BAYESIAN ANALYSIS IN LATENT FACTOR AND LONGITUDINAL MODELS

by

Hedibert Freitas Lopes

Institute of Statistics and Decision Sciences
Duke University

Date: _____

Approved:

_____
Dr. Mike West, Supervisor

_____
Dr. Peter Müller, Supervisor

_____
Dr. James O. Berger

_____
Dr. Gary L. Rosner

Dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy
in the Institute of Statistics and Decision Sciences
in the Graduate School of
Duke University

2000

ABSTRACT

(Multivariate Bayesian Methods)

BAYESIAN ANALYSIS IN LATENT FACTOR AND LONGITUDINAL MODELS

by

Hedibert Freitas Lopes

Institute of Statistics and Decision Sciences
Duke University

Date: _____

Approved:

_____
Dr. Mike West, Supervisor

_____
Dr. Peter Müller, Supervisor

_____
Dr. James O. Berger

_____
Dr. Gary L. Rosner

An abstract of a dissertation submitted in partial
fulfillment of the requirements for the degree
of Doctor of Philosophy in the
Institute of Statistics and Decision Sciences in the Graduate School of
Duke University

2000

Abstract

This thesis is a collection of studies in the field of multivariate Bayesian statistics.

In the first part we concentrate on model uncertainty in factor models by proposing a novel reversible jump MCMC algorithm that accounts for model uncertainty directly in the model setting. For comparison we consider a variety of strategies to compute normalizing constants. We study briefly cases where little prior information is available and default analysis must take place. We end with some simulated examples and a real application.

In the second part we use factor models to describe the covariance structure of time series, with special attention to financial time series where the factor variances have a multivariate stochastic volatility structure. We extend previous work by allowing the factor loadings, in the factor model structure, to have a time-varying structure. Simulation-based sequential analysis techniques are used in some real data applications, where predictive and financial performance are the main interest.

In the third and final part of the thesis, we propose a new way of combining information from related studies. We extend traditional random effects models to random *measure* models by allowing parameters in the model to be partially described by a probability measure common to all studies, and partially by a probability measure that is specific to each study. Both measures, common and specific, are represented by mixtures of normals. First we consider a model where the numbers of components in the mixtures are fixed; then we extend the model to an encompassing model where the number of components are treated as random in a second stage, in which case a reversible jump MCMC algorithm is needed to assess the posterior probability for the competing models. The motivation comes from meta analysis over related cancer studies.

Acknowledgements

I would like to profoundly thank professors and friends Mike West and Peter Müller for being extremely professional and challenging advisors, who always had their doors open and allowed me to freely express my views and unselfishly offered their knowledge. Their wisdom, pragmatism and guidance have positively changed my perception of scientific research and have expanded the bounds of this dissertation.

Thanks to professors Giovanni Parmigiani and Gary Rosner for all invaluable hours spent discussing various subjects and for giving me the opportunity to be involved in interesting research projects.

I especially acknowledge the friends I met at ISDS. In particular, I would like to thank my friends John, Lurdes, Marco, Gabriel, Viridiana, Omar, Jonathan, Raquel, Jenise, Courtney, Ana and Laura for being extremely supportive throughout my journey. I sincerely thank Cheryl McGhee, Patricia Johnson, Tara Nicole Scott, Lisa Squires and Sean O'Connell for all logistic assistance that allowed me to focus entirely on my research.

I dedicate this work to my wonderful daughter Anne, my parents, brothers, sisters and nephews for their unconditional love. Above all, I thank my mother, Neusa, a brilliant and lively woman that supported me in every single and imaginable aspect during all these years abroad. She has always been my source of inspiration, whose strength and passion for life have always influenced me spiritually and professionally.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The last decade has witnessed an unprecedented increase in the interest in the Bayesian paradigm. The Bayesian approach to statistical inference has been developed since quite a while, at least since the thirties, due to researchers like Bruno de Finetti, Harold Jeffreys and Leonard Savage. However, despite the fact that the Bayesian approach is intuitively simple, only recently, with the flourishing of powerful computers, has it gained researchers' and pratictioners' full attention. Because of such technology, in the late eighties and early nineties many statistical researchers became interested in Bayesian statistics and more specifically in Monte Carlo simulation. Markov chain Monte Carlo (MCMC) simulation made many practical and complex statistical problems accessible, and most importantly, is helping to spread the Bayesian viewpoint to all areas of knowledge. This thesis is a collection of studies in the field of multivariate Bayesian statistics, been exposed above. The thesis is divided into three parts, and we will briefly explain each of them.

The first part, which comprises chapters 2, 3 and 4, is focused on model uncertainty in factor models. Factor models have received enormous attention since the beginning of the century, but only recently have Bayesian researchers introduced sound inference techniques. Literature review, modeling issues and computing in

factor models are presented in Chapter 2. In Chapter 3 we propose a novel MCMC scheme that accounts for model uncertainty directly in the model setting. Also in Chapter 3 we apply to factor models a variety of recently developed strategies to compute normalizing constants required in Bayesian model assessment. We study briefly, in Chapter 4, cases where little prior information is available and default analysis must take place. We end with some simulated examples and a real application.

In the second part, correspoding to Chapters 5 and 6, we use factor models to describe the covariance structure of time series. More specifically, we concentrate on financial time series where the factor variances have a multivariate stochastic volatility structure. In Chapter 5, we extend previous work by allowing the factor loadings, in the factor model structure, to have a time-varying structure. Models are compared through their predictive and financial performance, the latter based on portfolio assessment. Simulation-based sequential analysis techniques are introduced in Chapter 6. In the two chapters, a real data application is extensively used for motivation and discussion.

Finally, Chapter 7 is part three of the thesis. We propose a new way of combining information from related studies. We extend traditional random effects models to random *measure* models by allowing parameters in the model to be partially described by a probability measure common to all studies, and partially by a probability measure that is specific to each study. Both measures, common and specific, are represented by mixtures of normals. The numbers of components in the mixtures are fixed in a first model and treated as random in an extended model, in which case a reversible jump MCMC (RJMCMC) is needed to assess the posterior probability for the competing models. The motivation comes from analyzing related cancer studies.

Chapter 2

Factor analysis

2.1 Introduction

Methodological innovations and real-world applications of factor analysis, and latent structure models more generally, have developed rapidly in recent years, partly due to increased access to appropriate computational tools. In particular, iterative MCMC simulation methods have very naturally opened up access to fully Bayesian treatments of factor analytic models, as developed and applied in, for example, Geweke and Zhou (1996), Polasek (1997), Arminger and Muthén (1998) and, with extensions to dynamic factor components in financial time series modelling (Aguilar and West, 2000; Pitt and Shephard, 1999b). The growing range of developments and creative applications in increasingly complex models, and with larger data sets in higher dimensions, justifies the view that computational advances have been critically enabling; the near future will very likely see much broader use of factor analysis in routine applied statistical work.

The above studies, and others, explore fully Bayesian inference in latent factor models in which the number of factors is a modelling choice; applied work typically studies sensitivity of predictions and variations/ambiguities of interpretations as the

number of factors is varied as a control parameter. Formal inference on the number of factors itself has been relatively ignored in the Bayesian literature, though there are ranges of standard likelihood and frequentist methods available. Some key additional references, Bayesian and non-Bayesian, include (in order of appearance) Lawley and Maxwell (1963), Joreskog (1967), Martin and McDonald (1981), Bartholomew (1981), Press (1982) (chapter 10), Lee (1981), Akaike (1987), Bartholomew (1987), Press and Shigemasu (1989), Press and Shigemasu (1994). The book by Bartholomew (1987) is an excellent overview of the field up to about ten years ago.

In this chapter we formally introduce the factor model along with some of its basic properties. Section 2.2 introduces the basic notation and the probabilistic framework in a k-factor model. Sections 2.3 and 2.4 discuss identification issues, invariance to linear transformation and the independence assumption of common factors in some details. We see, for instance, that assuming a nondiagonal covariance structure for unobserved common factor scores is irrelevant from an estimation viewpoint. The incorporation of prior information is touched in Section 2.5, while posterior analysis through Markov chain Monte Carlo is introduced in Section 2.6.

2.2 Basic model form

Data on m related variables are considered to arise through random sampling from a zero-mean multivariate normal distribution denoted by $N(\mathbf{0}, \mathbf{\Omega})$ where $\mathbf{\Omega}$ denotes the $m \times m$ non-singular variance matrix [1]. A random sample of size T is denoted by $\{\boldsymbol{y}_t, t = 1, \ldots, T\}$. For any specified positive integer $k \leq m$, the standard $k-$factor model relates each $\boldsymbol{y}_t$ to an underlying $k-$vector of random variables $\boldsymbol{f}_t$, the common factors, via

$$\boldsymbol{y}_t = \boldsymbol{\beta} \boldsymbol{f}_t + \boldsymbol{\epsilon}_t \tag{2.1}$$

[1] See Appendix A for the definition and some properties of the multivariate normal distribution.

where

- the factors $\boldsymbol{f}_t$ are independent with $\boldsymbol{f}_t \sim N(\mathbf{0}, \boldsymbol{I}_k)$,
- the $\boldsymbol{\epsilon}_t$ are independent normal $m-$vectors with $\boldsymbol{\epsilon}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma})$, and $\boldsymbol{\Sigma} = \text{diag}(\sigma_1^2, \cdots, \sigma_m^2)$,
- $\boldsymbol{\epsilon}_t$ and $\boldsymbol{f}_s$ are independent for all t and s,
- $\boldsymbol{\beta}$ is the $m \times k$ factor loadings matrix.

Under this model, the variance-covariance structure of the data distribution is constrained; we have $\boldsymbol{\Omega} = V(\boldsymbol{y}_t|\boldsymbol{\Omega}) = V(\boldsymbol{y}_t|\boldsymbol{\beta}, \boldsymbol{\Sigma})$ given by

$$\boldsymbol{\Omega} = \boldsymbol{\beta}\boldsymbol{\beta}' + \boldsymbol{\Sigma}. \tag{2.2}$$

The model implies that, conditional on the common factors, the observable variables are uncorrelated: hence the common factors explain all the dependence structure among the m variables. For any elements y_{it} and y_{jt} of $\boldsymbol{y}_t$, we have the characterising moments:

$$\begin{aligned}
\text{var}(y_{it}|\boldsymbol{\beta}, \boldsymbol{f}, \boldsymbol{\Sigma}) &= \sigma_i^2, \qquad \forall i, \\
\text{cov}(y_{it}, y_{jt}|\boldsymbol{\beta}, \boldsymbol{f}, \boldsymbol{\Sigma}) &= 0, \qquad \forall i, j, i \neq j, \\
\text{var}(y_{it}|\boldsymbol{\beta}, \boldsymbol{\Sigma}) &= \sum_{l=1}^{k} \beta_{il}^2 + \sigma_i^2, \qquad \forall i, \\
\text{cov}(y_{it}, y_{jt}|\boldsymbol{\beta}, \boldsymbol{\Sigma}) &= \sum_{l=1}^{k} \beta_{il}\beta_{jl}, \qquad \forall i, j, i \neq j.
\end{aligned}$$

In practical problems, especially with larger values of m, the number of factors k will often be small relative to m, so that much of the variance-covariance structure is explained by the common factors. The *uniquenesses,* or *idiosyncratic variances,* σ_i^2

measure the residual variability in each of the data variables once that contributed by the factors is accounted for.

The model (2.1) can be written as

$$\boldsymbol{y} = \boldsymbol{F}\boldsymbol{\beta}' + \boldsymbol{\epsilon} \tag{2.3}$$

where $\boldsymbol{y} = (\boldsymbol{y}_1, \cdots, \boldsymbol{y}_T)'$, $\boldsymbol{F} = (\boldsymbol{f}_1, \cdots, \boldsymbol{f}_T)'$ and $\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}_1, \cdots, \boldsymbol{\epsilon}_T)'$ are matrices of dimension $(T \times m), (T \times k)$ and $(T \times m)$, respectively. The elements $\boldsymbol{\epsilon}$ and $\boldsymbol{F}$ are mutually independent matrix variate normal random variables, as in Dawid (1981), Press (1982) and West and Harrison (1997) [2]. The notation, as in Dawid (1981), is simply $\boldsymbol{\epsilon} \sim N(\boldsymbol{0}, \boldsymbol{I}_T, \boldsymbol{\Sigma})$. We then have densities

$$p(\boldsymbol{y}|\boldsymbol{F}, \boldsymbol{\beta}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-T/2} \boldsymbol{etr}(-0.5\boldsymbol{\Sigma}^{-1}\boldsymbol{\epsilon}\boldsymbol{\epsilon}') \tag{2.4}$$

and, marginalising over $\boldsymbol{F}$,

$$p(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Omega}|^{-T/2} \boldsymbol{etr}(-0.5\boldsymbol{\Omega}^{-1}\boldsymbol{y}'\boldsymbol{y}) \tag{2.5}$$

where $\boldsymbol{etr}(\boldsymbol{A}) = \exp(\text{trace}(\boldsymbol{A}))$ for any square matrix $\boldsymbol{A}$. The likelihood function (2.4) will be subsequently used in Gibbs sampling for the parameters of a factor model with k fixed, whereas the likelihood form (2.5) will be extensively used in the RJMCMC algorithms and other techniques that also treat uncertainty about k to be presented in Chapter 3.

2.3 Model structure and identification issues

As is well-known, the $k-$factor model must be further constrained to define a unique model free from identification problems. First we address the standard issue that the model is invariant under transformations of the form $\boldsymbol{\beta}^* = \boldsymbol{\beta}\boldsymbol{P}'$ and $\boldsymbol{f}_t^* = \boldsymbol{P}\boldsymbol{f}_t$,

[2] See Appendix A for the definition and some properties of the matrix variate normal distribution.

where $\boldsymbol{P}$ is any orthogonal $k \times k$ matrix. There are many ways of identifying the model by imposing constraints on $\boldsymbol{\beta}$, including constraints to orthogonal $\boldsymbol{\beta}$ matrices, and constraints such that $\boldsymbol{\beta}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\beta}$ is diagonal (see Seber (1984), for example). The alternative preferred here is to constrain $\boldsymbol{\beta}$ to be a block lower triangular matrix, assumed to be of full rank. That is,

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_{11} & 0 & 0 & \cdots & 0 & 0 \\ \beta_{21} & \beta_{22} & 0 & \cdots & 0 & 0 \\ \beta_{31} & \beta_{32} & \beta_{33} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \beta_{k-1,1} & \beta_{k-1,2} & \beta_{k-1,3} & \cdots & \beta_{k-1,k-1} & 0 \\ \beta_{k,1} & \beta_{k,2} & \beta_{k,3} & \cdots & \beta_{k,k-1} & \beta_{k,k} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \beta_{m,1} & \beta_{m,2} & \beta_{m,3} & \cdots & \beta_{m,k-1} & \beta_{m,k} \end{pmatrix} \tag{2.6}$$

where the diagonal elements β_{ii} are strictly positive. This form is used, for example, in Geweke and Zhou (1996) and Aguilar and West (2000), and provides both identification and, often, useful interpretation of the factor model. In this form, the loadings matrix has $r = mk - k(k-1)/2$ free parameters. With m non-zero σ_i parameters, the resulting factor form of $\boldsymbol{\Omega}$ has $m(k+1) - k(k-1)/2$ parameters, compared with the total $m(m+1)/2$ in an unconstrained (or $k = m$) model. This leads to the constraint that

$$m(m+1)/2 - m(k+1) + k(k-1)/2 \geq 0 \tag{2.7}$$

which provides an upper bound on k. For example, $m = 6$ implies $k \leq 3$, $m = 12$ implies $k \leq 7$, $m = 20$ implies $k \leq 14$, $m = 50$ implies $k \leq 40$, and so on. Even for small m, the bound will often not matter as relevant k values will not be so large. In realistic problems, with m in double digits or more, the resulting bound will rarely matter. Finally, note that the number of factors can be increased beyond such bounds by setting one or more of the residual variances σ_i to zero. This is similar to rank restrictions usually present in simultaneous equations estimation of

econometric data. When k is larger than the maximum number of factors we have an overidentified model, in econometric terms, and Ω from equation 2.2 is not well defined.

A question arises about the full-rank assumption for $\boldsymbol{\beta}$. This was addressed in Geweke and Singleton (1980) who shown that, if $\boldsymbol{\beta}$ is rank deficient, then the model is unidentified. Specifically, if $\boldsymbol{\beta}$ has rank $r < k$ there exists a matrix $\boldsymbol{Q}$ such that $\boldsymbol{\beta Q} = \mathbf{0}$, $\boldsymbol{Q'Q} = \boldsymbol{I}$ and, for any orthogonal matrix $\boldsymbol{M}$,

$$\boldsymbol{\Omega} = \boldsymbol{\beta\beta'} + \boldsymbol{\Sigma} = (\boldsymbol{\beta} + \boldsymbol{MQ'})'(\boldsymbol{\beta} + \boldsymbol{MQ'}) + (\boldsymbol{\Sigma} - \boldsymbol{MM'}). \tag{2.8}$$

This translation invariance of $\boldsymbol{\Omega}$ under the factor model implies lack of identification and, in application, induces symmetries and potential multimodalities in resulting likelihood functions. This issue relates intimately to the question of uncertainty of the number of factors, discussed further below.

A final question concerns the ordering of the y_{it} variables and the connection between a chosen ordering and the specific form of the factor loading matrix above. The order of variables is a modelling decision that has no effect on the resulting theoretical model nor on predictive inferences under the model. Given the k-factor model (2.1) specified and appropriate for the $\boldsymbol{y}$ with variables in a specific order, alternative orderings are trivially produced via $\boldsymbol{Ay}_t$ for some rotation matrix $\boldsymbol{A}$. Model (2.1) then transforms to a similar factor model for the reordered data $\boldsymbol{Ay}_t$ with the same latent factors but transformed loadings matrix $\boldsymbol{A\beta}$. This new loadings matrix does not have the lower triangular structure. However, we can always find an orthonormal matrix $\boldsymbol{P}$ such that $\boldsymbol{A\beta P'}$ is lower triangular, and so simply recover the factor model in precisely the form (2.1) with the same probability structure for the underlying latent factors $\boldsymbol{Pf}_t$. This result confirms that the order of the variables in $\boldsymbol{y}_t$ is theoretically irrelevant assuming that k is properly chosen. However, when it comes to model estimation, the order of variables has a determining effect on the

choice of k, and the interaction between variable order and model fitting can be quite subtle, as we illustrate in examples below.

2.4 Independent common factors

In this section we show that whether $E(\boldsymbol{f}\boldsymbol{f}')$ is diagonal or not is irrelevant, as far as a static factor model is concerned.

Let us start assuming that $\boldsymbol{y}$ follows a k-factor model with dependent common factors, ie. $\boldsymbol{y} = \boldsymbol{\beta}\boldsymbol{f} + \boldsymbol{\epsilon}$ with $\boldsymbol{\epsilon} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma})$ and $\boldsymbol{f} \sim N(\boldsymbol{0}, \boldsymbol{H})$ where $\boldsymbol{H} > 0$ is not restricted to be diagonal.

Since $\boldsymbol{H} > 0$, $\boldsymbol{L}$ we can found such that $\boldsymbol{H} = \boldsymbol{L}\boldsymbol{L}'$ and $\boldsymbol{L}^{-1}\boldsymbol{H}(\boldsymbol{L}')^{-1} = \boldsymbol{L}^{-1}\boldsymbol{H}(\boldsymbol{L}^{-1})' = \boldsymbol{I}$. Then, the new factor model with $\boldsymbol{\beta}$ replaced by $\tilde{\boldsymbol{\beta}} = \boldsymbol{\beta}\boldsymbol{L}$ and the common factors replaced by $\tilde{\boldsymbol{f}} = \boldsymbol{L}^{-1}\boldsymbol{f}$, has independent common factor structure.

To recover the lower triangular property of $\boldsymbol{\beta}$, the following fact is used; there exists $\boldsymbol{P}$ such that $\boldsymbol{P}'\boldsymbol{P} = \boldsymbol{P}\boldsymbol{P}' = \boldsymbol{I}$ and

$$\boldsymbol{\beta}^* = \tilde{\boldsymbol{\beta}}\boldsymbol{P}' = \boldsymbol{\beta}\boldsymbol{L}\boldsymbol{P}'$$

is lower triangular with positive real numbers on the main diagonal.

A particular expression for $\boldsymbol{P}$ is $\boldsymbol{U}^{-1}\tilde{\boldsymbol{\beta}}_1$, where

$$\tilde{\boldsymbol{\beta}} = (\tilde{\boldsymbol{\beta}}_1', \tilde{\boldsymbol{\beta}}_2')'$$

and

$$\boldsymbol{U}\boldsymbol{U}' = \tilde{\boldsymbol{\beta}}_1\tilde{\boldsymbol{\beta}}_1'$$

It follows that,

$$\begin{aligned} \boldsymbol{P}'\boldsymbol{P} &= (\boldsymbol{U}^{-1}\tilde{\boldsymbol{\beta}}_1)'(\boldsymbol{U}^{-1}\tilde{\boldsymbol{\beta}}_1) = \tilde{\boldsymbol{\beta}}_1'(\boldsymbol{U}')^{-1}\boldsymbol{U}^{-1}\tilde{\boldsymbol{\beta}}_1 \\ &= \tilde{\boldsymbol{\beta}}_1'(\boldsymbol{U}\boldsymbol{U}')^{-1}\tilde{\boldsymbol{\beta}}_1 = \tilde{\boldsymbol{\beta}}_1'(\tilde{\boldsymbol{\beta}}_1\tilde{\boldsymbol{\beta}}_1')^{-1}\tilde{\boldsymbol{\beta}}_1 = \boldsymbol{I}, \end{aligned}$$

by the definition of $\boldsymbol{P}$; $\boldsymbol{P}$, and

$$\begin{aligned} \boldsymbol{P}\boldsymbol{P}' &= \boldsymbol{U}^{-1}\tilde{\boldsymbol{\beta}}_1\tilde{\boldsymbol{\beta}}_1'(\boldsymbol{U}^{-1})' \\ &= \boldsymbol{U}^{-1}\boldsymbol{U}\boldsymbol{U}'(\boldsymbol{U}')^{-1} = \boldsymbol{I}, \end{aligned}$$

This result has been overlooked by most researchers in factor analysis and simplifies matters considerably. For the rest of the next few chapters we will assume that the common factors are, *a priori*, independent. In the next section we set up the prior information.

2.5 Elements of prior specification

To complete the model specification we require classes of priors for the model parameters $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$. Our reported analyses are based on very diffuse but proper priors with the following ingredients. For the factor loadings, we take independent priors such that $\beta_{ij} \sim N(0, C_0)$ when $i \neq j$, and $\beta_{ii} \sim TN(0, C_0)$[3] for the upper-diagonal elements of positive loadings $i = 1, \cdots, k$. The latter simply truncates the basic normal prior to restrict the diagonal elements to positive values. Analysis now requires only that we specify the variance parameter C_0, which we take to be rather large in the studies below.

For each of the idiosyncratic variances σ_i^2 we assume a common inverse gamma prior, and take the variances to be independent. Specifically, the σ_i^2 are independently modelled as $\sigma_i^2 \sim IG(\nu/2, \nu s^2/2)$ with specified hyperparameters ν and s^2. Here s^2 is the prior mode of each σ_i^2 and ν the prior degrees of freedom hyperparameter[4]. Our examples below assume values of ν to produce diffuse though proper priors. Note that we eschew the use of standard improper reference priors $p(\sigma_i^2) \propto 1/\sigma_i^2$. Such

[3]See Appendix A for the definition and some properties of the truncated normal distribution.

[4]See Appendix A for the definition and some properties of the inverse gamma distribution.

priors lead to the Bayesian analogue of the so-called *Heywood problem* (Martin and McDonald, 1981; Ihara and Kano, 1995). In terms of these variance parameters, likelihood functions in factor models are bounded below away from zero as σ_i^2 tends to zero, so inducing singularities in the posterior at zero. Proper priors that decay to zero at the origin obviate this problem and induce proper posteriors.

2.6 MCMC methods in a $k-$factor model

With a specified $k-$factor model, Bayesian analyses using MCMC methods are straightforward. We simply summarise the main ingredients here, referring to Geweke and Zhou (1996), Polasek (1997), and Aguilar and West (2000) for further details. MCMC analysis involves iteratively simulating from sets of conditional posterior distributions which, in this model, are standard forms. A basic method simulates from the conditional posteriors for each of $\boldsymbol{F}$, $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ in turn, utilising the following sets of full conditional posteriors arising from our model as specified. These are as follows.

First, the factor model in (2.3) can be seen as a standard multivariate regression model with "parameters" $\boldsymbol{F}$ when $\boldsymbol{\beta}$, $\boldsymbol{\Sigma}$ and k are fixed (e.g., Press (1982), Box and Tiao (1973), Broemeling (1985) and Zellner (1971)). It easily follows that the full conditional posterior for $\boldsymbol{F}$ factors into independent normal distributions for the $\boldsymbol{f}_t$, namely

$$\boldsymbol{f}_t \sim N((\boldsymbol{I}_k + \boldsymbol{\beta}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\beta})^{-1}\boldsymbol{\beta}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y}_t, (\boldsymbol{I}_k + \boldsymbol{\beta}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\beta})^{-1})$$

independently for $t = 1, \ldots .T$.

Second, the full conditional posterior for $\boldsymbol{\beta}$ also factors into independent margins for the non-zero elements of the rows of $\boldsymbol{\beta}$, as follows. For rows $i = 1, \ldots, k$, write $\boldsymbol{\beta}_i = (\beta_{i1}, \ldots, \beta_{ii})'$ for just these non-zero elements. For the remaining rows $i = k+1, \ldots, m$, write $\boldsymbol{\beta}_i = (\beta_{i1}, \ldots, \beta_{ik})$. Similarly, for $i = 1, \ldots, k$ denote by $\boldsymbol{F}_i$ the $T \times i$ matrix containing the first i columns of $\boldsymbol{F}$, and for all i let $\boldsymbol{y}_i$ be the column i

of $\boldsymbol{y}$.

Finally, it is trivially deduced that full conditional posterior for the elements of $\boldsymbol{\Sigma}$ reduces to a set of m independent inverse gammas, with $\sigma_i^2 \sim IG((\nu+T)/2, (\nu s^2 + d_i)/2)$ where $d_i = (\boldsymbol{y}_i - \boldsymbol{F}\boldsymbol{\beta}_i')'(\boldsymbol{y}_i - \boldsymbol{F}\boldsymbol{\beta}_i')$.

Then we have full conditionals as follows:

- for $i = 1, \ldots, k$, $\boldsymbol{\beta}_i \sim N(\boldsymbol{m}_i, \boldsymbol{C}_i)\mathbf{1}(\beta_{ii} > 0)$ where $\boldsymbol{m}_i = \boldsymbol{C}_i \left(C_0^{-1}\mu_0 \mathbf{1}_i + \sigma_i^{-2}\boldsymbol{F}_i'\boldsymbol{y}_i\right)$ and $\boldsymbol{C}_i^{-1} = C_0^{-1}\boldsymbol{I}_i + \sigma_i^{-2}\boldsymbol{F}_i'\boldsymbol{F}_i$;
- for $i = k+1, \ldots, m$, $\boldsymbol{\beta}_i \sim N(\boldsymbol{m}_i, \boldsymbol{C}_i)$ where $\boldsymbol{m}_i = \boldsymbol{C}_i \left(C_0^{-1}\mu_0 \mathbf{1}_k + \sigma_i^{-2}\boldsymbol{F}'\boldsymbol{y}_i\right)$ and $\boldsymbol{C}_i^{-1} = C_0^{-1}\boldsymbol{I}_k + \sigma_i^{-2}\boldsymbol{F}'\boldsymbol{F}$.

These distributions are easily simulated.

2.7 Summary

In this chapter we reviewed the state of the art in Bayesian factor models along with model issues, such as invariance to linear transformation, identifiability constraints, prior information and MCMC methods for posterior inference analysis when the number of common factors, k is fixed. Chapter 3 explores simulated and real data applications that rely on the methodology here.

We have explored a variety of simulated and real datasets to test the MCMC algorithm. Some of them are fully explored in the next chapter, when model uncertainty for the number of factors is also considered. However, before moving forward, the following comments are worth making:

- When the number of factors is known to be correct and the factor loadings' prior information is relatively scarce (represented by large values for C_0, for instance), the MCMC algorithm converges quickly for fairly large datasets and posterior first moments converge to the classical maximum likelihood estimators.

- When the number of fitted factors is larger then what is really necessary, posterior multimodality shows up. We have anticipated this behavior in Section 2.3.

- In most of the applications, the convergence of the MCMC outputs were basically assessed by eyeballing trace plots from different starting values. In some particular cases, where more formal jugdment was needed, we used some of the well-known convergence diagnostic tests available in the BOA software[5]. Mengersen *et al.* (1999), and Cowles and Carlin (1996), amongst others, provide up-to-date and thorough reviews of MCMC convergence diagnostics.

[5]BOA stands for "Bayesian Output Analysis" Program. As of spring 2000, BOA is publicly available at http://www.public-health.uiowa.edu/boa

Chapter 3

Inference on the number of factors

3.1 Introduction

The focus of this chapter is inference on, and selection of the number of factors. Most recently, Polasek (1997) explored approaches to computing approximate posterior probabilities on the number of factors based on using MCMC methods for separate models differing only in the number of factors. Such an approach requires the computation of the marginal data densities (prior predictive densities) under each of these separate models, for it is just these values that define the (marginal) likelihood function for inference on the number of factors, and the resulting Bayes' factors for pairwise model comparisons. This computation lies at the heart of the model selection and comparison problem. There are serious practical questions about choice and specification of prior distributions within the individual models, but that is not our primary focus here initially.

A variety of methods are available for computing these marginal data density values – often referred to as the normalising constant problem. Some are specific to analyses based on MCMC methods for each individual sub-model, and some are generic and based on analytic and asymptotic arguments. A review of some stan-

dard methods appears in Kass and Raftery (1995), where the connections between various methods of approximating Bayes' factors using combinations of analytic and asymptotic arguments are explored. These standard methods are closely related to non-Bayesian model selection criteria, including the AIC, BIC/Schwartz criteria, and extensions of them using information-theoretic ideas, such as the ICOMP methods of Bozdogan and Ramirez (1987) and Bozdogan and Shigemasu (1998). In this dissertation we focus on methods of approximating the marginal data densities that utilize outputs from MCMC analyses of separate models. Some of the methods we consider below are: the so-called candidate formula (Chib, 1995), the harmonic mean estimator (Newton and Raftery, 1994), Gelfand and Dey's estimator (Gelfand and Dey, 1994), the Laplace-Metropolis estimator (Lewis and Raftery, 1997), and various novel approaches based on the recent innovative developments in bridge sampling (Meng and Wong, 1996). Additional useful references in this general area include, for example, Gilks *et al.* (1996), DiCiccio *et al.* (1997) and Godsill (1998), which study comparisons and connections between some of the various methods just referenced.

First, we introduce, in Sections 3.2 and 3.3, a customised reversible jump Markov chain Monte Carlo (hereafter RJMCMC, see Green (1995)) algorithm for moving between models with different numbers of factors. RJMCMC approaches avoid the need for computing marginal data densities by treating the number of factors as a parameter, but require ingenuity in designing appropriate jumping rules to produce computationally efficient and theoretically effective methods. To compare with this, we introduce, in Section 3.4 alternative methods based on bridge sampling ideas (Meng and Wong, 1996) that are specifically designed for computing the required marginal data densities in MCMC contexts. Section 3.4 also presents other alternative ways of computing normalizing constants based on MCMC outputs as mentioned in the last paragraph. Likelihood-based criteria, such as Akaike's information crite-

rion, are explored in Section 3.5. Simulated and real data applications are studied in Sections 3.6 and 3.7. Section 3.8 closes the chapter with some final comments about the simulated and real data application and perspectives concerning model uncertainty in factor analysis situations.

3.2 Preliminary parallel MCMC analyses

The last chapter provided the basis for posterior simulations in a model with k specified. For this and the following sections, we make explicit the dependence of the factor loading matrix on k by refining the notation, replacing $\boldsymbol{\beta}$ by $\boldsymbol{\beta}_k$ and $\boldsymbol{F}$ by $\boldsymbol{F}_k$. Further, we write $\boldsymbol{\theta}_k$ for the parameters $(\boldsymbol{\beta}_k, \boldsymbol{\Sigma})$ of a $k-$factor model. The number k now appears explicitly in the conditioning of all model density functions. Reversible jump MCMC (RJMCMC) methods are useful for exploring posterior distributions for model parameters in the context of uncertainty about k, and with k included as a parameter. As we move between models with different numbers of factors, the dimension and meaning of the model parameters change, and RJMCMC methods are designed for just such problems. See Appendix B for further detail and references about the RJMCMC algorithm.

Our method builds on a preliminary set of parallel MCMC simulations that are run over a set of prespecified values $k \in K$ for the number of factors. These chains produce a set of K posterior samples for $(\boldsymbol{\theta}_k, \boldsymbol{F}_k)$ that approximate the posterior distributions $p(\boldsymbol{\theta}_k, \boldsymbol{F}_k | k, \boldsymbol{y})$ for the sub-models $k = 1, \ldots, K$.

From these samples we compute posterior means and other summaries, and use these to guide choice of analytically specified distributions to be used to generate proposals in the RJMCMC algorithm. This component of the analysis operates only with the samples for the parameters $\boldsymbol{\theta}_k$, the simulated values of the actual factors $\boldsymbol{F}_k$ being relevant but incidental to moving between models with different values of

k. Write $\boldsymbol{b}_k$ and $\boldsymbol{B}_k$ for the approximate posterior mean and variance matrix of $\boldsymbol{\beta}_k$ from the MCMC analysis, and, for each $i = 1, \ldots, m$, write v_{ki}^2 for the approximate posterior mode of σ_i^2 from the analysis. In our current implementation we introduce the following analytic forms as components of a proposal distribution. For each model order $k \in K$,

- $q_k(\boldsymbol{\beta}_k) = N(\boldsymbol{b}_k, b\boldsymbol{B}_k)$ and
- for $i = 1, \ldots, m$, $q_k(\sigma_i^2) = IG(a, av_{ki}^2)$,

where a and b, two positive scale parameters to be specified. These density functions are combined to produce the distributions

$$q_k(\boldsymbol{\theta}_k) \equiv q_k(\boldsymbol{\beta}_k, \boldsymbol{\Sigma}) = q_k(\boldsymbol{\beta}_k) \prod_{i=1}^{m} q_k(\sigma_i^2), \qquad k \in K \tag{3.1}$$

for use as now described.

3.3 A reversible jump algorithm

Following the set of preliminary MCMC analyses for the sub-models, we explore the space of models as k varies using the following version of RJMCMC. The summary here is also schematically represented in Figure 3.1. In addition to the $k-$factor models and within model priors specified above, we need to specify the marginal prior probabilities $p(k)$ over $k \in K$. Then the RJMCMC analysis proceeds as follows.

0. Choose a starting value of k. Set the current values of $\boldsymbol{\theta}_k$ to a draw from the posterior $p(\boldsymbol{\theta}_k|k, \boldsymbol{y})$ by using one (or more) steps of the MCMC algorithm as described above and based on past sampled values from this $k-$factor model. Note that this step produces both new sampled values of $\boldsymbol{\theta}_k$ and the factors $\boldsymbol{F}_k$, though only the former are used in exploring moves to models with other k values.

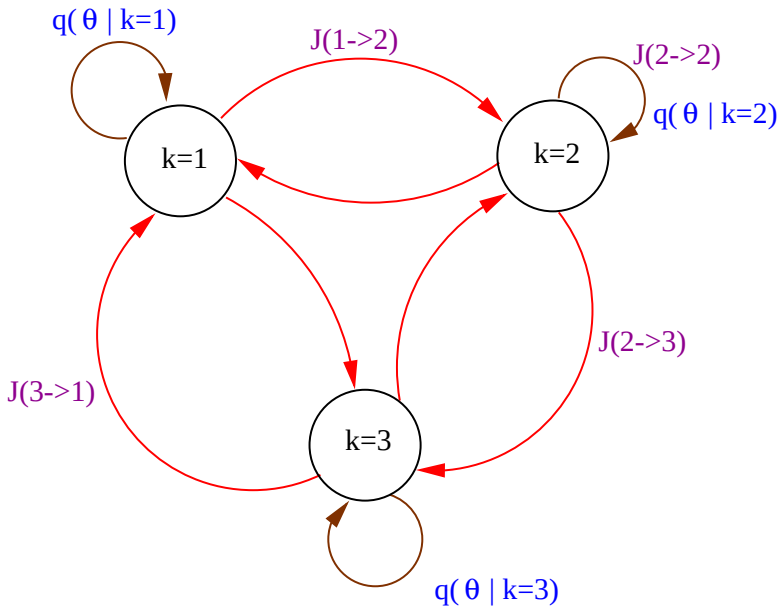


Figure 3.1: Representation of RJMCMC algorithm for exploring the space of $k-$factor models over a range of k values. The figure represents the example $k \in K = (1, 2, 3)$

1. *Between model* move step:

1.a Draw a candidate value of the number of factors k' from a proposal distribution defined by prespecified transition probabilities $Pr(k'|k) = J(k \to k')$.

1.b Given k', draw the parameters $\boldsymbol{\theta}_{k'}$ from the distribution $q_{k'}(\boldsymbol{\theta}_{k'})$ of equation (3.1).

1.c Compute the accept/reject ratio

$$\alpha = \min\left\{1, \frac{p(\boldsymbol{y}|k', \boldsymbol{\theta}_{k'})p(\boldsymbol{\theta}_{k'}|k')p(k')}{p(\boldsymbol{y}|k, \boldsymbol{\theta}_k)p(\boldsymbol{\theta}_k|k)p(k)} \frac{q_k(\boldsymbol{\theta}_k)J(k' \to k)}{q_{k'}(\boldsymbol{\theta}_{k'})J(k \to k')}\right\}. \tag{3.2}$$

Here, for each value $j \in (k, k')$, $p(\boldsymbol{y}|j, \boldsymbol{\theta}_j) = p(\boldsymbol{y}|j, \boldsymbol{\beta}_j, \boldsymbol{\Sigma})$ is the likelihood function in equation (2.5), $p(\boldsymbol{\theta}_j|j)$ is the prior density function for the parameters within the $j-$factor model, and $p(j)$ is the prior probability on k factors. With probability α, accept the jump to the $k'-$factor model

and the new parameter values $\boldsymbol{\theta}_{k'}$ just sampled as candidates.

2. *Within model* move step:

If the jump to model k' is accepted, run one step of the MCMC analysis in this $k'-$factor model, producing new sample values of the full set of quantities $(\boldsymbol{\theta}_{k'}, \boldsymbol{F}_{k'})$. Otherwise, remain in model k and use the MCMC to produce new values of $(\boldsymbol{\theta}_k, \boldsymbol{F}_k)$.

3. Repeat [1] and [2] until practical convergence is judged to have been achieved.

The chosen proposal distributions $q_k(\boldsymbol{\theta}_k)$ are not generally expected to provide globally accurate approximations to the conditional posteriors $p(\boldsymbol{\theta}_k|k, \boldsymbol{y})$. However, if that happened to be the case then the resulting accept/reject probabilities above reduce directly to Metropolis-type probabilities on the parameter k alone. Our algorithm is a particular case of what Dellaportas *et al.* (1998) and Godsill (1998) call the *Metropolized Carlin and Chib* method, where the proposal distributions generating both new model dimension and new parameters depend on the current state of the chain only through k. More specifically, our proposal densities, $p_k(\boldsymbol{\theta}_k)$ and $p_{k'}(\boldsymbol{\theta}_{k'})$ from equation (3.1) replace the pseudo-prior densities $p(\boldsymbol{\theta}_k|k')$ and $p(\boldsymbol{\theta}'_k|k)$ that appears in the metropolized Carlin and Chib method from equation (B.9) (see appendix B). This is true here as we use proposals based on the initial, auxilliary MCMC analyses. A more descriptive name is independence RJMCMC, analogous to the standard terminology for independence Metropolis-Hastings methods. See Appendix B for further details.

3.4 Computing normalizing constants

The RJMCMC technology is becoming standard in Bayesian work with competing models with differing numbers of parameters. By comparison with traditional

approaches based on nesting models in a "super-model", RJMCMC is often more efficient computationally and in terms of computing time to pratical convergence of the Markov chains, and it has an established theory that guarantees convergence of the chains in very general frameworks. Thus, in addition to having a very natural and direct specification in our factor model context, convergence of the chain to sampling from the full posterior across models, as well as for parameters and factors within models, is ensured. There are, however, ranges of existing methods for approximate inference on the number of factors, and we aim to compare these methods in examples below. This section provides a brief catalogue description of methods and model selection criteria, as well as introducing a novel approach based on recent work in bridge sampling.

In our Bayesian framework the within-model analysis provides, in theory, the marginal data density functions

$$p(\boldsymbol{y}|k) = \int p(\boldsymbol{y}|k, \boldsymbol{\theta}_k) p(\boldsymbol{\theta}_k|k) d\boldsymbol{\theta}_k \tag{3.3}$$

for each value of $k \in K$. If these could be computed, then inference on k follows from Bayes' theorem via $p(k|\boldsymbol{y}) \propto p(k)p(\boldsymbol{y}|k)$. The problem is computational: the marginal data densities are generally not easily computed and so much be approximated numerically. The following standard methods are of interest.

3.4.1 Candidate's estimator

The so-called candidate's estimator, first referred in Besag (1989) and fully analysed by Chib (1995), is of interest when the $k-$factor models are each analysed using MCMC. The approach observes that, for *any* value of $\boldsymbol{\theta}_k$ Bayes' theorem implies that

$$p(\boldsymbol{y}|k) = \frac{p(\boldsymbol{y}|k, \boldsymbol{\theta}_k) p(\boldsymbol{\theta}_k|k)}{p(\boldsymbol{\theta}_k|k, \boldsymbol{y})} \tag{3.4}$$

The idea is then to estimate the components of this equation that are not available analytically, then plug-in a chosen value of $\boldsymbol{\theta}_k$ to provide an estimate of $p(\boldsymbol{y}|k)$. Our analysis uses the posterior means from the MCMC analyses as these plug-in values.

Now, the numerator in equation (3.4) factors as $p(\boldsymbol{y}|k, \boldsymbol{\beta}_k, \boldsymbol{\Sigma})p(\boldsymbol{\beta}_k|k)p(\boldsymbol{\Sigma}|k)$ each component of which can be directly and easily evaluated. The posterior density function in the denominator requires approximation, and with some creativity. Theoretically, this is given by

$$p(\boldsymbol{\beta}_k, \boldsymbol{\Sigma}|k, \boldsymbol{y}) = p(\boldsymbol{\beta}_k|k, \boldsymbol{y})p(\boldsymbol{\Sigma}|k, \boldsymbol{\beta}_k, \boldsymbol{y}) \tag{3.5}$$

where the two terms in the right hand side are approximated, respectively by:

$$p(\boldsymbol{\beta}_k|k, \boldsymbol{y}) \approx \frac{1}{M}\sum_{m=1}^{M} p(\boldsymbol{\beta}_k|k, \boldsymbol{\Sigma}^{(m)}, \boldsymbol{F}_k^{(m)}, \boldsymbol{y}) \tag{3.6}$$

and

$$p(\boldsymbol{\Sigma}|k, \boldsymbol{\beta}_k, \boldsymbol{y}) \approx \frac{1}{M}\sum_{m=1}^{M} p(\boldsymbol{\Sigma}|k, \boldsymbol{\beta}_k, \boldsymbol{F}_{k1}^{(m)}, \boldsymbol{y}) \tag{3.7}$$

where the sum, in the first approximation, is over draws $(\boldsymbol{F}_k^{(m)}, \boldsymbol{\Sigma}^{(m)})$ from the MCMC analysis and can be easily performed, since it is a sum of multivariate normal distributions (truncated).

The second approximation, however, is a sum (of products of inverse gammas) over draws $\boldsymbol{F}_{k1}^{(m)}$ from an MCMC analysis conditional on $\boldsymbol{\beta}_k$. To do this we need to run additional MCMC chains in each $k-$factor model with the $\boldsymbol{\beta}_k$ fixed at its chosen value. This naturally introduces a significant additional computational burden, especially in larger models.

In studies below we denote by $\hat{p}_C$ the resulting approximation to $p(\boldsymbol{y}|k)$.

