio is reduced to 10, the distinction between exclusion and inclusion is smaller. Consequently, the chance that the true model being selected may be lower. On the other hand, the reduced ratio may lead to better inference if the prior correctly reflects the nature of the data-generating parameters. In applications of the stochastic search method, researchers should consider the trade-off in prior elicitation. In the absence of prior information of whether a variable should be included, we set the Bernoulli prior with probability of inclusion to .5. A more experienced researcher may prefer to set this probability different from .5 to reflect the number of anticipated model restrictions.

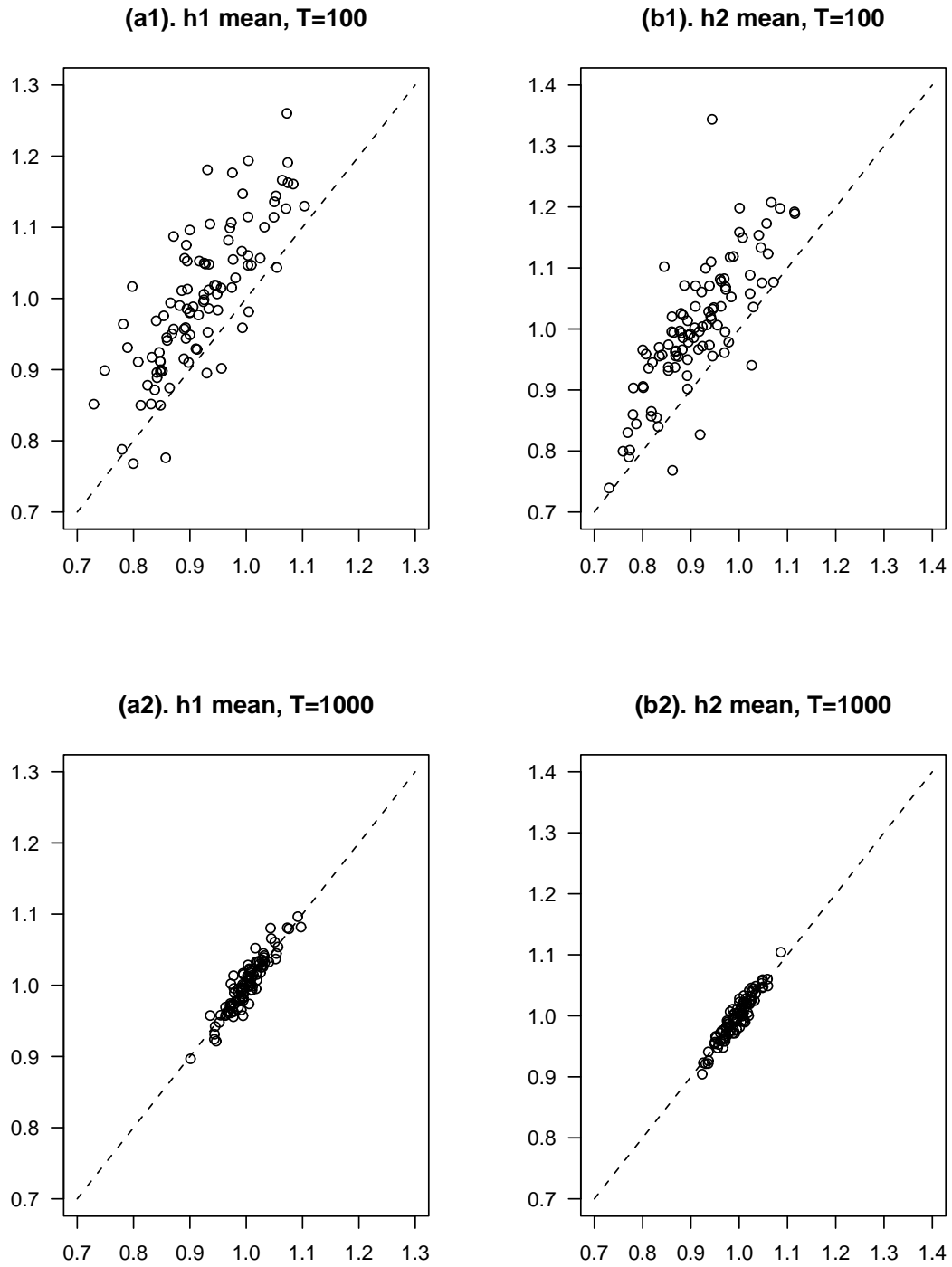
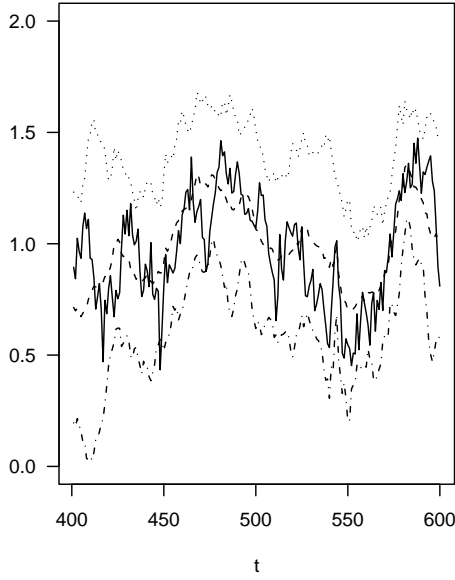


