BAYESIAN ANALYSIS OF MULTIVARIATE STOCHASTIC VOLATILITY AND DYNAMIC MODELS

A Dissertation

presented to

the Faculty of the Graduate School

University of Missouri-Columbia

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

---

by

ANTONELLO LODDO

Dr. Dongchu Sun, Dissertation Supervisor

AUGUST 2006

The undersigned, appointed by the Dean of the Graduate School, have examined the thesis entitled

BAYESIAN ANALYSIS OF MULTIVARIATE STOCHASTIC VOLATILITY AND DYNAMIC MODELS

presented by ANTONELLO LODDO

a candidate for the degree of Doctor of Philosophy

and hereby certify that in their opinion it is worthy of acceptance.

Professor Dongchu Sun _____

Professor Paul L. Speckman _____

Professor Christopher K. Wikle _____

Professor (Chong) Zhuoqiong He _____

Professor (Shawn) Xiaoguang Ni _____

ACKNOWLEDGEMENTS

I would like to address special thanks to my advisor Professor Sun for his support and advise during these years, and for being such an important figure in my growth as a statistician.

I also would like to thank my committee members for their insightful comments and suggestions: Professor Speckman, Professor Wikle, Professor He and Professor Ni.

I would like to express my gratitude to Professor Ni for his patience and help during my advisor's leave.

Many thanks to Walter Racugno, my first statistics professor and mentor.

This dissertation would have been much harder to be completed without the help of Richard Smemoe, ho always quickly solved any problem related to the server and network.

I am also very grateful to all the people that supported me during these years, without whose love and compassion I would probably not be writing this today: my family, my friend Riccardo, and Yasemin, who shared the worse, the best, and the rest of me during the last two and a half years. God sent you to me.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

Introduction

Research in quantitative methods for economics evolved dramatically over the last few decades. Earlier studies simply applied regression models, while today cutting edge time series analysis techniques are developed and employed. Parallel with these changes, econometricians face brand new puzzles: estimating unconstrained multivariate time series, dealing with heteroscedastic models and estimating latent processes appear to be crucial ones.

The necessity of estimating unconstrained multivariate time series comes from the desire to avoid any arbitrary classification between endogenous and exogenous variables. The question is this: when you have several correlated time series, which variables can be considered independent and which dependent? Simultaneous Equation (SE) models, which have been very popular for decades, are based on this distinction. Quoting Sims (1980):

> ...models end up with very different sets of variables on the right-side of these equations, they do so not by invoking economic theory, but (in the case of demand equations) by invoking an intuitive, econometrician's version of psychological and sociological theory.

Sims overcame this problem by introducing a new generation of models, known as Vector Autoregressive (VAR) models, in which a priori restrictions are reduced to the minimum:

> ...it should be feasible to estimate large scale macromodels as unrestricted reduced forms, treating all variables as endogenous. Of course, restrictions, if only on the lag length, are essential, so by "unrestricted" here I mean "without restrictions based on supposed a priori knowledge" (Sims (1980)).

The popularity of VAR models increased dramatically in the 1980s, and today they represent an almost indispensable tool for any quantitative economist. An overview of VAR models is presented in Section 2.1

Models for heteroscedastic time series can be dated back to the seminal work of Engle (1982), who extended traditional time series tools such as autoregressive moving average (ARMA) models (Box & Jenkins (1976)) for the mean to essentially analogous models for the variance. Autoregressive Conditional Heteroscedasticity (ARCH) models are now commonly used to forecast changes in the volatility of financial time series. For his contribution, Engle was awarded of the Nobel prize for economics in 2003. Since then, other models for time varying variances have been developed and applied, like Generalized ARCH (GARCH) models (Bollerslev (1986)), or stochastic volatility models. The necessity of studying the relations between the volatilities and co-volatilities of several markets contributed to the development of multivariate versions of these models. After an initial enthusiasm in the beginning of the 1990s and a period of relative tranquillity in the second half of the 1990s, Multivariate GARCH models "... are now experiencing again a quick expansion phase (Bauens et al. (2004))". An overview of time varying variance models can be found in Section 2.2.

Almost every science field has to deal to a certain extent with the discrepancy between the variables implied by the theoretical models and the actual quantities available or measured. Latent structures of time series have been found and analyzed in several fields,

including signal processing (Andrieu & Doucet (2002)), biology (Leroux & Puterman (1992)), genetics (Felsenstein & Churchill (1996)), psychology (Ho et al. (2003)) and sociology (Bollen (1989), Bollen & Long (1993)). Economics (Hamilton (1990), Schneider (1988)) and finance (Aguilar & West (2000)) are not exceptions. Economic growth is an example of an hidden process indirectly measured by the GNP expansion (see for example Hamilton (1990)), and it is not rare in economic literature to read contributions pointing out difficulties in using statistics for estimating certain theoretical quantities (see Quah (1993) among others). Related models have several names, according to the restrictions imposed to them: dynamic models, hidden Markov models, state space models, etc. ARMA, dynamical, and stochastic volatility models can all be represented in the form of a state space model, which allows for handling a wide range of time series models. An overview of state space models can be found in Chapter 4.

All of these models have been analyzed from both a frequentist and a Bayesian perspective. Effective model selection techniques can be successfully used here. They would allow one to start with a very general and unconstrained model and to impose later restrictions in a data-driven perspective, thereby stimulating an interaction between theoretical and quantitative research. The problems inherent in model selection with such frameworks though, have not been studied as much as problems related to estimation. One reason can be the computational burden imposed by the comparison of a very large number of possible candidates. This necessitates an efficient way to compare models and detect possible constraints.

The main object of this study is to propose and evaluate Bayesian algorithms for estimation and model selection of unrestricted multivariate time series models with time varying volatilities or latent variables. Classical model selection techniques are simply

not feasible in such contexts because of the large number of competing models, which makes them computationally too expensive. The algorithms for nested model comparison presented here are based on the stochastic search algorithm first developed by George & McCulloch (1993), which embeds the variable selection procedure inside the MCMC estimation algorithm. The aim is to produce procedures which are computationally efficient and that can be used by the data analyst.

This study is divided in three parts: in the first part (Chapter 2), we analyze a multivariate regression model with endogenous and exogenous variables, and with time varying volatilities. VAR, GARCH and stochastic volatility can be reconducted as special cases of the model presented, so a brief overview of them is provided. We propose different MCMC algorithms for Bayesian estimation and selection of nested models. The parametrization of the time varying covariance matrix is obtained through modified Cholesky decomposition (Pourahmadi (1999)), while different methods for the estimation of the latent volatilities are used and compared through simulations with artificially generated data. Simulation studies are also used to show how variable selection can improve the forecasting ability of the model itself. Finally, two examples of macroeconomic applications are given.

Chapter 3 deals with the selection of non-nested models, for which the technique described in Chapter 2 is not applicable. We utilize a different modified Cholesky decomposition to parameterize the covariance matrix of the model in Chapter 2. This alternative covariance decomposition has a different statistical interpretation and implies different assumptions on the structure of the time varying covariances. We apply the model selection technique developed in Chapter 2 to this model also, and then we compare the best models obtained with the two decompositions by estimating the Bayes

factor for the two competing models.

In Chapter 4, we focus on estimation and model selection for linear state space models. After providing an overview about the model and the common Bayesian estimation techniques, we introduce a model selection algorithm. We then compare the results of the common conjugate analysis methods with our model selection algorithm using artificially generated data.

The last chapter is dedicated to conclusions and future research.

Chapter 2

Multivariate Stochastic Volatility Models

2.1 Overview: VAR Models

Consider a p-dimensional column variable of interest, $\boldsymbol{y}_t$, determined by

$$\boldsymbol{y}_t = \boldsymbol{b}_0 + \boldsymbol{B}_{10}\boldsymbol{w}_t + \sum_{j=1}^{L} \boldsymbol{B}_{1j}\boldsymbol{y}_{t-j} + \boldsymbol{e}_t, \tag{2.1}$$

for $t = 1, \cdots, T$, where $\boldsymbol{y}_t$ is a $p \times 1$-vector variable, $\boldsymbol{b}_0$ is a $p \times 1$-vector of unknown parameters, $\boldsymbol{w}_t$ is a $q \times 1$-vector of known variables, $\boldsymbol{B}_{10} \equiv (\boldsymbol{b}_{10,1}, \cdots, \boldsymbol{b}_{10,q})$ is a $p \times q$-matrix of unknown parameters and, for $j = 1, \cdots, L$, $\boldsymbol{B}_{1j} \equiv (\boldsymbol{b}_{1j,1}, \cdots, \boldsymbol{b}_{1j,p})$ is a $p \times p$-matrix of unknown parameters, $\boldsymbol{\epsilon}_1, \cdots, \boldsymbol{\epsilon}_T$ are independent identically distributed $N_p(\boldsymbol{0}, \boldsymbol{\Sigma})$ errors, and $\boldsymbol{\Sigma}$ is an unknown $p \times p$ positive definite matrix. The above model, known as a vector autoregressive (VAR) model (often in absence of the exogenous variables $\boldsymbol{w}_t$), became widely used for time series analysis after the seminal work of Sims (1972, 1980).

Since then, scholars working on the development of VAR models have been interested

in how to put restrictions on the covariance matrix of the errors.

Several authors (e.g., Sims (1986), Blanchard (1989), Gordon & Leeper (1994), and Sims & Zha (1998)) proposed identification schemes that incorporate restrictions on the covariance matrix $\boldsymbol{\Sigma}$ or equivalently on the contemporaneous relationship among variables without restricting the lag coefficients. This model, called identified VAR, can be written as

$$\boldsymbol{A}_0\boldsymbol{y}_t = \boldsymbol{A}_0\boldsymbol{b}_0 + \boldsymbol{A}_0\boldsymbol{B}_{10}\boldsymbol{w}_t + \sum_{j=1}^{L} \boldsymbol{A}_0\boldsymbol{B}_{1j}\boldsymbol{y}_{t-j} + \boldsymbol{A}_0\boldsymbol{e}_t. \tag{2.2}$$

The major difference between Identified VAR models and the Simultaneous Equation (SE) models that had been used in practice by macroeconomists for several decades is that,

> instead of setting the covariance matrix to identity, SE models impose restrictions on the matrix $\boldsymbol{A}_0$ and regression coefficients in model (2.2) to eliminate a large number of parameters. George et al. (2005).

The advantages of the identified VAR approach over the traditional simultaneous equations approach are illustrated in Leeper et al. (1996).

Ordinary least Squares (OLS) or Maximum Likelihood (ML) methods are often used in applications to find estimates of $\boldsymbol{\Sigma}$ and of $\boldsymbol{B} = (\boldsymbol{b}_0, \boldsymbol{B}_{10}, \boldsymbol{B}_{11}, \cdots, \boldsymbol{B}_{1L})$ They present several drawbacks, as illustrated by Ni & Sun (2005):

- frequentist finite sample distributions of OLS or ML estimators for $\boldsymbol{\Sigma}$ and of $\boldsymbol{B}$ are unavailable;
- on the other hand, a typical VAR involves a large number of parameters, and the sample size of data is often not large enough to justify the use of asymptotic theory;

- asymptotic theory for nonlinear functions of the VAR coefficients (such as impulse response functions) involves approximation of nonlinear functions.

The Bayesian approach represents a valid alternative: Gibbs sampling (Gelfand & Smith (1990)) provides a feasibly implemented technique to find finite sample posterior estimates of VAR parameters. When a priori information is not available, or difficult to be parameterized, non-informative priors can be successfully used, as shown by Sun & Ni (2004), Ni & Sun (2003, 2005) and Ni et al. (2006). Also, Bayesian estimation of VAR models is available in some of the most popular econometric software, such as RATS.

2.2 Overview: Time varying variance models

Several formulations are available to model heteroscedastic time series. In the next two subsections, we are going to briefly present the most popular: ARCH, GARCH and stochastic volatility models.

2.2.1 ARCH and GARCH Models

Conditional on $\mathcal{F}_t$ (the information available at time t), a univariate GARCH(g,q) can be represented as follows:

$$e_t \,|\, \mathcal{F}_{\mathbf{t}} \sim \mathbf{N}(0, \sigma_t^2), \tag{2.3}$$

where the conditional variance is

$$\sigma_t^2 = \alpha_0 + \beta_1 \sigma_{t-1}^2 + \ldots, + \beta_g \sigma_{t-g}^2 + \alpha_1 e_{t-1}^2 + \ldots + \alpha_q e_{t-q}^2. \tag{2.4}$$

This is usually written in a more compact way using the lag operator $\boldsymbol{L}$

$$\sigma_t^2 = \lambda' \varphi_t \;=\; \alpha_0 + B(\boldsymbol{L})\sigma_t^2 + A(\boldsymbol{L})e_t^2,$$

where $A(\boldsymbol{L})\sigma_t^2 = \sum_{i=1}^{g} \sigma_{t-i}^2$ and $B(\boldsymbol{L})e_t^2 = \sum_{j=1}^{p} e_{t-j}^2$. Notice that the conditional variance is defined by an ARMA(g,q) process in the innovations e_t^2. The stationarity condition is that the root of the characteristic equation, $1 - B(\varphi) = 0$ must lie outside the unit circle. For this reason, the most common assumption is that $A(1) + B(1) < 1$. The implication is that the GARCH model is unconditionally covariance stationary. An ARCH(q) model is a model like (2.4) with $\beta_1 = \beta_2 = \cdots = \beta_g = 0$. The main reason why the GARCH formulation is preferred to ARCH models is that often a GARCH(1,1) is able to model the data as well as an ARCH(q), with $q >> 1$. Several modification to the basic model have been made, giving birth to formulations known as EGARCH, IGARCH and DGARCH, among others. An exaustive review of the univariate ARCH and GARCH literature can be found in Bollerslev et al. (1992).

Extensions to the multivariate case date back to the late 1980s and early 1990s: Bollerslev (1990) proposed a model in which the covariance matrix is composed of non constant variances and constant correlations. The advantage of this model is the ease of estimation, since it can be reconducted to a series of univariate GARCH models. Financial time series, though, often show time varying correlation as well, so recently some more general formulations have been presented: Engle (2002), for example, proposed a model for dynamic covariances and a quasi-likelihood (QML) based estimation procedure. A quite exaustive review about multivariate GARCH models can be found in Bauens et al. (2004).

One of the keys for the success of these models lies in the ease of evaluating the ARCH likelihood function. Maximum likelihood techniques have played the dominant role in GARCH models estimation so far, but starting from the nineties several authors provided Bayesian estimators for GARCH models. We could repeat here the same reason

listed in the VAR section to justify an eventual Bayesian choice. Also, we can point that

> the asymptotic properties of ML and QML estimators in multivariate GARCH models are not yet firmly established, and are difficult to derive from low level assumption (Bauens et al. (2004)).

For a review of Bayesian GARCH models see, among others, Muller & Pole (1998) and Vrontos et al. (2000). See Vrontos et al. (2003) for an application of both classical and Bayesian techniques to a full-factor multivariate GARCH model.

2.2.2 Stochastic Volatility Models

Contributions on stochastic volatilities date back to Clark (1973), who proposed an iid mixture model for the distribution of stock price changes, and can be found in both mathematical finance and econometric literature. According to Ghysels et al. (1996),

> The class of stochastic volatility (SV) models has its roots in both mathematical finance and financial econometrics. In fact, several variations of SV models originated from research looking at very different issues.

In option pricing literature, the fundamental Black & Scholes (1973) formula assumes constant volatility, but empirical studies show this assumption to be unrealistic. This motivated Hull & White (1987) to introduce an option pricing model with stochastic volatility, contributing to the development of the concept of implied volatility, which was first introduced by Latane & Rendleman (1976). We are more interested in the discrete time stochastic volatility models typical of the econometric literature: Tauchen & Pitts (1983) and Gallant et al. (1991) noted that if the information flows are autocorrelated, then a stochastic volatility model with time-varying and autocorrelated conditional variance might be appropriate for price-change series. Jacquier et al. (1994) pointed out the advantages of the stochastic volatility models compared to the ARCH/GARCH approaches, in terms of parsimony and performances with high kurtosis time series. In the

GARCH formulation, in fact the degree of kurtosis is tied to the roots of the variance equations, while it is independent from the latter in a SV model (see Ghysels et al. (1996) for details). Usually, if a series with high kurtosis has to be analyzed using a GARCH model, the problem is accomodated by using non normal innovation densities (commonly a t-Student, like in Nelson (1991)). For a review of early SV models literature see Taylor (1994) and Shephard (1996).

We can describe a stochastic volatility model as follows:

$$e_t \,|\, \mathcal{F}_{\mathbf{t}} \sim \mathbf{N}(0, \lambda_t),$$

where

$$\log \lambda_t \;=\; \alpha + \beta \log \lambda_{t-1} + \sigma \upsilon_t, \qquad \upsilon_t \;\sim\; \mathbf{N}(0,1). \tag{2.5}$$

Here the logarithm of the conditional volatility follows an autoregressive times series model; unlike in the ARCH and GARCH cases, both mean and log-volatility equations have separate error terms. Despite the above mentioned advantages, the popularity of SV models, although increasing, is still small if compared to GARCH. The nature of latent variable of the volatilities and the consequent difficulty of estimation play a major role in this situation. Chib et al. (1998) provide a likelihood based framework for the analysis of SV models. In the nineties different Bayesian techniques have been proposed: (Jacquier et al. (1994), Uhlig (1997) Jacquier et al. (2002), Aguilar & West (2000), Chib et al. (2002) and Chib et al. (2002, 2005) among others).

Comparison between the GARCH and SV formulations have been protracted using Monte Carlo simulations (Chib et al. (1998)) and empirical studies: Boscher et al. (2000) and Hol & Koopman (2002) show, for example, that SV models have better forecasting performances when compared with GARCH.

2.3 The Model

We consider the following p-vector multivariate model:

$$\boldsymbol{y}_t = \boldsymbol{b}_0 + \boldsymbol{B}_1 \boldsymbol{x}_t + \boldsymbol{e}_t, \tag{2.6}$$

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

$$\boldsymbol{\Sigma}_t = \boldsymbol{\Gamma} \boldsymbol{\Lambda}_t \boldsymbol{\Gamma}', \tag{2.7}$$

where $\boldsymbol{\Gamma}$ is a lower triangular matrix with unit diagonal elements and $\boldsymbol{\Lambda}_t = diag(\lambda_{1t}, \cdots, \lambda_{pt})$. Let $h_{jt} = \log \lambda_{jt}$ and $\boldsymbol{h}_t = (h_{1t}, \cdots, h_{pt})'$. We model $\boldsymbol{h}_t$ as follows:

$$\boldsymbol{h}_t \;\; = \;\; \boldsymbol{a}_0 + diag(\boldsymbol{h}_{t-1})\boldsymbol{\beta} + \boldsymbol{A}_1 \boldsymbol{z}_t + diag(\boldsymbol{\delta})\boldsymbol{v}_t, \tag{2.8}$$

where , $\boldsymbol{v}_t$ are iid $N_p(\boldsymbol{0}, \boldsymbol{I}_p)$, $\boldsymbol{z}_t = (z_{t,1}, \cdots, z_{t,r})'$ is a r vector observable exogenous variable uncorrelated with $\boldsymbol{v}_t$, $\boldsymbol{a}_0$ is a p unknown vector, $\boldsymbol{A}_1 = \{a_{jk}\},\ j = 1, \cdots, p,\ k = 1, \cdots, r$ is a matrix of unknown parameters, $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)'$ is a p-vector of unknown parameters, and $\boldsymbol{\delta} = (\delta_1, \cdots, \delta_p)'$ is a vector of unknown nonnegative parameters. If $\delta_i = 0$ then the ith variable is driven by a non-stochastic (but possibly time-varying) shock. It should be noticed that if $\boldsymbol{z}_t$ denotes the contemporaneous implied volatility measure, then be the formulation in (2.8) can be reconduct to SV model with embedded implied volatility (Hol & Koopman (2000)), while if $\delta_i = 0$ and $\boldsymbol{z}_t = \boldsymbol{e}_{t-1}^2$, then (2.8) is equivalent to a GARCH(1,1). Finally, if $\boldsymbol{x}_t = (\boldsymbol{w}_t, \boldsymbol{y}_{t-1}, \cdots, \boldsymbol{y}_{t-L})$ and $\boldsymbol{\Lambda}_1 = \boldsymbol{\Lambda}_2 = \cdots = \boldsymbol{\Lambda}_T$, then (2.8) is equivalent to the VAR(L) in (2.1).

2.3.1 Interpretation of the Cholesky Decomposition

Following and extending Pourahmadi (1999) and Pourahmadi & Dellaportas (2002), we consider the following AR structure of the cross section covariances of the covariance matrix:

$$\omega_{tj} \;=\; \sum_{k=1}^{j-1} \xi_{jk}\omega_{tk} + e_{tj}, \tag{2.9}$$

where $j = 1, \cdots, p$ and $t = 1, \cdots, T$, $var(e_{tj}) = \lambda_{jt}$. Writing (2.9) in matrix form, we obtain

$$\boldsymbol{\Xi}\boldsymbol{\omega}_t = \boldsymbol{e}_t, \tag{2.10}$$

where $\boldsymbol{\omega}_t = (\omega_{t1}, \cdots, \omega_{tj})'$, $\boldsymbol{e}_t = (e_{t1}, \cdots, e_{tj})'$ and

$$\boldsymbol{\Xi} \;=\; \begin{pmatrix} 1 & 0 & \cdots & 0 \\ -\xi_{21} & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ -\xi_{p1} & -\xi_{p2} & \cdots & 1 \end{pmatrix}. \tag{2.11}$$

From (2.10) it follows that the lower triangular matrix $\boldsymbol{\Xi}$ diagonalizes $\boldsymbol{\Sigma}_t$:

$$\boldsymbol{\Xi}\boldsymbol{\Sigma}_t\boldsymbol{\Xi}' = \boldsymbol{\Lambda}_t = diag(\lambda_{1t}, \cdots, \lambda_{pt}). \tag{2.12}$$

We can find notable differences with the stationary AR case:

(i). The difference equations in (2.9) do not involve initial values;

(ii). The innovation variances λ_{jt} are distinct;

(iii). The matrix $\boldsymbol{\Xi}$ is not banded or sparse, and its non-redundant entries are unconstrained.

Now consider the MA structure of the cross section covariances of the covariance matrix:

$$\omega_{tj} \;=\; e_{tj} + \sum_{i=1}^{j-1} \nu_{jk} e_{tk}, \tag{2.13}$$

where $j = 1, \cdots, p$ and $t = 1, \cdots, T$, and $var(e_{jt}) = \lambda_{jt}$. Explicit expressions for ν_{jk} can be obtained using the innovation algorithm (Brockwell & Davis (1991)). Writing (2.13) in matrix form we obtain

$$\boldsymbol{\omega}_t = \boldsymbol{\Gamma} \boldsymbol{e}_t, \tag{2.14}$$

where

$$\begin{pmatrix} 1 & 0 & \cdots & 0 \\ \nu_{21} & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ \nu_{p1} & \nu_{p2} & \cdots & 1 \end{pmatrix}, \tag{2.15}$$

from which (2.7) follows. From (2.12) and (2.7) it follows that

$$\boldsymbol{\Xi} \;=\; \boldsymbol{\Gamma}^{-1}. \tag{2.16}$$

Thus, once the AR parameters are given, one can compute the MA parameters and *vice versa*. It also follows that

$$\boldsymbol{\Sigma}_t^{-1} = \boldsymbol{\Xi}' \boldsymbol{\Lambda}_t^{-1} \boldsymbol{\Xi} = \boldsymbol{\Psi} \boldsymbol{\Lambda}_t^{-1} \boldsymbol{\Psi}', \tag{2.17}$$

where $\boldsymbol{\Psi} = \boldsymbol{\Xi}'$ is an upper triagular matrix whose diagonal elements are 1. Pourahmadi & Daniels (2002) used a similar decomposition to analyze dynamic models for longitudinal data. In this study, we decompose the time varying covariances $\boldsymbol{\Sigma}_t$ and assign to each $\boldsymbol{\lambda}_t$ the structure described in (2.8). Using (2.7), we then model the conditional variance of

each variable given the previous, and the marginal variance of variable one. Paying the price of imposing an order to the variables we can model each single volatility process as in (2.8) without imposing independence on the volatilities.

2.3.2 Stochastic Search Model Selection

For a given data set and choosing among a given class of models, the researcher must contemplate the trade off between the number of parameters in the model and the goodness of fit. If one considers a model space that includes all potential models of interest, then the model choice is equivalent to obtaining restrictions on the model space for reduction in the number of parameters (i.e., obtaining a submodel). Appropriate restrictions (submodels) often give rise to sharper finite sample inference and better forecasts, since unrestricted models tend to overfit the data. In this paper, we consider a multivariate linear regression framework where variables are potentially contemporaneously, as well as, serially correlated. In addition, some variables may exhibit time-varying, even stochastic, conditional volatilities. The unrestricted model, in general, consists of over-parameterized regression and volatility equations.

Even with constant volatilities, there is an intractable number of possible specifications of the regression model for exhaustive model comparison. The variable selection problem in a univariate model with m explanatory variables involves comparing 2^m number of competing models. The amount of computation is prohibitive for a moderate size m. George & McCulloch (1993) propose a Bayesian MCMC stochastic search algorithm in a univariate regression framework that greatly reduced the amount of computation, and George et al. (2005) extended the algorithm to VAR models with stationary errors.

We conduct stochastic model selection for time varying variance models. We develop

an MCMC algorithm for model selection on coefficients of both equations $(\boldsymbol{B}_1, \boldsymbol{A}_1)$ as well as on parameters in $\boldsymbol{\Sigma}_t$ $(\boldsymbol{\Psi}, \boldsymbol{\delta})$, where $\boldsymbol{\Psi} = \boldsymbol{\Gamma}^{-1'}$. There are 2^{pq} competing models for $\boldsymbol{B}_1$, 2^{pr} competing models for $\boldsymbol{A}_1$, $2^{\frac{p(p-1)}{2}}$ competing models for $\boldsymbol{\Psi}$ and 2^p competing models for $\boldsymbol{\delta}$. The total number of models is $2^{\frac{p(p-1)}{2}+p(r+q+1)}$. Following George & McCulloch (1993), we index a model by the tightness of a prior vector on each parameter. 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.

The stochastic search approach is fundamentally different from other approaches in the econometrics literature that rely on economic theory to justify parameter restrictions.

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 particular, the choice between a deterministic volatility process (such as in GARCH models) or a stochastic one (such as in SV models) is left to the researcher's expertise. Our model embeds both possibilities and makes possible the specification of hybrid models. Using stochastic search on the variance parameter of the volatility equation $(\boldsymbol{\delta})$, we propose a data driven selection of the time-varying volatilities' nature. This gives rise to 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 $\boldsymbol{\delta}$. We find the posterior of the model index on $\boldsymbol{\delta}$ to be quite sensitive to the hyperparameters of the mixed inverse gamma prior. We consider a hierarchal structure for the prior of $\boldsymbol{\delta}$, with a diffuse gamma prior on the

hyper-parameter of the mixed inverse gamma prior on $\boldsymbol{\delta}$.

2.4 Hierarchical Models

2.4.1 Likelihood Functions

We give several formulas for the likelihood functions. Define $etr(\boldsymbol{A}) = exp\{tr(\boldsymbol{A})\}$. First we know that the likelihood function of $(\boldsymbol{b}_0, \boldsymbol{B}_1, \boldsymbol{\Sigma}_1, \cdots, \boldsymbol{\Sigma}_T)$ based on $\boldsymbol{Y} = (\boldsymbol{y}_1, \cdots, \boldsymbol{y}_T)$, is

$$\begin{aligned} & [\boldsymbol{Y} \mid \boldsymbol{b}_0, \boldsymbol{B}_1, \boldsymbol{\Sigma}_1, \cdots, \boldsymbol{\Sigma}_T] \\ \propto & \prod_{t=1}^{T} |\boldsymbol{\Sigma}_t|^{-\frac{1}{2}} etr\Big\{-\frac{1}{2}\sum_{t=1}^{T}(\boldsymbol{y}_t - \boldsymbol{b}_0 - \boldsymbol{B}_1\boldsymbol{x}_t)'\boldsymbol{\Sigma}_t^{-1}(\boldsymbol{y}_t - \boldsymbol{b}_0 - \boldsymbol{B}_1\boldsymbol{x}_t)\Big\}. \end{aligned} \tag{2.18}$$

Using (2.7), we write

$$\boldsymbol{\Sigma}_t^{-1} = \boldsymbol{\Psi}\boldsymbol{\Lambda}_t^{-1}\boldsymbol{\Psi}', \tag{2.19}$$

where $\boldsymbol{\Psi} = \boldsymbol{\Gamma}^{-1'}$ is an upper unit triangular matrix. We then have the likelihood function of $(\boldsymbol{b}_0, \boldsymbol{B}_1, \boldsymbol{\Psi}, \boldsymbol{\Lambda}_1, \cdots, \boldsymbol{\Lambda}_T)$,

$$[\boldsymbol{Y} \mid \boldsymbol{b}_0, \boldsymbol{B}_1, \boldsymbol{\Psi}, \boldsymbol{\Lambda}_1, \cdots, \boldsymbol{\Lambda}_T] \propto |\boldsymbol{\Psi}|^T \prod_{t=1}^{T} |\boldsymbol{\Lambda}_t|^{-\frac{1}{2}} etr\Big\{-\frac{1}{2}\sum_{t=1}^{T}\boldsymbol{S}_t(\boldsymbol{B})\boldsymbol{\Psi}\boldsymbol{\Lambda}_t^{-1}\boldsymbol{\Psi}'\Big\}, \tag{2.20}$$

where, $\boldsymbol{B} = (\boldsymbol{b}_0, \boldsymbol{B}_1)$ and

$$\boldsymbol{S}_t(\boldsymbol{B}) = (\boldsymbol{y}_t - \boldsymbol{b}_0 - \boldsymbol{B}_1\boldsymbol{x}_t)(\boldsymbol{y}_t - \boldsymbol{b}_0 - \boldsymbol{B}_1\boldsymbol{x}_t)'. \tag{2.21}$$

This formula is useful in updating $(\boldsymbol{\Psi}, \boldsymbol{\Lambda}_1, \cdots, \boldsymbol{\Lambda}_T)$.

To see the formulas for $(\boldsymbol{b}_0, \boldsymbol{B}_1)$, we could rewrite (2.6) in the familiar matrix form

$$\boldsymbol{Y} \;=\; \boldsymbol{B}\boldsymbol{X} + \boldsymbol{E}, \tag{2.22}$$

where

$$\boldsymbol{Y} = (\boldsymbol{y}_1, \cdots, \boldsymbol{y}_T), \qquad \boldsymbol{B} = (\boldsymbol{b}_0, \boldsymbol{B}_1) = (\boldsymbol{b}_0, \boldsymbol{b}_1, \cdots, \boldsymbol{b}_q), \tag{2.23}$$

$$\boldsymbol{X} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ \boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_T \end{pmatrix}, \quad \boldsymbol{E} = (\boldsymbol{e}_1, \cdots, \boldsymbol{e}_T). \tag{2.24}$$

Here $\boldsymbol{Y}$, $\boldsymbol{E}$ are $p \times T$ matrices, $\boldsymbol{X}$ is a $(q+1) \times T$ matrix, and $\boldsymbol{B}$ is a $p \times (q+1)$ matrix.

It follows from formula (2.10) of Harville (1997) that $vec(\boldsymbol{BX}) = (\boldsymbol{X}' \otimes \boldsymbol{I}_p) vec(\boldsymbol{B})$. We define

$$\boldsymbol{y} = vec(\boldsymbol{Y}) = \begin{pmatrix} \boldsymbol{y}_1 \\ \vdots \\ \boldsymbol{y}_T \end{pmatrix}, \ \boldsymbol{e} = vec(\boldsymbol{E}) = \begin{pmatrix} \boldsymbol{e}_1 \\ \vdots \\ \boldsymbol{e}_T \end{pmatrix}, \ \boldsymbol{b} = vec(\boldsymbol{B}) = \begin{pmatrix} \boldsymbol{b}_0 \\ \boldsymbol{b}_1 \\ \vdots \\ \boldsymbol{b}_q \end{pmatrix}.$$

Then

$$(\boldsymbol{y} \mid \boldsymbol{b}, \boldsymbol{\Psi}, \boldsymbol{\Lambda}) \sim N_{pT}((\boldsymbol{X}' \otimes \boldsymbol{I}_p)\boldsymbol{b}, \ \boldsymbol{\Sigma}), \tag{2.25}$$

where

$$\boldsymbol{\Sigma} = diag(\boldsymbol{\Sigma}_1, \cdots, \boldsymbol{\Sigma}_T) = (\boldsymbol{I}_T \otimes \boldsymbol{\Gamma})\boldsymbol{\Lambda}(\boldsymbol{I}_T \otimes \boldsymbol{\Gamma}'), \tag{2.26}$$

$$\boldsymbol{\Lambda} = diag(\boldsymbol{\Lambda}_1, \cdots, \boldsymbol{\Lambda}_T) = diag(\lambda_{11}, \lambda_{21}, \cdots, \lambda_{p1}, \lambda_{12}, \cdots, \lambda_{pT}). \tag{2.27}$$

Clearly, (2.26) implies that $\boldsymbol{\Sigma}^{-1} = (\boldsymbol{I}_T \otimes \boldsymbol{\Psi})\boldsymbol{\Lambda}^{-1}(\boldsymbol{I}_T \otimes \boldsymbol{\Psi}')$. The likelihood of $(\boldsymbol{b}, \boldsymbol{\Psi}, \boldsymbol{\Lambda})$ is then

$$\begin{aligned} & L(\boldsymbol{b}, \boldsymbol{\Psi}, \boldsymbol{\Lambda}) = [\boldsymbol{y} \mid \boldsymbol{b}, \boldsymbol{\Psi}, \boldsymbol{\Lambda}] \\ \propto \ & |\boldsymbol{\Psi}|^T |\boldsymbol{\Lambda}|^{-\frac{1}{2}} etr\Big\{-\frac{1}{2}[\boldsymbol{y} - (\boldsymbol{X}' \otimes \boldsymbol{I}_p)\boldsymbol{b}]'(\boldsymbol{I}_T \otimes \boldsymbol{\Psi})\boldsymbol{\Lambda}^{-1}(\boldsymbol{I}_T \otimes \boldsymbol{\Psi}')[\boldsymbol{y} - (\boldsymbol{X}' \otimes \boldsymbol{I}_p)\boldsymbol{b}]\Big\} \\ = \ & |\boldsymbol{\Psi}|^T |\boldsymbol{\Lambda}|^{-\frac{1}{2}} etr\Big\{-\frac{1}{2}[(\boldsymbol{I}_T \otimes \boldsymbol{\Psi}')\boldsymbol{y} - (\boldsymbol{X}' \otimes \boldsymbol{\Psi}')\boldsymbol{b}]'\boldsymbol{\Lambda}^{-1}[(\boldsymbol{I}_T \otimes \boldsymbol{\Psi}')\boldsymbol{y} - (\boldsymbol{X}' \otimes \boldsymbol{\Psi}')\boldsymbol{b}]\Big\}. \end{aligned} \tag{2.28}$$

2.4.2 Priors

We employ independent priors on the elements of $(\boldsymbol{B}, \boldsymbol{\Psi}, \boldsymbol{a}_0, \boldsymbol{\beta}, \boldsymbol{A}_1, \boldsymbol{\delta})$ and independent priors on the indexes corresponding to the elements to be selected in $(\boldsymbol{B}, \boldsymbol{\Psi}, \boldsymbol{A}_1)$, $(\boldsymbol{\gamma}_b, \boldsymbol{\gamma}_\psi, \boldsymbol{\gamma}_a, \boldsymbol{\gamma}_\delta)$.

(i) Priors of $\boldsymbol{B}$.

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

$$b_{i0} \overset{ind}{\sim} N(b_{i0}^0, \xi_{i0}^0). \tag{2.29}$$

(ib) Priors of $\boldsymbol{B}_1$=$\{b_{ij}\}_{p\times q}$. We associate each element b_{ij} with an indicator variable $\gamma_{b,ij}$, and we assume that elements of $\boldsymbol{B}_1$ may be included in the model ($\gamma_{b,ij} = 1$) or may not ($\gamma_{b,ij} = 0$). By design, b_{ij} has a two stage prior. We assume independent Bernoulli priors for the index $\gamma_{b,ij}$: for fixed $p_{b,ij} \in (0, 1)$,

$$P(\gamma_{b,ij} = 1) = 1 - P(\gamma_{b,ij} = 0) = p_{b,ij},\ i = 1, \cdots, p,\ j = 1, \cdots, q. \tag{2.30}$$

For given $\boldsymbol{\gamma}_b = (\gamma_{b,11}, \gamma_{b,12}, \cdots, \gamma_{b,pq})'$, assume that

$$(b_{ij} \mid \gamma_{b,ij}) \overset{ind}{\sim} (1 - \gamma_{b,ij})N(0, \kappa_{b,ij}^2) + \gamma_{b,ij}N(0, c_{b,ij}^2\kappa_{b,ij}^2), \tag{2.31}$$

for $i = 1, \cdots, p$ and $j = 1, \cdots, q$, 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 $\boldsymbol{D}_{b,j} = diag((\eta_{b,1j}\kappa_{b,1j})^2, \cdots, (\eta_{b,pj}\kappa_{b,pj})^2)$, then (2.31) is equivalent to

$$(\boldsymbol{b}_j \mid \boldsymbol{\gamma}_{b,j}) \overset{ind}{\sim} N_p(\boldsymbol{0},\ \boldsymbol{D}_{b,j}), \quad \text{for } j = 1, \cdots, q. \tag{2.32}$$

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

$$(\boldsymbol{b} \mid \boldsymbol{\gamma}_b) \quad \sim \quad N(\bar{\boldsymbol{b}}, \bar{\boldsymbol{\Xi}}), \tag{2.33}$$

where

$$\begin{aligned} \bar{\boldsymbol{b}} &= (b_{10}^0, \cdots, b_{p0}^0, 0, \cdots, 0)', \\ \bar{\boldsymbol{\Xi}} &= diag(\xi_{10}^0, \cdots, \xi_{p0}^0, (\eta_{b,11}\kappa_{b,1})^2, \cdots, (\eta_{b,pq}\kappa_{b,pq})^2). \end{aligned}$$

(ii) Priors of $\boldsymbol{\Psi}$. We propose an approach of imposing priors on components of $\boldsymbol{\Psi}$ that utilizes the structure of the matrix to ease Bayesian computation. For $j = 2, \cdots, p$, let $\boldsymbol{\psi}_j$ be a vector containing the non-redundant elements of the j^{th} column of $\boldsymbol{\Psi}$, i.e. $\boldsymbol{\psi}_j = (\psi_{1j}, \cdots, \psi_{j-1,j})'$. Also, define a vector of indicators of length $j-1$, $\boldsymbol{\gamma}_{\psi,j} = (\gamma_{\psi,1j}, \cdots, \gamma_{\psi,j-1,j})'$. We assume that elements of $\boldsymbol{\psi}_j$ may be included in the model ($\gamma_{\psi,ij} = 1$) or may 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 $p_{\psi,ij} \in (0,1)$,

$$P(\gamma_{\psi,ij} = 1) = 1 - P(\gamma_{\psi,ij} = 0) = p_{\psi,ij}, \ i = 1, \cdots, j-1, \ j = 1, \cdots, p. \tag{2.34}$$

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

$$(\psi_{ij} \mid \gamma_{\psi,ij}) \quad \overset{ind}{\sim} \quad (1 - \gamma_{\psi,ij})N(0, \kappa_{\psi,ij}^2) + \gamma_{\psi,ij}N(0, c_{\psi,ij}^2\kappa_{\psi,ij}^2), \tag{2.35}$$

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

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

and $\boldsymbol{D}_{\psi,j} = diag((\eta_{\psi,1j}\kappa_{\psi,1j})^2, \cdots, (\eta_{\psi,j-1,j}\kappa_{\psi,j-1,j})^2)$, then (2.35) is equivalent to

$$(\boldsymbol{\psi}_j \mid \boldsymbol{\gamma}_{\psi,j}) \quad \overset{ind}{\sim} \quad N_{j-1}(\boldsymbol{0}, \ \boldsymbol{D}_{\psi,j}), \tag{2.36}$$

for $j = 2, \cdots, p$.

(iii) Priors of $(\boldsymbol{a}_0, \boldsymbol{\beta}, \boldsymbol{A}_1, \boldsymbol{\delta})$. We assume that $\boldsymbol{a}_0$, $\boldsymbol{\beta}$, $\boldsymbol{A}_1$, and $\boldsymbol{\delta}$ have mutually independent priors.

(iiia) Priors of $\boldsymbol{a}_0 = (a_{10}, \cdots, a_{p0})'$. For fixed $(\bar{a}_{j0}, \sigma_a)$, assume that

$$(a_{j0}) \overset{ind}{\sim} N\Big(\bar{a}_{j0}, \sigma_a\Big). \tag{2.37}$$

(iiib) Priors of $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)'$. For fixed $(\bar{\beta}_j, \sigma_\beta)$, assume that

$$(\beta_j) \overset{ind}{\sim} N\Big(\bar{\beta}_j, \sigma_\beta\Big). \tag{2.38}$$

(iiic) Priors of $\boldsymbol{A}_1$. The elements in $\boldsymbol{A}_1$ are of primary interest for model selection. Let the model index for a_{jk}, $\gamma_{a,jk}$, be independent Bernoulli $(p_{a,jk})$ random variables: for fixed $p_{a,jk} \in (0,1)$,

$$P(\gamma_{a,jk} = 1) \;=\; 1 - P(\gamma_{a,jk} = 0) = p_{a,jk}, \;\; for \;\; j = 1, \cdots, p, \;\; k = 1, \cdots, r. \tag{2.39}$$

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

$$(a_{jk} \mid \gamma_{a,jk}) \;\overset{ind}{\sim}\; (1 - \gamma_{a,jk}) N(0, \kappa_{a,jk}^2) + \gamma_{a,jk} N(0, c_{a,jk}^2 \kappa_{a,jk}^2). \tag{2.40}$$

where $\kappa_{a,jk}$ would be small and $c_{a,jk}$ would be large constants.

Later, it would be convenient to write $\boldsymbol{A}_1$ in terms of its row vectors: $\boldsymbol{A}_1 = (\widetilde{\boldsymbol{a}}_1', \cdots, \widetilde{\boldsymbol{a}}_p')'$. Here $\widetilde{\boldsymbol{a}}_j = (a_{j1}, \cdots, a_{jr})'$, $j = 1, \cdots, p$. Denote

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

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

$$(\widetilde{\boldsymbol{a}}_j \mid \boldsymbol{\gamma}_{a,j}) \overset{ind}{\sim} N_r(\boldsymbol{0}, \boldsymbol{D}_{a,j}). \tag{2.41}$$

Define $\widetilde{\boldsymbol{a}}_j^* = (a_{j0}, \beta_j, \widetilde{\boldsymbol{a}}_j')'$: combining (2.37) and (2.41) we can write

$$(\widetilde{\boldsymbol{a}}_j^* \mid \boldsymbol{\gamma}_{a,j}) \overset{ind}{\sim} N_{r+2}(\bar{\boldsymbol{a}}_j, \boldsymbol{\Omega}_j), \tag{2.42}$$

where

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

2.4.3 Prior for $\boldsymbol{\delta}$

The elements in $\boldsymbol{\delta} = (\delta_1, \cdots, \delta_p)'$ are objects of model selection. Selecting which δ_j are different than zero makes us learn from the data which volatilities come from a deterministic and which from a stochastic process. So far the choice between a deterministic process and a probabilistic one (such a stochastic volatility model) has been subjective and a priori. Also, in case of multivariate time series, all the generating processes for the volatilities had to be deterministic or stochastic. 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, $\gamma_{\delta j}$, be independent Bernoulli $(p_{\delta j})$ random variables: for fixed $p_{\delta j} \in (0, 1)$,

$$P(\gamma_{\delta j} = 1) = 1 - P(\gamma_{\delta j} = 0) = p_{\delta j},\ j = 1, \cdots, p. \tag{2.43}$$

For given $\gamma_{\delta j}$ $(j = 1, \cdots, p)$, we assume that δ_j is a mixture of independent inverse gamma, $(\delta_j^2 \mid \gamma_{\delta j}, q_j) \sim IG(v_{jo}; q_j s_{j0})$ with probability $\gamma_{\delta j}$ and $(\delta_j^2 \mid \gamma_{\delta j}, q_j) \sim IG(v_{jo}; s_{j0})$ with probability $1 - \gamma_{\delta j}$. In other words, the density of δ_j^2 has the form

$$[\delta_j^2 \mid \gamma_{\delta j}, q_j] \propto (\eta_{\delta j} s_{j0})^{v_{jo}} (\delta_j^2)^{-(v_{jo}+1)} \exp(-\frac{\eta_{\delta j} s_{j0}}{\delta_j^2}), \tag{2.44}$$

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_{jo} is a given positive constants and is larger than 2, the shape parameter s_{j0} is a small positive constant so that the mean and the variance of the prior with $\gamma_{\delta j} = 1$, $q_j s_{j0}/(v_{jo} - 1)$ and $(q_j s_{j0})^2/\{(v_{jo}-1)^2(v_{jo}-2)\}$, are large, while prior mean and variance corresponding to $\gamma_{\delta j} = 0$, $s_{j0}/(v_{jo} - 1)$ and $s_{j0}^2/\{(v_{jo} - 1)^2(v_{jo} - 2)\}$, 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 coefficients with normal distribution, like in George & McCulloch (1993) and George et al. (2005), but $\boldsymbol{\delta}$ follows an inverse gamma distribution, so even small changes in the values of its hyperparameters can have a more serious impact than in the previous cases (see Kim et al. (2002) for a discussion on normal gamma relations). Also, simulation studies show that no arbitrary value $\boldsymbol{q} = (q_1, \cdots, q_p)'$ seems to be universally effective. We then impose a diffuse prior on $\boldsymbol{q}$:

$$q_j \sim Ga(\alpha_q,\ \beta_q),\ \ for\ j = 1, \cdots, p. \tag{2.45}$$

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

2.5 Posterior Computation

The joint posterior of $(\boldsymbol{B}, \boldsymbol{\gamma}_b, \boldsymbol{\Psi}, \boldsymbol{\gamma}_\psi, \boldsymbol{a}_0, \boldsymbol{\beta}, \boldsymbol{A}_1, \boldsymbol{\gamma}_a, \boldsymbol{\delta}, \boldsymbol{\gamma}_\delta, \boldsymbol{q}, \boldsymbol{\Lambda})$ has the form,

$$\begin{aligned}[\boldsymbol{B}, \boldsymbol{\gamma}_b, \boldsymbol{\Psi}, \boldsymbol{\gamma}_\psi, \boldsymbol{a}_0, \boldsymbol{\beta}, \boldsymbol{A}_1, \boldsymbol{\delta}, \boldsymbol{\gamma}_a, \boldsymbol{\Lambda} \mid \boldsymbol{y}] \;&\propto\; [\boldsymbol{y} \mid \boldsymbol{B}, \boldsymbol{\Psi}, \boldsymbol{\Lambda}]\, [\boldsymbol{\Lambda} \mid \boldsymbol{a}_0, \boldsymbol{\beta}, \boldsymbol{A}_1, \boldsymbol{\delta}] \\ &\times\; [\boldsymbol{b}_0]\, [\boldsymbol{B}_1 \mid \boldsymbol{\gamma}_b]\, [\boldsymbol{\gamma}_b]\, [\boldsymbol{\Psi} \mid \boldsymbol{\gamma}_\psi]\, [\boldsymbol{\gamma}_\psi]\, [\boldsymbol{a}_0, \boldsymbol{\beta}]\, [\boldsymbol{A}_1 \mid \boldsymbol{\gamma}_a]\, [\boldsymbol{\gamma}_a]\, [\boldsymbol{\delta} | \boldsymbol{\gamma}_\delta, \boldsymbol{q}] [\boldsymbol{\gamma}_\delta] [\boldsymbol{q}],\end{aligned}$$

where $[\boldsymbol{y} \mid \boldsymbol{B}, \boldsymbol{\Psi}, \boldsymbol{\Lambda}]$ is the likelihood function, given by (2.28), $[\boldsymbol{\Lambda} \mid \boldsymbol{a}_0, \boldsymbol{\beta}, \boldsymbol{A}_1, \boldsymbol{\delta}]$ is given by (2.8), $[\boldsymbol{b}_0]$ is given by (2.29), $[\boldsymbol{B}_1 \mid \boldsymbol{\gamma}_b]$ is given by (2.32), $[\boldsymbol{\gamma}_b]$ is given by (2.30),

$[\boldsymbol{\Psi} \mid \boldsymbol{\gamma}_\psi]$ is given by (2.36), $[\boldsymbol{\gamma}_\psi]$ is given by (2.34), $[\boldsymbol{a}_0]$ is given by (2.37) $[\boldsymbol{\beta}]$ is given by (2.38), $[\boldsymbol{A}_1 \mid \boldsymbol{\gamma}_a]$ is given by (2.41), $[\boldsymbol{\gamma}_a]$ is given by (2.39), $[\boldsymbol{\delta} \mid \boldsymbol{\gamma}_\delta, \boldsymbol{q}]$ is given by (2.44), $[\gamma_\delta]$ is given by (2.43) and $[\boldsymbol{q}]$ is given by (2.45). To use an MCMC algorithm, we now derive the full conditional posteriors for $(\boldsymbol{B}, \boldsymbol{\Psi}, \boldsymbol{\Lambda}, \boldsymbol{a}_0, \boldsymbol{\beta}, \boldsymbol{A}_1, \boldsymbol{\delta}, \boldsymbol{\gamma}_b, \boldsymbol{\gamma}_\psi, \boldsymbol{\gamma}_a, \boldsymbol{\gamma}_\delta, \boldsymbol{q})$.