3.4.2 Harmonic mean estimator

In a similar spirit to the candidate's estimator, the harmonic mean estimator makes use of the identity

$$p(\boldsymbol{y}|k)^{-1} = \int p(\boldsymbol{y}|k, \boldsymbol{\theta}_k)^{-1} p(\boldsymbol{\theta}_k|k, \boldsymbol{y}) d\boldsymbol{\theta}_k. \tag{3.8}$$

As discussed in Newton and Raftery (1994), the resulting estimator is based on the importance sampling approximation to the integral using the exact posterior as importance sampling distribution. This results in the approximation $p(\boldsymbol{y}|k) \approx \hat{p}_H$ where

$$\hat{p}_H^{-1} = M^{-1} \sum_{m=1}^{M} p(\boldsymbol{y}|k, \boldsymbol{\theta}_k^{(m)})^{-1} \tag{3.9}$$

where the $\boldsymbol{\theta}_k^{(m)}$ are posterior samples from the MCMC analysis and the density evaluations are made using equation (2.5). Note that $\hat{p}_H$ is an harmonic mean of the likelihood values, hence the name. Newton and Raftery (1994) discuss the accuracy of $\hat{p}_H$ among other issues. Though it has been quite widely used, it can be unstable in some applications since small likelihood values can overly influence the resulting harmonic mean value.

3.4.3 Newton and Raftery's estimator

Partly motivated by the stability issues associated with $\hat{p}_H$, Newton and Raftery (1994) suggested estimators defined as follows. Let $g(\boldsymbol{\theta}_k) = \delta p(\boldsymbol{\theta}_k|k) + (1-\delta) p(\boldsymbol{\theta}_k|k, \boldsymbol{y})$ be a mixture of the prior and posterior for $\boldsymbol{\theta}_k$ for some small mixing probability δ. Sampling from $g(\cdot)$ is easy – simply randomly replace values in the available posterior sample by independent draws from the prior. Do this iteratively, repeatedly computing the sequence of γ values defined by

$$\gamma^{new} = A(\gamma^{old})/B(\gamma^{old})$$

where

$$A(\gamma) = \sum_{m=1}^{M} f_m\{\delta\gamma + (1-\delta)f_m\}^{-1}$$

$$B(\gamma) = \sum_{m=1}^{M} \{\delta\gamma + (1-\delta)f_m\}^{-1}$$

where the quantities f_m are the likelihood values $f_m = p(\boldsymbol{y}|k, \boldsymbol{\theta}_k^{(m)})$ evaluated at the current sample of M points $\boldsymbol{\theta}_k^{(m)}$ from the mixture $g(\cdot)$. Iterations of this procedure lead to a stable limiting value $\gamma = \hat{p}_{NR}$ that provides the proposed estimator of $p(\boldsymbol{y}|k)$. Notice that the case $\delta = 0$ implies that $\hat{p}_{NR} = \hat{p}_H$. Implementation of this method depends on the choice of δ.

The main problem with this estimator is that is depends on draws from both the posterior and the prior. Newton and Raftery (1994) have suggested combining the M draws from the posterior with $\delta M/(1-\delta)$ draws from the the prior, all of them with the same likelihood equal to their expected value, $p(\boldsymbol{y}|K)$, the predictive density. Once again, the solution is found iteratively by,

$$\gamma^{new} = A(\gamma^{old})/B(\gamma^{old})$$

where

$$A(\gamma) = \epsilon M + \sum_{m=1}^{M} f_m\{\delta\gamma + (1-\delta)f_m\}^{-1},$$

$$B(\gamma) = \epsilon M\gamma^{-1} + \sum_{m=1}^{M} \{\delta\gamma + (1-\delta)f_m\}^{-1},$$

and $\epsilon = \delta/(1-\delta)$.

3.4.4 Gelfand and Dey's estimator

Related estimators introduced by Gelfand and Dey (1994) are inspired by the identities

$$p(\boldsymbol{y}|k)^{-1} = \int g(\boldsymbol{\theta}_k)p(\boldsymbol{\theta}_k|k,\boldsymbol{y})\{p(\boldsymbol{\theta}_k|k)p(\boldsymbol{y}|k,\boldsymbol{\theta}_k)\}^{-1}d\boldsymbol{\theta}_k \tag{3.10}$$

that hold for arbitrary densities $g(\cdot)$. These authors study approximations $\hat{p}_{GD}$ given by

$$\hat{p}_{GD}^{-1} = M^{-1}\sum_{m=1}^{M} g(\boldsymbol{\theta}_k^{(m)})\{p(\boldsymbol{y}|k,\boldsymbol{\theta}_k^{(m)})p(\boldsymbol{\theta}_k^{(m)}|k)\}^{-1} \tag{3.11}$$

where, again, the $\boldsymbol{\theta}_k^{(m)}$ are posterior samples.

3.4.5 Laplace-Metropolis estimator

The Laplace-Metropolis estimator combines analytic posterior approximations with MCMC output to modify traditional Laplace asymptotics (Tierney and Kadane, 1986). Discussed in Lewis and Raftery (1997), the resulting estimator has the form

$$\hat{p}_{LM} = (2\pi)^{d/2}|\boldsymbol{\Psi}|^{1/2}p(\boldsymbol{y}|k,\tilde{\boldsymbol{\theta}}_k)p(\tilde{\boldsymbol{\theta}}_k|k) \tag{3.12}$$

where $\tilde{\boldsymbol{\theta}}_k$ maximises $p(\boldsymbol{y}|k,\boldsymbol{\theta}_k)p(\boldsymbol{\theta}_k|k)$ among the M posterior draws, $\boldsymbol{\Psi}$ is the MCMC approximation to the posterior variance of $\boldsymbol{\theta}_k$, and d is the dimension of $\boldsymbol{\theta}_k$. Variations on this method replace $\tilde{\boldsymbol{\theta}}_k$ by the MCMC approximation to the posterior mean.

3.4.6 Bridge sampling

Innovative methods based on bridge sampling have recently been studied by Meng and Wong (1996). In our context this applies as follows. If $g(\boldsymbol{\theta}_k)$ is any chosen proposal density function with the same support as the posterior $p(\boldsymbol{\theta}_k|k,\boldsymbol{y})$, note the identity $p(\boldsymbol{y}|k) = N/D$ where

$$N = \int \alpha(\boldsymbol{\theta}_k)g(\boldsymbol{\theta}_k)p(\boldsymbol{\theta}_k|k)p(\boldsymbol{y}|k,\boldsymbol{\theta}_k)d\boldsymbol{\theta}_k$$

and

$$D = \int \alpha(\boldsymbol{\theta}_k) g(\boldsymbol{\theta}_k) p(\boldsymbol{\theta}_k | k, \boldsymbol{y}) d\boldsymbol{\theta}_k.$$

Based on the MCMC sample values $\boldsymbol{\theta}_k^{(m)}$ from the posterior and a parallel importance sample of some L values $\boldsymbol{\theta}_k^{*(l)}$ from $g(\cdot)$, we now have an easy approximation $p(\boldsymbol{y}|k) \approx \hat{N}/\hat{D}$ where

$$\hat{N} = L^{-1} \sum_{l=1}^{L} \alpha(\boldsymbol{\theta}_k^{*(l)}) p(\boldsymbol{\theta}_k^{*(l)} | k) p(\boldsymbol{y} | k, \boldsymbol{\theta}_k^{*(l)})$$

and

$$\hat{D} = M^{-1} \sum_{m=1}^{M} \alpha(\boldsymbol{\theta}_k^{(m)}) g(\boldsymbol{\theta}_k^{(m)}).$$

Generally, we aim to choose $g(\cdot)$ to be as accurate an approximation to the posterior as possible, while remaining easy to compute and simulate.

Different choices of the "arbitrary" function $\alpha(\cdot)$ define different bridge sampling estimators. Some discussed by Meng and Wong (1996) are as follows.

- If $\alpha(\boldsymbol{\theta}_k) = (p(\boldsymbol{\theta}_k|k) p(\boldsymbol{y}|k, \boldsymbol{\theta}_k) g(\boldsymbol{\theta}_k))^{-1}$ the corresponding estimator resembles the harmonic mean estimator. We label this $\hat{p}_{GH}$ and note that it obtains when

  $$\hat{N} = L^{-1} \sum_{l=1}^{L} g(\boldsymbol{\theta}_k^{*(l)})^{-1}$$

  and

  $$\hat{D} = M^{-1} \sum_{l=1}^{M} \{p(\boldsymbol{\theta}_k^{(m)} | k) p(\boldsymbol{y} | k, \boldsymbol{\theta}_k^{(m)})\}^{-1}.$$

- If $\alpha = (p(\boldsymbol{\theta}_k|k) p(\boldsymbol{y}|k, \boldsymbol{\theta}_k) g(\boldsymbol{\theta}_k))^{-1/2}$ we have what is called the geometric estimator $\hat{p}_G$. This is given by

  $$\hat{N} = L^{-1} \sum_{l=1}^{L} \{p(\boldsymbol{\theta}_k^{*(l)} | k) p(\boldsymbol{y} | k, \boldsymbol{\theta}_k^{*(l)}) / g(\boldsymbol{\theta}_k^{*(l)})\}^{1/2}$$

and

$$\hat{D} = M^{-1} \sum_{l=1}^{M} \{ g(\boldsymbol{\theta}_k^{(m)}) / (p(\boldsymbol{\theta}_k^{(m)}|k) p(\boldsymbol{y}|k, \boldsymbol{\theta}_k^{(m)})) \}^{1/2}.$$

- The *optimal estimator* of Meng and Wong (1996), denoted by $\hat{p}_{opt}$, is obtained by an iterative procedure. Specify the initial value $r^{\cdot} = \hat{p}_G$ and, defining $s_1 = 1 - s_2 = M/(M+L)$, iterate the equation

$$r^{new} = A(r^{old}) / B(r^{old})$$

where

$$A(r) = \sum_{l=1}^{L} W_{2i} / (s_1 W_{2l} + s_2 r)$$

and

$$B(r) = \sum_{m=1}^{M} 1 / (s_1 W_{1m} + s_2 r)$$

where

$$W_{2l} = p(\boldsymbol{y}|k, \boldsymbol{\theta}_k^{*(l)}) p(\boldsymbol{\theta}_k^{*(l)}|k) / g(\boldsymbol{\theta}_k^{*(l)})$$

for $l = 1, \ldots, L$, and

$$W_{1m} = p(\boldsymbol{y}|k, \boldsymbol{\theta}_k^{(m)}) p(\boldsymbol{\theta}_k^{(m)}|k) / g(\boldsymbol{\theta}_k^{(m)})$$

for $m = 1, \ldots, M$.

3.5 Likelihood and information criteria

Traditional model selection criteria based on likelihood include variants of AIC Akaike (1987), the Schwartz or Bayesian criteria, or BIC, and related information criteria such as the ICOMP methods of Bozdogan and Ramirez (1987) and Bozdogan and Shigemasu (1998). Explicit equations for some of these criteria, that we use below

in comparative studies, are provided here. For each $k-$factor model, write $l_k = -2\log(p(\boldsymbol{y}|k,\hat{\boldsymbol{\theta}}_k))$ where $\hat{\boldsymbol{\theta}}_k$ is MLE of $\boldsymbol{\theta}_k = (\boldsymbol{\beta}_k, \boldsymbol{\Sigma})$ and the likelihood function is the standard form in equation (2.5). Write $\hat{\boldsymbol{\Omega}}_k = \hat{\boldsymbol{\beta}}_k\hat{\boldsymbol{\beta}}_k' + \hat{\boldsymbol{\Sigma}}$ for the corresponding MLE of $\boldsymbol{\Omega}_k$. It is easily deduced that

$$l_k = T\left\{m\log(2\pi) + \log|\hat{\boldsymbol{\Omega}}_k| + \text{trace}(\hat{\boldsymbol{\Omega}}_k^{-1}\boldsymbol{S})\right\}$$

where $\boldsymbol{S} = \boldsymbol{y}'\boldsymbol{y}/T$. The various model selection criteria are defined as follows:

- AIC$= l_k + 2p_k$
- BIC$= l_k + \log(T)p_k$
- BIC$^* = l_k + \log(\tilde{T})p_k$
- ICOMP$= l_k + C_1(\hat{\boldsymbol{\Sigma}}_k)$

where

$$\begin{aligned} p_k &= m(k+1) - k(k-1)/2, \\ \tilde{T} &= T - (2m+11)/6 - 2k/3, \\ C_1(\boldsymbol{\Sigma}_k) &= 2(k+1)\left((m/2)\log(\text{trace}\boldsymbol{\Sigma}_k/m) - 0.5\log|\boldsymbol{\Sigma}_k|\right) \end{aligned}$$

In the following two sections we study the empirical performance of the strategies presented in so far to compute normalizing constants along with the likelihood and information criteria and our proposed reservible jump strategy.

3.6 Simulation studies

Here we simulated two factor models with one and three common factors respectively. In the first simulation $m = 7$, which allows up to three factor to be fitted, while in the second simulation $m = 9$, or a maximum of five factors.

3.6.1 A first simulation study

An initial simulation study considers a one-factor model for a seven-dimensional problem generating one hundred observations; thus $m = 7, k = 1$ and $T = 100$. In each of a series of simulations, T observations were drawn from a one-factor models defined by parameters

$$\begin{aligned} \beta' &= (0.995, 0.975, 0.949, 0.922, 0.894, 0.866, 0.837), \\ diag(\Sigma) &= (0.01, 0.05, 0.10, 0.15, 0.20, 0.25, 0.30). \end{aligned}$$

Each such simulated data set was analysed using the MCMC and reversible jump methodologies, and also subject to study using the range of model selected criteria and methods described above. This study explored $k-$factor models for each data set, with up to three possible factors in each case.

MCMC analyses utilised the prior distributions based on the following hyper-parameter values: $m_0 = 0$ and $C_0 = 1$ define the prior distribution for $\boldsymbol{\beta}$, while σ_k^2, $\nu_{0i} = 2.2$ and $\nu_{0i}s_{0i}^2 = 0.1$ define the prior distribution for each σ_k^2 such that $E(\sigma_k^2) = 0.5$. The MCMC and reversible jump samplers were based on $M_0 = 10,000$ iterations as burn-in, followed by a further 10,000 iterates that were sampled every ten steps to produce a final MCMC sample of size 1,000. In generating proposals in the RJMCMC methods, we adopted $a = 18$, $b = 2$ and

$$J = \begin{pmatrix} 0.0 & 1.0 & 0.0 \\ 0.5 & 0.0 & 0.5 \\ 0.0 & 1.0 & 0.0 \end{pmatrix}.$$

Among the candidate methods for model selected, the "Newton and Raftery" technique requires the specification of a control parameter, δ; this was set at $\delta = 0.05$, and the number of iterations at 1,000.

Table 3.1 displays results from this simulation analysis. We repeated the model fitting exercises for 1,000 different data sets generated independently from the one-

	k		
Method	$k=1$	$k=2$	$k=3$
RJMCMC	1000	0	0
$\hat{p}_C$	954	46	0
$\hat{p}_H$	428	258	314
$\hat{p}_{NR}$	467	234	299
$\hat{p}_{GD}$	1000	0	0
$\hat{p}_{LM}$	1000	0	0
$\hat{p}_G$	1000	0	0
$\hat{p}_{opt}$	1000	0	0
Criterion	$k=1$	$k=2$	$k=3$
AIC	854	135	11
BIC	1000	0	0
BIC*	1000	0	0
ICOMP	607	296	97

Table 3.1: Comparison of model uncertainty assessment methods on simulated data set 1: $m = 7, k = 1$ and $T = 100$.

factor model. The table provides simple counts of the number of times that each k-factor model achieved the highest posterior probability. For example, the harmonic estimator method selected the one-factor model 428 times out of 1,000, and the three-factor model 314 times out of 1,000. Evidently, most of the approximation methods are very reliable in favouring the one-factor model, as is the RJMCMC (the "gold standard") approach. Bridge sampling methods agree with our RJMCMC approach. Relatively poor results are achieved by the harmonic mean method, Newton-Raftery estimator, AIC, ICOMP, and at some extent the candidate's estimator, which all tend to prefer higher numbers of factors a significant proportion of the time. In terms of model selection *per se*, as opposed to exploring model uncertainty more formally, the BIC methods are relatively accurate and, of course, rather easier to compute.

In analysis of a real data, we run into multi-modalities in posterior distributions that require some thought and explanation. In anticipation of this, we here ex-

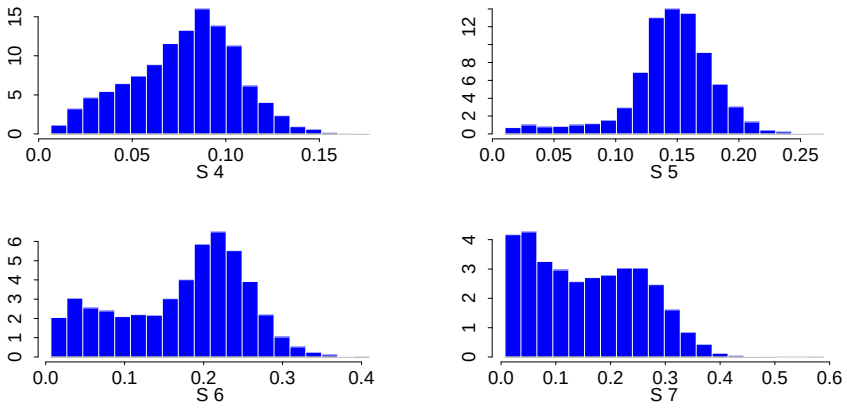


Figure 3.2: Marginal posteriors of the σ_i^2 $(i = 4, \ldots, 7)$ from analysis of the simulated data set from a one-factor structure but analysed using a model with $k = 3$ factors (See our first simulated study at Section 3.6.1).

plore some summaries of a three-factor model fitted to one of the simulated data sets arising from this true one-factor model. Figure 3.2 displays marginal posterior densities of some of the idiosyncratic variances from such an analysis. Note the multi-modality; marginal posteriors for elements of the $\boldsymbol{\beta}$ matrix exhibit corresponding multiple modes. This arises due to the mis-match between the model assumption of $k = 3$ and the data structure based on $k = 1$, and is driven by the identification issues arising in such cases, as discussed in Section 2.3 in the text around equation (2.8). Encountering such multi-modality in posteriors from a specified model can therefore be taken as a suggestion that the chosen value of k is too large.

3.6.2 A second simulation study

A second study follows the pattern of the above example, but now using data sets simulated from a model with $m = 9$ variables, $k = 3$ factors, and with a sample size

of just $T = 50$. The true model in this case has parameters

$$\beta' = \begin{pmatrix} 0.99 & 0.00 & 0.00 & 0.99 & 0.99 & 0.00 & 0.00 & 0.00 & 0.00 \\ 0.00 & 0.95 & 0.00 & 0.00 & 0.00 & 0.95 & 0.95 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.90 & 0.00 & 0.00 & 0.00 & 0.00 & 0.90 & 0.90 \end{pmatrix}$$

and

$$diag(\Sigma) = (0.02, 0.19, 0.36, 0.02, 0.02, 0.19, 0.19, 0.36, 0.36).$$

	k				
Method	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$
RJMCMC	7	0	993	0	0
$\hat{p}_C$	0	12	848	138	2
$\hat{p}_H$	0	0	650	228	122
$\hat{p}_{NR}$	0	0	615	258	127
$\hat{p}_{GD}$	0	0	998	2	0
$\hat{p}_{LM}$	0	1	999	0	0
$\hat{p}_G$	0	11	985	4	0
$\hat{p}_{opt}$	0	11	985	4	0
Criterion	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$
AIC	0	0	857	125	18
BIC	0	0	995	5	0
BIC*	0	0	993	7	0
ICOMP	0	0	886	97	17

Table 3.2: Comparison of model uncertainty assessment methods on simulated data set 2:$m = 9, k = 3$ and $T = 50$.

The analyses used the same hyper-parameters and MC sample size choices as in the first simulation study. As there, we summarise one aspect of performance of model selection methods by simply counting the number of times, out of a total of 1,000 analyses of simulated data sets, that each possible k-factor model received the highest posterior probability using each of the methods of computation. These summaries appear in Table 3.2. Again, it is clear that several of the approximation methods reliably identify the true model structure, which gives some indication of their likely

utility in real data analyses. Among the approximate Bayesian methods, those based on the candidate's estimtor, the harmonic mean estimator and the Newton-Raftery technique are the only failures, their performances being quite poor in comparison to the other Bayesian approaches and to the information criteria. Once again, AIC and ICOMP choose the wrong model in at least 12 percent of the simulations.

3.7 International exchange rate returns

We now explore the factor structure underlying the changes in monthly international exchanges rates using the data studied in West and Harrison (1997) (pp 610-618). These time series are the exchange rates in British pounds of the following $m = 6$ currencies: US dollar (US), Canadian dollar (CAN), Japanese yen (JAP), French franc (FRA), Italian lira (ITA), and the (West) German mark (GER). The data span the period from 1/1975 to 12/1986 inclusive, and the monthly changes in exchange rates appear in Figure 3.3. Each series has been standardised with respect to its sample mean and standard deviation over the period for comparability (this does not affect the modeling process and factor structure analysis). Earlier studies in West and Harrison (1997) used various principal component analyses that indicated up to three meaningful latent components, suggesting up to three factors may be relevance in our analyses.

Our illustrative analysis first explores uncertainty about the number of factors as in the foregoing simulated data analyses, and then investigates questions about the dependence of conclusions from such analyses on the chosen order of the series. This latter point is of interest as the particular factor model structure adopted – with the upper triangle of zero elements in the factor loading matrix – introduces an apparent order dependence.

Prior distributions are specified exactly as in the simulated examples, the general

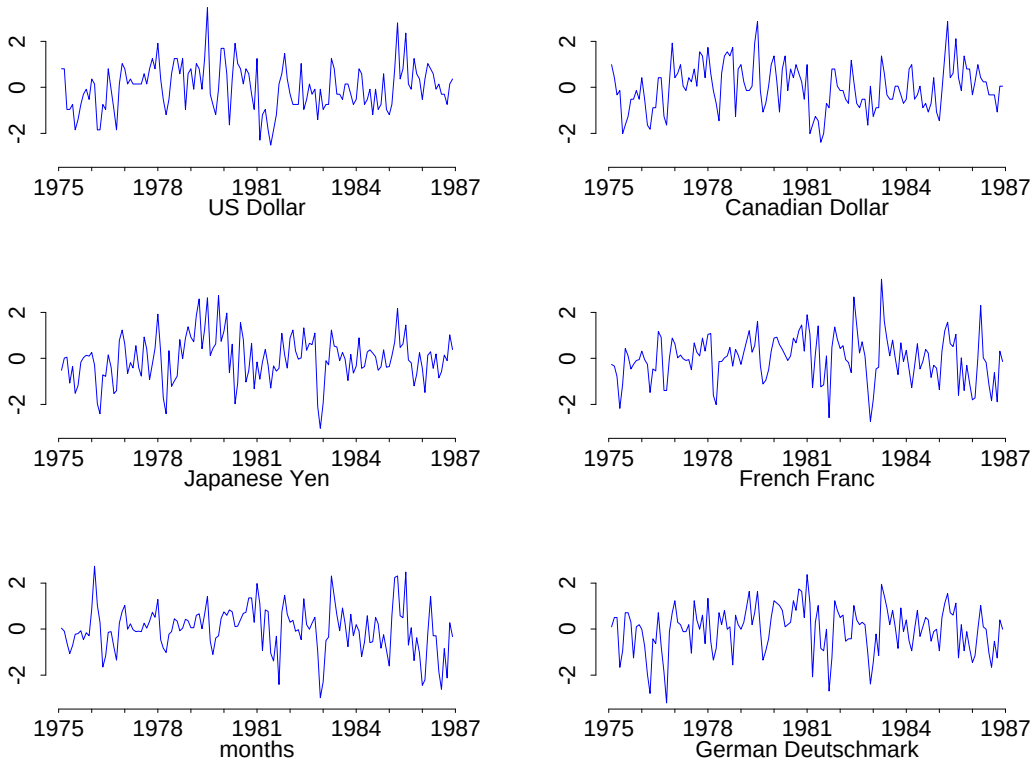


Figure 3.3: Standardized first differences of monthly observed exchange rates.

scales for all parameters now being comparable with those of the simulation examples as the data are modelled after standardisation. Specifically, we have $\mu_0 = 0$, $C_0 = 1$, $\nu_{0i} = 2.2$ and $\nu_{0i}s_{0i}^2 = 0.1$. For the Gibbs sampling and RJMCMC analyses we burn-in the algorithms for $10,000$ iterations, and then save equally spaced samples of $5,000$ draws from a longer run of $100,000$. Newton and Raftery (1994) suggest using δ small, so we decided to use $\delta = 0.05$. Alternative values were used and little or none variation was observed. Proposal distributions in the RJMCMC analysis are based on defining parameters $a = 18$ and $b = 2$, and the transition matrix is

$$J = \begin{pmatrix} 0.0 & 0.5 & 0.5 \\ 0.5 & 0.0 & 0.5 \\ 0.5 & 0.5 & 0.0 \end{pmatrix}.$$

The analyses were run on the data with currencies ordered by country as: US, CAN, JAP, FRA, ITA, GER. Table 3.3 provides summaries of the various approximate Bayesian and information criteria for assessment of the number of factors. The overall suggestions is that $k = 2$ is strongly favoured.

	$\log p(\boldsymbol{y}\|k)$			$p(k\|\boldsymbol{y})$		
Method	$k=1$	$k=2$	$k=3$	$k=1$	$k=2$	$k=3$
RJMCMC	—	—	—	0.00	**0.88**	0.12
$\hat{p}_C$	-1013.5	-935.3	**-925.5**	0.00	0.00	**1.00**
$\hat{p}_H$	-988.0	**-871.0**	-871.8	0.00	**0.71**	0.29
$\hat{p}_{NR}$	-991.9	**-880.1**	-881.4	0.00	**0.78**	**0.22**
$\hat{p}_{GD}$	-1017.7	-907.1	**-906.4**	0.00	0.34	**0.66**
$\hat{p}_{LM}$	-1014.8	-904.5	**-897.3**	0.00	0.00	1.00
$\hat{p}_G$	-1014.5	-903.7	-Inf	0.00	**1.00**	0.00
$\hat{p}_{opt}$	-1014.5	-903.7	-Inf	0.00	**1.00**	0.00

Criterion	$k=1$	$k=2$	$k=3$
p-value	0.00	**1.00**	1.00
AIC	1978.4	**1745.0**	1751.0
BIC	2013.9	**1795.4**	1813.2
BIC*	2013.6	**1794.8**	1812.3
ICOMP	1957.9	1776.1	**1724.0**

Table 3.3: Comparison of model uncertainty assessment methods from analyses of the international exchange rate time series.

From the MCMC analysis of the $k = 2$ factor model, we have the following posterior summaries:

- The posterior means of $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ parameters are, to two decimal places,

$$E(\boldsymbol{\beta}|\boldsymbol{y}) = \begin{pmatrix} 0.99 & 0 \\ 0.95 & 0.05 \\ 0.46 & 0.42 \\ 0.39 & 0.91 \\ 0.41 & 0.77 \\ 0.40 & 0.77 \end{pmatrix} \quad \text{and} \quad E(diag(\boldsymbol{\Sigma})|\boldsymbol{y}) = \begin{pmatrix} 0.05 \\ 0.13 \\ 0.62 \\ 0.04 \\ 0.25 \\ 0.28 \end{pmatrix}.$$

- The marginal posterior densities of the elements of $\boldsymbol{\beta}$ are displayed – in terms of histograms of the posterior samples – in Figure 3.4.

- The marginal posterior densities of the σ_i^2 parameters are displayed – also in terms of histograms of the posterior samples – in Figure 3.5.

- The time trajectories of the posterior means of the two factor time series are displayed in Figure 3.6. The first factor is plotted together with the US and CAN series, and the second factor is plotted together with the JAP and European currencies.

- For each currency series $i = 1, \ldots, 6$, the percentage of the conditional variance explained by each factor $j = 1, 2$ is simply $100(1 + \beta_{ij}^2/\sigma_i^2)$. Table 3.4 below provides estimated values of these quantities with the $\beta_\cdot$ and $\sigma_\cdot$ parameters estimated at their posterior means.

Country	Factor 1	Factor 2
US	95.1	0
CAN	87.6	0.2
JAP	20.5	17.6
FRA	14.7	81.8
ITA	16.4	58.6
GER	16.1	58.5

Table 3.4: Percentage of the variance of each series explained by each factor in analysis of the international exchange rate time series.

These summaries indicate the following broad conclusions. The first factor represents the value of sterling relative to a basket of currencies in which the North American currencies are dominant. US and CAN are roughly equally weighted, which is expected as CAN rates are heavily determined in international markets by US rates. This first factor may be termed the North American factor. The second factor may be

similarly termed the European Union (EU) factor. It represents a restricted basket of currencies dominated by the EU currencies, with a relatively reduced weighting on JAP. US is absent from this factor, by design of the factor loading matrix, and CAN is practically absent, with the posterior for $\beta_{2,2}$ indicating very small values. Inferences about the idiosyncratic variances strengthen and extend these conclusions. Those for US and GER are very small, indicating that these two currencies play determining roles in defining their sector factor. CAN, FRA and ITA have larger idiosyncratic variances, indicative of their departures from their sector factors. JAP has a large idiosyncratic variance, contributing about two-thirds of the overall conditional variance. A $k = 3$ factor model would move most of this variability over to the third, JAP factor, as further studies verify.

Notice that the marginal posteriors graphed are all unimodal. This is of interest in view of the earlier discussion about multiple posterior modes induced by multiple local maxima in the likelihood functions when the specified value of k is larger that is consistent with the data. Multiple modes appear in analysis of a 3-factor model, as shown in Figures 3.7 and 3.8. The margins from the 2-factor model analysis are consistent with the view that $k = 2$ is not too large, and therefore provide additional support for the 2-factor model.

It is of interest to explore possible dependence on the order of the series in this analysis. This is especially indicated here due to the high dependence between US and CAN and the resulting very small values of $\beta_{2,2}$, the diagonal factor loading element that is constrained to be positive. The above analysis was therefore re-run with the orders of CAN and JAP interchanged. The resulting posterior means of the

factor loadings and idiosyncratic variances are

$$E(\boldsymbol{\beta}|\boldsymbol{y}) = \begin{pmatrix} 0.98 & 0.00 \\ 0.45 & 0.42 \\ 0.95 & 0.03 \\ 0.39 & 0.91 \\ 0.41 & 0.77 \\ 0.41 & 0.77 \end{pmatrix} \quad \text{and} \quad E(diag(\boldsymbol{\Sigma})|\boldsymbol{y}) = \begin{pmatrix} 0.06 \\ 0.62 \\ 0.12 \\ 0.04 \\ 0.25 \\ 0.26 \end{pmatrix},$$

where the figures in bold font simply indicate JAP and CAN in the new order. Comparing with the original analysis we see that these numbers are in extremely close agreement, suggesting the order has essentially no effect on the estimation. To further explore this, Table 3.5 provides the summaries of model assessment methods based on this ordering of variables. There are differences in the numbers presented, though the overall conclusion of support for $k = 2$ factors stands.

	$\log p(\boldsymbol{y}\|k)$			$p(k\|\boldsymbol{y})$		
Method	$k=1$	$k=2$	$k=3$	$k=1$	$k=2$	$k=3$
RJMCMC	—	—	—	0.00	**0.98**	0.02
$\hat{p}_C$	-1013.5	**-934.5**	-985.8	0.00	**1.00**	0.00
$\hat{p}_H$	-988.3	-874.6	**-873.0**	0.00	0.16	**0.84**
$\hat{p}_{NR}$	-985.5	**-867.3**	-867.9	0.00	**0.65**	0.35
$\hat{p}_{GD}$	-1017.8	**-907.0**	-909.6	0.00	**0.93**	0.07
$\hat{p}_{LM}$	-1015.5	**-904.4**	-910.3	0.00	**1.00**	0.00
$\hat{p}_G$	-1014.5	**-903.5**	-Inf	0.00	**1.00**	0.00
$\hat{p}_{opt}$	-1014.5	**-903.5**	-Inf	0.00	**1.00**	0.00

Table 3.5: Comparison of model uncertainty assessment methods from the reanalyses of the international exchange rate time series under a different order of the currencies.

3.8 Summary

As discussed in the introduction of this chapter, our interest has been to study MCMC methods for factor models and novel RJMCMC and other methods for assessing the

issue of model uncertainty induced by an unknown number of factors. In doing so, we have explored empirical studies with two simulated and one real data example, highlighting ranges of methodological and modelling issues. A few additional comments are of interest in conclusion of the chapter.

- In the Newton and Raftery's estimator, the quantity δ, that defines the proportion of draws coming from the prior was studied in the context of the real data application. Table 3.6 shows that for δ between 0.01 and 0.10 the right model is always selected and convergence, of the iterative scheme, is achieved around 10 iterations. Another finding is that starting values are not relevant at all.

δ	$k=1$	$k=2$	$k=3$
0.00	-987.8	-871.5	-872.2
0.01	-986.6	-868.8	-869.6
0.02	-986.2	-868.2	-868.9
0.03	-985.9	-867.8	-868.5
0.04	-985.7	-867.5	-868.2
0.05	-985.5	-867.3	-867.9
0.06	-985.4	-867.1	-867.7
0.07	-985.2	-866.9	-867.5
0.08	-985.1	-866.8	-867.4
0.09	-985.0	-866.6	-867.2
0.10	-984.9	-866.5	-867.1

Table 3.6: Analysing the sensitiveness of Newton and Raftery's estimator in factor analysis, to the choice of δ in the international exchange rate time series under the second ordering of the currencies.

- Our customised RJMCMC method, using empirical proposal distributions based on parallel MCMC analyses for a range of models with specified numbers of factors, is effective and efficient, as tested in a range of simulation studies (beyond those reported here). Development of effective proposal distributions and jumping rules in models with even moderate dimensional parameters is usually very

challenging, and the approach used here will be useful in other such models. We note that the computation of approximate posterior probabilities on k using bridge sampling approaches, though also requiring some tuning, is similarly accurate in a range of studies, providing answers close to those from RJMCMC; these methods deserve further study too. By contrast, we have found that none of the other "standard" methods of approximation reviewed in Section 3.4 is consistently accurate in identifying correct models in ranges of simulation studies, and none consistently superior to the formal RJMCMC, Newton and Raftery and bridge sampling approaches. On purely empirical grounds, we find the BIC methods – though lacking in formal interpretation – generally provide more stable and reliable initial guides to the choice of k than the other standard methods.

- The selected order of data variables influences the likelihood function and hence posterior inferences about the number of factors. The effect can be marked, although inferences about the factor loadings and other parameters are generally relatively stable by comparison. Thus, very naturally, the order of variables is relevant in connection with the interpretation and the number of common factors. Of course, the conditional variance-covariance matrix of the variables is unaffected by the ordering, and hence so are the predictive inferences resulting from the model.

- In further empirical studies we have explored a range of predictive exercises. Two situations are illustrated here:

  - For example, in studies of simulated data and of the industrial stock indices analysed in Geweke and Zhou (1996) we have explored out-of-sample predictions based on sequential data analysis and one-step ahead predic-

tions, as are standard in time series work. These studies have indicated a general point; rather often, posterior uncertainty about the number of factors based on a set of historical data may understate the practically relevant uncertainties when forecasting ahead. For example, sequential forecasting performance using a four-factor model can often out-perform a three-factor model even when analysis of past data has given almost no posterior probability to the four-factor model. Table 3.7 and Figures 3.10 and 3.11 present the results we found when exploring Geweke and Zhou's industry portfolio groups (Figure 3.9).

	$p(k\|\boldsymbol{y})$						
Method	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$
RJMCMC	0.00	0.00	**0.99**	0.01	0.00	0.00	0.00
$\hat{p}_C$	**1.00**	0.00	0.00	0.00	0.00	0.00	0.00
$\hat{p}_H$	0.00	0.00	0.00	0.00	**0.79**	0.07	0.14
$\hat{p}_{GD}$	0.00	0.00	**1.00**	0.00	0.0	0.00	0.00
$\hat{p}_{LM}$	0.00	0.00	**0.84**	0.16	0.00	0.00	0.00
$\hat{p}_G$	0.00	0.00	**1.00**	0.00	0.00	0.00	0.00
$\hat{p}_{opt}$	0.00	0.00	**1.00**	0.00	0.00	0.00	0.00
Criterion	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$
AIC	2467.9	2353.9	2302.4	2288.2	**2271.3**	2278.5	2285.7
BIC	2534.8	2451.5	**2427.8**	2438.7	2444.1	2470.8	2494.8
BIC*	2533.5	2449.3	**2424.8**	2434.8	2439.1	2464.9	2487.9
ICOMP	2424.7	2319.4	2317.3	2289.5	**2246.5**	2347.6	2410.6

Table 3.7: Posterior model probabilities and information criteria for the $m = 12$ industry portfolio groups from Only a stretch of 12 years of data is used (from JAN/1984 to DEC/1995, or $T = 144$ months. See Figure 3.9).

– In the previous example, based on real data, we have seen that as far as predicting into the future is concerned a higher order factor model presented higher posterior predictive performance. We have applied the same idea to various simulation studies. Here we present one of them. A simulation with a one-factor model over $T = 200$ observations and

$m = 7$ variables. We fit factor models (k=1,2 and 3) for the first 180 observations and forecast into the future up to the 200th observation. This simulation data we will use later on on Section 4.4.1. Table 3.8 shows h-step ahead forecasts based on the k-factor model, for $k = 1, 2$ and 3. As it will be learned in Section 4.4.1 the one factor model has the highest posterior model probability. However, it can be seen that a 2-factor model is more appropriate for forecast purposes as opposed to the one-factor model, as also shown in the industry portfolio example. Even thought this result emphasizes our previous empirical findings, we believe that predictive comparisons and model mixing are worthy of further study.

		log posterior predictive		
h	k^*	k=1	k=2	k=3
1	1	-3.91	-4.17	-4.11
2	3	-12.29	-11.83	-11.72
3	2	-22.74	-21.87	-22.06
4	2	-26.26	-25.39	-25.60
5	2	-30.99	-30.03	-30.38
6	2	-36.95	-35.95	-36.06
7	2	-41.36	-40.49	-40.63
8	2	-43.65	-42.80	-42.97
9	2	-47.14	-46.25	-46.52
10	2	-53.32	-52.55	-52.87
15	2	-71.23	-70.92	-71.02
20	2	-87.75	-87.59	-87.75

Table 3.8: Posterior predictive analysis of the one-factor model. The models forecast performances were measured from time $T + 1$ to $T + h$, for $T = 180$ and $h = 1, 2, \ldots, 20$. The second column indicates the order of the factor model with highest posterior predictive value.

Beyond these issues, we note that related developments in dynamic factor modelling Aguilar and West (2000), in financial time series and portfolio studies are focussed almost exclusively on short-term forecasting and the potential improvements

available in forecasting moderate to high-dimensional time series using factor structures. Here the assessment of the number of factors is also a live issue, and one that is complicated by the time-varying nature of such models that leads to the notion of time-variation in the number of (practically relevant) factors. Some of those points are extended in Chapters 5 and 6.

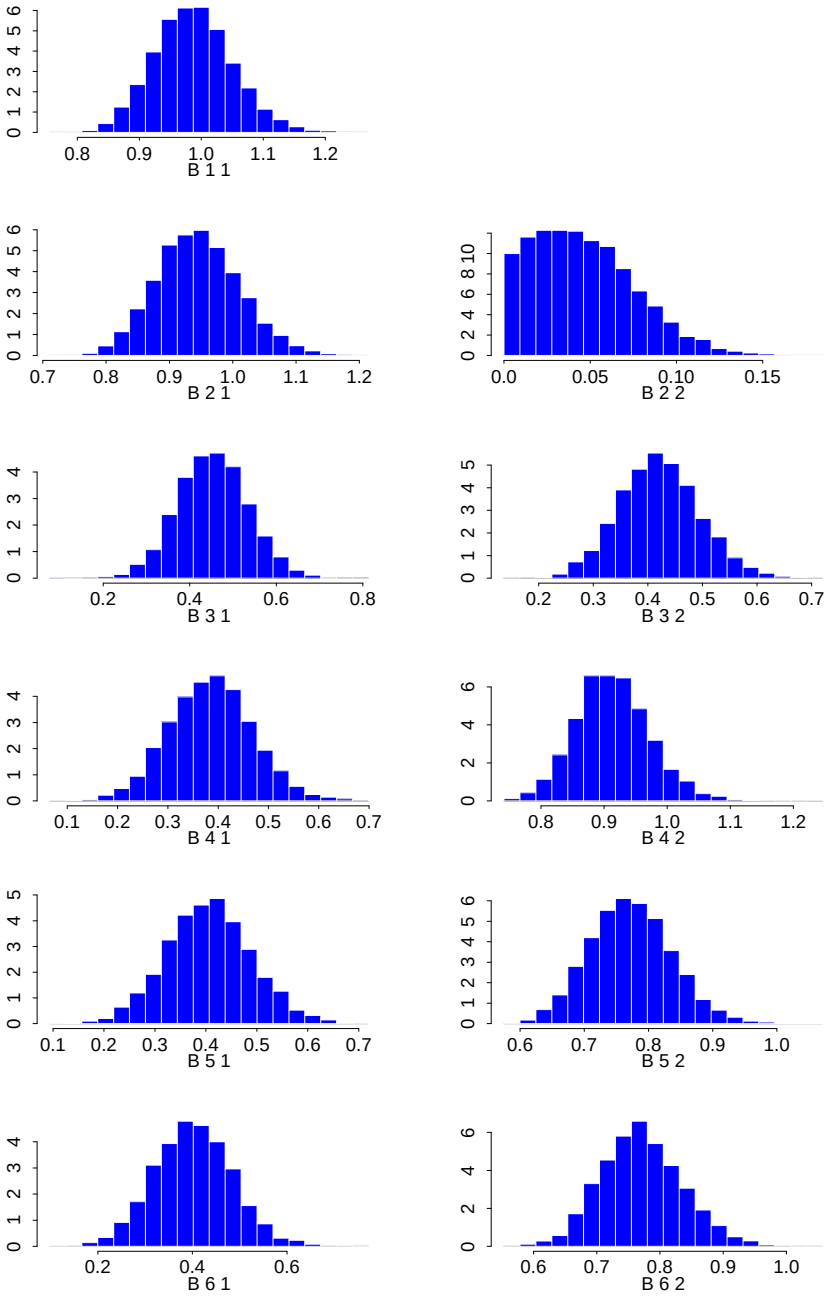


Figure 3.4: Marginal posteriors of the factor loadings when fitting a two-factor structure to the international exchange rates.

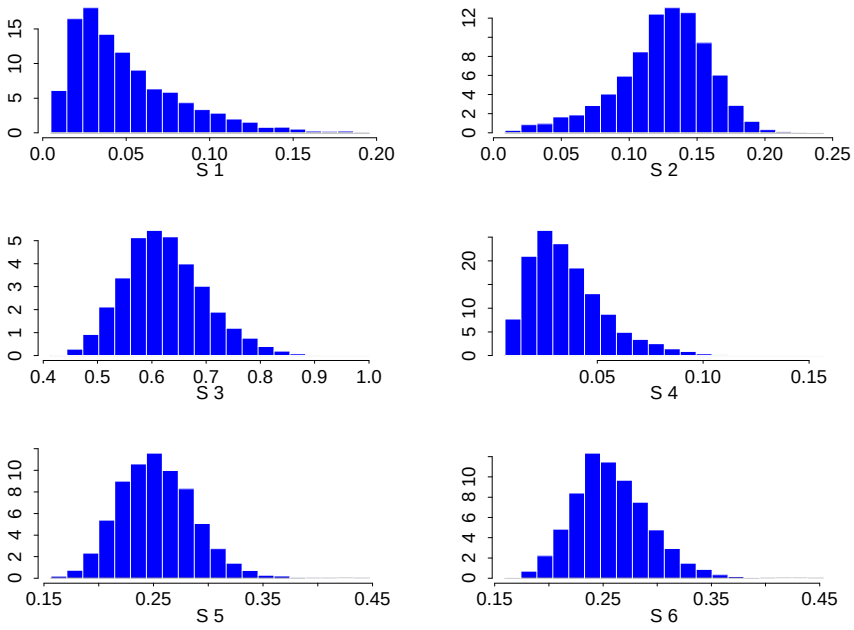


Figure 3.5: Marginal posteriors of the idiosyncratic variances when fitting a two-factor structure to the international exchange rates.