Figure 1: Posterior means and generated true values of sample average of h_1 with (a1) $T = 100$ and (a2) $T = 1,000$; and h_2 with (b1) $T = 100$ and (b2) $T = 1,000$. The horizontal axis represents the generated SV and the vertical axis plots the corresponding posterior mean of SV, both are averaged over time for each of the 100 samples. The dashed line is the 45 degree line.

(a) h_{1t} , $T=1000$, one sample, $t=400$ to 600



(b) h_{4t} , $T=1000$, one sample, $t=400$ to 600

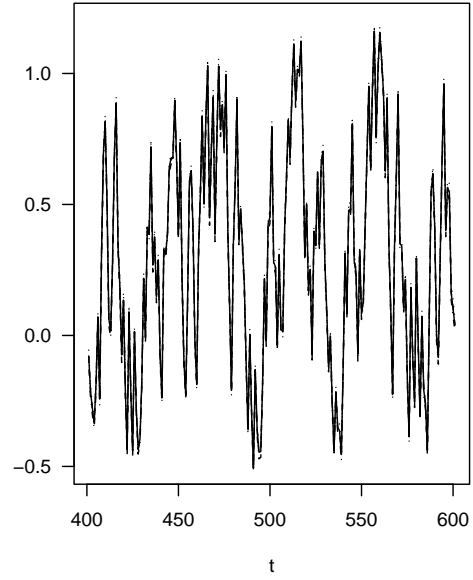
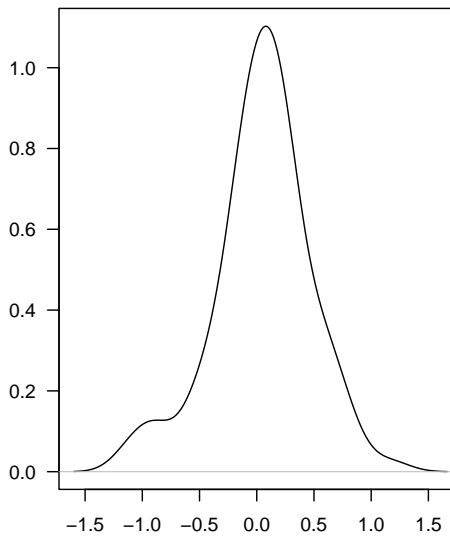


Figure 2: h_{1t} (in panel (a)) and h_{4t} (in panel (b)) for one sample, $t = 400, \dots, 600$. Solid=true h_t , dashed=posterior mean of h_t , dotted = posterior mean + 2 times standard posterior deviation of h_t , dot-dashed = posterior mean - 2 times standard posterior deviation of h_t .