2.5.1 Conditional Posterior for $\boldsymbol{B}$ and $\boldsymbol{\gamma}_b$

Fact 2.1. *(a) The conditional posterior distribution of* $\boldsymbol{b}$ *given* $(\boldsymbol{\gamma}_b, \boldsymbol{\Psi}, \boldsymbol{\gamma}_\psi, \boldsymbol{a}_0, \boldsymbol{\beta}, \boldsymbol{A}_1, \boldsymbol{\delta}, \boldsymbol{\gamma}_a, \boldsymbol{\Lambda}, \boldsymbol{q}; \boldsymbol{y})$ *depends only on* $(\boldsymbol{\Psi}, \boldsymbol{\Lambda}, \boldsymbol{\gamma}_b; \boldsymbol{y})$ *and has the form,*

$$(\boldsymbol{b} \mid \boldsymbol{\Psi}, \boldsymbol{\Lambda}, \boldsymbol{\gamma}_b; \boldsymbol{y}) \sim N_m(\widehat{\boldsymbol{b}}, \widehat{\boldsymbol{\Xi}}), \tag{2.46}$$

where

$$\begin{aligned} \widehat{\boldsymbol{\Xi}} &= \Big\{(\boldsymbol{X} \otimes \boldsymbol{\Psi})\boldsymbol{\Lambda}^{-1}(\boldsymbol{X}' \otimes \boldsymbol{\Psi}') + \bar{\boldsymbol{\Xi}}^{-1}\Big\}^{-1}, & (2.47)\\ \widehat{\boldsymbol{b}} &= \widehat{\boldsymbol{\Xi}}\Big\{(\boldsymbol{X} \otimes \boldsymbol{\Psi})\boldsymbol{\Lambda}^{-1}(\boldsymbol{I}_T \otimes \boldsymbol{\Psi}')\boldsymbol{y} + \bar{\boldsymbol{\Xi}}^{-1}\bar{\boldsymbol{b}}\Big\}. & (2.48) \end{aligned}$$

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

$$(\gamma_{b,ij} \mid \boldsymbol{B}_1) = (\gamma_{b,ij} \mid b_{ij}) \overset{ind}{\sim} \; Bernoulli\Big(\frac{u_{b,ij1}}{u_{b,ij1} + u_{b,ij2}}\Big), \tag{2.49}$$

where

$$\begin{aligned} u_{b,ij1} &= \frac{1}{c_{b,ij}} \exp\Big(-\frac{b_{ij}^2}{2c_{b,ij}^2\kappa_{b,ij}^2}\Big) p_{b,ij},\\ u_{b,ij2} &= \exp\Big(-\frac{b_{ij}^2}{2\kappa_{b,ij}^2}\Big)(1 - p_{b,ij}). & (2.50) \end{aligned}$$

Proof. Using the likelihood (2.28) part (a) is obvious. For part (b), recall that $\boldsymbol{\gamma}_b$ depends on data indirectly, then,

$$u_{b,ij1} \;\propto\; [\boldsymbol{b} \mid \boldsymbol{\gamma}_{b,(-ij)}, \gamma_{b,ij} = 1] p_{b,ij},$$

$$u_{b,ij2} \quad \propto \quad [\boldsymbol{b} \mid \boldsymbol{\gamma}_{b,(-ij)}, \gamma_{b,ij} = 0](1 - p_{b,ij});$$

the expression above, given prior independence of b_{ij}, gives the formula (2.50). □

Remark 2.1. Formula (2.47) involves computing products of matrices of dimensions proportional to sample size T. We use the following formula to reduce the dimension of matrices involved for multiplication:

$$(\boldsymbol{X} \otimes \boldsymbol{\Psi})\boldsymbol{\Lambda}^{-1}(\boldsymbol{X}' \otimes \boldsymbol{\Psi}') \quad = \quad \sum_{t=1}^{T} \begin{pmatrix} 1 \\ \boldsymbol{x}_t \end{pmatrix} (1, \boldsymbol{x}_t') \otimes \boldsymbol{\Sigma}_t^{-1}. \tag{2.51}$$

Similarly, to avoid multiplication of matrices of large dimensions in the formula for $\widehat{\boldsymbol{b}}$, we use the formula

$$(\boldsymbol{X} \otimes \boldsymbol{\Psi})\boldsymbol{\Lambda}^{-1}(\boldsymbol{I}_T \otimes \boldsymbol{\Psi}')\boldsymbol{y} \quad = \quad \sum_{t=1}^{T} \begin{pmatrix} 1 \\ \boldsymbol{x}_t \end{pmatrix} \otimes \boldsymbol{\Sigma}_t^{-1}\boldsymbol{y}_t. \tag{2.52}$$

2.5.2 Conditional Posterior for $\boldsymbol{\Psi}$ and $\boldsymbol{\gamma}_\psi$

To derive conditional distributions of $\boldsymbol{\Psi}$, we use the likelihood function (2.20) of $(\boldsymbol{B}, \boldsymbol{\Psi}, \boldsymbol{\Lambda}_1, \cdots, \boldsymbol{\Lambda}_T)$ and we adopt the algorithm derived by George et al. (2005) for VAR with time invariant covariances and adapt it to be used with a modified Cholesky decomposition: given $\boldsymbol{B} = (\boldsymbol{b}_0, \boldsymbol{B}_1)$, we know from (2.21) that $\boldsymbol{S}_t = \boldsymbol{S}_t(\boldsymbol{B})$ represents the covariance of residuals $\boldsymbol{e}_t$. Let $\boldsymbol{S}_{t,j}$ be the upper-left $j \times j$ submatrix of $\boldsymbol{S}_t(\boldsymbol{B})$. So $\boldsymbol{S}_t = \boldsymbol{S}_{t,p}$. We write the $(i,j)^{th}$ component of $\boldsymbol{S}_t(\boldsymbol{B})$ by $s_{t,ij}$. For $j = 2, \cdots, p$, define $\boldsymbol{s}_{t,j} = (s_{t,1j}, \cdots, s_{t,j-1,j})'$. Define

$$v_{t,1} = s_{t,11}, \; v_{t,j} = \frac{|\boldsymbol{S}_{t,j}|}{|\boldsymbol{S}_{t,j-1}|}, \text{ for } j = 2, \cdots, p. \tag{2.53}$$

It is well known that $v_{t,j} = s_{t,jj} - \boldsymbol{s}_{t,j}'\boldsymbol{S}_{t,j-1}^{-1}\boldsymbol{s}_{t,j} > 0$ for $i = 2, \cdots, p$. We define $\boldsymbol{\psi}_j$ as the $j-1$ elements of jth column of $\boldsymbol{\Psi}$ above the diagonal element. So for $j = 2, \cdots, p$,

$\boldsymbol{\psi}_j = (\psi_{1j}, \cdots, \psi_{j-1,j})'$. We then have a recursive formula,

$$tr(\boldsymbol{\Psi}_p' \boldsymbol{S}_{t,p} \boldsymbol{\Psi}_p \boldsymbol{\Lambda}_{t,p}^{-1}) = tr(\boldsymbol{\Psi}_{p-1}' \boldsymbol{S}_{t,p-1} \boldsymbol{\Psi}_{p-1} \boldsymbol{\Lambda}_{t,p-1}^{-1}) + \lambda_{t,p}^{-1} v_{t,p} + \lambda_{t,p}^{-1} \boldsymbol{g}_t' \boldsymbol{S}_{t,p-1} \boldsymbol{g}_t,$$

where $\boldsymbol{g}_t = \boldsymbol{\psi}_{p-1} + \boldsymbol{S}_{t,p-1}^{-1} \boldsymbol{s}_{t,p}$. The likelihood function $[\boldsymbol{y} \mid \boldsymbol{B}, \boldsymbol{\Psi}, \boldsymbol{\Lambda}_1, \cdots, \boldsymbol{\Lambda}_T]$ of (2.20) can then be written as

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

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

Fact 2.2. *(a) The conditional posterior distributions of $\boldsymbol{\psi}_2, \cdots, \boldsymbol{\psi}_p$ given $(\boldsymbol{\gamma}_b, \boldsymbol{B}, \boldsymbol{\gamma}_\psi, \boldsymbol{a}_0, \boldsymbol{\beta}, \boldsymbol{A}_1, \boldsymbol{\delta}, \boldsymbol{\gamma}_a, \boldsymbol{\Lambda}, \boldsymbol{q}; \boldsymbol{y})$ are independent and depend only on $(\boldsymbol{B}, \boldsymbol{\gamma}_\psi, \boldsymbol{\Lambda}; \boldsymbol{y})$. They have the form,*

$$(\boldsymbol{\psi}_j \mid \boldsymbol{B}, \boldsymbol{\gamma}_\psi, \boldsymbol{\Lambda}; \boldsymbol{y}) \overset{ind}{\sim} N_{j-1}(\boldsymbol{\mu}_j, \boldsymbol{\Delta}_j), \tag{2.54}$$

where

$$\begin{aligned} \boldsymbol{\Delta}_j &= \{ \sum_{t=1}^{T} \lambda_{jt}^{-1} \boldsymbol{S}_{t,j-1} + \boldsymbol{D}_{\psi,j}^{-1} \}^{-1}, \\ \boldsymbol{\mu}_j &= -\boldsymbol{\Delta}_j \sum_{t=1}^{T} \lambda_{jt}^{-1} \boldsymbol{s}_{t,j}. \end{aligned} \tag{2.55}$$

(b) For $j = 2, \cdots, p$ and $i = 1, \cdots, j-1$, denote the vector $\boldsymbol{\gamma}_{\psi,(-ij)} = (\gamma_{\psi,1j}, \cdots, \gamma_{\psi,i-1,j}, \gamma_{\psi,i+1,j}, \cdots, \gamma_{\psi,j-1,j})'$. Given prior independence for $(\psi_{1j}, \cdots, \psi_{j-1,j})$, the conditional posterior of $\gamma_{\psi,ij}$ given $(\boldsymbol{\gamma}_b, \boldsymbol{B}, \boldsymbol{\Psi}, \boldsymbol{\gamma}_{\psi,(-ij)}, \boldsymbol{a}_0, \boldsymbol{\beta}, \boldsymbol{A}_1, \boldsymbol{\delta}, \boldsymbol{\gamma}_a, \boldsymbol{\Lambda}, \boldsymbol{q}; \boldsymbol{y})$ depends only on $\boldsymbol{\Psi}$ and has the form,

$$(\gamma_{\psi,ij} \mid \boldsymbol{\Psi}) = (\gamma_{\psi,ij} \mid \psi_{ij}) \overset{ind}{\sim} Bernoulli\Big(\frac{u_{\psi,ij1}}{u_{\psi,ij1} + u_{\psi,ij2}}\Big), \tag{2.56}$$

where

$$\begin{aligned} u_{\psi,ij1} &= \frac{1}{c_{\psi,ij}} \exp\Big(-\frac{\psi_{ij}^2}{2 c_{\psi,ij}^2 \kappa_{\psi,ij}^2}\Big) p_{\psi,ij}, \\ u_{\psi,ij2} &= \exp\Big(-\frac{\psi_{ij}^2}{2 \kappa_{\psi,ij}^2}\Big) (1 - p_{\psi,ij}). \end{aligned} \tag{2.57}$$

Proof. The conditional posterior of $(\boldsymbol{\psi}_2, \cdots, \boldsymbol{\psi}_p)$, given $(\boldsymbol{B}, \boldsymbol{\Lambda}_1, \cdots, \boldsymbol{\Lambda}_T, \boldsymbol{\gamma}_\psi; \boldsymbol{y})$ is

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

where $\boldsymbol{\mu}_j$ and $\boldsymbol{\Delta}_j$ are defined in (2.55). Part (a) follows from direct computation. For part (b), recall the fact that $\boldsymbol{\gamma}_\psi$ depends on data indirectly, then,

$$\begin{aligned}
u_{\psi,ij1} &\propto [\boldsymbol{\psi}_j \mid \boldsymbol{\gamma}_{\psi,(-ij)}, \gamma_{\psi,ij} = 1]p_{\psi,ij}, \\
u_{\psi,ij2} &\propto [\boldsymbol{\psi}_j \mid \boldsymbol{\gamma}_{\psi,(-ij)}, \gamma_{\psi,ij} = 0](1 - p_{\psi,ij}).
\end{aligned}$$

The expression above, under prior independence of $\psi_{1j}, \cdots, \psi_{j-1,j}$, gives the formula (2.57). □

2.5.3 Conditional Posterior of $(\boldsymbol{a}_0, \boldsymbol{\beta}_j)$, $(\boldsymbol{A}_1, \boldsymbol{\gamma}_a)$ and $(\boldsymbol{\delta}, \boldsymbol{\gamma}_\delta, \boldsymbol{q})$

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

$$\boldsymbol{W}_j = \begin{pmatrix} 1 & h_{0j} & z_{11} & \cdots & z_{1r} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & h_{T-1,j} & z_{T1} & \cdots & z_{Tr} \end{pmatrix}. \tag{2.58}$$

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

Fact 2.3. *(a) The conditional posterior distributions of* $\widetilde{\boldsymbol{a}}_j^* = (a_{j0}, \beta_j, \widetilde{\boldsymbol{a}}_j')', j = 1, \cdots, p$ *given* $(\boldsymbol{\gamma}_b, \boldsymbol{B}, \boldsymbol{\gamma}_\psi, \boldsymbol{\Psi}, \boldsymbol{\delta}, \boldsymbol{\gamma}_a, \boldsymbol{H}, \boldsymbol{q}; \boldsymbol{y})$ *are independent and depend only on* $(\boldsymbol{H}, \boldsymbol{\gamma}_a, \boldsymbol{\delta})$.

$$(\widetilde{\boldsymbol{a}}_j^* \mid \boldsymbol{H}, \boldsymbol{\gamma}_a, \boldsymbol{\delta}) = (\widetilde{\boldsymbol{a}}_j^* \mid \widetilde{\boldsymbol{h}}_j, \boldsymbol{\gamma}_{a,j}, \delta_j) \overset{ind}{\sim} N_{r+2}(\boldsymbol{\nu}_j, \boldsymbol{\Upsilon}_j), \tag{2.59}$$

where

$$\begin{aligned}\boldsymbol{\nu}_j &= (\delta_j^{-2}\boldsymbol{W}_j'\boldsymbol{W}_j + \boldsymbol{\Omega}_j^{-1})^{-1}(\delta_j^{-2}\boldsymbol{W}_j'\widetilde{\boldsymbol{h}}_j + \boldsymbol{\Omega}_j^{-1}\bar{\boldsymbol{a}}_j),\\ \boldsymbol{\Upsilon}_j &= (\delta_j^{-2}\boldsymbol{W}_j'\boldsymbol{W}_j + \boldsymbol{\Omega}_j^{-1})^{-1}.\end{aligned}$$

(b) The conditional posterior of $(\delta_1, \cdots, \delta_p)$ *given* $(\boldsymbol{\gamma}_\delta, \boldsymbol{\gamma}_b, \boldsymbol{B}, \boldsymbol{\gamma}_\psi, \boldsymbol{\Psi}, \boldsymbol{a}_0, \boldsymbol{\beta}, \boldsymbol{A}_1, \boldsymbol{\gamma}_a, \boldsymbol{H}, \boldsymbol{q}; \boldsymbol{y})$ *are mutually independent and depend only on* $(\boldsymbol{\gamma}_\delta, \boldsymbol{H}, \boldsymbol{a}_0, \boldsymbol{\beta}, \boldsymbol{A}_1, \boldsymbol{q})$*,*

$$\begin{aligned}(\delta_j^2 \mid \boldsymbol{\gamma}_\delta, \boldsymbol{H}, \boldsymbol{a}_0, \boldsymbol{\beta}, \boldsymbol{A}_1) &= (\delta_j^2 \mid \gamma_{\delta j}, \widetilde{\boldsymbol{h}}_j, a_{j0}, \beta_j, \widetilde{\boldsymbol{a}}_j)\\ &\overset{ind}{\sim} IG\Big(v_{j0} + \frac{1}{2}(T-2), \eta_{\delta j}^2 s_{j0}^2 + \frac{1}{2}(\widetilde{\boldsymbol{h}}_j - \boldsymbol{W}_j\widetilde{\boldsymbol{a}}_j^*)'(\widetilde{\boldsymbol{h}}_j - \boldsymbol{W}_j\widetilde{\boldsymbol{a}}_j^*)\Big). \qquad (2.60)\end{aligned}$$

(c) For $j = 1, \cdots, p$*, the conditional posterior of* $\gamma_{\delta j}$ *given* $(\boldsymbol{\gamma}_b, \boldsymbol{B}, \boldsymbol{\Psi}, \boldsymbol{\gamma}_a, \boldsymbol{a}_0, \boldsymbol{\beta}, \boldsymbol{A}_1, \boldsymbol{\delta}, \boldsymbol{\gamma}_\psi, \boldsymbol{H}, \boldsymbol{q}; \boldsymbol{y})$ *depends only on* (δ_j, q_j)*, and has the form*

$$(\gamma_{\delta j} \mid \delta_j, q_j) \overset{ind}{\sim} Bernoulli\Big(\frac{u_{\delta j1}}{u_{\delta j1} + u_{\delta j2}}\Big), \qquad (2.61)$$

where

$$u_{\delta j1} = \exp\Big(-\frac{q_j s_{j0}}{\delta_j^2}\Big) p_{\delta j} {q_j}^{v_{j0}} \quad \textit{and} \quad u_{\delta j2} = \exp\Big(-\frac{s_{j0}}{\delta_j^2}\Big)(1 - p_{\delta j}). \qquad (2.62)$$

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

$$(\gamma_{a,jk} \mid \boldsymbol{A}_1) = (\gamma_{a,jk} \mid a_{jk}) \overset{ind}{\sim} Bernoulli\Big(\frac{u_{a,jk1}}{u_{a,jk1} + u_{a,jk2}}\Big), \qquad (2.63)$$

where

$$\begin{aligned}u_{a,jk1} &= \frac{1}{c_{a,jk}} \exp\Big(-\frac{a_{jk}^2}{c_{a,jk}^2 2\kappa_{a,jk}^2}\Big) p_{a,jk},\\ u_{a,jk2} &= \exp\Big(-\frac{a_{jk}^2}{2\kappa_{a,jk}^2}\Big)(1 - p_{a,jk}). \qquad (2.64)\end{aligned}$$

(e) For $j = 1, \cdots, p$, the conditional posterior of q_j given $(\boldsymbol{\gamma}_b, \boldsymbol{B}, \boldsymbol{\Psi}, \boldsymbol{\gamma}_a, \boldsymbol{a}_0, \boldsymbol{\beta}, \boldsymbol{A}_1, \boldsymbol{\delta}, \boldsymbol{\gamma}_\psi, \boldsymbol{H}, \boldsymbol{\gamma}_\delta; \boldsymbol{y})$ depends only on $(\delta_j, \boldsymbol{\gamma}_{\delta j})$, and has the form

$$(q_j \mid \delta_j, \boldsymbol{\gamma}_{\delta j}) \quad \sim \quad Ga(v_{jo}\gamma_{\delta j} + \alpha_q, \ \frac{s_{j0}}{\delta_j^2}\gamma_{\delta j} + \beta_q). \tag{2.65}$$

Proof. Parts (a) and (b) can be easily proved using regression theory results. For part (c), recall that $\boldsymbol{\gamma}_\delta$ depends on data indirectly, then,

$$u_{\delta j1} \propto [\widetilde{\delta}_j \mid \boldsymbol{\gamma}_{\delta j} = 1]p_{\delta j} \ \text{ and } \ u_{\delta j2} \propto [\widetilde{\delta}_j \mid \boldsymbol{\gamma}_{a,jk} = 0](1 - p_{\delta j}).$$

Substituting the density of Inverse Gamma to the expressions above gives the formula (2.62). Note that the scale parameter cancels out and does not affect the conditional posterior of the model index $\gamma_{\delta j}$. For part (d), recall that $\boldsymbol{\gamma}_a$ depends on data indirectly, then,

$$\begin{aligned} u_{a,jk1} &\propto [\widetilde{\boldsymbol{a}}_j \mid \boldsymbol{\gamma}_{a,(-jk)}, \boldsymbol{\gamma}_{a,jk} = 1]p_{a,jk}, \\ u_{a,jk2} &\propto [\widetilde{\boldsymbol{a}}_j \mid \boldsymbol{\gamma}_{a,(-jk)}, \boldsymbol{\gamma}_{a,jk} = 0](1 - p_{a,jk}). \end{aligned}$$

The two expressions, together with prior independence of $(a_{j1}, \cdots, a_{jr})$, give the formula (2.64). Part (e) comes from direct computation. □

2.5.4 Sampling $\boldsymbol{\Lambda}$

Updating the stochastic volatilities λ_{jt} is more complicated, because their full conditional does not have a closed form. Uhlig (1997), using a Beta distribution for the ratio of the volatilities, found an analytical solution for the full conditional of the volatilities, but his Gibbs sampler requires numerical integration step. Jacquier et al. (1994) use a Metropolis-Hasting for their univariate model, but such algorithm has not proven to be efficient for our model. Other methods have been proposed by the SV literature,

and here we implement some of them. As a preliminary step, we rewrite the likelihood function of $(\boldsymbol{B}, \boldsymbol{\Psi}, \boldsymbol{\Lambda}_1, \cdots, \boldsymbol{\Lambda}_T)$ as

$$\begin{aligned} & L(\boldsymbol{B}, \boldsymbol{\Psi}, \boldsymbol{\Lambda}_1, \cdots, \boldsymbol{\Lambda}_T) = [\boldsymbol{Y} \mid \boldsymbol{B}, \boldsymbol{\Psi}, \boldsymbol{\Lambda}_1, \cdots, \boldsymbol{\Lambda}_T] \\ \propto & \prod_{t=1}^{T} \prod_{j=1}^{p} \lambda_{jt}^{-\frac{1}{2}} etr\Big\{-\frac{1}{2} \sum_{t=1}^{T} \boldsymbol{\Psi} \boldsymbol{S}_t(\boldsymbol{B}) \boldsymbol{\Psi}' \boldsymbol{\Lambda}_t^{-1}\Big\} \\ = & \prod_{t=1}^{T} \prod_{j=1}^{p} \lambda_{jt}^{-\frac{1}{2}} etr\Big\{-\frac{1}{2} \sum_{t=1}^{T} \boldsymbol{Q}_t(\boldsymbol{B}, \boldsymbol{\Psi}) \boldsymbol{\Lambda}_t^{-1}\Big\}, \end{aligned} \tag{2.66}$$

where $\boldsymbol{S}_t(\boldsymbol{B})$ is given by (2.21) and

$$\boldsymbol{Q}_t(\boldsymbol{B}, \boldsymbol{\Psi}) \equiv \boldsymbol{\Psi} \boldsymbol{S}_t(\boldsymbol{B}) \boldsymbol{\Psi'} = (q_{t,ij})_{p \times p}. \tag{2.67}$$

It is more efficient to simulate the $h_{jt} = log(\lambda_{jt})$ instead of λ_{jt}, and there are several options to do so. In the next three subsections we discuss three of them: direct sampling using the Gilks adaptive rejection sampler, filtering and smoothing the volatilities with a particle filter, and filtering the volatilities using a rejection sampler.

Direct Sampling Using Gilks Adaptive Rejection Sampler

Consider the distribution of $\boldsymbol{h}_t$ given $\boldsymbol{h}_{t-1}$ and parameter vector $\boldsymbol{\theta} = (\boldsymbol{b}', \boldsymbol{\psi}_1', \cdots, \boldsymbol{\psi}_p', \widetilde{\boldsymbol{a}}_1', \cdots, \widetilde{\boldsymbol{a}}_p')'$,

$$(\boldsymbol{h}_t \mid \boldsymbol{h}_{t-1}, \boldsymbol{\theta}) \sim N_p\Big(\boldsymbol{a}_0 + diag(h_{1,t-1}, \cdots, h_{p,t-1})\boldsymbol{\beta} + \boldsymbol{A}_1 \boldsymbol{z}_t,\ diag(\delta_1^2, \cdots, \delta_p^2)\Big). \tag{2.68}$$

It is evident that this density can be decomposed as a product of univariate densities:

$$[\boldsymbol{h}_t \mid \boldsymbol{h}_{t-1}, \boldsymbol{\theta}] = \prod_{j=1}^{p} [h_{jt} \mid h_{j,t-1}, \boldsymbol{\theta}], \tag{2.69}$$

where

$$(h_{jt} \mid h_{j,t-1}, \boldsymbol{\theta}) \sim N\Big(u_{jt}, \delta_j^2\Big). \tag{2.70}$$

Here $u_{jt} = u_{jt}(h_{j,t-1}, \boldsymbol{\theta}) = a_{0j} + \beta_j h_{j,t-1} + \widetilde{\boldsymbol{a}}_j \boldsymbol{z}_t$. Using the previous result, we can derive a sampler for $\boldsymbol{h}_t$.

Fact 2.4. *(a) At time t, the posterior distributions of* $(h_{1t}, \cdots, h_{pt})$ *given* $(\boldsymbol{\gamma}_b, \boldsymbol{B}, \boldsymbol{\gamma}_\psi, \boldsymbol{\Psi}, \boldsymbol{a}_0, \boldsymbol{\beta}, \boldsymbol{A}_1, \boldsymbol{\gamma}_a, \boldsymbol{H}_{(-j)}, \boldsymbol{\delta}; \boldsymbol{y})$ *are independent. The conditional density of each* h_{jt} *depends only on* $(\boldsymbol{B}, \boldsymbol{\Psi}, \widetilde{\boldsymbol{a}}_j^*, \boldsymbol{\delta}, h_{j,t-1}, h_{j,t+1}; \boldsymbol{y})$ *and is given by*

$$[h_{jt} \mid \boldsymbol{B}, \boldsymbol{\Psi}, \widetilde{\boldsymbol{a}}_j^*, \boldsymbol{\delta}, h_{j,t-1}, h_{j,t+1}; \boldsymbol{y}] \propto \exp\Big\{-\frac{1}{2}\Big(h_{jt} + q_{t,jj} e^{-h_{jt}} + \frac{(h_{jt} - \varphi_{jt})^2}{\tau_j^2}\Big)\Big\}, \quad (2.71)$$

where

$$\begin{aligned} \tau_j^2 &= \delta_j^2/(1+\beta_j^2), \\ \varphi_{jt} &= \frac{1}{1+\beta_j^2}\Big[a_{j0}(1-\beta_j) + \beta_j(h_{j,t+1} + h_{j,t-1}) + \sum_{k=1}^{r} a_{jk}(z_{t-1,k} - \beta_j z_{t,k})\Big]. \end{aligned} \quad (2.72)$$

(b) The conditional density of h_{jt} *in (2.71) is log-concave.*

Proof. For part (a), we consider the series of univariate conditional densities,

$$\begin{aligned} & [h_{jt} \mid h_{j,t-1}, h_{j,t+1}, \boldsymbol{a}_0, \boldsymbol{\beta}, \boldsymbol{A}_1, \boldsymbol{\delta}, \boldsymbol{B}, \boldsymbol{\Psi}; \boldsymbol{y}] \\ \propto\ & [\boldsymbol{y} \mid \boldsymbol{H}, \boldsymbol{B}, \boldsymbol{\Psi}][h_{jt} \mid h_{j,t-1}, a_{j0}, \beta_j, \widetilde{\boldsymbol{a}}_j, \delta_j][h_{j,t+1} \mid h_{jt}, a_{j0}, \beta_j, \widetilde{\boldsymbol{a}}_j, \delta_j], \end{aligned}$$

which is proportional to (2.71). For part (b) consider the second derivative of the logarithm of (2.71):

$$\frac{d^2}{dh_{jt}^2} log(g(h_{jt}))\} = -\frac{1}{2} q_{t,jj} \exp(-h_{t,j}) - \frac{1}{\tau_j^2},$$

which is negative. □

The log-concavity of the conditional posterior of h_{jt} permits efficient simulation of h_{jt} using the adaptive rejection sampling algorithm given by Gilks & Wild (1992). To assure stationarity, we simulate h_{j0} and the $h_{j,T+1}$ from $N(a_{j0}/(1-\beta_j), \delta_j^2/(1-\beta_j^2))$

independently. This method is considerably less computationally intensive than the Metropolis-Hastings algorithms used in previous literature (Jacquier et al. (1994) among others).

Filtering and Smoothing $\mathbf{\Lambda}$

One drawback of using the rejection method based on the log-concavity is that the generated samples of stochastic volatility $\mathbf{\Lambda}$ exhibit a strong serial correlation and slow convergence to the stationary distribution. To deal with this problem, several authors proposed the use of particle filters for sequential importance sampling: see for example Chib et al. (2002, 2005). This method differs from the Kalman filter because of the nonlinearity of the model (note that the state equation concerns the logarithm of conditional variance). As noted earlier, conditional on data $\boldsymbol{Y}_t$ (available up to time t) and parameters $\boldsymbol{\theta}$, the density of $\mathbf{\Lambda}$ is

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

and

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

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

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

An importance sampling particle filter is as follows.

Algorithm F1

Suppose at *Stage t*, we have $(\boldsymbol{h}_{t-1}^1, \cdots, \boldsymbol{h}_{t-1}^M)$ drawn from $(\boldsymbol{h}_{t-1} \mid \boldsymbol{Y}_{t-1}, \boldsymbol{\theta})$.

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

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

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

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

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

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

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

$$[\boldsymbol{h}_T, \boldsymbol{h}_{T-1}, \cdots \boldsymbol{h}_1 \mid \boldsymbol{y}_T, \boldsymbol{y}_{T-1}, \cdots \boldsymbol{y}_1, \boldsymbol{\theta}] = [\boldsymbol{h}_T \mid \boldsymbol{y}_T, \boldsymbol{\theta}] \prod_{t=1}^{T-1} [\boldsymbol{h}_t \mid \boldsymbol{h}_{t+1}, \boldsymbol{y}_t, \boldsymbol{\theta}]. \tag{2.73}$$

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

$$[\boldsymbol{h}_t \mid \boldsymbol{h}_{t+1}, \boldsymbol{y}_t, \boldsymbol{\theta}] = [\boldsymbol{h}_t \mid \boldsymbol{y}_t, \boldsymbol{\theta}][\boldsymbol{h}_{t+1} \mid \boldsymbol{h}_t, \boldsymbol{\theta}],$$

where $[\boldsymbol{h}_{t+1} \mid \boldsymbol{h}_t, \boldsymbol{\theta}]$ is the Gaussian model by assumption of the stochastic volatility equation (2.8), and a numerical draw of $[\boldsymbol{h}_t \mid \boldsymbol{y}_t, \boldsymbol{\theta}]$ is the result of filtering (see Carter & Kohn (1994, 1996) for details). Smoothing for $(\boldsymbol{h} \mid \boldsymbol{Y}, \boldsymbol{\theta})$ is achieved by utilizing the recursive structure of (2.73) and implemented using the following algorithm.

Algorithm S

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

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

Because the above filters involve a multinomial resampling in Algorithm F1 *Step 3*, additional noise will be added and may produce outliers. Refinements of the particle filter includes Pitt & Shephard (1999) auxiliary sampling importance resampling filter, which samples the index of particles jointly with the state variable, then integrate out the the indices for filtering the state variable. The refined algorithm is less likely to produce outliers. The Pitt-Shephard algorithm is as follows.

Algorithm F2

Given M draws of $\boldsymbol{h}_{t-1}^l$ from the posterior $[\boldsymbol{h}_{t-1} | \boldsymbol{Y}_{t-1}, \boldsymbol{\theta}]$,

Steps 1 − 2. The same as Algorithm F1.

Step 3. Take a random sample of size R from $\{1, \cdots, M\}$ with relative weight w_t^l. Denote the result of the draws as $k_1, k_2, \cdots, k_R$.

Step 4. For each value k_i draw $\boldsymbol{h}_t^{*i}$ from $(\boldsymbol{h}_t \mid \boldsymbol{h}_{t-1}^{k_i}, \boldsymbol{\theta})$.

Step 5. Resample $\{\boldsymbol{h}_t^{*1}, \boldsymbol{h}_t^{*2}, \cdots, \boldsymbol{h}_t^{*R}\}$ M times with probabilities proportional to

$$\frac{[\boldsymbol{y}_t \mid \boldsymbol{h}_t^{*i}, \boldsymbol{\theta}]}{[\boldsymbol{y}_t \mid \boldsymbol{h}_t^{k_i}, \boldsymbol{\theta}]}.$$

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

Filtering the Volatilities via Rejection Sampling

Chib et al. (1998) criticize the Algorithm F1 to be inefficient, and instead propose a rejection sampling scheme to filter the volatilities of an univariate SV model. We adapt their algorithm for our multivariate model. Instead of sampling the vector $\boldsymbol{h}_t$ at once, we developed a method to sample h_{jt} elementwise, using the fact that for $j = 1, \cdots, p$ the conditional distributions of $(h_{jt} \mid h_{j,t-1}, \boldsymbol{\theta})$ are independent. In this way we reduce the rejection rate of the sampler. In fact, if we accept each h_{jt} with probability q_j, then the probability to accept the entire vector $\boldsymbol{h}_t$ at once is $\prod_{j=1}^{p} q_j$, which could be small and inefficient.

Fact 2.5. *Define* $\boldsymbol{\xi}_t = (\xi_{1t}, \cdots, \xi_{pt})' \equiv \boldsymbol{\Psi}'(\boldsymbol{y}_t - \boldsymbol{b}_0 - \boldsymbol{B}_1 \boldsymbol{x}_t)$ *and* $f(h_{jt}; \xi_{jt}) = \phi(\xi_{jt}; 0, e^{h_{jt}})$. *Here* $\phi(\cdot; \mu, \sigma^2)$ *is the density of normal with mean* μ *and variance* σ^2. *Suppose at Stage* t *we have* $(h_{1t}^1, \cdots, h_{1t}^M)$ *drawn from* $[h_{j,t-1} \mid \boldsymbol{Y}_{t-1}, \boldsymbol{\theta}]$, *then*

$$[\boldsymbol{h}_t \mid \boldsymbol{Y}_t, \boldsymbol{\theta}] \approx C \prod_{j=1}^{p} \Big\{ f(h_{jt}, \xi_{jt}) \frac{1}{M} \sum_{l=1}^{M} [h_{jt} \mid h_{j,t-1}^l, \boldsymbol{\theta}] \Big\}, \tag{2.74}$$

where $[h_{jt} \mid h_{j,t-1}^l, \boldsymbol{\theta}] = \phi(h_{jt}; u_{jt}^l, \delta_j^2)$. *Here* $u_{jt}^l = u_{jt}(h_{j,t-1}^l, \boldsymbol{\theta}) \;=\; a_{0j} + \beta_j h_{j,t-1}^l + \widetilde{\boldsymbol{a}}_j \boldsymbol{z}_t$.

Proof. From Bayes theorem we know that

$$[\boldsymbol{h}_t \mid \boldsymbol{Y}_t, \boldsymbol{\theta}] \propto [\boldsymbol{y}_t \mid \boldsymbol{h}_t, \boldsymbol{\theta}][\boldsymbol{h}_t \mid \boldsymbol{Y}_{t-1}, \boldsymbol{\theta}]. \tag{2.75}$$

Note that we can decompose (2.18) as

$$[\boldsymbol{y}_t \mid \boldsymbol{h}_t, \boldsymbol{\theta}] \;=\; \prod_{j=1}^{p} e^{-\frac{1}{2} h_{jt}} (2\pi)^{-\frac{1}{2}} \exp\Big(-\frac{1}{2} \xi_{jt}^2 e^{-h_{jt}} \Big) = \prod_{j=1}^{p} f(h_{jt}; \xi_{jt}). \tag{2.76}$$

Then, (2.75) can be written as

$$[\boldsymbol{h}_t \mid \boldsymbol{Y}_t, \boldsymbol{\theta}] \propto \prod_{j=1}^{p} f(h_{jt}, \xi_{jt}) [\boldsymbol{h}_t \mid \boldsymbol{Y}_{t-1}, \boldsymbol{\theta}]. \tag{2.77}$$

Also, using (2.69) and Fact 4 (a), we can write

$$[\boldsymbol{h}_t \mid \boldsymbol{Y}_{t-1}, \boldsymbol{\theta}] = \prod_{j=1}^{p} [h_{jt} \mid \boldsymbol{Y}_{t-1}, \boldsymbol{\theta}] = \prod_{j=1}^{p} \int [h_{jt} \mid h_{j,t-1}, \boldsymbol{\theta}][h_{j,t-1} \mid \boldsymbol{Y}_{t-1}, \boldsymbol{\theta}]\, dh_{j,t-1}. \quad (2.78)$$

Each of the above integrals can be approximated by $\frac{1}{M}\sum_{l=1}^{M}[h_{jt} \mid h_{j,t-1}^{l}, \boldsymbol{\theta}]$. Using (2.77) and (2.78), the proof is then obvious. □

From (2.74), $h_{1t}, \cdots, h_{pt}$ given $(\boldsymbol{Y}_t, \boldsymbol{\theta})$ are independent. Here is the particle filter rejection sampler to sample h_{jt} from (2.74). Taking the logarithm of $f(h_{jt}, \xi_{jt})$ we obtain

$$\log f(h_{jt}, \xi_{jt}) \;=\; -\frac{1}{2}\log(2\pi) - \frac{1}{2}h_{jt} - \frac{1}{2}J, \quad (2.79)$$

where

$$J \;=\; J(h_{jt}) \;=\; \xi_{jt}^2 e^{-h_{jt}}. \quad (2.80)$$

Note that the second derivative of J is always positive. For v_{jt} defined by (2.70), consider the first order Taylor expansion of J at u_{jt},

$$J \;=\; \xi_{jt}^2 e^{-u_{jt}} - \xi_{jt}^2 e^{-u_{jt}}(h_{jt} - u_{jt}). \quad (2.81)$$

Consequently, we have the following inequality,

$$\begin{aligned} f(h_{jt}, \xi_{jt}) \;&\leq\; (2\pi)^{-\frac{1}{2}} exp\Big\{-\frac{1}{2}h_{jt}(1 - \xi_{jt}^2 e^{-u_{jt}}) - \frac{1}{2}e^{-u_{jt}}(1 + \xi_{jt}^2 u_{jt})\Big\} \\ &=\; g(h_{jt}, u_{jt}, \xi_{jt}^2). \end{aligned} \quad (2.82)$$

Given a sample of draws $h_{t-1}^{1}, \cdots, h_{t-1}^{M}$, (2.82) and (2.74) imply that

$$[h_{jt} \mid \boldsymbol{Y}_t, \boldsymbol{\theta}] \;\leq\; \frac{1}{M}\sum_{l=1}^{M} g(h_{jt}, u_{jt}^{l}, \xi_{jt}^2)\phi(h_{jt}; u_{jt}^{l}, \delta_j). \quad (2.83)$$

Now, consider

$$\begin{aligned} & g(h_{jt}, u_{jt}^l, \xi_{jt}^2)\phi(h_{jt}; u_{jt}^l, \delta_j) \\ = & \frac{1}{2\pi\delta_j} exp\Big\{-\frac{1}{2\delta_j^2}(h_{jt} - c_{jt}^l)^2 + \frac{c_{jt}^l}{2\delta_j^2} - \frac{1}{2}e^{u_{jt}^l}(1 + u_{jt}^l\xi_{jt}^2) - \frac{(u_{jt}^l)^2}{2\delta_j^2}\Big\} \\ = & \widetilde{\pi}_{jt}^l \phi(h_{jt}; c_{jt}^l, \delta_j), \end{aligned} \tag{2.84}$$

where

$$\begin{aligned} c_{jt}^l &= u_{jt}^l - \frac{\delta_j^2}{2}(1 - e^{u_{jt}^l}\xi_{jt}^2) = c_{jt}^l(h_{j,t-1}^l, \boldsymbol{\theta}), \\ \widetilde{\pi}_{jt}^l &= exp\Big\{-\frac{u_{jt}^l + (c_{jt}^l)^2}{2\delta_j} - \frac{1}{2}e^{-u_{jt}^l}(1 + u_{jt}^l\xi_{jt}^2)\Big\}. \end{aligned}$$

Define

$$\pi_{jt}^l = \frac{\widetilde{\pi}_{jt}^l}{\sum_{l=1}^M \widetilde{\pi}_{jt}^l}, \tag{2.85}$$

then (2.84), along with (2.83), allows us to develop the following algorithm:

Algorithm F3

Step 1: Given $h_{t-1}^1, \cdots, h_{t-1}^M$, for any fixed $j = 1, \cdots, p$, sample $l \in \{1, \cdots, M\}$ with probability π_{jt}^l.

Step 2: Sample $\widetilde{h}_{jt} \sim N(c_{jt}^l, \delta_j)$, and independently $\widetilde{u} \sim Uniform[0, 1]$.

Step 3: If $\widetilde{u} \leq f(\widetilde{h}_{jt}, \xi_{jt}^2)/g(\widetilde{h}_{jt}, u_{jt}^l, \xi_{jt}^2)$, report $h_{jt} = \widetilde{h}_{jt}$.

2.5.5 MCMC Algorithms

From the result of the previous subsections, we define three different Gibbs sampling procedures to produce finite samples from the marginal posterior distributions of $(\boldsymbol{B}, \boldsymbol{\gamma}_b, \boldsymbol{\Psi}, \boldsymbol{\gamma}_\psi, \boldsymbol{a}_0, \boldsymbol{\beta}, \boldsymbol{A}_1, \boldsymbol{\gamma}_a, \boldsymbol{\delta}, \boldsymbol{\gamma}_\delta, \boldsymbol{q}, \boldsymbol{\Lambda})$.

Algorithm 1 (Direct Sampling)

Suppose in cycle k we have $(\boldsymbol{\Psi}^{(k-1)}, \boldsymbol{B}^{(k-1)}, \boldsymbol{a}_0^{(k-1)}, \boldsymbol{\beta}^{(k-1)}, \boldsymbol{A}_1^{(k-1)}\boldsymbol{\delta}^{(k-1)}, \boldsymbol{\gamma}_b^{(k-1)}, \boldsymbol{\gamma}_\psi^{(k-1)}, \boldsymbol{\gamma}_a^{(k-1)}, \boldsymbol{\gamma}_\delta^{(k-1)}, \boldsymbol{q}^{(k-1)}, \boldsymbol{\Lambda}^{(k-1)})$. Then have the following algorithm:

Step 1: Draw $(\boldsymbol{\gamma}_\delta^{(k)} \mid \boldsymbol{\delta}^{(k-1)}, \boldsymbol{q}^{(k-1)})$ from (2.62);

Step 2: Draw $(\boldsymbol{q}^{(k)} \mid \boldsymbol{\delta}^{(k-1)}, \boldsymbol{\gamma}_\delta^{(k)})$ from (2.65);

Step 3: Draw $(\boldsymbol{a}_0^{(k)}, \boldsymbol{\beta}^{(k)}, \boldsymbol{A}_1^{(k)})$: for $j = 1, \cdots, p$, draw $(\widetilde{\boldsymbol{a}}_j^{*(k)} \mid \boldsymbol{\Lambda}^{(k-1)}, \boldsymbol{\gamma}_a^{(k-1)}, \boldsymbol{\delta}^{(k-1)})$ from (2.59).

Step 4: Draw $(\boldsymbol{\delta}^{(k)} \mid \boldsymbol{\gamma}_\delta^{(k)}, \boldsymbol{\gamma}_a^{(k-1)}, \boldsymbol{\gamma}_b^{(k-1)}, \boldsymbol{\gamma}_\psi^{(k-1)}\boldsymbol{\Lambda}^{(k-1)}, \boldsymbol{a}_0^{(k)}, \boldsymbol{\beta}^{(k)}, \boldsymbol{A}_1^{(k)}, \boldsymbol{q}^{(k)})$ from the distribution (2.60).

Step 5: For $j = 1, \cdots, p$, draw $(\boldsymbol{\gamma}_{a,j}^{(k)} \mid \boldsymbol{A}_1^{(k)})$ from the Bernoulli distribution (2.64).

Step 6: Use the adaptive rejection method to draw $(\boldsymbol{\Lambda}^{(k)} \mid \boldsymbol{a}_0^{(k)}, \boldsymbol{\beta}^{(k)}, \boldsymbol{A}_1^{(k)}, \boldsymbol{B}^{(k-1)}, \boldsymbol{\delta}^{(k)}, \boldsymbol{\Psi}^{(k-1)}; \boldsymbol{y})$ using the log-concavity of the posterior of $[h_t \mid h_{t+1}, h_{t-1}\boldsymbol{a}_0, \boldsymbol{\beta}, \boldsymbol{A}_1, \boldsymbol{\delta}, \boldsymbol{B}, \boldsymbol{\Psi}; \boldsymbol{y}]$ as described in Section 3.4.

Step 7: Draw $\boldsymbol{\Psi}^{(k)}$: for $j = 2, \cdots, p$, draw $(\boldsymbol{\psi}_j^{(k)} \mid \boldsymbol{\Lambda}^{(k)}, \boldsymbol{B}^{(k-1)}, \boldsymbol{\gamma}_\psi^{(k-1)}; \boldsymbol{y})$ from the normal distribution (2.54).

Step 8: Draw $(\boldsymbol{\gamma}_\psi^{(k)} \mid \boldsymbol{\Psi}^{(k)})$ from the Bernoulli distribution (2.56).

Step 9: Draw $(\boldsymbol{B}^{(k)} \mid \boldsymbol{\Psi}^{(k)}, \boldsymbol{\Lambda}^{(k)}, \boldsymbol{\gamma}_b^{(k-1)}; \boldsymbol{y})$ from the normal distribution (2.46).

Step 10: Draw $(\boldsymbol{\gamma}_b^{(k)} \mid \boldsymbol{B}^{(k)})$ from the Bernoulli distribution (2.49).

Algorithm 2 (Filtering and Smoothing)

Same as Algorithm 1 except that in Step 6 Algorithm F1 or Algorithm F2 (in combination with Algorithm S) are used in place of the Gilks adaptive rejection sampler.

Algorithm 3 (Filtering via rejection Sampling)

Same as Algorithm 1 except that in Step 6 Algorithm F3 is used for simulating the volatilities λ_{jt}.

2.5.6 Choosing the Best Model

Consider, $\boldsymbol{\gamma} = (\boldsymbol{\gamma}'_\delta, \boldsymbol{\gamma}'_a, \boldsymbol{\gamma}'_\psi, \boldsymbol{\gamma}'_b)$. At the end of the MCMC simulation, a sample from the posterior of $\boldsymbol{\gamma}$ is produced. Each single draw from $\boldsymbol{\gamma}$ represents a particular subset of variables which has been chosen. In order to find the best subset, we can follow George & McCulloch (1993) and compute the sample posterior mode of $\boldsymbol{\gamma}$. The chosen subset of variables will be the most visited during the MCMC simulation. There are two practical drawbacks of this method: first, it may be computationally expensive for very large models, since it requires the storage of all the draws from the posterior of $\boldsymbol{\gamma}$ and calculation of their frequencies. George & McCulloch (1993) applied this method to a number of competing models much smaller than in our setup. For large p, or for long memory processes, the dimension of $\boldsymbol{\gamma}$ can be very high. Since each draw of $\boldsymbol{\gamma}$ is essentially a string of zeroes and ones, it can be read as a binary number. Thus, it can be univocally recoded to decimal or hexadecimal notation, which have a much more compact form. If the string is too long for the computer memory to transform it to decimal without rounding, the same can be cut in parts, and each of them converted. The second practical drawback is that, if the researcher wants to use stochastic search for model calibration, it is not easy to recognize which parameters are very important, which are somewhat important, and which are not. To find how many time parameters outside the modal subset have been chosen it is necessary to reconvert the model indicators to a binary coding and track the single parameter value, which is unfeasible especially for

big models.

Another solution can consist in computing the posterior mean of each element of $\boldsymbol{\gamma}$. The result would be a number between zero and one which gives the researcher an idea about the importance of the single parameter. The choice of the best subset can be made then by choosing a cutoff value for a parameter to be in the model, or simply by the expertise of the researcher. This method might present problems when the posterior means of the single indicators do not clearly approximate to zero or one, in which case the choice of the parameters to be considered in the model is somewhat aleatory. Also, the best model chosen via posterior mean can be a different model than the one chosen via posterior mode. In this case, a model which is not the most visited would be chosen.

2.6 Simulation Studies

In this section we report the results of several simulation studies with the aim to evaluate and compare the performances of the proposed algorithms. First we use Algorithm 1 and we do not perform selection on $\boldsymbol{\delta}$. Next we compare Algorithms 1, 2 and 3, then we analyze the problem related to the selection of the variance term of the stochastic volatility equation. Last, we propose an example using Algorithm 2 and performing selection on $\boldsymbol{\delta}$.

2.6.1 Numerical Examples using Algorithm 1

We propose four examples of the application of Algorithm 1 over simulated data.

Example 2.1. Consider a p variables VAR(1) model with one exogenous variable and the following parameters: $\psi_{ii} = 1$, $\psi_{1j} = .5$, $\gamma_{\psi,1j} = 1$, for $i = 1, \cdots, p$, and $j = 2, \cdots, p$;

$\boldsymbol{b}_0 = (\mathbf{1}'_{p/2}, (2)\mathbf{1}'_{p/2})'$, $\boldsymbol{b}_1 = (3,\ 1.5,\ 1,\ 0.75,\ 0.6,\ 0.5)'$, $b_{i+1,i} = 0.5$, $\gamma_{b,i1} = 1$, $\gamma_{b,i+1,i} = 1$ for $i = 1, \cdots, p$. All the other elements of ($\boldsymbol{\Psi}$, $\boldsymbol{\gamma}_\psi$, $\boldsymbol{B}$, $\boldsymbol{\gamma}_b$) are zero. The data for the stochastic volatility is generated based on:

$$\boldsymbol{A}_1 = (.1)\mathbf{1}_p, \quad \boldsymbol{\beta} = ((.4)\mathbf{1}'_{p/2}, (.6)\mathbf{1}'_{p/2})', \quad \boldsymbol{a}_0 = (.1)\mathbf{1}_p \ \ and \ \ \boldsymbol{\delta} = (.05)\mathbf{1}_p. \tag{2.86}$$

We generate the exogenous variables $\boldsymbol{x}$ and $\boldsymbol{z}$ as $x_{it} = \cos(t/2)$, $z_{it} = \sin(t^2)$, for $i = 1$, and $t = 1, \cdots, T$. The sample size is $T = 1,000$. The hyperparameters of the Bernoulli priors are $p = .5$. The prior on the intercepts is normal with zero mean and a large variance of 50. We let $v_{j0} = 6.0$ and $s^2_{j0} = 0.001$. We set the prior means α_{j0} and β_{j0} at 0 and standard deviation σ_α and σ_β at 10 (consequently the priors are not centered at the true values and are quite flat). Take $\kappa_{ij} = 0.1$, $d_{ij} = 50$, and fix for now the value of $\boldsymbol{\gamma}_\delta$ to one, so all the volatilities are considered stochastic.

We randomly generate 120 samples from the model above with $p = 6$, and for each sample we compute the posterior mean of all parameters using Algorithm 1. Estimates are obtained with $50,000$ MCMC draws after $10,000$ burn-ins. The following results are the average of the posterior means over all samples, while Table 2.1 reports the average over T of the true and estimated volatilities for a randomly chosen sample.

$$\widehat{\boldsymbol{\Psi}} \ = \ \begin{pmatrix} 1 & .5328 & .5127 & .5013 & .5062 & .5089 \\ 0 & 1 & -.0192 & -.0101 & -.0055 & -.0078 \\ 0 & 0 & 1 & -.0091 & -.0034 & -.0040 \\ 0 & 0 & 0 & 1 & .0004 & -.0004 \\ 0 & 0 & 0 & 0 & 1 & -.0025 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix},$$

$$\widehat{\boldsymbol{\gamma}}_{\psi} = \begin{pmatrix} * & .9996 & .9985 & .9976 & .9962 & .9968 \\ * & * & .0221 & .0216 & .0215 & .0217 \\ * & * & * & .0219 & .0215 & .0215 \\ * & * & * & * & .0215 & .0217 \\ * & * & * & * & * & .0212 \\ * & * & * & * & * & * \end{pmatrix},$$

$$\widehat{\boldsymbol{B}} = \begin{pmatrix} .9020 & 3.3394 & .4890 & .0073 & .0048 & .0048 & 0106 & .0083 \\ 1.0415 & 1.0666 & .0105 & .4661 & .0063 & -.0010 & -.0010 & 0.0004 \\ 1.0426 & .6593 & .0048 & .0051 & .4639 & .0035 & .0005 & -.0016 \\ 2.1374 & .4612 & .0071 & -.0008 & .0050 & .4646 & -.0039 & -.0018 \\ 2.1297 & .3444 & .0027 & .0047 & .0050 & -.0005 & .4582 & .0037 \\ 2.1613 & .2650 & .0030 & .0012 & .0016 & .0014 & -.0026 & .4581 \end{pmatrix},$$

$$\widehat{\boldsymbol{\gamma}}_{b} = \begin{pmatrix} 1.000 & .9995 & .0206 & .0209 & .0205 & .0209 & .0211 \\ 1.000 & .00204 & .9969 & .0213 & .0212 & .0209 & .0212 \\ 1.000 & .0203 & .0211 & .9959 & .0212 & .0213 & .0214 \\ .9791 & .0204 & .0208 & .0219 & .9963 & .0213 & .0213 \\ .7745 & .0202 & .0213 & .0214 & .0214 & .9946 & .0213 \\ .4506 & .0200 & .0210 & .0210 & .0212 & .0217 & .9948 \end{pmatrix},$$

$$\widehat{\boldsymbol{A}}_1 = (.1132,\ .1092,\ .1201,\ .1103,\ .1056,\ .1089)',$$

$$\widehat{\boldsymbol{a}}_0 = (.1062,\ .1066,\ .1063,\ .0822,\ .0814,\ .0822)',$$

$$\widehat{\boldsymbol{\beta}} = (.3622,\ .3618,\ .3612,\ .6525,\ .6552,\ .6512)',$$

$$\widehat{\boldsymbol{\delta}} = (.0407,\ .0409,\ .0407,\ .0392,\ .0394,\ .0393)'.$$

Estimates of ($\boldsymbol{\Psi}$, $\boldsymbol{\alpha}$, $\boldsymbol{A}_1$, $\boldsymbol{\delta}$) are very close to the correspondent true values. β_j tends to be underestimated when it is smaller than 0.5 and overestimated otherwise. The

Table 2.1: Comparison of True and Estimated Stochastic Volatilities for One Randomly Chosen Sample

variable	*mean* (*true* h)	*posterior mean* (h)	*S.d.* (*true* h)	*posterior s.d* (h)
1	0.1697	0.1702	0.0549	0.0535
2	0.1647	0.1655	0.0541	0.0514
3	0.1656	0.1677	0.0523	0.0505
4	0.2553	0.2410	0.0627	0.0685
5	0.2433	0.2320	0.0615	0.0653
6	0.2564	0.2412	0.0591	0.0642

second column of $\boldsymbol{B}$ ($\boldsymbol{b}_1$, relative to the exogenous variable of the model), is generally underestimated, while the error for the remaining b_{ij} elements is minimal. Also, the posterior mean of the $\boldsymbol{\gamma}_b$ elements correspondent to $\boldsymbol{b}_1$ is quite far from either zero or one. The results are in general very satisfactory though, especially considering the relatively small number of observations (for such models). The selection of both $\boldsymbol{B}$ and $\boldsymbol{\Psi}$ elements is very accurate, with the exception of few elements of $\boldsymbol{b}_1$. It would be easy to choose which parameters constrain to zero to produce a much simpler model.