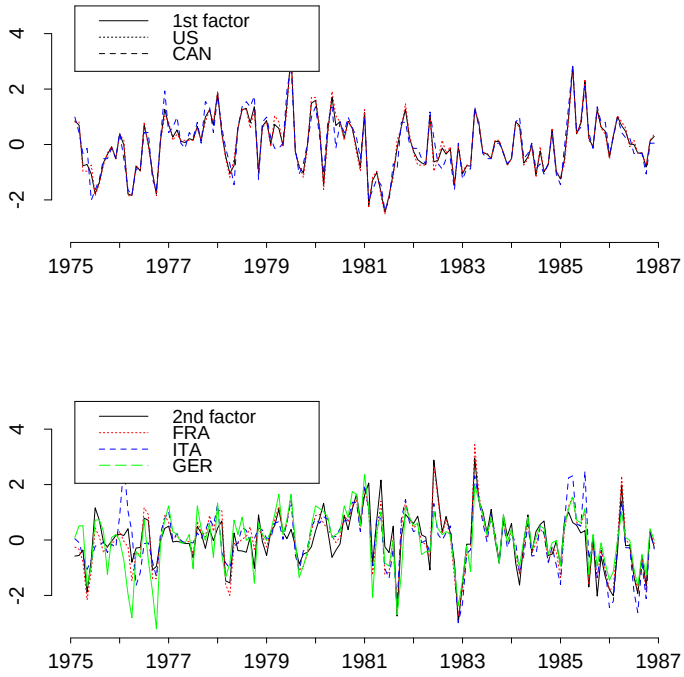


Figure 3.6: Posterior mean for the factor and actual exchange rates. First factor plus US and CAN (*upper frame*) and second factor plus FRA,ITA and GER (*lower frame*).

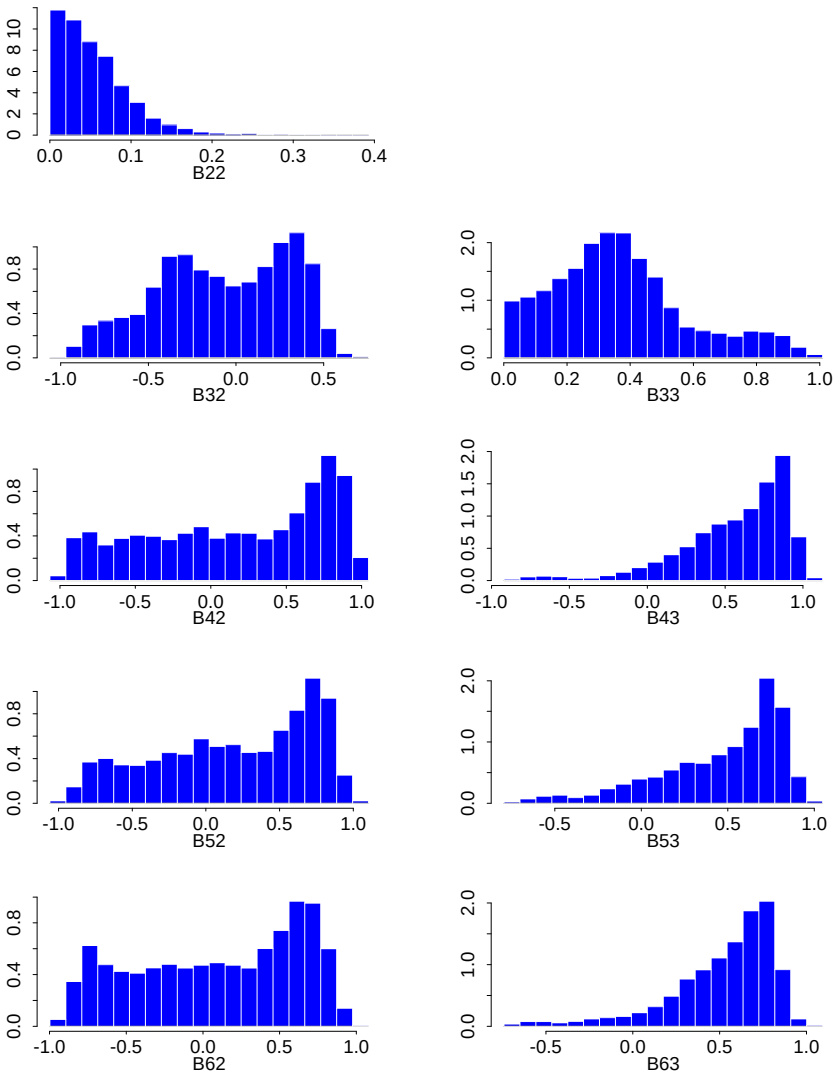


Figure 3.7: Marginal posteriors of the last two columns of the factor loadings matrix when fitting a three-factor structure to the international exchange rates. The first column of the factor loadings matrix (omitted here) is quite similar to its counterpart shown in figure 3.4.

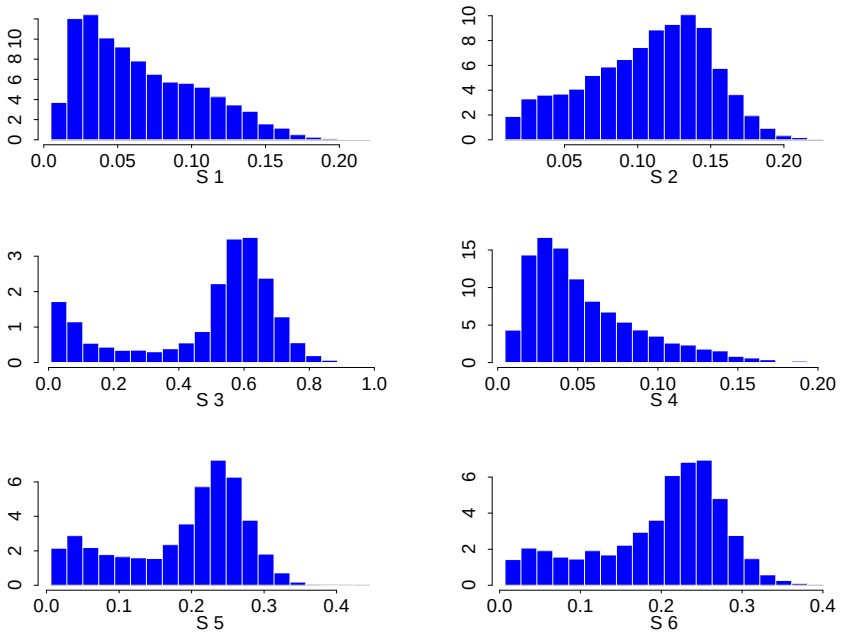


Figure 3.8: Marginal posteriors of the idiosyncratic variances when fitting a three-factor structure to the international exchange rates.

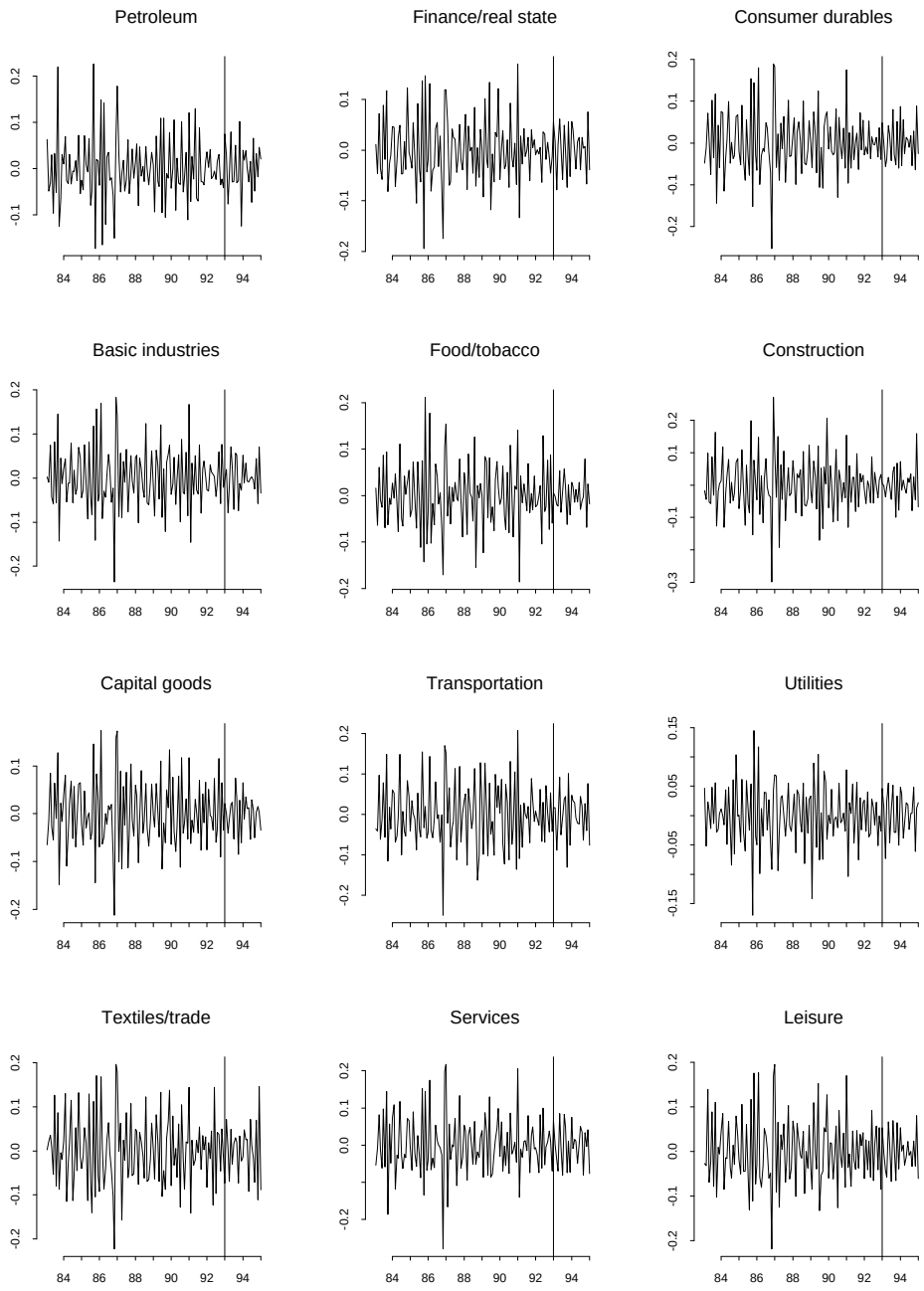


Figure 3.9: Monthly observed returns on industry portfolios from JAN/1984 to DEC/1995. The last two years of data were reserved for forecasting performance comparisons.

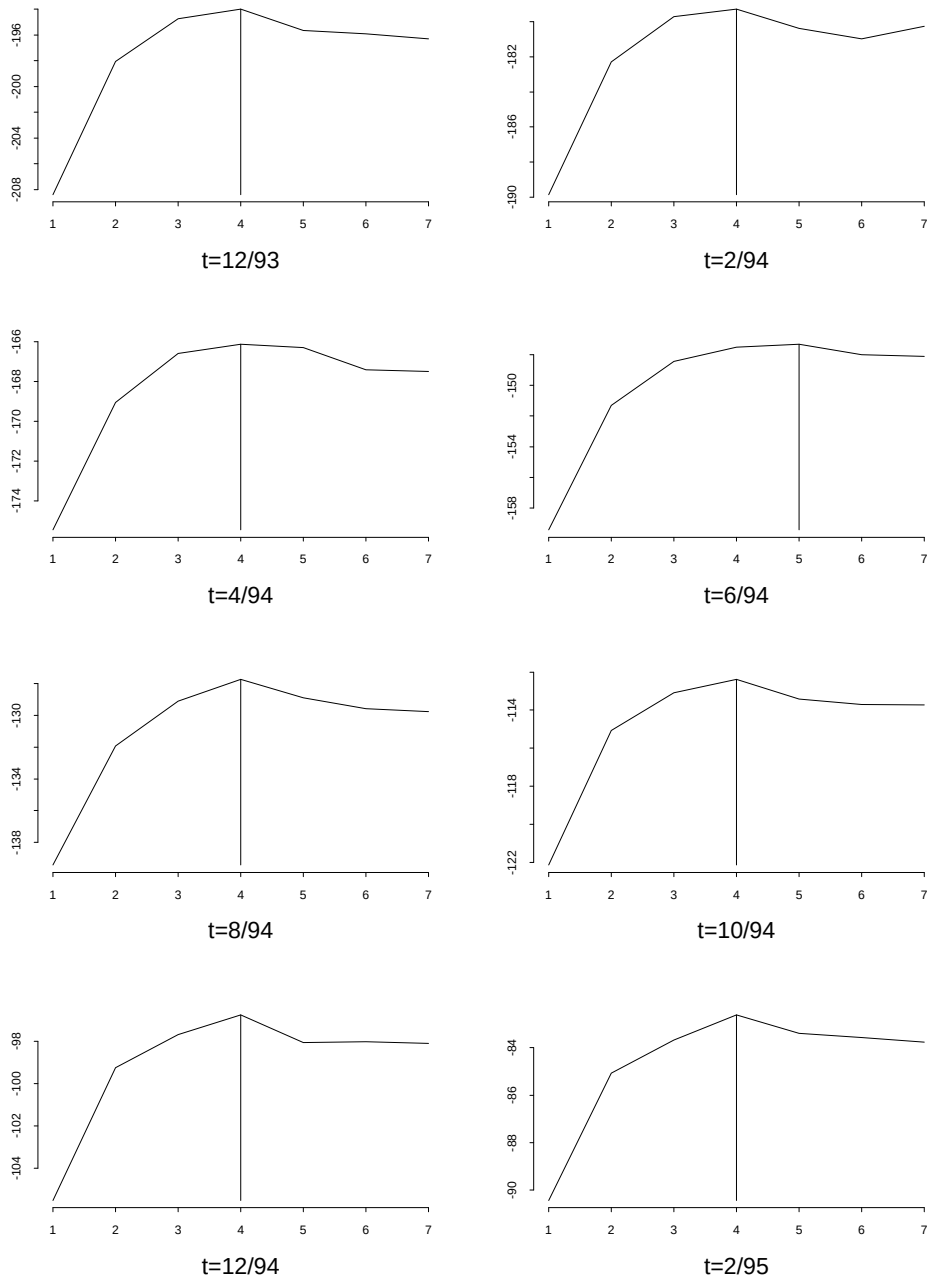


Figure 3.10: Factor analysis with k common factors ($k = 1, \ldots, 7$) were fitted to the dataset from JAN/84 to t and their forecast performances measured from time $t + 1$ to $T = 144$ (DEZ/95).

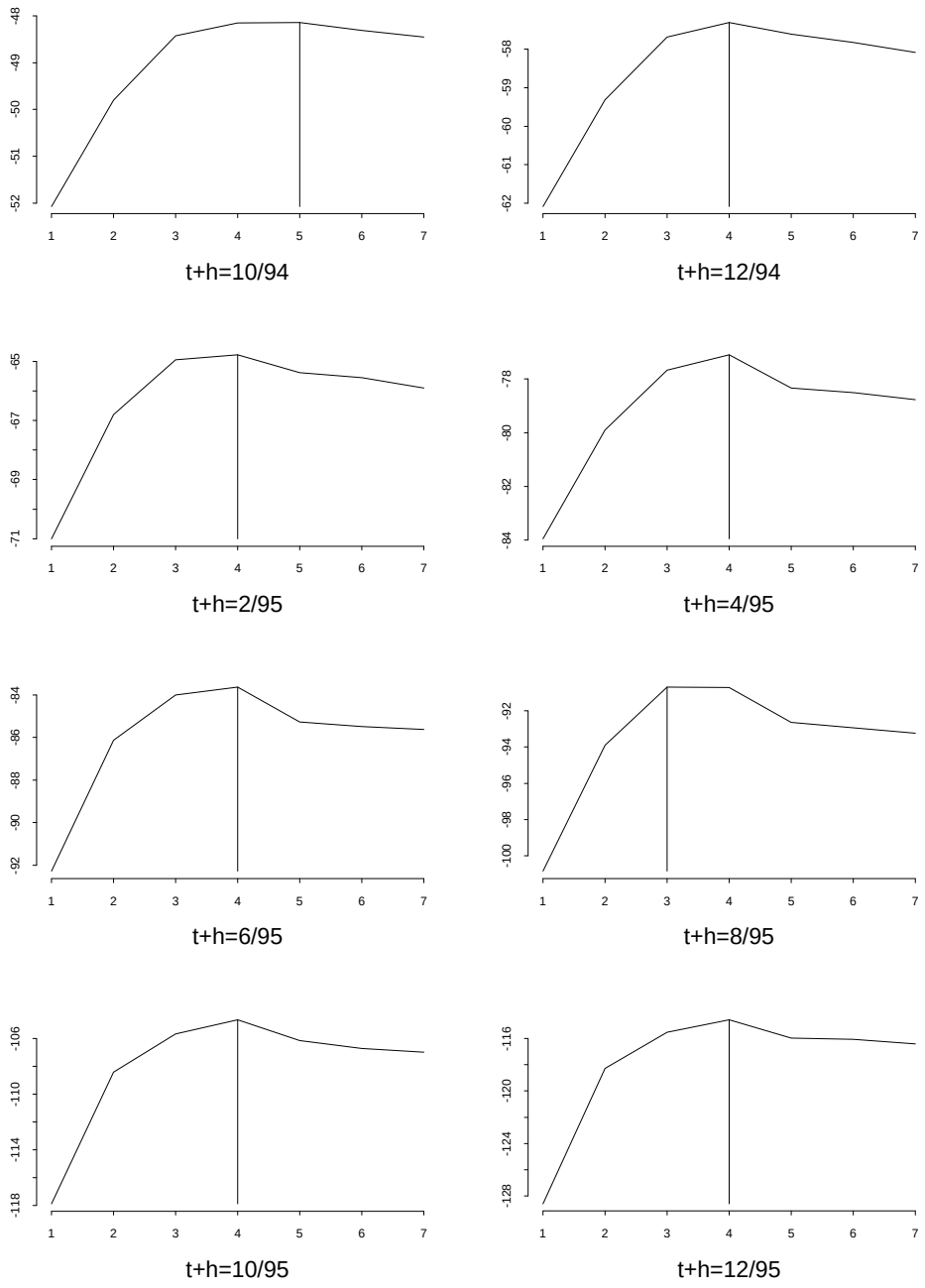


Figure 3.11: Factor analysis with k common factors ($k = 1, \ldots, 7$) were fitted to the dataset from JAN/84 to DEZ/93 ($t = 120$) and their forecast performances measured from time $t + 1$ to $t + h$, $h = 10, 12, \ldots, 24$.

Chapter 4

Expected posterior priors in factor analysis

4.1 Introduction

In this Chapter we explore default analysis of factor models using *expected posterior priors* (Pérez and Berger, 2000). Our main motivation is the difficulty in assessing prior information in factor models, both because it involves multivariate prior assessment and because of the highly nonlinear structure in a factor model.

More specifically, we want to study the sensitivity of the predictive distribution in factor models with respect to the values of the hyperparameters C_0, ν and νs^2 introduced in Section 2.5, or more generally, sensitivity with respect to the prior distribution chosen for the loadings, $\boldsymbol{\beta}$, and for the idiosyncratic variances, $\boldsymbol{\Sigma}$.

As far as estimation and posterior analysis is concerned, the effect of the prior decreases as the sample size increases (this can be found in any Bayesian textbook, such as Migon and Gamerman, 1999). Also, even in the case where improper priors are used posterior distribution are often proper. However, when it comes to model comparison, the importance of the prior does not decreases with sample size. Particularly, when priors are improper, Bayes' factor are arbitrarily as large or as small

as the decision makers might please. On the other hand, there are many situations where the investigator wants to use vague prior information, at least as an initial "benchmark" analysis.

In factor analysis there might be situations where a psychologist wants, for instance in a exploratory factor analysis, to identify possible traits that his patients share. In order to do so, we might want to consider factor models with different number of factors to base his conclusions. If the researcher has little prior information, she will find it hard to assess prior distributions and might be willing to use default methods, especially when the procedure is used routinely.

Many proposals have been made for such "objective" Bayesian model selection. Among others, Spiegelhalter and Smith's (1982) imaginary samples, Berger and Pericchi's (1996) intrinsic Bayes factors and O'Hagan's (1995) fractional Bayes factor, are just some examples of how important (partial) solutions for this problem might be.

The rest of this chapter is organized as follows. Section 4.2 reviews Pérez and Berger's expected posterior priors, while in Section 4.3 an algorithm is proposed to compute Bayes' factors in the factor analysis context by using MCMC algorithms for each entertained model. Simulated datasets are analysed in Section 4.4. We revisit the international exchange rate returns introduced in Section 3.7. Section 4.6 summarizes our empirical findings.

4.2 Expected posterior prior

In a recent article, Pérez and Berger (2000) propose *expected posterior priors* (or, *EP priors*). In the factor model, an expected posterior prior for $\boldsymbol{\theta}_k$ under m^* can be defined as

$$\pi_k^*(\boldsymbol{\theta}_k) = \int \pi_k^N(\boldsymbol{\theta}_k|\boldsymbol{y}^*)m^*(\boldsymbol{y}^*)d\boldsymbol{y}^* \tag{4.1}$$

where $\boldsymbol{y}^*$ is a training sample and m^* is a probability measure on the training sample space. Here $\pi_k^N(\boldsymbol{\theta}_k|\boldsymbol{y}^*)$ is the posterior density for $\boldsymbol{\theta}_k$ based on a specific "vague" or noninformative (usually improper) prior and under model $\mathcal{M}_k$. In our notation, $\boldsymbol{\theta}_k$ corresponds to the loading matrix and the uniquenesses variances in a k-factor model, as earlier.

The Bayes factor of $\mathcal{M}_k$ to $\mathcal{M}_{k'}$ is given by

$$\begin{aligned} B^*_{kk'}(\boldsymbol{y}) &= \frac{m_{\pi^*_k}(\boldsymbol{y})}{m_{\pi^*_{k'}}(\boldsymbol{y})} \\ &= \frac{\int p_k(\boldsymbol{y}|\boldsymbol{\theta}_k)\pi^*_k(\boldsymbol{\theta}_k)d\boldsymbol{\theta}_k}{\int p_{k'}(\boldsymbol{y}|\boldsymbol{\theta}_{k'})\pi^*_{k'}(\boldsymbol{\theta}_{k'})d\boldsymbol{\theta}_{k'}} \\ &= \frac{\int m^N_k(\boldsymbol{y}|\boldsymbol{y}^*)m^*(\boldsymbol{y}^*)d\boldsymbol{y}^*}{\int m^N_{k'}(\boldsymbol{y}|\boldsymbol{y}^*)m^*(\boldsymbol{y}^*)d\boldsymbol{y}^*} \end{aligned} \tag{4.2}$$

where $m^N_k(\boldsymbol{y}|\boldsymbol{y}^*) = m^N_k(\boldsymbol{y},\boldsymbol{y}^*)/m^N_k(\boldsymbol{y}^*)$.

Pérez and Berger (2000) list some properties of the EP priors, such as: (*i*) Only one m^* is used across competing models; (*ii*) Impropriety is immaterial, since any multiplicative constant factor in the priors cancel out; (*iii*) Coherent Bayes' factors are induced, i.e. $B^*_{kk'} = 1/B^*_{k'k}$, and $B^*_{kk'} = B^*_{kl}B^*_{lk'}$ (*iv*) The EP prior approach is, in certain cases, essentially equivalent to Berger and Pericchi's (1996) intrinsic Bayes Factor approach. (*v*) EP priors are easily embedded in a MCMC methodology.

It is this last property that will be given more attention in the next section. We present an approximation to Bayes' factors using EP priors in the factor analysis context. More detailed information about and applications of expected posterior priors can be found in Pérez (1998).

4.3 Computational issues

In order to compute the Bayes factor in (4.2) for the factor model at hand, we need first to obtain an approximation for $m_{\pi_k^*}$. But, firstly, let us rewrite $m_{\pi_k^*}$ as a function of m_k^N and m^*. Notice that

$$\begin{aligned} m_{\pi_k^*}(\boldsymbol{y}) &= \int p_k(\boldsymbol{y}|\boldsymbol{\theta}_k)\pi_k^*(\boldsymbol{\theta}_k)d\boldsymbol{\theta}_k \\ &= \int p_k(\boldsymbol{y}|\boldsymbol{\theta}_k)\left[\int \pi_k^N(\boldsymbol{\theta}_k|\boldsymbol{y}^*)m^*(\boldsymbol{y}^*)d\boldsymbol{y}^*\right]d\boldsymbol{\theta}_k \\ &= \underbrace{\int\left[\int \frac{p_k(\boldsymbol{y},\boldsymbol{y}^*|\boldsymbol{\theta}_k)\pi_k^N(\boldsymbol{\theta}_k)}{m_k^N(\boldsymbol{y}^*)}d\boldsymbol{\theta}_k\right]}_{m_k^N(\boldsymbol{y},\boldsymbol{y}^*)/m_k^N(\boldsymbol{y}^*)} m^*(\boldsymbol{y}^*)d\boldsymbol{y}^* \\ &= m_k^N(\boldsymbol{y})\int \frac{m_k^N(\boldsymbol{y}^*|\boldsymbol{y})}{m_k^N(\boldsymbol{y}^*)}m^*(\boldsymbol{y}^*)d\boldsymbol{y}^* \end{aligned} \tag{4.3}$$

and, accordingly, the Bayes factor $B_{kk'}^*(\boldsymbol{y})$ can be written as

$$B_{kk'}^*(\boldsymbol{y}) = \frac{m_k^N(\boldsymbol{y})}{m_{k'}^N(\boldsymbol{y})}\frac{\int \frac{m_k^N(\boldsymbol{y}^*|\boldsymbol{y})}{m_k^N(\boldsymbol{y}^*)}m^*(\boldsymbol{y}^*)d\boldsymbol{y}^*}{\int \frac{m_{k'}^N(\boldsymbol{y}^*|\boldsymbol{y})}{m_{k'}^N(\boldsymbol{y}^*)}m^*(\boldsymbol{y}^*)d\boldsymbol{y}^*} \tag{4.4}$$

Expression (4.3) gives us a hint of how to obtain an MCMC approximation for $m_{\pi_k^*}(\boldsymbol{y})$. Typically, $\boldsymbol{y}^*$ will be much smaller than $\boldsymbol{y}$, because it will be based on a minimal training sample (Pérez and Berger, 2000).

Initially, let $\boldsymbol{y}_1^*, \ldots, \boldsymbol{y}_R^*$ be random draws taken from $m^*(\boldsymbol{y}^*)$. An approximation for the integral in (4.3) is

$$\hat{I} = \frac{1}{R}\sum_{r=1}^{R}\frac{m_k^N(\boldsymbol{y}_r^*|\boldsymbol{y})}{m_k^N(\boldsymbol{y}_r^*)} \tag{4.5}$$

and according to well established MC results (Geweke, 1989), it can be shown that

$$\hat{I} \to \int \frac{m_k^N(\boldsymbol{y}^*|\boldsymbol{y})}{m_k^N(\boldsymbol{y}^*)}m^*(\boldsymbol{y}^*)d\boldsymbol{y}^*$$

as R goes to infinity. Therefore,

$$\hat{m}_{\pi_k^*}(\boldsymbol{y}) = \frac{m_k^N(\boldsymbol{y})}{R} \sum_{r=1}^{R} \frac{m_k^N(\boldsymbol{y}_r^*|\boldsymbol{y})}{m_k^N(\boldsymbol{y}_r^*)}$$

approximates $m_{\pi_k^*}(\boldsymbol{y})$ in the same sense. The question that remains now is how such predictive measures can be computed. Rephrasing it, the question is how normalizing constants can be computed. Section 3.4 described alternative ways of computing this quantity.

In order to compute $\hat{m}_{\pi_k^*}(\boldsymbol{y})$, we need first to compute $m_k^N(\boldsymbol{y})$, $m_k^N(\boldsymbol{y}_r^*|\boldsymbol{y})$ and $m_k^N(\boldsymbol{y}_r^*)$ for $r = 1, \ldots, R$. Firstly, it is easy to see that

$$m_k^N(\boldsymbol{y}^*|\boldsymbol{y}) = \int p(\boldsymbol{y}^*|\boldsymbol{\theta}_k)\pi_k^N(\boldsymbol{\theta}_k|\boldsymbol{y})d\boldsymbol{\theta}_k \tag{4.6}$$

where, having at hand draws from the posterior distribution under model $\mathcal{M}_k$, say $\boldsymbol{\theta}_k^{(1)}, \ldots, \boldsymbol{\theta}_k^{(M)}$, this integral can be approximated by

$$\hat{m}_k^N(\boldsymbol{y}^*|\boldsymbol{y}) = \frac{1}{M} \sum_{m=1}^{M} p(\boldsymbol{y}^*|\boldsymbol{\theta}_k^{(m)}) \tag{4.7}$$

which has the same asymptotic properties as $\hat{I}$. On the other hand, to compute $m_k^N(\boldsymbol{y})$ (or $m_k^N(\boldsymbol{y}^*)$) we need to rely on other kinds of approximations. However, now we need an estimator that (*i*) is minimally reliable, (*ii*) is based on a sample from the posterior distribution, and (*iii*) is fast enough to be used many times.

In the last Chapter, when comparing alternative methods of computing predictives, we have seen that the Laplace-Metropolis estimator was considerably reliable and easy to implement, therefore a natural candidate to compute predictives in an automatic way. Recalling equation (3.12),

$$\hat{m}_k^N(\boldsymbol{y}) = (2\pi)^{d/2}|\boldsymbol{\Psi}_k|^{1/2} p(\boldsymbol{y}|\tilde{\boldsymbol{\theta}}_k)\pi_k^N(\tilde{\boldsymbol{\theta}}_k) \tag{4.8}$$

where $\tilde{\boldsymbol{\theta}}_k$ is the posterior mode, $\boldsymbol{\Psi}_k$ is minus the inverse Hessian of $\log\left\{p(\boldsymbol{y}|\boldsymbol{\theta}_k)\pi_k^N(\boldsymbol{\theta}_k)\right\}$ evaluated at $\boldsymbol{\theta}_k = \tilde{\boldsymbol{\theta}}_k$, and d is the dimension of $\boldsymbol{\theta}_k$. Raftery (1996) suggests computing $\log\left\{p(\boldsymbol{y}|\boldsymbol{\theta}_k^{(i)})\pi_k^N(\boldsymbol{\theta}_k^{(i)})\right\}$ for each $t = 1, \ldots, M$ and estimating $\tilde{\boldsymbol{\theta}}_k$ as the value that maximizes the discrete log-posterior. $\boldsymbol{\Psi}_k$ is approximated by the sample variance of obtained from the posterior draws. The final approximation for $m_{\pi_k^*}(\boldsymbol{y})$ is

$$\hat{\hat{m}}_{\pi_k^*}(\boldsymbol{y}) = \frac{|\boldsymbol{\Psi}|^{1/2} p(\boldsymbol{y}|\tilde{\boldsymbol{\theta}}_k)\pi_k^N(\tilde{\boldsymbol{\theta}}_k)}{RM} \sum_{r=1}^{R} \frac{\sum_{m=1}^{M} p(\boldsymbol{y}_r^*|\boldsymbol{\theta}_{k,r}^{(m)})}{|\boldsymbol{\Psi}_{k,r}|^{1/2} p(\boldsymbol{y}_r^*|\tilde{\boldsymbol{\theta}}_{k,r})\pi_k^N(\tilde{\boldsymbol{\theta}}_{k,r})} \tag{4.9}$$

where $\tilde{\boldsymbol{\theta}}_{k,r}$ and $\boldsymbol{\Psi}_{k,r}$ are analogous to $\tilde{\boldsymbol{\theta}}_k$ and $\boldsymbol{\Psi}_k$, respectively, but based on the sample $\boldsymbol{\theta}_{k,r}^{(1)}, \ldots, \boldsymbol{\theta}_{k,r}^{(M)}$ taken from $\pi_k^N(\boldsymbol{\theta}|\boldsymbol{y}_r^*)$, for $r = 1, \ldots, R$.

Notice that $\pi_k^N(.)$ appears in both the numerator and denominator of (4.9), so cancelling out any multiplicative constant factor. In the next two sections we investigate the robustness of the expected posterior prior in three simulated situations and in the international exchange rate returns problem, extensively analysed in the last chapter.

4.4 Simulation studies

We studied three simulated dataset generated by: (i) one-factor model, (ii) a two-factor model, and (iii) a three-factor model. The main purpose is to investigate how sensitive to the choice of m^* is model comparison. We compare the EP prior approach with the naive analysis with proper but diffuse priors.

4.4.1 A one-factor model

We simulated $T = 200$ observations from a $k = 1$-factor model for a vector $\boldsymbol{y}$ of $m = 7$ variables. The factor loading matrix and the uniquenesses variances are fixed

at

$$\boldsymbol{\beta}_1 = (0.995, 0.975, 0.949, 0.922, 0.894, 0.866, 0.837)$$

and

$$diag(\boldsymbol{\Sigma}) = (0.01, 0.05, 0.10, 0.15, 0.20, 0.25, 0.30)$$

respectively. We have run the Gibbs sampler for $M_0 = 5000$ iterations, after which every tenth draw has been stored in a sample of size 1000 for further computation. In this model, the maximum number of parameters, i.e. when a three-factor model is fitted, is 25. Therefore, we have chosen to make $\boldsymbol{y}^*$ represent a sample with 26 observations taken from a multivariate normal distribution with zero mean. The covariance matrix used was the sample covariance matrix based on $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{200}$. Notice that this would correspond to a factor model with no factors at all, which in a sense represents the basic or null model, as suggested by Pérez and Berger (2000) as a natural candidate for m^*.

As prior hyperparameters we have tried three different sets, with $\mu_{ki} = 0$ and other parameters specified with values as specified in table 4.1 for $i = 1, \ldots, 7$ and $k = 1, \ldots, 3$. With these choices we believe we can express decreasing amounts of information present in the prior distributions, with (III) representing a rather noninformative prior distribution.

	C_{k0}	ν_{ki}	$\nu_{ki}s_{ki}^2$
I	1	2.2	0.1000
II	10	2.2	0.0100
III	100	2.2	0.0001

Table 4.1: Prior hyperparameters. C_{k0} represents the variance of the factor loadings and $\nu_{ki}/2$ and $\nu_{ki}s_{ki}^2/2$ represent the shape and scale of the prior for the idiosyncratic variances, σ_i^2.

Finally, we have chosen $R = 1000$ in order to compute the predictive distributions. Table 4.2 summarizes our findings. It can be seen that the predictive density

varies considerably when computed with different priors, while it is more stable or "robust" when based on expected posterior prior distributions. If the prior model probabilities are uniform, then posterior model probabilities, $\hat{P}_k^N(k)$ and $\hat{\hat{P}}_{\pi_k^*}(k)$ can also be computed. From table 4.2 it seems that in the context of the simulated example either prior set up leads to the same answer. In the next simulation we introduce a second but relatively unimportant factor in order to see whether or not the EP prior captures such characteristic.

Prior	$\hat{m}_k^N(\boldsymbol{y})$			$\hat{\hat{m}}_{\pi_k^*}(\boldsymbol{y})$		
	$k=1$	$k=2$	$k=3$	$k=1$	$k=2$	$k=3$
(I)	-933.7	-941.1	-948.6	-906.9	-908.4	-911.0
(II)	-950.2	-962.4	-970.9	-907.1	-911.1	-914.3
(III)	-989.4	-1009.3	-1030.4	-910.8	-916.3	-928.1
	$\hat{P}_k^N(k)$			$\hat{\hat{P}}_{\pi_k^*}(k)$		
	$k=1$	$k=2$	$k=3$	$k=1$	$k=2$	$k=3$
(I)	0.999	0.001	0.000	0.804	0.182	0.014
(II)	1.000	0.000	0.000	0.980	0.019	0.001
(III)	1.000	0.000	0.000	0.996	0.004	0.000

Table 4.2: Comparative results (one-factor model). $\hat{m}_k^N(\boldsymbol{y})$ and $\hat{\hat{m}}_{\pi_k^*}(\boldsymbol{y})$ are given by equations (4.8) and (4.9), respectively. Also, $\hat{P}_k^N(k)$ and $\hat{\hat{P}}_{\pi_k^*}(k)$ are the posterior model probabilities.

4.4.2 A two-factor model

In this example, we simulated again $T = 200$ observations, but from a $k = 2$-factor model for a vector $\boldsymbol{y}$ of $m = 7$ variables. The factor loading matrix and the uniquenesses variances are fixed at

$$\boldsymbol{\beta}' = \begin{pmatrix} 0.995 & 0.000 & 0.949 & 0.922 & 0.894 & 0.866 & 0.600 \\ 0.000 & 0.975 & 0.000 & 0.000 & 0.000 & 0.000 & 0.200 \end{pmatrix}$$

and $diag(\boldsymbol{\Sigma}) = (0.01, 0.05, 0.10, 0.15, 0.2, 0.25, 0.6)$, respectively. Notice that we chose these so that the second factor is virtually an independent factor. We ran the Gibbs sampler for $M_0 = 5000$ iterations, after which 10000 simulations where peformed and every tenth stored for further computation. In this model the maximum number of parameters is still 25 and, as before, $\boldsymbol{y}^*$ represents a sample of size 26 taken from a multivariate normal distribution with zero mean. The covariance matrix was computed accordingly. We used the same three different prior sets and $R = 100$. Our findings are summarized in tables 4.3 and 4.4. In table 4.4, m_k^* represents a multivariate normal distribution, with zero mean vector and covariance matrix defined as the maximum likelihood estimator obtained in a classical k-factor model. Technically, if $\hat{\boldsymbol{\beta}}_k$ and $\hat{\boldsymbol{\Sigma}}_k$ are the ML estimators of $\boldsymbol{\beta}_k$ and $\boldsymbol{\Sigma}_k$ in a k-factor model, then $m_k^* \sim N(\mathbf{0}, \hat{\boldsymbol{\beta}}_k \hat{\boldsymbol{\beta}}_k' + \hat{\boldsymbol{\Sigma}}_k)$. Notice that, as in the first example, a one-factor model is selected according to all priors when the EP prior approach is not used, even though the predictive estimates vary considerable for different prior hyperparameters. On the other hand, under the EP prior approach, the posterior model probabilities for a two-factor model is not negligible, being at least around 20%, while the estimated predictives are relatively similar across all hyperparameter specifications. Finally, as illustrated in Table 4.4, the choice of the probability measure, m_k^* seems to be unimportant since the results are virtually the same across alternative measures.

4.4.3 A three-factor model

We simulated $T = 200$ observations from a $k = 3$-factor model for a vector $\boldsymbol{y}$ of $m = 7$ variables. The factor loading matrix and the uniquenesses variances are fixed at

$$\boldsymbol{\beta}' = \begin{pmatrix} 0.795 & 0.000 & 0.000 & 0.722 & 0.000 & 0.000 & 0.722 \\ 0.000 & 0.775 & 0.000 & 0.000 & 0.794 & 0.000 & 0.000 \\ 0.000 & 0.000 & 0.749 & 0.000 & 0.000 & 0.766 & 0.000 \end{pmatrix}$$

Prior	$\hat{m}_k^N(\boldsymbol{y})$			$\hat{\hat{m}}_{\pi_k^*}(\boldsymbol{y})$		
	$k=1$	$k=2$	$k=3$	$k=1$	$k=2$	$k=3$
(I)	-1306.7	-1310.2	-1312.9	-1280.1	-1281.2	-1279.4
(II)	-1325.9	-1334.7	-1337.0	-1280.7	-1284.2	-1282.7
(III)	-1362.3	-1374.6	-1389.9	-1281.2	-1282.0	-1285.0
	$\hat{P}_k^N(k)$			$\hat{\hat{P}}_{\pi_k^*}(k)$		
	$k=1$	$k=2$	$k=3$	$k=1$	$k=2$	$k=3$
(I)	0.968	0.003	0.002	0.293	0.102	0.604
(II)	1.000	0.000	0.000	0.865	0.026	0.109
(III)	1.000	0.000	0.000	0.669	0.315	0.015

Table 4.3: Comparative results (two-factor model). $\hat{m}_k^N(\boldsymbol{y})$ and $\hat{\hat{m}}_{\pi_k^*}(\boldsymbol{y})$ are given by equations (4.8) and (4.9), respectively. Also, $\hat{P}_k^N(k)$ and $\hat{\hat{P}}_{\pi_k^*}(k)$ are the posterior model probabilities. $m^* \sim N(\mathbf{0}, \boldsymbol{S})$, where S is the sample covariance matrix.

and $diag(\boldsymbol{\Sigma}) = (0.368, 0.399, 0.439, 0.479, 0.370, 0.413, 0.479)$, respectively. We ran the Gibbs sampler for $M_0 = 1000$ iterations, after which $M = 1000$ draws were stored for further computation. In this model the maximum number of parameters is still 25 and, as before, $\boldsymbol{y}^*$ represents a sample of size 26 taken from a multivariate normal distribution with zero mean. The covariance matrix was computed accordingly. We used the same three different prior sets and $R = 1000$. Our findings are summarized in table 4.5. The posterior probabilities were all one hundred percent in favor of the $k = 3$-factor model. Notice that, even though all methods pointed out towards the correct model, more robust estimates were found for the various hyperparameter sets by using the expected posterior priors regardless of the choice of the probability measure, m_k^*.

4.5 International exchange rate returns (revisited)

In this section we revisited the real data application extensively analysed in Section 3.7 of last Chapter. As before, we set $\mu_{k0} = 0, C_{k0} = 1$ and $\nu_{ki} = 2.2, \nu_{ki} s_{ki}^2 = 0.1$

	$m^* = m_1^*$					
Prior	$\hat{\hat{m}}_{\pi_k^*}(\boldsymbol{y})$			$\hat{\hat{P}}_{\pi_k^*}(k)$		
	$k=1$	$k=2$	$k=3$	$k=1$	$k=2$	$k=3$
(I)	-1280.3	-1281.4	-1279.7	0.317	0.106	0.578
(II)	-1281.5	-1286.0	-1284.4	0.938	0.010	0.052
(III)	-1280.7	-1281.2	-1286.4	0.621	0.377	0.002
	$m^* = m_2^*$					
Prior	$\hat{\hat{m}}_{\pi_k^*}(\boldsymbol{y})$			$\hat{\hat{P}}_{\pi_k^*}(k)$		
	$k=1$	$k=2$	$k=3$	$k=1$	$k=2$	$k=3$
(I)	-1279.7	-1281.1	-1279.4	0.385	0.095	0.520
(II)	-1281.5	-1285.2	-1284.1	0.910	0.022	0.068
(III)	-1280.7	-1281.1	-1284.7	0.592	0.397	0.011
	$m^* = m_3^*$					
Prior	$\hat{\hat{m}}_{\pi_k^*}(\boldsymbol{y})$			$\hat{\hat{P}}_{\pi_k^*}(k)$		
	$k=1$	$k=2$	$k=3$	$k=1$	$k=2$	$k=3$
(I)	-1280.8	-1281.6	-1280.8	0.408	0.183	0.408
(II)	-1281.5	-1284.8	-1283.3	0.832	0.031	0.137
(III)	-1280.2	-1280.4	-1280.6	0.402	0.329	0.269

Table 4.4: Comparative results (two-factor model). $\hat{\hat{m}}_{\pi_k^*}(\boldsymbol{y})$'s are defined in (4.9), while $\hat{\hat{P}}_{\pi_k^*}(k)$'s the posterior model probabilities. m_k^* represents a multivariate normal distribution, with zero mean vector and covariance matrix defined as the maximum likelihood estimator obtained in a classical k-factor model.

for $i = 1, \ldots, 6$ and $k = 1, \ldots, 3$. The prior hyperparameters are the same ones used in the simulated cases, described in Section 3.6, also from last Chapter. The Gibbs sampler and the computation of normalizing constants were done in the same way, with $M_0 = M = 1000$ and $R = 1000$. Table (4.6) summarizes our findings.