(a)



(b)

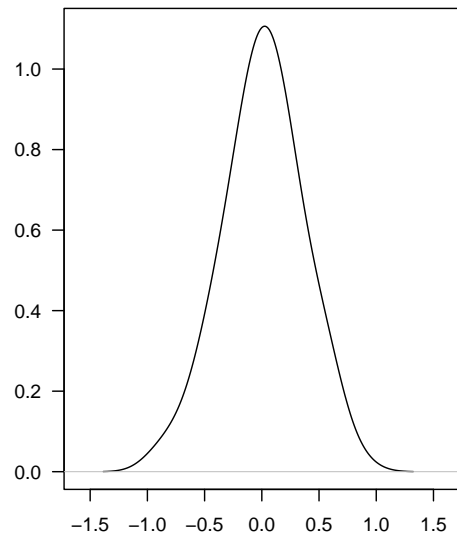


Figure 3: Smoothed Cross Sample Distribution of One-step-ahead Average Prediction Errors of y_{4T+1} Produced by Unrestricted and Restricted Models. The size of each sample $T = 100$. (a). The distribution of prediction errors of unrestricted model. (b). The distribution of prediction errors of restricted model obtained via stochastic search.

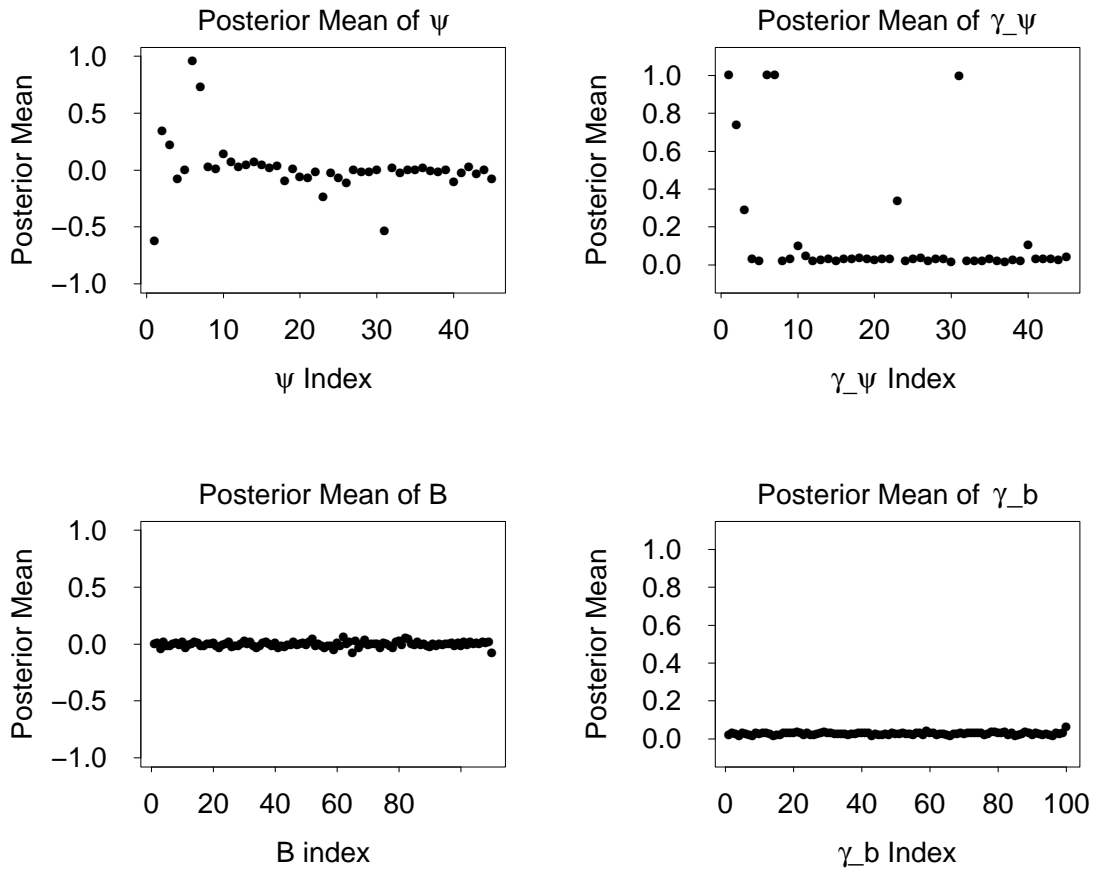


Figure 4: Stochastic Search of SV VAR(1) Using Daily Exchange Rate Data: Posterior Means for Ψ , γ_ψ , B , and γ_b . The matrices are vectorized by row.