We are now proposing some more examples of estimation using Algorithm 1. For every example we slightly change the structure of the true model, in order to see how this can modify the quality of the estimation and variable selection. The second example is to study whether the quality of the estimates depends on the variability of the data.

Example 2.2. Consider now a model with same (p, T, $\boldsymbol{X}$, $\boldsymbol{z}$) as in Example 2.1, but with a more complex parameter structure and a stronger random component of the

volatilities,

$$\boldsymbol{\Psi} = \begin{pmatrix} 1 & .5 & .5 & .6 & .6 & .3 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad \boldsymbol{\gamma}_\psi = \begin{pmatrix} * & 1 & 1 & 1 & 1 & 1 \\ * & * & 0 & 0 & 0 & 0 \\ * & * & * & 0 & 0 & 0 \\ * & * & * & * & 0 & 0 \\ * & * & * & * & * & 0 \\ * & * & * & * & * & * \end{pmatrix}, \tag{2.87}$$

$$\boldsymbol{B} = \begin{pmatrix} 1.00 & 3.00 & 0.30 & 0.00 & 0.20 & 0.00 & 0.20 & 0.00 \\ 1.00 & 1.50 & 0.00 & 0.30 & 0.00 & 0.00 & 0.00 & 0.00 \\ 1.00 & 1.00 & 0.00 & 0.00 & 0.30 & 0.00 & 0.00 & 0.00 \\ 2.00 & 0.75 & 0.00 & 0.00 & 0.00 & 0.30 & 0.00 & 0.00 \\ 2.00 & 0.60 & 0.00 & 0.00 & 0.00 & 0.00 & 0.30 & 0.00 \\ 2.00 & 0.50 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.30 \end{pmatrix}, \tag{2.88}$$

$$\boldsymbol{\gamma}_b = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \tag{2.89}$$

The data generating process for the stochastic volatility is given by

$$\boldsymbol{a}_0 = (.1)\mathbf{1}_6, \quad \boldsymbol{\beta} = (.3, .5, .3, .5, .3, .5)', \quad \boldsymbol{A}_1 = (.1)\mathbf{1}_6, \quad \boldsymbol{\delta} = (.3)\mathbf{1}_6. \tag{2.90}$$

This model is estimated using the same number of MCMC cycles and the same prior hyperparameters as in Example 2.1. From the estimates obtained, we can see how the variability of the latent variable is not incident to the quality of the parameter estimates.

Here are the results:

$$\widehat{\boldsymbol{\Psi}} = \begin{pmatrix} 1 & .5065 & .4963 & .5887 & .5926 & .2733 \\ 0 & 1 & -.0179 & -.0096 & -.0060 & -.0148 \\ 0 & 0 & 1 & -.0091 & -.0055 & -.0087 \\ 0 & 0 & 0 & 1 & -.0003 & -.0066 \\ 0 & 0 & 0 & 0 & 1 & -.0096 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix},$$

$$\widehat{\boldsymbol{\gamma}}_{\psi} = \begin{pmatrix} * & .9982 & .9967 & .9999 & .9999 & .4758 \\ * & * & .0219 & .0213 & .0211 & .0216 \\ * & * & * & .0217 & .0212 & .0215 \\ * & * & * & * & .0213 & .0217 \\ * & * & * & * & * & .0214 \\ * & * & * & * & * & * \end{pmatrix},$$

$$\widehat{\boldsymbol{B}} = \begin{pmatrix} .9884 & 3.1048 & .2921 & .0102 & .1915 & .0060 & .2010 & .0042 \\ 1.0003 & 1.1722 & .0064 & .2757 & .0112 & -.0032 & .0028 & -.0002 \\ 1.0099 & .7693 & .0031 & .0005 & .2795 & .0012 & .0042 & -.0020 \\ 2.0393 & .5590 & .0043 & -.0042 & .0122 & .2775 & .0020 & -.0028 \\ 2.0378 & .4439 & .0013 & .0007 & .0123 & -.0019 & .2754 & .0053 \\ 2.0763 & .3844 & .0027 & -.0021 & .0040 & -.0018 & .0022 & .2690 \end{pmatrix},$$

$$\widehat{\boldsymbol{\gamma}}_{b} = \begin{pmatrix} 1 & .5845 & .0207 & .1341 & .0205 & .1577 & .0212 \\ 1 & .0204 & .4867 & .0214 & .0212 & .0212 & .0215 \\ 1 & .0202 & .0210 & .5079 & .0212 & .0216 & .0216 \\ .9862 & .0203 & .0211 & .0221 & .4983 & .0216 & .0218 \\ .8606 & .0201 & .0213 & .0215 & .0214 & .4877 & .0214 \\ .8472 & .0200 & .0208 & .0210 & .0208 & .0215 & .4516 \end{pmatrix},$$

$$
\begin{aligned}
\widehat{\boldsymbol{a}}_0 &= (.1036, .1041, .1028, .1048, .1049, .1024)', \\
\widehat{\boldsymbol{\beta}} &= (.2637, .4561, .2666, .4555, .2695, .4564)', \\
\widehat{\boldsymbol{A}}_1 &= (.1127, .1112, .1221, .1085, .1096, .1111)', \\
\widehat{\boldsymbol{\delta}} &= (.2902, .2914, .2903, .2927, .2891, .2926).'
\end{aligned}
$$

The major difference between these estimates and those of Example 2.1 lies in $\widehat{\boldsymbol{\gamma}}_b$: the non-zero elements of the column of $\boldsymbol{B}$ relatives to the autoregressive terms ($\boldsymbol{b}_2$ and subsequent), have been selected only half of the time than in Example 2.1. The average of the posterior mean for the correspondent $\gamma_{b,ij}s$ is in fact .9996 in Example 2.1 and .4136 here. The Average of the posterior mean of $\gamma_{b,ij}s$ relative to the zero elements is instead similar for both examples and close to .02. Causes for this difference might be imputed to the higher variability of the latent process, or to the smaller values of the non-zero elements of $\boldsymbol{B}$. Also, there is a tendency of the algorithm to overestimate $\boldsymbol{a}_0$ and $\boldsymbol{A}_1$, and underestimate $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$.

Example 2.3. For this example, (p, T, $\boldsymbol{X}$, $\boldsymbol{z}$) remains the same as Examples 2.1 and 2.2. The main difference of this model is that, given the structure of $\boldsymbol{B}$ and $\boldsymbol{A}_1$, the exogenous variables $\boldsymbol{X}$ and $\boldsymbol{z}$ have respectively a limited and null influence on the outcomes $\boldsymbol{Y}$. Also, here the stochastic volatility generating process has quite a simple structure. Values of $\boldsymbol{B}$ are in between those of the previous examples, except the values of $\boldsymbol{b}_0$, which are much smaller. Also, the variance of the volatility generating process is

smaller than in Example 2.2, but still 3 times bigger than in Example 2.1:

$$\boldsymbol{\Psi} = \begin{pmatrix} 1 & .5 & .5 & .6 & .6 & .4 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad \boldsymbol{\gamma}_{\psi} = \begin{pmatrix} * & 1 & 1 & 1 & 1 & 1 \\ * & * & 0 & 0 & 0 & 0 \\ * & * & * & 0 & 0 & 0 \\ * & * & * & * & 0 & 0 \\ * & * & * & * & * & 0 \\ * & * & * & * & * & * \end{pmatrix}, \tag{2.91}$$

$$\boldsymbol{B} = \begin{pmatrix} 0.60 & 0.50 & 0.40 & 0.00 & 0.40 & 0.00 & 0.00 & 0.00 \\ 0.60 & 0.00 & 0.00 & 0.40 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.60 & 0.00 & 0.00 & 0.00 & 0.40 & 0.00 & 0.00 & 0.00 \\ 0.30 & 0.00 & 0.00 & 0.00 & 0.00 & 0.40 & 0.00 & 0.00 \\ 0.30 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.40 & 0.00 \\ 0.30 & 0.50 & 0.00 & 0.00 & 0.40 & 0.00 & 0.00 & 0.40 \end{pmatrix}, \tag{2.92}$$

$$\boldsymbol{\gamma}_{b} = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 \end{pmatrix}. \tag{2.93}$$

The data generating process for the stochastic volatility is given by:

$$\boldsymbol{a}_0 = (.2)\mathbf{1}_6, \;\; \boldsymbol{\beta} = (.2)\mathbf{1}_6, , \;\; \boldsymbol{A}_1 = (0)\mathbf{1}_6, \;\; \boldsymbol{\delta} = (.1)\mathbf{1}_6 \tag{2.94}$$

This model is estimated using the same number of MCMC cycles and the same prior

hyperparameters as in Example 2.1. Estimates obtained are as follows:

$$\widehat{\boldsymbol{\Psi}} = \begin{pmatrix} 1 & .4972 & .4996 & .5954 & .6014 & .4028 \\ 0 & 1 & -.0023 & .0023 & .0031 & -.0023 \\ 0 & 0 & 1 & -.0005 & .0009 & -.0003 \\ 0 & 0 & 0 & 1 & .0043 & .0026 \\ 0 & 0 & 0 & 0 & 1 & -.0003 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix},$$

$$\widehat{\boldsymbol{\gamma}}_{\psi} = \begin{pmatrix} * & .9984 & .9974 & 1.0000 & .9999 & .9161 \\ * & * & .0218 & .0213 & .0213 & .0215 \\ * & * & * & .0216 & .0212 & .0214 \\ * & * & * & * & .0216 & .0217 \\ * & * & * & * & * & .0213 \\ * & * & * & * & * & * \end{pmatrix},$$

$$\widehat{\boldsymbol{B}} = \begin{pmatrix} .6321 & .5248 & .3782 & .0052 & .3856 & .0033 & .0105 & .0084 \\ .6182 & -.0117 & .0017 & .3656 & .0107 & -.0045 & -.0043 & .0011 \\ .6131 & -.0121 & .0050 & .0017 & .3723 & .0022 & -.0014 & .0008 \\ .2950 & -.0070 & .0050 & -.0046 & .0142 & .3678 & -.0050 & -.0016 \\ .3121 & -.0060 & -.0052 & -.0002 & .0123 & -.0042 & .3583 & .0041 \\ .3181 & .4005 & .0125 & .0039 & .3834 & .0049 & .0028 & .3721 \end{pmatrix},$$

$$\widehat{\boldsymbol{\gamma}}_{b} = \begin{pmatrix} .9979 & .9243 & .0211 & .9452 & .0207 & .0212 & .0209 \\ .0268 & .0226 & .8934 & .0217 & .0215 & .0211 & .0211 \\ .0256 & .0221 & .0218 & .9077 & .0215 & .0217 & .0213 \\ .0271 & .0221 & .0215 & .0226 & .9002 & .0217 & .0213 \\ .0261 & .0219 & .0219 & .0218 & .0218 & .8683 & .0211 \\ .9259 & .0220 & .0215 & .9374 & .0211 & .0219 & .9194 \end{pmatrix},$$

$$
\begin{aligned}
\widehat{\boldsymbol{a}}_0 &= (.2056,\ .2054,\ .2046,\ .2057,\ .2048,\ .2046)', \\
\widehat{\boldsymbol{\beta}} &= (.1755,\ .1790,\ .1790,\ .1787,\ .1813,\ .1801)', \\
\widehat{\boldsymbol{A}}_1 &= (.0215,\ -.0460,\ .0065,\ -.0322,\ .0276)', \\
\widehat{\boldsymbol{\delta}} &= (.0956,\ .0951,\ .0956,\ .0956,\ .0953,\ .0955)'.
\end{aligned}
$$

We summarize here the major differences between the results of this example and the previouses. The $\boldsymbol{\Psi}$ setup in this example and that of Example 2.3 differ only for $\psi_{1,6}$, which is .4 here and .3 in the previous case: $\psi_{1,6}$ has been selected 47 percent of the time in Example 2.2, and 92 percent of the time here. Selection of the elements of $\boldsymbol{B}_1$ is quite successful (almost comparable with the results on Example 2.1), since non-zero elements have been selected on average relative frequency of 91 percent, while zero elements with one of 2 percent. In the stochastic volatility generation process, $\boldsymbol{a}_0$, $\boldsymbol{A}_1$ are quite accurately estimated, while $\boldsymbol{\delta}$ and $\boldsymbol{\beta}$ are generally underestimated, with errors on the order of 5 and 11 percent of the true value, respectively.

Example 2.4. We are proposing a fourth example, with (p, T, $\boldsymbol{X}$, $\boldsymbol{z}$), as in Examples 2.1−2.3. Here $\boldsymbol{b}_0 = \boldsymbol{0}$, while $\boldsymbol{\Psi}$ and $\boldsymbol{\gamma}_\psi$ are as in (2.91), and the parameters of the stochastic volatility generating process are the same as in (2.94). The other parameters are:

$$
\boldsymbol{B} = \begin{pmatrix}
0.00 & 0.50 & 0.40 & 0.00 & 0.40 & 0.00 & 0.00 & 0.00 \\
0.00 & 1.00 & 0.00 & 0.40 & 0.00 & 0.00 & 0.00 & 0.00 \\
0.00 & 1.00 & 0.00 & 0.00 & 0.40 & 0.00 & 0.00 & 0.00 \\
0.00 & 1.00 & 0.00 & 0.00 & 0.00 & 0.40 & 0.00 & 0.00 \\
0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.00 & 0.40 & 0.00 \\
0.00 & 0.50 & 0.00 & 0.00 & 0.40 & 0.00 & 0.00 & 0.40
\end{pmatrix}, \quad (2.95)
$$

$$\boldsymbol{\gamma}_b = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 \end{pmatrix}. \tag{2.96}$$

Using the same number of MCMC cycles and the same prior hyperparameters as in Example 2.1, we get the following estimates:

$$\widehat{\boldsymbol{\Psi}} = \begin{pmatrix} 1 & .5164 & .5134 & .6035 & .6010 & .4033 \\ 0 & 1 & -.0130 & -.0081 & .0052 & -.0073 \\ 0 & 0 & 1 & -.0109 & .0031 & -.0047 \\ 0 & 0 & 0 & 1 & .0062 & -.0022 \\ 0 & 0 & 0 & 0 & 1 & .0005 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix},$$

$$\widehat{\boldsymbol{\gamma}}_\psi = \begin{pmatrix} * & .9993 & .9984 & 1.0000 & .9999 & .9172 \\ * & * & .0220 & .0215 & .0213 & .0216 \\ * & * & * & .0218 & .0212 & .0214 \\ * & * & * & * & .0216 & .0217 \\ * & * & * & * & * & .0213 \\ * & * & * & * & * & * \end{pmatrix},$$

$$\widehat{\boldsymbol{B}} = \begin{pmatrix} .0009 & .7800 & .3845 & .0076 & .3861 & .0058 & .0111 & .0102 \\ -.0026 & .6781 & .0040 & .3690 & .0147 & -.0016 & -.0049 & -.0015 \\ -.0001 & .6765 & .0071 & .0027 & .3745 & .0034 & -.0013 & .0001 \\ -.0040 & .6628 & .0058 & -.0017 & .0187 & .3707 & -.0061 & -.0049 \\ -.0019 & -.1242 & -.0073 & -.0005 & .0128 & -.0044 & .3580 & .0029 \\ -.0017 & .3142 & .0148 & .0054 & .3846 & .0061 & .0028 & .3720 \end{pmatrix},$$

$$\widehat{\gamma}_b \;=\; \begin{pmatrix} 1.00 & .9435 & .0210 & .9471 & .0205 & .0212 & .0210 \\ 1.00 & .0219 & .9056 & .0215 & .0213 & .0213 & .0211 \\ 1.00 & .0216 & .0213 & .9171 & .0213 & .0218 & .0212 \\ 1.00 & .0217 & .0212 & .0225 & .9127 & .0218 & .0215 \\ .0760 & .0214 & .0216 & .0216 & .0215 & .8681 & .0213 \\ .6645 & .0215 & .0212 & .9415 & .0209 & .0219 & .9184 \end{pmatrix}.$$

This example is very close to Example 2.3. The only differences lie in the first two columns of the matrix $\boldsymbol{B}$. Estimates of the stochastic volatility generating process parameters are not reported since they are essentially the same as in Example 2.3. The elements of $\boldsymbol{B}$ relative to the autoregressive structure ($\boldsymbol{b}_2$ and subsequent) have been chosen selected with an accuracy similar to that in Example 2.3 (these columns share exactly the same structure in the two examples). $\boldsymbol{b}_1$ has only one zero element, which has been selected 7 percent of the time. All the non-zero elements of $\boldsymbol{b}_1$ have been always selected to be in the model, except the last one, which has been selected only 66 percent of the time.

2.6.2 Comparison Between Gilks'Adaptive Sampling and Particle Filtering

In this section we perform a simulation study to compare the methods adopted in this study for sampling the stochastic volatilities. We experimented with both filtering techniques presented in Algorithms F1 and F2. The auxiliary sampling-resampling filter provided results very similar to the simpler particle filter, and it required longer computing time. Furthermore, we did not experience weight degeneration problems, and so the choice to use the simpler filtering method has been made. The rejection sampler showed

acceptance rates around 80 percent, and running times higher than Algorithms 1 and 2. From Table 2.2, notice that estimation using the Gilks' adaptive sampler (Algorithm 1) is faster than estimation using filtering (Algorithms 2 and 3)[2], even for a moderate number of particles. All the simulations are obtained on a Pentium 4, 3.20 Ghz Intel processor with 1 GByte of Ram.

Table 2.2: Running Times for Algorithms 1, 2 and 3 ($T = 1,000$, $M = 100$)

	$p = 4$			$p = 10$		
draws	Alg 1	Alg 2	Alg 3	Alg 1	Alg 2	Alg 3
1,100	2'48"	5'22"	6'29"	9'51"	16'12"	19'24"
11,000	21'36"	52'53"	1h2'13"	1h40'44"	2h44'33"	3h20'14"
50,000	1h42'46"	4h05'36"	4h48'18"	7h51'03"	12h49'21"	15h21'42"

The ratio between the running times of Algorithms 1 and 2 is not constant: differences become smaller as the model becomes more complex. This might be due to the fact that the particle filter samples vectorwise, while the Gilks' sampler operates elementwise.

Comparing algorithms just through computing times might be misleading. If an MCMC algorithm produces very correlated draws, it might require much longer chains to effectively explore the posterior domain, thus resulting in a larger total cpu time in spite of a smaller time over iteration ratio. We measure the correlation among MCMC outcomes through an inefficiency factor, as in Chib et al. (1998) and (2005). These factors can be estimated as the ratio between the variance of the sample mean obtained using the actual MCMC scheme and the one obtained with an hypothetical independent sampling scheme. A more inefficient algorithm will require longer MCMC chains to well

[2]If auxiliary particle filter is used for Algorithm 2, running times result about 5 times larger.

explore the posterior sample space. We can evaluate this ratio as

$$\widehat{R} \;=\; 1+\frac{2Ba}{Ba-1}\sum_{n=1}^{Ba}K\Big(\frac{n}{Ba}\Big)\widehat{\rho}(n), \tag{2.97}$$

where Ba is known as the bandwidth, $K(.)$ is known as the Parzen kernel and is defined as

$$K(x)=\begin{cases}1-6x^2-6x^3, & \text{if } x\in\left[0,\frac{1}{2}\right]\\ 2(1-x)^3, & \text{if } x\in\left[\frac{1}{2},1\right]\\ 0, & \text{otherwise.}\end{cases}$$

Also,

$$\widehat{\rho}(n) \;=\; \frac{\widehat{\Gamma(n)}}{\widehat{\Gamma(0)}},$$

where

$$\widehat{\Gamma(n)} \;=\; \frac{1}{N}\sum_{i=n+1}^{N}\Big(z^{(i)}-\overline{z}\Big)\Big(z^{(i-n)}-\overline{z}\Big),$$

where N is the number of MCMC iterations, $z^{(i)}$ is the outcome for the parameter of interest at iteration i and $\overline{z}$ is the average over the N iterations. Inefficiency factors for the element of $\boldsymbol{B}$ are summarized in Table 2.3. The models are parameterized as in Example 2.1. It is clear that the Gilks' adaptive sampler leads to much more correlated draws for $\boldsymbol{B}$, and thus necessitates longer MCMC chains.

Table 2.3: Inefficiency Factors for Algorithms 1, 2 and 3 ($T=1,000$, $M=100$)

	$p=4$			$p=10$		
Ba=	Alg1	Alg2	Alg3	Alg1	Alg2	Alg3
200	80.3698	1.0673	0.9416	52.5249	0.8907	0.8948
2000	802.7149	0.9468	0.8112	518.9325	1.0292	1.0356

Possible consequences are shown in the next example. Algorithms 1 and 2 are employed with starting values which are far away from the true parameters. The two algorithm show similar results only for chains long enough, otherwise Algorithm 1 presents considerable estimation errors. This minimum length depends on the model. For a model with four variables, $T = 1,000$ and a relatively simple structure of $\mathbf{\Psi}$, we found $10,000$ MCMC cycles after $1,000$ burn-ins to be sufficient for the two algorithms to produce similar estimates, while $1,000$ MCMC cycles after 100 burn-ins not to (see next example). For a model with $p = 10$, $T = 1,000$ and $M = 100$, $10,000$ cycles with $1,000$ burn-ins are still not enough for Algorithm 1 to produce estimates comparable to those obtained with Algorithm 2 (see Figure 2.1).

Example 2.5. Consider a model with the following $\mathbf{\Psi}$ and $\boldsymbol{B}$ parameters:

$$\mathbf{\Psi} = \begin{pmatrix} 1 & .5 & .5 & .6 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \boldsymbol{B} = \begin{pmatrix} 0 & .5 & .4 & 0 & .4 & 0 \\ 0 & 1 & .5 & .4 & 0 & .4 \\ 0 & 1 & 0 & .5 & .4 & 0 \\ 0 & 1 & 0 & 0 & 0 & .4 \end{pmatrix}.$$

We perform Bayesian estimation of the previous model using both Algorithms 1 and 2, and starting the MCMC simulation from "bad" starting values. Using $1,000$ MCMC with 100 burn-ins we obtain the following estimates:

$$\widehat{\mathbf{\Psi}}_{Alg-1} = \begin{pmatrix} 1 & .1714 & .8133 & .4115 \\ 0 & 1 & -.0136 & -.0274 \\ 0 & 0 & 1 & -.0836 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \widehat{\mathbf{\Psi}}_{Alg-2} = \begin{pmatrix} 1 & .4431 & .5145 & .6104 \\ 0 & 1 & .0090 & .0159 \\ 0 & 0 & 1 & .0175 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

$$\widehat{\boldsymbol{B}}_{Alg-1} = \begin{pmatrix} .2184 & .6117 & .3005 & -.0145 & .4488 & .0118 \\ .1079 & .7602 & .6144 & .3847 & -.1567 & .5384 \\ -.0929 & .6399 & -.1354 & .6737 & .3373 & -.0036 \\ .0455 & .2377 & -.0361 & -.0359 & -.0508 & .3413 \end{pmatrix},$$

$$\widehat{\boldsymbol{B}}_{Alg-2} = \begin{pmatrix} -.0335 & .7003 & .4066 & -.0019 & .3956 & .0349 \\ .1011 & .9739 & .4596 & .3796 & .0010 & .3430 \\ -.0028 & .6923 & -.0168 & .4852 & .4046 & -.0406 \\ .1114 & .6506 & -.0203 & -.0144 & .0368 & .3200 \end{pmatrix}.$$

It is quite evident that Algorithm 1 underestimates ψ_{12} and overestimate ψ_{13}. It also performs poorly in estimating the first 2 columns of $\boldsymbol{B}$. Algorithm 2 generally performs much better.

Using $10,000$ MCMC with $1,000$ burn-ins instead, we obtain estimates from the two algorithm which are very close and generally satisfactory.

$$\widehat{\boldsymbol{\Psi}}_{Alg-1} = \begin{pmatrix} 1 & .4703 & .5293 & .5994 \\ 0 & 1 & .0205 & .0152 \\ 0 & 0 & 1 & .0241 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \widehat{\boldsymbol{\Psi}}_{Alg-2} = \begin{pmatrix} 1 & .4617 & .5200 & .6014 \\ 0 & 1 & .0230 & .0142 \\ 0 & 0 & 1 & .0251 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

$$\widehat{\boldsymbol{B}}_{Alg-1} = \begin{pmatrix} -.0311 & .5471 & .4185 & -.0110 & .4054 & .0241 \\ .0949 & 1.0031 & .4585 & .3956 & -.0128 & .3649 \\ -.0094 & .9880 & -.0270 & .4838 & .4145 & -.0458 \\ .1046 & .7216 & -.0257 & -.0114 & .0276 & .3271 \end{pmatrix},$$

$$\widehat{\boldsymbol{B}}_{Alg-2} = \begin{pmatrix} -.0328 & .5476 & .4196 & -.0105 & .4057 & .0241 \\ .0963 & 1.0013 & .4596 & .3966 & -.0142 & .3661 \\ -.0072 & .9796 & -.0256 & .4822 & .4145 & -.0431 \\ .1075 & .7212 & -.0282 & -.0101 & .0301 & .3249 \end{pmatrix}.$$

2.6.3 Model Selection via Posterior Mode

In the previous examples, we used the posterior mean of the indicators as an instrument for model selection. Here, we use the posterior model of $\boldsymbol{\gamma}$ instead. Consider the model of Example 2.5. For $T = 1,000$, $10,000$ MCMC cycles after $1,000$

burn-ins, the most visited model (41 percent of the time) is the one with decimal marker 16161687205. This number, once reconverted to binary, corresponds to $\widehat{\boldsymbol{\gamma}} = 11110000110100111111000110101001 01$, which happens to be the true $\boldsymbol{\gamma}$.

If we estimate $\boldsymbol{\gamma}$ using the posterior mean, we get

$$\begin{aligned}\widehat{\boldsymbol{\gamma}} = (&.9875, .9897, .9925, .9925, .0094, .0213, .0197, .0095, .9977, .9999, .0224, 1,\\ &.0216, .0202, 1, 1, 1, 1, .9753, .9987, .0224, .0198, .0199, .9533, .9993, .0206,\\ &.9726, .0197, .9805, .0206, .0236, .8637, .0249, .7470),\end{aligned}$$

which would lead to the choice of the very same model using any cutting point lower than .74. Here the number of parameters subject to stochastic search is 34, for a total number of 2^{34} competing models. The algorithm visited a total of 494 different models. Results depend on the number of observations available. For $T = 500$ results are almost identical, with the best model visited 41 of the time and a total number of 587 visited models. When $T = 300$, the best model is visited with a relative frequency of 7.9 percent only, and the number of visited models is $1,678$. If we reduce the number of observations to 100, we see that the most visited model is visited only 0.9 percent of the time, and does not correspond to the true one, which has been visited only 0.016 percent of the time. The total number of visited models, in this case, is 4740, meaning that the algorithm selects the models almost randomly, and so that $T = 100$ is not a sufficient sample size for this type of variable selection. For $T = 100$, if we estimate $\boldsymbol{\gamma}$ using the posterior mean we get

$$\begin{aligned}\widehat{\boldsymbol{\gamma}} = (&.9032, .8817, .8015, .9112, .0526, .0771, .0929, .0122, .4237, .9992, .0302,\\ &.6014, .0667, .1168, .9233, .9685, .3947, .4949, .3673, .9768, .0619, .0265,\\ &.0435, .6678, .9748, .0378, .9466, .0356, .6988, .0251, .0445, .7974, .0263, .646).\end{aligned}$$

This leads to the choice of the true model for any cutting point between 0.12 and 0.36 (which are not natural choices). The major difference with the estimate obtained with $T = 1,000$ is that the parameters in the true model (some of them at least) have been chosen less frequently, while the parameters not in the model continue to have very low posterior mean. In other words, for smaller sample sizes, the algorithm tends to choose simpler models. A barplot of the frequencies models have been visited can be found in Figure 2.2. The plot shows that, with $T = 1,000$ the most visited subset has much an higher frequency than the second most visited one, this making the choice of the best model very easy. For $T = 100$ though, the choice is not easy, and we saw that it would lead to a wrong model. So, when the number of observations is limited, the posterior mean might be preferable: the researcher can consider some parameters to be almost surely in or out of the model (those with an estimate very close to zero or one), and choose the other subjectively. The computation of the posterior mean can also be used for preliminary runs, in order to later "fix" some parameter and run the selection on a reduced space.

We repeated the experiment simulating from the same model of Example 2.1, with $p = 6$ and using $10,000$ MCMC cycles after $1,000$ burn-ins. The number competing models is larger, but the model structure relatively simpler, many parameters being equal to zero. Barplots of the frequencies are reported in Figure 2.3, while Table 2.4 shows a summary of the results.

2.6.4 Model Selection for $\boldsymbol{\delta}$

In Examples 2.1−2.5 we did not select the elements of $\boldsymbol{\delta}$. Unlike model selection for other parameters in the model, the selection of volatility variances is sensitive to the

Table 2.4: Summary of Posterior Modes with p=6. The Total Number of Visits is $10,000$

T	number of visited models	freq. most visited	freq. true
$T = 100$	8516	35	2
$T = 300$	4835	245	245
$T = 500$	3265	1274	1274
$T = 1000$	2698	1924	1924

choice of prior hyperparameters. In Table 2.5 we present simulation results obtained using a model with $p = 10$ and the same parameter structure as in Example 2.1, except for $\boldsymbol{\delta}$. The posterior mean of $\boldsymbol{\gamma}_\delta$ varies significantly depending on $\boldsymbol{q}$: big values of q_j

Table 2.5: Model Selection of $\boldsymbol{\delta}$, When $T = 1,000$, and $p = 10$

$\boldsymbol{\delta}$	.0000	.0200	.0400	.0000	.0600	.0800	.1000	.1200	.1500	.2000
$\boldsymbol{\gamma}_\delta$	0	1	1	0	1	1	1	1	1	1
$\widehat{\boldsymbol{\gamma}}_\delta, q$=1.1	.0015	.5689	.5585	.0035	.5904	.5273	.4897	.5868	.5987	.5476
$\widehat{\boldsymbol{\gamma}}_\delta, q$=1.5	.0012	.4795	.6265	.0024	.7186	.6107	.6562	.7094	.6982	.6883
$\widehat{\boldsymbol{\gamma}}_\delta, q$=3	.0007	.0856	.7086	.0014	.9087	.9101	.8678	.9154	.9107	.8655
$\widehat{\boldsymbol{\gamma}}_\delta, q$=5	.0003	.0005	.6464	.0003	.9202	.9577	.9203	.9498	.9409	.9415
$\widehat{\boldsymbol{\gamma}}_\delta, q$=10	.0000	.0001	.2718	.0000	.9198	.9764	.9499	.9989	.9856	.9745
$\widehat{\boldsymbol{\gamma}}_\delta, q$=20	.0000	.0000	.0000	.0000	.7702	.9654	.9719	.9956	.9972	.9902
$\widehat{\boldsymbol{\gamma}}_\delta, q$=50	.0000	.0000	.0000	.0000	.0000	.4815	.9177	.9985	.9997	.9967
$\widehat{\boldsymbol{\gamma}}_\delta, q$=100	.0000	.0000	.0000	.0000	.0000	.0000	.1703	.9400	.9879	.9996
$\widehat{\boldsymbol{\gamma}}_\delta, q$=200	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0623	.8886	.9905

lead to an unreasonable number of zeros in $\gamma_{\delta j}$, while small values of q_j lead to posterior means for the $\gamma_{\delta j}$s which are never close to one. Fixing q_j accordingly to the magnitude of δ_j is impossible since δ_j is unknown. Instead, we specify a prior and estimate its

posterior distribution. The prior specification and conditional posterior computation are discussed in Sections 2.4 and 2.5, respectively, while simulation results are in Table 2.6. These are obtained with the same model that produced the results in Table 2.5. It is clear that this hierarchical formulation delivers better results than fixing $\boldsymbol{q}$ at a

Table 2.6: Model Selection of $\boldsymbol{\delta}$ Under Prior on $\boldsymbol{q}$

$\boldsymbol{\delta}$	.0000	.0200	.0400	.0000	.0600	.0800	.1000	.1200	.1500	.2000
$\boldsymbol{\gamma}_\delta$	0	1	1	0	1	1	1	1	1	1
$\widehat{\gamma}_\delta$	.0000	.9723	.9911	.0000	.937	.9941	.9941	.9944	.9943	.9947
$\widehat{q}_j$	3.971	6.711	8.704	3.999	9.486	9.713	9.655	9.845	9.853	9.969

constant value.

2.6.5 Examples Using Algorithm 2 and Selecting $\boldsymbol{\delta}$

The higher efficiency of Algorithm 2 is more advantegeous for larger models, as we saw previously in the chapter. We present an example with 10 variables, performing stochastic search on $\boldsymbol{\delta}$ as well. Because of the high number of parameters, we present the results for $\boldsymbol{B}$ and $\boldsymbol{\Psi}$ in graphical form (Figure 2.1).

Example 2.6. We study a VAR(1) with $p = 10$ and $T = 1,000$. The model presents stochastic volatilities and two exogenous variables, $\boldsymbol{x}$ and $\boldsymbol{z}$, generated as in Example 2.1. The first two columns of $\boldsymbol{B}$ have all the elements different than zero. The matrix obtained from the remaining columns $(\boldsymbol{b}_2, \cdots, \boldsymbol{b}_q)$ is relative to the AR coefficients of the model, and has all the diagonal elements different from zero, as well as some off diagonal ones. Also, all the non-redundant first row elements of $\boldsymbol{\Psi}$ are different than zero with values varying from .3 to .6, as well as some sparse non-redundant elements of

the subsequent rows. The true parameters of the stochastic volatility generating process are shown below.

$$\boldsymbol{a}_0 = (.2)\mathbf{1}_{10},\ \boldsymbol{\beta} = ((.7)\mathbf{1}_6',\ .2,\ .2,\ .5,\ .5)',\ \boldsymbol{A}_1 = (0,\ 0,\ .2,\ 0,\ 0,\ (.2)\mathbf{1}_4')',$$

$$\boldsymbol{\delta} = (0,\ .02,\ .04,\ 0,\ .06,\ .08,\ .1,\ 0,\ .15,\ .2)',$$

$$\boldsymbol{\gamma}_a = (0,\ 0,\ 1,\ 0,\ 0,\ \mathbf{1}_4')',\ \boldsymbol{\gamma}_\delta = (0,\ 1,\ 1,\ 0,\ \mathbf{1}_3',\ 0,\ 1,\ 1)'.$$

This model is quite more complex than those previously discussed.

The hyperparameters of the Bernoulli priors are always $p = .5$. The prior on the intercepts is normal with zero mean and a large variance of 50. We let $v_{j0} = 6.0$ and $s_{j0}^2 = 0.001$. We set the prior means α_{j0} and β_{j0} at 0 and standard deviation σ_α and σ_β at 10. We take $\kappa_{ij} = 0.1$, $d_{ij} = 50$ and $M = 100$. For $\boldsymbol{\delta}$ selection, we take $v_{jo} = 3$ and $s_{j0} = .001$ for all js, while we estimate $\boldsymbol{q}$, having $\alpha_q = 2$ and $\beta_q = 0.5$. Here we present the results obtained sampling 10 different samples from the model, and for each computing the posterior mean of the parameters using $10,000$ cycles after $1,000$ burn-ins.

$$\begin{aligned}
\widehat{\boldsymbol{a}}_0 &= (.2038,\ .2008,\ .2121,\ .2000,\ .2030,\ .2042,\ .2013,\ .2012,\ .2042,\ .2027)',\\
\widehat{\boldsymbol{\beta}} &= (.6943,\ .6826,\ .7000,\ .6933,\ .6926,\ .1974,\ .1954,\ .4823,\ .4966)',\\
\widehat{\boldsymbol{A}}_1 &= (.0000,\ .0017,\ .1895,\ .0000,\ .0054,\ .1898,\ -.0043,\ .0000,\ -.0.25,\ -.0188)',\\
\widehat{\boldsymbol{\delta}} &= (.0000,\ .0202,\ .0402,\ .0000,\ .0598,\ .0809,\ .1003,\ .0000,\ .1499,\ .1922)',\\
\widehat{\boldsymbol{\gamma}}_a &= (.0185,\ .0224,\ .9589,\ .0208,\ .0705,\ .9920,\ .0182,\ .0221,\ .0226,\ .0209)',\\
\widehat{\boldsymbol{\gamma}}_\delta &= (.0000,\ .9708,\ .9911,\ .0000,\ .9930,\ .9940,\ .9942,\ .0000,\ .9946,\ .9944)',\\
\widehat{\boldsymbol{q}} &= (3.971,\ 6.711,\ 8.704,\ 3.999,\ 9.486,\ 9.713,\ 9.655,\ 9.845,\ 9.853,\ 9.969)'.
\end{aligned}$$

Figure 2.4 presents the differences between the true and estimated values of $\boldsymbol{B}$ and $\boldsymbol{\Psi}$ standardized over the range of the true values. Table 2.7 reports the average estimates

for the $\gamma_{b,j}$ and $\gamma_{\psi,j}$ with a true value of 0 and 1, respectively.

Table 2.7: Estimates of the Indicator Variables

	$Average\, estimate\, when\, true = 0$	$Average\, estimate\, when\, true = 1$
$\boldsymbol{B}$	0.02563556	0.93872
$\boldsymbol{\Psi}$	0.02147241	0.801025

Finally, Table 2.8 shows the average true and estimated values of h_{jt}. Some param-

Table 2.8: Comparison of True and Estimated Stochastic Volatilities for One Randomly Chosen Sample. $T = 1,000$, $p = 10$.

$variable$	$mean\,(true\, h)$	$posterior\, mean\,(h)$	$S.d.\,(true\, h)$	$posterior\, s.d\,(h)$
1	0.6632	0.6609	0.0707	0.0738
2	0.6642	0.6642	0.0763	0.0738
3	0.6647	0.6661	0.0763	0.0738
4	0.6550	0.6545	0.0704	0.0706
5	0.6636	0.6625	0.0741	0.0768
6	0.6677	0.6694	0.0734	0.0754
7	0.2522	0.2524	0.0489	0.0507
8	0.2492	0.2492	0.0516	0.0510
9	0.4009	0.4004	0.0574	0.0582
10	0.3982	0.3988	0.0597	0.0596

eters of $\boldsymbol{B}$ are still quite far from the true value, but generally the matrix is accurately estimated. The estimates of $\boldsymbol{a}_0$ and $\boldsymbol{\beta}$ look respectively positively and negatively biased, but the dimension of this bias is to be considered paltry, since the largest error is on the order of 2 percent of the true value for $\boldsymbol{\beta}$ and 6 percent for $\boldsymbol{a}_0$. It is interesting to notice that here estimation errors for $\boldsymbol{\beta}$ are smaller than estimation errors for $\boldsymbol{a}_0$, while

it has always been the opposite in the previous examples. The estimation of $\boldsymbol{\delta}$ and the relative model selection gives almost perfect results.

2.6.6 Forecasting

So far we discussed about the advantage of the variable selection technique in terms of the ease of interpreting a simpler model. In Section 2.3, we anticipated about the possible advantages of a calibrated model over an overparameterized unrestricted model in terms of prediction and forecasting as well. The unrestricted model might present problems of over-fitting which adversely affect model forecasts. Variable selection might be able to overcome this problems.

In this section, we compare the forecasting ability between an unrestricted VAR model with stochastic volatilities and the correspondent restricted model obtained with our model selection algorithm.

Consider the simulated data of Example 2.6: we obtain Bayesian estimates of both the restricted and unrestricted model. To obtain Bayesian estimates of the latter we simply fix all the model indexes to be one. Restricted model is chosen using the posterior mean of the model indicators and fixing at .2 the threshold for the correspondent parameters to be selected. We then use these estimates to make a forecast of $h_{j,T+1}$ and $y_{j,T+1}$, for $j = 1, \cdots, 10$. We compute the mean square prediction error for the true, restricted and unrestricted models over 50 simulated samples of $T = 1,000$. We define MSE_{model} to be the mean of the the squared errors $(y_{j,T+1} - \widehat{y}_{j,T+1}^{(model)})$, and MSE^*_{model} to be the mean of the the squared errors $(\widehat{y}_{j,T+1}^{(true)} - \widehat{y}_{j,T+1}^{(model)})$. Here are the results for the forecast of $y_{j,T+1}$:

$$\frac{MSE_{res}}{MSE_{true}} = (1.056,\ 1.096,\ 1.379,\ 1.121,\ 1.011,\ 1.748,\ 1.327,\ 1.012,\ 1.048,\ 1.152),$$

$$\frac{MSE^*_{unres}}{MSE^*_{res}} = (2.489,\ 1.181,\ 2.607,\ 1.755,\ 2.179,\ 1.682,\ 2.984,\ 1.830,\ 2.012,\ 2.146).$$

The restricted model provides good forecasts. The highest ratio between true and restricted models' MSE is 1.74 (variable 6). The unrestricted model performs well for some variables (e.g., variable 2), but significantly worse than the restricted one for others (e.g., variable 7). Here are the results for the forecast of $h_{j,T+1}$:

$$\frac{MSE_{res}}{MSE_{true}} = (1.010,\ 1.828,\ 3.614,\ 1.02,\ 1.249,\ 3.553,\ 3.415,\ 1.012,\ 1.421,\ 1.667),$$
$$\frac{MSE^*_{unres}}{MSE^*_{res}} = (1.286,\ 5.10,\ 5.184,\ 8.591,\ 13.941,\ 1.699,\ 4.510,\ 1.175,\ 1.34,\ 7.35).$$

The restricted model works best for the variables with $\delta_j = 0$. Variables 3 and 5 present the worst results, while variables 9 and 10 are the best among those with $\delta_j > 0$. In general, the restricted model selected by stochastic search algorithm provides better forecasting than the correspondent unrestricted model, for both $\boldsymbol{y}$ and $\boldsymbol{h}$.

2.6.7 Numerical Simulations with Smaller Sample Sizes

Previous examples utilize a sample size $T = 1,000$. Although this is a reasonably small sample size for most financial econometrics applications, many macroeconomic models must be estimated with quite fewer observations. The numerical examples in this subsection illustrate the performance of the stochastic search algorithm with samples sizes $T < 1,000$. For each example, we simulate one thousand samples and conduct stochastic search model selection with $10,000$ MCMC cycles of Algorithm 1 (after $1,000$ burn-in runs).

Example 2.7. Consider a three-variable model with the following parameters,

$$\boldsymbol{\Psi} = \begin{pmatrix} 1 & .5 & .5 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \boldsymbol{\gamma}_\psi = \begin{pmatrix} * & 1 & 1 \\ * & * & 0 \\ * & * & * \end{pmatrix}, \tag{2.98}$$

$$\boldsymbol{B} = \begin{pmatrix} 0 & .3 & .5 & .2 & .2 \\ 0 & 0 & 0 & .5 & 0 \\ 0 & 0 & 0 & 0 & .5 \end{pmatrix}, \quad \boldsymbol{\gamma}_b = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{2.99}$$

The data generating process for the stochastic volatility is given by:

$$\boldsymbol{A}_1 = (.1)\mathbf{1}_3, \;\; \boldsymbol{a}_0 = (.2)\mathbf{1}_3 \;\; \boldsymbol{\beta} = (.5)\mathbf{1}_3, \;\; \boldsymbol{\delta} = (.1)\mathbf{1}_3. \tag{2.100}$$

The sample size is $T = 300$. The hyperparameters of the Bernoulli priors are all 0.5. The prior on the intercepts is normal with zero mean variance 25. We let $v_{j0} = 1.0$, $s_{j0}^2 = 0.001$ (meaning the prior for σ_j is quite diffuse). α_{j0} and β_{j0} are 0, while σ_α and σ_β are 10. Finally, $\kappa_{ij} = 0.1$ and $d_{ij} = 50$. Here we present the estimate obtained for this model:

$$\widehat{\boldsymbol{B}} = \begin{pmatrix} .0009 & .2610 & .4667 & .2345 & .1864 \\ .0022 & -.0008 & -.0014 & .4840 & -.0060 \\ .0030 & -.0011 & .0013 & -.0149 & .4777 \end{pmatrix}, \quad \widehat{\boldsymbol{\Psi}} = \begin{pmatrix} 1 & .4830 & .5384 \\ 0 & 1 & .0548 \\ 0 & 0 & 1 \end{pmatrix},$$

$$\widehat{\boldsymbol{\gamma}}_b = \begin{pmatrix} .9667 & .9889 & .9313 & .9156 \\ .0444 & .0223 & .9667 & .0556 \\ .0347 & .0112 & .0123 & 1.000 \end{pmatrix}, \quad \widehat{\boldsymbol{\gamma}}_\psi = \begin{pmatrix} * & .9307 & .9920 \\ * & * & .0791 \\ * & * & * \end{pmatrix}.$$

Estimates of the latent volatilities generation process parameters are shown in Table 2.9, while plots with the true and estimated log-volatilities are shown in Figure 2.5.

The model presented here is quite simple. It can be seen that the results are good but not perfect. The reason might be found in the small sample size. In particular, while

Table 2.9: Posterior Mean for Stochastic Volatility generation process parameters

$\widehat{\boldsymbol{A}}_1$	.1201	.1119	.0897
$\widehat{\boldsymbol{a}}_0$	.2188	.2404	.2231
$\widehat{\boldsymbol{\beta}}$	.4614	.5587	.5337
$\widehat{\boldsymbol{\delta}}$	.0961	.1240	.0882

the estimates of the parameters in the main equation and of the correspondent indicator variables have errors comparable or even smaller than in the previous examples, the parameters in the stochastic volatility equation show errors up to 20 percent of the true parameter value.

Example 2.8. In this example we present a more complex model, estimated using a larger sample size. Consider a four-variable model with the following parameters,

$$\boldsymbol{\Psi} = \begin{pmatrix} 1 & .5 & .5 & .2 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \boldsymbol{\gamma}_\psi = \begin{pmatrix} * & 1 & 1 & 1 \\ * & * & 0 & 0 \\ * & * & * & 0 \\ * & * & * & * \end{pmatrix}, \tag{2.101}$$

$$\boldsymbol{B} = \begin{pmatrix} 0 & .3 & .5 & .2 & .2 & .4 \\ 0 & 0 & 0 & .5 & 0 & 0 \\ 0 & .1 & 0 & 0 & .5 & 0 \\ 0 & 0 & 0 & 0 & 0 & .5 \end{pmatrix}, \ \boldsymbol{\gamma}_b = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \tag{2.102}$$

The data generating process for the stochastic volatility is given by:

$$\boldsymbol{A}_1 = (0.1, 0.2, 0.3, 0.4)', \ \ \boldsymbol{a}_0 = (.1)\mathbf{1}_4, \ \ \boldsymbol{\beta} = (.8)\mathbf{1}_4, \ \ \boldsymbol{\delta} = (.1)\mathbf{1}_4. \tag{2.103}$$

The sample size is $T = 600$. The prior hyperparameters are as in the previous example. Here are the estimates obtained with $10,000$ MCMC samples of Algorithm 1 (after 1,000

burn-ins):

$$\widehat{\boldsymbol{B}} = \begin{pmatrix} .0128 & .2556 & .4314 & .1671 & .1743 & .3567 \\ -.0045 & -.0017 & -.0791 & .5670 & -.0411 & -.0738 \\ -.0049 & .0923 & -.0044 & -.0435 & .5420 & -.0817 \\ .0028 & -.0004 & .0140 & .0012 & .0019 & .4820 \end{pmatrix},$$

$$\widehat{\boldsymbol{\gamma}}_b = \begin{pmatrix} .9407 & .9818 & .9807 & .8943 & .9034 \\ .0731 & .0324 & .9323 & .\ 0228 & .0300 \\ .8775 & .0378 & .0228 & .9264 & .0322 \\ .0630 & .0218 & .0209 & .0205 & .9332 \end{pmatrix},$$

$$\widehat{\boldsymbol{\Psi}} = \begin{pmatrix} 1 & .5178 & .4800 & .2193 \\ 0 & 1 & -.0118 & -.0490 \\ 0 & 0 & 1 & -.0622 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \widehat{\boldsymbol{\gamma}}_\psi = \begin{pmatrix} * & .9100 & .9051 & .8341 \\ * & * & .0427 & .0268 \\ * & * & * & .0344 \\ * & * & * & * \end{pmatrix}.$$

Estimates of the latent volatilities are shown in Table 2.10, while plots with the true and estimated log-volatilities are shown in Figure 2.6.

Table 2.10: Posterior mean for stochastic volatility generation process parameters

$\widehat{\boldsymbol{A}}_1$	.1531	.2519	.3361	.4368
$\widehat{\boldsymbol{a}}_0$	.1374	.1032	.0842	.0749
$\widehat{\boldsymbol{\beta}}$	.7491	.7830	.7975	.8036
$\widehat{\boldsymbol{\delta}}$	.0857	.1321	.1215	.0789

Compared with the previous example, here we have one more variable and $T = 600$. Again, we have generally satisfactory estimates for the main equation parameters, while some of the estimates in the stochastic volatility equation still present big errors, here even bigger (in percentage) than in the previous example.

Example 2.9. The last example of the section shows how simply changing the parameters of the stochastic volatility process can change the quality of the results when the sample size is not very big. Consider the same model of Example 2.8, just with a different data generating process:

$$\boldsymbol{A}_1 = (.1, .2, .3, .4)', \;\; \boldsymbol{a}_0 \;=\; \boldsymbol{\beta} \;=\; \boldsymbol{\delta} \;=\; (.1)\mathbf{1}_4. \tag{2.104}$$

The sample size is $T = 500$, while prior hyperparameter and number of MCMC draws are the same as in Example 2.8. The estimates obtained are:

$$\widehat{\boldsymbol{B}} = \begin{pmatrix} .0128 & .3141 & .5060 & .2110 & .2147 & .4365 \\ -.0045 & -.0023 & -.0065 & .5146 & -.0018 & -.0266 \\ -.0049 & .1117 & -.0306 & -.0205 & .4873 & -.0377 \\ .0028 & -.0005 & .0022 & .0137 & -.0049 & .5330 \end{pmatrix},$$

$$\widehat{\boldsymbol{\gamma}}_b = \begin{pmatrix} .9959 & .9985 & .9793 & .8916 & .9753 \\ .0630 & .0215 & .9700 & .0207 & .0249 \\ .8637 & .0225 & .0220 & .9155 & .0258 \\ .0616 & .0201 & .0207 & .0208 & .9741 \end{pmatrix},$$

$$\widehat{\boldsymbol{\Psi}} = \begin{pmatrix} 1 & .5031 & .5031 & .1915 \\ 0 & 1 & -.0010 & -.0044 \\ 0 & 0 & 1 & -.0216 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \widehat{\boldsymbol{\gamma}}_\psi = \begin{pmatrix} * & .9984 & .9990 & .8859 \\ * & * & .0222 & .0220 \\ * & * & * & .0224 \\ * & * & * & * \end{pmatrix}.$$

Estimates of the latent volatilities are shown in Table 2.11. Plots of the true and estimated volatilities can be found in Figure 2.7. Here estimates are still not perfect but are generally better than in the previous example. It can be seen that just modifying β_j to be 0.1 for all js leads to more precise estimates of ($\boldsymbol{\Psi}$, $\boldsymbol{a}_0$, $\boldsymbol{\delta}$), even if the number of observations decreased from $T = 600$ to $T = 500$.