As can be seen, there is not much difference when we look at the posterior probability for each model (here, the number of factors). However, it is clear how unstable the predictive is when based on noninformative priors indicating, as one would have expected, that the predictive is arbitrary in this case. On the other hand, by using expected posterior prior distributions the predictives are fairly robust. We have also

tried $m^* \sim N(\mathbf{0}, \widehat{\mathbf{\Omega}})$, where $\widehat{\mathbf{\Omega}}$ is the maximum likelihood estimator of $\boldsymbol{\beta}\boldsymbol{\beta}' + \mathbf{\Sigma}$ based on a one factor model. Despite the fact that intuition would suggest that such choice would rather favor a one-factor model, we have found, at least in this simulation, that this is not the case. As a matter of fact, the results are virtually the same as the one in table 4.6.

From the simulated and real data situations presented in this chapter, it can be seen, at least empirically, that Pérez and Berger's expected posterior prior performs considerably well in factor analysis context when prior information is lacking.

4.6 Summary

In this chapter we have specialized Pérez and Berger's (2000) expected posterior priors to the factor analysis context. Particular interest was focused on the sensitivity of Bayes factors (posterior model probabilities) to the choice of m^*. Despite the fact that, in our simulations, we notice that virtually all methods pointed out towards the truth (correctly assessing the number of factors), more robust estimates were found for the various hyperparameter sets by using the expected posterior priors. Alternative choices for m^*, the training sample size, the shape of the expected posterior prior distributions, among other issues, need further and deeper investigation and we anticipate their important practical impact into the field of latent factor models.

	$m^* \sim N(\mathbf{0}, \boldsymbol{S})$					
Prior	$\hat{m}_k^N(\boldsymbol{y})$			$\hat{\hat{m}}_{\pi_k^*}(\boldsymbol{y})$		
	$k=1$	$k=2$	$k=3$	$k=1$	$k=2$	$k=3$
(I)	-1954.4	-1915.9	-1884.5	-1923.5	-1887.2	-1853.5
(II)	-1978.8	-1944.3	-1918.6	-1928.6	-1886.7	-1856.2
(III)	-2022.1	-1988.2	-1970.5	-1926.3	-1874.4	-1840.5
	$m^* = m_1^*$					
Prior	$\hat{m}_k^N(\boldsymbol{y})$			$\hat{\hat{m}}_{\pi_k^*}(\boldsymbol{y})$		
	$k=1$	$k=2$	$k=3$	$k=1$	$k=2$	$k=3$
(I)	-1954.4	-1915.9	-1884.5	-1929.0	-1893.0	-1862.9
(II)	-1978.8	-1944.3	-1918.6	-1921.6	-1896.2	-1868.5
(III)	-2022.1	-1988.2	-1970.5	-1929.6	-1880.2	-1855.7
	$m^* = m_2^*$					
Prior	$\hat{m}_k^N(\boldsymbol{y})$			$\hat{\hat{m}}_{\pi_k^*}(\boldsymbol{y})$		
	$k=1$	$k=2$	$k=3$	$k=1$	$k=2$	$k=3$
(I)	-1954.4	-1915.9	-1884.5	-1922.1	-1881.5	-1858.0
(II)	-1978.8	-1944.3	-1918.6	-1925.7	-1880.7	-1861.4
(III)	-2022.1	-1988.2	-1970.5	-1924.9	-1874.2	-1850.5
	$m^* = m_3^*$					
Prior	$\hat{m}_k^N(\boldsymbol{y})$			$\hat{\hat{m}}_{\pi_k^*}(\boldsymbol{y})$		
	$k=1$	$k=2$	$k=3$	$k=1$	$k=2$	$k=3$
(I)	-1954.4	-1915.9	-1884.5	-1924.0	-1884.0	-1850.0
(II)	-1978.8	-1944.3	-1918.6	-1922.4	-1887.2	-1858.4
(III)	-2022.1	-1988.2	-1970.5	-1922.6	-1874.2	-1844.8

Table 4.5: Comparative results (three-factor model). $\hat{m}_k^N(\boldsymbol{y})$ and $\hat{\hat{m}}_{\pi_k^*}(\boldsymbol{y})$ are given by equations (4.8) and (4.9), respectively. Also, $\hat{P}_k^N(k)$ and $\hat{\hat{P}}_{\pi_k^*}(k)$ are the posterior model probabilities. $m^* \sim N(\mathbf{0}, \boldsymbol{S})$, where S is the sample covariance matrix.

Prior	$\hat{m}_k^N(\boldsymbol{y})$			$\hat{\hat{m}}_{\pi_k^*}(\boldsymbol{y})$		
	$k=1$	$k=2$	$k=3$	$k=1$	$k=2$	$k=3$
(I)	-1015.5	-904.4	-910.3	-994.4	-878.4	-884.1
(II)	-1034.7	-924.0	-936.2	-994.3	-870.3	-888.5
(III)	-1071.6	-960.4	-973.4	-996.0	-855.7	-859.4
	$\hat{P}_k^N(k)$			$\hat{\hat{P}}_{\pi_k^*}(k)$		
	$k=1$	$k=2$	$k=3$	$k=1$	$k=2$	$k=3$
(I)	0.0	0.997	0.003	0.0	0.997	0.003
(II)	0.0	1.000	0.000	0.0	1.000	0.000
(III)	0.0	0.994	0.006	0.0	0.976	0.024

Table 4.6: Comparative results for the exchange rate data. $\hat{m}_k^N(\boldsymbol{y})$ and $\hat{\hat{m}}_{\pi_k^*}(\boldsymbol{y})$ are given by equations (4.8) and (4.9), respectively. Also, $\hat{P}_k^N(k)$ and $\hat{\hat{P}}_{\pi_k^*}(k)$ are the posterior model probabilities. $m^* \sim N(\mathbf{0}, \boldsymbol{S})$, where S is the sample covariance matrix.

Chapter 5

Factor stochastic volatility and time-varying loadings

5.1 Introduction

The first part of this thesis, which comprises chapters 2, 3 and 4, were primarily devoted to studying model uncertainty issues in static factor models. It has been argued that, conditional on the number of common latent factors, a straightforward Gibbs sampler is available for posterior analysis (Geweke and Zhou, 1996; Aguilar and West, 2000; Pitt and Shephard, 1999b), as studied in chapter 2. For the case in which the number of common factors is unknown, an extensive list of alternative procedures to compute posterior model probability where extensively studied with theoretical and empirical findings summarized in chapters 3 and 4.

Factor models have received great academic attention since the early days at the beginning of the century with the work of Spearman (Bartholomew, 1995). However, until recently, with the advent of powerful and fast computer machinery, factor models have received less than full attention of practitioners. These models are now emerging in problems in the fields as diverse as geology, credit analysis, and financial markets. In this chapter and the next, we concentrate on factor models to character-

ize covariance structures in certain classes of multivariate stochastic volatility models, or more specifically, factor stochastic volatility models (FSV), with a particular view to applications in finance.

Stochastic volatility models are basically a class of time-series models that allow the time-series variances and covariances to evolve with time as stochastic functionals of past variances, covariances and possibly other information available. Further details about univariate stochastic volatility models, as well as comparisons with the well-known class of autoregressive conditionally heterokedastic (ARCH) models, can be found in Shephard (1996) and Kim *et al.* (1998). Although generalizations to multivariate situations are theoretically and conceptually simple, their implementation makes them practically infeasible to handle. Many attempts have being made to overcome dimensionality problems and the factor analysis model seems to be a natural candidate for the same reasons stressed above.

Diebold and Nerlove (1989) introduce the latent factor ARCH models, which is further explored and compared with other variance models in Sentana (1998) and Giakoumatos *et al.* (1999). The former studied the differences between Diebold and Nerlove's latent factor ARCH models and Engle (1987)'s factor ARCH models, while the latter compared a latent one-factor ARCH model with Shephard's unobserved ARCH model.

The works of Harvey *et al.* (1994) followed by Jacquier *et al.* (1995) and Kim *et al.* (1998), and more recently by Aguilar and West (2000) and Pitt and Shephard (1999b), form the basis for the model developments we consider in the next two chapters of the thesis. They basically model the levels of a set of time-series by a factor model where both the common factor variances and the specific (or idiosyncratic) variances follow multivariate and univariate first order stochastic volatility structures, respectively.

We extend these works in some theoretical and practical directions, our main

contributions consisting of *(i)* allowing the factor loadings to evolve in time, and *(ii)* estimating sequentially the parameters of the model, as opposed to fixing them at certain levels.

The first main contribution will be fully described in this chapter. To some extent, we believe that by allowing the factor loadings to change over time we maintain the factor scores interpretability virtually the same across time. In other words, the weight that some factors have on a particular time series might change with time, mimicking real financial/economic scenarios. One such example is when a country (or countries) enter/leave a particular market, and when such a market is been represented by a group of stable latent factors. The second main contribution is explored in the next chapter where filtering techniques recently proposed for sequential analysis of dynamic systems are adapted to the factor stochastic volatility context.

The rest of the Chapter is organized as follows. Section 5.2 sets up the model which can be split in two parts. In the first part a factor model is used to represent the level of time series dependence structure, while in the second one the variances of common factor follow a multivariate stochastic volatility model. Also in this section we lay down the prior distributions for the model's parameters. Section 5.3 presents in full details the posterior analysis, performed by an MCMC algorithm (hybrid Gibbs and Metropolis-Hastings algorithms) on both parts of the model. We close the chapter with an application in Section 5.4. For illustrative and comparative purposes we revisit the same dataset used in Aguilar and West (2000); these data are the returns on weekday closing spot prices currencies of six countries relative to the US dollar. We start with static factor analysis to gain some exploratory intuition about the time series covariance structure. We then implement factor stochastic volatility models with and without stochastic volatility structure for the specific variances. We close the chapter with some comments regarding the application and the overall

research on factor stochastic volatility models.

5.2 Model and prior

As in equation (2.1) from Chapter 2, we assume that $\boldsymbol{y}_t$ is a m-dimensional vector of time series, whose levels follow a k-factor model,

$$(\boldsymbol{y}_t|\boldsymbol{f}_t, \boldsymbol{\gamma}_t, \boldsymbol{\beta}_t, \boldsymbol{\Sigma}_t) \sim N(\boldsymbol{\gamma}_t + \boldsymbol{\beta}_t \boldsymbol{f}_t; \boldsymbol{\Sigma}_t) \tag{5.1}$$

where, again, $\boldsymbol{\gamma}_t$ is the m-dimensional mean level vector, $\boldsymbol{\beta}_t$ is the $m \times k$ factor loading matrix, $\boldsymbol{f}_t$ is the $k \times 1$ vector of common factors and $\boldsymbol{\Sigma}_t$ is the $m \times m$ diagonal matrix with the specific or idiosyncratic variances. The main differences lie in the time structure of the parameters in $\boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\Sigma}$ and the variances of $\boldsymbol{f}$:

- We assume that the mean level, $\boldsymbol{\gamma}_t$ follows a simple multivariate random walk process of the following form,

  $$(\boldsymbol{\gamma}_t|\boldsymbol{\gamma}_{t-1}, \boldsymbol{W}_t^{\gamma}) \sim N(\boldsymbol{\gamma}_{t-1}, \boldsymbol{W}_t^{\gamma}) \tag{5.2}$$

  to capture constant local (myopic) levels in the series. We will further assume that the evolution matrices $\boldsymbol{W}_t^{\gamma}$ are completely specified by a single and known discount factor $\delta_{\gamma} \in (0, 1)$. We refer to West and Harrison (1997) and to developments from Section 5.3 for further details.

- We assume that the common factors are independent and normally distributed over time, conditional on $\boldsymbol{H}_t$,

  $$(\boldsymbol{f}_t|\boldsymbol{H}_t) \sim N(\boldsymbol{0}; \boldsymbol{H}_t) \tag{5.3}$$

  with $\boldsymbol{H}_t = diag(h_{1t}, \ldots, h_{kt})$

- The factor log-variances, or simply $\lambda_{it} = log(h_{it})$, will follow a multivariate first-order autoregressive (VAR) model,

$$(\boldsymbol{\lambda}_t|\boldsymbol{\lambda}_{t-1}, \boldsymbol{\alpha}, \boldsymbol{\phi}, \boldsymbol{U}) \sim N(\boldsymbol{\alpha} + \boldsymbol{\phi}(\boldsymbol{\lambda}_{t-1} - \boldsymbol{\alpha}); \boldsymbol{U}) \tag{5.4}$$

for $\boldsymbol{\lambda}_t = (\lambda_{1t}, \ldots, \lambda_{kt})'$, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_k)'$, and $\boldsymbol{\phi} = diag(\phi_1, \ldots, \phi_k)$. Aguilar and West (2000) argue that allowing $\boldsymbol{U}$ to have nonzero covariances has important practical implications. We will follow this parametrization in this thesis. We also assume that $0 < \phi_i < 1$ for $i = 1, \ldots, k$, to guarantee nonexplosive behaviour in the final time series variances and covariances. This restriction also implies that changes in variances in a period of time are most certainly followed by consecutive changes in the same direction in the near future, an assumption that is fairly reasonable and observed quite oftenly in practice. Under such constraints it is easy to see that $\boldsymbol{\lambda}_1 \sim N(\boldsymbol{\mu}, \boldsymbol{W})$, where $\boldsymbol{W}$ satisfies $\boldsymbol{W} = \boldsymbol{\phi}\boldsymbol{W}\boldsymbol{\phi} + \boldsymbol{U}$ and results in an unknown full conditional for $\boldsymbol{\phi}$, even when the prior is conjugate. See Aguilar and West (2000) for further details and the discussion in Section 5.3.

- We also assume univariate stochastic volatility structures for the nonzero elements of $\boldsymbol{\Sigma}_t = diag(\sigma_{1t}^2, \ldots, \sigma_{mt}^2)$. More specifically, the idiosyncratic log-variances, $\eta_{it} = log(\sigma_{it}^2)$ follow standard first-order autoregressive models,

$$(\boldsymbol{\eta}_t|\boldsymbol{\eta}_{t-1}, \tilde{\boldsymbol{\alpha}}, \boldsymbol{\rho}, \boldsymbol{S}) \sim N(\tilde{\boldsymbol{\alpha}} + \boldsymbol{\rho}(\boldsymbol{\eta}_{t-1} - \tilde{\boldsymbol{\alpha}}); \boldsymbol{S}) \tag{5.5}$$

for $\boldsymbol{\eta}_t = (\eta_{1t}, \ldots, \eta_{mt})'$, $\tilde{\boldsymbol{\alpha}} = (\tilde{\alpha}_1, \ldots, \tilde{\alpha}_m)$, $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_m)$ and
$\boldsymbol{S} = diag(s_1, \ldots, s_m)$. As for the common factor variance equations, we assume that $0 < \rho_i < 1$ for $i = 1, \ldots, m$.

- Finally the loading matrices, $\boldsymbol{\beta}_t$, are constrained for identification reasons in the same manner described in Chapter 2; that is the upper diagonal elements

equal zero, $\beta_{ij,t} = 0$ for all $j > i$. The main difference lies in the fact that the diagonal elements are now set to one, $\beta_{ii,t} = 1$ for all $i = 1, \ldots, k$ to avoid confounding them with the common factor variances.

The unconstrained elements of $\boldsymbol{\beta}$ are stacked up in $\tilde{\boldsymbol{\beta}}_t = (\beta_{21,t}, \beta_{31,t}, \ldots, \beta_{n,k,t})$, a $d = mk - k(k-1)/2$ dimensional vector. As for the mean level vector, $\boldsymbol{\gamma}$, we will assume a first-order autoregression evolution for the unconstrained loadings,

$$(\tilde{\boldsymbol{\beta}}_t | \tilde{\boldsymbol{\beta}}_{t-1}, \boldsymbol{\zeta}, \boldsymbol{\Delta}, \boldsymbol{W}_t^{\beta}) \sim N(\boldsymbol{\zeta} + \boldsymbol{\Delta}\tilde{\boldsymbol{\beta}}_{t-1}; \boldsymbol{W}_t^{\beta}) \tag{5.6}$$

with $\boldsymbol{\zeta} = (\zeta_1, \ldots, \zeta_d), \boldsymbol{\Delta} = diag(\delta_1, \ldots, \delta_d)$. We will further assume that the evolution matrices $\boldsymbol{W}_t^{\beta}$ are completely specified by a single and known discount factor $\delta_{\beta} \in (0, 1)$.

It is worth mentioning that this model structure generalizes a number of currently available factor model structures:

1. $\delta_{\gamma} = \delta_{\beta} = 1$: Aguilar and West's dynamic factor model;
2. $u_{ij} = 0$ for $i \neq j$, $\delta_{\gamma} = \delta_{\beta} = 1$: Pitt and Shephard's factor stochastic volatility model;
3. $\boldsymbol{U} = \boldsymbol{0}, \boldsymbol{S} = \boldsymbol{0}$ recovers the traditional static factor model (Geweke and Zhou, 1996).

In order to complete the requisites for the Bayesian analysis, prior distributions must be defined and in this thesis will be restricted to conditionally conjugate prior distributions to facilitate the already complicated posterior analysis. To begin with, the prior distribution for the time series mean level at time $t = 1$ is

$$\boldsymbol{\gamma}_1 \sim N(\boldsymbol{\gamma}_0, \boldsymbol{\Sigma}_0) \tag{5.7}$$

for $\boldsymbol{\gamma}_0$ and $\boldsymbol{\Sigma}_0$ known hyperparameters. Analogously, the prior distribution for the unconstrained loadings at time $t = 1$ is,

$$\tilde{\boldsymbol{\beta}}_0 \sim N(\boldsymbol{m}_0, \boldsymbol{C}_0) \tag{5.8}$$

with $\boldsymbol{m}_0$ and $\boldsymbol{C}_0$ known hyperparameters. In our applications we will focus on the particular case where $\boldsymbol{m}_0 = \zeta_0 \mathbf{1}_d$ and $\boldsymbol{C}_0 = C_0 \boldsymbol{I}_d$, for m_0 and C_0 scalars. The prior distributions for components of $\boldsymbol{\zeta}$ and the nonzero components $\boldsymbol{\Delta}$ are

$$\zeta_j \sim N(\zeta_{0j}, C_{0j}) \tag{5.9}$$

$$\delta_j \sim N(\delta_{0j}, V_{0j}) \tag{5.10}$$

respectively, for $j = 1, \ldots, d$.

For the parameters that define both the factors's log-volatility equations, $\boldsymbol{\alpha}, \boldsymbol{\phi}, \boldsymbol{U}$, and the idiosyncratic factors's log-volatility,$\tilde{\boldsymbol{\alpha}}, \boldsymbol{\rho}, \boldsymbol{s}$, we follow Aguilar and West (2000) suggestions. They assume independent normal priors for the univariate terms of $\boldsymbol{\alpha}$ and $\boldsymbol{\alpha}$ and independent truncated normal priors for the terms in $\boldsymbol{\phi}$ and $\boldsymbol{\rho}$. Inverted Wishart and inverted gamma distributions are assigned to $\boldsymbol{U}$ and to the diagonal elements of $\boldsymbol{S}$, respectively. More specifically,

$$\boldsymbol{U} \sim IW(r_0, r_0 \boldsymbol{R}_0) \tag{5.11}$$

$$\boldsymbol{s}_i \sim IG(\nu_{0i}/2, \nu_{0i} s_{0i}^2/2) \qquad i = 1, \ldots, m \tag{5.12}$$

with $r_0, \boldsymbol{R}_0, \nu_{0i}$ and s_{0i}^2 given hyperparameters.

5.3 Posterior analysis

This section presents the proposed MCMC algorithm to obtain samples from the smoothed posterior disbribution of the parameters and the states (for $t \leq T$), i.e. the posterior distribution that are obtained when conditioning on all the observations

up to time T. In the following chapter we will relax this assumption. There we use algorithms that produce samples of the parameters and states sequentially, i.e. at each time $t \leq T$.

[*Distribution of the mean vector*]

Defining $\boldsymbol{z}_t = \boldsymbol{y}_t - \boldsymbol{\beta}_t \boldsymbol{f}_t$, it is easily seen from equations (5.1) and (5.2) that

$$\begin{aligned} \boldsymbol{z}_t &= \boldsymbol{\gamma}_t + \boldsymbol{\varepsilon}_t \qquad \boldsymbol{\varepsilon}_t \sim N(0, \boldsymbol{\Sigma}_t) \\ \boldsymbol{\gamma}_t &= \boldsymbol{\gamma}_{t-1} + \boldsymbol{\omega}_t \qquad \boldsymbol{\omega}_t \sim N(\mathbf{0}, \boldsymbol{W}_t^{\gamma}) \end{aligned}$$

defines a multivarite dynamic linear model. Using (5.7) as the distribution of $\boldsymbol{\gamma}_0$ given initially and assuming that $\boldsymbol{C}_0$ is diagonal, an algorithm to sample $\boldsymbol{\gamma}_t$ for all t can be built as simply as follows:

Forward: For each $t = 1, \ldots, T$,

$$(\boldsymbol{\gamma}_t | D_t) \sim N(\boldsymbol{m}_t, \boldsymbol{C}_t)$$

with $\boldsymbol{m}_t$ and $\boldsymbol{C}_t$ given by,

$$\begin{aligned} \boldsymbol{m}_t &= \boldsymbol{m}_{t-1} + \boldsymbol{A}_t(\boldsymbol{z}_t - \boldsymbol{m}_{t-1}) \\ \boldsymbol{C}_t &= \boldsymbol{C}_{t-1}/\delta_\gamma - \boldsymbol{A}_t \boldsymbol{Q}_t \boldsymbol{A}_t' \end{aligned}$$

with $\boldsymbol{Q}_t = \boldsymbol{C}_{t-1}/\delta_\gamma + \boldsymbol{\Sigma}_t$ and $\boldsymbol{A}_t = \boldsymbol{C}_{t-1} \left[\boldsymbol{C}_{t-1}/\delta_\gamma + \boldsymbol{\Sigma}_t\right]^{-1} /\delta_\gamma$

It is worthnoting that, being $\boldsymbol{C}_0$ diagonal, so will $\boldsymbol{Q}_t$ and $\boldsymbol{C}_t$, which simplifies considerably the calculations, which is crucial in a problem that already has so many parameters. For further details see West and Harrison (1997), pages 582-583.

Backward: Sample $\boldsymbol{\gamma}_t, t = 1, \ldots, T$ as follows:

1. For $t = T$ draw $\boldsymbol{\gamma}_T$ from $N(\boldsymbol{m}_T, \boldsymbol{C}_T)$;

2. For all other $t < T$ draw $\boldsymbol{\gamma}_t$ recursively and backwards from $N(\boldsymbol{h}_t^*, \boldsymbol{H}_t^*)$ where,

$$\begin{aligned} \boldsymbol{h}_t^* &= (1-\delta_\gamma)\boldsymbol{m}_t + \delta_\gamma \boldsymbol{\gamma}_{t+1} \\ \boldsymbol{H}_t^* &= (1-\delta_\gamma)\boldsymbol{C}_t \end{aligned}$$

   using results from West and Harrison (1997), pages 112-115, about filtering recurrences.

When $\boldsymbol{\gamma}_t = \boldsymbol{\gamma}$ for all t, the full conditional for $\boldsymbol{\gamma}$ can be easily obtained by combining the prior distribution (5.7) and the likelihood (5.1). In this case, the full conditional distribution of $\boldsymbol{\gamma}$ is multivariate normal with mean and variance given by

$$\left(\sum_{t=1}^{T} \boldsymbol{\Sigma}_t^{-1} + \boldsymbol{\Sigma}_0^{-1}\right)^{-1} \left(\sum_{t=1}^{T} \boldsymbol{\Sigma}_t^{-1}(\boldsymbol{y}_t - \boldsymbol{\beta}_t \boldsymbol{f}_t) + \boldsymbol{\Sigma}_0^{-1}\boldsymbol{\gamma}_0\right) \quad \text{and} \quad \left(\sum_{t=1}^{T} \boldsymbol{\Sigma}_t^{-1} + \boldsymbol{\Sigma}_0^{-1}\right)^{-1}$$

respectively.

[*Distribution of the common factor scores*]

Combining the prior distribution (5.3) and the likelihood (5.1) it is also easy to see that the full conditional distribution of $\boldsymbol{f}_t$ is multivariate normal with mean vector and variance-covariance matrix given, respectively, by:

$$(\boldsymbol{H}_t^{-1} + \boldsymbol{\beta}_t'\boldsymbol{\Sigma}_t^{-1}\boldsymbol{\beta}_t)^{-1}\boldsymbol{\beta}_t'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_t - \boldsymbol{\gamma}_t) \quad \textit{and} \quad (\boldsymbol{H}_t^{-1} + \boldsymbol{\beta}_t'\boldsymbol{\Sigma}_t^{-1}\boldsymbol{\beta}_t)^{-1}$$

independently for $t = 1, \ldots . T$.

[*Distribution of the factor loading equations*]

Conditionally on $\boldsymbol{W}_t^{\beta}, \tilde{\boldsymbol{\beta}}_t$ for all t, equation (5.6) when combined with the prior distribution for $\boldsymbol{\zeta}$ and $\boldsymbol{\Delta}$, (5.9) and (5.10), respectively, gives rise to a multivariate regression model, $\boldsymbol{Y} \sim N(\boldsymbol{X}\boldsymbol{\gamma}, \boldsymbol{W})$ with prior $\boldsymbol{\gamma} \sim N(\boldsymbol{a}_0, \boldsymbol{B}_0)$ where $\boldsymbol{Y} = (\tilde{\boldsymbol{\beta}}_1', \ldots, \tilde{\boldsymbol{\beta}}_T')'$, $\boldsymbol{\gamma} = (\zeta_1, \delta_1, \ldots, \zeta_d, \delta_d)'$, $\boldsymbol{X} = diag(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_T)$ with elements $\boldsymbol{X}_t = (\boldsymbol{1}_d, \tilde{\boldsymbol{\beta}}_t)$ for all t and $\boldsymbol{1}_d$ is a vector with d ones. Similarly, $\boldsymbol{W} = diag(\boldsymbol{W}_1^{\beta}, \ldots, \boldsymbol{W}_T^{\beta})$, $\boldsymbol{a}_0 = (\zeta_{01}, \delta_{01}, \ldots, \zeta_{0d}, \delta_{0d})'$ and $\boldsymbol{B}_0 = diag(C_{01}, V_{01}, \ldots, C_{0d}, V_{0d})$. Here, $\boldsymbol{W}_t^{\beta}$ is given by equation (5.16).

Therefore, from standard multivariate regression analysis (Zellner, 1971; Press, 1982; Broemeling, 1985), it follows that the posterior distribution of $\boldsymbol{\gamma}$ (in our case, a full conditional distribution) is also multivariate normal with mean vector and covariance matrix given by $\boldsymbol{a}_1 = \boldsymbol{B}_1(\boldsymbol{X}'\boldsymbol{W}^{-1}\boldsymbol{Y} + \boldsymbol{B}_0^{-1}\boldsymbol{a}_0)$ and $\boldsymbol{B}_1 = (\boldsymbol{X}'\boldsymbol{W}^{-1}\boldsymbol{X} + \boldsymbol{B}_0^{-1})^{-1}$ respectively, which can be easily computed if we further notice that $\boldsymbol{X}'\boldsymbol{W}^{-1}\boldsymbol{X} = \sum_{t=1}^{T} \boldsymbol{X}_t'\boldsymbol{W}_t^{-1}\boldsymbol{X}_t$ and $\boldsymbol{X}'\boldsymbol{W}^{-1}\boldsymbol{Y} = \sum_{t=1}^{T} \boldsymbol{X}_t'\boldsymbol{W}_t^{-1}\boldsymbol{Y}_t$.

[*Distribution of the factor loadings*]

Notice that equation (5.1) can be rewritten as $y_{1t} = \gamma_{1t} + f_{1t} + \epsilon_{1t}$. Also, define $\boldsymbol{z}_t = (z_{1t}, \ldots, z_{m-1,t})$ by

$$z_{it} = \begin{cases} y_{i+1,t} - \gamma_{i+1,t} - f_{i+1,t} & \text{if } \; i = 1, \ldots, k-1 \\ y_{i+1,t} - \gamma_{i+1,t} & \text{if } \; i = k, \ldots, m-1, \end{cases} \tag{5.13}$$

and $\boldsymbol{F}_t = diag(\boldsymbol{F}_{1t}, \ldots, \boldsymbol{F}_{N-1,t})$ with

$$\boldsymbol{F}_{it} = \begin{cases} (f_{1t}, \ldots, f_{it}) & \text{if } \; i = 1, \ldots, k-1 \\ (f_{1t}, \ldots, f_{Kt}) & \text{if } \; i = k, \ldots, m-1. \end{cases} \tag{5.14}$$

Notice that $\boldsymbol{F}_t$ is a $d \times (m-1)$ matrix. Equations (5.13) and (5.6) when combined with (5.8) form, conditionally on $\boldsymbol{F}_t, \boldsymbol{\Sigma}_t' s, \boldsymbol{\zeta}, \boldsymbol{\Delta}$ and $\boldsymbol{W}_t^{\beta}$, an well-known multivariate dynamic linear model as presented in West and Harrison (1997), pages 112-115 and

582-583. Recalling,

$$\begin{aligned} \boldsymbol{z}_t &= \boldsymbol{F}_t'\tilde{\boldsymbol{\beta}}_t + \boldsymbol{\varepsilon}_t \qquad \boldsymbol{\varepsilon}_t \sim N(\boldsymbol{0}, \boldsymbol{V}_t) \\ \tilde{\boldsymbol{\beta}}_t &= \boldsymbol{\zeta} + \boldsymbol{\Delta}\tilde{\boldsymbol{\beta}}_{t-1} + \boldsymbol{\xi}_t \qquad \boldsymbol{\xi}_t \sim N(\boldsymbol{0}, \boldsymbol{W}_t^{\beta}) \end{aligned} \tag{5.15}$$

with $\tilde{\boldsymbol{\beta}}_0 \sim N(\boldsymbol{m}_0, \boldsymbol{C}_0)$, $\boldsymbol{\varepsilon}_t = (\epsilon_{2t}, \ldots, \epsilon_{mt})'$, and $\boldsymbol{V}_t = diag(\sigma_{2t}^2, \ldots, \sigma_{mt}^2)$. Therefore, an algorithm to sample the $\tilde{\boldsymbol{\beta}}_t$ for all t can be performed as follows:

Forward: For each $t = 1, \ldots, T$,

$$(\tilde{\boldsymbol{\beta}}_t | D_t) \sim N(\boldsymbol{m}_t, \boldsymbol{C}_t)$$

with $\boldsymbol{m}_t = \boldsymbol{a}_t + \boldsymbol{A}_t(\boldsymbol{z}_t - \boldsymbol{F}_t'\boldsymbol{a}_t)$, $\boldsymbol{C}_t = \boldsymbol{R}_t - \boldsymbol{A}_t\boldsymbol{Q}_t\boldsymbol{A}_t'$, $\boldsymbol{a}_t = \boldsymbol{\zeta} + \boldsymbol{\Delta}\boldsymbol{m}_{t-1}$, $\boldsymbol{R}_t = \boldsymbol{\Delta}\boldsymbol{C}_{t-1}\boldsymbol{\Delta} + \boldsymbol{W}_t^{\beta}$, $\boldsymbol{Q}_t = \boldsymbol{F}_t'\boldsymbol{R}_t\boldsymbol{F}_t + \boldsymbol{V}_t$ and $\boldsymbol{A}_t = \boldsymbol{R}_t\boldsymbol{F}_t\boldsymbol{Q}_t^{-1}$. In this step $\boldsymbol{W}_t^{\beta}$ is computed indirectly as a function of the discount factor δ_β as

$$\boldsymbol{W}_t^{\beta} = \frac{1-\delta_\beta}{\delta_\beta}\boldsymbol{\Delta}\boldsymbol{C}_{t-1}\boldsymbol{\Delta} \tag{5.16}$$

Further details can be obtained from West and Harrison (1997), pages 582-583.

Backward: Sample $\tilde{\boldsymbol{\beta}}_t, t = 1, \ldots, T$ as follows:

1. For $t = T$ draw $\tilde{\boldsymbol{\beta}}_T$ from $N(\boldsymbol{m}_T, \boldsymbol{C}_T)$;
2. For all other $t < T$ draw $\tilde{\boldsymbol{\beta}}_t$ recursively and backwards from $N(\boldsymbol{h}_t^*, \boldsymbol{H}_t^*)$ where,$\boldsymbol{h}_t^* = \boldsymbol{m}_t + \boldsymbol{B}_t(\tilde{\boldsymbol{\beta}}_{t+1} - \boldsymbol{a}_{t+1})$ and $\boldsymbol{H}_t^* = \boldsymbol{C}_t - \boldsymbol{B}_t\boldsymbol{R}_{t+1}\boldsymbol{B}_t'$, with $\boldsymbol{B}_t = \boldsymbol{C}_t\boldsymbol{\Delta}'\boldsymbol{R}_{t+1}^{-1}$. Once again, additional details are found in West and Harrison (1997), pages 112-115.

When $\boldsymbol{\beta}_t = \boldsymbol{\beta}$ for all t, combining the likelihood from (5.15) with the prior from (5.8) the full conditional is a multivariate normal with mean vector and variance-covariance matrix given, respectively, by,

$$\left(\boldsymbol{C}_0^{-1} + \sum_{t=1}^{T}\boldsymbol{F}_t'\boldsymbol{V}_t^{-1}\boldsymbol{F}_t\right)^{-1}\left(\boldsymbol{C}_0^{-1}\boldsymbol{m}_0 + \sum_{t=1}^{T}\boldsymbol{F}_t'\boldsymbol{V}_t^{-1}\boldsymbol{z}_t\right)$$

and

$$\left(\boldsymbol{C}_0^{-1} + \sum_{t=1}^{T} \boldsymbol{F}_t' \boldsymbol{V}_t^{-1} \boldsymbol{F}_t\right)^{-1}$$

[*Distribution of the stochastic volatility parameters*]

This section is based on Appendix B from Aguilar and West (2000).

$\boldsymbol{\alpha}$: Conditional on values of each of $\boldsymbol{\lambda}_t$, $\boldsymbol{\phi}$ and $\boldsymbol{U}$, equation (5.4) can be thought of as describing a multivariate normal linear regression. When combined with conjugate priors for $\boldsymbol{\mu}$, it produces a multivariate normal as full conditional, which in turn is easy to sample from.

$\boldsymbol{\phi}$: Sampling $\boldsymbol{\phi}$ is carried out through a Metropolis-Hastings algorithm that takes advantages of $\boldsymbol{\phi}$'s full conditional distribution; see Aguilar and West (2000).

$\boldsymbol{U}$: As in the case of $\boldsymbol{\phi}$, the full conditional for $\boldsymbol{U}$ is affected by the nonlinearity of $\boldsymbol{W} = \boldsymbol{\phi}\boldsymbol{W}\boldsymbol{\phi} + \boldsymbol{U}$, the unconditional variance of $\boldsymbol{\lambda}_t$ (see equation 5.4). Therefore, Aguilar and West (2000) propose a Metropolis-Hastings algorithm that takes advantages of $\boldsymbol{U}$'s full conditional distribution.

$\tilde{\boldsymbol{\alpha}}$: Conditional on values of each of $\boldsymbol{\nu}_t$, $\boldsymbol{\rho}$ and $\boldsymbol{S}$, equation (5.5) can be thought of as describing a multivariate normal linear regression. When combined with conjugate priors for $\tilde{\boldsymbol{\alpha}}$, it produces a multivariate normal as full conditional. Notice that the very nature of $\boldsymbol{S}$ makes the multivariate normal factor in products of marginal univariate normals for each component of $\tilde{\boldsymbol{\alpha}}$.

$\boldsymbol{\rho}$: Conditional on values of each of $\boldsymbol{\nu}_t$, $\boldsymbol{\mu}$ and $\boldsymbol{S}$, equation (5.5) can be thought of as describing independent univarite normal linear regression which combined with a conjugate prior for ρ_i produces another univariate normal as full conditional. Since ρ_i is restricted to be in $(0, 1)$ those normal are truncated.

***S*:** Finally, conditionally on $\boldsymbol{\nu}_t$, $\tilde{\boldsymbol{\alpha}}$ and $\boldsymbol{\rho}$, equation (5.5) can be thought of as describing normal linear regressions, which combined with 5.12 produces inverse gammas as full conditionals.

[*Distribution of the common and idiosyncratic factors' volatilities*]

When the i-th common factor equation (expression 5.3)

$$\log(f_{it}^2) = \lambda_{it} + \nu_{it} \qquad \nu_{it} \sim log - \chi_1^2 \tag{5.17}$$

is combined with the equation for $\boldsymbol{\lambda}_i$ (expression 5.4) a non-Gaussian linear model is formed (Kim *et al.*, 1998; Aguilar and West, 2000). Kim *et al.* (1998) propose an algorithm to sample the states in this non-Gaussian linear model,while Aguilar and West (2000) extends it to the multivariate case.

Their idea, roughly speaking, is the following: *(i)* approximate the distribution of $\boldsymbol{\nu}_{it}$, a log-χ_1^2, by a specified finite mixture of normals, *(ii)* then introduce T indicator variables (traditional in estimating mixture of normals Diebolt and Robert, 1994) in order to identify from which term of the mixture each of the $\boldsymbol{\nu}_{it}$ comes; and *(iii)* finally apply, conditionally on the set of indicators, a forward-filtering, backward sampling (FFBS) algorithm (Carter and Kohn, 1994; Frühwirth-Schnatter, 1994) in order to sample $\boldsymbol{\lambda}_i = (\lambda_{i1}, \ldots, \lambda_{iT})$. Notice that the FFBS algorithms were already used when sampling $\boldsymbol{\gamma}_t$ and $\tilde{\boldsymbol{\beta}}_t$ (Section 5.3).

The same strategy is applied to sample the idiosyncratic volatilities, $\boldsymbol{\eta}_t$. Firstly, define $\boldsymbol{\varepsilon}_t = \boldsymbol{y}_t - \boldsymbol{\gamma}_t - \boldsymbol{\beta}_t \boldsymbol{f}_t$ from equation (5.1). Then when the i-th idiosyncratic factor equation (expression 5.1),

$$\log(\varepsilon_{it}^2) = \eta_{it} + \nu_{it} \qquad \nu_{it} \sim log - \chi_1^2 \tag{5.18}$$

is combined with the equation for $\boldsymbol{\eta}_i$ (expression 5.5) forms a non-Gaussian linear

model and the whole procedure used to sample the $\tilde{\boldsymbol{\beta}}_t$'s can be identically be used to sample $\boldsymbol{\eta}_t$'s.

In the next section we apply this machinary to a real problem that involve six countries's exchange rates returns and over ten years of daily observations.

5.4 Daily exchange rate returns

We analyse the same dataset used in Aguilar and West (2000), namely the returns on weekday closing spot prices for six currencies relative to the US dollar. The dataset contains 2872 observations that range from 12/31/1987 to 01/01/1999. The 2872 observations will be transformed in order to use one-day-ahead returns, resulting in the lost of the first observation. Figure 5.1 shows the transformed time series in the order used by Aguilar and West (2000): German Mark (DEM), British Pound (GBP), Japanese Yen (JPY), French Franc (FRF), Canadian Dollar (CAD) and Spanish Peseta (ESP).

5.4.1 Exploratory analysis: the whole dataset

We first develop traditional exploratory analyses. To gain more knowledge about the data and to set the ground for further and deeper investigations with more complex factor models.

Figure 5.1 presents the 2871 observations. The two vertical lines represent 1/1/1992 and 10/31/95, respectively. We will focus on the final two thirds of the data to keep the analysis as close as possible to that presented in Aguilar and West (2000). The

following matrix represents the time series sample correlations:

$$\begin{pmatrix} 1.000 & 0.732 & 0.541 & 0.935 & 0.010 & 0.783 \\ 0.732 & 1.000 & 0.430 & 0.721 & 0.063 & 0.625 \\ 0.541 & 0.430 & 1.000 & 0.527 & -0.011 & 0.408 \\ 0.935 & 0.721 & 0.527 & 1.000 & 0.021 & 0.769 \\ 0.010 & 0.063 & -0.011 & 0.021 & 1.000 & 0.026 \\ 0.783 & 0.625 & 0.408 & 0.769 & 0.026 & 1.000 \end{pmatrix}$$

Figures 5.2 and 5.3 present the sample autocorrelogram for the time series and their square transformations.

Classical static factor analysis is performed and the following estimates for the factor loading matrix and idiosyncratic variances were found when fitting a $k = 3$-factor model (expression 2.1 of Session 2.2 might be helpful to recall the notation and the model used here):

$$\boldsymbol{\beta} = \begin{pmatrix} 1.00 & 0.00 & 0.00 \\ 0.77 & 1.00 & 0.00 \\ 0.57 & 0.04 & 1.00 \\ 0.98 & 0.01 & -0.05 \\ 0.01 & 0.12 & -0.34 \\ 0.82 & 0.05 & -7.86 \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma} = \text{diag} \begin{pmatrix} 1.975419e-06 \\ 2.944112e-11 \\ 4.137319e-05 \\ 3.393270e-06 \\ 9.090150e-06 \\ 3.211583e-06 \end{pmatrix}$$

These are crude estimates which do not take into consideration any time-varying structure for the time series covariance. Nevertheless, they point out some interesting directions; we summarize below.

- The GBP, DEM, FRF and ESP time series are overall fairly correlated, with a slight decrease after 10/31/95 (827 observations, not graphically presented here);
- The CAN time series does not seem to be correlated to any of the other returns.
- The JPY time series seems to be correlated with DEM, FRF and ESP, to some extent. Once again, this is somewhat diminished after 10/31/95.

- The first factor weights (first column of the factor loading matrix) has basically the same structure when one, two or three factor models are fitted to the data. Basically, DEM, GBP, FRF and ESP dominate this factor, with JPY having minor influence and CAD placing no influence at all.

- The second and third factors seem to be less important and represent an unique time series each, GBP and JPY, respectively;

- The sample autocorrelation functions for the squared time series suggest the presence of marked correlations in the variances of the time series.

5.4.2 Exploratory analysis: the shorter dataset

In this second part of our exploratory analysis we focus on the time series from 1/1/1992 to 10/31/1995, the same portion used by Aguilar and West (2000) in their work. This correspond to 1000 observations. They used this dataset to perform what they call *retrospective model analyses*, in which posterior inference for the states and parameters of the model, as defined at the beginning of this chapter, are assessed based on the whole information set, as opposed to *sequential analysis*, discussed in the next chapter. However, following analysis of the previous section, we will perform some exploratory investigation to learn more about some aspects of the data.

Figure (5.4) magnifies the time series within the period analysed, while

$$\begin{pmatrix} 1.000 & 0.442 & 0.377 & 0.966 & -0.021 & 0.876 \\ 0.442 & 1.000 & 0.185 & 0.441 & 0.031 & 0.415 \\ 0.377 & 0.185 & 1.000 & 0.366 & 0.108 & 0.325 \\ 0.966 & 0.441 & 0.366 & 1.000 & -0.014 & 0.877 \\ -0.021 & 0.031 & 0.108 & -0.014 & 1.000 & -0.009 \\ 0.876 & 0.415 & 0.325 & 0.877 & -0.009 & 1.000 \end{pmatrix}$$

represents their sample correlations. Also for that period, Figures 5.5 and 5.6 show the time series and its squared transformation's sample autocorrelation functions.

Static factor model analysis was performed and the following estimates found:

$$\boldsymbol{\beta} = \begin{pmatrix} 1.00 & 0.00 & 0.00 \\ 0.74 & 1.00 & 0.00 \\ 0.72 & -0.39 & 1.00 \\ 0.96 & 0.05 & 0.00 \\ -0.05 & 0.24 & -0.08 \\ 0.87 & 0.39 & -0.08 \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma} = \text{diag} \begin{pmatrix} 2.921264e-06 \\ 9.963879e-06 \\ 2.656160e-06 \\ 2.338258e-06 \\ 8.535363e-06 \\ 1.268561e-05 \end{pmatrix}.$$

The results are quite similar to those stated when using the whole dataset. Next our model is fitted to the shortened dataset. We decided to fit a three-factor model with stochastic volatilty for the common factor variances.