Table 2.11: Posterior Mean for Stochastic Volatility Generation Process Parameters

$\widehat{\boldsymbol{A}}_1$	.1324	.2449	.3254	.4412
$\widehat{\boldsymbol{a}}_0$	.1216	.1134	.0983	.0935
$\widehat{\boldsymbol{\beta}}$	.1086	.0899	.0905	.0834
$\widehat{\boldsymbol{\delta}}$	.0841	.0840	.0840	.0938

2.7 Macroeconomic Applications

We analyze two real data examples. The first example is on daily exchange rates of US dollar, and is known to exhibit stochastic volatilities. The other example is about a set of macroeconomic indicators, measured on a monthly basis. Prior hyperparameters are chosen as in Example 2.6, except $s_0 = 10^{-8}$, $\alpha_q = 0.1$, and $\beta_q = 0.05$.

2.7.1 Daily Exchange Rates

Our data consist of daily exchange rates between US dollar and ten major currencies between January 2001 and November 2005, for a total of $1,222$ observations.

We examine the daily exchange rates of USD with, respectively, EUR, GBP, JPY, CAD, MXN, BRL, CHF, NOK, SGD and INR[3]. We transform the data using a function of the log differences, precisely

$$y_{jt} = 100 \times \Big\{ \log r_{jt} - \log r_{j,t-1} - \frac{1}{T} \sum_{i=1}^{T} (\log r_{ji} - \log r_{j,i-1}) \Big\}, \tag{2.105}$$

where r_{jt} denotes the exchange rate j at time t. Chib et al. (2002) use this transformation for modeling univariate stochastic volatilities. Their parametrization does not include exogenous or endogenous variables in the right end side of the data generating equation,

[3]Currency codes are according to the iso4217 standard.

which is then composed only by the error.

We apply Model (2.6), with $\boldsymbol{X}$ composed of lag 1 values of y_t. Exogenous variables are not present in the stochastic volatility generating process either. For every parameter we present the posterior mean obtained simulating $50,000$ MCMC cycles of Algorithm 2 ($M = 1,000$) and discarding the first $10,000$ as burn-in. Figure 2.8 gives a scatter plot of the posterior mean of the non-redundant elements of ($\boldsymbol{\Psi}$, $\boldsymbol{\gamma}_\psi$, $\boldsymbol{B}$ $\boldsymbol{\gamma}_b$). It is noticeable that the posterior mean of $\boldsymbol{\gamma}_\psi$ easily allows us to distinguish between the zero and non-zero elements of the matrix $\boldsymbol{\Psi}$, being all the elements close to zero or one. Also, all the elements of the matrix $\boldsymbol{B}_1$ present corresponding indicators in $\boldsymbol{\gamma}_b$ very close to zero. This signifies that the data generating process is not autoregressive, and so that data can be modeled as in Chib et al. (2002). What they assumed a priori though, we reached via model selection.

Below are the posterior means of the stochastic volatility generation process parameters:

$$\begin{aligned}
\widehat{\boldsymbol{a}}_0 &= (-.0012, .0002, -.1103, -.005, -.1247, -.0122, -.9583, .0331, -.055, -.2671)', \\
\widehat{\boldsymbol{\beta}} &= (.4581, -.1815, -.5509, .1666, -.1042, .8823, .6137, -.3741, -.1151, .1391)', \\
\widehat{\boldsymbol{\delta}} &= (.0002, .0010, .0147, .0018, .0129, .1693, .0158, .0068, .0141, .0089)', \\
\widehat{\boldsymbol{\gamma}}_\delta &= (.8930, 1.0000, 1.0000, 1.0000, 1.0000, 1.0000, 1.0000, 1.0000, .9780, 1.0000)', \\
\widehat{\boldsymbol{q}} &= (28.495, 57.524, 60.4645, 59.414, 60.976, 57.206, 61.010, 60.442, 28.166, 59.862)'.
\end{aligned}$$

All δ_js are selected to be in the model, no matter how small the actual value is. This is concordant with well known results about the nature of exchange rates. The value of the hyperparameter q_j changes to adapt to the different magnitude of δ_j values. Most of the variables present values of $a_{0,j}$ very close to zero, with the exception of Variables 7 and 8, which show strong negative values. Variables 1, 4, 6, 7, and 10 have positive β_j

coefficient, while the remaining a negative one. Table 2.12 reports the posterior mean and standard deviation of h_{jt} averaged over t.

Table 2.12: Posterior Mean for $\boldsymbol{h}$, $\boldsymbol{\lambda}$, Posterior SD for $\boldsymbol{h}$, averaged over t

variable	*posterior mean* (h)	*posterior mean* (λ)	*posterior s.d.* (h)
1	-0.0016	0.9984	0.0008
2	0.0002	1.00022	0.0037
3	-0.0691	0.9335	0.0241
4	-0.0056	0.9944	0.0241
5	-0.0970	0.9083	0.0417
6	-0.3843	0.9521	0.8168
7	-2.2683	0.1400	0.6860
8	0.0232	1.0235	0.0106
9	-0.0443	0.9573	0.0352
10	-0.2690	0.7689	0.1083

2.7.2 Monthly Macroeconomic Indicators

In the second real data example we use a set of monthly measurements of macroeconomic indicators. Unlike the previous case, the nature of the conditional volatilities has not between thoroughly studied for this type of data. We use our model selection technique to see whether some or all the variables present a conditional stochastic variability. The variables are PP index of crude materials, CP index, index of industrial production, Standard and Poor adjusted close value, federal funds rate, weighted exchange rate of US dollar against some major foreign currencies. The sample period is from January 1973 to November 2005, for a total of 393 observations. All the variables have been log transformed with the exception of federal funds rate. The data generating process has

been modeled as a VAR(12) without exogenous variables, while the volatility generating process is an AR(1) with no exogenous variables. We perform stochastic search selection on all the parameters of interest, using $40,000$ MCMC cycles of Algorithm 2 (M=$1,000$) after $10,000$ burn-ins.

First, we report the values of the first column of $\boldsymbol{B}$ ($\boldsymbol{b}_0$), which is not subject to model selection:

$$\widehat{\boldsymbol{b}_0} \;=\; (-4.4490,\; -0.0849,\; 4.0647,\; -2.4985,\; -2.3704,\; 8.3390)'.$$

Then, since $\boldsymbol{B}_1$ has 438 elements, we display it using a plot. Figure 2.9 gives the scatter plot of the posterior means for $\boldsymbol{B}_1$ and the correspondent values of $\boldsymbol{\gamma}_b$. Since we have seen, from simulation studies, that for relatively small sample sizes the stochastic search algorithm tends to oversimplify the model, we have chosen a low cutoff point for $\boldsymbol{\gamma}_b$: Table 2.13 reports the posterior mean and SD of the $\boldsymbol{B}_1$ elements with relative γ_b bigger than 0.2.

Only 14 γ_{bj} are larger than 0.2, which leads to a considerable simplification of a very complex starting model. All the variables show correlation with their own first lag except Variable 5 (federal funds rate). Variable 4 (S&P index) is also correlated with its lag 2 term. Some lag 1 relations are also present between Variables 3 and 2, 4 and 2, 4 and 3 and 5 and 2, and some higher order relations between Variables 3 and 2, 4 and 1 and 4 and 6. Variable 4 (S&P close value) shows to be the most influenced by the other series, being influenced by the index of crude materials, CP index, index of industrial production and exchange rates. It should be noticed that changes in the CP index and in the index of industrial production will influence the S&P close value in a shorter time than changes in the index of crude materials. The variables that shows less relations is Variable 5 (federal funds rate), while the variable influencing most indicators

Table 2.13: Posterior Mean and SD of $\boldsymbol{B}_1$ Elements With a Correspondent $\widehat{\gamma}_{bj} \geq 0.2$

$\widehat{\gamma_{bj}}$	$\widehat{b_{ij}}$	$SD\,of\,\widehat{b_{ij}}$	$Relates$
1.000	0.9019	0.0082	$y_{1t}\ and\ y_{1,t-1}$
1.000	0.7980	0.0069	$y_{2t}\ and\ y_{2,t-1}$
0.5470	-0.1489	0.0150	$y_{3t}\ and\ y_{2,t-1}$
0.9662	0.4944	0.0062	$y_{3t}\ and\ y_{3,t-1}$
0.3712	0.0245	0.0053	$y_{3t}\ and\ y_{2,t-2}$
0.6043	0.3143	0.0154	$y_{3t}\ and\ y_{3,t-2}$
0.2435	-0.1768	0.0198	$y_{4t}\ and\ y_{2,t-1}$
0.2742	-0.1823	0.0124	$y_{4t}\ and\ y_{3,t-1}$
1.000	0.8875	0.0108	$y_{4t}\ and\ y_{4,t-1}$
0.5535	-0.2888	0.0296	$y_{4t}\ and\ y_{1,t-7}$
0.2915	-0.2218	0.0382	$y_{4t}\ and\ y_{6,t-8}$
0.2670	0.2742	0.0326	$y_{5t}\ and\ y_{2,t-1}$
1.000	0.8287	0.0086	$y_{6t}\ and\ y_{6,t-1}$

is Variable 2 (CP index), which influences index of industrial production, S&P close value and federal funds rate. We also present the posterior mean for the parameters in the stochastic volatility equation, while a plot of the estimated volatilities os shown in Figure 2.11:

$$\begin{aligned}
\widehat{\boldsymbol{a}}_0 &= (-0.025,\ 0.002,\ 0.054,\ -0.013,\ 0.057,\ -0.026)', \\
\widehat{\boldsymbol{\beta}} &= (0.0444,\ -0.3790,\ -0.0074,\ 0.1784,\ -0.6547,\ -0.6149)', \\
\widehat{\boldsymbol{\delta}} &= (0.0015,\ 0.0181,\ 0.0053,\ 0.0356,\ 0.0013,\ 0.0218)', \\
\widehat{\boldsymbol{\gamma}}_\delta &= (0.8260,\ 0.5960,\ 0.9270,\ 0.8250,\ 0.6180,\ 0.7670)', \\
\widehat{\boldsymbol{q}} &= (10.0350,\ 21.5314,\ 42.1637,\ 14.2045,\ 24.0582,\ 20.4372)'.
\end{aligned}$$

Also in this dataset all the variables seem to show stochastic volatility.

We compare the out of sample forecast ability of the model chosen via model selection with that of an unrestricted model. In order to do so, we divide the data set in two: a training sample with 357 observations and a predicting (or forecasting) sample with 24 observations. The MSE is calculated over the forecasting period, for each variable. The ratio for the MSE computed with the unrestricted and the one computed with the restricted model is:

$$\frac{MSE_{unres}}{MSE_{res}} = (1.5423,\ 1.3526,\ 2.5489,\ 1.2256,\ 0.9632,\ 0.8772).$$

Here the result is different than in the simulations, since for two of the variables the unrestricted model seems to perform better. Results are not really comparable though, since in the simulation we were comparing a one step ahead forecast over fifty samples, while here we are performing an out of sample forecasting over a very long period.

It should be anyway noticed the asymmetry in forecast errors between the restricted and unrestricted models: the smallest ratio is .8772 (Variable 6), while the biggest ratio

is 2.5489 (Variable 3), so for the variables better forecasted by the unrestricted model the difference is smaller than for the variables forecasted better by the restricted one.

2.8 Variables' Order

One of the limitations of the modified Cholesky decomposition is that estimation results are not invariant to the order of variables. Interpretation of the decomposition parameters can facilitate the variable ordering: using (2.7), we model the marginal variance of variable 1, the conditional variance of variable 2 given variable 1, the conditional variance of variable 3 given variables 1 and 2, and so on. The order of the variables can be then controlled by the researcher according to theoretical reasoning. For example, macroeconomic theory has been used to justify the specific variables' order in the two real data examples before. Here, we present the data analysis of the same dataset of monthly macroeconomic indicators analyzed in the previous section, the only difference consisting in the order of variables which has been reversed. The order is now: weighted exchange rate of US dollar against some major foreign currencies, federal funds rate, Standard and Poor 500 stock price adjusted close value, Index of Industrial Production, Consumer Price Index (CPI), and Producer Price Index (PPI) of crude materials. All the variables have been log transformed with the exception of federal funds rate. The data generating process has been modeled as a VAR(12) without exogenous variables, while the volatility generating process is an AR(1) with no exogenous variables, exactly as in the previous section. First, we report the values of the first column of $\boldsymbol{B}$, $\boldsymbol{b}_0$:

$$\widehat{\boldsymbol{b}_0} \;=\; (7.5442,\; -2.6646,\; -5.3813,\; 2.9355,\; 0.7153,\; -3.6233)'.$$

Then, Figure 2.10 gives the scatter plot of the posterior means for $\boldsymbol{B}_1$ and the correspondent values of $\boldsymbol{\gamma}_b$. The plot shows three clusters of γ_{bj} values. One is concentrated on 0, one from 0.2 to 0.4 and one from 0.6 to 1.0. Table 2.14 shows the posterior mean and SD of the $\boldsymbol{B}_1$ elements with relative γ_b bigger than 0.2.

Table 2.14: Posterior Mean and SD of $\boldsymbol{B}_1$ Elements With a Correspondent $\widehat{\gamma}_{bj} \geq 0.2$

$\widehat{\gamma_{bj}}$	$\widehat{b_{ij}}$	$SD\,of\,\widehat{b_{ij}}$	$Relates$
1.000	0.8395	0.0061	$y_{1t}\ and\ y_{1,t-1}$
1.000	0.7233	0.0071	$y_{2t}\ and\ y_{2,t-1}$
0.2730	0.2178	0.0126	$y_{2t}\ and\ y_{4,t-1}$
0.4170	-0.2502	0.0068	$y_{3t}\ and\ y_{2,t-1}$
1.0000	0.8851	0.0009	$y_{3t}\ and\ y_{3,t-1}$
0.2450	0.1586	0.1054	$y_{3t}\ and\ y_{5,t-7}$
0.3430	-0.2448	0.0024	$y_{3t}\ and\ y_{6,t-7}$
0.2640	-0.2173	0.0064	$y_{3t}\ and\ y_{3,t-9}$
0.8850	0.4320	0.0286	$y_{4t}\ and\ y_{4,t-1}$
0.4420	0.2789	0.0296	$y_{4t}\ and\ y_{4,t-2}$
0.9880	0.8758	0.0082	$y_{5t}\ and\ y_{5,t-1}$
0.2670	0.1785	0.0326	$y_{6t}\ and\ y_{4,t-1}$
0.6950	-0.5376	0.2040	$y_{6t}\ and\ y_{5,t-1}$
1.000	0.8529	0.2040	$y_{6t}\ and\ y_{6,t-1}$

Only 15 elements of $\boldsymbol{B}_1$ have a correspondent γ_{bj} bigger than 0.2. All the variables show correlation with their own first lag, Variable 4 (index of industrial production), also with its lag 2 term, and Variable 3 (S&P index) with its lag 9. Some lag 1 relations are also present between Variables 2 and 4, 3 and 2, 4 and 6 and 5 and 6, and some higher order relations between Variables 3 and 5 and 3 and 6. Variables 3 (S&P close value) and 6 (PP index of crude materials) show to be the most influenced by the other series.

Variable 2 (federal funds rate) is influenced by the previous month index of industrial production, while the remaining three variables seem to be influenced only by their past values. We also present the posterior mean for the parameters in the stochastic volatility equation:

$$\begin{aligned}
\widehat{\boldsymbol{a}}_0 &= (0.0009,\ -0.0020,\ 0.0010,\ 0.0121,\ -0.0755,\ 0.0035)', \\
\widehat{\boldsymbol{\beta}} &= (-0.1337,\ -0.4095,\ -0.2271,\ -0.8471,\ 0.8245,\ -0.3497)', \\
\widehat{\boldsymbol{\delta}} &= (0.0001,\ 0.0031,\ 0.0115,\ 0.0093,\ 0.0282,\ 0.0089)', \\
\widehat{\boldsymbol{\gamma}}_\delta &= (0.8800,\ 0.9970,\ 0.8240,\ 0.8210,\ 0.9810,\ 0.7240)', \\
\widehat{\boldsymbol{q}} &= (1.5478,\ 62.0149,\ 25.5232,\ 18.3992,\ 57.9823,\ 27.5507)'.
\end{aligned}$$

Again, all the variables show stochastic volatility. Comparing these estimates with those of the previous analysis, we can see that even if the results are similar, there are some substantial differences. For example, the AR(1) coefficient of federal funds rate is significant only with this variable ordering, and the higher order coefficients are quite different.

The out of sample forecast ability of the model, compared with an unrestricted model, provides similar results with both orders:

$$\frac{MSE_{unres}}{MSE_{res}} = (0.8217,\ 0.9562,\ 1.2652,\ 2.6513,\ 1.3367,\ 1.7726).$$

A plot of the forecasting errors is provided in Figure 2.12.

2.9 Conclusions

In this chapter we develop and implement algorithms for Bayesian model selection of multivariate stochastic volatility models. We extend the stochastic search method

of George & McCulloch (1993) to multivariate regression models with time varying covariances, providing a computationally feasible method to select among a large number of competing models.

We presented different algorithms to perform Bayesian estimation and stochastic search model selection, differing on the technique used to simulate the latent variables h_{jt}: direct sampling via Gilks adaptive sampler, particle filtering and smoothing, and filtering using a rejection sampler. Simulation studies show that the particle filter algorithm is preferable in terms of the quality of the estimates and efficiency, while Gilks adaptive sampler is effective for simple models. Real data examples show how the stochastic search algorithm can greatly simplify complex models and allows the researcher to gain insights on the problem he, or she, is analyzing. We compare forecasting ability using both simulated and real data under unrestricted models and restricted models obtained through stochastic search. We find that restricted models generally outperform unrestricted models.

We compare the choice of the best model via posterior mean and posterior mode of $\boldsymbol{\gamma}$, and pointed advantages and disadvantages of both methods. We have seen how for relatively large sample sizes the choice of the best model is easy using both methods, while as the number of observations decreases, further attention is necessary in order not to be mislead.

The model can still be improved, and further developments and generalizations are possible. Those will be discussed in the chapter relative to further research.

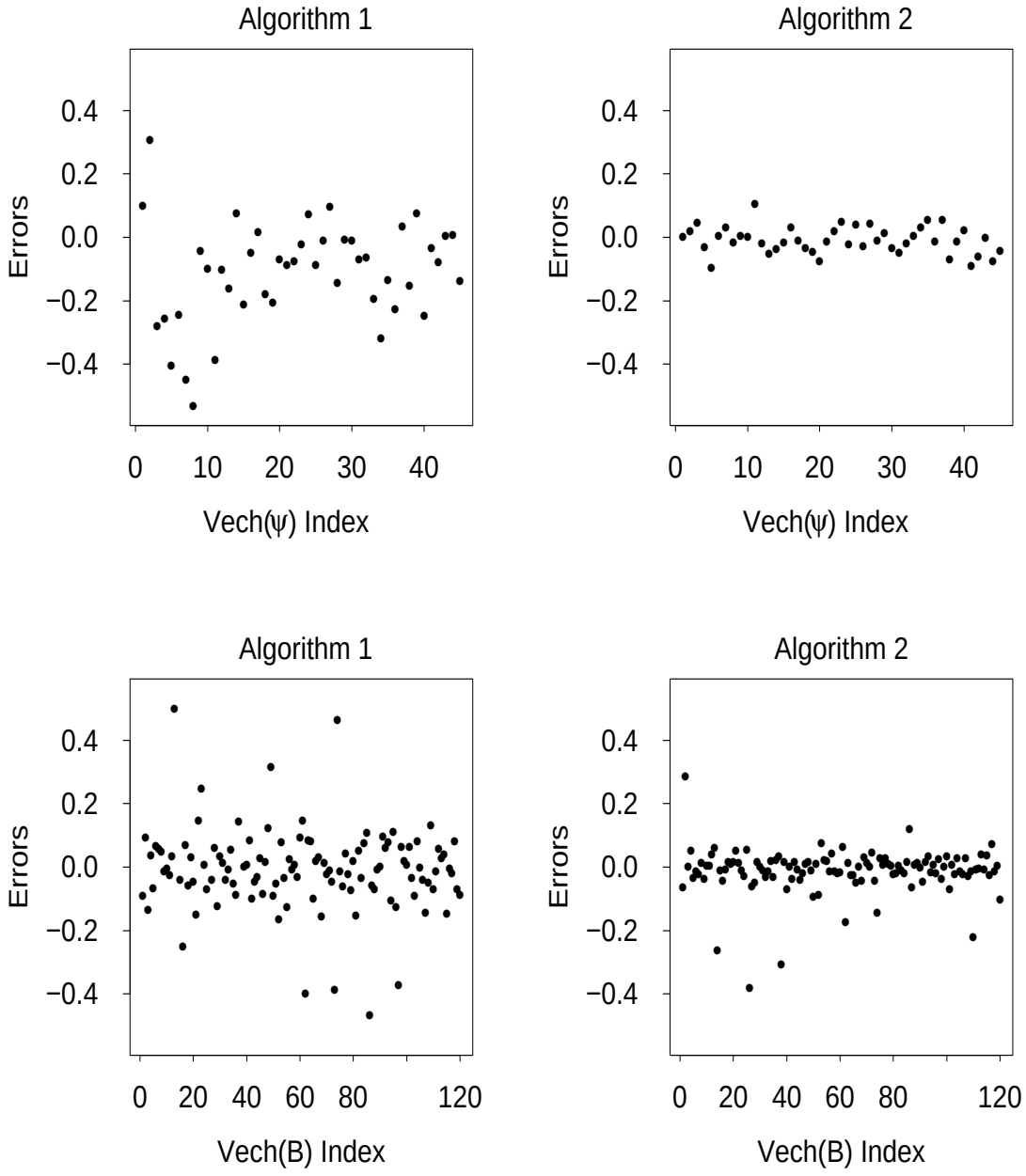

Figure 2.1: Comparison Between Algorithms 1 and 2. We compare the results obtained with $p = 10$, $10,000$ MCMC cycles after $1,000$ burn-ins and using starting values far from the true ones. Results are presented plotting the standardized differences between the true parameter value and the correspondent estimate. Differences are standardized by the range of the true parameter values. Both $\boldsymbol{\Psi}$ and $\boldsymbol{B}$ and their correspondent estimates have been vectorized before being plotted.

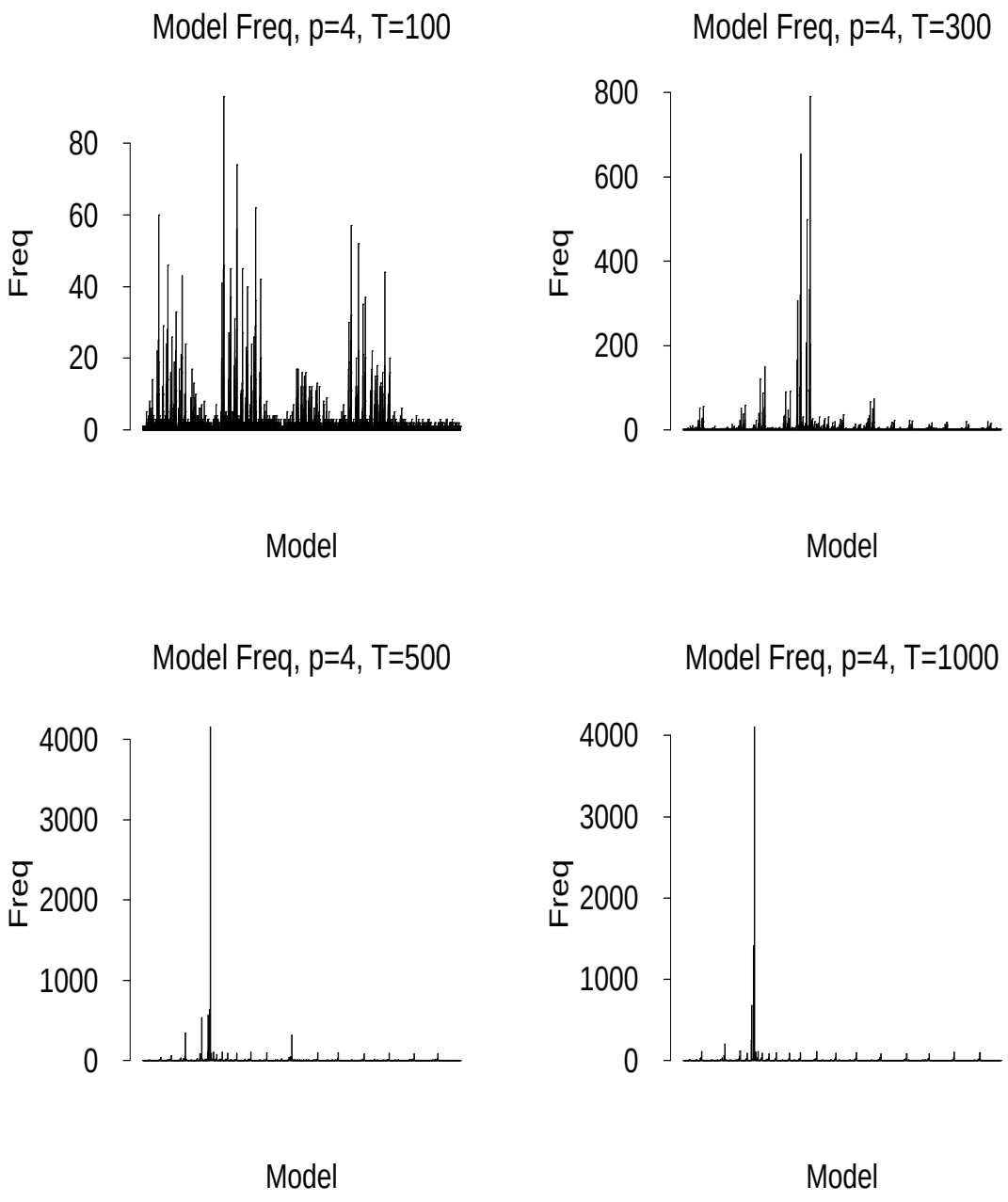

Figure 2.2: Barplot of the Most Visited Model. $p = 4$

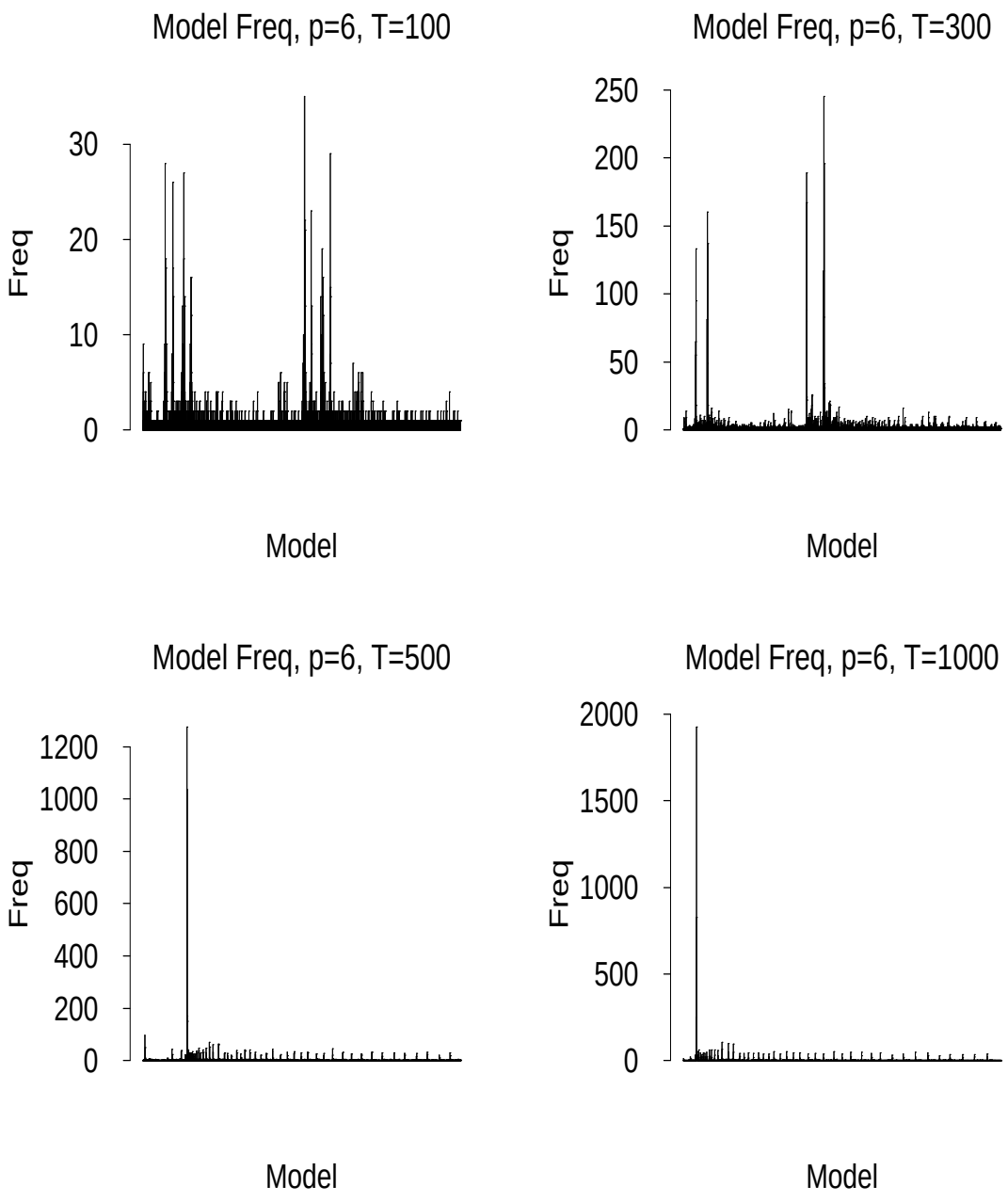

Figure 2.3: Barplot of the Most Visited Model. $p = 6$

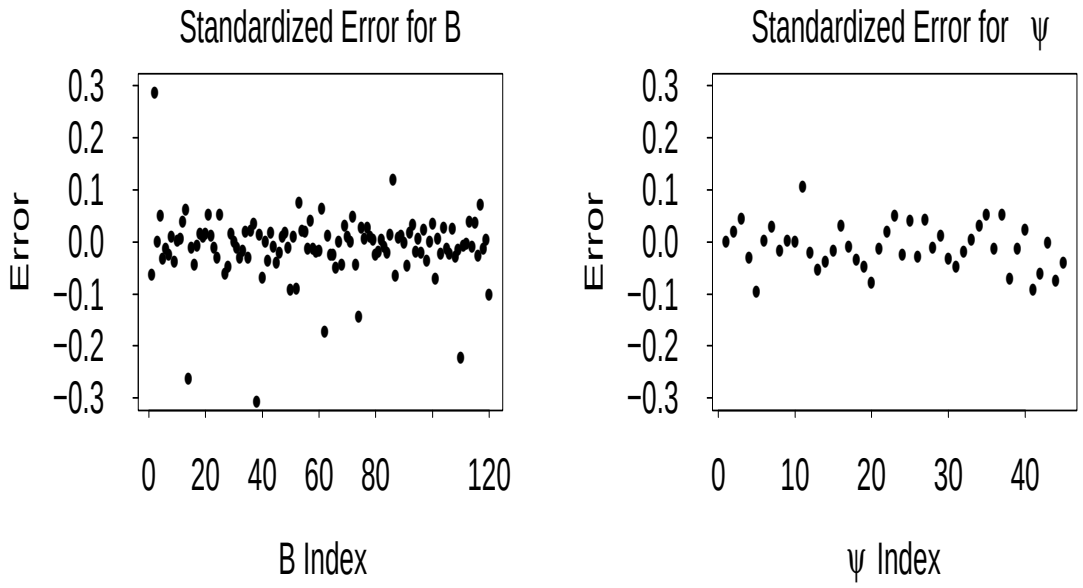

Figure 2.4: Standardized Estimation Errors for B and $\mathbf{\Psi}$. The matrices are vectorized by row and the errors are standardized by the true value range

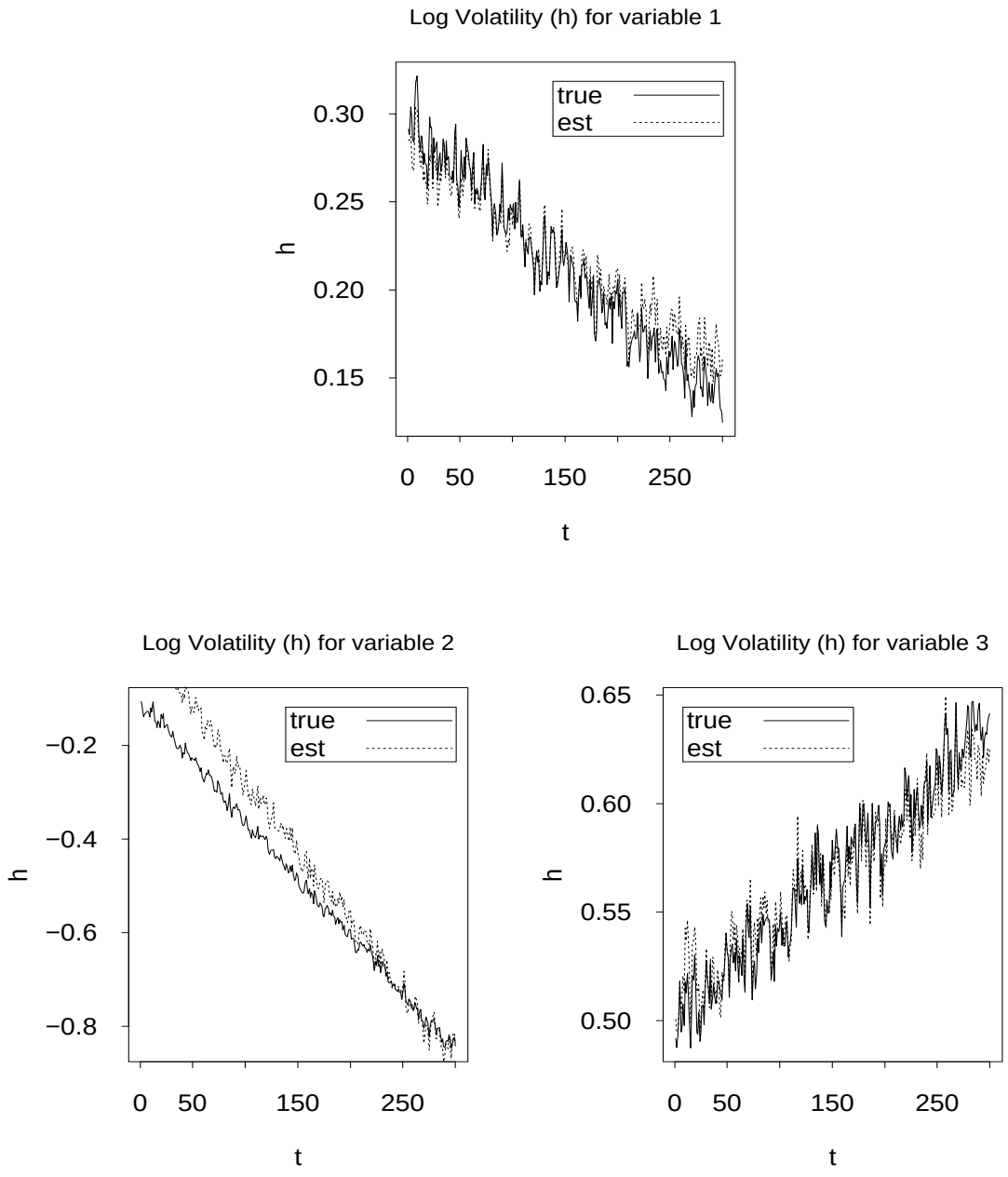

Figure 2.5: Comparison of the True and Estimated Log-Volatilities for One Random Sample.

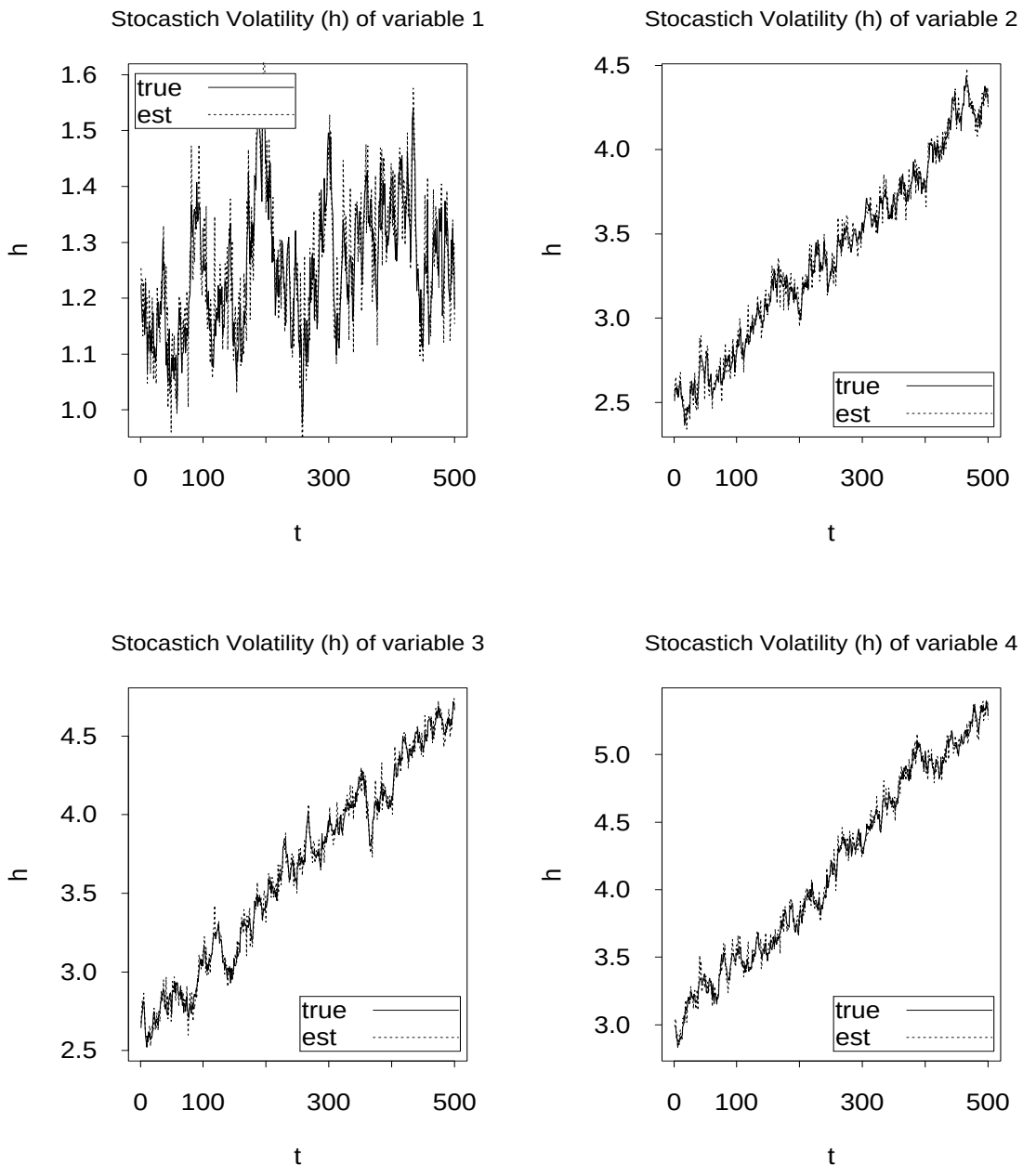

Figure 2.6: Comparison of the True and Estimated Log-Volatilities for One Random Sample.

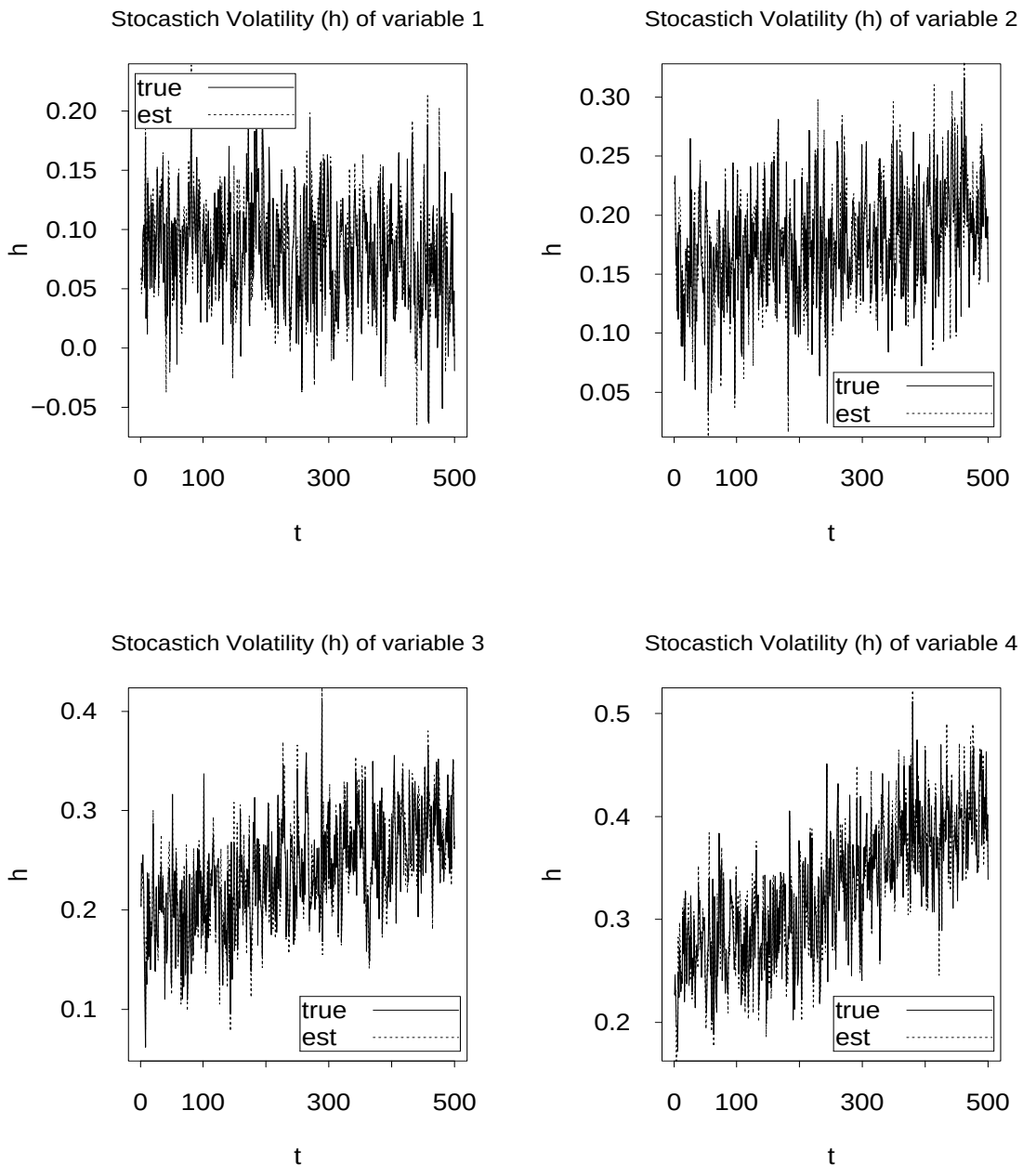

Figure 2.7: Comparison of the True and Estimated Log-Volatilities for One Random Sample.

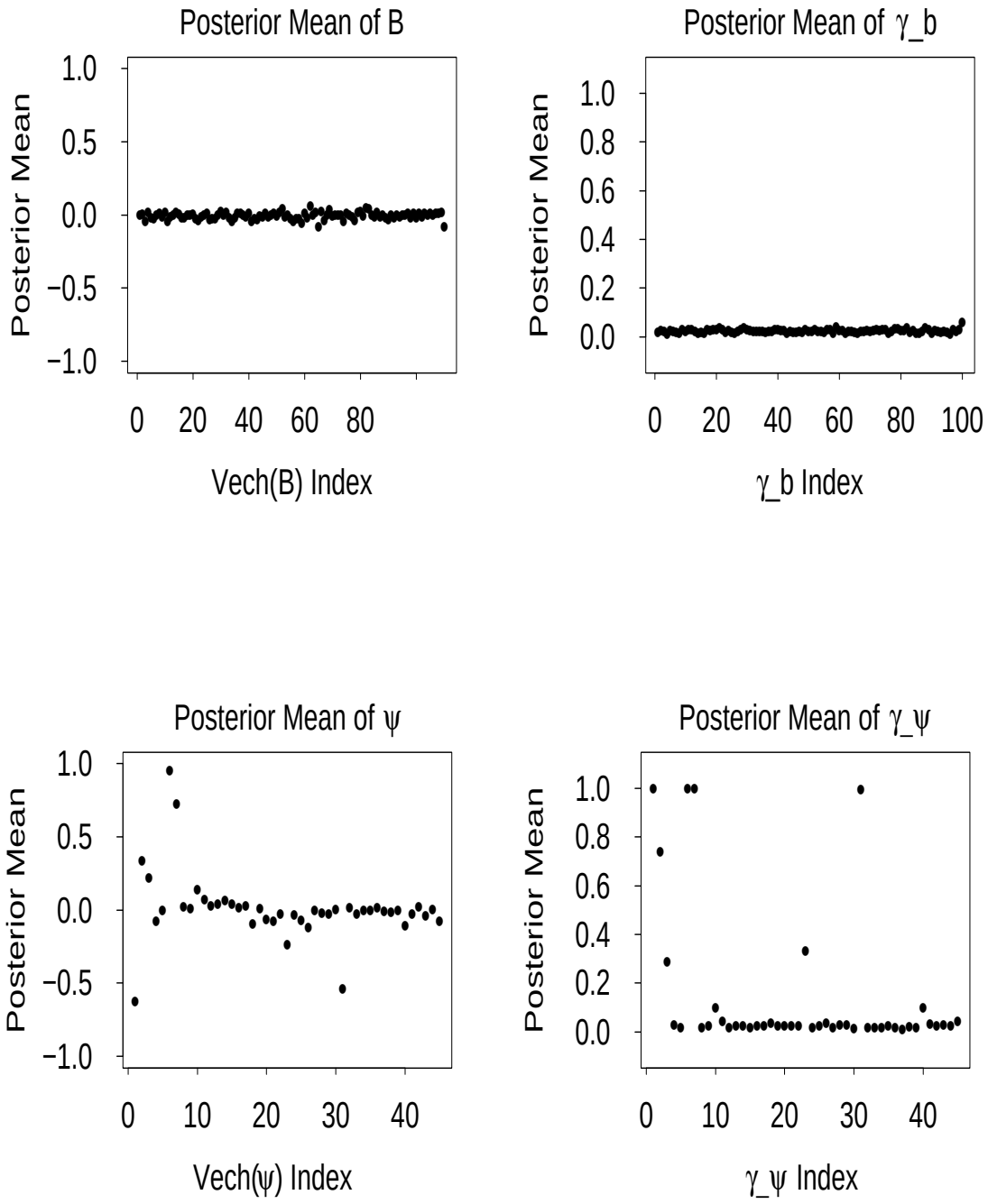

Figure 2.8: (Daily Exchange Rates) Posterior Means for $\boldsymbol{\Psi}$, $\boldsymbol{\gamma}_\psi$, $\boldsymbol{B}$ and $\boldsymbol{\gamma}_b$. The matrices are vectorized by row

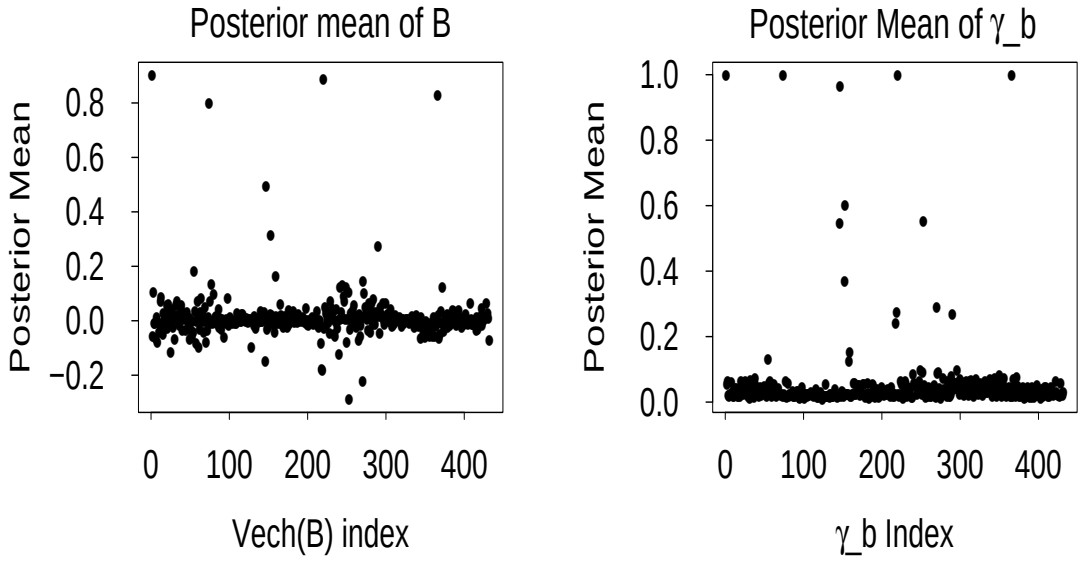

Figure 2.9: (Macroeconomic Indicators) Posterior Means for $\boldsymbol{B}$ and $\boldsymbol{\gamma}_b$. The matrices are vectorized by row

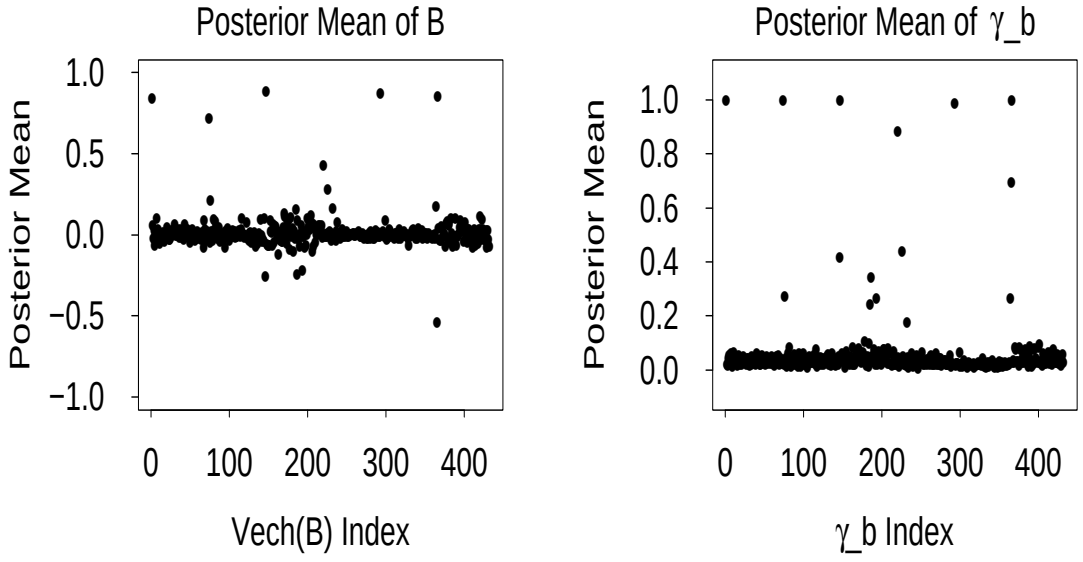

Figure 2.10: (Macroeconomic Indicators Inverted) Posterior Means for $\boldsymbol{B}$ and $\boldsymbol{\gamma}_b$. The matrices are vectorized by row

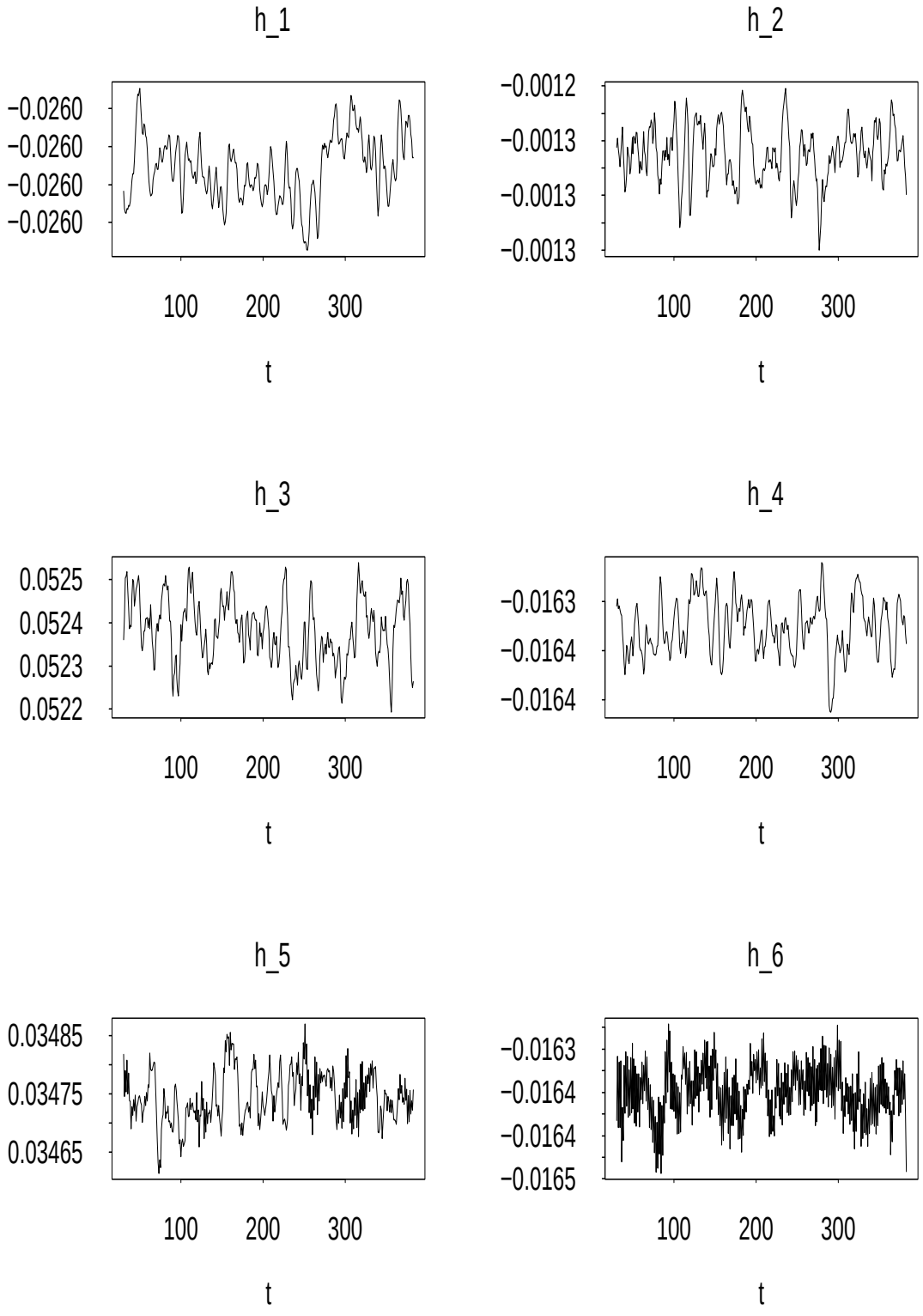

Figure 2.11: (Macroeconomic Indicators) Posterior Means of h_{jt}. Plots are centered on the average of h_{jt} and all have the same range (0.25)

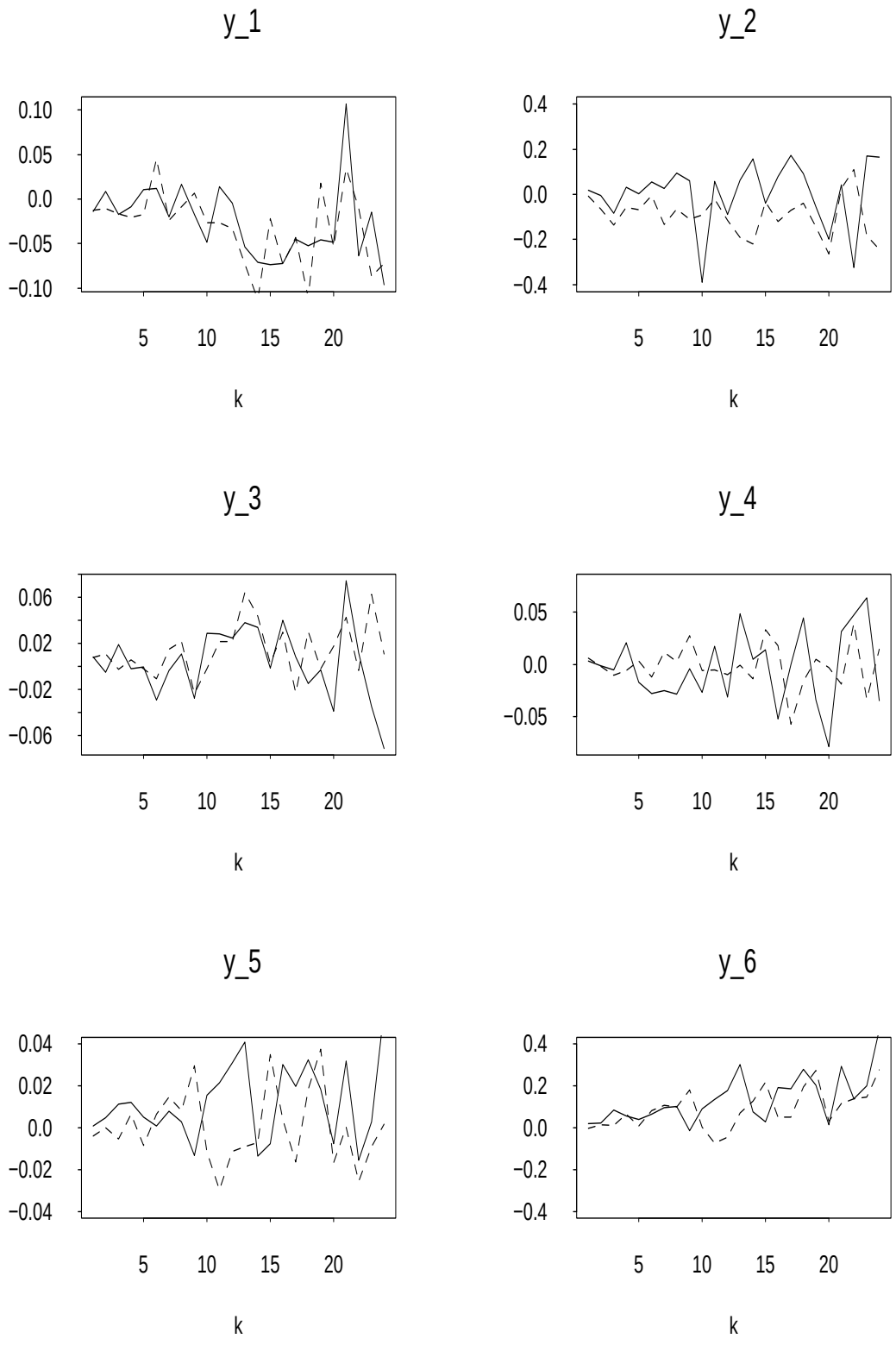

Figure 2.12: Forecast Error for the Unrestricted and Restricted Model. Solid lines represent the unrestricted model errors, while dashed lines represent the the restricted's.

Chapter 3

An Alternative Cholesky Decomposition

3.1 Introduction

Following and expanding Pourahmadi (1999), so far we parameterized $\mathbf{\Sigma}_t$ using (2.7). Several other options are available though. According to Pourahmadi (2005), we can divide covariance decompositions in three categories, in increasing order of effectiveness: variance-correlation (Barnard et al. (2000)), spectral (Chiu et al. (1996)), and Cholesky (Pinheiro et al. (1996), Pourahmadi (1999), Smith & Kohn (2002), Chen & Dunson (2003)). In this study, we are concentrating only on Cholesky decompositions. These have the unique characteristic of "providing unconstrained and statistically meaningful reparametrization of a covariance matrix at the expense of imposing an order (coordinate) among the variables" (Pourahmadi (2005)). In the previous chapter, we used a dynamic version of the Pourahmadi (1999) decomposition to parameterize our time varying covariance matrix. Imposing that particular decomposition is equivalent to modeling

the conditional variance of each variable j at time t given variables 1 to $j-1$ at time t. Chen & Dunson (2003) instead, proposed a different modified Cholesky decomposition, which permits to separate the variance and correlation parameters of the covariance matrix. Here we adapt their parametrization to a time varying variance model by imposing

$$\boldsymbol{\Sigma}_t = \boldsymbol{\Lambda}_t \boldsymbol{\Gamma} \boldsymbol{\Gamma}' \boldsymbol{\Lambda}_t, \tag{3.1}$$

where $\boldsymbol{\Lambda}_t$ is a diagonal and $\boldsymbol{\Gamma}$ a unit lower triangular matrix.

Using either (2.7) or (3.1), we obtain a diagonal matrix $\boldsymbol{\Lambda}_t$ whose elements are constrained to be positive, and a unit triangular $\boldsymbol{\Gamma}$ matrix whose below diagonal elements are unconstrained. The parametrization and interpretation of the two is quite different though. We showed in the previous chapter that (2.7) models the variance of y_{jt} conditional on $(y_{1t}, \cdots, y_{j-1,t})$:

$$y_{jt} \quad = \quad e_{jt} + \sum_{k=1}^{j-1} \gamma_{jk} e_{kt},$$

where γ_{jk} correspond to the non-redundant elements of $\boldsymbol{\Gamma}$, while the variance of e_{jt} is equal to λ_{jt}. The covariance between two contemporaneous observations is

$$cov(y_{jt}, y_{st}) \quad = \quad \sum_{k=1}^{s \wedge j} \gamma_{jk} \gamma_{sk} \lambda_{kt}. \tag{3.2}$$

This implies that the correlation between variable j and variable s depends on the variance of the innovations λ_{jt} and therefore, in our model, is time varying. To interpret the decomposition in (3.1), we can follow Pourahmadi (2005) and notice that the covariance matrix of $\boldsymbol{\Lambda}_t^{-1} \boldsymbol{y}_t$ is equal to $\boldsymbol{\Gamma}\boldsymbol{\Gamma}'$, and so

$$cov(\boldsymbol{\Gamma}^{-1} \boldsymbol{\Lambda}_t^{-1} \boldsymbol{y}_t) = \boldsymbol{I}_p. \tag{3.3}$$

Thus, the variance of y_{jt} can be written as

$$var(y_{jt}) \quad = \quad \lambda_{jt}^2 \sum_{k=1}^{j} \gamma_{jk}, \tag{3.4}$$

where λ_{jt} and γ_{jk} are the generic elements of $\mathbf{\Lambda}_t$ and $\mathbf{\Gamma}$ respectively. Also,

$$corr(y_{jt}, y_{st}) = \frac{\sum_{k=1}^{s\wedge j} \gamma_{sk}\gamma_{jk}}{\sqrt{\sum_{k=1}^{s} \gamma_{sk}^2 \sum_{k=1}^{s} \gamma_{jk}^2}}, \tag{3.5}$$

Since the correlation between two different variables is just a function of the elements of $\mathbf{\Gamma}$, the decomposition (3.1) implies that the contemporaneous correlations are not time varying.

Thus, performing model selection between a model parameterized using (2.7) and another parameterized using (3.1) can be a fast way to test whether the correlation between several time series are constant or time varying. It should be noticed that time varying variance models assumed constant correlations at first (Bollerslev (1990)), but later empirical studies showed this assumption to be unrealistic for many financial time series. Then, time varying correlation models had been developed. A general methodology to test correlation invariance still does not exist.

We consider the following p-vector multivariate model:

$$\boldsymbol{y}_t = \boldsymbol{b}_0 + \boldsymbol{B}_1 \boldsymbol{x}_t + \boldsymbol{e}_t, \tag{3.6}$$

for $t = 1, \cdots, T$, where $\boldsymbol{y}_t = (y_{1t}, \cdots, y_{pt})'$ is a $p \times 1$-vector of variables, $\boldsymbol{b}_0$ is a $p \times 1$-vector of unknown parameters, $\boldsymbol{x}_t$ is a $q \times 1$-vector of known endogenous or exogenous variables, $\boldsymbol{B}_1 \equiv (\boldsymbol{b}_1, \cdots, \boldsymbol{b}_q)$ is a $p \times q$-matrix of unknown parameters, $\boldsymbol{e}_t$ are independent $N_p(\mathbf{0}, \mathbf{\Sigma}_t)$ errors, and $\mathbf{\Sigma}_t$ is an unknown $p \times p$ positive definite matrix which allows (3.1). In order to model the evolution of λ_{jt}, it is convenient to consider its logarithm: let $h_{jt} = \log \lambda_{jt}$, and $\boldsymbol{h}_t = (h_{1t}, \cdots, h_{pt})'$. We model $\boldsymbol{h}_t$ as follows:

$$\boldsymbol{h}_t = \boldsymbol{a}_0 + diag(\boldsymbol{h}_{t-1})\boldsymbol{\beta} + \boldsymbol{A}_1 \boldsymbol{z}_t + diag(\boldsymbol{\delta})\boldsymbol{v}_t, \tag{3.7}$$

where $\boldsymbol{z}_t = (z_{t,1}, \cdots, z_{t,r})'$ is an r vector observable exogenous variable uncorrelated with $\boldsymbol{v}_t$, $\boldsymbol{a}_0$ is a p unknown vector, $\boldsymbol{A}_1 = \{a_{jk}\}$, $j = 1, \cdots, p$, $k = 1, \cdots, r$ is a matrix of

unknown parameters, $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)'$ is a vector of unknown parameters, $\boldsymbol{v}_t$ are iid $N_p(\mathbf{0}, \boldsymbol{I}_p)$, $\boldsymbol{\delta} = (\delta_1, \cdots, \delta_p)'$ is a vector of unknown nonnegative parameters.

The model here proposed is then exactly the same model proposed in Chapter 2, with the only difference in the covariance decomposition and parametrization. We are mainly interested in Bayesian estimation for this model, stochastic search model selection and comparison between (2.6) and (3.6). This last step cannot be implemented through stochastic search, so we are proposing a comparison using estimated Bayes factors.

The chapter is organized as follows: Section 3.2 defines the likelihood functions and the hierarchical priors for the model. Section 3.3 derives the conditional posterior distributions and proposes a MCMC algorithm for Bayesian estimation and variable selection. Section 3.4 is about comparisons of non-nested multivariate stochastic volatility models using Bayes factors. Section 3.5 reports simulation results.

3.2 The Hierarchical Model

3.2.1 Likelihood Functions

We give several formulas for the likelihood functions. First we know that the likelihood function of $(\boldsymbol{b}_0, \boldsymbol{B}_1, \boldsymbol{\Sigma}_1, \cdots, \boldsymbol{\Sigma}_T)$ based on $\boldsymbol{Y} = (\boldsymbol{y}_1, \cdots, \boldsymbol{y}_T)$, is exactly (2.18). In order to derive more convenient forms of the likelihood function, we define the decomposition (3.1) in terms of the precision matrix:

$$\boldsymbol{\Sigma}_t^{-1} = \boldsymbol{\Lambda}_t^{-1} \boldsymbol{\Psi} \boldsymbol{\Psi}' \boldsymbol{\Lambda}_t^{-1}, \tag{3.8}$$

where $\boldsymbol{\Psi} = \boldsymbol{\Gamma}^{-1'}$. Clearly $\boldsymbol{\Psi}$ is a unit upper triangular matrix. Then, we have the likelihood function of $(\boldsymbol{b}_0, \boldsymbol{B}_1, \boldsymbol{\Psi}, \boldsymbol{\Lambda}_1, \cdots, \boldsymbol{\Lambda}_T)$,

$$[\boldsymbol{Y} \mid \boldsymbol{b}_0, \boldsymbol{B}_1, \boldsymbol{\Psi}, \boldsymbol{\Lambda}_1, \cdots, \boldsymbol{\Lambda}_T] \propto |\boldsymbol{\Psi}|^T \prod_{t=1}^{T} |\boldsymbol{\Lambda}_t|^{-\frac{1}{2}} etr\Big\{-\frac{1}{2}\sum_{t=1}^{T} \boldsymbol{S}_t(\boldsymbol{B})\boldsymbol{\Lambda}_t^{-1}\boldsymbol{\Psi}\boldsymbol{\Psi}'\boldsymbol{\Lambda}_t^{-1}\Big\}, \quad (3.9)$$

where, $\boldsymbol{B} = (\boldsymbol{b}_0, \boldsymbol{B}_1)$ and $\boldsymbol{S}_t(\boldsymbol{B})$ is defined by (2.21)

In order to derive the posterior distribution for $(\boldsymbol{b}_0, \boldsymbol{B}_1)$, it is convenient to rewrite (3.6) as (2.22). Vectorizing $\boldsymbol{Y}$, we can then write

$$(\boldsymbol{y} \mid \boldsymbol{b}, \boldsymbol{\Psi}, \boldsymbol{\Lambda}) \sim N_{pT}((\boldsymbol{X}' \otimes \boldsymbol{I}_p)\boldsymbol{b},\ \boldsymbol{\Sigma}), \quad (3.10)$$

where

$$\boldsymbol{\Sigma} = diag(\boldsymbol{\Sigma}_1, \cdots, \boldsymbol{\Sigma}_T) = \boldsymbol{\Lambda}(\boldsymbol{I}_T \otimes \boldsymbol{\Gamma})(\boldsymbol{I}_T \otimes \boldsymbol{\Gamma}')\boldsymbol{\Lambda} = \boldsymbol{\Lambda}(\boldsymbol{I}_T \otimes \boldsymbol{\Gamma}\boldsymbol{\Gamma}')\boldsymbol{\Lambda}, \quad (3.11)$$

$$\boldsymbol{\Lambda} = diag(\boldsymbol{\Lambda}_1, \cdots, \boldsymbol{\Lambda}_T) = diag(\lambda_{11}, \lambda_{21}, \cdots, \lambda_{p1}, \lambda_{12}, \cdots, \lambda_{pT}). \quad (3.12)$$

Clearly, (3.11) implies that $\boldsymbol{\Sigma}^{-1} = \boldsymbol{\Lambda}^{-1}(\boldsymbol{I}_T \otimes \boldsymbol{\Psi}\boldsymbol{\Psi}')\boldsymbol{\Lambda}^{-1}$. The likelihood of $(\boldsymbol{b}, \boldsymbol{\Psi}, \boldsymbol{\Lambda})$ is then

$$\begin{aligned} & [\boldsymbol{y} \mid \boldsymbol{b}, \boldsymbol{\Psi}, \boldsymbol{\Lambda}] \\ \propto\ & |\boldsymbol{\Lambda}|^{-1} etr\Big\{-\frac{1}{2}[\boldsymbol{y} - (\boldsymbol{X}' \otimes \boldsymbol{I}_p)\boldsymbol{b}]'\boldsymbol{\Lambda}^{-1}(\boldsymbol{I}_T \otimes \boldsymbol{\Psi}\boldsymbol{\Psi}')\boldsymbol{\Lambda}^{-1}[\boldsymbol{y} - (\boldsymbol{X}' \otimes \boldsymbol{I}_p)\boldsymbol{b}]\Big\}. \end{aligned} \quad (3.13)$$

3.2.2 Priors

We employ independent priors on the elements of $(\boldsymbol{B}, \boldsymbol{\Psi}, \boldsymbol{a}_0, \boldsymbol{\beta}, \boldsymbol{A}_1, \boldsymbol{\delta})$ and independent priors on the indexes corresponding to the elements to be selected in $(\boldsymbol{B}, \boldsymbol{\Psi}, \boldsymbol{A}_1)$, $(\boldsymbol{\gamma}_b, \boldsymbol{\gamma}_\psi, \boldsymbol{\gamma}_a, \boldsymbol{\gamma}_\delta)$.

(i) Priors of $\boldsymbol{B}$.

(ia) Priors of $\boldsymbol{b}_0 = (b_{10}, \cdots, b_{p0})'$. We assume that the intercept b_{i0} is always included in the model, and also that priors for b_{i0} are independent:

$$b_{i0} \overset{ind}{\sim} N(b_{i0}^0, \xi_{i0}^0). \tag{3.14}$$

(ib) Priors of $\boldsymbol{B}_1 = \{b_{ij}\}_{p\times q}$. Each element b_{ij} is associated with an indicator variable $\gamma_{b,ij}$, and we assume that elements of $\boldsymbol{B}_1$ may be included in the model ($\gamma_{b,ij} = 1$) or may not ($\gamma_{b,ij} = 0$). Independent Bernoulli priors are assumed for the index $\gamma_{b,ij}$: for fixed $p_{b,ij} \in (0, 1)$,

$$P(\gamma_{b,ij} = 1) = 1 - P(\gamma_{b,ij} = 0) = p_{b,ij},\ i = 1, \cdots, p,\ j = 1, \cdots, q. \tag{3.15}$$

For given $\boldsymbol{\gamma}_b = (\gamma_{b,11}, \gamma_{b,12}, \cdots, \gamma_{b,pq})'$, assume that

$$(b_{ij} \mid \gamma_{b,ij}) \overset{ind}{\sim} (1 - \gamma_{b,ij})N(0, \kappa_{b,ij}^2) + \gamma_{b,ij}N(0, c_{b,ij}^2\kappa_{b,ij}^2), \tag{3.16}$$

for $i = 1, \cdots, p$ and $j = 1, \cdots, q$, 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 $\boldsymbol{D}_{b,j} = diag((\eta_{b,1j}\kappa_{b,1j})^2, \cdots, (\eta_{b,pj}\kappa_{b,pj})^2)$, then (3.16) is equivalent to

$$(\boldsymbol{b}_j \mid \boldsymbol{\gamma}_{b,j}) \overset{ind}{\sim} N_p(\boldsymbol{0},\ \boldsymbol{D}_{b,j}), \quad \text{for } j = 1, \cdots, q. \tag{3.17}$$

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

$$(\boldsymbol{b} \mid \boldsymbol{\gamma}_b) \sim N(\bar{\boldsymbol{b}}, \bar{\boldsymbol{\Xi}}), \tag{3.18}$$

where

$$\begin{aligned} \bar{\boldsymbol{b}} &= (b_{10}^0, \cdots, b_{p0}^0, 0, \cdots, 0)', \\ \bar{\boldsymbol{\Xi}} &= diag(\xi_{10}^0, \cdots, \xi_{p0}^0, (\eta_{b,11}\kappa_{b,1})^2, \cdots, (\eta_{b,pq}\kappa_{b,pq})^2). \end{aligned}$$

(ii) Priors of $\boldsymbol{\Psi}$. For $j = 2, \cdots, p$, let $\boldsymbol{\psi}_j$ be a vector containing the non-redundant elements of the j^{th} column of $\boldsymbol{\Psi}$, i.e. $\boldsymbol{\psi}_j = (\psi_{1j}, \cdots, \psi_{j-1,j})'$. Define also $\boldsymbol{\gamma}_{\psi,j} = (\gamma_{\psi,1j}, \cdots, \gamma_{\psi,j-1,j})'$, a vector of lenght $j-1$ containind model indicators for $\boldsymbol{\psi}_j$. Let the model index for ψ_{ij}, $\gamma_{\psi,ij}$, be independent Bernoulli $(p_{\psi,ij})$ random variables: for fixed $p_{\psi,ij} \in (0,1)$,

$$P(\gamma_{\psi,ij} = 1) = 1 - P(\gamma_{\psi,ij} = 0) = p_{\psi,ij},\ i = 1, \cdots, j-1,\ j = 1, \cdots, p. \tag{3.19}$$

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

$$(\psi_{ij} \mid \gamma_{\psi,ij}) \overset{ind}{\sim} (1-\gamma_{\psi,ij})N(0, \kappa^2_{\psi,ij}) + \gamma_{\psi,ij}N(0, c^2_{\psi,ij}\kappa^2_{\psi,ij}), \tag{3.20}$$

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

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

and $\boldsymbol{D}_{\psi,j} = diag((\eta_{\psi,1j}\kappa_{\psi,1j})^2, \cdots, (\eta_{\psi,j-1,j}\kappa_{\psi,j-1,j})^2)$, then (3.20) is equivalent to

$$(\boldsymbol{\psi}_j \mid \boldsymbol{\gamma}_{\psi,j}) \overset{ind}{\sim} N_{j-1}(\mathbf{0},\ \boldsymbol{D}_{\psi,j}), \tag{3.21}$$

for $j = 2, \cdots, p$.

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

(iiia) Priors of $\boldsymbol{a}_0 = (a_{10}, \cdots, a_{p0})'$. For fixed $(\bar{a}_{j0}, \sigma_a)$, assume that

$$(a_{j0}) \overset{ind}{\sim} N\Big(\bar{a}_{j0}, \sigma_a\Big). \tag{3.22}$$

(iiib) Priors of $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)'$. For fixed $(\bar{\beta}_j, \sigma_\beta)$, assume that

$$(\beta_j) \overset{ind}{\sim} N\Big(\bar{\beta}_j, \sigma_\beta\Big). \tag{3.23}$$

(iiic) Priors of $\boldsymbol{A}_1$. Let the model index for a_{jk}, $\gamma_{a,jk}$, be independent Bernoulli $(p_{a,jk})$: for fixed $p_{a,jk} \in (0,1)$, $j = 1, \cdots, p$ and $k = 1, \cdots, r$

$$P(\gamma_{a,jk} = 1) \;=\; p_{a,jk},\; P(\gamma_{a,jk} = 0) = 1 - p_{a,jk},. \tag{3.24}$$

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

$$(a_{jk} \mid \gamma_{a,jk}) \;\overset{ind}{\sim}\; (1 - \gamma_{a,jk})N(0, \kappa_{a,jk}^2) + \gamma_{a,jk}N(0, c_{a,jk}^2\kappa_{a,jk}^2). \tag{3.25}$$

where $\kappa_{a,jk}$ would be small and $c_{a,jk}$ be would large constants.

We write $\boldsymbol{A}_1$ in terms of its row vectors: $\boldsymbol{A}_1 = (\widetilde{\boldsymbol{a}}_1', \cdots, \widetilde{\boldsymbol{a}}_p')'$. Here $\widetilde{\boldsymbol{a}}_j = (a_{j1}, \cdots, a_{jr})'$, $j = 1, \cdots, p$. Denote

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

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

$$(\widetilde{\boldsymbol{a}}_j \mid \boldsymbol{\gamma}_{a,j}) \overset{ind}{\sim} N_r(\boldsymbol{0}, \boldsymbol{D}_{a,j}). \tag{3.26}$$

Define $\widetilde{\boldsymbol{a}}_j^* = (a_{j0}, \beta_j, \widetilde{\boldsymbol{a}}_j')'$: combining (3.22) and (3.26) we can write

$$(\widetilde{\boldsymbol{a}}_j^* \mid \boldsymbol{\gamma}_{a,j}) \overset{ind}{\sim} N_{r+2}(\bar{\boldsymbol{a}}_j, \boldsymbol{\Omega}_j), \tag{3.27}$$

where

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

(iv) Prior for $\boldsymbol{\delta}$ Let the model index for δ_j, $\gamma_{\delta j}$, be independent Bernoulli $(p_{\delta j})$ random variables: for fixed $p_{\delta j} \in (0,1)$,

$$P(\gamma_{\delta j} = 1) \;=\; p_{\delta j},\; P(\gamma_{\delta j} = 0) = 1 - p_{\delta j},\; j = 1, \cdots, p. \tag{3.28}$$

For given $\gamma_{\delta j}$ $(j = 1, \cdots, p)$, we assume that δ_j is a mixture of independent Inverse Gamma, $\delta_j^2 \sim \mathbf{IG}(v_{jo}; q_j s_{j0})$ with probability $\gamma_{\delta j}$ and $\delta_j^2 \sim \mathbf{IG}(v_{jo}; s_{j0})$ with probability $1 - \gamma_{\delta j}$. In other words,

$$\delta_j^2 \mid \gamma_{\delta j}, q_j \overset{ind}{\sim} \mathbf{IG}(v_{jo}; \eta_{\delta j} s_{j0}) \propto (\eta_{\delta j} s_{j0})^{v_{jo}} (\delta_j^2)^{-(v_{jo}+1)} \exp(-\frac{\eta_{\delta j} s_{j0}}{\delta_j^2}), \tag{3.29}$$

where

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

where the scale parameter v_{jo} is a given positive constants and is larger than 2, the shape parameter s_{j0} is a small positive constant so that the mean and the variance of the prior with $\gamma_{\delta j} = 1$, $q_j s_{j0}/(v_{jo} - 1)$ and $(q_j s_{j0})^2/(v_{jo} - 1)^2(v_{jo} - 2)$, are large, while prior mean and variance corresponding to $\gamma_{\delta j} = 0$, $s_{j0}/(v_{jo} - 1)$ and $s_{j0}^2/(v_{jo} - 1)^2(v_{jo} - 2)$, are close to zero. We impose a diffuse prior on $\boldsymbol{q}$:

$$q_j \sim Ga(\alpha_q, \; \beta_q), \tag{3.30}$$

for all $j's$, so that each single δ_j will have a different data driven posterior value of q_j.

3.3 Posterior Computation

The joint posterior of $(\boldsymbol{B}, \boldsymbol{\gamma}_b, \boldsymbol{\Psi}, \boldsymbol{\gamma}_\psi, \boldsymbol{a}_0, \boldsymbol{\beta}, \boldsymbol{A}_1, \boldsymbol{\gamma}_a, \boldsymbol{\delta}, \boldsymbol{\gamma}_\delta, \boldsymbol{q}, \boldsymbol{\Lambda})$ has the form,

$$\begin{aligned} [\boldsymbol{B}, \boldsymbol{\gamma}_b, \boldsymbol{\Psi}, \boldsymbol{\gamma}_\psi, \boldsymbol{a}_0, \boldsymbol{\beta}, \boldsymbol{A}_1, \boldsymbol{\delta}, \boldsymbol{\gamma}_a, \boldsymbol{\Lambda} \mid \boldsymbol{y}] \;\; &\propto \;\; [\boldsymbol{y} \mid \boldsymbol{B}, \boldsymbol{\Psi}, \boldsymbol{\Lambda}] \; [\boldsymbol{\Lambda} \mid \boldsymbol{a}_0, \boldsymbol{\beta}, \boldsymbol{A}_1, \boldsymbol{\delta}] \\ &\times \;\; [\boldsymbol{b}_0] \; [\boldsymbol{B}_1 \mid \boldsymbol{\gamma}_b] \; [\boldsymbol{\gamma}_b] \; [\boldsymbol{\Psi} \mid \boldsymbol{\gamma}_\psi] \; [\boldsymbol{\gamma}_\psi] \; [\boldsymbol{a}_0, \boldsymbol{\beta}] \; [\boldsymbol{A}_1 \mid \boldsymbol{\gamma}_a] \; [\boldsymbol{\gamma}_a] \; [\boldsymbol{\delta} | \boldsymbol{\gamma}_\delta, \boldsymbol{q}] [\boldsymbol{\gamma}_\delta] [\boldsymbol{q}], \end{aligned}$$

where $[\boldsymbol{y} \mid \boldsymbol{B}, \boldsymbol{\Psi}, \boldsymbol{\Lambda}]$ is the likelihood function, given by (3.13), $[\boldsymbol{\Lambda} \mid \boldsymbol{a}_0, \boldsymbol{\beta}, \boldsymbol{A}_1, \boldsymbol{\delta}]$ is given by (3.7), $[\boldsymbol{b}_0]$ is given by (3.14), $[\boldsymbol{B}_1 \mid \boldsymbol{\gamma}_b]$ is given by (3.17), $[\boldsymbol{\gamma}_b]$ is given by (3.15),

$[\boldsymbol{\Psi} \mid \boldsymbol{\gamma}_\psi]$ is given by (3.21), $[\boldsymbol{\gamma}_\psi]$ is given by (3.19), $[\boldsymbol{a}_0]$ is given by (3.22) $[\boldsymbol{\beta}]$ is given by (3.23), $[\boldsymbol{A}_1 \mid \boldsymbol{\gamma}_a]$ is given by (3.26), $[\boldsymbol{\gamma}_a]$ is given by (3.24), $[\boldsymbol{\delta} \mid \boldsymbol{\gamma}_\delta, \boldsymbol{q}]$ is given by (3.29), $[\gamma_\delta]$ is given by (3.28) and $[\boldsymbol{q}]$ is given by (3.30). To use an MCMC algorithm, we now derive the full conditional posteriors for $(\boldsymbol{B}, \boldsymbol{\Psi}, \boldsymbol{\Lambda}, \boldsymbol{a}_0, \boldsymbol{\beta}, \boldsymbol{A}_1, \boldsymbol{\delta}, \boldsymbol{\gamma}_b, \boldsymbol{\gamma}_\psi, \boldsymbol{\gamma}_a, \boldsymbol{\gamma}_\delta, \boldsymbol{q})$.

3.3.1 Conditional Posterior for $\boldsymbol{B}$ and $\boldsymbol{\gamma}_b$

Fact 3.1. *(a) The conditional posterior distribution of* $\boldsymbol{b}$ *given* $(\boldsymbol{\gamma}_b, \boldsymbol{\Psi}, \boldsymbol{\gamma}_\psi, \boldsymbol{a}_0, \boldsymbol{\beta}, \boldsymbol{A}_1, \boldsymbol{\delta}, \boldsymbol{\gamma}_a, \boldsymbol{\Lambda}, \boldsymbol{q}; \boldsymbol{y})$ *depends only on* $(\boldsymbol{\Psi}, \boldsymbol{\Lambda}, \boldsymbol{\gamma}_b; \boldsymbol{y})$ *and has the form,*

$$(\boldsymbol{b} \mid \boldsymbol{\Psi}, \boldsymbol{\Lambda}, \boldsymbol{\gamma}_b; \boldsymbol{y}) \sim N_m(\widehat{\boldsymbol{b}}, \widehat{\boldsymbol{\Xi}}), \tag{3.31}$$

where

$$\widehat{\boldsymbol{\Xi}} = \Big\{(\boldsymbol{X} \otimes \boldsymbol{I}_p)\boldsymbol{\Lambda}^{-1}(\boldsymbol{I}_T \otimes \boldsymbol{\Psi}\boldsymbol{\Psi}')\boldsymbol{\Lambda}^{-1}(\boldsymbol{X}' \otimes \boldsymbol{I}_p) + \bar{\boldsymbol{\Xi}}^{-1}\Big\}^{-1}, \tag{3.32}$$

$$\widehat{\boldsymbol{b}} = \widehat{\boldsymbol{\Xi}}\Big\{(\boldsymbol{X} \otimes \boldsymbol{I}_p)\boldsymbol{\Lambda}^{-1}(\boldsymbol{I}_T \otimes \boldsymbol{\Psi}\boldsymbol{\Psi}')\boldsymbol{\Lambda}^{-1}\boldsymbol{y} + \bar{\boldsymbol{\Xi}}^{-1}\bar{\boldsymbol{b}}\Big\}. \tag{3.33}$$

(b) Denote $\boldsymbol{\gamma}_{b,(-ij)} = (\gamma_{b,kl} : (k,l) \neq (i,j))$. *Given prior independence for* b_{ij}, *the conditional posterior distribution of* $\boldsymbol{\gamma}_b$ *given* $(\boldsymbol{B}, \boldsymbol{\Psi}, \boldsymbol{\Lambda}, \boldsymbol{a}_0, \boldsymbol{\beta}, \boldsymbol{A}_1, \boldsymbol{\delta}, \boldsymbol{\gamma}_{b,(-ij)}, \boldsymbol{\gamma}_\psi, \boldsymbol{\gamma}_a, \boldsymbol{q}; \boldsymbol{y})$ *depends only on* $\boldsymbol{B}_1$,

$$(\gamma_{b,ij} \mid \boldsymbol{B}_1) = (\gamma_{b,ij} \mid b_{ij}) \overset{ind}{\sim} \; Bernoulli\Big(\frac{u_{b,ij1}}{u_{b,ij1} + u_{b,ij2}}\Big), \tag{3.34}$$

where

$$\begin{aligned} u_{b,ij1} &= \frac{1}{c_{b,ij}} \exp\Big(-\frac{b_{ij}^2}{2c_{b,ij}^2\kappa_{b,ij}^2}\Big)p_{b,ij}, \\ u_{b,ij2} &= \exp\Big(-\frac{b_{ij}^2}{2\kappa_{b,ij}^2}\Big)(1 - p_{b,ij}). \end{aligned} \tag{3.35}$$

Proof. Using the likelihood (3.13) part (a) is obvious. For part (b), recall that $\boldsymbol{\gamma}_b$ depends on data indirectly, then,

$$\begin{aligned} u_{b,ij1} &\propto [\boldsymbol{b} \mid \boldsymbol{\gamma}_{b,(-ij)}, \gamma_{b,ij} = 1] p_{b,ij}, \\ u_{b,ij2} &\propto [\boldsymbol{b} \mid \boldsymbol{\gamma}_{b,(-ij)}, \gamma_{b,ij} = 0](1 - p_{b,ij}); \end{aligned}$$

the expression above, given prior independence of b_{ij}, gives the formula (3.35). □

It should be noted that, for computational purposes, we do not adoperate the previous formulas for MCMC simulations, but we recompose $\boldsymbol{\Sigma}_t$ in order not to deal with matrices of dimension proportional to T, as we did in (2.51) and (2.52).

3.3.2 Conditional Posterior for $\boldsymbol{\Psi}$ and $\boldsymbol{\gamma}_\psi$

To derive conditional distributions of $\boldsymbol{\Psi}$, we use the likelihood function (3.9) of $(\boldsymbol{B}, \boldsymbol{\Psi}, \boldsymbol{\Lambda}_1, \cdots, \boldsymbol{\Lambda}_T)$. Given $\boldsymbol{B} = (\boldsymbol{b}_0, \boldsymbol{B}_1)$, we know from (2.21) that $\boldsymbol{S}_t = \boldsymbol{S}_t(\boldsymbol{B})$ represents the covariance of residuals $\boldsymbol{e}_t$. Following Harville (1997) we write

$$tr(\boldsymbol{S}_t \boldsymbol{\Lambda}_t^{-1} \boldsymbol{\Psi} \boldsymbol{\Psi}' \boldsymbol{\Lambda}_t^{-1}) = tr(\boldsymbol{\Psi}' \boldsymbol{\Lambda}_t^{-1} \boldsymbol{S}_t \boldsymbol{\Lambda}_t^{-1} \boldsymbol{\Psi}).$$

Then, we can write (3.9) as

$$[\boldsymbol{Y} \mid \boldsymbol{b}_0, \boldsymbol{B}_1, \boldsymbol{\Psi}, \boldsymbol{\Lambda}_1, \cdots, \boldsymbol{\Lambda}_T] = \prod_{t=1}^{T} |\boldsymbol{\Lambda}_t| etr\Big\{ -\frac{1}{2} \sum_{t=1}^{T} \boldsymbol{\Psi}' \boldsymbol{R}_t(\boldsymbol{B}, \boldsymbol{\Lambda}_t) \boldsymbol{\Psi} \Big\}, \quad (3.36)$$

where $\boldsymbol{R}_t(\boldsymbol{B}, \boldsymbol{\Lambda}_t) = \boldsymbol{\Lambda}_t^{-1} \boldsymbol{S}_t \boldsymbol{\Lambda}_t^{-1}$. We can now compute the conditional posterior of $\boldsymbol{\Psi}$ similarly as in George et al. (2005). Let $\boldsymbol{R}_{t,j}$ be the upper-left $j \times j$ submatrix of $\boldsymbol{R}_t(\boldsymbol{B}, \boldsymbol{\Lambda}_t)$. So $\boldsymbol{R}_t = \boldsymbol{R}_{t,p}$. We write the $(i,j)^{th}$ component of $\boldsymbol{R}_t(\boldsymbol{B}, \boldsymbol{\Lambda}_t)$ by $r_{t,ij}$. For $j = 2, \cdots, p$, define $\boldsymbol{r}_{t,j} = (r_{t,1j}, \cdots, r_{t,j-1,j})'$. Define

$$v_{t,1} = r_{t,11}, \; v_{t,j} = \frac{|\boldsymbol{R}_{t,j}|}{|\boldsymbol{R}_{t,j-1}|}, \text{ for } j = 2, \cdots, p. \quad (3.37)$$

It is well known that $v_{t,j} = r_{t,jj} - \boldsymbol{r}'_{t,j}\boldsymbol{R}^{-1}_{t,j-1}\boldsymbol{r}_{t,j} > 0$ for $i = 2, \cdots, p$. We define $\boldsymbol{\psi}_j$ as the $j-1$ elements of jth column of $\boldsymbol{\Psi}$ above the diagonal element. So for $j = 2, \cdots, p$, $\boldsymbol{\psi}_j = (\psi_{1j}, \cdots, \psi_{j-1,j})'$. The likelihood function $[\boldsymbol{y} \mid \boldsymbol{B}, \boldsymbol{\Psi}, \boldsymbol{\Lambda}_1, \cdots, \boldsymbol{\Lambda}_T]$ of (3.36) can then be written as

$$\begin{aligned} & [\boldsymbol{y} \mid \boldsymbol{B}, \boldsymbol{\Psi}, \boldsymbol{\Lambda}_1, \cdots, \boldsymbol{\Lambda}_T] \\ \propto \; & \Big(\prod_{t=1}^{T}\prod_{j=1}^{p}\frac{1}{\lambda_{jt}}\Big)\exp\Big[-\frac{1}{2}\Big\{\sum_{t=1}^{T}\sum_{j=1}^{p}v_{t,j} + \sum_{t=1}^{T}\sum_{j=2}^{p}(\boldsymbol{\psi}_j + \boldsymbol{R}^{-1}_{t,j-1}\boldsymbol{r}_{t,j})' \\ \times \; & \boldsymbol{R}_{t,j-1}(\boldsymbol{\psi}_j + \boldsymbol{R}^{-1}_{t,j-1}\boldsymbol{r}_{t,j})\Big\}\Big]. \end{aligned}$$

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

Fact 3.2. *(a) The conditional posteriors of $\boldsymbol{\psi}_2, \cdots, \boldsymbol{\psi}_p$ given $(\boldsymbol{\gamma}_b, \boldsymbol{B}, \boldsymbol{\gamma}_\psi, \boldsymbol{a}_0, \boldsymbol{\beta}, \boldsymbol{A}_1, \boldsymbol{\delta}, \boldsymbol{\gamma}_a, \boldsymbol{\Lambda}, \boldsymbol{q}; \boldsymbol{y})$ are independent and depend only on $(\boldsymbol{B}, \boldsymbol{\gamma}_\psi, \boldsymbol{\Lambda}; \boldsymbol{y})$. They have the form,*

$$(\boldsymbol{\psi}_j \mid \boldsymbol{B}, \boldsymbol{\gamma}_\psi, \boldsymbol{\Lambda}; \boldsymbol{y}) \overset{ind}{\sim} N_{j-1}(\boldsymbol{\mu}_j, \boldsymbol{\Delta}_j), \tag{3.38}$$

where

$$\begin{aligned} \boldsymbol{\Delta}_j &= \{\sum_{t=1}^{T}\boldsymbol{R}_{t,j-1} + \boldsymbol{D}^{-1}_{\psi,j}\}^{-1}, \\ \boldsymbol{\mu}_j &= -\boldsymbol{\Delta}_j\sum_{t=1}^{T}\boldsymbol{r}_{t,j}. \end{aligned} \tag{3.39}$$

(b) For $j = 2, \cdots, p$ and $i = 1, \cdots, j-1$, given prior independence for $(\psi_{1j}, \cdots, \psi_{j-1,j})$, the conditional posterior of $\gamma_{\psi,ij}$ given $(\boldsymbol{\gamma}_b, \boldsymbol{B}, \boldsymbol{\Psi}, \boldsymbol{\gamma}_{\psi,(-ij)}, \boldsymbol{a}_0, \boldsymbol{\beta}, \boldsymbol{A}_1, \boldsymbol{\delta}, \boldsymbol{\gamma}_a, \boldsymbol{\Lambda}, \boldsymbol{q}; \boldsymbol{y})$ depends only on $\boldsymbol{\Psi}$ and has the form,

$$(\gamma_{\psi,ij} \mid \boldsymbol{\Psi}) = (\gamma_{\psi,ij} \mid \psi_{ij}) \overset{ind}{\sim} Bernoulli\Big(\frac{u_{\psi,ij1}}{u_{\psi,ij1} + u_{\psi,ij2}}\Big), \tag{3.40}$$

where

$$\begin{aligned} u_{\psi,ij1} &= \frac{1}{c_{\psi,ij}}\exp\Big(-\frac{\psi^2_{ij}}{2c^2_{\psi,ij}\kappa^2_{\psi,ij}}\Big)p_{\psi,ij}, \\ u_{\psi,ij2} &= \exp\Big(-\frac{\psi^2_{ij}}{2\kappa^2_{\psi,ij}}\Big)(1 - p_{\psi,ij}). \end{aligned} \tag{3.41}$$

Proof. The conditional posterior density of $(\boldsymbol{\psi}_2, \cdots, \boldsymbol{\psi}_p)$, given $(\boldsymbol{B}, \boldsymbol{\Lambda}_1, \cdots, \boldsymbol{\Lambda}_T, \boldsymbol{\gamma}_\psi;\ \boldsymbol{y})$ has the form

$$\begin{aligned}
& [\boldsymbol{\psi}_2, \cdots, \boldsymbol{\psi}_p \mid \boldsymbol{B}, \boldsymbol{\gamma}_\psi, \boldsymbol{\Lambda}_1, \cdots, \boldsymbol{\Lambda}_T; \boldsymbol{y}] \\
\propto\ & \exp\Big\{-\frac{1}{2}\sum_{j=2}^{p}\sum_{t=1}^{T}(\boldsymbol{\psi}_j + \boldsymbol{R}_{t,j-1}^{-1}\boldsymbol{r}_{t,j})'\boldsymbol{R}_{t,j-1}(\boldsymbol{\psi}_j + \boldsymbol{R}_{t,j-1}^{-1}\boldsymbol{r}_{t,j}) - \frac{1}{2}\sum_{j=2}^{p}\boldsymbol{\psi}_j'\boldsymbol{D}_{\psi,j}^{-1}\boldsymbol{\psi}_j\Big\} \\
\propto\ & \exp\Big\{-\frac{1}{2}\sum_{j=2}^{p}(\boldsymbol{\psi}_j - \boldsymbol{\mu}_j)'\boldsymbol{\Delta}_j^{-1}(\boldsymbol{\psi}_j - \boldsymbol{\mu}_j)\Big\},
\end{aligned}$$

where $\boldsymbol{\mu}_j$ and $\boldsymbol{\Delta}_j$ are defined in (3.39). Part (a) follows from direct computation. For part (b), recall the fact that $\boldsymbol{\gamma}_\psi$ depends on data indirectly, then,

$$\begin{aligned}
u_{\psi,ij1} &\propto [\boldsymbol{\psi}_j \mid \boldsymbol{\gamma}_{\psi,(-ij)}, \gamma_{\psi,ij} = 1]p_{\psi,ij}, \\
u_{\psi,ij2} &\propto [\boldsymbol{\psi}_j \mid \boldsymbol{\gamma}_{\psi,(-ij)}, \gamma_{\psi,ij} = 0](1 - p_{\psi,ij}).
\end{aligned}$$

These expressions, under prior independence of $(\psi_{1j}, \cdots, \psi_{j-1,j})$, give the formula (3.41).

□

3.3.3 Conditional Posterior of $(\boldsymbol{a}_0, \boldsymbol{\beta}_j)$, $(\boldsymbol{A}_1, \boldsymbol{\gamma}_a)$ and $(\boldsymbol{\delta}, \boldsymbol{\gamma}_\delta, \boldsymbol{q})$

Recall $h_{jt} = \log(\lambda_{jt})$, $\boldsymbol{H} = (h_{jt})_{p\times T}$ and $\widetilde{\boldsymbol{h}}_j = (h_{j1}, \cdots, h_{jT})'$. Recall that $\widetilde{\boldsymbol{a}}_j'$ is the j^{th} row of $\boldsymbol{A}_1$ and also $\boldsymbol{W}_j$ as defined by (2.58). The conditional posterior of $\boldsymbol{a}_0, \boldsymbol{\beta}, \boldsymbol{A}_1, \boldsymbol{\gamma}_a$ and $\boldsymbol{\delta}$ are as follows:

Fact 3.3. *(a) The conditional posterior distributions of* $\widetilde{\boldsymbol{a}}_j^* =_(a_{j0}, \beta_j, \widetilde{\boldsymbol{a}}_j')', j = 1, \cdots, p$ *given* $(\boldsymbol{\gamma}_b, \boldsymbol{B}, \boldsymbol{\gamma}_\psi, \boldsymbol{\Psi}, \boldsymbol{\delta}, \boldsymbol{\gamma}_a, \boldsymbol{H}, \boldsymbol{q}; \boldsymbol{y})$ *are independent and depend only on* $(\boldsymbol{H}, \boldsymbol{\gamma}_a, \boldsymbol{\delta})$*,*

$$(\widetilde{\boldsymbol{a}}_j^* \mid \boldsymbol{H}, \boldsymbol{\gamma}_a, \boldsymbol{\delta}) = (\widetilde{\boldsymbol{a}}_j^* \mid \widetilde{\boldsymbol{h}}_j, \boldsymbol{\gamma}_{a,j}, \delta_j) \overset{ind}{\sim} N_{r+2}(\boldsymbol{\nu}_j, \boldsymbol{\Upsilon}_j), \tag{3.42}$$

where

$$\begin{aligned}
\boldsymbol{\nu}_j &= \{\delta_j^{-2}\boldsymbol{W}_j'\boldsymbol{W}_j + \boldsymbol{\Omega}_j^{-1}\}^{-1}(\delta_j^{-2}\boldsymbol{W}_j'\widetilde{\boldsymbol{h}}_j + \boldsymbol{\Omega}_j^{-1}\bar{\boldsymbol{a}}_j), \\
\boldsymbol{\Upsilon}_j &= \{\delta_j^{-2}\boldsymbol{W}_j'\boldsymbol{W}_j + \boldsymbol{\Omega}_j^{-1}\}^{-1}.
\end{aligned}$$

(b) The conditional posteriors of $(\delta_1, \cdots, \delta_p)$ *given* $(\boldsymbol{\gamma}_\delta, \boldsymbol{\gamma}_b, \boldsymbol{B}, \boldsymbol{\gamma}_\psi, \boldsymbol{\Psi}, \boldsymbol{a}_0, \boldsymbol{\beta}, \boldsymbol{A}_1, \boldsymbol{\gamma}_a, \boldsymbol{H}, \boldsymbol{q}; \boldsymbol{y})$ *are mutually independent and depend only on* $(\boldsymbol{\gamma}_\delta, \boldsymbol{H}, \boldsymbol{a}_0, \boldsymbol{\beta}, \boldsymbol{A}_1, \boldsymbol{q})$*,*

$$\begin{aligned}(\delta_j^2 \mid \boldsymbol{\gamma}_\delta, \boldsymbol{H}, \boldsymbol{a}_0, \boldsymbol{\beta}, \boldsymbol{A}_1) &= (\delta_j^2 \mid \gamma_{\delta j}, \widetilde{\boldsymbol{h}}_j, a_{j0}, \beta_j, \widetilde{\boldsymbol{a}}_j) \\ &\overset{ind}{\sim} \mathbf{IG}\Big(v_{j0} + \frac{T-2}{2}, \eta_{\delta j}^2 s_{j0}^2 + \frac{(\widetilde{\boldsymbol{h}}_j - \boldsymbol{W}_j \widetilde{\boldsymbol{a}}_j^*)'(\widetilde{\boldsymbol{h}}_j - \boldsymbol{W}_j \widetilde{\boldsymbol{a}}_j^*)}{2}\Big); \end{aligned} \tag{3.43}$$

(c) For $j = 1, \cdots, p$*, the conditional posterior of* $\gamma_{\delta j}$ *given* $(\boldsymbol{\gamma}_b, \boldsymbol{B}, \boldsymbol{\Psi}, \boldsymbol{\gamma}_a, \boldsymbol{a}_0, \boldsymbol{\beta}, \boldsymbol{A}_1, \boldsymbol{\delta}, \boldsymbol{\gamma}_\psi, \boldsymbol{H}, \boldsymbol{q}; \boldsymbol{y})$ *depends only on* (δ_j, q_j)*, and has the form*

$$(\gamma_{\delta j} \mid \delta_j, q_j) \overset{ind}{\sim} Bernoulli\Big(\frac{u_{\delta j1}}{u_{\delta j1} + u_{\delta j2}}\Big), \tag{3.44}$$

where

$$\begin{aligned} u_{\delta j1} &= \exp\Big(-\frac{q_j s_{j0}}{\delta_j^2}\Big) p_{\delta j} {q_j}^{v_{j0}}, \\ u_{\delta j2} &= \exp\Big(-\frac{s_{j0}}{\delta_j^2}\Big)(1 - p_{\delta j}). \end{aligned} \tag{3.45}$$

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

$$(\gamma_{a,jk} \mid \boldsymbol{A}_1) = (\gamma_{a,jk} \mid a_{jk}) \overset{ind}{\sim} Bernoulli\Big(\frac{u_{a,jk1}}{u_{a,jk1} + u_{a,jk2}}\Big), \tag{3.46}$$

where

$$\begin{aligned} u_{a,jk1} &= \frac{1}{c_{a,jk}} \exp\Big(-\frac{a_{jk}^2}{c_{a,jk}^2 2\kappa_{a,jk}^2}\Big) p_{a,jk}, \\ u_{a,jk2} &= \exp\Big(-\frac{a_{jk}^2}{2\kappa_{a,jk}^2}\Big)(1 - p_{a,jk}). \end{aligned} \tag{3.47}$$

(e) For $j = 1, \cdots, p$, the conditional posterior of q_j given $(\boldsymbol{\gamma}_b, \boldsymbol{B}, \boldsymbol{\Psi}, \boldsymbol{\gamma}_a, \boldsymbol{a}_0, \boldsymbol{\beta}, \boldsymbol{A}_1, \boldsymbol{\delta}, \boldsymbol{\gamma}_\psi, \boldsymbol{H}, \boldsymbol{\gamma}_\delta; \boldsymbol{y})$ depends only on $(\delta_j, \boldsymbol{\gamma}_{\delta j})$, and has the form

$$(q_j \mid \delta_j, \boldsymbol{\gamma}_{\delta j}) \sim \begin{cases} Ga(\alpha_q,\ \beta_q), & \text{if } \gamma_{\delta j} = 0, \\ Ga(v_{jo} + \alpha_q,\ \frac{s_{j0}}{\delta_j^2} + \beta_q), & \text{if } \gamma_{\delta j} = 1. \end{cases} \tag{3.48}$$

Proof. Parts (a) and (b) can be easily proved using regression theory results. For part (c), recall that $\boldsymbol{\gamma}_\delta$ depends on data indirectly, then,

$$\begin{aligned} u_{\delta j1} &\propto [\widetilde{\delta}_j \mid \boldsymbol{\gamma}_{\delta j} = 1] p_{\delta j}, \\ u_{\delta j2} &\propto [\widetilde{\delta}_j \mid \boldsymbol{\gamma}_{a,jk} = 0](1 - p_{\delta j}). \end{aligned}$$

Substituting the density of Inverse Gamma to the expressions above gives the formula (3.45). Note that the scale parameter cancels out and does not affect the conditional posterior of the model index $\gamma_{\delta j}$.