5.4.3 Factor stochastic volatility model

In this section we use the model structure developed in Section 5.2 for the thousand observations on the international exchange rate returns data from 1/1/1992 to 10/31/1995. Relatively vague prior were implemented for all model parameters following Aguilar and West (2000). For this section we will assume that $\boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}$ for all t, such its i-th component, σ_i^2, follows, *a priori*, an inverse gamma distribution with parameters $\nu_{0i} = 3$ and $\nu_{0i} s_{0i}^2 = 1$ (following the notation from Section 2.5, from Chapter 2). Notice that the prior is set up for the data in standardized form. The hyperparameters for the time series mean level, $\boldsymbol{\gamma}_i$ in expression (5.7) are $\boldsymbol{\gamma}_{0i} = 0$ and $\boldsymbol{\Sigma}_0 = 100,000\boldsymbol{I}_6$, describing quite vaguely the information about the time series locations. The unconstrained elements of $\boldsymbol{\beta}_0$ are normally distributed with zero mean and unit variance. Finally, the prior for the stochastic volatility regression parameters, α_i, ϕ_i are $N(0, 25)$ and $2Be(20; 1.5) - 1$, respectively. For $\boldsymbol{U}$ we chose $\boldsymbol{R}_0 = 0.0015\boldsymbol{I}_3$ and $r_0 = 20$, in Aguilar and West's notation. Other combinations of $\boldsymbol{R}_0$ and r_0 where tested resulting in most of the parameters being unaffected, apart from those concerning the stochastic volatility equations. The discount factors δ_β and δ_γ were both set equal to 0.9975 representing slight moves on the levels of the series and on

the factor loadings, respectively. Similar results were achieved with lower values for the discount factors, such as 0.99. The MCMC was run for 35,000 iterations, being the last 5,000 used for perform the analysis. Different starting values were tried as well as different MCMC burn-in lengths. In general the results were pretty much the same, with the chain pratically converging after 20,000 iterations.

Tables 5.1 and 5.2 present posterior estimates (posterior means and standard deviations), for the time series levels and their idiosyncratic variances. Also, an estimate of the correlation matrix at time $t = T = 1000$ is

$$\begin{pmatrix} 0.004 & & & & & \\ 0.629 & 0.003 & & & & \\ 0.419 & 0.352 & 0.006 & & & \\ 0.726 & 0.640 & 0.425 & 0.004 & & \\ -0.119 & -0.101 & -0.073 & -0.120 & 0.003 & \\ 0.726 & 0.643 & 0.418 & 0.737 & -0.119 & 0.004 \end{pmatrix}$$

with standard deviations along the main diagonal. Notice that similarity between this estimate with the one obtained in the preliminary analysis in Section 5.4.2. Figure 5.7 present the time series standard deviations and correlations through time. Also, the following exhibit shows point estimates (posterior means) for $\boldsymbol{\beta}_t$ for $t = 1, 500, 1000$, while the parameters defining the stochastic volatility equations are summarized in Table 5.3.

$$\hat{\boldsymbol{\beta}}_t = \left(\begin{array}{ccc|ccc|ccc} & t=1 & & & t=500 & & & t=1000 & \\ 1.00 & 0.00 & 0.00 & 1.00 & 0.00 & 0.00 & 1.00 & 0.00 & 0.00 \\ 0.85 & 1.00 & 0.00 & 0.75 & 1.00 & 0.00 & 0.63 & 1.00 & 0.00 \\ 0.54 & 0.24 & 1.00 & 0.63 & -0.04 & 1.00 & 0.79 & -0.20 & 1.00 \\ 0.94 & 0.10 & 0.52 & 0.95 & 0.03 & -0.02 & 0.96 & 0.02 & -0.03 \\ 0.04 & 0.15 & 0.15 & -0.05 & 0.03 & -0.12 & -0.10 & 0.01 & -0.20 \\ 0.93 & 0.08 & -2.77 & 0.93 & 0.02 & -3.40 & 0.92 & 0.02 & -3.40 \end{array}\right)$$

Finally, some graphical results are presented in Figures 5.8 to 5.10. Figure 5.8 shows the posterior means for the factor loadings through time, while Figure 5.9

presents the posterior means for the common factors itself and their standard deviations. Finally, Figure 5.10 present the proportion of the time series variances explained by each of the factors, common and specific ones.

Some interesting points arise from this analysis.

- It is interesting to note how similar the first factor loadings are to their counterparts obtained in Section 5.4.2.

- Since the discount factor δ_β was set at a very high level (0.9975), very smooth trends describe the evolution of the factor loadings. Nonetheless, the second and third factor demonstrate non-constancy in some of their loadings. More specifically, the second factor's importance in explaining the Japanese Yen decreases with time.

- To some extent, the third factor seems to provide little or no contribution to the fit of the model. In the next section, we discuss stochastic volatility structure for the idiosyncratic variances, which sheds light at the interpretation of the third factor. It seems that the second and third factors are responsible for time evolving movements in the variances of two currencies, the British Pound and the Japanese Yen, respectively.

- Looking at the common factor standard deviations, it can be argued that the first two factors are indeed more important than the third one. As far as forecasting performance is concerned, it might be that the third factor is in fact important, even though it may seem not from a model fitting standpoint.

- The high correlation between the log-volatilities (see Table 5.3, might indicate that a three-factor model is not needed to explain the codependence among the time series.

t	DEM	GBP	JPY	FRF	CAD	ESP
1	1.86	1.16	2.96	2.03	-2.15	1.28
100	2.27	1.09	2.81	2.30	-2.20	1.05
300	2.36	1.07	4.04	2.08	-1.80	0.99
500	2.31	1.06	2.79	2.08	-1.47	1.04
700	2.33	0.98	1.61	2.06	-0.75	1.18
900	2.24	0.93	0.20	2.08	-0.25	1.31
1000	2.27	0.93	-0.32	2.08	-0.16	1.32

Table 5.1: Three-factor model: retrospective posterior means for $\theta_t(\times 10^{-4})$, $t = 1, 100, 300, 500, 700, 900$ and 1000 (1/1/1992 to 10/ 31/ 1995)

- From Figure 5.10 it can be observed that the first common factor explains at least 60 percent of all time series variances at all times, for all but the Canadian currency. As a matter of fact, that currency's variance is completely determined by its idiosyncratic variance. Also according to this plot, the second factor is responsible for about 40 percent of the British currency's variability and is meaningless for the others, while the third factor exhibits the same behavior with respect to the Spanish currency's variance. These results reinforce our view that including stochastic volatility structure for the specific variances can reduce the number of factors by two, since the second and third factors are basic describing individual countries' variabilities.

Overall it seems that a two-factor model would suffice to explain the exchange rate returns comovements. However, before more exploration we will structure the idiosyncrasies by allowing them to follow independent first-order stochastic volatility autoregressive models. This is done in the next section.

5.4.4 Stochastic volatility model for the idiosyncrasies

In this part, we allow for heterokedasticity in the idiosyncratic variances, σ_i^2, as described in Section .

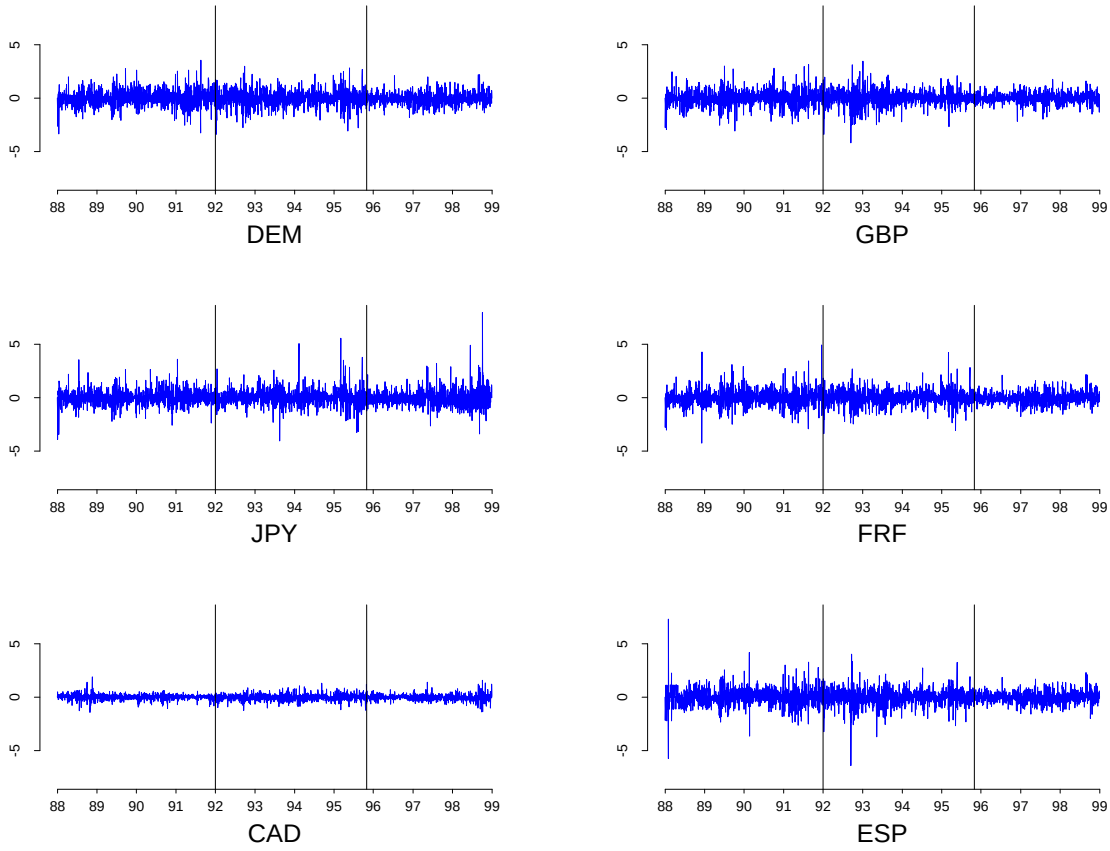


Figure 5.1: Returns from 1/1/1988 to 01/01/1999 (2871 observations). The vertical lines indicates 1/1/1992 and 10/31/95, respectively.

We set the same hyperparameters and MCMC quantities as fully described in Section 5.4.3. The prior for $\tilde{\alpha}_i, \rho_i$ are relatively vague, while for s_i we have an inverse gamma with mean 0.0004 and 25 degrees of freedom. As before, Tables 5.4 and 5.5 present the time series levels and idiosyncratic variances posterior estimates for some

Country	$E(\sigma^2)$	$SD(\sigma^2)$
DEM	3.69	0.42
GBP	2.20	0.44
JPY	32.37	1.51
FRF	2.98	0.34
CAD	8.75	0.40
ESP	1.71	0.23

Table 5.2: Three-factor model: retrospective posterior means and standard deviations for $\sigma_i^2(\times 10^{-6})$ (1/1/1992 to 10/31/1995)

i	α_i	ϕ_i	$\boldsymbol{U}$		
1	-10.021(0.479)	0.977(0.010)	0.058(0.017)	*0.979(0.011)*	*0.969(0.018)*
2	-11.106(1.744)	0.994(0.003)	0.110(0.026)	0.198(0.048)	*0.989(0.005)*
3	-13.951(3.396)	0.994(0.004)	0.225(0.060)	0.397(0.076)	0.737(0.190)

Table 5.3: Three factor model: retrospective posterior means and standard deviations (in parenthesis) for the stochastic volatility parameters. The upper diagonal entries for $\boldsymbol{U}$ represent correlations. (1/1/1992 to 10/ 31/ 1995)

periods of time. An estimate of the correlation matrix at time $t = T = 1000$ is

$$\begin{pmatrix} 0.004 & & & & & \\ 0.664 & 0.003 & & & & \\ 0.584 & 0.500 & 0.004 & & & \\ 0.683 & 0.600 & 0.520 & 0.005 & & \\ -0.098 & -0.087 & -0.117 & -0.088 & 0.003 & \\ 0.719 & 0.639 & 0.541 & 0.647 & -0.092 & 0.004 \end{pmatrix}$$

which is a function of the MCMC samples' ergodic average of $\boldsymbol{\beta}_T \boldsymbol{H}_t \boldsymbol{\beta}_T + \boldsymbol{\Sigma}_T$.

Also, the following exhibit shows point estimates (posterior means) for $\boldsymbol{\beta}_t$ for

$t = 1, 500, 1000,$

$$\hat{\boldsymbol{\beta}}_t = \left(\begin{array}{ccc|ccc|ccc} & t=1 & & & t=500 & & & t=1000 & \\ 1.00 & 0.00 & 0.00 & 1.00 & 0.00 & 0.00 & 1.00 & 0.00 & 0.00 \\ 0.80 & 1.00 & 0.00 & 0.70 & 1.00 & 0.00 & 0.58 & 1.00 & 0.00 \\ 0.53 & 0.34 & 1.00 & 0.63 & 0.03 & 1.00 & 0.79 & -0.06 & 1.00 \\ 0.96 & 0.03 & -0.01 & 0.95 & 0.02 & -0.01 & 0.95 & 0.02 & -0.01 \\ 0.04 & 0.11 & -0.03 & -0.05 & 0.02 & -0.09 & -0.09 & -0.01 & -0.09 \\ 0.92 & 0.13 & 0.03 & 0.92 & 0.05 & -0.02 & 0.89 & 0.06 & -0.03 \end{array}\right)$$

The inferences for parameters defining the stochastic volatility equations for both the common factor variances and for the idiosyncratic variances are summarized in the Tables 5.6 and 5.7, respectively.

As in the case where the idiosyncratic variances were static, graphical summaries are presented in Figures 5.12 to 5.10. Figure 5.12 shows the posterior means for the factor loadings through time, while Figure 5.13 presents the posterior means for the common factors itself and their standard deviations. Finally, Figure 5.14 present the proportion of the time series variances explained by each of the factors, common and specific.

5.5 Summary

In the next chapter we will compare the predictive performance of two models. However, some interesting facts show up in the analysis performed up to this point and deserve to be mentioned. They are as follows:

- Allowing the idiosyncratic variances to follow stochastic volatility models has little or impact on the point estimates for most of the parameters. Apart from some of the entries in $\boldsymbol{\beta}_t$, most of the quantities of interest remains virtually at the same estimates.

- The extended model just exarcebates our previous comment about the necessity of fewer common factors. Previously, a common factor that represented only one the variables could be understood as a adaptation of the model to allow some of the series to have specific variances that varied with time. Now, however, by allowing the specific variances of the series to follow independent stochastic volatility models, we force the factors to fully represent real interactions (possibly dynamic) between the time series at hand.

- An identification problem arises when a particular factor is represented by just one of the time series. The reason is quite simple. In such a situation, the model cannot distinguish between that common factor and the series' idiosyncratic variance. Some preliminary analysis with just two, or even one, common factor revealed that the first factor is unchanged, but the other two seems to be unnecessary in the model. Once again, to confirm this some sort of model comparison based on predictive and/or forecast issues must be performed.

- The previous comment is exacerbated if we compare Figures 5.10 and 5.14 describing the relevance of each factor, common and specific, in each one of the time series. The proportion of the variance explained by the first common factor is virtually unchanged, the same being said about the second common factor. The third factor is the one that sheds more light on our previous comments. For instance, when the idiosyncratic variances were not modeled through time (see Figure 5.10), the variability of the Japanese currency was explained by the first common factor (40%), the third common factor (40%), and by its idiosyncratic factor (20%), for most of the time period. When a stochastic volatility is included to explain its specific variance (see Figure 5.14), all the weight attributed to the third factor is relocated to its specific factor, indicating that, in fact, the third factor is unimportant. The Spanish Peseta has a similar

behaviour too.

- Following the previous discussion we have considered a one and two-factor model. In the one-factor model, the posterior distribution for the parameters defining the factor stochastic volatility are such that: $E(\alpha) = -9.904$, $SD(\alpha) = 0.313$, $E(\phi) = 0.9891, SD(\phi) = 0.0057$ and $E(U) = 0.0037, SD(U) = 0.0015$. Also, the following exhibit shows point estimates (posterior means) for $\boldsymbol{\beta}_t$ for $t = 1, 500, 1000$

$$\hat{\boldsymbol{\beta}}_t = \left(\begin{array}{c|c|c} t=1 & t=500 & t=1000 \\ 1.00 & 1.00 & 1.00 \\ 0.84 & 0.74 & 0.65 \\ 0.46 & 0.55 & 0.69 \\ 0.96 & 0.96 & 0.96 \\ 0.05 & -0.04 & -0.10 \\ 0.93 & 0.92 & 0.90 \end{array}\right)$$

for the one-factor model and,

$$\hat{\boldsymbol{\beta}}_t = \left(\begin{array}{cc|cc|cc} \multicolumn{2}{c|}{t=1} & \multicolumn{2}{c|}{t=500} & \multicolumn{2}{c}{t=1000} \\ 1.00 & 0.00 & 1.00 & 0.00 & 1.00 & 0.00 \\ 0.73 & 1.00 & 0.73 & 1.00 & 0.73 & 1.00 \\ 0.41 & 0.32 & 0.54 & 0.00 & 0.67 & -0.11 \\ 0.96 & 0.03 & 0.96 & 0.03 & 0.96 & 0.03 \\ 0.03 & 0.12 & -0.05 & 0.04 & -0.10 & 0.01 \\ 0.91 & 0.15 & 0.92 & 0.06 & 0.90 & 0.08 \end{array}\right)$$

for the two-factor model. As we expected, the time series variances explained by the second and third factors moves to the idiosyncratic components in the one-factor model. The second and third factors are basically responsible for part of the variances of GBP and JPY, respectively, and are those two currencies that have their idiosyncratic variances changed as the number of factors is decreased from three to one. Changes on DEM, FRF, CAN and JPY's idiosyncratic variances are immaterial, suggesting that an one-factor model is enough to

t	DEM	GBP	JPY	FRF	CAD	ESP
1	2.41	1.49	0.25	2.87	-1.81	2.06
100	2.48	1.50	0.32	2.87	-1.82	1.70
300	2.62	1.50	0.43	2.85	-1.38	1.53
500	2.68	1.50	0.37	2.84	-1.18	1.48
700	2.76	1.49	0.31	2.83	-0.55	1.58
900	2.76	1.48	0.23	2.84	-0.20	1.58
1000	2.77	1.48	0.20	2.84	-0.19	1.56

Table 5.4: Three-factor model: retrospective posterior means for $\theta_t(\times 10^{-4})$, $t = 1, 100, 300, 500, 700, 900$ and 1000 when the idiosyncrasies have stochastic volatility structures (1/1/1992 to 10/31/1995)

t	DEM	GBP	JPY	FRF	CAD	ESP
1	0.36	0.2	1.13	0.23	8.00	3.86
100	0.44	0.2	1.13	0.25	6.82	0.76
300	0.54	0.2	1.13	0.16	8.30	39.36
500	0.84	0.2	1.13	2.93	10.89	15.15
700	0.88	0.2	1.13	0.71	7.07	5.51
900	1.88	0.2	1.13	0.70	8.52	9.45
1000	2.58	0.2	1.13	9.83	12.07	4.49

Table 5.5: Three-factor model: retrospective posterior means for $\sigma_i^2(\times 10^{-6})$ when the idiosyncrasies have stochastic volatility structures (1/1/1992 to 10/31/1995).

explain the co-movements in those currencies. We will discuss further this issue later on in the next chapter when conducting sequential predictive and portfolio comparisons amongst alternative models.

- Finally, it is interesting to observe, from figure 5.14, that the importance of the European factor in explaining Great Britain's volatility decreases by the end of 1992 the year in which England withdrew from the European Union agreement. From that point on, Britain's volatility is partially explained by the second common factor.

i	α_i	ϕ_i	$\boldsymbol{U}$		
1	-10.191(0.376)	0.978(0.008)	0.053(0.015)	*0.990(0.005)*	*0.990(0.008)*
2	-11.828(1.351)	0.991(0.004)	0.076(0.020)	0.060(0.016)	*0.994(0.004)*
3	-10.942(0.866)	0.988(0.005)	0.115(0.030)	0.089(0.023)	0.078(0.024)

Table 5.6: Three-factor model: retrospective posterior means and standard deviations (in parenthesis) for the stochastic volatility parameters when the idiosyncrasies have stochastic volatility structures . The upper diagonal entries for $\boldsymbol{U}$ represent correlations (1/1/1992 to 10/31/1995).

i	$\tilde{\alpha}_i$	ρ_i	s_i
1	-13.665(1.507)	0.999(0.001)	0.001(0.000)
2	-15.446(0.212)	0.843(0.105)	0.000(0.000)
3	-13.758(0.323)	0.879(0.093)	0.000(0.000)
4	-14.351(0.411)	0.922(0.023)	0.648(0.181)
5	-11.651(0.190)	0.992(0.005)	0.001(0.000)
6	-11.967(0.248)	0.937(0.020)	0.201(0.061)

Table 5.7: Three-factor model: retrospective posterior means and standard deviations (in parenthesis) for the stochastic volatility parameters when the idiosyncrasies have stochastic volatility structures (1/1/1992 to 10/31/1995).

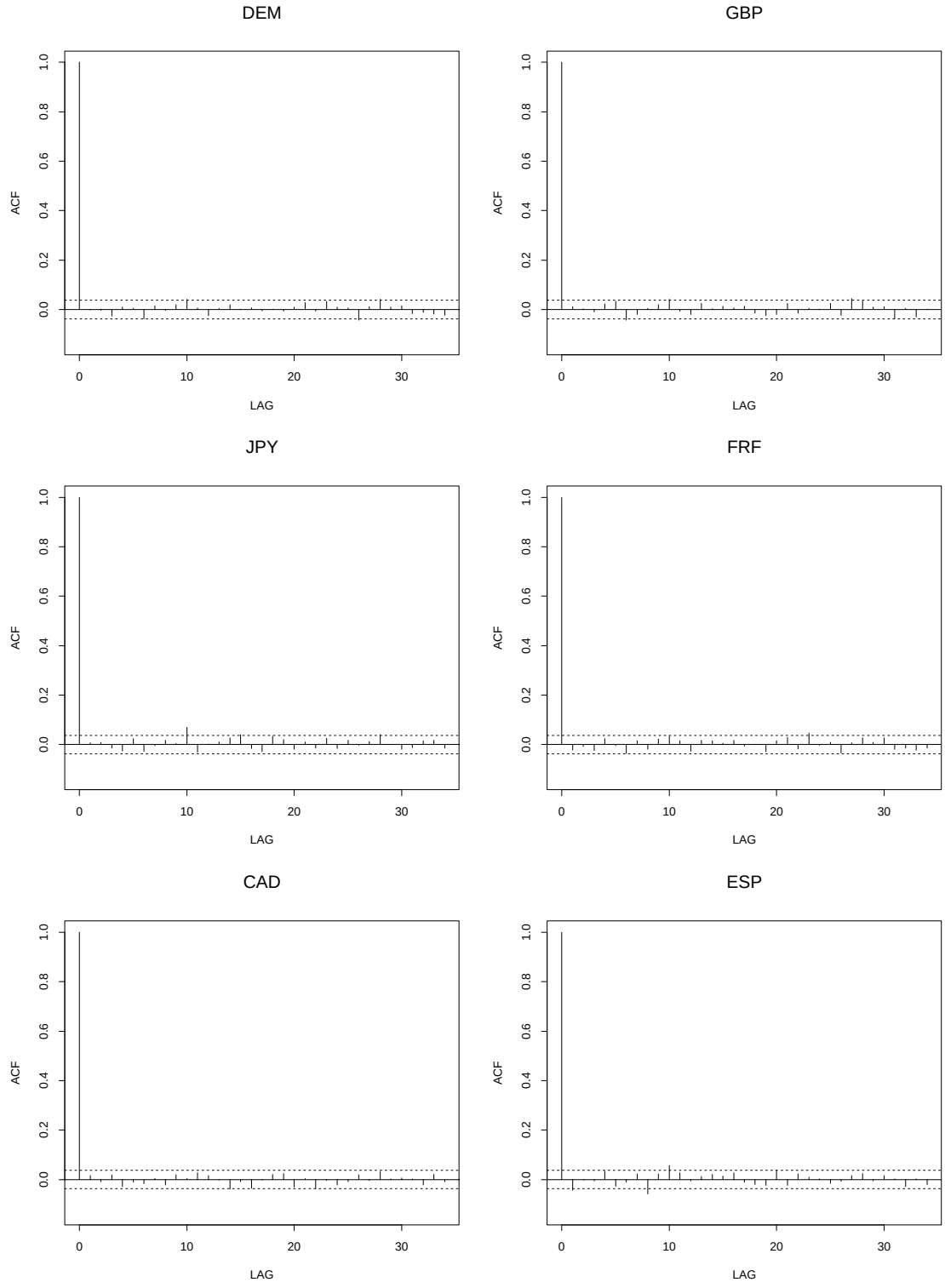


Figure 5.2: Returns autocorrelograms based on data from 1/1/1988 to 01/01/1999 (2871 observations).

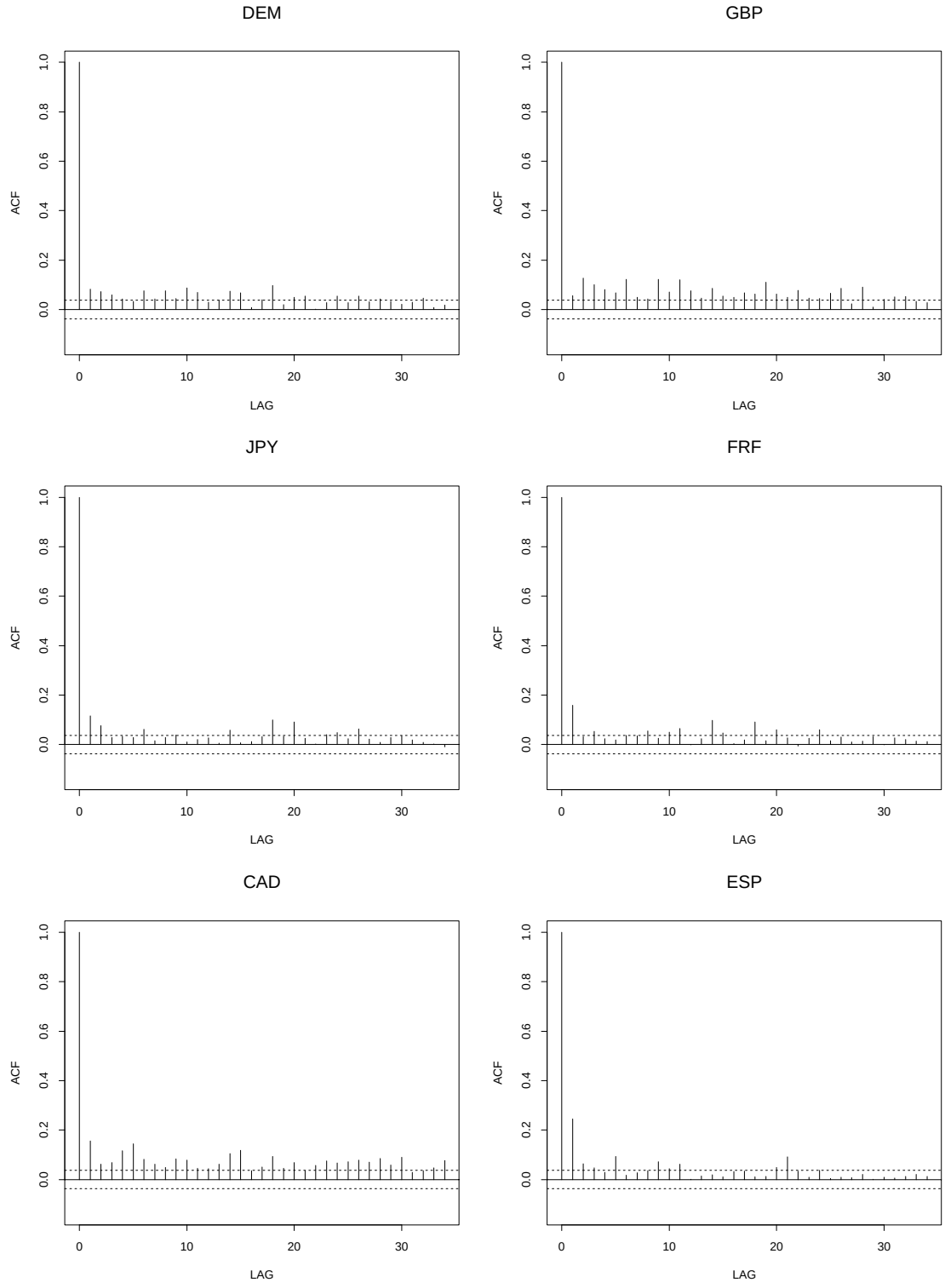


Figure 5.3: Squared returns autocorrelograms based on data from 1/1/1988 to 01/01/1999 (2871 observations).

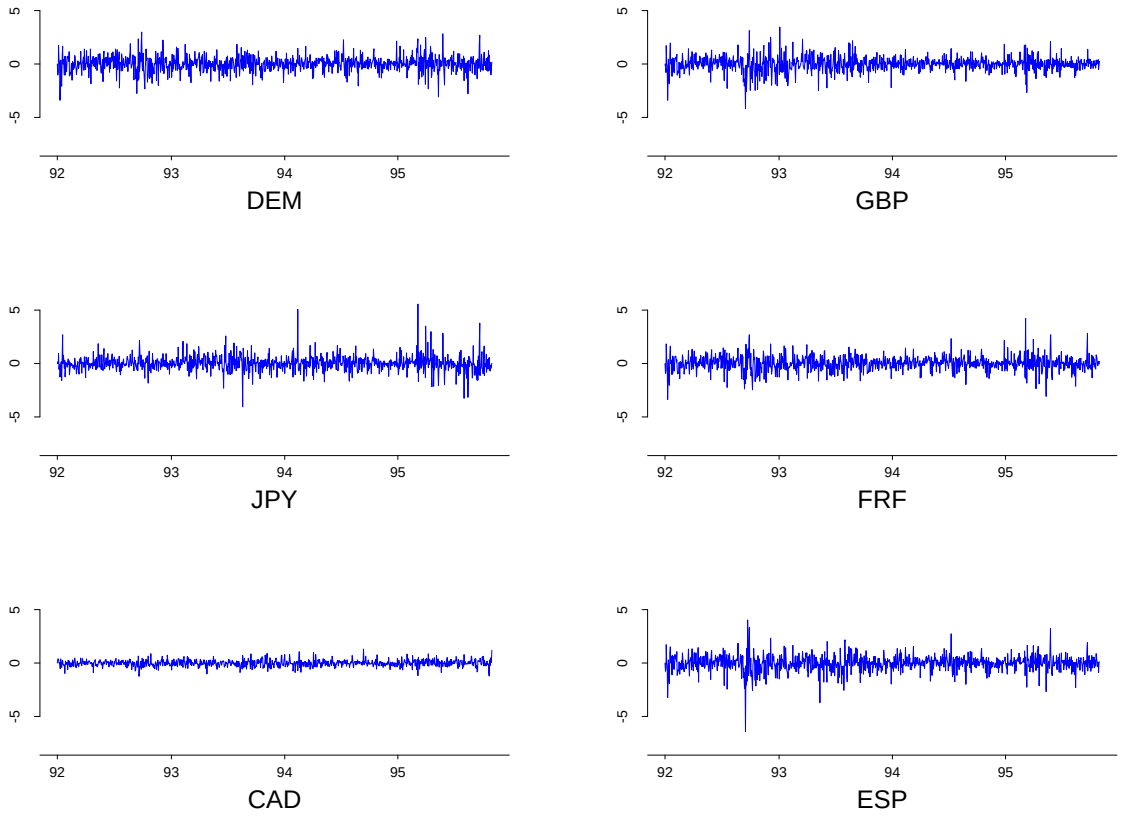


Figure 5.4: Returns from 1/1/1992 to 10/31/1995 (1000 observations).

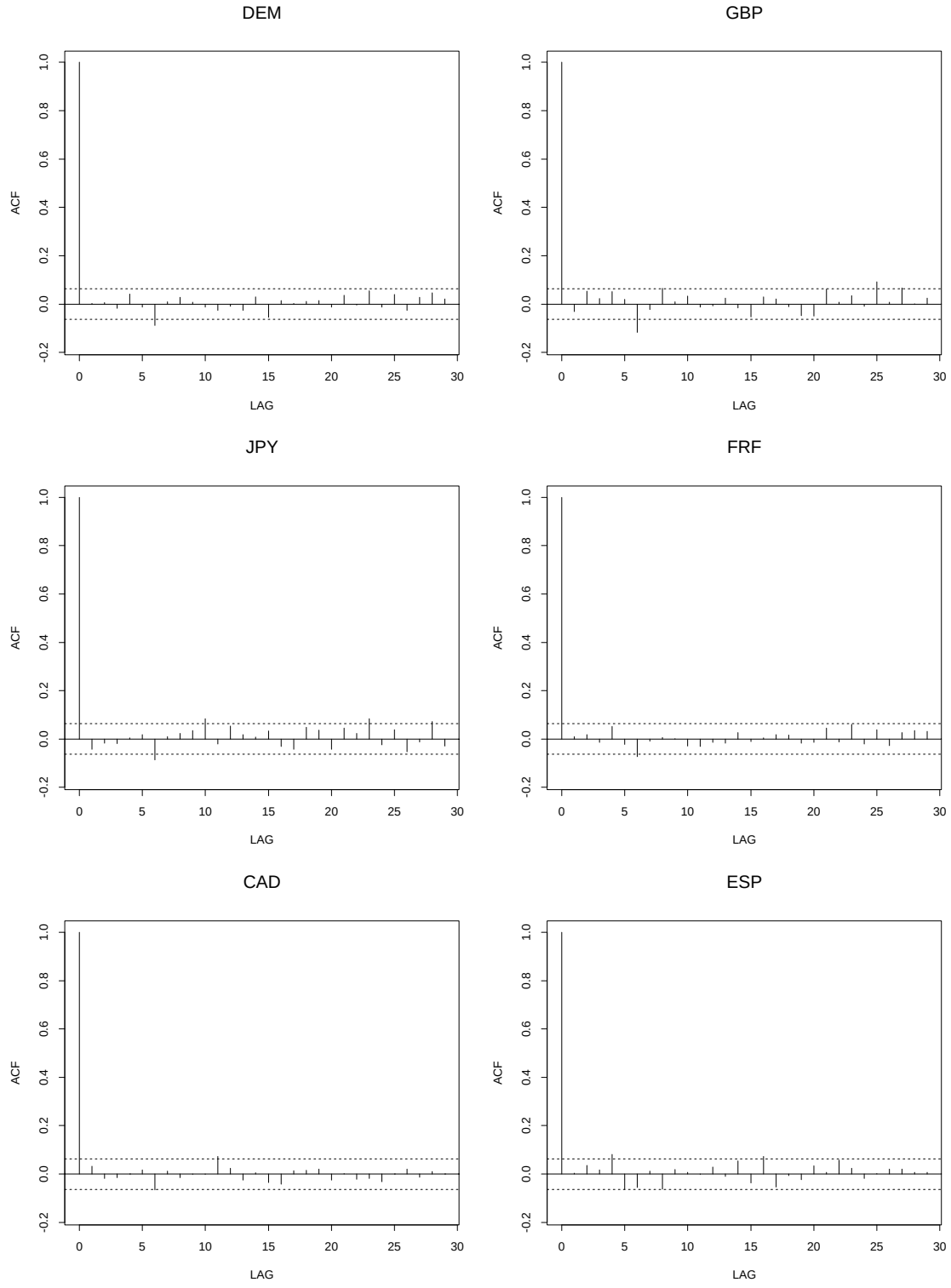


Figure 5.5: Returns autocorrelograms based on data from 1/1/1992 to 10/31/1995 (1000 observations).

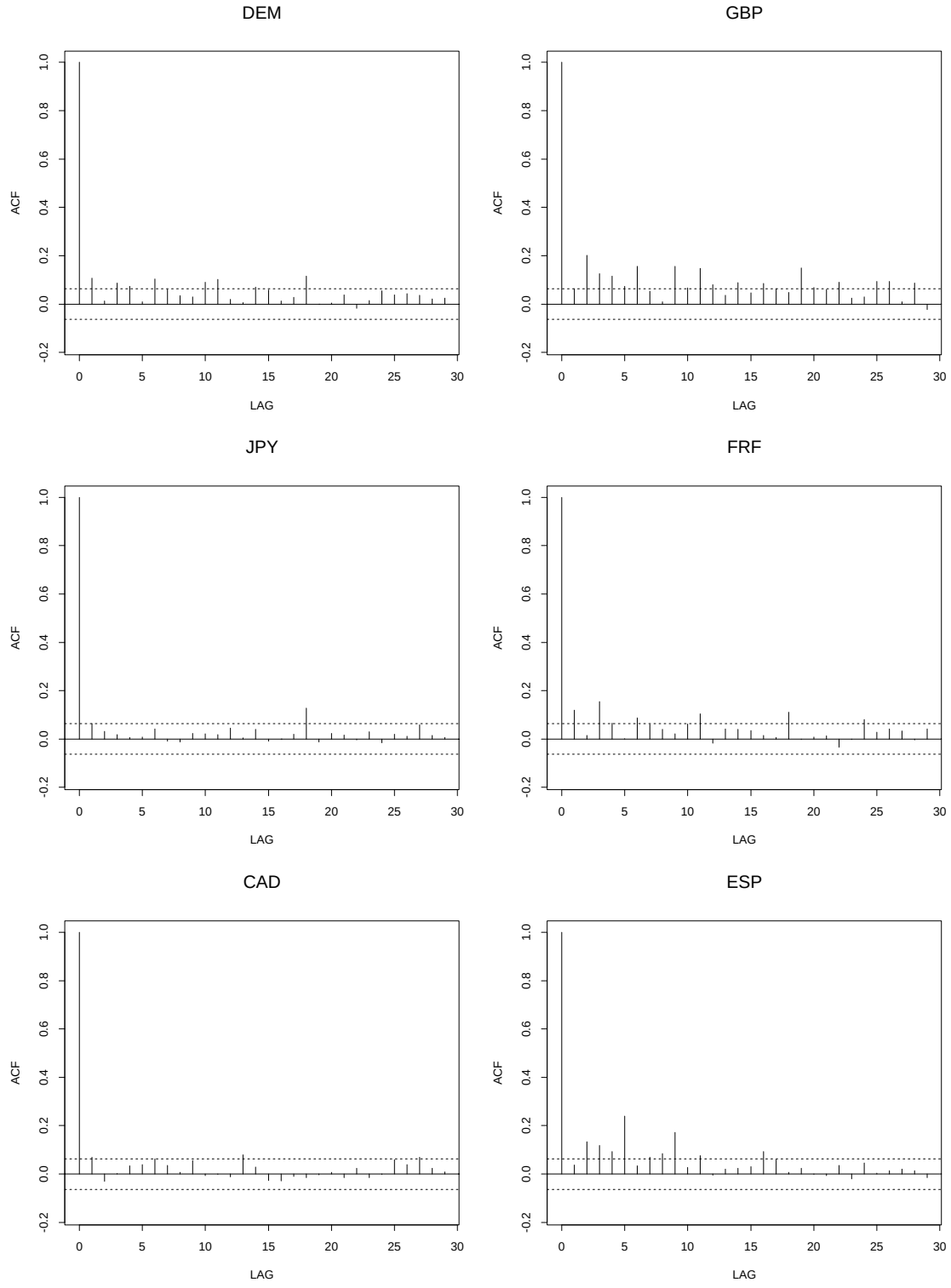


Figure 5.6: Squared returns autocorrelograms based on data from 1/1/1992 to 10/31/1995 (1000 observations).

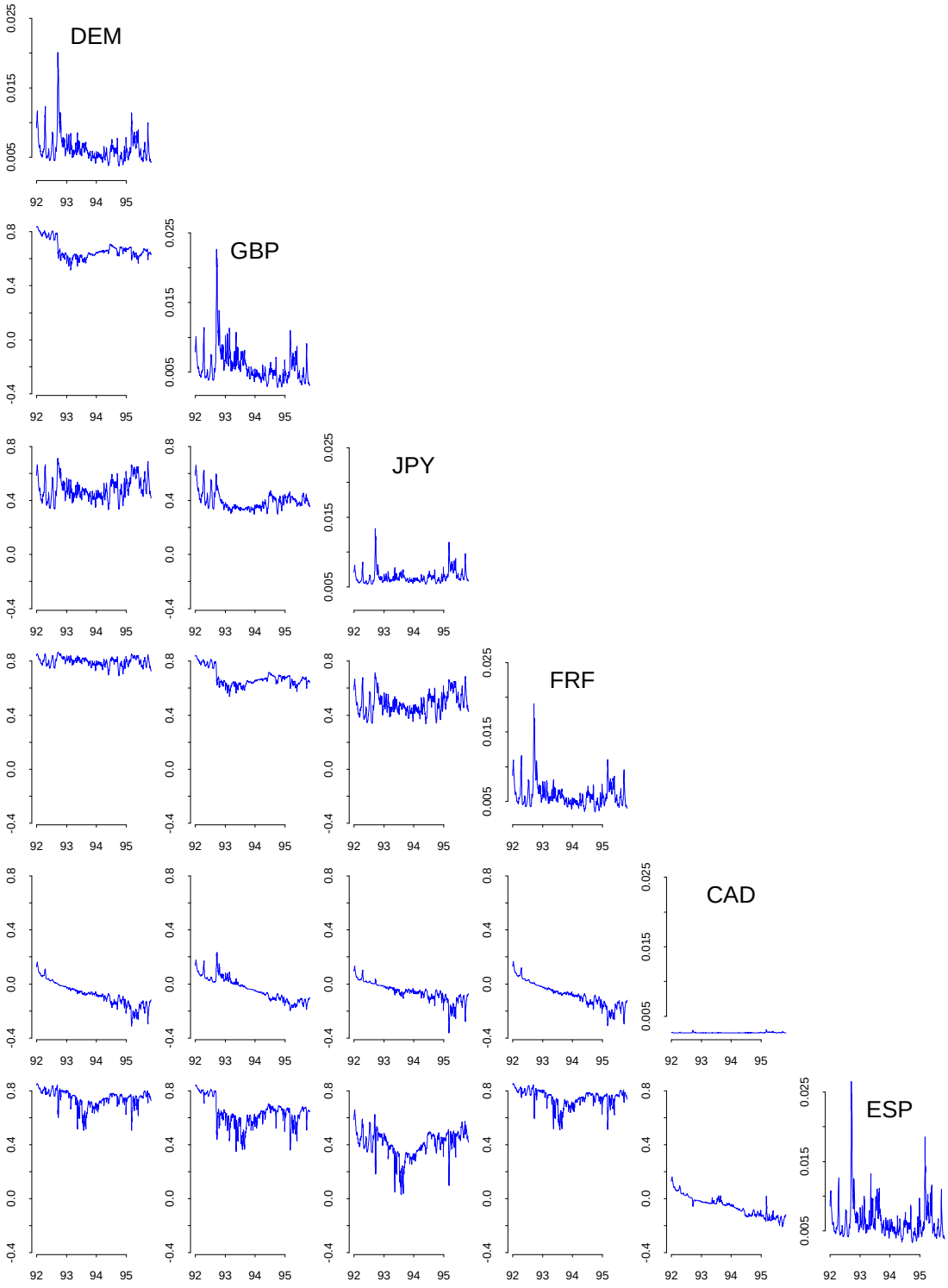


Figure 5.7: Three factor model: time series codependence structure from 1/1/1992 to 10/31/1995. The diagonal elements are the time series standard deviations and the lower diagonal elements are the correlations.