For part (d), recall that $\boldsymbol{\gamma}_a$ depends on data indirectly, then,

$$\begin{aligned} u_{a,jk1} &\propto [\widetilde{\boldsymbol{a}}_j \mid \boldsymbol{\gamma}_{a,(-jk)}, \boldsymbol{\gamma}_{a,jk} = 1] p_{a,jk}, \\ u_{a,jk2} &\propto [\widetilde{\boldsymbol{a}}_j \mid \boldsymbol{\gamma}_{a,(-jk)}, \boldsymbol{\gamma}_{a,jk} = 0](1 - p_{a,jk}). \end{aligned}$$

The two expressions, together with prior independence of $(a_{j1}, \cdots, a_{jr})$, give the formula (3.47). Part (e) comes from direct computation. □

3.3.4 MCMC Algorithm

From the result of the previous section, we can define the Gibbs MCMC sampling procedure as follows. Suppose in cycle k we have $(\boldsymbol{\Psi}^{(k-1)}, \boldsymbol{B}^{(k-1)}, \boldsymbol{a}_0^{(k-1)}, \boldsymbol{\beta}^{(k-1)}, \boldsymbol{A}_1^{(k-1)}$ $\boldsymbol{\delta}^{(k-1)}, \boldsymbol{\Lambda}^{(k-1)})$. Then have the following algorithm:

Step 1: Draw $(\boldsymbol{\gamma}_\delta^{(k)} \mid \boldsymbol{\delta}^{(k-1)}, \boldsymbol{q}^{(k-1)})$ from (3.45);

Step 2: Draw $(\boldsymbol{q}^{(k)} \mid \boldsymbol{\delta}^{(k-1)}, \boldsymbol{\gamma}_\delta^{(k)})$ from (3.48);

Step 3: Draw $(\boldsymbol{a}_0^{(k)}, \boldsymbol{\beta}^{(k)}, \boldsymbol{A}_1^{(k)})$: for $j = 1, \cdots, p$, draw $(\widetilde{\boldsymbol{a}}_j^{*(k)} \mid \boldsymbol{\Lambda}^{(k-1)}, \boldsymbol{\gamma}_a^{(k-1)}, \boldsymbol{\delta}^{(k-1)})$ from (3.42).

Step 4: Draw $(\boldsymbol{\delta}^{(k)} \mid \boldsymbol{\gamma}_\delta^{(k)}, \boldsymbol{\gamma}_a^{(k-1)}, \boldsymbol{\gamma}_b^{(k-1)}, \boldsymbol{\gamma}_\psi^{(k-1)} \boldsymbol{\Lambda}^{(k-1)}, \boldsymbol{a}_0^{(k)}, \boldsymbol{\beta}^{(k)}, \boldsymbol{A}_1^{(k)}, \boldsymbol{q}^{(k)})$ from the distribution (3.43).

Step 5: For $j = 1, \cdots, p$, draw $(\boldsymbol{\gamma}_{a,j}^{(k)} \mid \boldsymbol{A}_1^{(k)})$ from the Bernoulli distribution (3.47).

Step 6: Draw $(\boldsymbol{\Lambda}^{(k)} \mid \boldsymbol{a}_0^{(k)}, \boldsymbol{\beta}^{(k)}, \boldsymbol{A}_1^{(k)}, \boldsymbol{B}^{(k-1)}, \boldsymbol{\delta}^{(k)}, \boldsymbol{\Psi}^{(k-1)}; \boldsymbol{y})$ using the particle filter algorithm F2 and the likelihood (3.9).

Step 7: Draw $\boldsymbol{\Psi}^{(k)}$: for $j = 2, \cdots, p$, draw $(\boldsymbol{\psi}_j^{(k)} \mid \boldsymbol{\Lambda}^{(k)}, \boldsymbol{B}^{(k-1)}, \boldsymbol{\gamma}_{\psi}^{(k-1)}; \boldsymbol{y})$ from the normal distribution (3.38).

Step 8: Draw $(\boldsymbol{\gamma}_{\psi}^{(k)} \mid \boldsymbol{\Psi}^{(k)})$ from the Bernoulli distribution (3.40).

Step 9: Draw $(\boldsymbol{B}^{(k)} \mid \boldsymbol{\Psi}^{(k)}, \boldsymbol{\Lambda}^{(k)}, \boldsymbol{\gamma}_b^{(k-1)}; \boldsymbol{y})$ from the normal distribution (3.31).

Step 10: Draw $(\boldsymbol{\gamma}_b^{(k)} \mid \boldsymbol{B}^{(k)})$ from the Bernoulli distribution (3.34).

3.4 Non−Nested Model Selection

Different options are available to parameterize $\boldsymbol{\Sigma}_t$. Following Chen & Dunson (2003), we use (3.1). We discussed in the chapter's introduction similarities and differences between (3.1) and (2.7), which has been used in Chapter 2. In this section we perform model selection to compare these two different decompositions. Since we are comparing non-nested models, we cannot use the stochastic search algorithm. Several options are available, including BIC (Schwarz (1978)), DIC (Spiegelhalter et al. (2002)), and Bayes factors (Jeffreys (1935, 1961), see Kass & Raftery (1995) for a discussion). We will use this last technique to compare the two models.

In order to compute a Bayes factor for two competing models, we need to evaluate their joint posterior distributions, which in our case are of unknown form. Following Chib (1995), we estimate the log of the Bayes factor for two non-nested models (M_1 and M_2) as

$$\log[\boldsymbol{y} \mid M_1] - \log[\boldsymbol{y} \mid M_2] = \log[\boldsymbol{y} \mid M_1, \boldsymbol{\theta}_1^*] + \log[\boldsymbol{\theta}_1^* \mid M_1] - \log[\boldsymbol{\theta}_1^* \mid M_1, \boldsymbol{y}]$$

$$-\Big\{\log[\boldsymbol{y} \mid M_2, \boldsymbol{\theta}_2^*] + \log[\boldsymbol{\theta}_2^* \mid M_2] - \log[\boldsymbol{\theta}_2^* \mid M_2, \boldsymbol{y}]\Big\}, \quad (3.49)$$

where $\boldsymbol{\theta}_1^*$ and $\boldsymbol{\theta}_2^*$ are the Bayesian estimates of the parameters of interest of M_1 and M_2, respectively. This estimation technique has been already applied to multivariate stochastic volatility models and resulted quite effective (see Chib et al. (2002)). Note that $[\boldsymbol{\theta}_1^* \mid M_1]$ and $[\boldsymbol{\theta}_2^* \mid M_2]$ can be easily obtained from the prior distributions, while more complex is to efficiently compute the marginal likelihood and the joint posterior densities. The next two subsections are dedicated to this problem.

3.4.1 Posterior Density

We can decompose the posterior density as

$$\begin{aligned}
[\boldsymbol{\theta}^* \mid M, \boldsymbol{y}] &= [\boldsymbol{b}^*, \boldsymbol{\Psi}^*, \boldsymbol{a}_0^*, \boldsymbol{\beta}^*, \boldsymbol{A}_1^*, \boldsymbol{\delta}^* \mid M, \boldsymbol{y}] \\
&= [\boldsymbol{b}^* \mid M, \boldsymbol{y}]\,[\boldsymbol{\Psi}^* \mid M, \boldsymbol{b}^*, \boldsymbol{y}]\,[\boldsymbol{a}_0^*, \boldsymbol{\beta}^*, \boldsymbol{A}_1^* \mid M, \boldsymbol{y}] \\
&\times\ [\boldsymbol{\delta}^* \mid M, \boldsymbol{a}_0^*, \boldsymbol{\beta}^*, \boldsymbol{A}_1^*, \boldsymbol{y}]. \qquad (3.50)
\end{aligned}$$

The density function of each of these posteriors is unknown. We are able, though, to simulate from any of them. Chib & Jeliazkov (2001) suggest a way to estimate the joint posterior densities using reduced MCMC runs: the Gibbs sampler algorithm in Section 3.3 samples from $[\boldsymbol{b} \mid M, \boldsymbol{y}]$ and $[\boldsymbol{a}_0, \boldsymbol{\beta}, \boldsymbol{A}_1 \mid M, \boldsymbol{y}]$. To sample from $[\boldsymbol{\Psi} \mid M, \boldsymbol{b}^*, \boldsymbol{y}]$ and $[\boldsymbol{\delta} \mid M, \boldsymbol{a}_0^*, \boldsymbol{\beta}^*, \boldsymbol{A}_1^*, \boldsymbol{y}]$, fix $\boldsymbol{b} = \boldsymbol{b}^*$, $(\boldsymbol{a}_0, \boldsymbol{\beta}, \boldsymbol{A}_1) = (\boldsymbol{a}_0^*, \boldsymbol{\beta}^*, \boldsymbol{A}_1^*)$ and run the MCMC simulation for another G iterations. Each density, then, can be estimated via kernel smoothing. This method is computationally very intensive. Also, it is affected by the curse of dimensionality of kernel smoothers, that can only be partially overcome by dimension reduction techniques such as factor decomposition.

Alternatively, we can rely on the Blackwell theorem to obtain an estimate of $[\boldsymbol{\theta}^* \mid$

$M, \boldsymbol{y}]$. First, the posterior density of $\boldsymbol{b}^*$ can be obtained as

$$[\boldsymbol{b}^* \mid M, \boldsymbol{y}] \approx \frac{1}{K}\sum_{k=1}^{K}[\boldsymbol{b}^* \mid \boldsymbol{\Psi}^k, \boldsymbol{\Lambda}^k, M, \boldsymbol{y}] \tag{3.51}$$

where $(\boldsymbol{\Psi}^k, \boldsymbol{\Lambda}^k)$ is the k^{th} Gibbs cycle draw of $(\boldsymbol{\Psi}, \boldsymbol{\Lambda})$. The density in the right side of the equation is the full conditional density of $\boldsymbol{b}$ evaluated at $\boldsymbol{b}^*$. Similarly, $[\boldsymbol{\Psi}^*, \boldsymbol{a}_0^*, \boldsymbol{\beta}^*, \boldsymbol{A}_1^*, \boldsymbol{\delta}^* \mid M, \boldsymbol{y}]$ is approximated by

$$\begin{aligned} & [\boldsymbol{\Psi}^*, \boldsymbol{a}_0^*, \boldsymbol{\beta}^*, \boldsymbol{A}_1^*, \boldsymbol{\delta}^* \mid M, \boldsymbol{y}] \\ \approx \; & \frac{1}{K}\sum_{k=1}^{K}[\boldsymbol{\Psi}^* \mid \boldsymbol{b}^*, \boldsymbol{\Lambda}^k, M, \boldsymbol{y}][\boldsymbol{a}_0^*, \boldsymbol{\beta}^*, \boldsymbol{A}_1^* \mid \boldsymbol{\delta}^*, \boldsymbol{\Lambda}^k, M, \boldsymbol{y}][\boldsymbol{\delta}^* \mid \boldsymbol{a}_0^*, \boldsymbol{\beta}^*, \boldsymbol{A}_1^*, \boldsymbol{\Lambda}^k, M, \boldsymbol{y}]. \end{aligned} \tag{3.52}$$

Combining (3.51) and (3.52) we obtain an estimate of (3.50).

3.4.2 Marginal likelihood

In order to obtain an estimate of the marginal likelihood for given $\boldsymbol{\theta}^*$ we need to integrate out the volatilities λ_{jt}. We can obtain an estimate of the marginal likelihood via auxiliary particle filter, similarly as in Pitt & Shephard (1999). This method involves sampling and resampling of particles drawn from a proposal distribution of $\boldsymbol{h}$. The algorithm is as follows:

Step f1. For each $\boldsymbol{h}_{t-1}^l$, $l = 1, \cdots, M$, draw $\boldsymbol{h}_t^l$ from the proposal distribution

$$(\boldsymbol{h}_t \mid \boldsymbol{h}_{t-1}^l, \boldsymbol{\theta}) \sim N_p\Big(\boldsymbol{a}_0 + diag(\boldsymbol{h}_{t-1}^l)\boldsymbol{\beta} + \boldsymbol{A}_1\boldsymbol{z}_t,\ diag(\delta_1^2, \cdots, \delta_p^2)\Big). \tag{3.53}$$

Step f2. For $l = 1, ..M$, Compute $\boldsymbol{\Sigma}_t^l = \boldsymbol{\Gamma}\boldsymbol{\Lambda}_t^l\boldsymbol{\Gamma}'$ and the importance weight for the l^{th} sample

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

Step f3. Sample M draws of $\boldsymbol{h}_t$ from $(\boldsymbol{h}_t^1, \cdots, \boldsymbol{h}_t^M)$ with replacement and relative weight $\boldsymbol{w}_t = (w_t^1, \cdots, w_t^M)$.

Step f4. Choose $R > M$, say $R = 5M$, and sample R draws of index $1, 2, \cdots, M$ with relative weights w_t^l. Denote the result of the draws as $k_1, k_2, \cdots, k_R$.

Step f5. For each r, draw $\boldsymbol{h}_t^{*r}$ from $(\boldsymbol{h}_t \mid \boldsymbol{h}_{t-1}^{k_r}, \boldsymbol{\theta})$.

Step f6. For $q = 1, \cdots, M$, sample $\widetilde{\boldsymbol{h}}^{*q}$ from $\{\boldsymbol{h}_t^{*1}, \boldsymbol{h}_t^{*2}, \cdots, \boldsymbol{h}_t^{*R}\}$ with probabilities proportional to

$$w_{2,t}^r = \frac{[\boldsymbol{y}_t \mid \boldsymbol{h}_t^{*r}, \boldsymbol{\theta}]}{[\boldsymbol{y}_t \mid \boldsymbol{h}_t^{*k_j}, \boldsymbol{\theta}]}, \quad r = 1, \cdots, R.$$

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

Consider the weights w_t^l and $w_{2,t}^r$. It can be shown (see Pitt (2001)) that, given $\boldsymbol{Y}_{t-1}$ (the data available up to time $t-1$), then

$$\widehat{f}(\boldsymbol{y}_t \mid M, \boldsymbol{Y}_{t-1}, \boldsymbol{\theta}^*) = \Big(\frac{1}{M}\sum_{l=1}^{M} f(\boldsymbol{y}_t \mid \boldsymbol{\Sigma}_t^l, \boldsymbol{\theta})\Big)\Big(\frac{1}{R}\sum_{r=1}^{R} w_{2,t}^r\Big), \tag{3.54}$$

converges to $f(\boldsymbol{y}_t \mid M, \boldsymbol{Y}_{t-1}, \boldsymbol{\theta}^*)$ in probability as M and R go to infinity. Combining the estimates for each t, we can obtain an estimate of the likelihood $[\boldsymbol{y} \mid M, \boldsymbol{\theta}^*]$.

3.5 Simulation Studies

In this section, we use simulated data to test the potentialities of the stochastic search variable selection. Suppose we have a p variables VAR(1) model with one exogenous variable and the following parameters: $\psi_{ii} = 1$, $\psi_{1j} = .5$, $\gamma_{\psi,1j} = 1$, for $i = 1, \cdots, p$, and $j = 2, \cdots, p$; $\boldsymbol{b}_0 = (1,\ 1,\ 2,\ 2)'$, $\boldsymbol{b}_1 = (3,\ 1.5,\ 1,\ .75,\ .6,\ .5)'$, $b_{i+1,i} = .5$, $\gamma_{b,i1} = 1$,

$\gamma_{b,i+1,i} = 1$ for $i = 1, \cdots, p$. All the other elements of $\boldsymbol{\Psi}$, $\boldsymbol{\gamma}_\psi$, $\boldsymbol{B}$ and $\boldsymbol{\gamma}_b$ are zero. The stochastic volatilities are generated based on:

$$\boldsymbol{A}_1 = (.1)\mathbf{1}_p, \quad \boldsymbol{\beta} = ((.4)\mathbf{1}'_{p/2}, (.6)\mathbf{1}'_{p/2})', \quad \boldsymbol{\alpha} = (.1)\mathbf{1}_p \;\; and \;\; \boldsymbol{\delta} = (.05)\mathbf{1}_p.$$

We generate the exogenous variables $\boldsymbol{x}$ and $\boldsymbol{z}$ as $x_{it} = \cos(t/2)$, $z_{it} = \sin(t^2)$, for $i = 1$, and $t = 1, \cdots, T$. The sample size is $T = 1,000$. The hyperparameters of the Bernoulli priors are $p = .5$. The prior on the intercepts is normal with zero mean variance of 50. We let $v_{j0} = 6.0$ and $s^2_{j0} = .001$. We set the prior means α_{j0} and β_{j0} at 0 and standard deviation σ_α and σ_β at 10, κ_{ij} at .1, and d_{ij} at 50.

We randomly generate 120 samples from the model above with $p = 4$ and $p = 10$ respectively, and for each sample we estimate the posterior mean of all parameters. Estimates are obtained with $10,000$ MCMC draws after $1,000$ burn-ins. The following results are the average of the posterior means over all samples. When $p = 4$, we present the estimates in matrix form, while when $p = 10$, due to the high dimensionality of the matrixes, we rely on graphical representation of the results.

Example 3.1. Consider the results of the Bayesian estimation and model selection when $p = 4$, and the value of $\boldsymbol{\gamma}_\delta$ is fixed to one, so all the volatilities are considered stochastic. Estimates are the following:

$$\widehat{\boldsymbol{B}} = \begin{pmatrix} 1.035 & 2.933 & 0.491 & 0.019 & -0.029 & 0.031 \\ 1.029 & 1.508 & -0.019 & 0.512 & 0.003 & -0.025 \\ 2.012 & 1.055 & 0.008 & -0.038 & 0.501 & -0.006 \\ 2.095 & 0.704 & 0.007 & 0.024 & -0.002 & 0.470 \end{pmatrix},$$

$$\widehat{\boldsymbol{\Psi}} = \begin{pmatrix} 1.000 & 0.429 & 0.516 & 0.556 \\ 0 & 1.000 & 0.040 & 0.062 \\ 0 & 0 & 1.000 & 0.062 \\ 0 & 0 & 0 & 1.000 \end{pmatrix},$$

$$\widehat{\boldsymbol{\gamma}}_b = \begin{pmatrix} 1.000 & 1.000 & 0.020 & 0.021 & 0.019 \\ 1.000 & 0.020 & 1.000 & 0.020 & 0.021 \\ 1.000 & 0.021 & 0.025 & 1.000 & 0.020 \\ 1.000 & 0.019 & 0.022 & 0.020 & 0.999 \end{pmatrix}, \; \widehat{\boldsymbol{\gamma}}_\psi = \begin{pmatrix} * & .729 & .718 & .900 \\ * & * & .053 & .083 \\ * & * & * & .078 \\ * & * & * & * \end{pmatrix}.$$

$$\begin{aligned} \widehat{\boldsymbol{a}}_0 &= (.098, .106, .093, .100)', \; \widehat{\boldsymbol{\beta}} = (.408, , .376, .638, .602)', \\ \widehat{\boldsymbol{A}}_1 &= (.121, .082, .119, .092)', \; \widehat{\boldsymbol{\delta}} = (.051, .048, .047, .048)'. \end{aligned}$$

The most evident result here is that all the non-zero elements of $\boldsymbol{B}_1$ have been always chosen to be in the model, except one which has been selected 99.9 percent of the time. The zero elements have been chosen on average 2 percent of the time, in line with the results of all the simulations in Chapter 2. On the other hand, the posterior mean of elements of $\boldsymbol{\gamma}_\psi$ is generally lower than the posterior means obtained in Chapter 2, and the estimates of $\boldsymbol{\Psi}$ not as accurate. It should be noticed that $\boldsymbol{\Psi}$s are not directly comparable, since they derive from two different covariance decompositions. The major difference between the estimates of the stochastic volatility parameters here obtained and those from the simulation study in Chapter 2 is that while the previous chapter's algorithm showed to always underestimate or overestimate the parameters, here the estimates are both smaller and larger than the true value. Table 3.1 shows the average over T for the true and estimated values of $\boldsymbol{h}$, for a randomly chosen sample.

Example 3.2. Consider now a more complex model. The parameters are exactly as in Example 3.1, but $p = 10$ and $\boldsymbol{\gamma}_\delta$ is not fixed. Prior hyperparameters and number of MCMC draws are the same as in the previous example. Figure 3.1 reports the standardized errors for vectorized $\boldsymbol{B}$ and $\boldsymbol{\Psi}$, Table 3.2 shows the average posterior mean for $\gamma_{b,ij}$ and $\gamma_{\psi,ij}$ conditional on the true value of the parameter.

Table 3.3 presents the average over T for the true and estimated $\boldsymbol{h}$ over a randomly

Table 3.1: Comparison of True and Estimated Stochastic Volatilities for One Randomly Chosen Sample. $T = 1,000$, $p = 4$.

variable	*mean (true h)*	*posterior mean (h)*	*S.d. (true h)*	*posterior s.d (h)*
1	.1648	.1673	.0572	.0559
2	.1714	.1706	.0539	.0522
3	.1654	.1648	.0541	.0534
4	.2507	.2488	.0633	.0611

Table 3.2: Estimates of the Indicator Variables

	Average estimate when true $= 0$	*Average estimate when true* $= 1$
$\boldsymbol{B}$	0.05235	0.99981
$\boldsymbol{\Psi}$	0.07428	0.69955

chosen sample, while results for the stochastic volatility generation process parameters are given below.

$$
\begin{aligned}
\widehat{\boldsymbol{a}}_0 &= (.0971, .1056, .0923, .1007, .0859, .1053, .1062, .0987, .1012, .1014)', \\
\widehat{\boldsymbol{\beta}} &= (.4109, .3727, .4344, .3952, .4601, .5773, .5732, .6078, .5976, .5852)', \\
\widehat{\boldsymbol{A}}_1 &= (.1020, .1015, .0906, .1053, .1000, .9732, -.1057, .1078, -.1027, -.1250)', \\
\widehat{\boldsymbol{\delta}} &= (.0471, .0506, .0502, .0493, .0491, .0467, .0513, .0508, .0485, .0487)', \\
\widehat{\boldsymbol{\gamma}}_a &= (.9875, .9721, .9663, .9256, .9854, .9721, .9823, .9466, .9774, .9634)', \\
\widehat{\boldsymbol{\gamma}}_\delta &= (.9986, .9566, .9952, .9981, .9927, .9886, .9935, .9746, .9813, .9433)'.
\end{aligned}
$$

Again, the posterior mean of the $\gamma_{b,ij}$s relative to the non-zero elements of $\boldsymbol{B}_1$ is very high, while the opposite can be said for the $\gamma_{\psi,ij}$ relative to the non-zero elements of $\boldsymbol{\Psi}$.

Table 3.3: Comparison of True and Estimated Stochastic Volatilities for One Randomly Chosen Sample. $T = 1,000$, $p = 10$.

variable	*mean* (*true* h)	*posterior mean* (h)	*S.d.* (*true* h)	*posterior s.d* (h)
1	0.163	0.169	0.052	0.052
2	0.169	0.170	0.055	0.054
3	0.164	0.164	0.057	0.055
4	0.248	0.250	0.062	0.061
5	0.250	0.251	0.068	0.067
6	0.248	0.249	0.059	0.058
7	0.247	0.249	0.064	0.063
8	0.248	0.249	0.065	0.064
9	0.248	0.250	0.062	0.060
10	0.242	0.242	0.061	0.060

Also, it is clear from Figure 3.1 how the algorithm overestimates $\boldsymbol{\Psi}$. The parameters of the stochastic volatility process show estimates both larger and smaller than the true values they refer to, and the selection of the elements of both $\boldsymbol{A}_1$ and $\boldsymbol{\delta}$ is quite accurate.

3.6 Comparison of $\boldsymbol{\Sigma}_t$ Decompositions

In this section we report the result of non-nested model selection over simulated and real data. Precisely, we compare two different covariance decompositions: (2.7) and (3.1). First, we sample 50 random samples of size $T = 1,000$ according to (3.6) and (3.1). Then, for each sample we compute the Bayes estimators and perform stochastic search model selection assuming first (3.1) as true, and then (2.7) as true. Finally, we compare the selected model under each assumption by estimating the logarithm of the Bayes factors for the two models as described in Section 3.4. After, we repeat the same

procedure but with data generated according to (2.6) and (2.7).

Consider the model in Example 2.1, with $p = 4$. For both decomposition we use the same structure of $\mathbf{\Psi}$. For each sample, we estimate the model and perform variable selection using $10,000$ MCMC draws after $1,000$ burn-ins. The hyperparameters of the Bernoulli priors are $p = .5$. The prior on the intercepts is normal with mean zero and variance 50. Also, $v_{j0} = 6$ and $s_{j0}^2 = .001$, α_{j0} and β_{j0} are 0, and the prior standard deviations for σ_α and σ_β are 10. We assume $\kappa_{ij} = .1$ and $d_{ij} = 50$.

Let M_1 be the model derived assuming (2.7) and M_2 the model derived assuming (3.1). If the data is generated according to M_1, then

$$[\boldsymbol{y} \mid M_1]/[\boldsymbol{y} \mid M_2] \;=\; 236.4918.$$

If M_2 is the model used for data generation, instead,

$$[\boldsymbol{y} \mid M_1]/[\boldsymbol{y} \mid M_2] \;=\; 768.711^{-1}.$$

The same model, but with $p = 6$, gave the following results: For data generated by M_1

$$[\boldsymbol{y} \mid M_1]/[\boldsymbol{y} \mid M_2] \;=\; 759.2856.$$

If M_2 is true, instead,

$$[\boldsymbol{y} \mid M_1]/[\boldsymbol{y} \mid M_2] \;=\; 1629.187^{-1}.$$

In addition, we compare the same models when estimated without performing variable selection. Table 3.4 shows the results for two unconstrained models with $p = 4$ and $p = 6$.

As we can see, the results are very similar for constrained and unconstrained models: the evidence in favor of the true model is always very strong. More, the evidence in

Table 3.4: Bayes' Factors : $[\boldsymbol{y} \mid M_1]/[\boldsymbol{y} \mid M_2]$

	M_1 *is true*	$M_n 2$ *is true*
$p = 4$	186.7867	709.296^{-1}
$p = 6$	799.7985	1625.708^{-1}

favor of M_2 when M_2 is true is always stronger than the evidence in favor of M_1 when M_1 is true.

We apply the same comparison to real data. The dataset considered is the same as the first real data example of Chapter 2, data consisting of log differences of daily exchange rates of USD with, respectively, EUR, GBP, JPY, CAD, MXN, BRL, CHF, NOK, SGD and INR. Previous empirical studies show that this kind of data exhibits time varying correlations, so we should expect evidence in favor of the decomposition (2.7). Models based on (2.7) and (3.1) are both estimated performing stochastic search. The Bayes factor obtained is $[\boldsymbol{y} \mid M_1]/[\boldsymbol{y} \mid M_2] = 97.567$, thus providing strong evidence in favor of the modified Cholesky decomposition (2.7).[4]

3.7 Conclusions

We presented in this chapter an alternative modified Cholesky decomposition for time varying covariances. The decomposition has been first developed by Chen & Dunson (2003) and applied to time invariant covariances in a mixed model framework. Particularity of this decomposition is the separation between the correlations, which are completely defined by the unit triangular matrix $\boldsymbol{\Gamma}$, and the scale parameters, defined

[4]This particular result will need further study, since some MCMC iterations presented inversion of matrices close to unit.

by $\mathbf{\Lambda}$. Modeling the elements of $\mathbf{\Lambda}$ as time varying according to a stochastic volatility model, we implicitly are imposing time invariance of the correlations. We applied the stochastic search variable selection algorithm to this model as well, and studied its performances using artificially simulated data. Results show the algorithm to be quite effective. If compared with the results obtained with the decomposition of Pourahmadi (1999), they show a higher accuracy in selecting the non-zero elements of $\boldsymbol{B}_1$, and a positive bias in the estimation of the non-zero elements of $\mathbf{\Psi}$. We performed Bayesian model selection between a multivariate stochastic volatility model whose time varying covariance is modeled following Pourahmadi (1999), and a competing model with a covariance modeled according to Chen & Dunson (2003). Because of the implicit assumption of the two decompositions, this can be seen as a way to test whether a particular multivariate time series exhibits time varying correlations. Models are selected using Bayes factor, which are estimated using the method first proposed by Chib (1995). Previous studies had shown the accuracy of such Bayes factor estimate (Chib et al. (2002)). Chib's method requires the estimation of the marginal likelihood and the posterior ordinate of the Bayesian estimator, which have been computed using auxiliary particle filter and the output of the MCMC simulations in combination with the Rao-Blackwell theorem respectively. Application to artificially simulated data show the Bayes factor to strongly support the true model used for data generation. Real data applications have been made using daily exchange rates, resulting in a strong evidence in favor of the time varying correlations model. Extension of this study will be discussed in the chapter relative to future research.

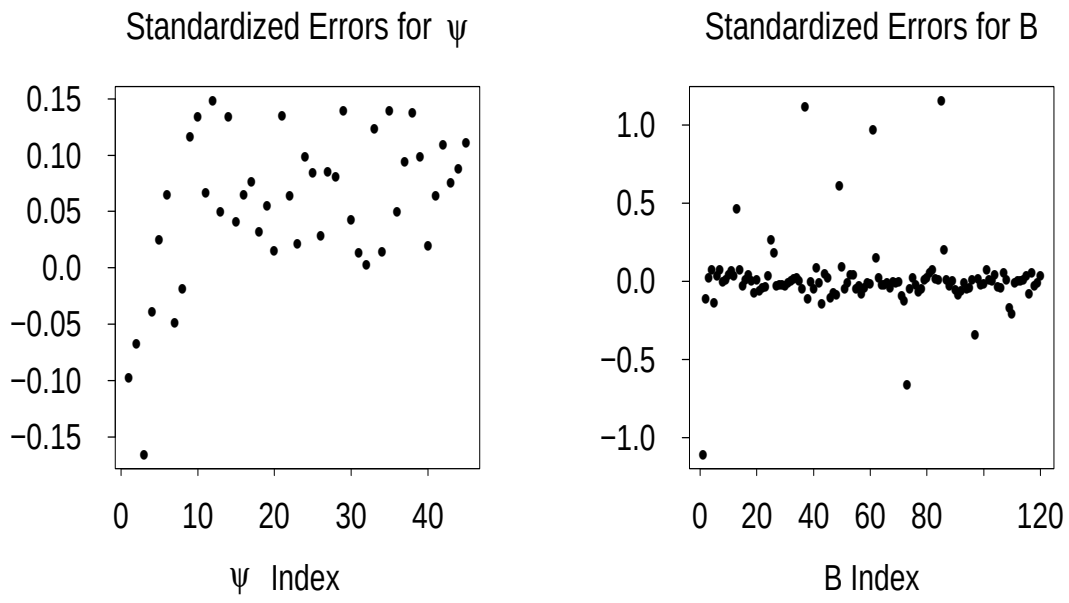

Figure 3.1: Standardized Estimation Errors for B and $\mathbf{\Psi}$. The matrices are vectorized by row and the errors are standardized by the true value range

Chapter 4

Dynamic Models

4.1 Overview

Dynamic models have been studied both from a frequentist (Shumway & Stoffer (2000)) or Bayesian point of view (West & Harrison (1997), Aguilar et al. (1999), Scott (2002)). Accurate reviews of this kind of model can be found in Harvey (1990), (1993) and Shumway & Stoffer (2000).

A particular form of dynamic models are the hidden Markov models. They are characterized by a measured process $\boldsymbol{z}_t$ whose evolution is described by a latent Markovian process $\boldsymbol{y}_t$. It is normally represented in graphical form as in Figure 4.1.

In this chapter, we will concentrate on state space model which, in their most general form, can be written in two stages. The first stage is the so called measurement, or observation, equation:

$$\boldsymbol{z}_t = \boldsymbol{A}_t\boldsymbol{y}_t + \boldsymbol{d}_t + \boldsymbol{e}_t, \tag{4.1}$$

where $t = 1, \cdots, T$, $\boldsymbol{z}_t$, $\boldsymbol{d}_t$ and $\boldsymbol{e}_t$ are $p \times 1$ vectors, $\boldsymbol{y}_t$ is a $q \times 1$ vector and $\boldsymbol{A}_t$ is a $p \times q$

matrix. The second stage is called the state or transition equation:

$$\boldsymbol{y}_t = \boldsymbol{B}_t \boldsymbol{y}_{t-1} + \boldsymbol{c}_t + \boldsymbol{P}_t \boldsymbol{h}_t, \tag{4.2}$$

where $\boldsymbol{h}_t$ is $g \times 1$, $\boldsymbol{P}_t$ is $q \times g$ and $\boldsymbol{B}_t$ is $q \times q$. $\boldsymbol{e}_t$ and $\boldsymbol{h}_t$ are the stochastic terms. In particular,

$$\boldsymbol{e}_t \sim N_p(\boldsymbol{0}, \boldsymbol{\Sigma}_{\mathbf{e},\mathbf{t}}), \tag{4.3}$$

$$\boldsymbol{h}_t \sim N_g(\boldsymbol{0}, \boldsymbol{\Sigma}_{\mathbf{h},\mathbf{t}}), \tag{4.4}$$

where $\boldsymbol{d}_t$, $\boldsymbol{A}_t$, $\boldsymbol{B}_t$, $\boldsymbol{c}_t$ and $\boldsymbol{P}_t$ are non stochastic and are known as system matrices. Also, $\boldsymbol{z}_t$ is the vector of measurements of the process of interest, while $\boldsymbol{y}_t$ is a the true latent process of interest.

If all the system matrices are known, as well as the initial value for the latent variable ($\boldsymbol{y}_o$) and its covariance ($\boldsymbol{\Sigma}_{h,0}$), it is possible to update the system via Kalman filtering (Kalman (1960)). Usually one or several system matrices are unknown, so they need to be estimated. If all the system matrices of the form given above are unknown, the system is clearly indeterminate, so it is impossible to be updated. In the common practice some assumptions are made to restrict the number of unknown quantities, while the remaining are estimated usually via EM algorithm.

Most of the common time series models can be written in a state space form. For example, a VAR(1) model can be written as:

$$\boldsymbol{z}_t = \boldsymbol{\mu}_t + \boldsymbol{e}_t, \qquad \boldsymbol{e}_t \sim N_p(\boldsymbol{0}, \boldsymbol{\Sigma}_{\mathbf{e}}), \tag{4.5}$$

$$\boldsymbol{\mu}_t = \boldsymbol{\Phi}\boldsymbol{\mu}_{t-1} + \boldsymbol{h}_t, \qquad \boldsymbol{h}_t \sim N_p(\boldsymbol{0}, \boldsymbol{\Sigma}_{\mathbf{h}}). \tag{4.6}$$

The two above equations describe a state space model in which $m = t = g$ and $\boldsymbol{y}_t = \boldsymbol{\mu}_t$ for every t, while for $t = 1, \cdots, T$

$$\boldsymbol{A}_t = \boldsymbol{P}_t = I_{p\times p}, \quad \boldsymbol{B}_t = \boldsymbol{\Phi}, \quad \boldsymbol{d}_t = \boldsymbol{c}_t = \boldsymbol{0}, \quad \boldsymbol{\Sigma}_{\mathbf{e},\mathbf{t}} = \boldsymbol{\Sigma}_{\mathbf{e}}, \quad \boldsymbol{\Sigma}_{\mathbf{h},\mathbf{t}} = \boldsymbol{\Sigma}_{\mathbf{h}}.$$

More generally, a VAR(L) model

$$\boldsymbol{z}_t = \boldsymbol{H}_t\boldsymbol{Y}_t + \boldsymbol{e}_t, \qquad \boldsymbol{e}_t \sim N_p(\boldsymbol{0}, \boldsymbol{\Sigma}_{\mathbf{e}}), \tag{4.7}$$

$$\boldsymbol{\mu}_t = \boldsymbol{B}\boldsymbol{Y}_{t-1} + \boldsymbol{h}_t^*, \qquad \boldsymbol{h}_t^* \sim N_Lp(\boldsymbol{0}, \boldsymbol{\Sigma}_{\mathbf{h}^*}), \tag{4.8}$$

can be written as a state space model in which

$$\boldsymbol{Y}_t = (\boldsymbol{y}_t', \cdots, \boldsymbol{y}_{t-L}'), \quad \boldsymbol{h}_t = (\boldsymbol{h}_t', \boldsymbol{0}', \cdots, \boldsymbol{0}')$$

$$\boldsymbol{H}_t = [\boldsymbol{A}_t, \boldsymbol{0}, \cdots, \boldsymbol{0}], \quad \boldsymbol{\Sigma}_h = \begin{pmatrix} \boldsymbol{\Sigma}_h & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}.$$

Usually, in the VAR literature $\boldsymbol{z}_t = \boldsymbol{y}_t$. Obviously, also in this case we can write the model in a state space form by imposing restrictions on $\boldsymbol{A}_t$ and $\boldsymbol{e}_t$. It should also be noticed that stochastic volatility models are often written as a non linear state space model:

$$\boldsymbol{z}_t = \boldsymbol{A}_t e^{\frac{\boldsymbol{y}_t}{2}} \boldsymbol{e}_t, \tag{4.9}$$

$$\boldsymbol{y}_t = \boldsymbol{B}_t\boldsymbol{y}_{t-1} + \boldsymbol{h}_t. \tag{4.10}$$

The state space framework can be used then as a general unifying notation for the study of multivariate time series. Also, they incorporate measurement error and latent variable models. For more details see Harvey (1990) and Shumway & Stoffer (2000) among others.

The problematic inherent model selection in the state space framework has not been discussed as much as that regarding estimation. The stochastic search algorithm proposed in Chapter 2 can be easily extended to state space models, as will be shown in this chapter, which is organized as follows: Section 2 defines the model and Section 3 outlines the likelihood functions. Section 4 describes a well known way to perform Bayesian estimation on state space models, while Section 5 extends the stochastic search methodology

of Chapter 2. Finally, Section 6 compares the two analyses using artificially generated data.

4.2 Model

We consider a latent process of interest and a measured quantity that is stochastically related. Consider the two stage equation:

$$\begin{aligned} \boldsymbol{z}_t &= \boldsymbol{A}\boldsymbol{y}_t + \boldsymbol{e}_t, & \boldsymbol{e}_t \sim N_p(\mathbf{0}, \boldsymbol{\Sigma}_\mathbf{e}), \qquad (4.11) \\ \boldsymbol{y}_t &= \boldsymbol{B}\boldsymbol{y}_{t-1} + \boldsymbol{h}_t, & \boldsymbol{h}_t \sim N_g(\mathbf{0}, \boldsymbol{\Sigma}_\mathbf{h}). \qquad (4.12) \end{aligned}$$

This is a state space model with $\boldsymbol{P}_t = I_{g\times g}$, $\boldsymbol{A}_t = \boldsymbol{A}$, $\boldsymbol{B}_t = \boldsymbol{B}$, $\boldsymbol{\Sigma}_{e,t} = \boldsymbol{\Sigma}_e$, $\boldsymbol{\Sigma}_{h,t} = \boldsymbol{\Sigma}_h$, and $\boldsymbol{c}_t = \boldsymbol{d}_t = \mathbf{0}$ for every t. Note that here $g = m$, so $\boldsymbol{z}_t$ is $p \times 1$, $\boldsymbol{y}_t$ is $g \times 1$, $\boldsymbol{A}$ is $p \times g$ and $\boldsymbol{B}$ is $g \times g$. The fact that $\boldsymbol{c}_t$ is set to zero assumes that $\boldsymbol{z}_t$ is an unbiased measurement of $\boldsymbol{y}_t$. We observe $\boldsymbol{z}_t$ for $t = 1, \cdots, T$, while we estimate all the other quantities.

4.2.1 Likelihood Functions

Consider the likelihood function of $(\boldsymbol{A}, \boldsymbol{\Sigma}_e, \boldsymbol{B}, \boldsymbol{\Sigma}_h)$:

$$\begin{aligned} &\prod_{t=1}^{T} [\boldsymbol{z}_t \mid \boldsymbol{y}_t, \boldsymbol{A}, \boldsymbol{\Sigma}_e]\, [\boldsymbol{y}_t \mid \boldsymbol{y}_{t-1}\boldsymbol{B}, \boldsymbol{\Sigma}_h] \\ &\quad \propto |\boldsymbol{\Sigma}_e|^{-T/2} |\boldsymbol{\Sigma}_h|^{-T/2} \exp\Big\{ -\frac{1}{2} \sum_{t=1}^{T} (\boldsymbol{z}_t - \boldsymbol{A}\boldsymbol{y}_t)' \boldsymbol{\Sigma}_e^{-1} (\boldsymbol{z}_t - \boldsymbol{A}\boldsymbol{y}_t) \\ &\quad - \frac{1}{2} \sum_{t=1}^{T} (\boldsymbol{y}_t - \boldsymbol{B}\boldsymbol{y}_{t-1})' \boldsymbol{\Sigma}_h^{-1} (\boldsymbol{y}_t - \boldsymbol{B}\boldsymbol{y}_{t-1}) \Big\}. \qquad (4.13) \end{aligned}$$

This formula is useful for the derivation of the conditional posteriors of $\boldsymbol{y}_1, \cdots, \boldsymbol{y}_T$. Alternatively, we can rewrite (4.11) in a familiar matrix form

$$\begin{cases} \boldsymbol{Z} = \boldsymbol{A}\boldsymbol{Y} + \boldsymbol{E}, \\ \boldsymbol{Y} = \boldsymbol{B}\boldsymbol{W} + \boldsymbol{H}, \end{cases} \tag{4.14}$$

where

$$\boldsymbol{Z} = (\boldsymbol{z}_1, \cdots, \boldsymbol{z}_T), \quad \boldsymbol{Y} = (\boldsymbol{y}_1, \cdots, \boldsymbol{y}_T), \quad \boldsymbol{W} = (\boldsymbol{y}_0, \cdots, \boldsymbol{y}_{T-1}),$$

$$\boldsymbol{E} = (\boldsymbol{e}_1, \cdots, \boldsymbol{e}_T), \quad \boldsymbol{H} = (\boldsymbol{h}_1, \cdots, \boldsymbol{h}_T).$$

Here $\boldsymbol{Z}$ and $\boldsymbol{E}$ are $p \times T$ matrices, $\boldsymbol{Y}$, $\boldsymbol{W}$ and $\boldsymbol{H}$ are $g \times T$ matrices, and $\boldsymbol{A}$ and $\boldsymbol{B}$ are as before. The likelihood function of $(\boldsymbol{A}, \boldsymbol{\Sigma}_e, \boldsymbol{B}, \boldsymbol{\Sigma}_h)$ is

$$\begin{aligned} [\boldsymbol{Z} \mid \boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{\Sigma}_e]\,[\boldsymbol{Y} \mid \boldsymbol{W}, \boldsymbol{B}, \boldsymbol{\Sigma}_h] \;\propto\; & |\boldsymbol{\Sigma}_e|^{-T/2}|\boldsymbol{\Sigma}_h|^{-T/2}\mathrm{etr}\Big\{-\frac{1}{2}(\boldsymbol{Z}-\boldsymbol{A}\boldsymbol{Y})'\boldsymbol{\Sigma}_e^{-1}(\boldsymbol{Z}-\boldsymbol{A}\boldsymbol{Y}) \\ - \;& \frac{1}{2}(\boldsymbol{Y}-\boldsymbol{B}\boldsymbol{W})'\boldsymbol{\Sigma}_h^{-1}(\boldsymbol{Y}-\boldsymbol{B}\boldsymbol{W})\Big\}. \end{aligned} \tag{4.15}$$

This form is useful in the derivation of the conditional posterior of $\boldsymbol{\Sigma}_e$ and $\boldsymbol{\Sigma}_h$. To derive the conditional posterior of $\boldsymbol{A}$ and $\boldsymbol{B}$, it is useful instead to write the model (4.14) in its vector form:

$$\begin{cases} \boldsymbol{z} = (\boldsymbol{Y}' \otimes \boldsymbol{I}_p)\boldsymbol{a} + \boldsymbol{e}, \\ \boldsymbol{y} = (\boldsymbol{W}' \otimes \boldsymbol{I}_g)\boldsymbol{b} + \boldsymbol{h}, \end{cases} \tag{4.16}$$

where

$$\boldsymbol{z} = vec(\boldsymbol{Z}), \quad \boldsymbol{y} = vec(\boldsymbol{Y}), \quad \boldsymbol{a} = vec(\boldsymbol{A}), \quad \boldsymbol{b} = vec(\boldsymbol{B}).$$

Also

$$\boldsymbol{e} \;\sim\; N_{pg}\Big(\boldsymbol{0}, \widetilde{\boldsymbol{\Sigma}}_{\mathbf{e}}\Big), \tag{4.17}$$

$$\boldsymbol{h} \;\sim\; N_{g^2}\Big(\boldsymbol{0}, \widetilde{\boldsymbol{\Sigma}}_{\mathbf{h}}\Big), \tag{4.18}$$

where

$$\widetilde{\boldsymbol{\Sigma}}_e = \boldsymbol{I}_T \otimes \boldsymbol{\Sigma}_e, \quad \widetilde{\boldsymbol{\Sigma}}_h = \boldsymbol{I}_T \otimes \boldsymbol{\Sigma}_h.$$

Then, the likelihood function of $(\boldsymbol{A}, \boldsymbol{\Sigma}_e, \boldsymbol{B}, \boldsymbol{\Sigma}_h)$ is

$$\begin{aligned} & [\boldsymbol{z} \mid \boldsymbol{y}, \boldsymbol{a}, \boldsymbol{\Sigma}_e]\,[\boldsymbol{y} \mid \boldsymbol{W}, \boldsymbol{b}, \boldsymbol{\Sigma}_h] \\ \propto\ & |\boldsymbol{\Sigma}_e|^{-T/2} |\boldsymbol{\Sigma}_h|^{-T/2} \exp\Big\{ -\frac{1}{2} (\boldsymbol{z} - (\boldsymbol{Y}' \otimes \boldsymbol{I}_p)\boldsymbol{a})' \widetilde{\boldsymbol{\Sigma}}_e^{-1} (\boldsymbol{z} - (\boldsymbol{Y}' \otimes \boldsymbol{I}_p)\boldsymbol{a}) \\ -\ & \frac{1}{2} (\boldsymbol{y} - (\boldsymbol{W}' \otimes \boldsymbol{I}_g)\boldsymbol{b})' \widetilde{\boldsymbol{\Sigma}}_h^{-1} (\boldsymbol{y} - (\boldsymbol{W}' \otimes \boldsymbol{I}_g)\boldsymbol{b}) \Big\}. \end{aligned} \tag{4.19}$$

4.3 Conjugate Analysis

Here we present a well-known Bayesian analysis of the state-space model based on conjugate priors.

4.3.1 Priors

(i) Priors of $\boldsymbol{A}$ and $\boldsymbol{B}$. In practice it convenient to consider $\boldsymbol{a} = vec(\boldsymbol{A})$ and $\boldsymbol{b} = vec(\boldsymbol{B})$, respectively a pg and a g^2 vector. We assume

$$\boldsymbol{a} \sim N_{pg}(a_0, \boldsymbol{\Sigma}_{a,0}), \tag{4.20}$$

$$\boldsymbol{b} \sim N_{g^2}(b_0, \boldsymbol{\Sigma}_{b,0}). \tag{4.21}$$

(ii) Priors of $\boldsymbol{\Sigma}_e$ and $\boldsymbol{\Sigma}_h$. We are using common conjugate priors for the covariance terms:

$$\boldsymbol{\Sigma}_e \sim IW_p(\nu_e, \boldsymbol{\Sigma}_{e,0}), \tag{4.22}$$

$$\boldsymbol{\Sigma}_h \sim IW_g(\nu_h, \boldsymbol{\Sigma}_{h,0}). \tag{4.23}$$

(iii) Prior for $\boldsymbol{y}_0$. We need a prior for the first value of the latent process, $\boldsymbol{y}_0$:

$$\boldsymbol{y}_0 \;\sim\; N_p(\boldsymbol{y}_0^0, \boldsymbol{\Sigma}_0^0). \tag{4.24}$$

4.3.2 Joint Posterior

The joint posterior of $(\boldsymbol{A}, \boldsymbol{\Sigma}_e, \boldsymbol{Y}, \boldsymbol{B}, \boldsymbol{\Sigma}_h)$ has the form

$$\begin{aligned}
&[\boldsymbol{A}, \boldsymbol{\Sigma}_e, \boldsymbol{Y}, \boldsymbol{B}, \boldsymbol{\Sigma}_h \mid \boldsymbol{z}] \\
&\quad\propto \prod_{t=1}^{T}[\boldsymbol{z}_t \mid \boldsymbol{A}, \boldsymbol{y}_t, \boldsymbol{\Sigma}_e] \prod_{t=1}^{T}[\boldsymbol{y}_t \mid \boldsymbol{B}, \boldsymbol{y}_{t-1}, \boldsymbol{\Sigma}_h]\,[\boldsymbol{y}_0]\,[\boldsymbol{A}]\,[\boldsymbol{\Sigma}_e]\,[\boldsymbol{B}]\,[\boldsymbol{\Sigma}_h]\,,
\end{aligned} \tag{4.25}$$

where $[\boldsymbol{z}_t \mid \boldsymbol{A}, \boldsymbol{y}_t, \boldsymbol{\Sigma}_e]$ is given by (4.11), $[\boldsymbol{y}_t \mid \boldsymbol{B}, \boldsymbol{y}_{t-1}, \boldsymbol{\Sigma}_h]$ is given by (4.12), $[\boldsymbol{y}_0]$ is given by (4.24), $[\boldsymbol{A}]$ is given by (4.20), $[\boldsymbol{\Sigma}_e]$ is given by (4.22), $[\boldsymbol{B}]$ is given by (4.21) and $[\boldsymbol{\Sigma}_h]$ is given by (4.23). In order to specify an MCMC algorithm, we now derive the full conditional posteriors for $(\boldsymbol{A}, \boldsymbol{\Sigma}_e, \boldsymbol{Y}, \boldsymbol{B}, \boldsymbol{\Sigma}_h)$.