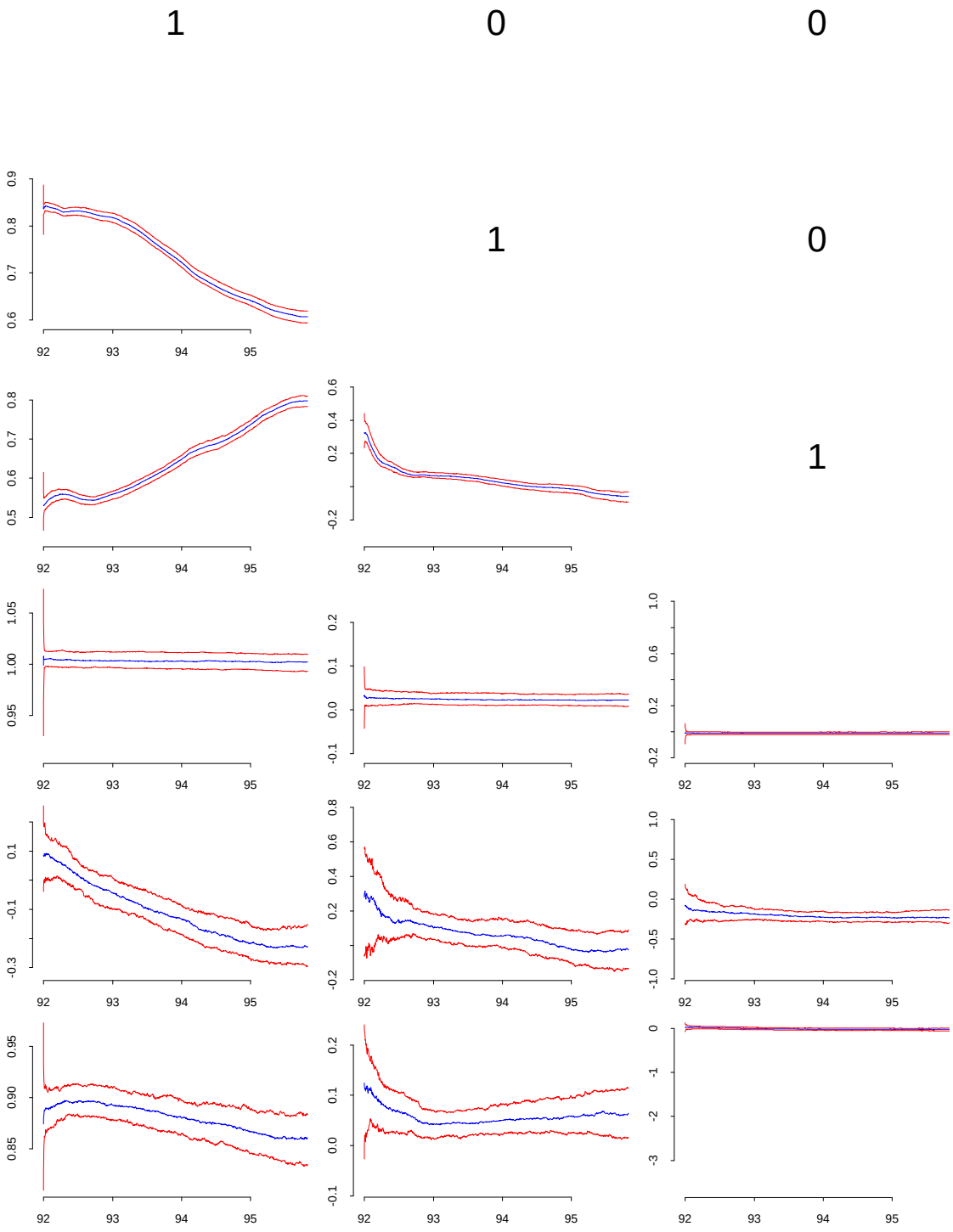


Figure 5.8: Three factor model: retrospective posterior means for the unconstrained elements of $\boldsymbol{\beta}_t$ (1/1/1992 to 10/31/1995).

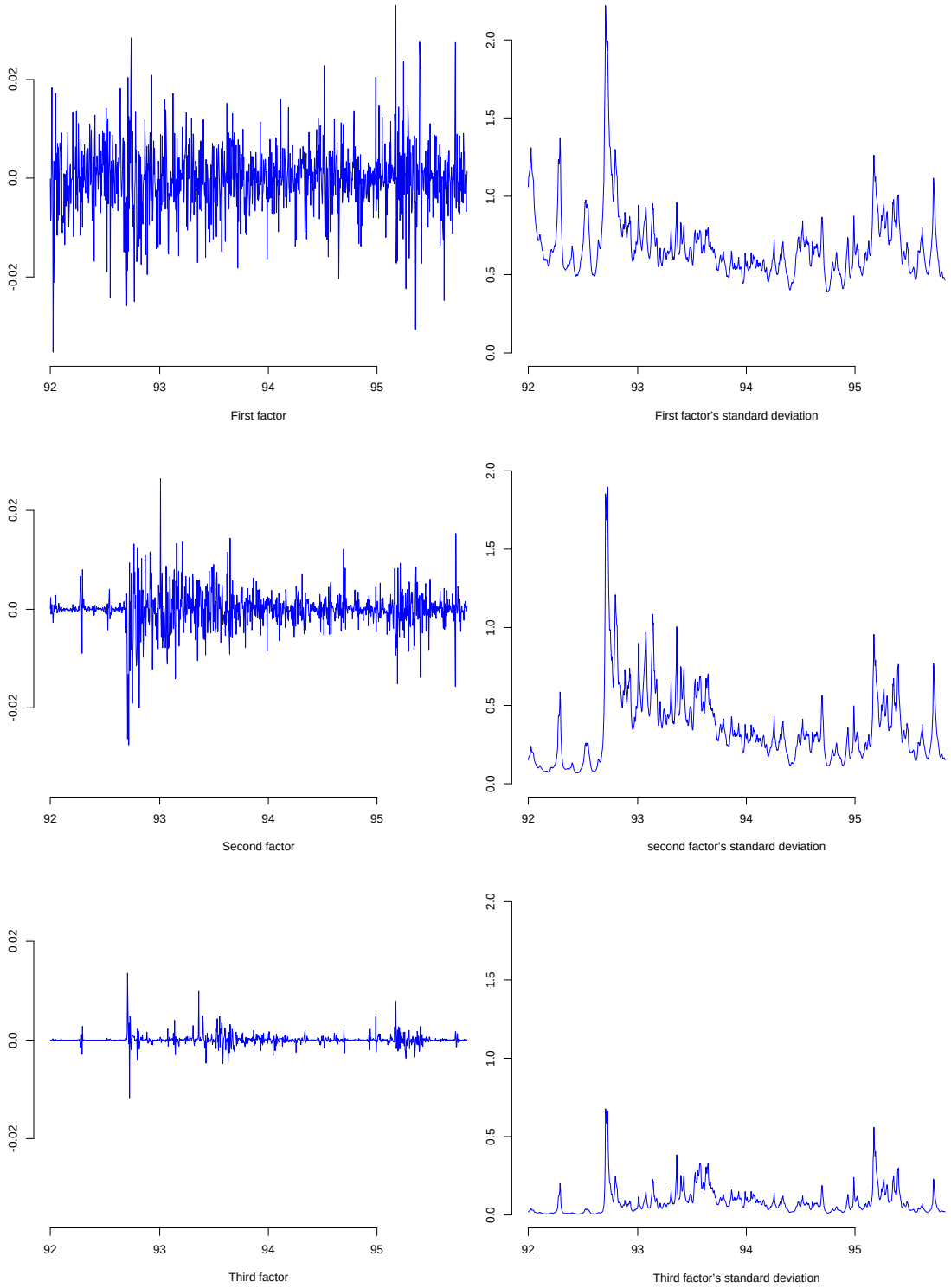


Figure 5.9: Three factor model: retrospective posterior means for the common factors, $\boldsymbol{f}_t$, and their standard deviations ($\times 100$), the square root of the diagonal elements of $\boldsymbol{H}_t$ (1/1/1992 to 10/31/1995).

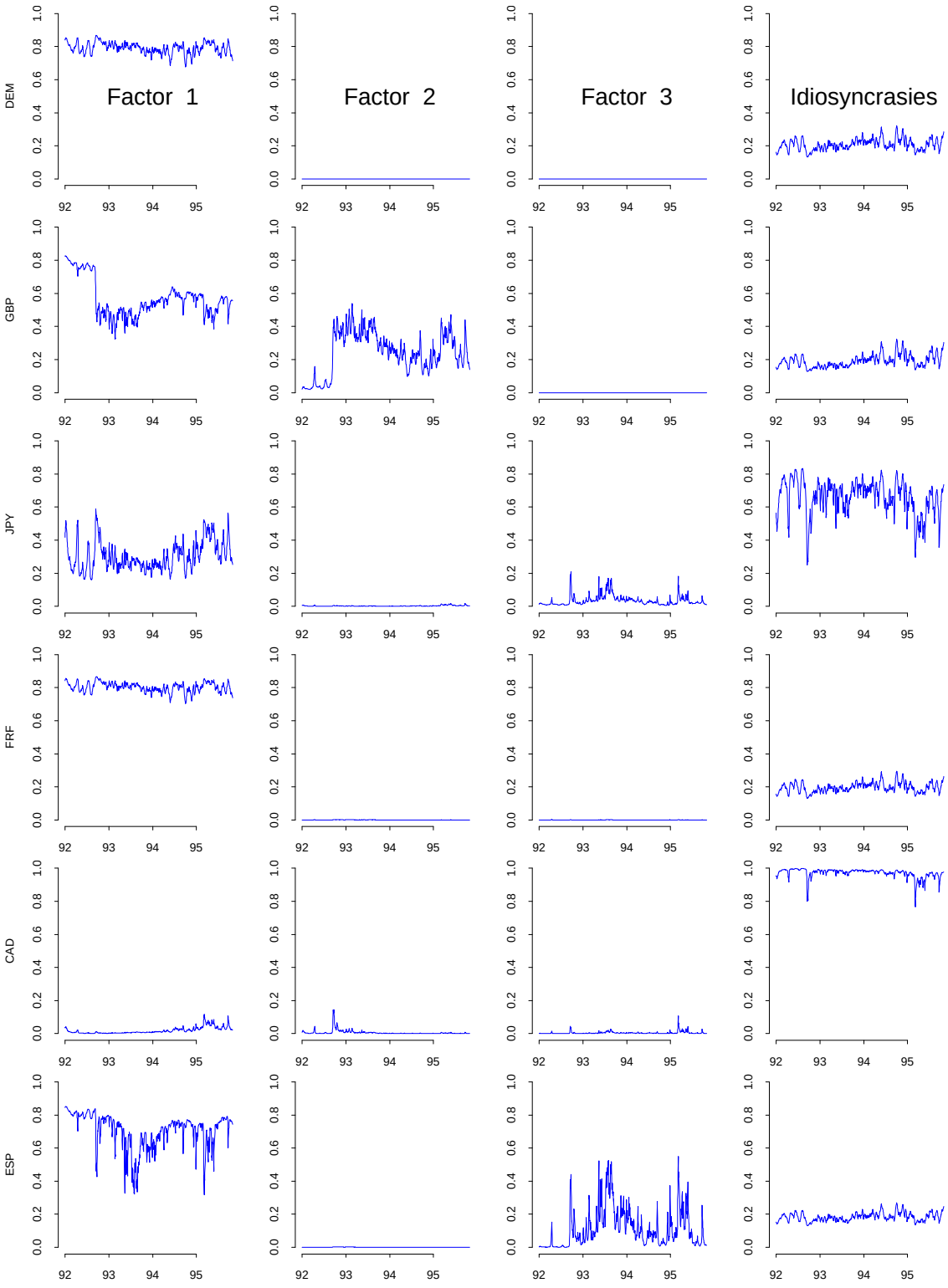


Figure 5.10: Three factor model: proportion of the time series variances explained by each of the factors (common and specific), from 1/1/1992 to 10/ 31/ 1995.

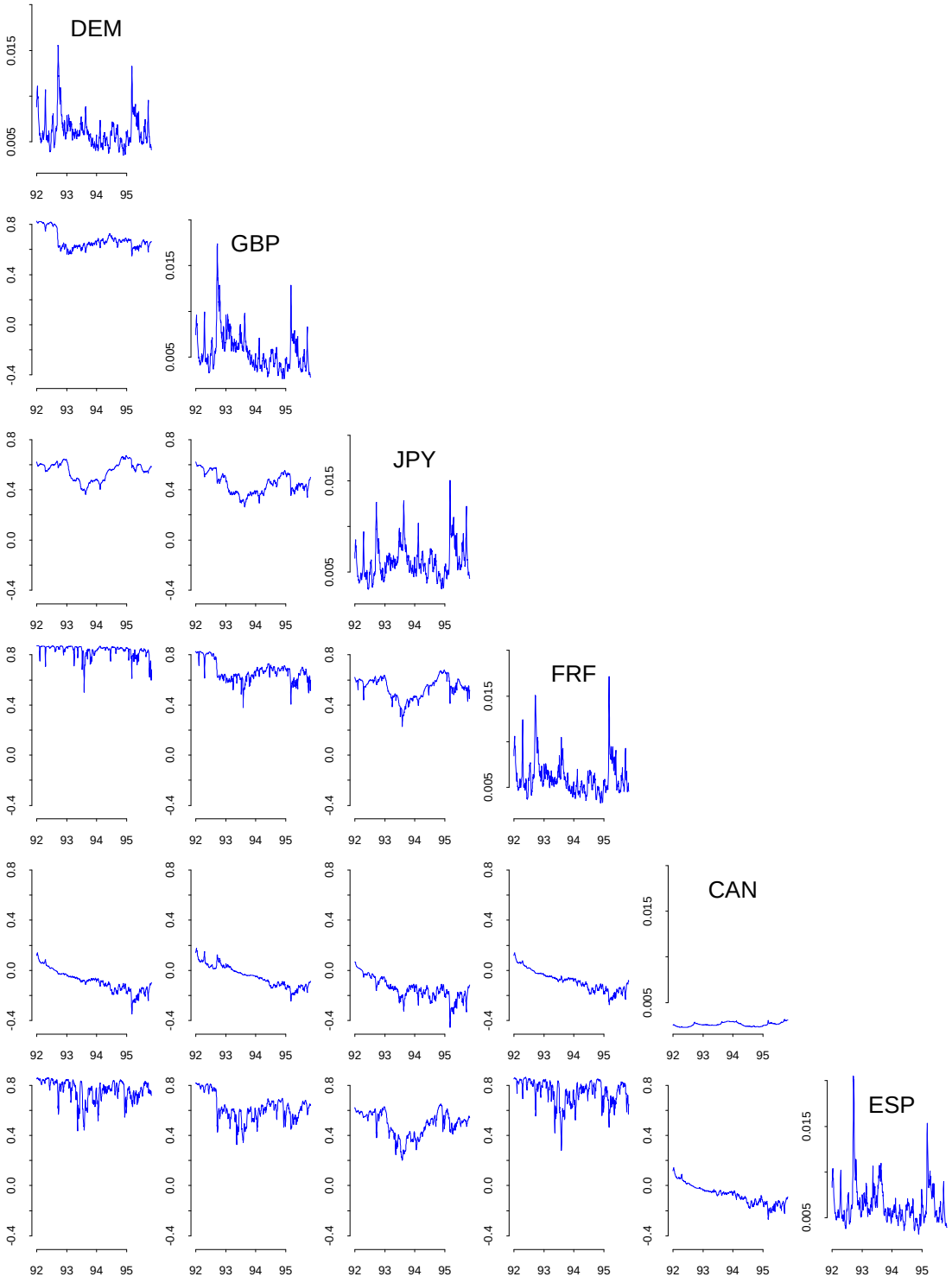


Figure 5.11: Time series codependence structure from 1/1/1992 to 10/31/1995. The diagonal elements are the time series standard deviations and the lower diagonal elements are the correlations, when the idiosyncrasies have stochastic volatility structures.

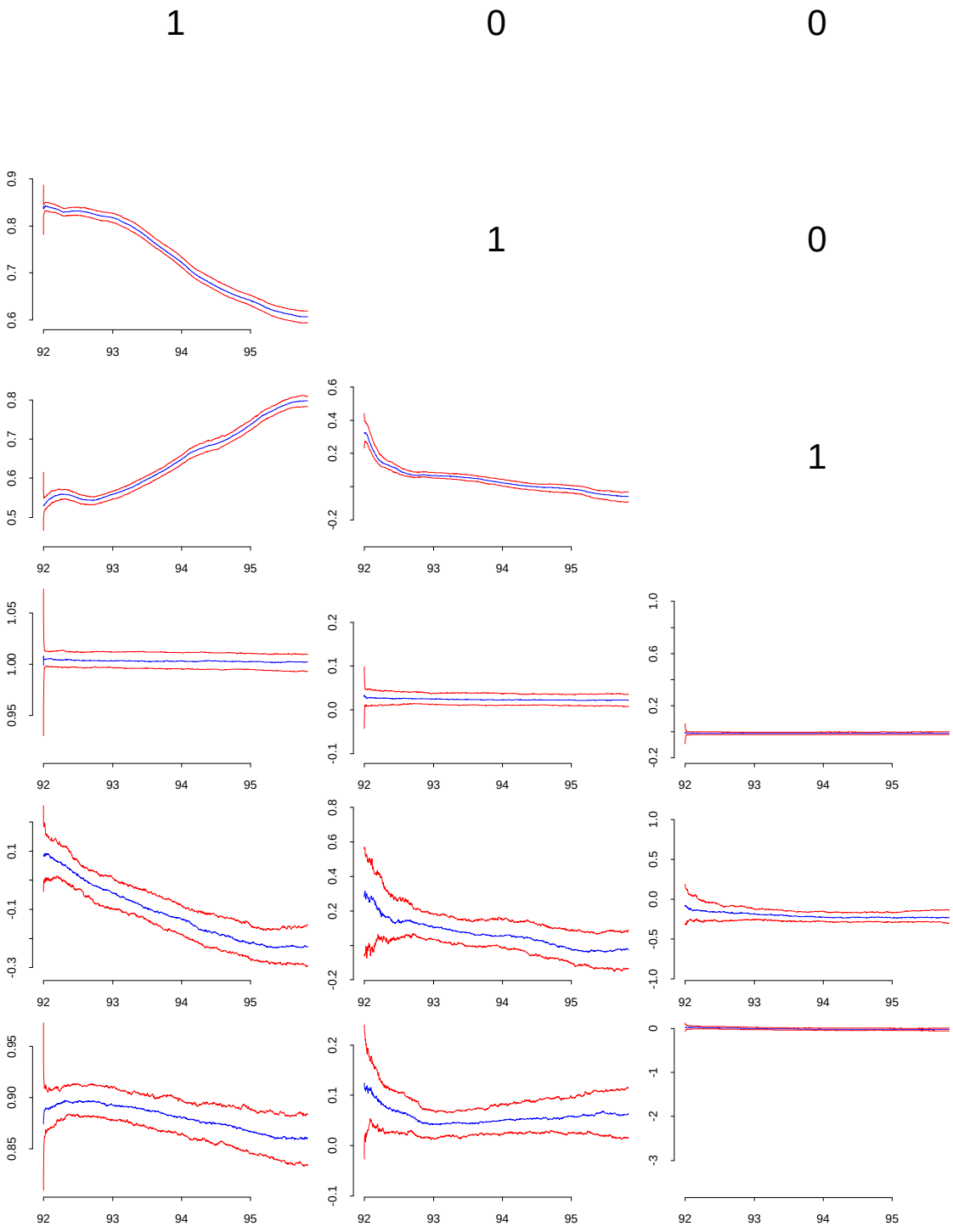


Figure 5.12: Retrospective posterior means for the unconstrained elements of $\boldsymbol{\beta}_t$ when the idiosyncrasies have stochastic volatility structures (1/1/1992 to 10/31/1995).

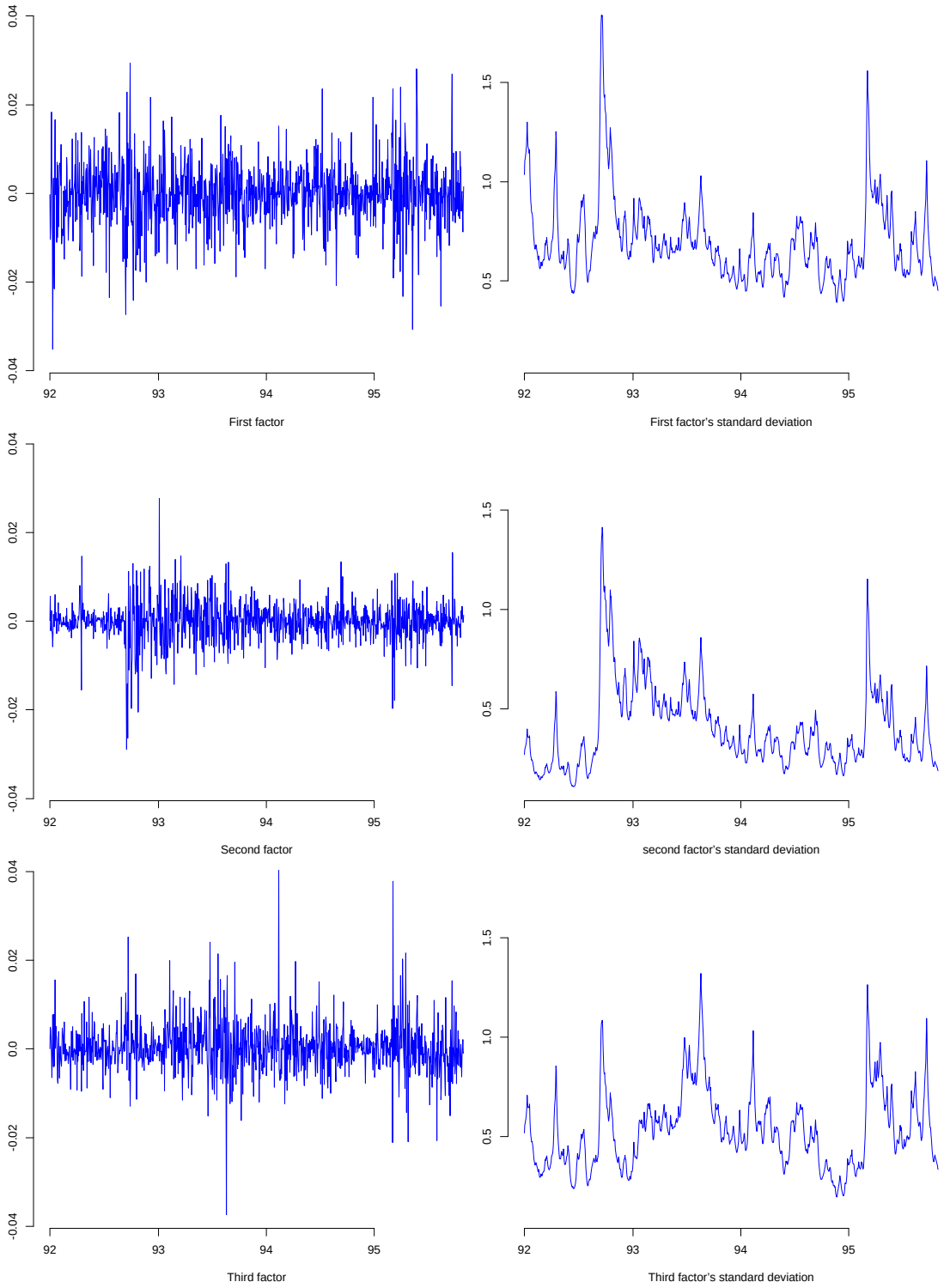


Figure 5.13: Retrospective posterior means for the common factors, $\boldsymbol{f}_t$, and their standard deviations ($\times 100$), the square root of the diagonal elements of $\boldsymbol{H}_t$, when the idiosyncrasies have stochastic volatility structures (1/1/1992 to 10/31/1995).

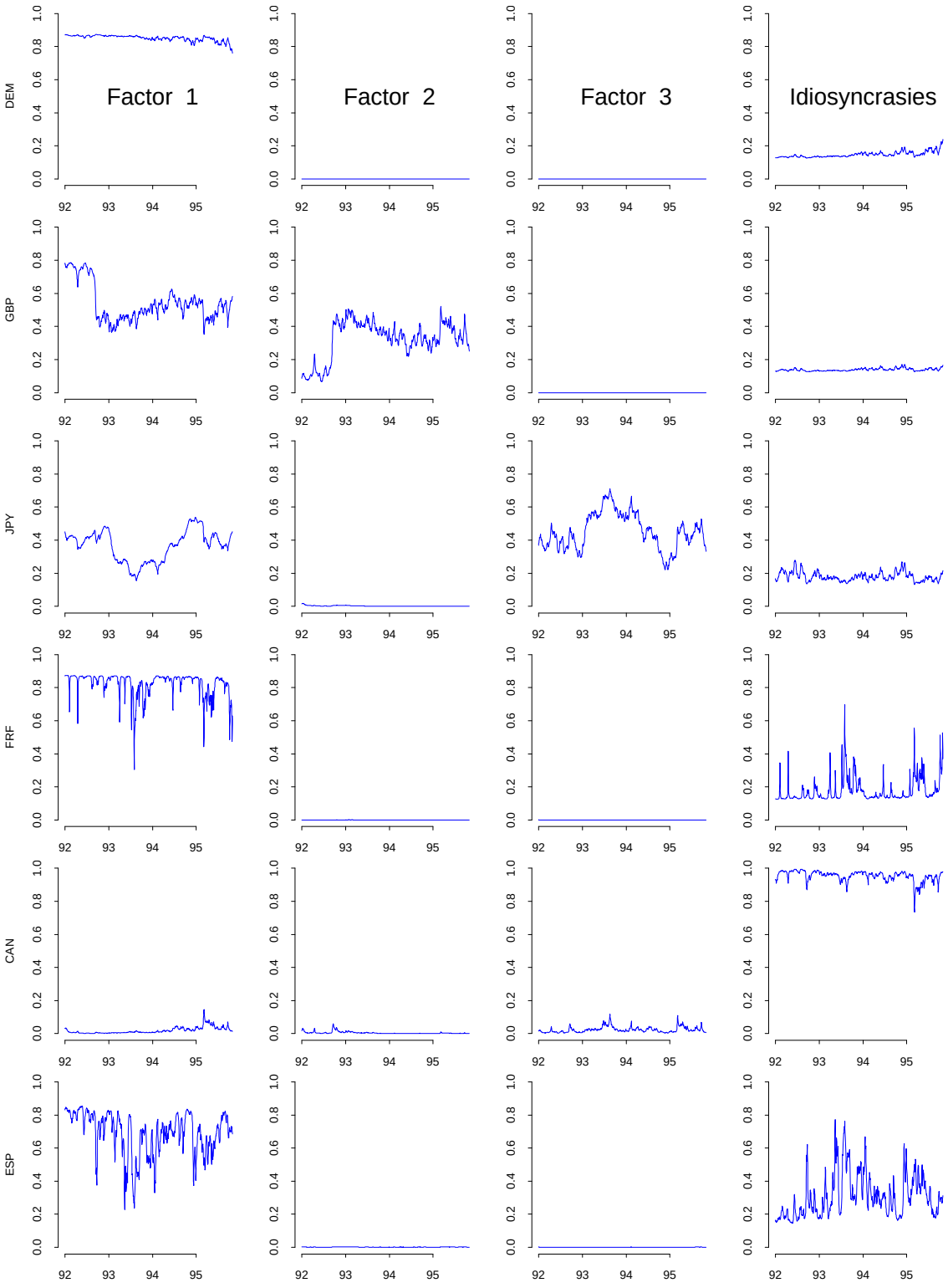


Figure 5.14: Proportion of the time series variances explained by each of the factors (common and specific), when the idiosyncrasies have stochastic volatility structures (1/1/1992 to 10/31/1995).

Chapter 6

Simulation-based sequential analysis

6.1 Introduction

Smoothed estimation of factor models with multivariate stochastic volatilities were extensively explored in the previous chapter. However, in many applications such as financial time series analysis, where data arrive almost continuously, it is imperative to have statistical approaches that allow sequential analysis and forecasting.

Sequential analysis and estimation are old topics amongst researchers, especially engineers, applied economists and statisticians. Kalman filters or dynamic models are particularly famous and widely used within and outside academy. West and Harrison (1997) and Harvey (1989) are two good examples of the extent of attention that sequential methods have taken in time series analysis. The field has been extensively investigated, especially when dealing with normal linear dynamic models.

However, when it comes to larger nonlinear and nonnormal dynamic models, there had not, until the 1990s, been sufficient statistical development. Not until recently, with the revival of Monte Carlo techniques, that simulation-based sequential analysis has come to the fore. Simultaneous appearance of related approaches can be tracked back to West (1993a, 1993b) and Gordon *et al.* (1995). Broadly speaking, they

propose approximating and sampling from the posterior distribution of the state space vector at a certain period in time based on discrete random approximations of the prior distribution at previous times. The samples that summarize the state space distribution have been commonly called *particles*. More recent research has explored and generalized in many directions the particle filter algorithms, such as Liu and Chen (1995), Berzuini *et al.* (1997), Kitagawa (1998b), Doucet (1998), Pitt and Shephard (1999a) Liu and West (2000), and Godsill *et al.* (2000), to cite a few.

In this chapter we apply some of the most recent developments in simulation-based sequential analysis for the factor model with multivarite stochastic volatility problem. We improve upon methods in Aguilar and West (2000), and Pitt and Shephard (1999b), by allowing the model parameters to be estimated sequentially, since this is very important when comparing competing models, either by their predictive capacity or by their performance when measured by other mechanisms, such as porfolio returns in financial settings. On one hand, we rely on Pitt and Shephard (1999a)'s auxiliary particle filters ideas to deal with the state parameters in our model structure. On the other hand, the fixed parameters are sampled sequentially according to a neat and new idea developed in Liu and West (2000), following previous work from West (1993a, 1993b).

We start the chapter with a review of the general notation for dynamic models in Section 6.2, followed by the presentation of the auxiliary particle filter in Section 6.3. Section 6.4 treats the fixed parameters and describe the general sequential algorithm. Predictive analysis is rapidly presented in Section 6.5, while Section 6.6 specializes the sequential procedure for the factor analysis with multivariate stochastic volatility context.

6.2 General notation

We adopt the following general notation for a Markovian dynamic model:

$$\text{Observation Equation} \quad : \quad p(\boldsymbol{y}_t|\boldsymbol{x}_t, \boldsymbol{\theta}) \tag{6.1}$$

$$\text{Evolution Equation} \quad : \quad p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}, \boldsymbol{\theta}). \tag{6.2}$$

where $\boldsymbol{y}_t$ is the vector of observable time series, $\boldsymbol{x}_t$ is the vector of unobservable state parameters and $\boldsymbol{\theta}$ is the vector of static parameters, sometimes referred to as the *hyperparameters*.

For the moment suppose that $\boldsymbol{\theta}$ is known and omitted from the notation. Later on we will show how to include $\boldsymbol{\theta}$ in the analysis. Therefore, and using the terminology of West and Harrison (1997), the *evolution equation* at each time t is:

$$p(\boldsymbol{x}_t|D_{t-1}) = \int p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})p(\boldsymbol{x}_{t-1}|D_{t-1})d\boldsymbol{x}_{t-1} \tag{6.3}$$

while the *updating equation* at each time t is

$$p(\boldsymbol{x}_t|D_t) \propto p(\boldsymbol{x}_t|D_{t-1})p(\boldsymbol{y}_t|\boldsymbol{x}_t). \tag{6.4}$$

Apart from the traditional Gaussian case, extensively studied in West and Harrison (1997), the filtering densities cannot be obtained analytically and numerical methods must be used. There are many alternative methods proposed in the literature for approximating these densities. We will focus our attention to recent developments on auxiliary particle filtering by Pitt and Shephard (1999a) combined with an ingenious algorithm proposed by Liu and West (2000) that takes into account uncertainty driven by fixed parameters, such as $\boldsymbol{\theta} = \{\boldsymbol{\Sigma}, \boldsymbol{\alpha}, \boldsymbol{\phi}\}$ in our case.

6.3 Auxiliary particle filtering

Here we will describe the rationale behind the auxiliary particle filtering we will use later on. Detailed explanation and applications can be found in Pitt and Shephard

(1999a). First, here we are at time $t-1$. We want to use equations (6.3) and (6.4) to update to time t posterior. As mentioned before, apart from very simple cases, neither the prior in (6.3) nor the posterior in (6.4) are analytically tractable. Instead, we will assume that a weighted sample $\boldsymbol{x}_{t-1}^{(1)}, \ldots, \boldsymbol{x}_{t-1}^{(M)}$ with weights $w_{t-1}^{(1)}, \ldots, w_{t-1}^{(M)}$ can be taken from the states posterior distribution at time $t-1$, $p(\boldsymbol{x}_{t-1}|D_{t-1})$. In other words, the distribution of the state space vector $\boldsymbol{x}_t$ is approximated discrete *particles* with discrete probability mass. We will use the set $\Xi_{t-1} = \{(\boldsymbol{x}_{t-1}, w_{t-1})^{(j)},\ j = 1, \ldots M\}$ to summarize $p(\boldsymbol{x}_{t-1}|D_{t-1})$. Hence, a natural Monte Carlo approximation (as $M \to \infty$) for the prior (6.3) is

$$\hat{p}(\boldsymbol{x}_t|D_{t-1}) = \sum_{j=1}^{M} p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}^{(j)}) w_{t-1}^{(j)} \tag{6.5}$$

which, following Pitt and Shephard's (1999a) terminology, is called the *empirical prediction density*. Combining this approximate prior with the observation equation produces, by Bayes' theorem, the following approximation for the state space vector posterior distribution at time t.

$$\begin{aligned} \hat{p}(\boldsymbol{x}_t|D_t) &\propto p(\boldsymbol{y}_t|\boldsymbol{x}_t) \sum_{j=1}^{M} p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}^{(j)}) w_{t-1}^{(j)} \\ &\propto \sum_{j=1}^{M} p(\boldsymbol{y}_t|\boldsymbol{x}_t) p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}^{(j)}) w_{t-1}^{(j)}. \end{aligned} \tag{6.6}$$

the *empirical filtering density* according to Pitt and Shephard (1999a). A sampling scheme from the approximate posterior (6.6) is needed in order to complete the evolution/update cycle. More specifically, the empirical filtering density should be summarized by a set of particles and corresponding probability masses. Gordon *et al.* (1995), Kitagawa (1998a) and Berzuini *et al.* (1997), for instance, suggest using *sampling/importance resampling* (SIR) where samples are drawn from the prior

$\hat{p}(\boldsymbol{x}_t|D_{t-1})$ and reweighted according to the likelihood $p(\boldsymbol{y}_t|\boldsymbol{x}_t)$. As well-known the SIR method becomes ineffective either when the prior is relatively diffuse or the likelihood is highly informative. Adaptive SIR, rejection and MCMC sampling methods are possible alternatives (Pitt and Shephard, 1999a, Section 2.2).

Pitt and Shephard improve on particle filter methods by addressing to practical and importante issues: *(i)* Efficiently sampling from the approximate posterior distribution (equation 6.6) and *(ii)* Efficiently approximating tails' behavior of the approximate prior (equation 6.5). They developed a generic filtering algorithm that is currently well known as *auxiliary particle filtering*. The basic feature of a auxiliary particle filter is to take advantage of the mixture of densities (6.6) to obtain draws from $p(\boldsymbol{x}_t|D_t)$ by introducing latent indicator variables to identify the terms in the mixture (an idea commonly used in mixture modeling, Diebolt and Robert, 1994). In other words, if $(\boldsymbol{x}_t, k)$ is sampled from the following joint density

$$p(\boldsymbol{x}_t, k) \propto p(\boldsymbol{y}_t|\boldsymbol{x}_t)p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}^{(k)})w_{t-1}^{(k)} \tag{6.7}$$

and the index discarded, the resulting $\boldsymbol{x}_t$ is a sample from (6.6). Pitt and Shephard describe various sampling methods. In this work we focus our attention to their SIR algorithm, which can be summarized in the following steps:

Step 1: *Sample* $\{(\boldsymbol{x}_t, w_t)^{(l)},\ l = 1, \dots L\}$ *from an importance function, say* $g(\boldsymbol{x}_t, k|D_t)$*;*

Step 2: *Compute the weights*

$$w_t^{(l)} \propto \frac{p(\boldsymbol{y}_t|\boldsymbol{x}_t^{(l)})p(\boldsymbol{x}_t^{(l)}|\boldsymbol{x}_{t-1}^{(k^l)})}{g(\boldsymbol{x}_t^{(l)}, k^l|D_t)}, \qquad l = 1, \dots, L. \tag{6.8}$$

Therefore, $\boldsymbol{\Xi}_t = \{(\boldsymbol{x}_t, w_t)^{(l)},\ l = 1, \dots L\}$ summarizes $p(\boldsymbol{x}_t|D_t)$. A third step can be included depending on whether or not an iid sample is desired, in which case L

is usually much larger than M. In our applications we use $M = L = 10000$ draws to compute posterior and predictive quantities.

Pitt and Shephard (1999a) suggest using the following generic importance function,g:

$$g(\boldsymbol{x}_t, k|D_t) \propto p(\boldsymbol{y}_t|\boldsymbol{\mu}_t^{(k)})p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}^{(k)})w_{t-1}^{(k)}, \qquad k = 1, \ldots, M. \tag{6.9}$$

where $\boldsymbol{\mu}_t^{(k)}$ is an estimate of $\boldsymbol{x}_t$ given $\boldsymbol{x}_{t-1}^{(k)}$. In our applications, $\boldsymbol{\mu}_t = E(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})$, even though modes or any other likely value from $p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})$ would be as convenient. Therefore, the form of $g(k|D_t)$ is

$$\begin{aligned} g(k|D_t) &\propto w_{t-1}^{(k)} \int p(\boldsymbol{y}_t|\boldsymbol{\mu}_t^{(k)})p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}^{(k)})d\boldsymbol{x}_t \\ &\propto p(\boldsymbol{y}_t|\boldsymbol{\mu}_t^{(k)})w_{t-1}^{(k)}. \end{aligned} \tag{6.10}$$

The previous two-step algorithm can then be rewritten as:

Step 1a: *Sample k^l with probability proportional to $g(k|D_t) \propto p(\boldsymbol{y}_t|\boldsymbol{\mu}_t^{(k)})w_{t-1}^{(k)}$;*

Step 1b: *Sample $\boldsymbol{x}_t^{(l)}$ from $p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}^{(k^l)})$;*

Step 2: *Compute the weights*

$$w_t^{(l)} \propto \frac{p(\boldsymbol{y}_t|\boldsymbol{x}_t^{(l)})}{p(\boldsymbol{y}_t|\boldsymbol{\mu}_t^{(k^l)})}, \qquad l = 1, \ldots, L. \tag{6.11}$$

Choosing $g(\cdot)$ is a nontrivial task and is inherent in all Monte Carlo methods (SIR, adaptive,MCMC,etc.). Pitt and Shephard's wise argument is that the simulation algorithm will favour particles with larger predictive likelihoods. By doing so, they continue, the resampling step will have lower computational cost and will improve on statistical efficiency of the procedure.

In the next section we incorporate the parameter vector, $\boldsymbol{\theta}$ in the filtering analysis. We consider Liu and West'(2000) approach to this problem.

6.4 Treatment of model parameters

In this section we extend the auxiliary particle filter from last section to incorporate uncertainty about the fixed parameter vector, $\boldsymbol{\theta}$. This and the next sections are based on Liu and West (2000). They combine kernel density estimation techniques with artificial parameter evolution and propose a novel algorithm to sequentially treat fixed parameters in general dynamic model settings.

Initially, let (6.4) be rewritten as

$$p(\boldsymbol{x}_t, \boldsymbol{\theta}|D_t) \propto p(\boldsymbol{y}_t|\boldsymbol{x}_t, \boldsymbol{\theta})p(\boldsymbol{x}_t|\boldsymbol{\theta}, D_{t-1})p(\boldsymbol{\theta}|D_{t-1}) \tag{6.12}$$

where the uncertainty about $\boldsymbol{\theta}$ is explicited by adding the term $p(\boldsymbol{\theta}|D_{t-1})$. As before, and conditional on $\boldsymbol{\theta}$, the evolution density $p(\boldsymbol{x}_t|\boldsymbol{\theta}, D_{t-1})$ can be approximated by

$$\hat{p}(\boldsymbol{x}_t|\boldsymbol{\theta}, D_{t-1}) = \sum_{j=1}^{M} p(\boldsymbol{x}_t|\boldsymbol{\theta}, \boldsymbol{x}_{t-1}^{(j)})w_{t-1}^{(j)} \tag{6.13}$$

where $\boldsymbol{\Xi}_{t-1} = \{(\boldsymbol{x}_{t-1}, \boldsymbol{\theta}_{t-1}, w_{t-1})^{(j)},\ j = 1, \ldots, M\}$ summarizes $p(\boldsymbol{x}_{t-1}, \boldsymbol{\theta}|D_{t-1})$.

A natural solution, firstly explored by Gordon *et al.* (1995), is to pretend that the fixed parameters are states in the dynamic modeling, for instance, by adding small random disturbances to artificial evolutions, and proceed the analysis with the auxiliary particle filters presented in the previous section. Such artificial evolution reduces the sample degeneracy problems, however it imputes unnecessary uncertainty into the model and also creates artificial loss of information resulting on overdispersion of the posterior distributions.

Liu and West (2000) reinterpret Gordon *et al.*'s artificial parameter evolution idea and combine it with West's kernel smoothing techniques. Approximations for $p(\boldsymbol{\theta}|D_t)$ based on mixtures of multivariate normals were suggested by West (1993b),

$$\hat{p}(\boldsymbol{\theta}|D_{t-1}) = \sum_{j=1}^{M} N(\boldsymbol{\theta}|\boldsymbol{m}_{t-1}^{(j)}, h^2\boldsymbol{V}_{t-1})w_{t-1}^{(j)} \tag{6.14}$$

where h is a smoothing parameter, $\boldsymbol{V}_{t-1} = Var(\boldsymbol{\theta}|D_{t-1})$, and $\boldsymbol{\mu}_{t-1}^{(j)}$ are the locations of the components of the mixture. In standard kernel methods, $\boldsymbol{\mu}_{t-1}^{(j)} = \boldsymbol{\theta}_{t-1}^{(j)}$. Also, for large M, it is also common practice h as a decreasing function of M. West (1993b) introduces a shrinkage rule for the locations,

$$\boldsymbol{m}_{t-1}^{(j)} = a\boldsymbol{\theta}_{t-1}^{(j)} + (1-a)\overline{\boldsymbol{\theta}}_{t-1} \tag{6.15}$$

where $\overline{\boldsymbol{\theta}}_{t-1} = E(\boldsymbol{\theta}_{t-1}|D_{t-1})$. The variance of the resulting mixture of normals is $(a^2+h^2)\boldsymbol{V}_{t-1}$, which is always larger then $\boldsymbol{V}_{t-1}$ for $a^2+h^2 > 1$. West suggests using $a^2 = 1-h^2$ to guarantee that the correct variance is used in the approximation, crucial in sequential schemes.

Liu and West (2000) show that if δ is the discount factor used in Gordon *et al.*'s artificial evolution method, then defining

$$h^2 = 1 - ((3\delta - 1)/2\delta)^2 \tag{6.16}$$

produces an algorithm that links the kernel estimation (equation 6.13) with the shrinkage idea (equation 6.15).

Liu and West (2000) apply their novel strategy in a couple of situations, including a simpler version of Aguilar and West's (2000) dynamic factor model for financial time series, which are similar to the models we explored in the last chapter. Amongst other empirical findings, they argue that MCMC methods should be combined with sequential algorithm in real applications. They show that, when performed for longer periods of time the filtering algorithm starts to deteriorate and diverges from the "gold standard" MCMC results.

Combining Pitt and Shephard's (1999a) auxiliary particle filters with Liu and West's (2000) developments on parameter and state estimation can be schematically summarized in the following algorithm. *For* $t = 1, \ldots, T$:

1. $\boldsymbol{\Xi}_{t-1} = \{(\boldsymbol{x}_{t-1}, \boldsymbol{\theta}_{t-1}, w_{t-1})^{(j)},\ j = 1, \ldots, M\}$ *summarizes* $p(\boldsymbol{x}_{t-1}, \boldsymbol{\theta}|D_{t-1})$.

2. *Compute* $\boldsymbol{V}_{t-1} = \sum_{j=1}^{M}(\boldsymbol{\theta}_{t-1}^{(j)} - \overline{\boldsymbol{\theta}}_{t-1})(\boldsymbol{\theta}_{t-1}^{(j)} - \overline{\boldsymbol{\theta}}_{t-1})' w_{t-1}^{(j)}$ *and* $\overline{\boldsymbol{\theta}}_{t-1} = \sum_{j=1}^{M} \boldsymbol{\theta}_{t-1}^{(j)} w_{t-1}^{(j)}$

3. *Compute* $\boldsymbol{\mu}_t^{(j)} = E(\boldsymbol{x}_t | \boldsymbol{x}_{t-1}^{(j)}, \boldsymbol{\theta}^{(j)})$ *and* $\boldsymbol{m}_{t-1}^{(j)} = a\boldsymbol{\theta}_{t-1}^{(j)} + (1-a)\overline{\boldsymbol{\theta}}_{t-1}$

4. *For* $l = 1, \ldots, M$

I. *Sample* k^l *with* $Pr(k^l = j) \propto w_{t-1}^{(j)} p(\boldsymbol{y}_t | \boldsymbol{\mu}_t^{(j)}, \boldsymbol{m}_{t-1}^{(j)})$

II. *Sample* $\boldsymbol{\theta}_t^{(l)}$ *from* $N(\boldsymbol{m}_{t-1}^{(k^l)}; h^2 \boldsymbol{V}_{t-1})$

III. *Sample* $\boldsymbol{x}_t^{(l)}$ *from the* $p(\boldsymbol{x}_t | \boldsymbol{x}_{t-1}^{(k^l)}, \boldsymbol{\theta}_t^{(l)})$

IV. *Calculate* $w_t^{(l)} \propto p(\boldsymbol{y}_t | \boldsymbol{x}_t^{(l)}, \boldsymbol{\theta}_t^{(l)}) / p(\boldsymbol{y}_t | \boldsymbol{\mu}_t^{(k^l)}, \boldsymbol{m}_{t-1}^{(k^l)})$

Looping through I-IV, for each time t, *produces* $\boldsymbol{\Xi}_t = \{(\boldsymbol{x}_t, \boldsymbol{\theta}_t, w_t)^{(j)},\ j = 1, \ldots, M\}$ *summarizing* $p(\boldsymbol{x}_t, \boldsymbol{\theta} | D_t)$.

The next section includes an extra step in the algorithm in order to obtain predictive distributions.

6.5 Computing predictive densities

Suppose, as before, that $\boldsymbol{\Xi}_{t-1} = \{(\boldsymbol{x}_{t-1}, \boldsymbol{\theta}_{t-1}, w_{t-1})^{(j)},\ j = 1, \ldots, M\}$ summarizing $p(\boldsymbol{x}_{t-1}, \boldsymbol{\theta} | D_{t-1})$. Then the one-step predictive density

$$p(\boldsymbol{y}_t | D_{t-1}) = \int p(\boldsymbol{y}_t | \boldsymbol{x}_t, \boldsymbol{\theta}) p(\boldsymbol{x}_t, \boldsymbol{\theta} | D_{t-1}) d\boldsymbol{x}_t$$

can be approximated by

$$\hat{p}(\boldsymbol{y}_t | D_{t-1}) = \frac{1}{L} \sum_{j=1}^{L} p(\boldsymbol{y}_t | \tilde{\boldsymbol{x}}_t^{(l)}, \tilde{\boldsymbol{\theta}}_t^{(l)})$$

where $\{(\tilde{\boldsymbol{x}}_t, \tilde{\boldsymbol{\theta}}_t)^{(l)},\ l = 1, \ldots, L\}$ form a sample from the states and parameters' joint prior distribution at time t, $p(\boldsymbol{x}_t, \boldsymbol{\theta} | D_{t-1})$. *For* $l = 1, \ldots, L$,

1. *Sample an indicator* $k^l = j$ *with probability* $w_t^{(j)}$*;*

2. $\tilde{\boldsymbol{\theta}}_t^{(l)} = \boldsymbol{\theta}_{t-1}^{(k^l)}$*;*

3. *Sample* $\tilde{\boldsymbol{x}}_t^{(l)}$ *from the evolution equation (6.2), conditional on* $\boldsymbol{x}_{t-1}^{(k^l)}$ *and* $\tilde{\boldsymbol{\theta}}_t^{(l)}$*.*

Predictive moments can be approximated similarly by

$$\hat{E}(\boldsymbol{y}_t|D_{t-1}) = \frac{1}{L}\sum_{l=1}^{L} E(\boldsymbol{y}_t|\tilde{\boldsymbol{x}}_t^{(l)}, \tilde{\boldsymbol{\theta}}_t^{(l)}) \tag{6.17}$$

and

$$\hat{V}(\boldsymbol{y}_t|D_{t-1}) = \frac{1}{L}\sum_{l=1}^{L} V(\boldsymbol{y}_t|\tilde{\boldsymbol{x}}_t^{(l)}, \tilde{\boldsymbol{\theta}}_t^{(l)}). \tag{6.18}$$

These quantities will be used in the next chapter when comparing models based on their forecasts and porfolio performances.

6.6 Sequential analysis with stochastic volatility and time dependent loadings

In this section we specialize the previous algorithms to our factor stochastic volatility model. The state vector $\boldsymbol{x}_{t-1}$ now comprises $\tilde{\boldsymbol{\beta}}_{t-1}, \boldsymbol{\lambda}_{t-1}, \boldsymbol{\gamma}_{t-1}$ and $\boldsymbol{\eta}_{t-1}$, according to equations 5.6, 5.4, 5.2, and 5.5, respectively. It has, therefore, $p_1 = d + k + 2m$ elements, where $d = mk - k(k+1)/2$. The parameter vector $\boldsymbol{\theta}$ comprises elements $\boldsymbol{\alpha}, \boldsymbol{\phi}, \tilde{\boldsymbol{\alpha}}, \boldsymbol{\rho}, vech(\boldsymbol{U}), \boldsymbol{s}, \boldsymbol{\zeta}$ and $\boldsymbol{\Delta}$, following the same equations as for the state vectors. In our applications we have $\boldsymbol{\zeta} = \mathbf{0}$ and $\boldsymbol{\Delta} = \mathbf{1}$, so the factor loadings follow random walk processes. Also, the full form of the evolution equation at time t is

$$\begin{aligned} p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}, \boldsymbol{\theta}) &= p(\tilde{\boldsymbol{\beta}}_t|\boldsymbol{\zeta}, \boldsymbol{\Delta}, \tilde{\boldsymbol{\beta}}_{t-1}, \boldsymbol{W}_{t-1}^{\beta}) p(\boldsymbol{\lambda}_t|\boldsymbol{\lambda}_{t-1}, \boldsymbol{\alpha}, \boldsymbol{\phi}, \boldsymbol{U}) \\ &\times\ p(\boldsymbol{\gamma}_t|\boldsymbol{\gamma}_{t-1}, \boldsymbol{W}_{t-1}^{\gamma}) p(\boldsymbol{\eta}_t|\boldsymbol{\eta}_{t-1}, \tilde{\boldsymbol{\alpha}}, \boldsymbol{\rho}, \boldsymbol{s}). \end{aligned}$$

The technical details about each step of Liu and West's method are now presented.

3. New state vector

$$\boldsymbol{\mu}_t^{(j)} = \begin{pmatrix} \boldsymbol{\zeta}_{t-1}^{(j)} + \boldsymbol{\Delta}_{t-1}^{(j)} \tilde{\boldsymbol{\beta}}_{t-1}^{(j)} \\ \boldsymbol{\alpha}_{t-1}^{(j)} + \boldsymbol{\phi}_{t-1}^{(j)} (\boldsymbol{\lambda}_{t-1}^{(j)} - \boldsymbol{\alpha}_{t-1}^{(j)}) \\ \boldsymbol{\gamma}_{t-1}^{(j)} \\ \tilde{\boldsymbol{\alpha}}_{t-1}^{(j)} + \boldsymbol{\rho}_{t-1}^{(j)} (\boldsymbol{\eta}_{t-1}^{(j)} - \tilde{\boldsymbol{\alpha}}_{t-1}^{(j)}) \end{pmatrix}$$

4.I. $(\boldsymbol{y}_t | \boldsymbol{\mu}_t^{(j)}, \boldsymbol{m}_{t-1}^{(j)}) \sim N(\boldsymbol{\mu}_{t,[d+k+1,\cdots,d+k+m]}^{(j)}; \boldsymbol{\Omega}_{t-1}^{(j)})$ where

$$\begin{aligned} \boldsymbol{\Omega}_{t-1}^{(j)} &= f(\boldsymbol{\mu}_{t,[1,\ldots,d]}^{(j)}) \boldsymbol{\Psi}_t^{(j)} f(\boldsymbol{\mu}_{t,[1,\ldots,d]}^{(j)})' + \boldsymbol{S}_t^{(j)} \\ \boldsymbol{\Psi}_t^{(j)} &= diag(exp(\boldsymbol{\mu}_{t,[d+1,\ldots,d+k]}^{(j)})) \\ \boldsymbol{S}_t^{(j)} &= diag(exp(\boldsymbol{\mu}_{t-1,[d+k+m+1,\ldots,p_1]}^{(j)})) \end{aligned}$$

with $f(\tilde{\boldsymbol{\beta}}) = \boldsymbol{\beta}$ being a m by k matrix. In other words, f transforms a d-dimensional vector back to the lower triangular structure of $\boldsymbol{\beta}$ (see equation 2.6 with the diagonal elements set to one). Here notation $\boldsymbol{x}_{[a,\ldots,b]}$ stands for $(x_a, \ldots, x_b)'$.

4.II. We have transformed some of the parameters to better accomodate the mixture of normal approximations in the kernel method. That is, we apply the multivariate normal kernel method in the parametrisation defined on:

- $log(\phi_i/(1-\phi_i)), \qquad i = 1, \ldots, k;$
- $log(u_{ii}), \qquad i = 1, \ldots, k;$
- $log(\rho_i/(1-\rho_i)), \qquad i = 1, \ldots, m;$
- $log(s_i), \qquad i = 1, \ldots, m.$

4.III. This step is divided into four parts.

(a) Sample $\tilde{\boldsymbol{\beta}}_t^{(l)}$ from $N(\boldsymbol{a}_t^{(l)}, \boldsymbol{W}_t^{\beta})$, where

$$\boldsymbol{a}_t^{(l)} = \boldsymbol{\theta}_{t,[p_2+1,\ldots,p_2+d]}^{(l)} + \boldsymbol{\Xi}_t^{(l)} \tilde{\boldsymbol{\beta}}_{t-1}^{(k^l)},$$

and $\boldsymbol{W}_t^{\beta}$ is estimated by

$$\frac{1-\delta_\beta}{\delta_\beta}\boldsymbol{\Xi}_t^{(l)}\left[\sum_{j=1}^{M}(\tilde{\boldsymbol{\beta}}_{t-1}^{(j)}-\overline{\tilde{\boldsymbol{\beta}}}_{t-1})(\tilde{\boldsymbol{\beta}}_{t-1}^{(j)}-\overline{\tilde{\boldsymbol{\beta}}}_{t-1})'w_{t-1}^{(j)}\right]\boldsymbol{\Xi}_t^{(l)},$$

with $\overline{\tilde{\boldsymbol{\beta}}}_{t-1}=\sum_{j=1}^{M}\tilde{\boldsymbol{\beta}}_{t-1}^{(j)}w_{t-1}^{(j)}$, $\boldsymbol{\Xi}_t^{(l)}=diag(\boldsymbol{\theta}_{t,[p_2+d+1,\ldots,p_2+2d]}^{(l)})$.

(b) Sample $\boldsymbol{\lambda}_t^{(l)}$ from $N(\boldsymbol{a}_t^{(l)},\boldsymbol{\Xi}_t^{(l)})$ where

$$a_{i,t}^{(l)}=\theta_{t,i}^{(l)}+\theta_{t,k+i}^{(l)}(\lambda_{it}^{(k^l)}-\theta_{t,i}^{(l)}),\qquad i=1,\ldots,k,$$

and

$$\boldsymbol{\Xi}_t^{(l)}=vech^{-1}(\boldsymbol{\theta}_{t,[2k+2m+1,\ldots,2k+2m+k(k+1)/2]}^{(l)}).$$

(c) Sample $\boldsymbol{\gamma}_t^{(l)}$ from $N(\boldsymbol{\gamma}_t^{(k^l)},\boldsymbol{W}_t^{\gamma})$, with $\boldsymbol{W}_t^{\gamma}$ estimated by

$$\frac{1-\delta_\gamma}{\delta_\gamma}\left[\sum_{j=1}^{M}(\boldsymbol{\gamma}_{t-1}^{(j)}-\overline{\boldsymbol{\gamma}}_{t-1})(\boldsymbol{\gamma}_{t-1}^{(j)}-\overline{\boldsymbol{\gamma}}_{t-1})'w_{t-1}^{(j)}\right],$$

with $\overline{\boldsymbol{\gamma}}_{t-1}=\sum_{j=1}^{M}\boldsymbol{\gamma}_{t-1}^{(j)}w_{t-1}^{(j)}$.

(d) Sample $\boldsymbol{\eta}_t^{(l)}$ from $N(\boldsymbol{a}_t^{(l)},\boldsymbol{\Xi}_t^{(l)})$ where

$$a_{i,t}^{(l)}=\theta_{t,2k+i}^{(l)}+\theta_{t,2k+m+i}^{(l)}(\eta_{it}^{(k^l)}-\theta_{t,2k+i}^{(l)}),\qquad i=1,\ldots,m,$$

and

$$\boldsymbol{\Xi}_t^{(l)}=diag(\boldsymbol{\theta}_{t,[2k+2m+k(k+1)/2+1,\ldots,p_2]}^{(l)}).$$

In the next Section, we combine the methods presented and developed in this chapter and in the previous one. We reanalyse the financial time series of Aguilar and West (2000) in that setting.

6.7 Daily exchange rate returns: revisited

As opposed to Chapter 5, where we basically discussed modeling issues and focused on smoothed analysis of multivariate financial time series, here we entertain two competing factor stochastic volatility models (see equations (5.1) through (5.6) from the previous chapter). The difference between the two models lies in whether or not the unconstrained elements of the factor loading matrix, $\boldsymbol{\beta}$, follow univariate random walk processes. In other words, models with different discount factors, δ_β, are compared (see the discussion after equation (5.6)). The constant factor loading model, $\delta_\beta = 1$, is the benchmark model.

6.7.1 Implementation

The model structure considered here is the same as presented in Section 5.4.3. Recalling, a $k = 3$-factor stochastic volatility model (Chapter 5) is fitted to the thousand observations on the international exchange rate returns ($m = 6$), from 1/1/1992 to 10/31/1995 ($T = 1000$). Posterior inferences, at time T, are based on a MCMC sample of size $M = 10,000$ from the joint posterior. The state and parameter vectors are

$$\boldsymbol{x}_t = (\tilde{\boldsymbol{\beta}}_t', \boldsymbol{\lambda}_t', \boldsymbol{\gamma}_t', \boldsymbol{\eta}_t')'$$

and

$$\boldsymbol{\theta}_t = (\boldsymbol{\alpha}', \boldsymbol{\phi}', \tilde{\boldsymbol{\alpha}}', \boldsymbol{\rho}', vech(\boldsymbol{U})', \boldsymbol{s}')'$$

of size $p_1 = 27$ and $p_2 = 30$, respectively. In order to investigate the impact of time-varying loadings in the factor stochastic model, we fixed all other parameters at their (approximate) posterior expected values for time $t = 0$ ($T = 1000$ in the previous

analysis). Those values, from Tables 5.4, 5.6 and 5.7, are

$$(\boldsymbol{\alpha}', \boldsymbol{\phi}')' = \left(\begin{array}{r} -10.19 \\ -11.83 \\ -10.94 \\ \hline 0.98 \\ 0.99 \\ 0.99 \end{array}\right), (\tilde{\boldsymbol{\alpha}}, \boldsymbol{\rho}, \boldsymbol{s}) = \left(\begin{array}{rrr} -13.67 & 0.99 & 0.001 \\ -15.45 & 0.84 & 0.001 \\ -13.76 & 0.88 & 0.001 \\ -14.35 & 0.92 & 0.648 \\ -11.65 & 0.99 & 0.001 \\ -11.97 & 0.94 & 0.201 \end{array}\right),$$

$$\boldsymbol{U} = \left(\begin{array}{rrr} 0.053 & 0.076 & 0.060 \\ 0.076 & 0.115 & 0.089 \\ 0.060 & 0.089 & 0.078 \end{array}\right), \overline{\boldsymbol{\gamma}} = \left(\begin{array}{r} 0.000277 \\ 0.000148 \\ 0.000020 \\ 0.000284 \\ -0.000019 \\ 0.000156 \end{array}\right),$$

where $\overline{\boldsymbol{\gamma}} = E(\boldsymbol{\gamma}_0|D_0)$ is the time series (approximate) posterior expected value based on the first thousand observations. We argue that by doing this we narrow down the sources of discrepancies between the alternative models only to the components of $\boldsymbol{x}_t$ and $\boldsymbol{\theta}_t$ that contribute to the computation of the financial time series one-step ahead forecast covariance matrices. Such matrices play important roles in both the computation of posterior predictive densities and in assessing portfolio performance, as will be clear in the following section. Identical steps were performed for the one and two-factor models, and these are compared to the three-factor model in the next section. Finally, the discount factor from Section 6.4 was fixed at $\delta = 0.83$, which correspond to a shrinkage parameter $a = 0.9$ and a smoothing parameter $h = 0.4$. Basically, when centering the terms in the kernel mixture, 90% of the weight goes to the sampled values from the posterior at time $t-1$, while the components variances are 40% of the variances of the draws. This is an arbritary choice and other values should be considered if robustness in the forecast/portfolio performance is a major concern. The results, measured in goodness of fit and forecast performance terms, were virtually the same when δ was chosen to be in $(0.8, 0.99)$.

6.7.2 Predictive analysis

In the above framework, we perform sequential analysis based on the filtering algorithm presented earlier in the chapter. We use data from 11/01/95 ($t = 1$) up to 10/31/97 ($t = 523$). We have seen, in Section 6.5, how the one-step ahead forecast distributions are computed based on the parameter and state prior samples. Equations (6.17) and (6.18) are particularly important in performing portfolio allocation.

For illustrative purposes, five models were entertained, for each $k = 1, 2, 3$:

Model 1: The loading matrix, $\boldsymbol{\beta}$, is fixed at its posterior mean at $t = 0$ and kept constant throughout the sample period.

Model 2: The loading matrix, $\boldsymbol{\beta}$, does not evolve with time, even though $p(\boldsymbol{\theta})$ is updated as new observations arrive. In other words, $\delta_\beta = 1$.

Model 3: $\delta_\beta = 0.9975$.

Model 4: $\delta_\beta = 0.99$.

Model 5: $\delta_\beta = 0.95$.

Table 6.1 shows the log-predictive densities when performing the sequential procedure for a period of two years, i.e. 523 daily observations. According to the table, a $k = 2$ factor model has the highest posterior predictive density for all five alternative models of the factor loadings evolution. Therefore, if one-step ahead forecast performance is the analyst's ultimate goal, the two factor model with $\delta_\beta = 0.95$ would provide the best results, among the fifteen entertained models. Figure 6.1 emphasizes the argument. The number of factors seems to be unimportant up to mid-April, 1996. After that, and more evidently after December, 1996, the one and three-factor models deteriorate as far as predictive capability is concerned. The one and three factor models exhibit fairly similar predictive performances.

Model	$k=1$	$k=2$	$k=3$
1	-353.4	-196.3	-276.6
2	-353.0	-200.1	-270.5
3	-294.8	-182.5	-256.6
4	-272.1	-137.8	-257.0
5	-238.5	-96.8	-228.8

Table 6.1: Predictive analysis for the k-factor stochastic volatility model (k=1, 2, 3) based on data from 11/01/95 to 10/31/97. *Model 1*: $\boldsymbol{\beta}_t = E(\boldsymbol{\beta}|D_0)$; *Model 2*: $\delta_\beta = 1.0000$; *Model 3*: $\delta_\beta = 0.9975$; *Model 4*: $\delta_\beta = 0.9900$; *Model 5*: $\delta_\beta = 0.9500$.

Another way of comparing models, at least in the present context, is by assessing practical portfolio performance. Some results are presented in the following section.

6.7.3 Portfolio comparisons

In many situations statistical performance is not the only interest. In our application, where the modeler or decision maker is, for instance, a risk manager or portfolio analyst, more interest might be given to cumulative returns of competing portfolios, generally based upon competing models. In the previous section we compared fifteen different models in terms of their predictive performance. In this section we will compare their performance based on portfolio cumulative returns as explored in Aguilar and West (2000), for instance.

Broadly speaking, a portfolio is a vector $\boldsymbol{a}$ (in our case of size m) that comprises the amount invested on a set of risky assets (exchange rates in our application). At time $t-1$, a portfolio $\boldsymbol{a}_t$ with target return m, i.e. $\boldsymbol{a}_t'\boldsymbol{\gamma}_t = m$, is obtained by minimizing its one-step ahead variance. In our application we simply take one of Aguilar and West's (2000) target values, $m = 0.0016$. Hence, the investor's problem can be seen as solving a quadratic programming problem:

$$\min_{\boldsymbol{a}_t} \boldsymbol{a}_t'\boldsymbol{\Omega}_t\boldsymbol{a}_t \quad \text{subject to} \quad \begin{cases} \boldsymbol{a}_t'\boldsymbol{\gamma}_t = m \\ \boldsymbol{a}_t'\mathbf{1} = 1 \end{cases} \tag{6.19}$$

where $\mathbf{1}$ is the unity vector of size m and $\mathbf{\Omega}_t = Var(\boldsymbol{y}_t|D_{t-1})$. The second constraint can be interpreted as forcing the investor to fix the total amount invested at each time. Without the second constraint, the investor is allowed to take long and short positions across the currencies in order to optimize the portfolio.

Let the unconstrained fixed-target portfolio, i.e. the solution of the quadratic programming equation (6.19) without the sum-one constraint, be denoted by $\tilde{\boldsymbol{a}}_t^{(m)}$. The closed form solutions for the dynamic programming problem are (Markowitz, 1959):

$$\boldsymbol{a}_t^{(m)} = \mathbf{\Omega}_t^{-1}\left(\mathbf{1}'\boldsymbol{e}_t - \boldsymbol{\gamma}_t'\boldsymbol{e}_t\mathbf{1}\right)) \tag{6.20}$$

$$\tilde{\boldsymbol{a}}_t^{(m)} = \mathbf{\Omega}_t^{-1}\boldsymbol{\gamma}_t/\boldsymbol{\gamma}_t'\mathbf{\Omega}_t^{-1}\boldsymbol{\gamma}_t \tag{6.21}$$

where $\boldsymbol{e}_t = \mathbf{\Omega}_t^{-1}(\mathbf{1}m - \boldsymbol{\gamma}_t)/d_t$ and $d_t = (\mathbf{1}'\mathbf{\Omega}_t^{-1}\mathbf{1})(\boldsymbol{\gamma}_t'\mathbf{\Omega}_t^{-1}\boldsymbol{\gamma}_t) - (\mathbf{1}'\mathbf{\Omega}_t^{-1}\boldsymbol{\gamma}_t)^2$.

Besides being the minimum-variance fixed-target portfolio, the optimal portfolios, $\boldsymbol{a}_t^{(m)}$ and $\tilde{\boldsymbol{a}}_t^{(m)}$, are the ones that maximize one-step ahead expected returns amongst all portfolios with common variances.

Without any constraints, the minimization problem from equation (6.19) yields two well-known portfolios. The *minimum variance* portfolio,

$$\boldsymbol{a}_t^{minvar} = \frac{\mathbf{\Omega}_t^{-1}\mathbf{1}}{\mathbf{1}'\mathbf{\Omega}_t^{-1}\mathbf{1}} \tag{6.22}$$

From a statistical viewpoint, Polson and Tew (1997) argue that the global minimum variance portfolio, which does not depend on $\boldsymbol{\gamma}_t$, is suitable for comparing models that parameterize $\mathbf{\Omega}_t$ differently, which turns out to be our main concern in this application. Further details can be found in Aguilar and West (2000, Section 5.3) and Polson and Tew (1997, Section 2) and their references.

Figure 6.2 and Table 6.2 focus on comparing the performance of three portfolios when three of the entertained models were fitted to one, two or three common factors.

Portfolio	Number of factors								
	k=1			k=2			k=3		
	After half a year: 4/30/96								
	M_1	M_4	M_5	M_1	M_4	M_5	M_1	M_4	M_5
1	-0.42	-1.34	-1.38	0.56	-1.59	-1.67	-1.42	-1.75	-2.23
2	1.55	0.92	0.54	1.79	0.55	0.18	0.09	-0.07	-0.37
3	-2.11	-2.06	-1.71	-1.76	-1.73	-1.50	-1.47	-1.70	-1.75
	After one year: 10/31/96								
	M_1	M_4	M_5	M_1	M_4	M_5	M_1	M_4	M_5
1	8.98	3.21	3.41	10.04	2.42	2.19	3.67	3.40	3.32
2	4.12	3.93	3.78	4.05	3.43	3.06	3.68	3.56	3.17
3	1.15	1.30	1.89	1.17	1.22	2.26	1.14	1.36	2.84
	After one year and a half: 4/30/96								
	M_1	M_4	M_5	M_1	M_4	M_5	M_1	M_4	M_5
1	14.63	0.51	0.21	16.90	-1.00	-2.27	3.42	2.61	1.32
2	11.05	5.33	5.21	11.36	5.16	3.56	8.29	7.94	5.26
3	-3.10	-2.50	-2.43	-3.72	-3.26	-2.62	-2.61	-2.38	-0.15
	After two year: 10/31/97								
	M_1	M_4	M_5	M_1	M_4	M_5	M_1	M_4	M_5
1	18.43	1.89	1.07	20.47	-0.25	-1.57	4.70	3.96	2.98
2	12.47	6.04	5.29	13.15	5.99	3.79	9.29	9.15	6.48
3	-2.42	-1.18	-1.39	-3.87	-3.07	-2.15	-2.17	-1.87	1.12

Table 6.2: Comparing portfolio performances for the k-factor stochastic volatility models (k=1,2,3) based on data from 11/01/95 to 10/31/97. Portfolio one is the constrained fixed-target (6.20), two is the unconstrained fixed-target (6.21) and three is the global minimum variance portfolio (6.22). $M_1 \ : \ \boldsymbol{\beta}_t = E(\boldsymbol{\beta}|D_0)$; $M_2 \ : \ \delta_\beta = 0.99$; $M_5 \ : \ \delta_\beta = 0.95$.

Figure 6.3 presents the performance of the three portfolios strategies discussed above. Portfolio one is the constrained fixed-target (6.20), two is the unconstrained fixed-target (6.21), and three is the global minimum variance portfolio (6.22). The rows represent the five models entertained (Section 6.7.2). As far as cumulative returns are concerned, the mean-variance efficient portfolio and the minimum variance portfolio have the lowest returns, as is to be expected. Additionally, the constrained fixed-target portfolio appears to exhibit the best performance across the models.

The constrained fixed-target portfolio with fixed factor loadings and a two-factor

structure seems to be the best overall model. However, when comparing the global minimum variance portfolios, the models with time-varying loadings give better results especially in the two or three-factor models. Figure (6.4) shows the portfolio composition through time for two competing models (M_1 and M_5) when two-factor models are fitted and three different portfolio strategies performed: constrained fixed-target portfolio, unconstrained fixed-target portfolio (middle), and global minimum variance portfolio. The elements of $\boldsymbol{\alpha}, \boldsymbol{\phi}, \tilde{\boldsymbol{\alpha}}, \boldsymbol{\rho}, \boldsymbol{U}, \boldsymbol{s}$ and $\overline{\boldsymbol{\gamma}}$ were fixed at,

$$\begin{pmatrix} \alpha \\ \phi \\ U \end{pmatrix} = \begin{pmatrix} -9.90 \\ 0.99 \\ 0.00369 \end{pmatrix}, \begin{pmatrix} \tilde{\boldsymbol{\alpha}}' \\ \boldsymbol{\rho}' \\ \boldsymbol{s}' \end{pmatrix}' = \begin{pmatrix} -13.51 & 0.99 & 0.001 \\ -11.31 & 0.99 & 0.020 \\ -10.48 & 0.99 & 0.002 \\ -14.51 & 0.99 & 0.693 \\ -11.64 & 0.93 & 0.001 \\ -11.97 & 0.99 & 0.220 \end{pmatrix},$$

and $10^4\overline{\boldsymbol{\gamma}}' = (2.60, 0.75, 0.64, 2.65, -0.25, 1.55)$ for the two-factor model, and

$$\begin{pmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\phi} \\ vech(\boldsymbol{U}) \end{pmatrix} = \left(\begin{array}{r} -10.18 \\ -11.34 \\ \hline 0.63 \\ 0.88 \\ \hline 0.382 \\ 0.303 \\ 0.248 \end{array}\right), (\tilde{\boldsymbol{\alpha}}, \boldsymbol{\rho}, \boldsymbol{s}) = \begin{pmatrix} -13.55 & 0.99 & 0.001 \\ -19.10 & 0.87 & 0.001 \\ -10.52 & 0.99 & 0.002 \\ -14.49 & 0.92 & 0.741 \\ -11.64 & 0.99 & 0.001 \\ -11.96 & 0.93 & 0.210 \end{pmatrix},$$

and $10^4\overline{\boldsymbol{\gamma}}' = (3.52, 2.41, 0.94, 3.57, -0.28, 2.36)$, for the three-factor model.

6.8 Summary

In this chapter simulation-based sequential time series analysis were and adapted to a class of factor stochastic volatility models where the factor loading matrices are allowed to evolve with time. We have revisited the international exchange rate application discussed in Chapter 5. We note some interesting summary findings.

- When the comparison amongst competing models, particularly models M_1 and M_5, were made with respect to their predictive capability, we found increasing performance when the factor loading's discount, δ_β, was diminished. In other words, the faster the factor loadings aadapt the better is the one-step ahead forecasting performance. This behavior is presented in one, two or three factor models, with more emphasis on the one factor model (see Figure 6.1). Unfortunately, this empirical result can not be state in general, for as δ_β decreases the one-step forecasting accuracy will eventually get very bad, for small values of δ_β.

- The fixed-target portfolios (constrained and unconstrained) indicate that models M_1 has the best financial performance, as measured by percen cumulative returns. On the other hand, M_5 is the best when the global minimum variance portfolio is the chosen strategy (see Figure 6.2). We argue that, when comparing different models, the minimum variance portfolio should be the one to be used because its is insensitive to changes in the levels of the time series. In other words, since this portfolio strategy is a function of $\mathbf{\Omega}_t$ (and not of $\boldsymbol{\gamma}_t$), we are able to see the effect of allowing the factor loadings evolve through time without such effect being confounded with variations in the levels.

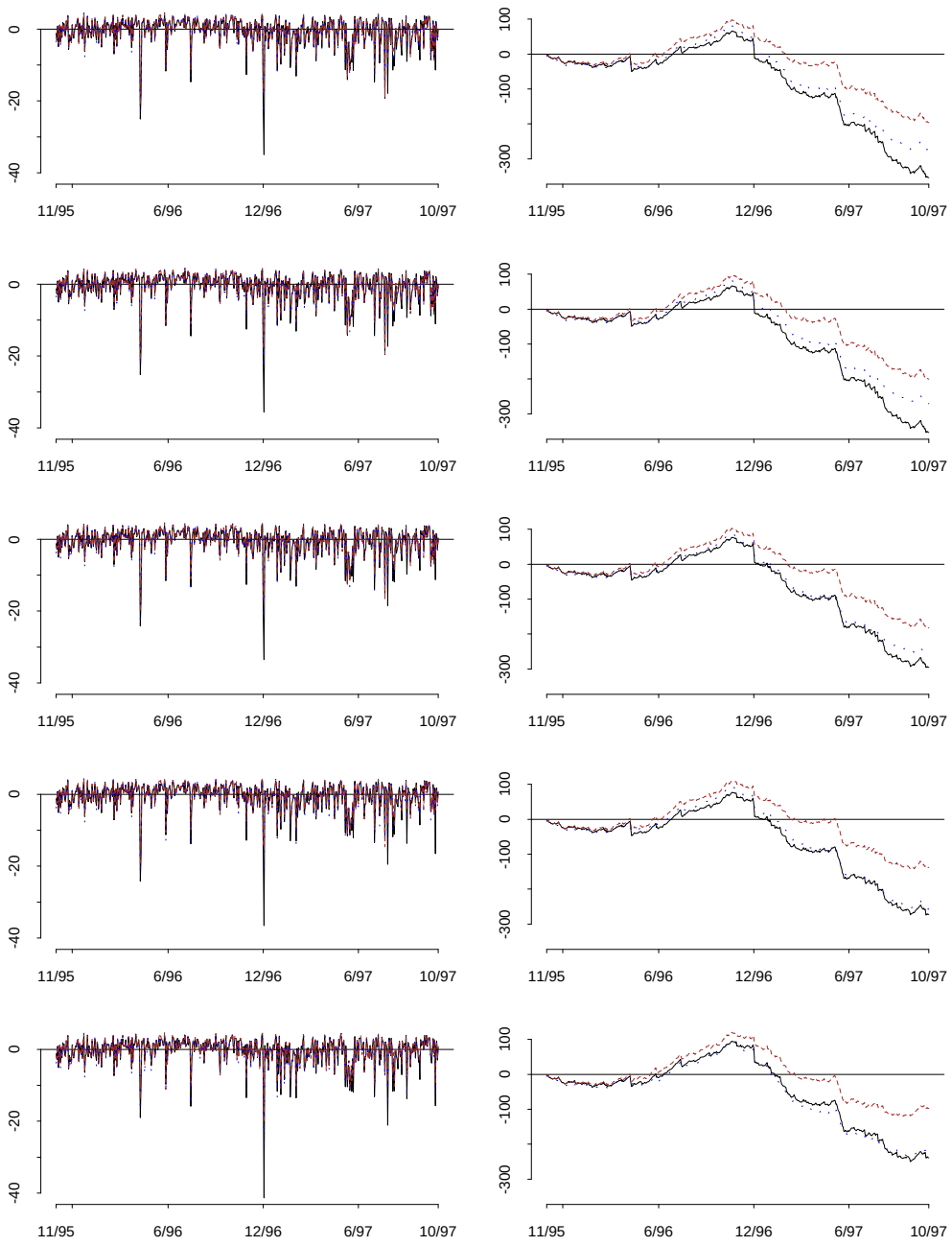


Figure 6.1: Predictive analysis for the fifteen competing model based on data from 11/01/95 to 10/31/97. Row i corresponds to the $log(p(\boldsymbol{y}_t|D_{t-1}, M_i))$ (left) and $log(\prod_{l=1}^{t} p(\boldsymbol{y}_l|D_{l-1}, M_i))$ (right), where M_i is model i $(i = 1, \ldots, 5)$. In each frame, the *solid line* represents the one factor stochastic volatility model, while the *dashed line* and the *dotted line* indicates the two and three factor models, respectively.

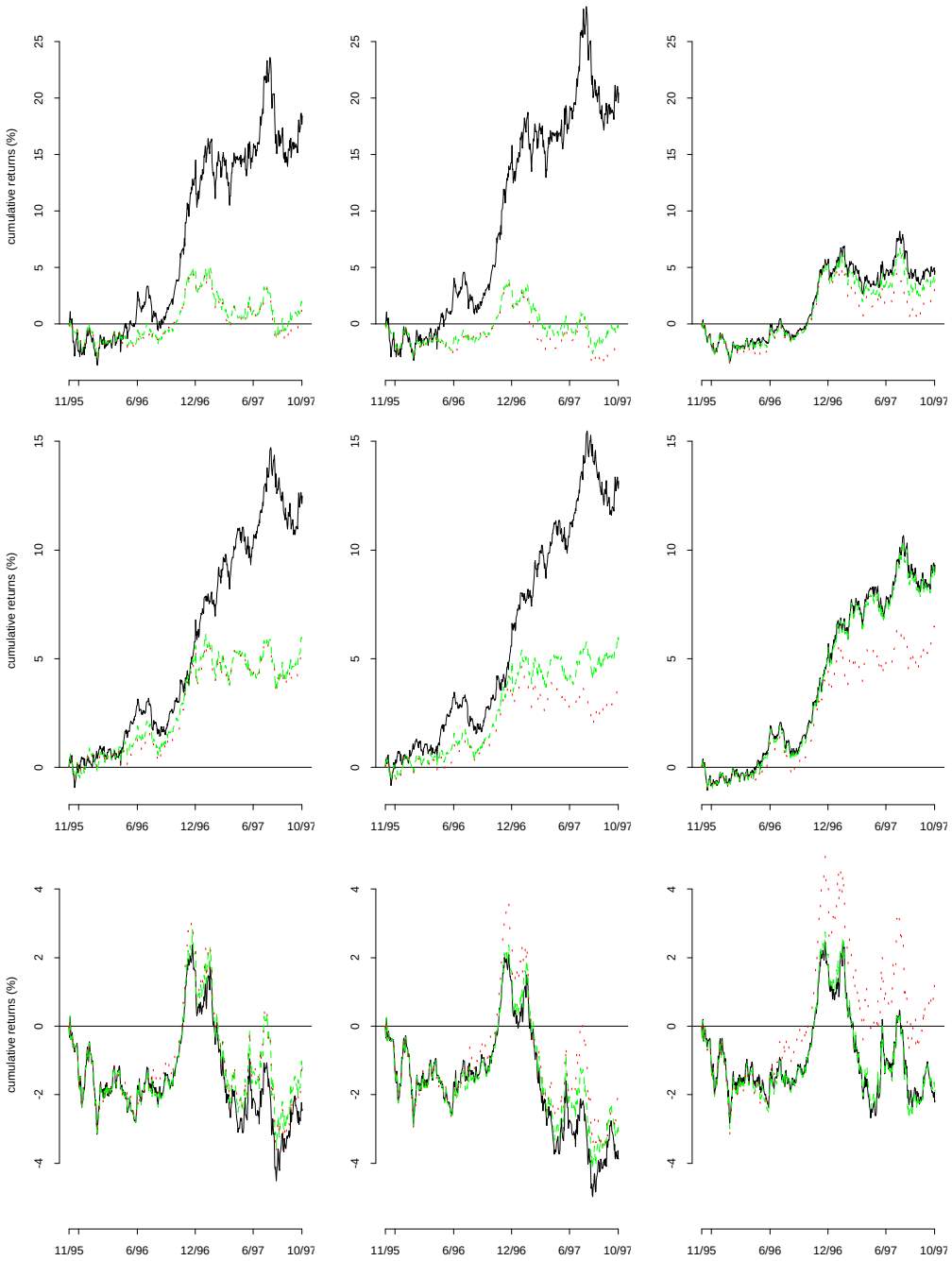


Figure 6.2: Comparing portfolio performances for the k-factor stochastic volatility models (k=1,2,3) based on data from 11/01/95 to 10/31/97. The columns are the number of common factors: $k = 1$(left), $k = 2$(middle) and $k = 3$(right). The rows are the different portfolio strategies: constrained fixed-target portfolio (top), unconstrained fixed-target portfolio (middle), and global minimum variance portfolio (bottom). Each frame has the cumulative returns for three competing models: $\boldsymbol{\beta}_t = E(\boldsymbol{\beta}|D_0)$ (solid line), $\delta_\beta = 0.99$ (dashed line), and $\delta_\beta = 0.95$ (dotted line).

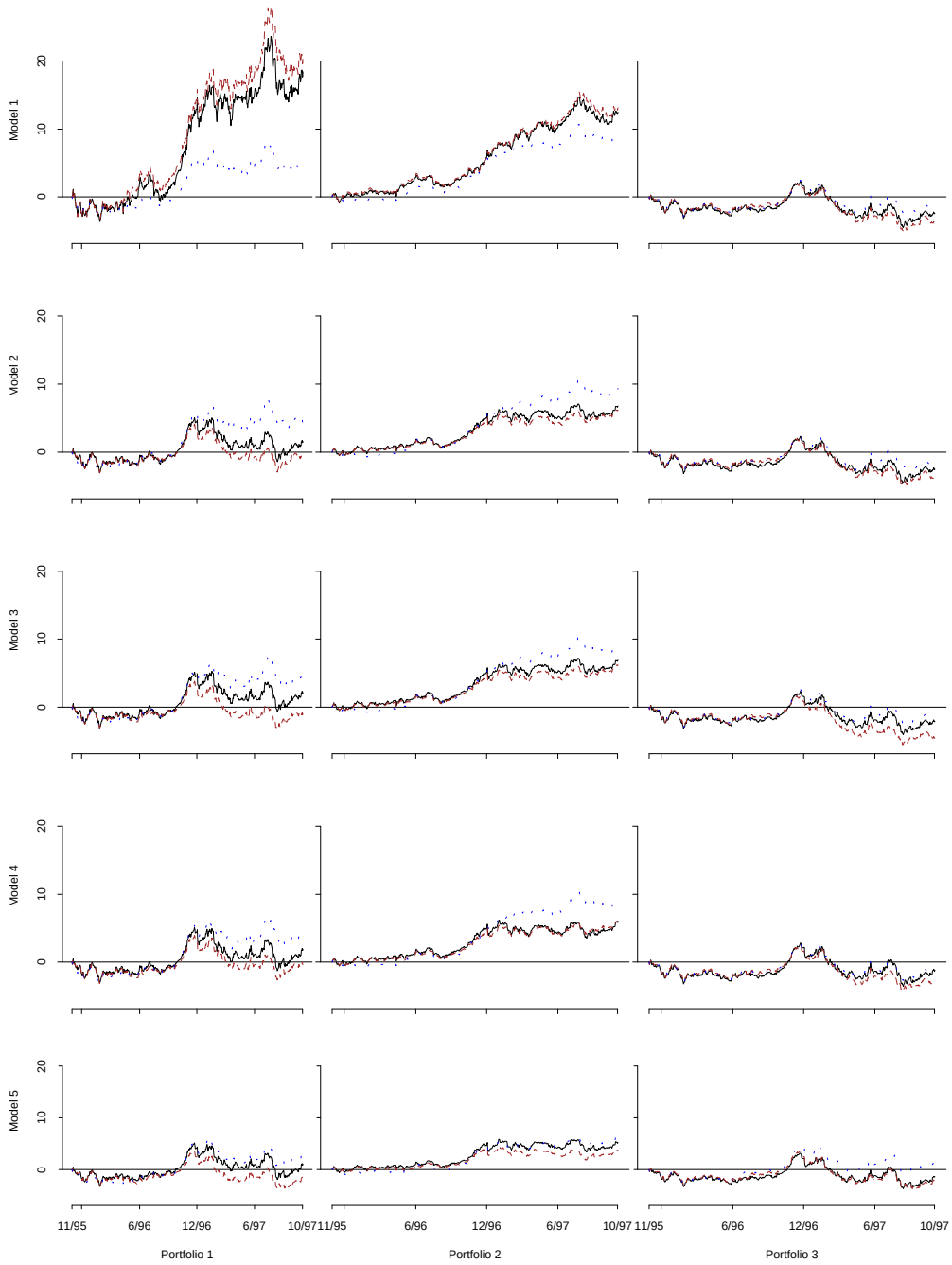


Figure 6.3: Portfolio comparisons. Each column represents the one-step ahead performance of a particular portfolio. For instance, portfolio 1 (left column) is represented by equation 6.21, while portfolios 2 (middle column) and 3 (right column) are described by equations (6.20) and (6.22), respectively. Models 1 to 5 are presented in the text (Section 6.7.2). In each frame, the *solid line* represents the one factor stochastic volatility model, while the *dashed line* and the *dotted line* indicates the two and three factor models, respectively.