4.3.3 Conditional Posterior of $\boldsymbol{Y}$

It is convenient to consider conditional distribution of each $\boldsymbol{y}_t$ separately. For our purpose we also need to assume a distributional value for $\boldsymbol{y}_0$. Define $\boldsymbol{Y}_{(-t)} = (\boldsymbol{y}_0, \boldsymbol{y}_1, \cdots, \boldsymbol{y}_{t-1}, \boldsymbol{y}_{t+1}, \cdots, \boldsymbol{y}_T)$, $t = 0, \cdots, T$.

Fact 4.1. *(a) Given $(\boldsymbol{A}, \boldsymbol{\Sigma}_e, \boldsymbol{Y}_{(-t)}, \boldsymbol{B}, \boldsymbol{\Sigma}_h; \boldsymbol{z})$, the posterior distribution of $\boldsymbol{y}_0$ depends only on $(\boldsymbol{y}_1, \boldsymbol{B}, \boldsymbol{\Sigma}_h)$ and has the form*

$$(\boldsymbol{y}_0 \mid \boldsymbol{A}, \boldsymbol{\Sigma}_e, \boldsymbol{y}_1, \boldsymbol{B}, \boldsymbol{\Sigma}_h) \;\sim\; N_g(\boldsymbol{\gamma}_0, \boldsymbol{\Sigma}_0), \tag{4.26}$$

where

$$\begin{aligned}
\boldsymbol{\Sigma}_0 &= \Big\{(\boldsymbol{\Sigma}_0^0)^{-1} + \boldsymbol{B}'\boldsymbol{\Sigma}_h^{-1}\boldsymbol{B}\Big\}^{-1}, \\
\boldsymbol{\gamma}_0 &= \boldsymbol{\Sigma}_0\Big\{\boldsymbol{B}'\boldsymbol{\Sigma}_h^{-1}\boldsymbol{y}_1 + (\boldsymbol{\Sigma}_0^0)^{-1}\boldsymbol{y}_0^0\Big\}.
\end{aligned}$$

(b) Given $(\boldsymbol{A}, \boldsymbol{\Sigma}_e, \boldsymbol{Y}_{(-t)}, \boldsymbol{B}, \boldsymbol{\Sigma}_h; \boldsymbol{z})$*, the posterior distribution of* $\boldsymbol{y}_t$ $(t = 1, \cdots, T-1)$ *depends only on* $(\boldsymbol{A}, \boldsymbol{\Sigma}_e, \boldsymbol{y}_{t-1}, \boldsymbol{y}_{t+1}, \boldsymbol{B}, \boldsymbol{\Sigma}_h; \boldsymbol{z})$ *and has the form*

$$(\boldsymbol{y}_t \mid \boldsymbol{A}, \boldsymbol{\Sigma}_e, \boldsymbol{y}_{t-1}, \boldsymbol{y}_{t+1}, \boldsymbol{B}, \boldsymbol{\Sigma}_h) \;\sim\; N_g(\boldsymbol{\gamma}_t, \boldsymbol{\Sigma}_t), \tag{4.27}$$

where

$$\begin{aligned} \boldsymbol{\Sigma}_t &= \Big\{\boldsymbol{\Sigma}_h^{-1} + \boldsymbol{B}'\boldsymbol{\Sigma}_h^{-1}\boldsymbol{B} + \boldsymbol{A}'\boldsymbol{\Sigma}_e^{-1}\boldsymbol{A}\Big\}^{-1}, \\ \boldsymbol{\gamma}_t &= \boldsymbol{\Sigma}_t\Big\{\boldsymbol{\Sigma}_h^{-1}\boldsymbol{B}\boldsymbol{y}_{t-1} + \boldsymbol{B}'\boldsymbol{\Sigma}_h^{-1}\boldsymbol{y}_{t+1} + \boldsymbol{A}'\boldsymbol{\Sigma}_e^{-1}\boldsymbol{z}_t\Big\}. \end{aligned}$$

(c) Given $(\boldsymbol{A}, \boldsymbol{\Sigma}_e, \boldsymbol{Y}_{(-t)}, \boldsymbol{B}, \boldsymbol{\Sigma}_h; \boldsymbol{z})$*, the posterior distribution of* $\boldsymbol{y}_T$ *depends only on* $(\boldsymbol{A}$, $\boldsymbol{\Sigma}_e, \boldsymbol{y}_{T-1}, \boldsymbol{B}, \boldsymbol{\Sigma}_h; \boldsymbol{z})$ *and has the form*

$$(\boldsymbol{y}_T \mid \boldsymbol{A}, \boldsymbol{\Sigma}_e, \boldsymbol{y}_{T-1}, \boldsymbol{B}, \boldsymbol{\Sigma}_h) \;\sim\; N_g(\boldsymbol{\gamma}_T, \boldsymbol{\Sigma}_T), \tag{4.28}$$

where

$$\begin{aligned} \boldsymbol{\Sigma}_T &= \Big\{\boldsymbol{A}'\boldsymbol{\Sigma}_e^{-1}\boldsymbol{A} + \boldsymbol{\Sigma}_h^{-1}\Big\}^{-1}, \\ \boldsymbol{\gamma}_T &= \boldsymbol{\Sigma}_T\Big\{\boldsymbol{\Sigma}_h^{-1}\boldsymbol{B}\boldsymbol{y}_{T-1} + \boldsymbol{A}'\boldsymbol{\Sigma}_e^{-1}\boldsymbol{z}_T\Big\}. \end{aligned}$$

Proof. For $t = 1, \cdots, T$, we can decompose the posterior of $\boldsymbol{y}_t$ as

$$\begin{aligned} [\boldsymbol{y}_t \mid \boldsymbol{A}, \boldsymbol{\Sigma}_e, \boldsymbol{Y}_{(-t)}, \boldsymbol{B}, \boldsymbol{\Sigma}_h] &= [\boldsymbol{y}_t \mid \boldsymbol{B}, \boldsymbol{y}_{t-1}, \boldsymbol{\Sigma_h}][\boldsymbol{y}_{t+1} \mid \boldsymbol{B}, \boldsymbol{y}_t, \boldsymbol{\Sigma_h}][\boldsymbol{z}_t \mid \boldsymbol{A}, \boldsymbol{y}_t] \\ &\propto \exp\Big\{-\frac{1}{2}(\boldsymbol{y}_t - \boldsymbol{B}\boldsymbol{y}_{t-1})'\boldsymbol{\Sigma}_h^{-1}(\boldsymbol{y}_t - \boldsymbol{B}\boldsymbol{y}_{t-1}) \\ &\quad -\frac{1}{2}(\boldsymbol{y}_{t+1} - \boldsymbol{B}\boldsymbol{y}_t)'\boldsymbol{\Sigma}_h^{-1}(\boldsymbol{y}_{t+1} - \boldsymbol{B}\boldsymbol{y}_t) \\ &\quad -\frac{1}{2}(\boldsymbol{z}_t - \boldsymbol{A}\boldsymbol{z}_t)'\boldsymbol{\Sigma}_e^{-1}(\boldsymbol{z}_t - \boldsymbol{A}\boldsymbol{y}_t)\Big\}. \end{aligned}$$

Part (a) is obvious. Parts (b) and (c) follow from the fact that they are the extreme of the process, *i.e.* $\boldsymbol{y}_{-1}$ and $\boldsymbol{Y}_{T+1}$ do not exist. □

Alternatively, y_t can be filtered using a Kalman filter. Filtering the latent variable y_t is computationally more expensive than the method proposed in this section but leads

to less correlated MCMC draws and so to faster mixing MCMC chains. For reference about filtering inside a Gibbs sampler for state space models see Carter & Kohn (1994).

4.3.4 Conditional Posterior of A and B

We derive the posterior distributions of the vector forms of $\boldsymbol{A}$ and $\boldsymbol{B}$, namely $\boldsymbol{a} = vec(\boldsymbol{A})$ and $\boldsymbol{b} = vec(\boldsymbol{B})$.

Fact 4.2. *(a) The conditional posterior distribution of* $\boldsymbol{a}$ *given* $(\boldsymbol{A}, \boldsymbol{\Sigma}_e, \boldsymbol{Y}, \boldsymbol{\Sigma}_h; \boldsymbol{z})$ *depends only on* $(\boldsymbol{Y}, \boldsymbol{\Sigma}_e, \boldsymbol{z})$ *and has the form,*

$$(\boldsymbol{a} \mid \boldsymbol{Y}, \boldsymbol{\Sigma}_e) \sim N_{pg}(\boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a), \tag{4.29}$$

where

$$\begin{aligned} \boldsymbol{\Sigma}_a &= \Big\{ \boldsymbol{Y}\boldsymbol{Y}' \otimes \boldsymbol{\Sigma}_e^{-1} + (\boldsymbol{\Sigma}_{a,0})^{-1} \Big\}^{-1}, \\ \boldsymbol{\mu}_a &= \boldsymbol{\Sigma}_a \Big\{ (\boldsymbol{Y} \otimes \boldsymbol{\Sigma}_e^{-1})\boldsymbol{z} + \boldsymbol{\Sigma}_{a,0}^{-1}\boldsymbol{b}_0 \Big\}. \end{aligned}$$

(b) The conditional posterior distribution of $\boldsymbol{b}$ *given* $(\boldsymbol{A}, \boldsymbol{\Sigma}_e, \boldsymbol{Y}, \boldsymbol{\Sigma}_h; \boldsymbol{z})$ *depends only on* $(\boldsymbol{Y}, \boldsymbol{\Sigma}_h)$ *and has the form,*

$$(\boldsymbol{b} \mid \boldsymbol{Y}, \boldsymbol{\Sigma}_h) \sim N_{g^2}(\boldsymbol{\mu}_b, \boldsymbol{\Sigma}_b), \tag{4.30}$$

where

$$\begin{aligned} \boldsymbol{\Sigma}_b &= \Big\{ \boldsymbol{W}\boldsymbol{W}' \otimes \boldsymbol{\Sigma}_h^{-1} + (\boldsymbol{\Sigma}_{b,0})^{-1} \Big\}^{-1}, \\ \boldsymbol{\mu}_b &= \boldsymbol{\Sigma}_b \Big\{ (\boldsymbol{W} \otimes \boldsymbol{\Sigma}_h^{-1})\boldsymbol{y} + \boldsymbol{\Sigma}_{b,0}^{-1}\boldsymbol{b}_0 \Big\}. \end{aligned}$$

Proof. From direct computation, we can see that

$$\boldsymbol{\Sigma}_a = \Big\{ (\boldsymbol{Y}' \otimes \boldsymbol{I}_p)' \widetilde{\boldsymbol{\Sigma}}_e^{-1} (\boldsymbol{Y}' \otimes \boldsymbol{I}_p) + (\boldsymbol{\Sigma}_{a,0})^{-1} \Big\}^{-1},$$

$$
\begin{aligned}
\boldsymbol{\mu}_a &= \boldsymbol{\Sigma}_a\Big\{(\boldsymbol{Y}'\otimes \boldsymbol{I}_p)'\widetilde{\boldsymbol{\Sigma}}_e^{-1}\boldsymbol{z} + \boldsymbol{\Sigma}_{a,0}^{-1}\boldsymbol{a}_0\Big\},\\
\boldsymbol{\Sigma}_b &= \Big\{(\boldsymbol{W}'\otimes \boldsymbol{I}_g)'\widetilde{\boldsymbol{\Sigma}}_h^{-1}(\boldsymbol{W}'\otimes \boldsymbol{I}_g) + (\boldsymbol{\Sigma}_{b,0})^{-1}\Big\}^{-1},\\
\boldsymbol{\mu}_b &= \boldsymbol{\Sigma}_b\Big\{(\boldsymbol{W}'\otimes \boldsymbol{I}_g)'\widetilde{\boldsymbol{\Sigma}}_h^{-1}\boldsymbol{y} + \boldsymbol{\Sigma}_{b,0}^{-1}\boldsymbol{b}_0\Big\}.
\end{aligned}
$$

The formulas for part (a) and (b) come from basic properties of the Kronecker operator.

□

4.3.5 Conditional Posterior of $\boldsymbol{\Sigma}_e$ and $\boldsymbol{\Sigma}_h$

The posteriors of $\boldsymbol{\Sigma}_e$ and $\boldsymbol{\Sigma}_h$ can be easily computed from conjugate theory. Write

$$
\boldsymbol{S}_a(\boldsymbol{A}) = (\boldsymbol{Z}-\boldsymbol{Y}\boldsymbol{A})(\boldsymbol{Z}-\boldsymbol{Y}\boldsymbol{A})', \tag{4.31}
$$

$$
\boldsymbol{S}_b(\boldsymbol{B}) = (\boldsymbol{Y}-\boldsymbol{W}\boldsymbol{B})(\boldsymbol{Y}-\boldsymbol{W}\boldsymbol{B})'. \tag{4.32}
$$

Fact 4.3. *(a) The conditional posterior distribution of* $\boldsymbol{\Sigma_e}$ *given* $(\boldsymbol{A},\boldsymbol{B},\boldsymbol{Y},\boldsymbol{\Sigma}_h;\boldsymbol{z})$ *depends only on* $(\boldsymbol{Y},\boldsymbol{A};\boldsymbol{z})$*,*

$$
(\boldsymbol{\Sigma_e} \mid \boldsymbol{Y},\boldsymbol{A};\boldsymbol{z}) \sim IW_p\Big(\nu_e + T, (\boldsymbol{\Sigma}_{e,0} + \boldsymbol{S}_a(\boldsymbol{A}))^{-1}\Big). \tag{4.33}
$$

(b) The conditional posterior distribution of $\boldsymbol{\Sigma}_h$ *given* $(\boldsymbol{A},\boldsymbol{B},\boldsymbol{Y},\boldsymbol{\Sigma}_h;\boldsymbol{z})$ *depends only on* $(\boldsymbol{Y},\boldsymbol{B};\boldsymbol{z})$*,*

$$
(\boldsymbol{\Sigma}_h \mid \boldsymbol{Y},\boldsymbol{B};\boldsymbol{z}) \sim IW_g\Big(\nu_h + T, (\boldsymbol{\Sigma}_{h,0} + \boldsymbol{S}_b(\boldsymbol{B}))^{-1}\Big). \tag{4.34}
$$

Proof. It is easy. □

4.3.6 MCMC Algorithm

Based on the above discussions we have the following Gibbs MCMC sampling procedure. Suppose at the beginning of cycle k, we have $(\boldsymbol{Y}^{(k-1)}, \boldsymbol{A}^{(k-1)}, \boldsymbol{B}^{(k-1)}, \boldsymbol{\Sigma}_e^{(k-1)}, \boldsymbol{\Sigma}_h^{(k-1)})$ given. Then we have the following algorithm:

1. Draw $(\boldsymbol{y}_0^{(k)} \mid \boldsymbol{A}^{(k-1)}, \boldsymbol{\Sigma}_e^{(k-1)}, \boldsymbol{y}_1^{(k-1)}, \boldsymbol{B}^{(k-1)}, \boldsymbol{\Sigma}_h^{(k-1)})$ from (4.26).

2. For $t = 1, \cdots, T-1$, draw $(\boldsymbol{y}_t^{(k)} \mid \boldsymbol{y}_{t-1}^{(k)}, \boldsymbol{y}_{t+1}^{k-1}, \boldsymbol{A}^{(k-1)}, \boldsymbol{B}^{(k-1)}, \boldsymbol{\Sigma}_e^{(k-1)}, \boldsymbol{\Sigma}_h^{(k-1)})$ from (4.27).

3. Draw $(\boldsymbol{y}_T^{(k)} \mid \boldsymbol{A}^{(k-1)}, \boldsymbol{\Sigma}_e^{(k-1)}, \boldsymbol{y}_{T-1}^{(k)}, \boldsymbol{B}^{(k-1)}, \boldsymbol{\Sigma}_h^{(k-1)})$ from (4.28).

4. Draw $(\boldsymbol{A}^{(k)} \mid \boldsymbol{\Sigma}_e^{(k-1)}; \boldsymbol{z})$ from normal distribution (4.29).

5. Draw $(\boldsymbol{B}^{(k)} \mid \boldsymbol{y}^{(k)}, \boldsymbol{\Sigma}_h^{(k-1)})$ from normal distribution (4.30).

6. Draw $(\boldsymbol{\Sigma}_e^{(k)} \mid \boldsymbol{y}^{(k)}; \boldsymbol{z})$ from distribution (4.33).

7. Draw $(\boldsymbol{\Sigma}_h^{(k)} \mid \boldsymbol{y}^{(k)})$ from distribution (4.34).

4.4 Stochastic Search Model Selection for State Space Models

State space models generally involve a very large number of parameters, making the imposition of a priori restrictions very common. Here we take a different perspective on model selection. We do not impose zero restrictions on any coefficient, meaning that we do not rule out any model a priori. Instead, we use a method that allows comparison of all possible models. In order to perform a model selection on the covariance matrix, we decompose the precision matrix in the state space equation (4.12) using a Cholesky decomposition:

$$\boldsymbol{\Sigma}_h^{-1} = \boldsymbol{\Psi}\boldsymbol{\Psi}', \tag{4.35}$$

where $\boldsymbol{\Psi}$ is the $g \times g$ upper-triangular matrix with ψ_{ij} as its $(i,j)th$ entry, so $\psi_{ij} = 0$ for $i > j$. Note that there is no restriction on off-diagonal elements ψ_{ij}, but ψ_{jj} should never change sign, so we can assume that $\psi_{jj} > 0$. Such a matrix $\boldsymbol{\Psi}$ without restriction

on ψ_{ij} is called normalized and just-identified by Sims & Zha (1998).

4.4.1 Model and Prior

Consider again model (4.14). We apply a data driven model selection for the innovation and covariance matrix of the state process, $(\boldsymbol{B}, \boldsymbol{\Sigma}_h)$, inserting in the model latent indicator variables $\boldsymbol{\gamma}_b$ and $\boldsymbol{\gamma}_\psi$.

(i) Priors of $\boldsymbol{A}$. In practice it is convenient to consider $\boldsymbol{a} = vec(\boldsymbol{A})$. We consider

$$\boldsymbol{a} \sim N_{pg}(a_0, \boldsymbol{\Sigma}_{a,0}). \tag{4.36}$$

(ii) Priors of $\boldsymbol{\Sigma}_e$. We are using common conjugate priors for the covariance terms of the measurement process:

$$\boldsymbol{\Sigma}_e \sim W_p^{-1}(\nu_e, \boldsymbol{\Sigma}_{e,0}). \tag{4.37}$$

(iii) Priors of $\boldsymbol{B}$ and $\boldsymbol{\gamma}_b$. Consider $\boldsymbol{B} = (b_{ij})_{g\times g}$. We associate each element b_{ij} with an indicator variable $\gamma_{b,ij}$, and we assume that elements of $\boldsymbol{B}$ may be included in the model ($\gamma_{b,ij} = 1$) or may not ($\gamma_{b,ij} = 0$). Thus, b_{ij} has a two stage prior. We assume independent Bernoulli priors for the index $\gamma_{b,ij}$: for for fixed $p_{b,ij} \in (0, 1)$,

$$P(\gamma_{b,ij} = 1) = 1 - P(\gamma_{b,ij} = 0) = p_{b,ij},\ i = 1, \cdots, g,\ j = 1, \cdots, g. \tag{4.38}$$

For given $\boldsymbol{\gamma}_b = (\gamma_{b,11}, \gamma_{b,12}, \cdots, \gamma_{b,gg})'$, assume that

$$(b_{ij} \mid \gamma_{b,ij}) \overset{ind}{\sim} (1 - \gamma_{b,ij})N(0, \kappa_{b,ij}^2) + \gamma_{b,ij}N(0, c_{b,ij}^2\kappa_{b,ij}^2), \tag{4.39}$$

for $i = 1, \cdots, g$ and $j = 1, \cdots, g$, 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 $\boldsymbol{D}_b = diag((\eta_{b,11}\kappa_{b,11})^2, \cdots, (\eta_{b,gg}\kappa_{b,gg})^2)$, then (4.39) is equivalent to

$$(\boldsymbol{b} \mid \boldsymbol{\gamma}_b) \overset{ind}{\sim} N_{g^2}(\boldsymbol{0},\ \boldsymbol{D}_b), \tag{4.40}$$

(iv) Priors of $\boldsymbol{\Psi}$ and $\boldsymbol{\gamma}_\psi$. For $j = 2, \cdots, g$, let $\boldsymbol{\phi}_j$ be a vector containing the upper diagonal elements of the j^{th} column of $\boldsymbol{\Psi}$, i.e. $\boldsymbol{\phi}_j = (\psi_{1j}, \cdots, \psi_{j-1,j})'$. Also, define a vector of indicators of length $j-1$, $\boldsymbol{\gamma}_{\psi,j} = (\gamma_{\psi,1j}, \cdots, \gamma_{\psi,j-1,j})'$. We assume that elements of $\boldsymbol{\phi}_j$ may be included in the model ($\gamma_{\psi,ij} = 1$) or may 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 $p_{\psi,ij} \in (0,1)$,

$$P(\gamma_{\psi,ij} = 1) = 1 - P(\gamma_{\psi,ij} = 0) = p_{\psi,ij},\ i = 1, \cdots, j-1,\ j = 1, \cdots, g. \tag{4.41}$$

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

$$(\psi_{ij} \mid \gamma_{\psi,ij}) \overset{ind}{\sim} (1-\gamma_{\psi,ij})N(0, \kappa^2_{\psi,ij}) + \gamma_{\psi,ij}N(0, c^2_{\psi,ij}\kappa^2_{\psi,ij}), \tag{4.42}$$

for $i = 1, \cdots, j-1$ and $j = 2, \cdots, g$, 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 $\boldsymbol{D}_{\psi,j} = diag((\eta_{\psi,1j}\kappa_{\psi,1j})^2, \cdots, (\eta_{\psi,j-1,j}\kappa_{\psi,j-1,j})^2)$, then (4.42) is equivalent to

$$(\boldsymbol{\phi}_j \mid \boldsymbol{\gamma}_{\psi,j}) \overset{ind}{\sim} N_{j-1}(\boldsymbol{0},\ \boldsymbol{D}_{\psi,j}), \tag{4.43}$$

for $j = 2, \cdots, g$.

(v) Priors of $\boldsymbol{\psi} = (\psi_{11}, \cdots, \psi_{pp})'$. Assume that $\psi^2_{jj} \overset{ind}{\sim}$ gamma (α_j, β_j) distributions. Here (α_j, β_j) are positive constants. So for $j = 1, \cdots, g$, ψ_{jj} has the density

$$[\psi_{jj}] = \frac{2\beta_j^{\alpha_j}}{\Gamma(\alpha_j)}\psi_{jj}^{2(\alpha_j-1)}\exp(-\beta_j\psi^2_{jj}), \quad \text{for } \psi_{jj} > 0. \tag{4.44}$$

(vi) Prior for $\boldsymbol{y}_0$. We need a prior for the initial value of the latent process, $\boldsymbol{y}_0$:

$$\boldsymbol{y}_0 \;\sim\; N_p(\boldsymbol{y}_0^0, \boldsymbol{\Sigma}_0^0). \tag{4.45}$$

4.4.2 Joint Posterior

The joint posterior of $(\boldsymbol{A}, \boldsymbol{\Sigma}_e, \boldsymbol{Y}, \boldsymbol{B}, \boldsymbol{\Sigma}_h, \boldsymbol{\gamma}_b, \boldsymbol{\gamma}_\psi)$ has the form,

$$\begin{aligned}
&[\boldsymbol{A}, \boldsymbol{\Sigma}_e, \boldsymbol{Y}, \boldsymbol{B}, \boldsymbol{\Sigma}_h, \boldsymbol{\gamma}_b, \boldsymbol{\gamma}_\psi \mid \boldsymbol{z}] \\
&\quad\propto \prod_{t=1}^{T}[\boldsymbol{z}_t \mid \boldsymbol{A}, \boldsymbol{y}_t, \boldsymbol{\Sigma}_e] \prod_{t=1}^{T}[\boldsymbol{y}_t \mid \boldsymbol{B}, \boldsymbol{y}_{t-1}, \boldsymbol{\Sigma}_h]\,[\boldsymbol{y}_0]\,[\boldsymbol{A}]\,[\boldsymbol{\Sigma}_e]\,[\boldsymbol{B} \mid \boldsymbol{\gamma}_b]\,[\boldsymbol{\gamma}_b] \\
&\quad\times \prod_{j=2}^{g}\{[\boldsymbol{\phi}_j \mid \boldsymbol{\gamma}_{\psi,j}]\,[\boldsymbol{\gamma}_{\psi,j}]\}\,[\boldsymbol{\psi}]\,,
\end{aligned} \tag{4.46}$$

where $[\boldsymbol{z}_t \mid \boldsymbol{A}, \boldsymbol{y}_t, \boldsymbol{\Sigma}_e]$ is given by (4.11), $[\boldsymbol{y}_t \mid \boldsymbol{B}, \boldsymbol{y}_{t-1}, \boldsymbol{\Sigma}_h]$ is given by (4.12), $[\boldsymbol{y}_0]$ is given by (4.45), $[\boldsymbol{A}]$ is given by (4.36), $[\boldsymbol{\Sigma}_e]$ is given by (4.22), $[\boldsymbol{B} \mid \boldsymbol{\gamma}_b]$ is given by (4.40), $[\boldsymbol{\gamma}_b]$ is given by (4.38), $[\boldsymbol{\phi}_j \mid \boldsymbol{\gamma}_{\psi,j}]$ is given by (4.43), $[\boldsymbol{\gamma}_{\psi,j}]$ is given by (4.41) and $[\boldsymbol{\psi}]$ is given by (4.44). In order to update an MCMC algorithm, we now derive the full conditional posteriors for $(\boldsymbol{A}, \boldsymbol{\Sigma}_e, \boldsymbol{Y}, \boldsymbol{B}, \boldsymbol{\Sigma}_h, \boldsymbol{\gamma}_b, \boldsymbol{\gamma}_\psi)$.

4.4.3 Conditional Posterior of $(\boldsymbol{A}, \boldsymbol{\Sigma}_e, \boldsymbol{y})$

Fact 4.4. *The conditional posterior distribution of $\boldsymbol{\Sigma_e}$ given $(\boldsymbol{B}, \boldsymbol{Y}, \boldsymbol{\Sigma}_h, \delta; \boldsymbol{z})$ and $\boldsymbol{y}_t$ $(t = 0, \cdots, T)$ given $(\boldsymbol{\Sigma}_e, \boldsymbol{Y}_{(-t)}, \boldsymbol{B}, \boldsymbol{\Sigma}_h, \delta; \boldsymbol{z})$ have the same form than in section 4.3.*

Proof. It is obvious. □

4.4.4 Conditional Posterior for $(\boldsymbol{B}, \boldsymbol{\gamma}_b)$

Fact 4.5. *(a) The conditional posterior distribution of* $\boldsymbol{b}$ *given* $(\boldsymbol{A}, \boldsymbol{\Sigma}_e, \boldsymbol{Y}, \boldsymbol{\Sigma}_h, \boldsymbol{\gamma}_b, \boldsymbol{\gamma}_\psi; \boldsymbol{z})$ *depends only on* $(\boldsymbol{\Sigma}_h, \boldsymbol{\gamma}_b, \boldsymbol{y})$ *and has the form,*

$$(\boldsymbol{b} \mid \boldsymbol{Y}, \boldsymbol{\Sigma}_h) \sim N_{g^2}(\boldsymbol{\mu}_b, \boldsymbol{\Sigma}_b), \tag{4.47}$$

where

$$\begin{aligned} \boldsymbol{\Sigma}_b &= \Big\{ \boldsymbol{W}\boldsymbol{W}' \otimes \boldsymbol{\Sigma}_h^{-1} + (\boldsymbol{D}_b)^{-1} \Big\}^{-1}, \\ \boldsymbol{\mu}_b &= \boldsymbol{\Sigma}_b \Big\{ (\boldsymbol{W} \otimes \boldsymbol{\Sigma}_h^{-1}) \boldsymbol{y} \Big\}. \end{aligned}$$

(b) Denote $\boldsymbol{\gamma}_{b,(-ij)} = (\gamma_{b,kl} : (k,l) \neq (i,j))$. *Given prior independence for* b_{ij}, *the conditional posterior distributions of* $\boldsymbol{\gamma}_b$ *for given* $(\boldsymbol{B}, \boldsymbol{\Psi}, \boldsymbol{\Lambda}, \boldsymbol{a}_0, \boldsymbol{\beta}, \boldsymbol{A}_1, \boldsymbol{\delta}, \boldsymbol{\gamma}_{b,(-ij)}, \boldsymbol{\gamma}_\psi, \boldsymbol{\gamma}_a)$ *depends only on* $\boldsymbol{B}$,

$$(\gamma_{b,ij} \mid \boldsymbol{B}) = (\gamma_{b,ij} \mid b_{ij}) \overset{ind}{\sim} \; Bernoulli\Big(\frac{u_{b,ij1}}{u_{b,ij1} + u_{b,ij2}} \Big), \tag{4.48}$$

where

$$\begin{cases} u_{b,ij1} = \frac{1}{c_{b,ij}} \exp\Big(-\frac{b_{ij}^2}{2c_{b,ij}^2 \kappa_{b,ij}^2} \Big) p_{b,ij}, \\ u_{b,ij2} = \exp\Big(-\frac{b_{ij}^2}{2\kappa_{b,ij}^2} \Big) (1 - p_{b,ij}). \end{cases} \tag{4.49}$$

Proof. Using the likelihood (4.19) part (a) is obvious. For part (b), recall that $\boldsymbol{\gamma}_b$ depends on data indirectly, then,

$$\begin{cases} u_{b,ij1} \propto [\boldsymbol{b} \mid \boldsymbol{\gamma}_{b,(-ij)}, \gamma_{b,ij} = 1] p_{b,ij}, \\ u_{b,ij2} \propto [\boldsymbol{b} \mid \boldsymbol{\gamma}_{b,(-ij)}, \gamma_{b,ij} = 0] (1 - p_{b,ij}); \end{cases}$$

the expression above, given prior independence of b_{ij}, gives the formula (4.49). □

4.4.5 Conditional Posterior for $(\boldsymbol{\Psi}, \boldsymbol{\gamma}_\psi)$

To derive the conditional posterior distribution of $(\boldsymbol{\Psi}, \boldsymbol{\gamma}_\psi)$, we rewrite the likelihood function (4.15) as

$$\begin{aligned}
&[\boldsymbol{Z} \mid \boldsymbol{A}, \boldsymbol{\Sigma}_e, \boldsymbol{Y}, \boldsymbol{B}, \boldsymbol{\Sigma}_h] \\
&\quad \propto \ |\boldsymbol{\Sigma}_e|^{-T/2} |\boldsymbol{\Psi}|^T \mathrm{etr}\Big\{ -\frac{1}{2}\Big((\boldsymbol{Z} - \boldsymbol{AY})' \boldsymbol{\Sigma}_e^{-1} (\boldsymbol{Z} - \boldsymbol{AY}) + \boldsymbol{\Psi}' \boldsymbol{S}_b \boldsymbol{\Psi} \Big\}. \qquad (4.50)
\end{aligned}$$

where $\boldsymbol{S}_b$ is given by (4.32). Then, we use the algorithm of George et al. (2005): write $\boldsymbol{S}_b = (s_{ij})$. For $j = 2, \cdots, g$, define $\boldsymbol{s}_j = (s_{1j}, \cdots, s_{j-1,j})'$. Let $\boldsymbol{S}_j$ be the upper-left $j \times j$ submatrix of $\boldsymbol{S}_b$. It derives that $\boldsymbol{S}_b = \boldsymbol{S}_g$. Define $v_1 = s_{11}$ and $v_j = |\boldsymbol{S}_j| / |\boldsymbol{S}_{j-1}|$ for $j = 2, \cdots, g$. It is well known that $v_j = s_{jj} - \boldsymbol{s}_j' \boldsymbol{S}_{j-1}^{-1} \boldsymbol{s}_i > 0$ for $j = 2, \cdots, g$. Then (4.50) equals to

$$\begin{aligned}
&[\boldsymbol{Y} \mid \boldsymbol{\phi}, \boldsymbol{\Psi}] \\
&\propto \ \prod_{j=1}^{g} \psi_{jj}^T \exp\Big[-\frac{1}{2}\Big\{ \sum_{j=1}^{g} \psi_{jj}^2 v_j + \sum_{j=2}^{g} (\boldsymbol{\phi}_j + \psi_{jj} \boldsymbol{S}_{j-1}^{-1} \boldsymbol{s}_j)' \boldsymbol{S}_{j-1} (\boldsymbol{\phi}_j + \psi_{jj} \boldsymbol{S}_{j-1}^{-1} \boldsymbol{s}_j) \Big\} \Big]. \qquad (4.51)
\end{aligned}$$

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

Fact 4.6. *(a) For given $(\boldsymbol{A}, \boldsymbol{\Sigma}_e, \boldsymbol{Y}, \boldsymbol{B}, \boldsymbol{\gamma}_b, \boldsymbol{\gamma}_\psi, \boldsymbol{\psi}; \boldsymbol{z})$, the posterior distributions of $\boldsymbol{\phi}_2$ $\cdots, \boldsymbol{\phi}_g$ are independent and depend only on $(\boldsymbol{Y}, \boldsymbol{B}, \boldsymbol{\gamma}_\psi, \boldsymbol{\psi})$,*

$$(\boldsymbol{\phi}_j \mid \boldsymbol{B}, \boldsymbol{\gamma}_\psi, \boldsymbol{\psi}, \boldsymbol{Y}) \ \overset{ind}{\sim} \ N_{j-1}(\boldsymbol{\mu}_j, \boldsymbol{\Delta}_j), \qquad (4.52)$$

where

$$\begin{aligned}
\boldsymbol{\mu}_j &= -\psi_{jj} \{ \boldsymbol{S}_{j-1} + (\boldsymbol{D}_{\psi,j})^{-1} \}^{-1} \boldsymbol{s}_j, \qquad (4.53) \\
\boldsymbol{\Delta}_j &= \{ \boldsymbol{S}_{j-1} + (\boldsymbol{D}_{\psi,j})^{-1} \}^{-1}. \qquad (4.54)
\end{aligned}$$

(b) For given $(\boldsymbol{A}, \boldsymbol{\Sigma}_e, \boldsymbol{Y}, \boldsymbol{B}, \boldsymbol{\gamma}_b, \boldsymbol{\gamma}_\psi, \boldsymbol{\phi}_1, \cdots, \boldsymbol{\phi}_p; \boldsymbol{z})$, the posterior distributions of ψ_{jj}, $j = 1, \cdots, g$ are independent and depend only on $(\boldsymbol{Y}, \boldsymbol{B}, \boldsymbol{\gamma}_\psi)$,

$$(\psi_{jj}^2 \mid (\boldsymbol{Y}, \boldsymbol{B}, \boldsymbol{\gamma}_\psi)) \overset{ind}{\sim} gamma\Big(\alpha_j + \frac{1}{2} T, H_j \Big), \qquad (4.55)$$

where

$$H_j \;=\; \begin{cases} \beta_1 + \frac{1}{2}s_{11}, & \text{if } \; j=1, \\ \beta_j + \frac{1}{2}\Big\{ s_{jj} - \boldsymbol{s}_j'[\boldsymbol{S}_{j-1} + (\boldsymbol{D}_{\psi,j})^{-1}]^{-1}\boldsymbol{s}_j \Big\}, & \text{if } \; j=2,\cdots,g. \end{cases} \tag{4.56}$$

(c) For $j = 2,\cdots,g$ *and* $i = 1,\cdots,j-1$, *denote* $\boldsymbol{\gamma}_{\psi,(-i),j} = (\gamma_{\psi,1j},\cdots,\gamma_{\psi,i-1,j},\gamma_{\psi,i+1,j},$ $\cdots,\gamma_{\psi,j-1,j})'$. *For given* $(\boldsymbol{A},\boldsymbol{\Sigma}_e,\boldsymbol{Y},\boldsymbol{B},\boldsymbol{\gamma}_b,\boldsymbol{\gamma}_{\psi,(-i),j},\boldsymbol{\phi}_1,\cdots,\boldsymbol{\phi}_p,\boldsymbol{\psi};\boldsymbol{z})$, $\gamma_{\psi,ij}$ *depends only on* $\boldsymbol{\phi}_j$,

$$(\gamma_{\psi,ij} \mid \boldsymbol{\phi}_j) = (\gamma_{\psi,ij} \mid \psi_{ij}) \overset{ind}{\sim} Bernoulli\Big(\frac{u_{ij1}}{u_{ij1}+u_{ij2}}\Big), \tag{4.57}$$

where

$$\begin{cases} u_{ij1} = \frac{1}{d_{ij}}\exp\Big(-\frac{\psi_{ij}^2}{2d_{ij}^2\kappa_{ij}^2}\Big)p_{\psi,ij}, \\ u_{ij2} = \exp\Big(-\frac{\psi_{ij}^2}{2\kappa_{ij}^2}\Big)(1-p_{\psi,ij}). \end{cases}$$

Proof. For $j = 2,\cdots,g$, the conditional posterior density of $(\psi_{jj}^2,\ j=1,\cdots,g;\boldsymbol{\phi}_j,\ j=2,\cdots,g)$ given $(\boldsymbol{A},\boldsymbol{\Sigma}_e,\boldsymbol{Y},\boldsymbol{B},\boldsymbol{\gamma}_b,\boldsymbol{\gamma}_\psi;\boldsymbol{z})$ is given by

$$\begin{aligned} & [\psi_{jj}^2, j=1,\cdots,g;\boldsymbol{\phi}_j, j=2,\cdots,g \mid \boldsymbol{\phi},\boldsymbol{\omega};\boldsymbol{Y}] \\ \propto\ & \Big\{\prod_{j=1}^{g}(\psi_{jj}^2)^{\alpha_j+\frac{T}{2}-1}\Big\}\exp\Big[-\Big\{\sum_{j=1}^{g}\psi_{jj}^2(H_j)\Big\}\Big]\exp\Big\{-\frac{1}{2}\sum_{j=2}^{g}\boldsymbol{\phi}_j'\boldsymbol{D}_{\psi,j}^{-1}\boldsymbol{\phi}_j\Big\} \\ \times\ & \exp\Big\{-\frac{1}{2}\sum_{j=2}^{g}(\boldsymbol{\phi}_j+\psi_{jj}\boldsymbol{S}_{j-1}^{-1}\boldsymbol{s}_j)'\boldsymbol{S}_{j-1}(\boldsymbol{\eta}_j+\psi_{jj}\boldsymbol{S}_{j-1}^{-1}\boldsymbol{s}_j)\Big\} \\ =\ & \Big\{\prod_{j=1}^{g}(\psi_{jj}^2)^{\alpha_j+\frac{T}{2}-1}\Big\}\exp\Big\{-\sum_{j=1}^{g}\psi_{jj}^2H_i - \frac{1}{2}\sum_{j=2}^{g}(\boldsymbol{\phi}_j-\boldsymbol{\mu}_j)'\boldsymbol{\Delta}_j^{-1}(\boldsymbol{\phi}_j-\boldsymbol{\mu}_j)\Big\}, \end{aligned}$$

where $\boldsymbol{\mu}_j$ is defined in (4.53) and $\boldsymbol{\Delta}_j$ in (4.54). Part (a) is obvious. For part (b), the case when $j=1$ holds clearly; when $j=2,\cdots,g$ the result follows by integrating out $\boldsymbol{\phi}_j$. Part (c) is similar to part (b) of fact 5. □

4.4.6 MCMC Algorithm

Based on the above discussion we design the following Gibbs MCMC sampling procedure. Suppose at the beginning of cycle k, we have $(\boldsymbol{Y}^{(k-1)},\boldsymbol{A}^{(k-1)},\boldsymbol{B}^{(k-1)},\boldsymbol{\Sigma}_e^{(k-1)},$

$\boldsymbol{\Sigma}_h,^{(k-1)}, \boldsymbol{\gamma}_b^{(k-1)}, \boldsymbol{\gamma}_\psi^{(k-1)})$. Then we have the following algorithm:

1. Draw $(\boldsymbol{y}_0^{(k)} \mid \boldsymbol{A}^{(k-1)}, \boldsymbol{\Sigma}_e^{(k-1)}, \boldsymbol{y}_1^{(k-1)}, \boldsymbol{B}^{(k-1)}, \boldsymbol{\Sigma}_h^{(k-1)})$ from (4.26).

2. For $t = 1, \cdots, T-1$, draw $(\boldsymbol{y}_t^{(k)} \mid \boldsymbol{y}_{t-1}^{(k)}, \boldsymbol{y}_{t+1}^{(k-1)}, \boldsymbol{A}^{(k-1)}, \boldsymbol{B}^{(k-1)}, \boldsymbol{\Sigma}_e^{(k-1)}, \boldsymbol{\Sigma}_h^{(k-1)})$ from (4.27).

3. Draw $(\boldsymbol{y}_T^{(k)} \mid \boldsymbol{A}^{(k-1)}, \boldsymbol{\Sigma}_e^{(k-1)}, \boldsymbol{y}_{T-1}^{(k)}, \boldsymbol{B}^{(k-1)}, \boldsymbol{\Sigma}_h^{(k-1)})$ from (4.28).

4. Draw $(\boldsymbol{A}^{(k)} \mid \boldsymbol{\Sigma}_e^{(k-1)}; \boldsymbol{z})$ from normal distribution (4.29).

5. Draw $(\boldsymbol{B}^{(k)} \mid \boldsymbol{y}^{(k)}, \boldsymbol{\Sigma}_h^{(k-1)}, \boldsymbol{\gamma}_b^{(k-1)})$ from normal distribution (4.47).

6. For $j = 2, \cdots, g$, draw $(\boldsymbol{\phi}_j^{(k)} \mid \boldsymbol{B}^{(k)}, \boldsymbol{\gamma}_\psi^{(k-1)}, \boldsymbol{\psi}^{(k-1)}, \boldsymbol{y}^{(k)})$ from distribution (2.54).

7. Draw $(\gamma_b^{(k)} \mid \boldsymbol{B}^{(k)})$ from distribution (4.48).

8. For $j = 1, \cdots, g$, draw $(\psi_{jj}^2 \mid \boldsymbol{y}^{(k)}, \boldsymbol{B}^{(k)}, \boldsymbol{\gamma}_\psi^{(k-1)})$, from distribution (4.55), then construct $\boldsymbol{\Sigma}_h^{(k)}$.

9. Draw $(\boldsymbol{\gamma}_\psi^{(k)} \mid \boldsymbol{\phi}_1^{(k)}, \cdots, \boldsymbol{\phi}_p^{(k)})$ from distribution (4.57).

4.5 Numerical Simulations

In order to study the performances of the stochastic search algorithm, we use artificially simulated data and compare the results obtained with results from a common conjugate analysis, as described in Section 4.3. We simulate one thousand samples and for each we estimate the parameters using $10,000$ cycles (after $1,000$ burn-in runs).

Example 4.1. Consider a four-variable model with the following parameters,

$$\boldsymbol{\Psi} = \begin{pmatrix} 1.5 & 2.5 & 2.5 & 2.0 \\ 0 & 1.5 & 0 & 0 \\ 0 & 0 & 1.5 & 0 \\ 0 & 0 & 0 & 1.5 \end{pmatrix}, \ \boldsymbol{\Sigma}_h = \begin{pmatrix} .444 & -.741 & -.741 & -.593 \\ -.741 & 1.679 & 1.235 & .988 \\ -.741 & 1.235 & 1.679 & .988 \\ -.593 & .988 & .988 & 1.235 \end{pmatrix},$$

$$\boldsymbol{\Sigma}_e = \begin{pmatrix} .250 & -.375 & -.750 & -.313 \\ -.375 & .813 & 1.125 & .469 \\ -.750 & 1.125 & 3.250 & .938 \\ -.313 & .469 & .938 & .641 \end{pmatrix},$$

$$\boldsymbol{A} = \begin{pmatrix} .1 & 0 & 0 & 0 \\ 0 & .1 & 0 & 0 \\ 0 & 0 & .1 & 0 \\ 0 & 0 & 0 & .1 \end{pmatrix}, \ \boldsymbol{B} = \begin{pmatrix} .3 & .2 & -.5 & -.1 \\ 0 & .4 & -.2 & 0 \\ 0 & 0 & .8 & -.2 \\ 0 & 0 & 0 & .8 \end{pmatrix}.$$

The sample size $T = 200$. For the conjugate analysis, we used inverse Wishart with 5 degrees of freedom and identity covariance matrix. The prior mean for $\boldsymbol{A}$ and $\boldsymbol{B}$ are the true values, with a diagonal covariance matrix whose elements are equal to .1 (so they are very informative). For the model selection part, the hyper-parameters of the priors are as follows: $p_i = .5$, $q_{ij} = .5$. $\kappa_{ij} = .1$, $d_{ij} = 50$. We have chosen an unrealistically good prior for $\boldsymbol{B}$ to use the conjugate results as a benchmark. Here we present the estimate obtained for this model:

1. Conjugate Analysis

$$\widehat{\boldsymbol{\Sigma}}_h = \begin{pmatrix} .478 & -.726 & -.773 & -.613 \\ -.726 & 1.647 & 1.211 & .975 \\ -.773 & 1.211 & 1.696 & .991 \\ -.613 & .975 & .991 & 1.254 \end{pmatrix}, \quad \widehat{\boldsymbol{\Sigma}}_e = \begin{pmatrix} .256 & -.396 & -.731 & -.299 \\ -.396 & .919 & 1.134 & .473 \\ -.731 & 1.134 & 3.233 & .899 \\ -.299 & .473 & .899 & .617 \end{pmatrix},$$

$$\widehat{\boldsymbol{B}} = \begin{pmatrix} .3031 & .2015 & -.4980 & -.0985 \\ .0002 & .4010 & -.1987 & .0011 \\ .0023 & .0025 & .7994 & -.1995 \\ -.0009 & -.0005 & .0014 & .7984 \end{pmatrix},$$

$$\widehat{\boldsymbol{A}} = \begin{pmatrix} .9962 & -.0002 & -.0026 & -.0010 \\ -.0010 & .9970 & -.0005 & .0002 \\ -.0007 & .0014 & 1.0008 & .0011 \\ -.0016 & .0035 & .0020 & 1.0023 \end{pmatrix}.$$

The estimates of the variances are already good but not excellent with such sample size, the estimates of $\boldsymbol{A}$ and $\boldsymbol{B}$ are very good, thanks also to the highly informative priors.

2. Model selection

$$\widehat{\boldsymbol{\Sigma}}_h = \begin{pmatrix} .470 & -.688 & -.757 & -.589 \\ -.688 & 1.590 & 1.158 & .931 \\ -.757 & 1.158 & 1.689 & .961 \\ -.589 & .931 & .961 & 1.230 \end{pmatrix}, \widehat{\boldsymbol{\gamma}}_\psi = \begin{pmatrix} * & 1.000 & .990 & .912 \\ * & * & .585 & .578 \\ * & * & * & .521 \\ * & * & * & * \end{pmatrix},$$

$$\widehat{\boldsymbol{\Sigma}}_e = \begin{pmatrix} .270 & -.425 & -.749 & -.312 \\ -.425 & 1.017 & 1.197 & .511 \\ -.749 & 1.197 & 3.282 & .922 \\ -.312 & .511 & .922 & .646 \end{pmatrix},$$

$$\widehat{\boldsymbol{B}} = \begin{pmatrix} .3111 & .1969 & -.4895 & -.0949 \\ .0084 & .3911 & -.1945 & .0025 \\ .0090 & .0146 & .7918 & -.1994 \\ -.0006 & .0073 & .0051 & .7860 \end{pmatrix}, \widehat{\boldsymbol{\gamma}}_b = \begin{pmatrix} 1.000 & .992 & 1.000 & .884 \\ .358 & 1.000 & .985 & .334 \\ .352 & .353 & 1.000 & .991 \\ .343 & .347 & .318 & 1.000 \end{pmatrix},$$

$$\widehat{\boldsymbol{A}} = \begin{pmatrix} .9958 & .0006 & -.0034 & -.0015 \\ -.0002 & .9971 & -.0003 & .0004 \\ -.0014 & .0007 & .9999 & .0005 \\ -.0019 & .0032 & .0016 & 1.0016 \end{pmatrix}.$$

Figures 4.2 and 4.3 present a scatterplot of true and estimated values of the latent process $\boldsymbol{y}_t$ for one randomly chosen sample using conjugate priors and for one randomly

chosen sample using the model selection technique. The parameter estimates obtained using the model selection technique are overall satisfactory, especially considered the small sample size, and particularly for $\boldsymbol{B}$. Although, some parameters are poorly estimated. For example, one of the elements in $\boldsymbol{\Sigma}_e$ is 37 percent bigger than the true value.

When compared with the results from the simulation study in Chapter 2, we can notice a different behaviour of $\widehat{\boldsymbol{\gamma}}_b$ and $\widehat{\boldsymbol{\gamma}}_\psi$: the values relative to the non-zero elements of $\boldsymbol{B}$ and $\boldsymbol{\Psi}$ are all very close to one, but the values relative to the zero elements of the same matrices are much bigger than zero, around .35 in $\widehat{\boldsymbol{\gamma}}_b$ and .5 in $\widehat{\boldsymbol{\gamma}}_\psi$. In other words, the stochastic search algorithm, when applied to the stochastic volatility model in Chapter 2, tends to choose models simpler than the true, while here tends to choose models which are more complex than the true. Figure 4.4 shows a comparison of the estimation error between the results of the benchmark conjugate analysis and those obtained with the model selection technique. The biggest differences are in the estimates of $\boldsymbol{B}$, where the conjugate analysis obtained almost perfect results, while the estimation error of the variances are of the same magnitude.

Example 4.2. Consider the same model as Example 4.1, but this time with $T = 500$. We compare the results obtained with the conjugate and the model selection techniques. For the conjugate case, we use two different priors for $\boldsymbol{B}$. In the first case we use the same benchmark prior as in the previous example, in the second case we choose a more realistic prior, having mean equal to .3 in correspondence of the positive elements of the true $\boldsymbol{B}$, and $-.3$ in correspondence of the negatives, with variances ten times bigger than in the previous example. All the other priors as set as before. The estimates are as follows:

1. Conjugate Analysis, best prior

$$\widehat{\boldsymbol{\Sigma}}_h = \begin{pmatrix} .468 & -.730 & -.763 & -.609 \\ -.730 & 1.653 & 1.216 & .979 \\ -.763 & 1.216 & 1.693 & .990 \\ -.609 & .979 & .990 & 1.261 \end{pmatrix}, \ \widehat{\boldsymbol{\Sigma}}_e = \begin{pmatrix} .250 & -.385 & -.733 & -.298 \\ -.385 & .873 & 1.129 & .462 \\ -.733 & 1.129 & 3.227 & .911 \\ -.298 & .462 & .911 & .612 \end{pmatrix},$$

$$\widehat{\boldsymbol{B}} = \begin{pmatrix} .3029 & .2006 & -.4983 & -.0999 \\ .0001 & .4006 & -.1997 & .0005 \\ .0023 & .0022 & .8002 & -.2001 \\ -.0001 & -.0004 & .008 & .7992 \end{pmatrix},$$

$$\widehat{\boldsymbol{A}} = \begin{pmatrix} .9971 & -.0002 & -.0009 & -.0005 \\ -.0008 & .9978 & -.0003 & .0002 \\ -.0007 & .0011 & 1.0007 & .0002 \\ -.0016 & .0035 & .0016 & 1.0018 \end{pmatrix}.$$

2. Conjugate Analysis, other prior

$$\widehat{\boldsymbol{\Sigma}}_h = \begin{pmatrix} .455 & -.671 & -.748 & -.598 \\ -.671 & 1.512 & 1.129 & .924 \\ -.748 & 1.129 & 1.696 & .979 \\ -.598 & .924 & .979 & 1.257 \end{pmatrix}, \ \widehat{\boldsymbol{\Sigma}}_e = \begin{pmatrix} .263 & -.429 & -.743 & -.299 \\ -.429 & 1.001 & 1.195 & .504 \\ -.743 & 1.195 & 3.214 & .914 \\ -.299 & .504 & .914 & .608 \end{pmatrix},$$

$$\widehat{\boldsymbol{B}} = \begin{pmatrix} .3308 & .1688 & -.4767 & -.0825 \\ .0013 & .4176 & -.1962 & .0059 \\ .0233 & .0525 & .7824 & -.2238 \\ -.0340 & -.0393 & .0202 & .7714 \end{pmatrix},$$

$$\widehat{\boldsymbol{A}} = \begin{pmatrix} .9958 & -.0010 & -.0020 & -.0012 \\ -.0060 & .9960 & -.0005 & .0003 \\ -.0016 & .0014 & 1.0030 & .0011 \\ -.0019 & .0038 & .0014 & 1.0027 \end{pmatrix}.$$

3. Model selection

$$\widehat{\boldsymbol{\Sigma}}_h = \begin{pmatrix} .466 & -.714 & -.763 & -.608 \\ -.714 & 1.602 & 1.193 & .966 \\ -.763 & 1.193 & 1.702 & .993 \\ -.608 & .966 & .993 & 1.261 \end{pmatrix}, \ \widehat{\boldsymbol{\gamma}}_\psi = \begin{pmatrix} * & 1.000 & .999 & .974 \\ * & * & .495 & .501 \\ * & * & * & .411 \\ * & * & * & * \end{pmatrix},$$

$$\widehat{\boldsymbol{\Sigma}}_e = \begin{pmatrix} .255 & -.405 & -.738 & -.301 \\ -.405 & .935 & 1.161 & .479 \\ -.738 & 1.161 & 3.218 & .916 \\ -.301 & .479 & .916 & .611 \end{pmatrix},$$

$$\widehat{\boldsymbol{B}} = \begin{pmatrix} .3108 & .1936 & -.4928 & -.0976 \\ .0073 & .4018 & -.1958 & .0025 \\ .0048 & .0148 & .7985 & -.2005 \\ -.0003 & .0073 & .0009 & .7908 \end{pmatrix}, \ \widehat{\boldsymbol{\gamma}}_b = \begin{pmatrix} 1.000 & .998 & 1.000 & .930 \\ .323 & 1.000 & .995 & .226 \\ .308 & .304 & 1.000 & 1.000 \\ .300 & .287 & .255 & 1.000 \end{pmatrix},$$

$$\widehat{\boldsymbol{A}} = \begin{pmatrix} .9963 & .0001 & -.0032 & -.0011 \\ -.0002 & .9975 & -.0001 & .0000 \\ -.0020 & .0007 & 1.0011 & .0002 \\ -.0011 & .0031 & .0015 & 1.0012 \end{pmatrix}.$$

Again, the stochastic search algorithm tends to select models which are more complex than the true. Posterior means of $\boldsymbol{\gamma}_b$ and $\boldsymbol{\gamma}_\psi$ express in fact similar values as in the previous example. The estimation is generally more precise, with the biggest estimation error for a non-zero element in $\boldsymbol{B}$, $\boldsymbol{\Sigma}_e$ and $\boldsymbol{\Sigma}_h$ being respectively 3, 15 and 4 percent. All the elements of $\widehat{\boldsymbol{B}}$ present estimation error smaller to those of the $\widehat{\boldsymbol{B}}$ relative to the non benchmark conjugate analysis. That is not the same if we consider the variance estimate, where the model selection estimator does not perform unifromly better, maybe because of the indirect estimation of $\boldsymbol{\Sigma}_h$. A graphical comparison of the estimates obtained with

the model selection technique and the non benchmark conjugate analysis can be found in Figure 4.5.

Example 4.3. Consider now a five-variable model with $T = 500$ and the following parameters,

$$\boldsymbol{\Psi} = \begin{pmatrix} 1.5 & 2.5 & 2.5 & 2.0 & 2.0 \\ 0 & 1.5 & 0 & 0 & 0 \\ 0 & 0 & 1.5 & 0 & 0 \\ 0 & 0 & 0 & 1.5 & 0 \\ 0 & 0 & 0 & 0 & 1.5 \end{pmatrix}, \tag{4.58}$$

$$\boldsymbol{\Sigma}_h = \begin{pmatrix} .444 & -.741 & -.741 & -.593 & -.593 \\ -.741 & 1.679 & 1.235 & .988 & .988 \\ -.741 & 1.235 & 1.679 & .988 & .988 \\ -.593 & .988 & .988 & 1.235 & .790 \\ -.593 & .988 & .988 & .790 & 1.235 \end{pmatrix}, \tag{4.59}$$

$$\boldsymbol{\Sigma}_e = \begin{pmatrix} .250 & -.375 & -.750 & -.313 & -.313 \\ -.375 & .813 & 1.125 & .469 & .469 \\ -.750 & 1.125 & 3.250 & .938 & .938 \\ -.313 & .469 & .938 & .641 & .391 \\ -.313 & .469 & .938 & .391 & .641 \end{pmatrix}, \tag{4.60}$$

$$\boldsymbol{B} = \begin{pmatrix} .3 & .2 & -.5 & -.1 & 0 \\ 0 & .4 & -.2 & 0 & .1 \\ 0 & 0 & .8 & -.2 & 0 \\ 0 & 0 & 0 & .8 & -.2 \\ 0 & 0 & 0 & 0 & .5 \end{pmatrix}, \ \boldsymbol{A} = \begin{pmatrix} .1 & 0 & 0 & 0 & 0 \\ .5 & .1 & 0 & 0 & 0 \\ 0 & .5 & .1 & 0 & 0 \\ 0 & 0 & .5 & .1 & 0 \\ 0 & 0 & 0 & .5 & .1 \end{pmatrix}. \tag{4.61}$$

Here we present the estimates obtained using a "non benchmark" conjugate analysis and the stochastic search algorithm for model selection.

1.Conjugate Analysis

$$\widehat{\boldsymbol{\Sigma}}_h = \begin{pmatrix} .469 & -.727 & -.785 & -.597 & -.589 \\ -.727 & 1.591 & 1.213 & .965 & .958 \\ -.785 & 1.213 & 1.778 & .987 & .991 \\ -.597 & .965 & .987 & 1.248 & .759 \\ -.589 & .958 & .991 & .759 & 1.216 \end{pmatrix},$$

$$\widehat{\boldsymbol{\Sigma}}_e = \begin{pmatrix} .248 & -.393 & -.737 & -.297 & -.300 \\ -.393 & .912 & 1.215 & .509 & .495 \\ -.737 & 1.215 & 3.257 & .919 & .917 \\ -.297 & .509 & .919 & .627 & .389 \\ -.300 & .495 & .917 & .389 & .646 \end{pmatrix},$$

$$\widehat{\boldsymbol{B}} = \begin{pmatrix} .3079 & .2293 & -.4850 & -.1021 & -.0070 \\ .0161 & .3772 & -.1896 & .0131 & .1209 \\ -.0260 & .0220 & .7685 & -.2308 & -.0123 \\ -.0008 & -.0106 & .0004 & .7819 & -.2010 \\ .0085 & -.0057 & -.0024 & .0159 & .4926 \end{pmatrix},$$

$$\widehat{\boldsymbol{A}} = \begin{pmatrix} .9952 & .0023 & -.0044 & -.0008 & -.0014 \\ .5015 & .9949 & -.0007 & .0004 & .0005 \\ -.0015 & .5006 & .9994 & .0001 & .0003 \\ -.0005 & .0006 & .4994 & 1.0006 & -.00073 \\ -.0024 & .0044 & .0027 & .5017 & 1.0024 \end{pmatrix}.$$

2. Model selection

$$\widehat{\boldsymbol{\Sigma}}_h = \begin{pmatrix} .482 & -.759 & -.792 & -.615 & -.602 \\ -.759 & 1.685 & 1.249 & .987 & .981 \\ -.792 & 1.249 & 1.759 & 1.023 & .998 \\ -.615 & .987 & 1.023 & 1.268 & .775 \\ -.602 & .981 & .998 & .775 & 1.232 \end{pmatrix},$$

$$\widehat{\boldsymbol{\gamma}}_{\psi} = \begin{pmatrix} * & 1.000 & 1.000 & .973 & .960 \\ * & * & .547 & .561 & .582 \\ * & * & * & .523 & .530 \\ * & * & * & * & .557 \\ * & * & * & * & * \end{pmatrix},$$

$$\widehat{\boldsymbol{\Sigma}}_{e} = \begin{pmatrix} .240 & -.369 & -.709 & -.285 & -.292 \\ -.369 & .854 & 1.138 & .474 & .473 \\ -.709 & 1.138 & 3.178 & .871 & .898 \\ -.285 & .474 & .871 & .605 & .384 \\ -.292 & .473 & .898 & .384 & .646 \end{pmatrix},$$

$$\widehat{\boldsymbol{B}} = \begin{pmatrix} .3045 & .2168 & -.4924 & -.0945 & -.0060 \\ .0052 & .3842 & -.1942 & .0055 & .0973 \\ -.0051 & .0033 & .7922 & -.2023 & .0023 \\ -.0009 & -.0057 & -.0047 & .7907 & -.1979 \\ -.0026 & -.0023 & -.0019 & .0082 & .4952 \end{pmatrix},$$

$$\widehat{\boldsymbol{\gamma}}_{b} = \begin{pmatrix} 1.000 & .999 & 1.000 & .928 & .347 \\ .419 & 1.000 & .994 & .344 & .871 \\ .402 & .406 & 1.000 & .998 & .408 \\ .404 & .405 & .365 & 1.000 & .991 \\ .392 & .376 & .356 & .298 & 1.000 \end{pmatrix},$$

$$\widehat{\boldsymbol{A}} = \begin{pmatrix} .9944 & .0006 & -.0045 & -.0019 & -.0022 \\ .5008 & .9953 & -.0003 & .0006 & .0004 \\ -.0019 & .5012 & 1.0000 & -.0001 & .0007 \\ -.0005 & .0010 & .4999 & 1.0011 & -.0005 \\ -.0025 & .0041 & .0023 & .5015 & 1.0022 \end{pmatrix}.$$

The major difference between this example and the previous two lies in the structure of

$\boldsymbol{A}$, which now is not an identity matrix. Again, results are generally satisfactory[5]. The posterior means of $\boldsymbol{\gamma}_b$ and $\boldsymbol{\gamma}_\psi$ relative to the non-zero elements of $\boldsymbol{B}_1$ and $\boldsymbol{\Psi}$ are still quite larger than zero, and for $\boldsymbol{\gamma}_b$ are even larger than in the four variable examples. Nevertheless, it is still possible to easy distinguish the zero and non-zero elements of the matrix. The estimates of $\boldsymbol{B}$ are generally more precise than the estimates of the covariance matrices, like in the two previous examples. A graphical comparison of the results obtained with conjugate analysis and model selection can be found in Figure 4.6. We can see how the model selection procedure generally produces better estimates of $\boldsymbol{B}$, but worse estimates of $\boldsymbol{\Sigma}_h$.

4.6 Conclusions

Estimation techniques for dynamic models require feasibility and computational efficiency. Often dynamic modeling relies on a priori restrictions and assumptions that are not always easily justifiable using theoretical reasoning. Also, speaking about the Bayesian choice, being able to elicitate reasonable priors is not a trivial problem at all, especially for the latent process' evolution. Extending the results of Chapter 2, we derived a variable selection approach for state space models which allows us to start with a very general setup and later impose data driven restrictions. The stochastic search algorithm is fast enough and computationally cheap to be used as an instrument of model calibration in large state space frameworks. Scholars can then focus on all potentially useful models. We applied the algorithm to artificially generated data, and compared the estimates with those obtained from a more common Bayesian MCMC algorithm based on conjugate priors. Results show the efficacy of the stochastic search algorithm.

[5]Simulations performed with more complex structures of $\boldsymbol{A}$ happened to be much harder to be correctly estimated.

Much still needs to be done. Improvements and further developments will be discussed in the next chapter.

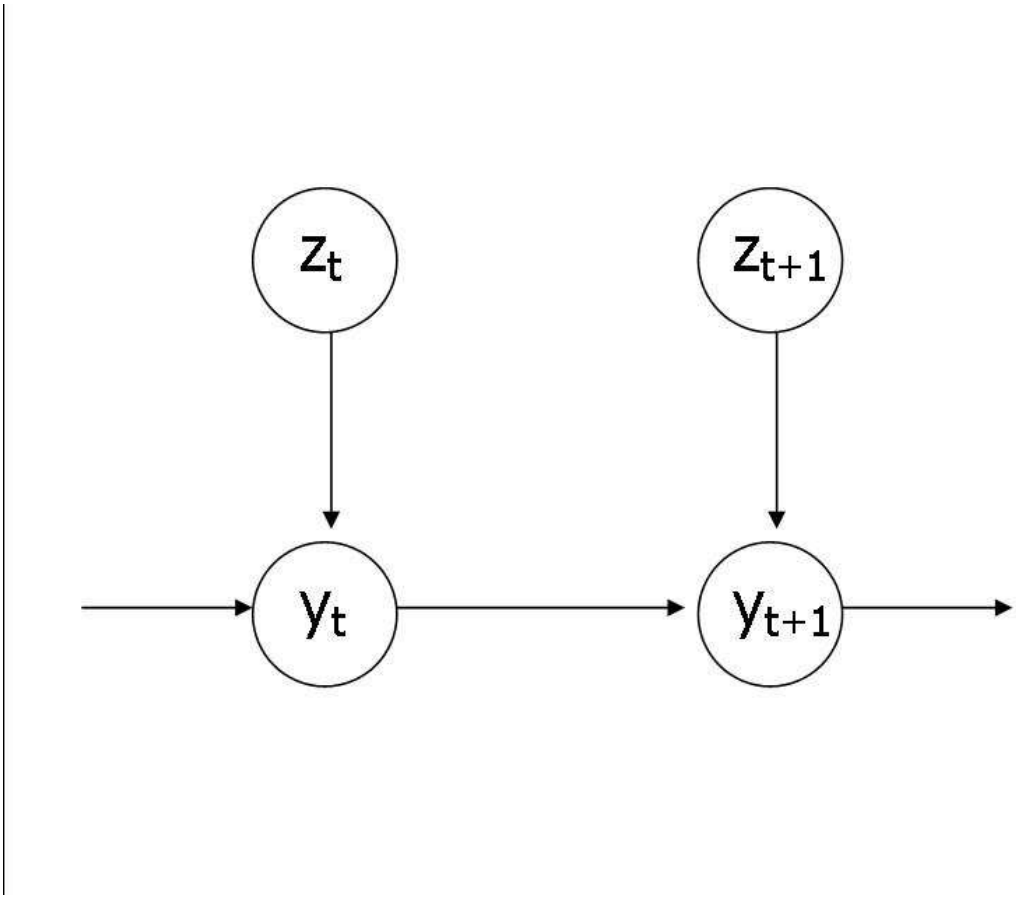

Figure 4.1: Dependence Structure of a Hidden Markov Models. The observed variable is $\boldsymbol{z}_t$ while the latent variable is $\boldsymbol{y}_t$

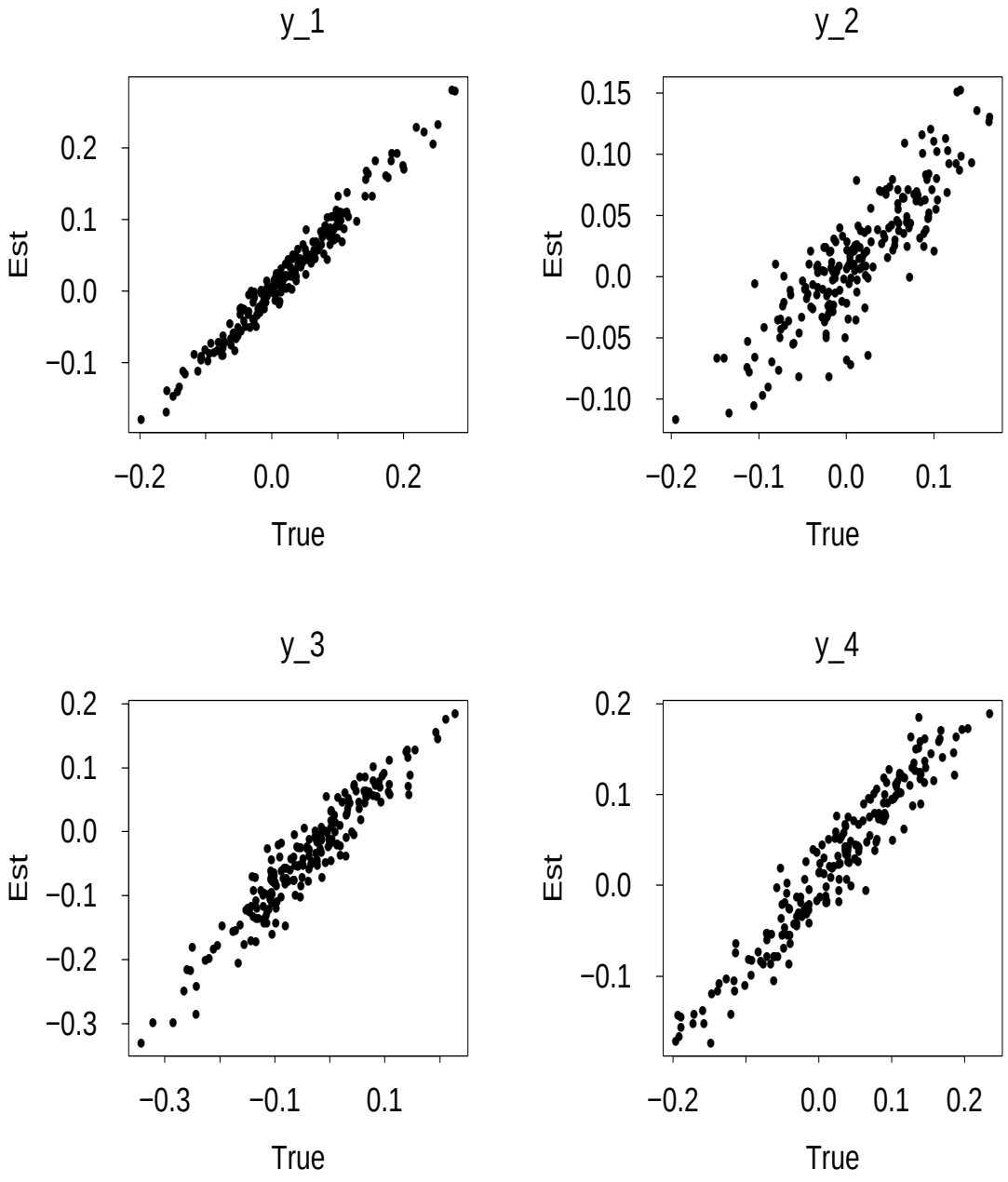

Figure 4.2: Comparison of the True and Estimated Latent Process y for One Sample. Results are obtained estimating the parameters as in Section 4.3

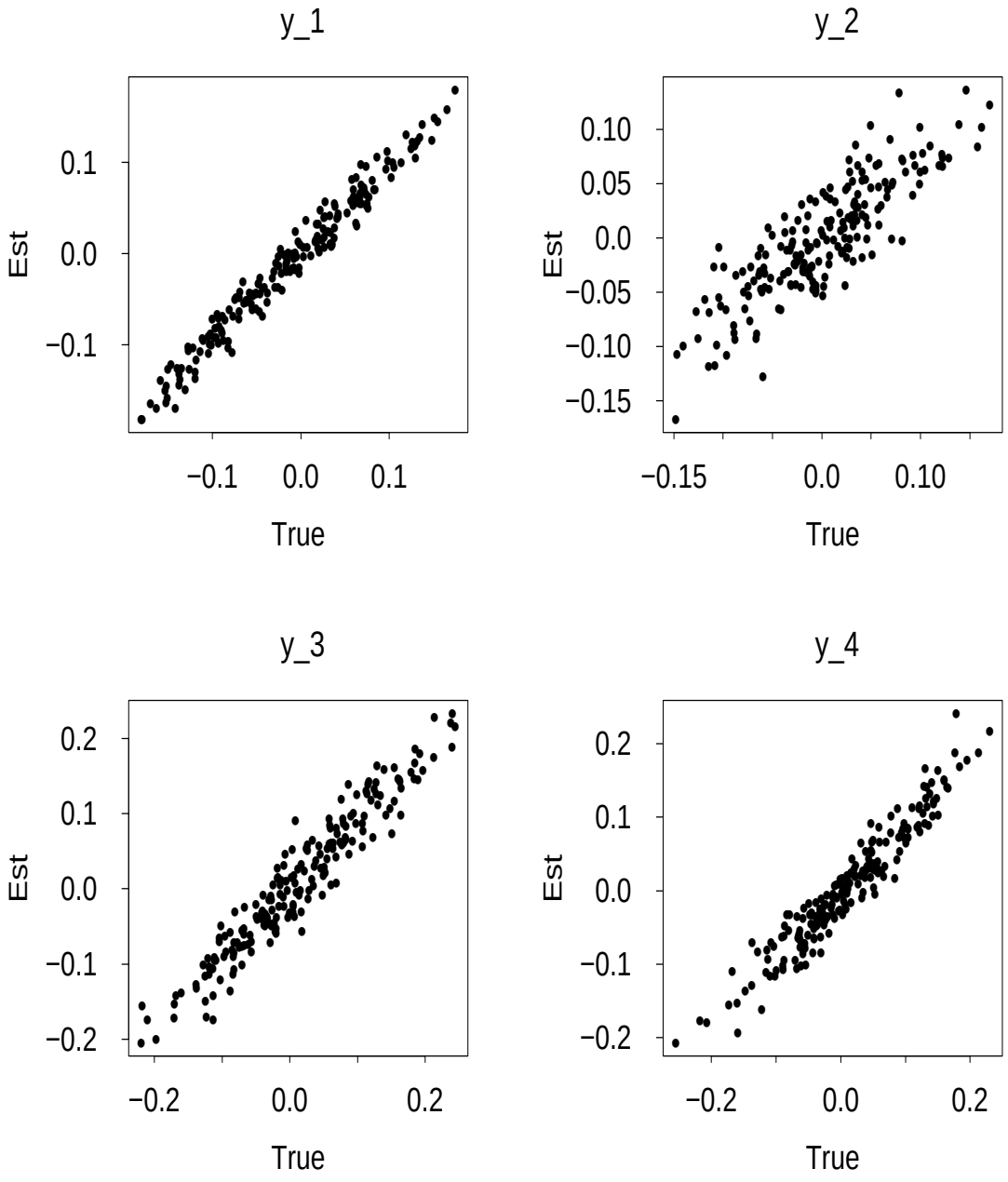

Figure 4.3: Comparison of the True and Estimated Latent Process y for One Sample. Results are obtained estimating the parameters as in Section 4.4

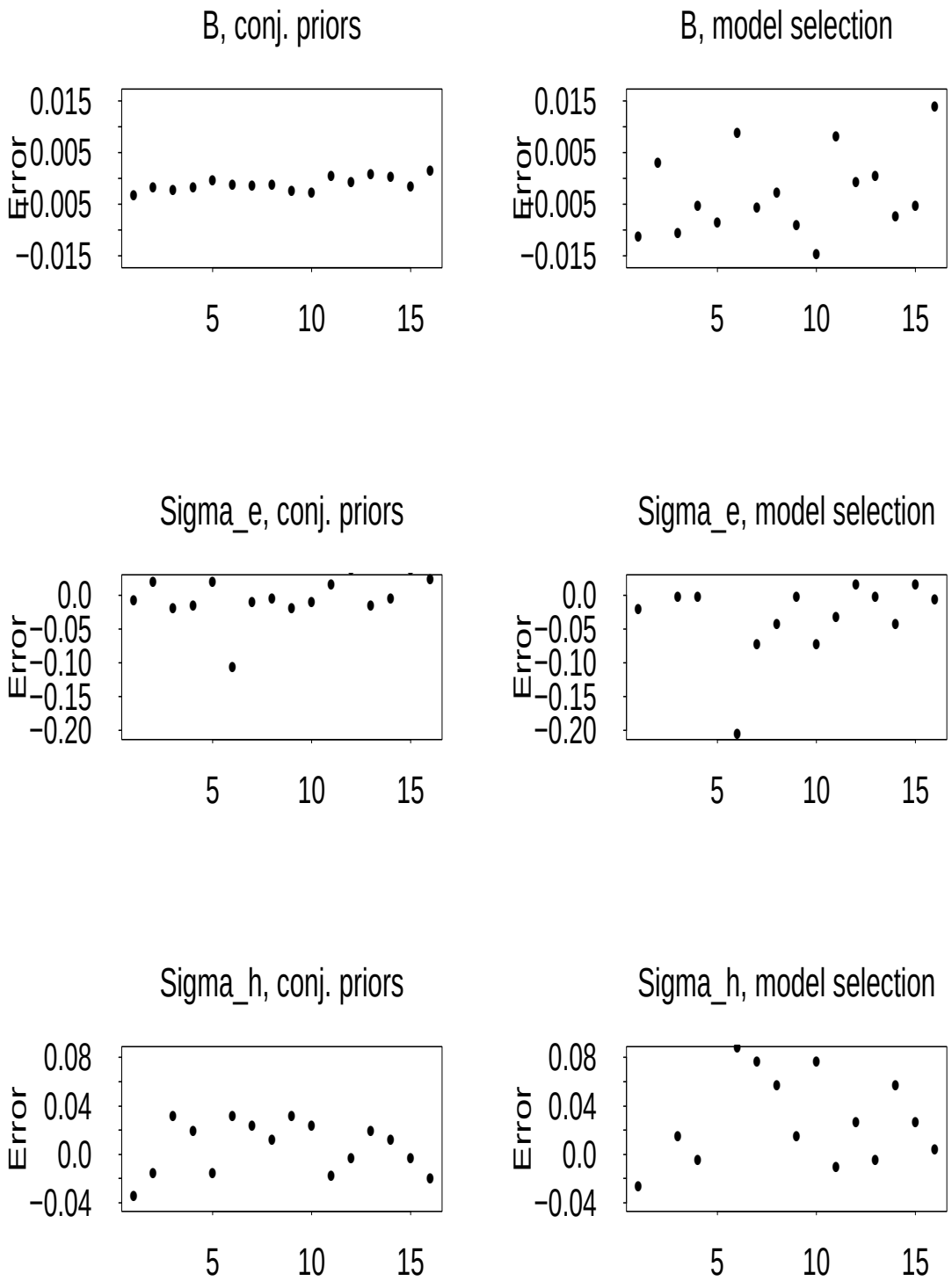

Figure 4.4: Comparison of the estimates obtained with the methods of Section 4.3 and 4.4

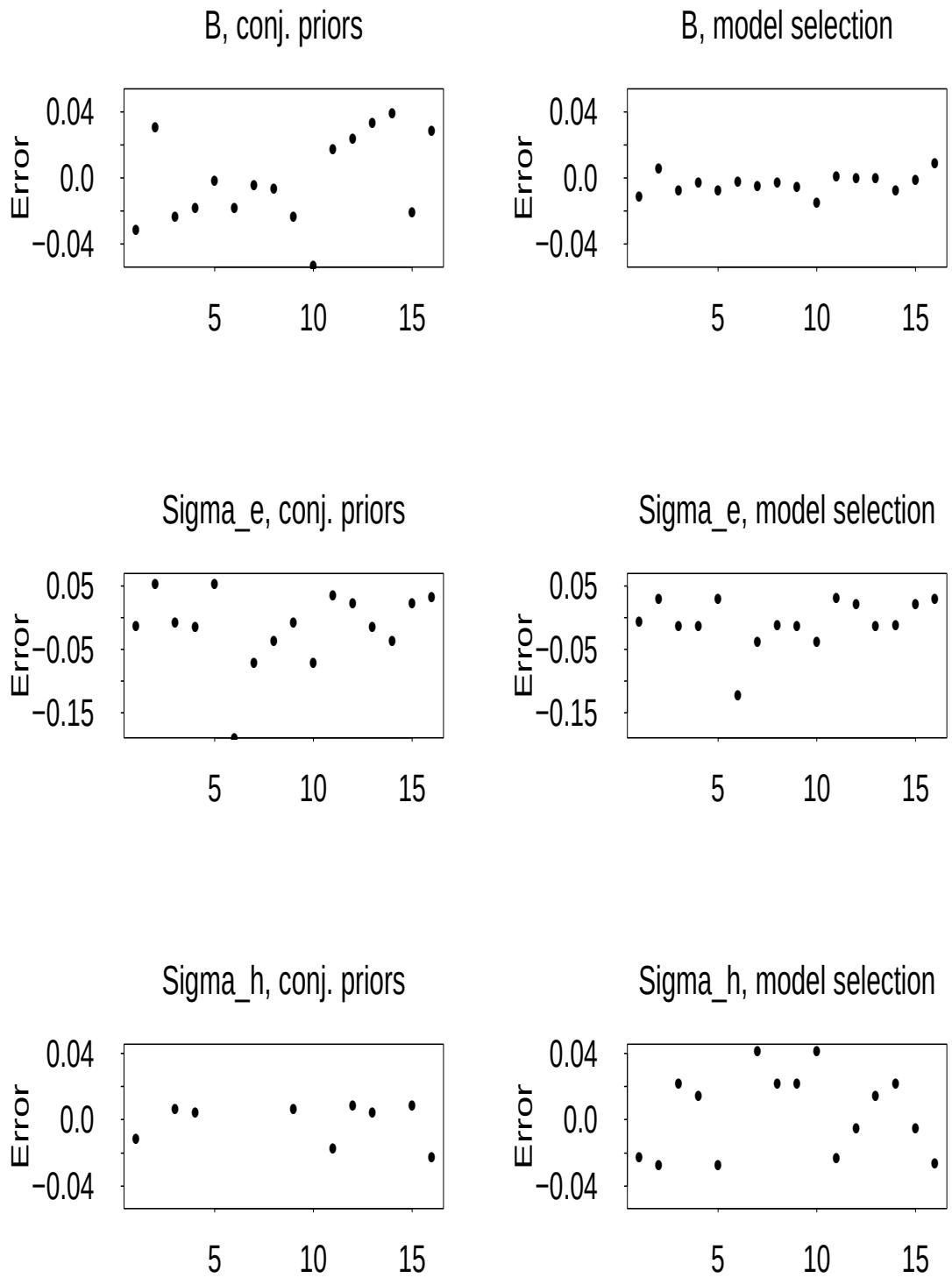

Figure 4.5: Comparison of the estimates obtained with the methods of Section 4.3 and 4.4

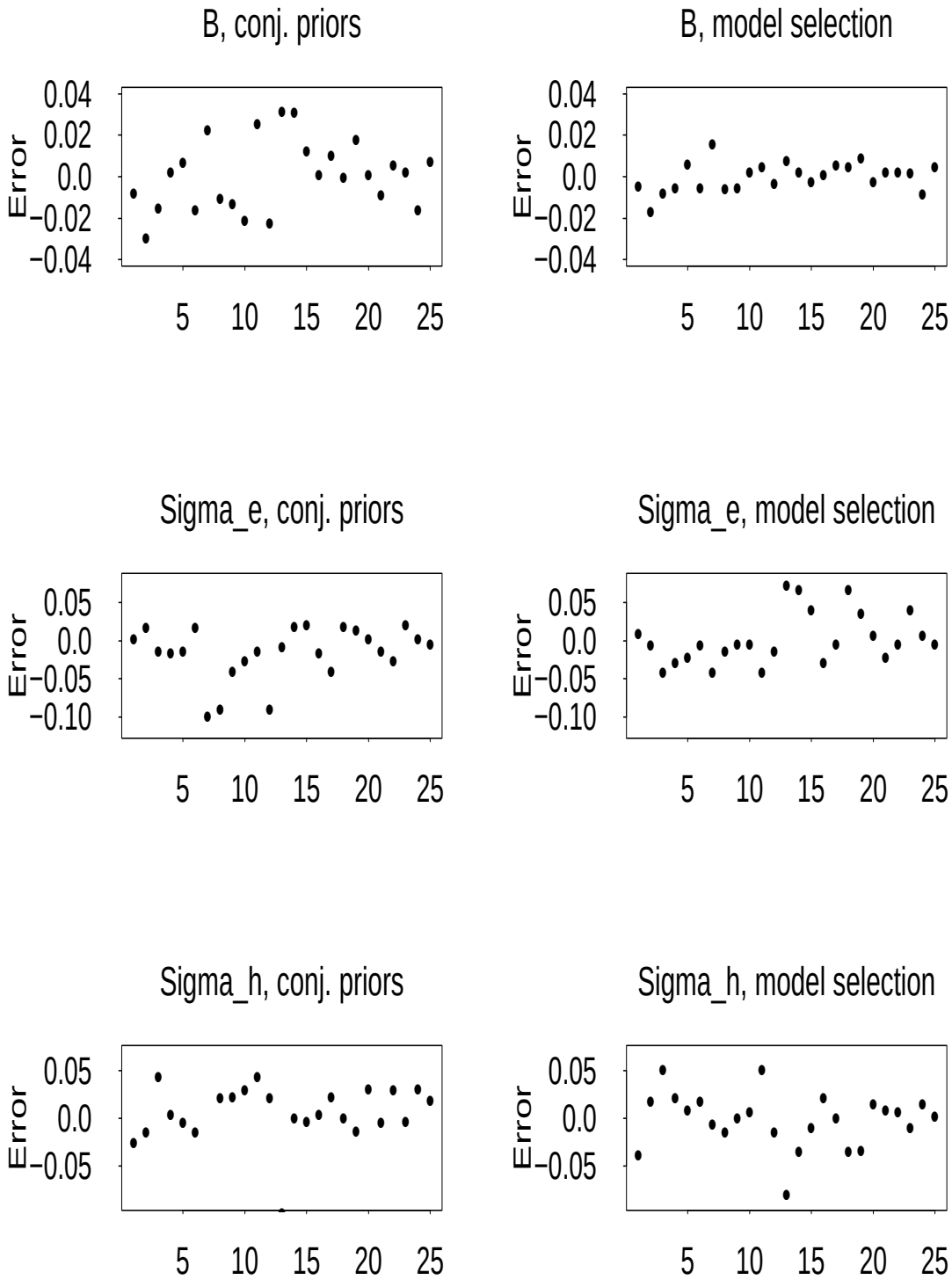

Figure 4.6: Comparison of the estimates obtained with the methods of Section 4.3 and 4.4

Chapter 5

Discussion and Future Research

In this study, we presented Bayesian algorithms for estimation and model selection of multivariate stochastic volatility and dynamic models. All the algorithms utilize and extend the stochastic search variable selection method first developed by George & McCulloch (1993), who applied the method to linear regression. We showed in this study how the same method can be successfully applied to multivariate time series models with latent variables, such as stochastic volatility or state space models. We performed stochastic search model selection for both the regression coefficient and the covariance parameters. For the stochastic volatility framework, we also performed selection of the regression coefficient and variance parameters of the volatility equation. Selection of the non-zero variances of the stochastic volatility generating process in particular allows us to choose between time varying variance models with stochastic or deterministic volatilities, a choice that up to now has been made a priori by the researcher. Also, new challenges are brought to the forefront: unlike all other parameters subject to model selection, variances of the volatility equation are not normal but inverse gamma parameters. Selection of these parameters happens to be very sensitive to the choice

of prior hyperparameters. We found a solution by imposing a flat prior on them, and simulation studies showed the solution to be successful, but we believe that it could still be improved, and this will be one of the points of our future research.

Computational efficiency of the stochastic search algorithm lies in embedding model selection in the same Gibbs sampler used for parameter estimation. In the context of state space models, all the estimation can be carried out using Gibbs sampling, although filtering the latent variables will lead to a faster mixing MCMC chain. In the context of time varying variance models, the unknown form of the stochastic volatility full conditional necessitates a different sampling procedure to be introduced in the Gibbs cycle. We utilized Gilks sampler, particle filter and smoother, and filter based on rejection sampling. Alternative methods available (Metropolis Hastings and auxiliary particle filter among others), did not produced results that were as effective. Metropolis Hasting has already proven to provide low acceptance rates and slowly mixing MCMC chains in such frameworks (Carter & Kohn (1994), Chib et al. (1998)), while auxiliary particle filter has proven to be very useful in particular situations which have not been manifest in our study. The methods chosen have been compared in terms of speed and efficiency. The efficiency of the computation can still be improved in many ways, and this will be another point for future research. In particular, Sun & Berger (2006) developed an algorithm for posterior computation using Cholesky decompositions which allows us to use only the below diagonal elements of the sample covariance instead of the whole matrix. This method can be easily adapted for the modified Cholesky decompositions utilized here, and will greatly reduce the amount of computation necessary to sample from the covariances' posteriors.

We employed different types of Cholesky decompositions for efficient parametrization

of the covariance matrices. Stochastic search algorithms could be embedded in all the resulting models with only slight modifications. In particular, two different modified Cholesky decompositions (Pourahmadi (1999) and Chen & Dunson (2003)) have been used for time varying covariances parametrization, and the resulting models have been compared with estimated Bayes factors. For both decompositions, we imposed the resulting diagonal matrices to be time varying, while the lower unit triangular matrix to be time invariant. This is equivalent to imposing time invariant correlations if we utilize Chen & Dunson decomposition. By performing model selection between the two different parametrizations, we can also test the assumption of constant correlations. We used Bayes factors to compare two restricted models chosen using the stochastic search algorithm, and so utilizing the two selection techniques in a complementary way. Both modified Cholesky decomposition can successfully be employed in a state space model framework. The Chen & Dunson decomposition would allow us to separately model scale and correlation parameters of the latent variable's covariance matrix. Also, different types of covariance decompositions can be utilized for both the models studied here. Spectral decompositions (for example Chiu et al. (1996)) can obviate the problem of variable ordering, even if they have some drawbacks when compared to Cholesky type decompositions.

In addition to the improvements and developments mentioned in this chapter, there are further avenues for future research. In this study we always adopted independent priors. The structure of (2.54) can be used to develop better priors for $\mathbf{\Psi}$. Also, non independent priors for the model indicators can greatly improve the flexibility of the model. The researcher could impose priority on some variables, or easily develop hierarchical selection structures, just modifying the prior hyperparameters. The cost of this is essentially computational: prior independence allows the full conditional posterior for

the model indicators to have closed form, thus not requiring an additional Metropolis step.

The multivariate stochastic volatility model in Chapter 2 can be generalized to allow fat tail distributions of the errors and jumps, similarly to Chib et al. (2005), while the stochastic search algorithm can be successfully applied to a much broader family of dynamic structures than the state space models presented here.

In conclusion, this work presents an efficient and computationally feasible method for performing model selection for large multivariate time series, which present a very high number of competing models. The method is fully Bayesian and can be easily applied to different frameworks. It can be used to choose the best submodel for each framework, which can then be compared using traditional model selection techniques. Efficient parametrization of the covariance matrix is crucial for the efficiency of the Bayesian estimation, and Cholesky decompositions are found to be very effective. Much work must still be done to improve the efficiency of the algorithm and the generality of the model, but the methods studied here already represent a useful tool for model building and calibration, a tool that allows the researcher to start with a very general model and then impose restrictions based on statistical considerations, thus catalyzing a feedback between theory and quantitative analysis in economics and social sciences in general.

Bibliography

Aguilar, O., Huerta, G., Prado, R. & West, M. (1999), Bayesian inference on latent structure in time series, *in* 'Bayesian Statistics 6 – Proceedings of the Sixth Valencia International Meeting', pp. 3–26.

Aguilar, O. & West, M. (2000), 'Bayesian dynamic factor models and portfolio allocation', *Journal of Business and Economic Statistics* **18**(3), 338–357.

Andrieu, C. & Doucet, A. (2002), 'Particle filtering for partially observed Gaussian state space models', *Journal of the Royal Statistical Society, Series B, Methodological* **64**(4), 827–836.

Barnard, J., McCulloch, R. & Meng, X.-L. (2000), 'Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage', *Statistica Sinica* **10**(4), 1281–1311.

Bauens, L., Laurent, S. & Rombouts, J. V. K. (2004), Multivariate garch models: a survey. CORE Discussion Paper.

Black, F. & Scholes, M. (1973), 'The pricing options of corporate liabilities', *Journal of Political Economy* **81**, 637–654.

Blanchard, O. J. (1989), 'A traditional interpretation of macroeconomic fluctuations', *American Economic Review* pp. 1146–1164.

Bollen, K. A. (1989), *Structural equations with latent variables*, John Wiley and Sons.

Bollen, K. A. e. & Long, J. S. e. (1993), *Testing structural equation models*, Sage Publications Inc.

Bollerslev, T. (1986), 'Generalized autoregressive conditional heteroskedasticity', *Journal of Econometrics* **31**, 307–327.

Bollerslev, T. (1990), 'Modelling the coherence in short-run nominal exchange rates: a multivariate genearlized arch model', *The review of economics and statistics* **72**, 498–505.

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 [Formerly: Statistischen Hefte]* **41**(4), 409–422.

Box, G. E. P. & Jenkins, G. M. (1976), *Time series analysis: Forecasting and control (Rev ed)*, Holden-Day Inc.

Brockwell, P. J. & Davis, R. A. (1991), *Time series. Theory and methods (Second edition)*, Springer-Verlag Inc.

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. (1995), 'Marginal likelihood from the Gibbs output', *Journal of the American Statistical Association* **90**, 1313–1321.

Chib, S. & Jeliazkov, I. (2001), 'Marginal likelihood from the Metropolis-Hastings output', *Journal of the American Statistical Association* **96**(453), 270–281.

Chib, S., Nardari, F. & Shephard, N. (1998), 'Stochastic volatility: likelihood inference and comparison with arch models', *Review of Economic Studies* **65**, 361–394.

Chib, S., Nardari, F. & Shephard, N. (2002), 'Markov chain Monte Carlo methods for stochastic volatility models', *Journal of Econometrics* **108**(2), 281–316.

Chib, S., Nardari, F. & Shephard, N. (2005), 'Analysis of high dimensional multivariate stochastic volatility models', *Journal of Econometrics* **forthcoming**.

Chiu, T. Y. M., Leonard, T. & Tsui, K.-W. (1996), 'The matrix-logarithmic covariance model', *Journal of the American Statistical Association* **91**, 198–210.

Clark, P. K. (1973), 'A subordinated stochastic process model with finite variance for speculative prices (Com: p157-159)', *Econometrica* **41**, 135–155.

Engle, R. (1982), 'Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation (STMA V24 501)', *Econometrica* **50**, 987–1007.

Engle, R. (2002), 'Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models', *Journal of Business and Economic Statistics* **20**(3), 339–350.

Felsenstein, J. & Churchill, G. A. (1996), 'A hidden Markov model approach to variation among sites in rate of evolution', *Molecular Biology and Evolution* **13**, 93–104.

Gallant, A., Hsieh, D. & Tauchen, G. (1991), 'On fitting a recalcitrant series: The pound/dollar exchange rates (Com: p1974-83)', *Nonparametric and Semiparametric Methods in Econometric and Statistics, Proceedings of the Fifth International Symposium in Economic Theory and Econometrics* pp. 1974–1983.

Gelfand, A. E. & Smith, A. F. M. (1990), 'Sampling-based approaches to calculating marginal densities', *Journal of the American Statistical Association* **85**, 398–409.

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. (2005), Stochastic search model selection for restricted var models. manuscript.

Ghysels, E., Harvey, A. C. & Renault, E. (1996), Stochastic volatility, *in* 'Handbook of Statistics Volume 14: Statistical Methods in Finance', pp. 119–191.

Gilks, W. R. & Wild, P. (1992), 'Adaptive rejection sampling for Gibbs sampling', *Applied Statistics* **41**, 337–348.

Gordon, D. B. & Leeper, E. M. (1994), 'The dynamic impacts of monetary policy: An exercise in tentative identification', *Journal of Political Economy* **102**, 1228–1247.

Hamilton, J. D. (1990), 'Analysis of time series subject to changes in regime', *Journal of Econometrics* **45**, 39–70.

Harvey, A. C. (1990), *Forecasting, structural time series models, and the Kalman filter*, Cambridge University Press.

Harvey, A. C. (1993), *Time series models*, MIT Press.

Harville, D. A. (1997), *Matrix Algebra from a Statistician's Perspective*, Springer-Verlag Inc.

Ho, M., Hombao, H. & Shumway, R. H. (2003), Investigating brain dynamics in fmri experiments by state-space models, *in* 'Joint statistial meeting proceedings'.

Hol, E. & Koopman, S. J. (2000), Forecasting the variability of stock index returns with stochastic volatility models and implied volatility. Tinbergen Institute Discussion Paper.

Hol, E. & Koopman, S. J. (2002), Stock index volatility forecasting high frequency data. Tinbergen Institute Discussion Paper.

Hull, J. & White, A. (1987), 'The pricing options on assets with stochastic volatilies', *Journal of Finance* **42**, 281–300.

Jacquier, E., Polson, N. G. & Rossi, P. E. (1994), 'Bayesian analysis of stochastic volatility models (Disc: p389-417)', *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**(1), 69–87.

Jeffrey, H. (1961), *Theory of Probability*, Oxford University Press.

Jeffreys, H. (1935), 'Some test of significance, treated by the theory of probability', *Proceedings of Cambridge Philosophical Society* **31**, 203–222.

Kalman, R. (1960), 'A new approach to linear filtering and prediction problems', *Transaction of the ASME–Journal of Basic Engineering* **March**, 35–45.

Kass, R. E. & Raftery, A. E. (1995), 'Bayes factors', *Journal of the American Statistical Association* **90**, 773–795.

Kim, H., Sun, D. & Tsutakawa, R. K. (2002), 'Lognormal vs. gamma: Extra variations', *Biometrical Journal* **44**(3), 305–323.

Latane, H. & Rendleman, R. J. (1976), 'Standard deviations and stock price ratios implied in option prices', *Journal of Finance* **31**, 369–381.

Leeper, E. M., Sims, C. A. & Zha, T. (1996), 'What does monetary policy do', *Brooking Papers on Economic Activity* **2**, 1–63.

Leroux, B. G. & Puterman, M. L. (1992), 'Maximum-penalized-likelihood estimation for independent and Markov-dependent mixture models', *Biometrics* **48**, 545–558.

Muller, P. & Pole, A. (1998), 'Monte Carlo posterior integration in GARCH models', *Sankhya, Series B, Indian Journal of Statistics* **60**, 127–144.

Nelson, D. B. (1991), 'Conditional heterosckdasticity in asset returns: A new approach', *Econometrica* **59**, 347–370.

Ni, S. & Sun, D. (2003), 'Noninformative priors and frequentist risks of Bayesian estimators of vector-autoregressive models', *Journal of Econometrics* **115**, 159–197.

Ni, S. & Sun, D. (2005), 'Alternative Bayesian estimators for vector-autoregressive models', *Journal of Business and Economic Statistics* **23**, 105–117.

Ni, S., Sun, D. & Sun, X. (2006), 'Bayesian analysis for vector-autoregressive models under the intrinsic loss', *Journal of Business and Economic Statistics* . In press.

Pinheiro, J. C., Pinheiro, J. C. & Bates, D. M. (1996), 'Unconstrained parametrizations for variance-covariance matrices', *Statistics and Computing* **6**, 289–296.

Pitt, M. K. (2001), Smooth particle filters for likelihood maximization. manuscript.

Pitt, M. K. & Shephard, N. (1999), 'Filtering via simulation: Auxiliary particle filters', *Journal of the American Statistical Association* **94**, 590–599.

Pourahmadi, M. (1999), 'Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation', *Biometrika* **86**, 677–690.

Pourahmadi, M. (2005), A note on cholesky decompositions and maximum likelihood estimation of multivariate normal covariance matrix. manuscript.

Pourahmadi, M. & Daniels, M. J. (2002), 'Dynamic conditionally linear mixed models for longitudinal data', *Biometrics* **58**(1), 225–231.

Pourahmadi, M. & Dellaportas, P. (2002), Contemporaneous arma structures for time varying covariances with application to finance. note.

Quah, D. (1993), 'Galton's fallacy and test of the convergence hipothesis', *Scandinavian Journal of Economics* **95**(4), 427–443.

Schneider, W. (1988), 'Analytical uses of Kalman filtering in econometrics: A survey', *Statistical Papers [Formerly: Statistischen Hefte]* **29**, 3–33.

Schwarz, G. (1978), 'Estimating the dimension of a model', *The Annals of Statistics* **6**, 461–464.

Scott, S. L. (2002), 'Bayesian methods for hidden Markov models: Recursive computing in the 21st century', *Journal of the American Statistical Association* **97**(457), 337–351.

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.

Shumway, R. H. & Stoffer, D. S. (2000), *Time series analysis and its applications*, Springer-Verlag Inc.

Sims, C. A. (1972), 'The role of approximate prior restrictions in distributed lag estimation', *Journal of the American Statistical Association* **67**, 169–175.

Sims, C. A. (1980), 'Macroeconomics and reality', *Econometrica* **48**, 1–48.

Sims, C. A. (1986), 'Are forecast models usable for policy analysis?', *Quarterly Review of Federal Reserve Bank of Minneapolis* **Winter**, 2–16.

Sims, C. A. & Zha, T. (1998), 'Bayesian methods for dynamic multivariate models', *International Economic Review* **39**, 949–968.

Smith, M. & Kohn, R. (2002), 'Parsimonious covariance matrix estimation for longitudinal data', *Journal of the American Statistical Association* **97**(460), 1141–1153.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P. & van der Linde, A. (2002), 'Bayesian measures of model complexity and fit (Pkg: p583-639)', *Journal of the Royal Statistical Society, Series B, Methodological* **64**(4), 583–616.

Sun, D. & Berger, J. O. (2006), Objective priors for a multivariate normal model. *To Appear in Bayesian Statistic 8 (with discussion).*

Sun, D. & Ni, S. (2004), 'Bayesian analysis of vector-autoregressive models with noninformative priors', Journal of Statistical Planning and Inference **121**, *291–309.*

Tauchen, G. & Pitts, M. (1983), 'The price variability-volume relatioship on speculative markets (Com: p485-505)', Econometrica **51**, *485–505.*

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 (STMA V38 5003)', Econometrica **65**, *59–73.*

Vrontos, I. D., Dellaportas, P. & Politis, D. N. (2000), 'Full Bayesian inference for GARCH and EGARCH models', Journal of Business and Economic Statistics **18***(2), 187–198.*

Vrontos, I. D., Dellaportas, P. & Politis, D. N. (2003), 'A full factor multivariate garch model', Econometrics Journal **6**, *312–334.*

West, M. & Harrison, J. (1997), Bayesian forecasting and dynamic models, *Springer-Verlag Inc.*

VITA

I was born in Cagliari, Italy, on the fifth of March 1976. I completed my high school and college in the same city, graduating in Economics at the University of Cagliari in July 2000, with a thesis titles "Asymmetries of the Italian Economic System", under the advise of Prof. Beniamino Moro.

After, I worked as an econometrician for a transportation research center (CRIMM), at the University of Cagliari. Center was directed by Dr. Italo Meloni. My interest for quantitative methods convinced me to pursue a Ph.D. in statistics, and so I started the doctoral program at the University of Missouri-Columbia in January 2002, after a brief parenthesis as a visiting scholar in the same university from October to December 2001.

With my advisor, Dr. Dongchu Sun, I worked on problems related to Bayesian estimation and model selection for stochastic volatility and dynamic models. I am defending my dissertation on June 28, 2006, and then I will start working as a statistician for Capital One. The start date should be July 31, 2006.