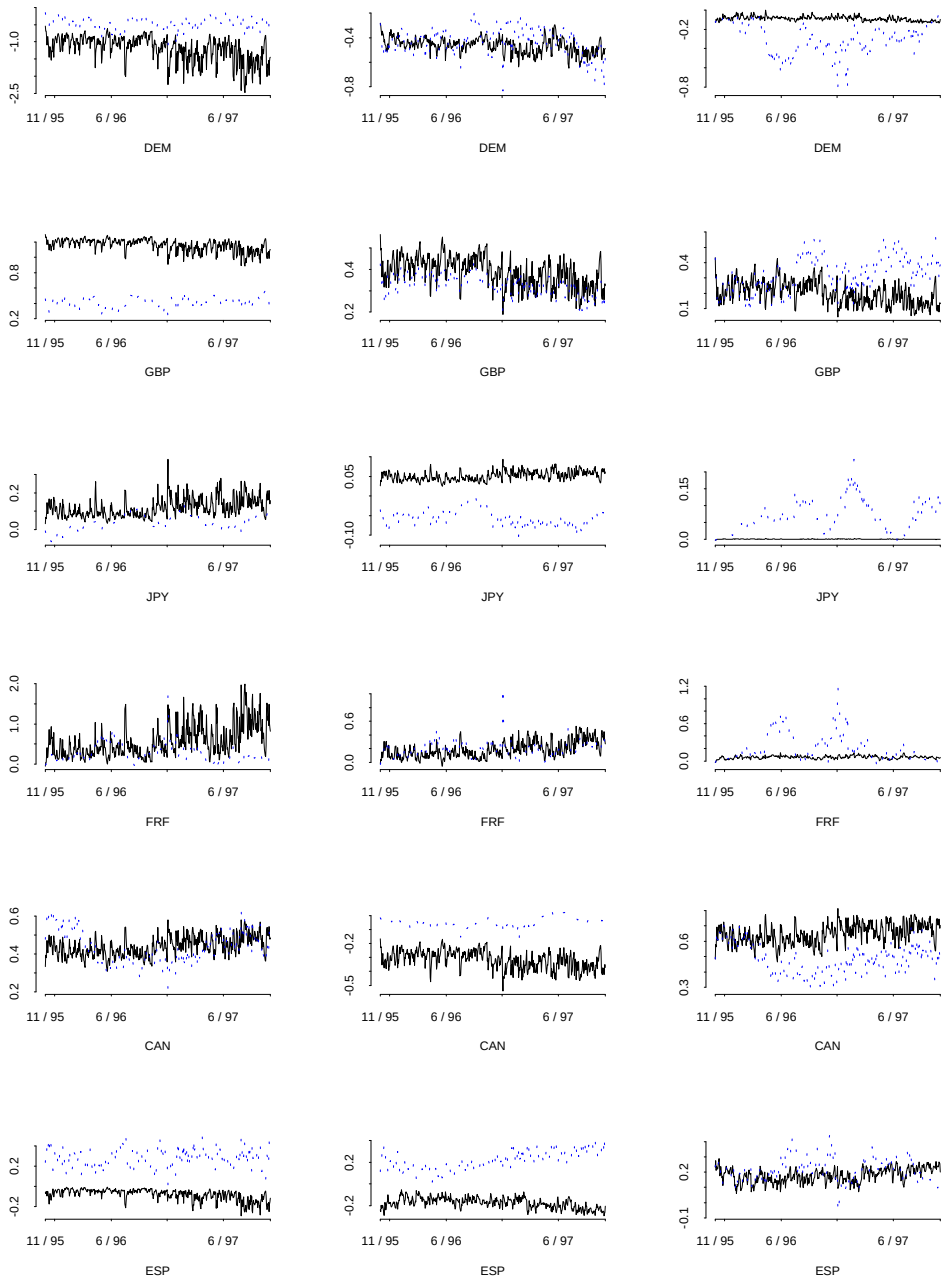


Figure 6.4: Portfolio weights for the two-factor stochastic volatility model based on data from 11/01/95 to 10/31/97. The columns are the different portfolio strategies: constrained fixed-target portfolio (left), unconstrained fixed-target portfolio (middle), and global minimum variance portfolio (right). The rows correspond to the six countries: DEM, GBP, JPY, FRF, CAN and ESP. Each frame has the weights for two competing models: $\boldsymbol{\beta}_t = E(\boldsymbol{\beta}|D_0)$ (solid line) and $\delta_\beta = 0.95$ (dotted line).

Chapter 7

Longitudinal modeling using mixture priors

7.1 Introduction

We propose a class of longitudinal data models with random effects that generalizes currently used models in two important aspects. First, the random effects model is a flexible mixture of multivariate normals, accomodating population heterogeneity, outliers and non-linearity in regression on subject-specific covariates. Second, we extend the model to allow for meta-analysis over related studies.

Models for Bayesian inference in longitudinal data models with random effects are reviewed, for example, in Wakefield *et al.* (1998), with a focus on population pharmacokinetic/pharmacodynamic (PK/PD) studies similar to the motivating application in this chapter. Let y_{ij} denote the j-th measurement on the i-th patient, let θ_i denote a random effects vector for patient i, and let x_i denote patient-specific covariates, including treatment dose. The usual structure of population PK/PD models is

$$p(y_{ij}|\theta_i), \quad p(\theta_i|x_i,\phi), \quad p(\phi). \tag{7.1}$$

Here $p(y_{ij}|\theta_i)$ is typically a parametric non-linear regression for expected response

over time. For example, θ_i could be the parameters in a compartmental model for drug concentrations. The second level of the model specifies the prior distribution for the random effects vectors θ_i, possibly including a regression on patient-specific covariates x_i, with ϕ denoting the hyperparameters. Bayesian models similar to (7.1) have been considered in Zeger and Karim (1991) for generalized linear mixed models, in Wakefield *et al.* (1994) for the general population model assuming a multivariate normal population distribution, in Dellaportas and Smith (1993), and in Wakefield (1996) with multivariate t priors.

Heterogeneity in the patient population, outliers and over-dispersion make a strict parametric model for the population distribution $p(\theta_i|x_i, \phi)$ unreasonable in our application. Instead, we consider an essentially non-parametric extension. In maximum likelihood-based inference, popular non-parametric extensions to the population model (7.1) are NPML (Mallet, 1986), with no restrictions on the distribution of the random-effects in the model and yielding a discrete estimate of this distribution; and SNP (Davidian and Gallant 1992, 1993), a method that assumes the model's random effects have a "smooth" density and produces estimates from such a class of densities. Bayesian approaches to non-parametric extensions are described in Rosner and Müller (1997), Müller and Rosner (1997), Kleinman and Ibrahim(1998a,b), Walker and Wakefield (1994). Walker and Wakefield (1994) use Dirichlet process priors, the other references use Dirichlet process mixtures. In this chapter we propose an alternative approach, based on finite mixture of normal models. A similar idea is proposed by Magder and Zeger (1996), where the mixing distributions in univariate and multivariate linear mixed-effects models are estimated by using mixtures of gaussians. Unrelated with the application in population models, Richardson and Green (1997) and Neal (1998) argue for the use of finite mixture of normal models in place of Dirichlet process mixtures, citing issues of computational efficiency, flexibility of

prior specifications, and interpretability.

Another important extension of the basic model (7.1) is to allow joint analysis of several related studies meta-analysis. Meta-analysis is a popular theme in statistical inference, often modeled with a hierarchy as in (7.1). But there is little work on meta-analysis over related non-parametric models. One approach is discussed in Müller *et al.* (1999), which extends Dirichlet process mixtures to allow for hierarchical extensions suitable for meta-analysis.

The chapter is organized as follows. In Section 7.2 we introduce the motivating application. The random-effects model based on finite mixtures of multivariate normals is presented and discussed in Section 7.3. In this section, we also propose a new hierarchical model across related studies to allow for meta-analysis. Posterior analysis is presented in Section 7.4. In section 7.5 we extend the model to include uncertainty analysis about the number of terms in the finite mixtures of multivariate normals. Section 7.6 presents results based on the motivating application, and Section 7.7 concludes with a final discussion.

7.2 Data: Meta-analysis over related studies

The motivating application is the analysis of two studies carried out by the Cancer and Leukemia Group B (CALGB, Lichtman *et al.*, 1993). CALGB 8881 was a phase I study that sought the highest dose of the anti-cancer agent cyclophosphamide one could give cancer patients every two weeks. Patients also received the drug GM-CSF to help reduce the ill effects of cyclophosphamide on the patients marrow. The other study, CALGB 9160, built upon the experience gained in 8881. The drug amifostine had been shown in some studies to reduce some of the toxic side effects of anticancer therapy, such as cyclophosphamide and radiation therapy (Spencer and Goa, 1995). A common toxicity of cancer therapy is myelosuppression, in which the immune

system is suppressed by the therapy's killing cells involved in immune functions. The objective of CALGB 9160 was to determine if adding amifostine would reduce the myelosuppressive side effects of aggressive chemotherapy with cyclophosphamide and GM-CSF. CALGB 9160 randomized patients to receive amifostine or not, along with cyclophosphamide (3 grams per square meter of body surface area) and GM-CSF (5 micrograms per kilogram of body weight). The main study question in CALGB 9160 concerned the effect of amifostine on various measures of hematologic toxicity, such as nadir (i.e., minimum) blood cell counts or days of life-threatening myelosuppression. Since only 46 patients entered the randomized trial, we wished to use data already gathered in the earlier study to help make inference in CALGB 9160 more precise.

Let $K = 2$ be the number of studies under consideration, and n_k be the number of patients in study k. In study 8881, we have data on $n_1 = 52$ patients. The other study includes data on $n_2 = 46$ patients. In both studies, the main response was white blood cell count (WBC) for each patient over time. We will use y_{kij} to denote the j-th blood count measurement on the i-th patient in study k on day t_{kij}, recorded on a log scale of thousands/microliter, i.e., $y_{kij} = \log(\text{WBC}/1000)$, and t_{kij} denotes the corresponding times of measurements. In CALGB 8881 and 9160, we had a total of 674 and 706 observations, respectively, with the number of observations for one patient varying between 2 and 19.

Rosner and Müller (1997) used a non-linear regression model

$$y_{kij} \sim N\left[f(\boldsymbol{\theta}_{ki}, t_{kij}), \sigma^2\right]. \tag{7.2}$$

to fit the data. The vector $\boldsymbol{\theta} = (z_1, z_2, z_3, \tau_1, \tau_2, \beta_0, \beta_1)$ parametrizes a mean function $f(\boldsymbol{\theta}, t)$ that is defined piecewise as: (i) a horizontal line for $0 \leq t < \tau_1$; (ii) a straight line connecting parts (i) and (iii) ($\tau_1 \leq t < \tau_2$); and (iii) a shifted and scaled logistic

curve ($t \geq \tau_2$):

$$f(\boldsymbol{\theta}, t) = \begin{cases} z_1 & t < \tau_1 \\ rz_1 + (1-r)g(\boldsymbol{\theta}, \tau_2) & \tau_1 \leq t < \tau_2 \\ g(\boldsymbol{\theta}, t) & t \geq \tau_2 \end{cases} \tag{7.3}$$

where $r = (\tau_2 - t)/(\tau_2 - \tau_1)$ and $g(\boldsymbol{\theta}, t) = z_2 + z_3/(1 + \exp\{-\beta_0 - \beta_1(t - \tau_2)\})$. The horizontal line (i) represents the initial base-line count; the steep decline (ii) corresponds to the start of the WBC decline; and (iii) models an S-shaped recovery. Figure 7.2 illustrates the nonlinar function $f(\boldsymbol{\theta}, \cdot)$.

7.3 The random effects model

7.3.1 A mixture of normal models

We start by describing the random effects model for just one study, i.e., assuming $K = 1$. To simplify notation, we will drop the k subindex until we discuss extension to $K > 1$. Also, we shall first consider a prior distribution without a regression on covariates x_i.

To generalize a multivariate normal prior $p(\boldsymbol{\theta}_i|\phi)$, we use a mixture-of-normal model $p(\boldsymbol{\theta}_i|\phi) = \sum_{k=1}^{L} \pi_l N(\boldsymbol{\mu}+\boldsymbol{d}_l, \boldsymbol{S})$. The mixture is parametrized by $\phi = (\pi_l, \boldsymbol{\mu}, \boldsymbol{d}_l,\ l = 1, \ldots, L)$, including an overall location parameter $\boldsymbol{\mu}$ and offsets $\boldsymbol{d}_l$ for the individual terms, with the constraint $\boldsymbol{d}_1 = 0$. Under such parametrization, $\boldsymbol{\mu}$ can be assigned a non-informative prior, since it is present in all terms of the mixture. The same parametrization is used, for example, in Mengersen and Robert (1995) and Roeder and Wasserman (1997). As Celeux *et al.* (1999) point out, the parametrization of mixture models can be critical for convergence and meaningful interpretation of posterior simulation.

With sufficiently large L, the mixture-of-normal model can approximate any desired random effects distribution (Dalal and Hall, 1983; Diaconis and Ylvisaker, 1985).

As in any non-linear, non-normal modeling context, the specific choice of L is guided by two competing principles. Low L leads to parsimony and easier estimation. In the extreme case of a single normal distribution, $L = 1$, it reduces to the single multivariate normal prior. On the other hand, by choosing a large L, one can increase flexibility when fitting the population distribution $p(\boldsymbol{\theta}_i|\phi)$. Later, in Section 7.5, we discuss formal selection of L as a model selection problem.

We follow an approach used, for example, in Mallet *et al.* (1988). We now consider a model extension to include a regression on covariates $\boldsymbol{x}_i$ in $p(\boldsymbol{\theta}_i|\boldsymbol{x}_i, \boldsymbol{\phi})$ (Müller and Rosner, 1997). We augment the prior mixture of normal model to a probability model in $(\boldsymbol{\theta}_i, \boldsymbol{x}_i)$ jointly, i.e.

$$p(\boldsymbol{\theta}_i, \boldsymbol{x}_i|\phi) = \sum_{l=1}^{L} \pi_l N(\boldsymbol{\mu} + \boldsymbol{d}_l, \boldsymbol{S}). \tag{7.4}$$

The implied conditional distribution $p(\boldsymbol{\theta}_i|\boldsymbol{x}_i, \phi)$ formalizes the desired regression and takes the form of a locally weighted mixture of linear regressions.

Finally, we extend the mixture of normal random effects model to a hierarchical model across related studies to allow the desired meta-analysis. The construction of this hierarchical extension is driven by the following considerations. First, the model should include the extreme cases of one common random effects distribution on one hand, and entirely different random effects distributions for each study on the other hand. Second, the hierarchical extension should not unreasonably complicate posterior simulation. And third, the hierarchical model should allow interpretation of the additional parameters. Based on these considerations, we propose a model

$$p(\boldsymbol{\theta}_{ki}, x_{ki}|\phi) = \varepsilon\, p_c(\boldsymbol{\theta}_{ki}, x_{ki}|\phi_0) + (1 - \varepsilon)\, p_k(\boldsymbol{\theta}_{ki}, x_{ki}|\phi_k), \tag{7.5}$$

where p_c represents a commom measure shared among all studies and p_k, $k = 1, \ldots, K$, is a study-specific measure. The vector of hyperparameters ϕ is split into subvectors ϕ_0 and ϕ_k, $k = 1, \ldots, K$. The additional mixing parameter ε determines

the amount of borrowing strength across the related studies. We shall refer to p_c as the common measure, and p_k as the idiosyncratic measure. Figure 7.3 illustrates this split of the random effects distributions for the K studies. By assuming a prior distribution $p(\varepsilon)$ with support including 0 and 1 we could *a priori* include the extreme cases: total lack of or complete exchangeability across all studies, respectively.

For p_k and p_c we assume mixtures of multivariate normal models,

$$p_c(\boldsymbol{\theta}_{ki}, \boldsymbol{x}_{ki}|\phi_0) = \sum_{l=1}^{L_1} \pi_l N(\boldsymbol{\mu} + \boldsymbol{d}_l, \boldsymbol{S}) \tag{7.6}$$

$$p_k(\boldsymbol{\theta}_{ki}, \boldsymbol{x}_{ki}|\phi_k) = \sum_{l=1}^{L_2} \pi_{kl} N(\boldsymbol{\mu} + \boldsymbol{d}_{kl}, \boldsymbol{S}) \tag{7.7}$$

with possibly different numbers (L_1 and L_2) of components in the mixtures. For the moment we assume fixed size mixtures, i.e., L_1 and L_2 are fixed hyperparameters. Later, in Section 7.5, we shall extend the model to random size mixtures. Also, $\boldsymbol{\mu}$ represents the overall mean, while $\boldsymbol{d}_l$ (and $\boldsymbol{d}_{kl}$) are deviations from $\boldsymbol{\mu}$.

To avoid identifiability problems, $\boldsymbol{d}_1$ is set to 0, and the variance $\boldsymbol{S}$ is assumed equal across terms of the mixtures, even though we recognize that letting $\boldsymbol{S}$ vary across studies would not be a major computational problem.

Using a mixture of normal model to represent the unknown measures $p_c(.)$ and $p_k(.)$ we face identification problems common to any mixture model. We impose an order constraint on the weights π_l and π_{kl} to avoid lack of likelihood identifiability due to arbitrary permutations of indices.

There are at least two more sources of possible identifiability concerns. The additional level of mixture introduced in (7.5) allows alternative representations of the same random effects distribution $p(.)$ by (i) including a idiosyncratic term in $p_c(\cdot)$, and/or (ii) duplicating the same term in all idiosyncratic measures, $p_k(\cdot)$. The latter concern can be addressed by choosing as informative prior for L_2, one that favors

smaller size idiosyncratic mixtures. Also, by replacing one term in p_c by K copies in each of the idiosyncratic measures, p_k increases the model's complexity unnecessarily. The Bayesian paradigm inherently favors the more parsimonious alternative ((i) on the previous page) by intrinsically applying *Occam's razor principle* (Jefferys and Berger, 1992). The former concern, i.e. including a idiosyncratic term in the common measure, is not a problem of likelihood identifiabilty because such representation actually has *lower* likelihood than the alternative one, in which the idiosyncratic term is in the idiosyncratic measure, where it belongs.

7.3.2 Implementation

A common device in posterior simulation with mixture models like the proposed random effects model is the introduction of indicator variables to break the mixture (Diebolt and Robert, 1994). Consider indicators r_{ki} to break the mixture in (7.5), (7.6) and (7.7), such that

$$Pr\{r_{ki} = l\} = \begin{cases} \pi_l & \text{if } \; 1 \leq l \leq L_1 \\ \pi_{k,l-L_1} & \text{if } \; L_1 < l \leq L_1 + L_2 \end{cases} \tag{7.8}$$

Conditional on the indicators r_{ki}

$$p(\boldsymbol{\theta}_{ki}, \boldsymbol{x}_{ki} | r_{ki} = l) = \begin{cases} N(\boldsymbol{\mu} + \boldsymbol{d}_l, \boldsymbol{S}) & \text{if } \; 1 \leq l \leq L_1 \\ N(\boldsymbol{\mu} + \boldsymbol{d}_{kl}, \boldsymbol{S}) & \text{if } \; L_1 < l \leq L_1 + L_2 \end{cases}$$

In words, $l \leq L_1$ indicates that the random effect is sampled from the common, while $l > L_1$ indicates a sample from the idiosyncratic part of the random effects distribution.

We complete the model with prior distributions on σ^2, $\boldsymbol{S}$, $\boldsymbol{\mu}$, $\boldsymbol{d}_l$, and $\boldsymbol{d}_{kl}$. For σ^2 we assume a conditionally conjugate inverse gamma distribution, $\sigma^2 \sim IG(\alpha_0/2, \beta_0/2)$ with α_0 and β_0 positive hyperparameters. The prior on $\boldsymbol{\Phi} = \boldsymbol{S}^{-1}$ is a Wishart distribution with hyperparameters ν_0 and $\boldsymbol{\Phi}_0$, i.e., $\boldsymbol{\Phi} \sim W(\nu_0, \nu_0^{-1}\boldsymbol{\Phi}_0)$. For ε we use

a beta prior, $\varepsilon \sim Beta(a_0, b_0)$. Alternatively, we could include prior point mass at 0 and 1 without significantly complicating posterior simulations.

For $\boldsymbol{\mu}$ and $\boldsymbol{d}$ we choose conjugate multivariate normal priors, $\boldsymbol{\mu} \sim N(\tilde{\boldsymbol{\mu}}, \delta\tilde{\boldsymbol{V}})$, $\boldsymbol{d}_m \sim N(\tilde{\boldsymbol{d}}, \tilde{\boldsymbol{V}})$, $\boldsymbol{d}_{kl} \sim N(\tilde{\boldsymbol{d}}_k, \tilde{\boldsymbol{V}}_k)$ and $\tilde{\boldsymbol{d}}_k \sim N(\tilde{\tilde{\boldsymbol{d}}}, \tilde{\tilde{\boldsymbol{V}}})$, with $\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{d}}, \tilde{\tilde{\boldsymbol{d}}}, \tilde{\boldsymbol{V}}, \tilde{\boldsymbol{V}}_k$ and $\tilde{\tilde{\boldsymbol{V}}}$ fixed hyperparameters (See Section 7.6). Finally, the prior distributions for $\boldsymbol{\pi}$ are Dirichlet, $(\pi_{k1}, \ldots, \pi_{kL_2}) \sim Dir(\boldsymbol{\alpha}_k)$ for $k = 1, \ldots, K$ and $l = 1, \ldots, L_2$; while $(\pi_1, \ldots, \pi_{L_1}) \sim Dir(\boldsymbol{\alpha})$, with fixed hyperparameters $\boldsymbol{\alpha}_k = (\alpha_{k1}, \ldots, \alpha_{kL_2})$ and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_{L_1})$.

In the following section we describe in full details the Markov chain Monte Carlo algorithm designed for simulating from the full posterior and predictive distributions.

7.4 Posterior analysis

The posterior distribution for all parameters involved in the longitudinal model is

$$\begin{aligned} p(\boldsymbol{\theta}, \boldsymbol{r}, \sigma^2, \boldsymbol{\Xi}, \boldsymbol{\Phi}, \boldsymbol{\pi} | \boldsymbol{y}, \boldsymbol{t}, \boldsymbol{x}) \quad &\propto \quad p(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{t}, \sigma^2) p(\boldsymbol{\theta}|\boldsymbol{r}, \boldsymbol{\Xi}, \boldsymbol{\Phi}, \boldsymbol{x}) p(\sigma^2|\alpha_0, \beta_0) \\ &\times \quad p(\boldsymbol{r}|\boldsymbol{\pi}, \varepsilon) p(\boldsymbol{\Xi}|\delta, \tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{d}}, \tilde{\boldsymbol{V}}, \tilde{\boldsymbol{V}}_1, \ldots, \tilde{\boldsymbol{V}}_K, \tilde{\tilde{\boldsymbol{V}}}) \\ &\times \quad p(\boldsymbol{\Phi}|\nu_0, \boldsymbol{\Phi}_0) p(\boldsymbol{\pi}|\boldsymbol{\alpha}, \boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_K) p(\varepsilon|a_0, b_0) \quad (7.9) \end{aligned}$$

where,

- $\boldsymbol{\theta} = \{\boldsymbol{\theta}_{ki} \quad k = 1, \ldots, K \quad i = 1, \ldots, n_k\}$
- $\boldsymbol{\Xi} = \{\boldsymbol{\mu}, \boldsymbol{d}_2, \ldots, \boldsymbol{d}_{L_1}, \boldsymbol{d}_{kl}, \tilde{\boldsymbol{d}}_k \quad k = 1, \ldots, K \quad l = 1, \ldots, L_2\}$
- $\boldsymbol{\pi} = \{\pi_1, \ldots, \pi_{L_1}, \pi_{kl} \quad k = 1, \ldots, K \quad l = 1, \ldots, L_2\}$
- $\boldsymbol{x} = \{x_{ki} \quad k = 1, \ldots, K \quad i = 1, \ldots, n_k\}$
- $\boldsymbol{r} = \{r_{ki} \quad k = 1, \ldots, K \quad i = 1, \ldots, n_k\}$

Closed form solutions for posterior integrals with respect to the posterior in (7.9) can not be achieved, even when conditionally conjugate prior distributions are used. We use Markov chain Monte Carlo simulation in order to overcome such problems. In this section we detail the sampler used in order to implement posterior inference in our proposed model. We are essentially using a Gibbs sampler. Most of the full conditionals are easy to find and to sample. Metropolis-Hastings steps are proposed for those parameters in $\boldsymbol{\theta}_{ki}$. Further details can be found, for instance, in Müller and Rosner (1997).

[*Distribution of* σ^2]

The conditional posterior for σ^2 is obtained by combining its inverse-gamma prior and the likelihood from equation (7.2),

$$p(\sigma^2|\cdots) \quad \propto \quad p(\sigma^2|\alpha_0, \beta_0)p(\boldsymbol{y}|\boldsymbol{\theta}, \boldsymbol{t}, \sigma^2)$$

which is recognized as the kernel of an inverse gamma distribution with parameters $(\alpha_0 + n)/2$ and $\left\{\beta_0 + \sum_{k=1}^{K}\sum_{i=1}^{n_k}\sum_{j=1}^{n_{ki}}\left(y_{kij} - f(\boldsymbol{\theta}_{ki}, t_{kij})\right)^2\right\}/2$.

[*Distribution of* $\boldsymbol{\Phi}$]

The conditional posterior for $\boldsymbol{\Phi}$ is obtained by combining its Wishart prior and the likelihood from equation (7.2). Let $\tilde{\boldsymbol{\theta}}_{ki} = (\boldsymbol{\theta}_{ki}, \boldsymbol{x}_{ki})'$ and $d = dim(\tilde{\boldsymbol{\theta}}_{ki})$. Then,

$$p(\boldsymbol{\Phi}|\cdots) \quad \propto \quad p(\boldsymbol{\Phi}|\nu_0, \boldsymbol{\Phi}_0)p(\boldsymbol{\theta}|\boldsymbol{r}, \boldsymbol{\Xi}, , \boldsymbol{\Phi}, \boldsymbol{x})$$

is recognized as the kernel of a Wishart distribution with $\nu_0 + n_p$ degrees of freedom and scale matrix $\left[\nu_0\boldsymbol{\Phi}_0^{-1} + \sum_{k=1}^{K}\sum_{i=1}^{n_k}(\tilde{\boldsymbol{\theta}}_{ki} - \boldsymbol{\mu}_{ki})(\tilde{\boldsymbol{\theta}}_{ki} - \boldsymbol{\mu}_{ki})'\right]$.

The $\boldsymbol{\mu}_{ki}$ are functions of the indicators, whose full conditional distributions are given below. Such indicators, as well known in the literature of mixture models,

are used to facilitate the exposition and computation of the likelihood of the model. More specifically,

$$\boldsymbol{\mu}_{ki} \equiv \boldsymbol{\mu}_{ki}(r_{ki} = l) = \begin{cases} \boldsymbol{\mu} + \boldsymbol{d}_l & \text{if } 1 \leq l \leq L_1 \\ \boldsymbol{\mu} + \boldsymbol{d}_{k(l-L_1)} & L_1 < l \leq L_1 + L_2 \end{cases}$$

[*Distribution of* $\boldsymbol{r}_{ki}$]

Initially, let us define $\tilde{\boldsymbol{r}} = (\tilde{r}_1, \ldots, \tilde{r}_{n_p})$ as the collection of indicators for all patients under all studies. Notice that the $\tilde{r}$'s are the r_{ki}'s from above without discriminating from studies. Analogously, $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{n_p})$ corresponds to the patient specific parameters. The objective is to compute

$$Pr(\tilde{r}_i = m | \tilde{\boldsymbol{r}}_{(i)}, \boldsymbol{\theta}, \boldsymbol{\pi}, \boldsymbol{\Phi}, \varepsilon)$$

where $\tilde{\boldsymbol{r}}_{(i)} = (\tilde{r}_1, \ldots, \tilde{r}_{i-1}, \tilde{r}_{i+1}, \ldots, \tilde{r}_{n_p})$. Some results used in this section are presented in the Appendix C.

$\underline{m = 1, \ldots, L_1}$ It can be shown that

$$Pr(\tilde{r}_i = m | \tilde{\boldsymbol{r}}_{(i)}, \boldsymbol{\theta}, \boldsymbol{\pi}, \boldsymbol{\Phi}, \varepsilon) \propto \varepsilon \pi_m dN(\boldsymbol{\theta}_i | \boldsymbol{\mu}_{m,i}, \boldsymbol{S} + \boldsymbol{V}_{m,i})$$

where

$$\begin{aligned} \boldsymbol{V}_{1,i}^{-1} &= n_{1,i}\boldsymbol{\Phi} + \tilde{\boldsymbol{V}}^{-1}/\delta \\ \boldsymbol{\mu}_{1,i} &= \boldsymbol{V}_{1,i}\left(n_{1,i}\boldsymbol{\Phi}\tilde{\boldsymbol{\theta}}_{1,i} + \tilde{\boldsymbol{V}}^{-1}\tilde{\boldsymbol{\mu}}/\delta\right) \\ \boldsymbol{V}_{m,i}^{-1} &= n_{m,i}\boldsymbol{\Phi} + \tilde{\boldsymbol{V}}^{-1}/(1+\delta) \qquad m = 2, \ldots, L_1 \\ \boldsymbol{\mu}_{m,i} &= \boldsymbol{V}_{m,i}\left(n_{m,i}\boldsymbol{\Phi}\tilde{\boldsymbol{\theta}}_{m,i} + \tilde{\boldsymbol{V}}^{-1}(\tilde{\boldsymbol{\mu}} + \tilde{\boldsymbol{d}})/(1+\delta)\right) \qquad m = 2, \ldots, L_1 \end{aligned}$$

with $I_{m,i} = \{l : \tilde{r}_l = m, l = 1, \ldots, n_p \text{ and } l \neq i\}$, $n_{m,i} = card(I_{m,i})$ and $n_{m,i}\tilde{\boldsymbol{\theta}}_{m,i} = \sum_{l \in I_{m,i}} \boldsymbol{\theta}_l$.

$\underline{m = L_1 + 1, \ldots, L_1 + L_2}$ It can be shown that

$$Pr(\tilde{r}_i = m | \tilde{\boldsymbol{r}}_{(i)}, \boldsymbol{\theta}, \boldsymbol{\pi}, \boldsymbol{\Phi}, \varepsilon) \propto (1 - \varepsilon) \pi_{k, m - L_1} dN(\boldsymbol{\theta}_i | \boldsymbol{\mu}_{m,i,k}, \boldsymbol{S} + \boldsymbol{V}_{m,i,k})$$

where k indexes those observations from study k, and

$$\begin{aligned} \boldsymbol{V}^{-1}_{m,i,k} &= n_{m,i,k} \boldsymbol{\Phi} + (\delta \tilde{\boldsymbol{V}} + \tilde{\boldsymbol{V}}_k)^{-1} \\ \boldsymbol{\mu}_{m,i,k} &= \boldsymbol{V}_{m,i,k} (n_{m,i,k} \boldsymbol{\Phi} \tilde{\boldsymbol{\theta}}_{m,i,k} + (\delta \tilde{\boldsymbol{V}} + \tilde{\boldsymbol{V}}_k)^{-1} (\tilde{\boldsymbol{\mu}} + \tilde{\boldsymbol{d}}_k)) \end{aligned}$$

with $I_{m,i,k} = \{l : \tilde{r}_l = m, l = 1, \ldots, n_p, l \neq i \text{ and study } k\}$, $n_{m,i,k} = card(I_{m,i,k})$ and $n_{m,i,k} \tilde{\boldsymbol{\theta}}_{m,i,k} = \sum_{l \in I_{m,i,k}} \boldsymbol{\theta}_l$.

[*Distribution of* ε]

The conditional posterior for ε is obtained by combining its Beta prior distribution and the likelihood from equation (7.2),

$$p(\varepsilon | \cdots) \propto p(\varepsilon | a_0, b_0) p(\boldsymbol{r} | \boldsymbol{\pi}, \varepsilon)$$

which is recognized as the kernel of a beta distribution with parameters $a_0 + \eta$ and $b_0 + n_p - \eta$, for $\eta = \sum_{m=1}^{L_1} \eta_m$, $\eta_m = \sum_{k=1}^{K} \sum_{i=1}^{n_k} \boldsymbol{I}(\tilde{r}_{ki} = m)$, $m = 1, \ldots, L_1$, and $\boldsymbol{I}(A)$ is the indicator function that equals one if A is true and zero otherwise.

[*Distribution of* $(\pi_1, \ldots, \pi_{L_1})$]

Analogously, the conditional posterior for $(\pi_1, \ldots, \pi_{L_1})$ is obtained by combining its Dirichlet prior distribution and the likelihood from equation (7.2),

$$p(\pi_1, \ldots, \pi_{L_1} | \cdots) \propto p(\pi_1, \ldots, \pi_{L_1} | \boldsymbol{\alpha}) p(\boldsymbol{r} | \boldsymbol{\pi}, \varepsilon)$$

which is recognized as the kernel of a Dirichlet distribution with parameter $\boldsymbol{\alpha} + \boldsymbol{\eta}$.

[*Distribution of* $(\pi_{k1}, \ldots, \pi_{kL_2})$]

Analogously, for each $k = 1, \ldots, K$, the conditional posterior for $(\pi_{k1}, \ldots, \pi_{kL_2})$ is also a Dirichlet distribution with parameter $\boldsymbol{\alpha}_k + \boldsymbol{\eta}_k$, where $\boldsymbol{\eta}_k = (\eta_{k1}, \ldots, \eta_{kL_2})$ and $\eta_{k,m-L_1} = \sum_{i=1}^{n_k} \boldsymbol{I}(\tilde{r}_{ki} = m)$, for $m = L_1 + 1, \ldots, L_1 + L_2$.

[*Distribution of* $\boldsymbol{\mu}$]

The conditional posterior for the baseline mean, $\boldsymbol{\mu}$, is obtained by combining its multivariate normal prior distribution and the likelihood from equation (7.2). Notice that according to equation (7.5) and conditional on $r_{ki} = m$,

$$\boldsymbol{\theta}_{ki} - d^*_{ki} \sim N(\boldsymbol{\mu}, \boldsymbol{S}).$$

where $d^*_{ki} = \boldsymbol{d}_m$ if $m \leq L_1$ and $d^*_{ki} = \boldsymbol{d}_{k(m-L_1)}$ if $m > L_1$. Once again, the traditional multivariate Bayesian linear model arises when combining the previous model for $\boldsymbol{\mu}$ with its multivariate normal prior, i.e. $N(\tilde{\boldsymbol{\mu}}, \delta\tilde{\boldsymbol{V}})$. Therefore, it can be shown that the conditional posterior for $\boldsymbol{\mu}$ is multivariate normal with mean vector and variance-covariance matrix given, respectively, by

$$\left(\tilde{\boldsymbol{V}}^{-1}/\delta + n_p\boldsymbol{S}^{-1}\right)^{-1}\left(\tilde{\boldsymbol{V}}^{-1}\tilde{\boldsymbol{\mu}}/\delta + \boldsymbol{S}^{-1}\sum_{k=1}^{K}\sum_{i=1}^{n_k}(\boldsymbol{\theta}_{ki} - d^*_{ki})\right)$$

and

$$\left(\tilde{\boldsymbol{V}}^{-1}/\delta + n_p\boldsymbol{S}^{-1}\right)^{-1}$$

[*Distribution of* $\boldsymbol{d}_2, \ldots, \boldsymbol{d}_{L_1}$ *and* $\boldsymbol{d}_{11}, \ldots, \boldsymbol{d}_{KL_2}$]

Analogously, the full conditionals for $\boldsymbol{d}_2, \ldots, \boldsymbol{d}_{L_1}$ and $\boldsymbol{d}_{11}, \ldots, \boldsymbol{d}_{KL_2}$ can be found by focusing only on those $\boldsymbol{\theta}_{ki}$ such that $r_{ki} = m$ and $m = 1, \ldots, L_1$ or $r_{ki} = m$ and $m = L_1 + 1, \ldots, L_1 + L_2$.

[*Distribution of* $\tilde{\boldsymbol{d}}_k$]

For each $k = 1, \ldots, K$, the conditional posterior for $\tilde{\boldsymbol{d}}_k$ is obtained by combining its multivariate normal prior, $N(\tilde{\boldsymbol{d}}_k|\tilde{\tilde{\boldsymbol{d}}}, \tilde{\tilde{\boldsymbol{V}}})$, and the likelihood, $\prod_{m=1}^{L_2} dN(\boldsymbol{d}_{km}|\tilde{\boldsymbol{d}}_k, \tilde{\boldsymbol{V}}_k)$, which also has a multivariate normal kernel. Therefore, conditional on all other unknown parameters, the full conditional of $\tilde{\boldsymbol{d}}_k$ is a multivariate normal with mean vector and covariance matrix given by,

$$\tilde{\tilde{\boldsymbol{d}}} + \tilde{\tilde{\boldsymbol{V}}}(\tilde{\boldsymbol{V}}_k/L_2 + \tilde{\tilde{\boldsymbol{V}}})^{-1}(\overline{\boldsymbol{d}}_k - \tilde{\tilde{\boldsymbol{d}}})$$

and

$$\tilde{\tilde{\boldsymbol{V}}} - \tilde{\tilde{\boldsymbol{V}}}(\tilde{\boldsymbol{V}}_k/L_2 + \tilde{\tilde{\boldsymbol{V}}})^{-1}\tilde{\tilde{\boldsymbol{V}}}$$

respectively.

[*Distribution of* z_1, z_2, z_3]

For the sake of notation we will omit the subscript of $\boldsymbol{\theta}$ and of all results for the rest of this section, since we will focus on each patient's random effects. Initially, let $\boldsymbol{y}$ be an n dimensional vector containing the log WBC for the patient.

Conditionally on all the other parameters, such as τ_1, τ_2, β_0 and β_1 it can be seen that $\boldsymbol{y}$ can be split in three parts $\boldsymbol{y} = (\boldsymbol{y}_1, \boldsymbol{y}_2, \boldsymbol{y}_3)'$ as follows

$$\begin{aligned}
\boldsymbol{y}_1 &= \mathbf{1}z_1 + \mathbf{0}z_2 + \mathbf{0}z_3 + \boldsymbol{\varepsilon}_1 \\
\boldsymbol{y}_2 &= \boldsymbol{r}z_1 + (\mathbf{1} - \boldsymbol{r})z_2 + (\mathbf{1} - \boldsymbol{r})\frac{e^{\beta_0}}{1 + e^{\beta_0}}z_3 + \boldsymbol{\varepsilon}_2 \\
\boldsymbol{y}_3 &= \mathbf{0}z_1 + \mathbf{1}z_2 + \frac{e^{\beta_0 + \beta_1(\boldsymbol{t} - \tau_2)}}{1 + e^{\beta_0 + \beta_1(\boldsymbol{t} - \tau_2)}}z_3 + \boldsymbol{\varepsilon}_3
\end{aligned}$$

or more concisely,

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{z} + \boldsymbol{\varepsilon}$$

for $\boldsymbol{z} = (z_1, z_2, z_3)'$. Therefore, if the prior for $\boldsymbol{z}$ is a normal distribution with mean vector $\boldsymbol{a}$ and variance covariance matrix $\boldsymbol{B}$, then its full conditional is also a multi-

variate normal with mean vector and variance-covariance matrix given by,

$$\begin{aligned} \boldsymbol{m} &= \boldsymbol{C}\left\{\boldsymbol{B}^{-1}\boldsymbol{a} + \boldsymbol{X}'\boldsymbol{y}/\sigma^2\right\} \\ \boldsymbol{C} &= \left\{\boldsymbol{B}^{-1} + (\boldsymbol{X}'\boldsymbol{X})/\sigma^2\right\}^{-1} \end{aligned}$$

respectively.

[*Distribution of* τ_1 *and* τ_2]

The following steps are performed to update τ_1 :

1. Draw a candiate τ_1^* from $N(a, B)$ where a and B are the conditional mean and variance for τ_1, derived from equation (7.5) when conditioning on the other coordinates being equal to the current values.

2. Accept τ_1^* with probability $min\left\{1, \frac{p(\boldsymbol{y}|\boldsymbol{\theta}^*)}{p(\boldsymbol{y}|\boldsymbol{\theta})}\right\}$, otherwise keep it unchanged. Here $\boldsymbol{\theta}$ and $\boldsymbol{\theta}^*$ differ only with respect to τ_1.

See the appendix from Müller and Rosner (1997) for more details. τ_2 is updated analogously.

[*Distribution of* β_0 *and* β_1]

We use here Müller and Rosner's idea in which the moments of the normal approximation to the likelihood based on the linearized logistic link function in a logistic regression with Gaussian error is used. For all y_i in $\boldsymbol{y}$ such that $\tau_1 \leq t_i < \tau_2$ the transformation is

$$g_i = \frac{y_i - r_i z_1 - (1 - r_i) z_2}{(1 - r_i) z_3}$$

while for all y_i such that $t_i \geq \tau_2$ the transformation is,

$$g_i = \frac{y_i - z_2}{z_3}$$

and defining $y_i^* = log(g_i/(1-g_i))$,

$$\boldsymbol{y}^* = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1)'$ and $\boldsymbol{X}$ is the design matrix (the first column is filled with ones and the second with zeros or $t_i - \tau_2$ depending on whether t_i is greater than τ_2 or not).

As for the τ's, we will assume that the prior for this model will be the conditional distribution obtained by equation (7.5) when conditioning on the other coordinates being equal to the current values; this distribution turns out to be a bivariate normal distribution with mean vector and covariance matrix, say, $\boldsymbol{a}_1$ and $\boldsymbol{B}_1$. Looking at this (approximate) likelihood as a function of $\boldsymbol{\beta}$ it can be seen that it has the kernel of a multivariate normal with mean vector and variance covariance matrix, say, $\boldsymbol{a}_2$ and $\boldsymbol{B}_2$. This kernel when combined with the prior yields an approximate posterior distribution that is also normal with mean and variance, $\boldsymbol{a}_3$ and $\boldsymbol{B}_3$ and that will be used as the probing distribution for β_0 and β_1. In other words, $\boldsymbol{\beta}^*$ is sampled from the (approximate) posterior, $N(\boldsymbol{a}_3, \boldsymbol{B}_3)$, and accepted with probability

$$min\left\{1, \frac{p(\boldsymbol{y}|\boldsymbol{\theta}^*)}{p(\boldsymbol{y}|\boldsymbol{\theta})} \frac{\phi(\boldsymbol{\beta}; \boldsymbol{a}_2, \boldsymbol{B}_2)}{\phi(\boldsymbol{\beta}^*; \boldsymbol{a}_2, \boldsymbol{B}_2)}\right\}$$

where, again, $\boldsymbol{\theta}$ and $\boldsymbol{\theta}^*$ differ only with respect to $\boldsymbol{\beta}$. $\phi(\boldsymbol{\beta}; \boldsymbol{a}, \boldsymbol{B})$ denotes a normal density with moments $\boldsymbol{a}$ and $\boldsymbol{B}$ (for further details see the appendix of Müller and Rosner (1997).

In the next section we discuss model selection and model averaging. There the number of terms in the mixtures defining the common and specific measures, L_1 and L_2, will be also considered uncertain.

7.5 Selecting the number of terms in the mixture

There are two main aspects to model assessment. The first is related to model selection. In many situations, even if no particular model is thought to be the true one, it is convenient to select one for scientific reporting. On the other hand, in many applications there are quantities of interest, like predictive inference for future patients, that do not depend on a particular model and might be averaged across models. General issues of model averaging and model selection are discussed in Raftery *et al.* (1997), Wasserman (1997) and Clyde (1999). Hoeting *et al.* (1999) provide a recent tutorial on Bayesian model averaging.

Let $\mathcal{M} = \{1, 2, \ldots, M\}$ denote the set of indices representing all models under consideration and assume that Δ is an outcome of interest, such as the future profile of a new patient from the population, or the time at which the white blood cell counts of a new patient drops below a critical threshold. Let $\boldsymbol{\theta}_m$ denote the parameter vector under model m. The posterior distribution for Δ is $p(\Delta|D) = \sum_{m=1}^{I} p(\Delta|m, D)Pr(m|D)$ where D denotes the data,

$$Pr(m|D) = p(D|m)Pr(m) \left\{ \sum_{\widetilde{m}=1}^{M} p(D|\widetilde{m})\ Pr(\widetilde{m}) \right\}^{-1}$$

is the posterior probability of model m and $p(D|m) = \int p(D|\boldsymbol{\theta}_m, m)p(\boldsymbol{\theta}_m|m)\ d\boldsymbol{\theta}_m$ is the marginal likelihood for model m.

In the proposed population model, neither $p(\theta_m|m, D)$ nor $p(\Delta|m, D)$ are analytically tractable. Additionally, $p(D|m)$, the predictive distribution under m, involves the solution of a possibly high-dimensional integral. Possible approaches for computing marginals $p(D|m)$ are discussed, among others, in Gelfand and Dey (1994), Chib (1995) and Meng and Wong (1996). Chib uses the candidate formula to compute the predictive density based on draws from the posterior distribution; Gelfand and

Dey use importance resampling; and Meng and Wong suggest a bridge estimator that links an importance sampling density and the target function, usually the posterior distribution. These and other approaches were presented and extensively studied in the factor analysis context in Chapter 3.

As in Chapter 3, we will use again the a reversible jump MCMC sampler. Recall that a RJMCMC algorithm is a Markov chain simulation that jumps between models of different dimensions (Green, 1995). Dellaportas *et al.* (1998) and Godsill (1998) have shown the relationship with alternative methods of model selection. The frequency with which the algorithm visits each particular model approximates $Pr(m|D)$. Detailed information about the RJMCMC is also available in appendix B.

Green (1995) and Richardson and Green (1997) propose RJMCMC algorithms for general mixture-of-normal models. However, these general algorithms are difficult to apply in the possibly high dimensional context of our application. Instead, we suggest below an RJMCMC scheme that exploits the fact that we have only few competing models to consider.

We will use posterior draws from fitting each model $\mathcal{M}_m$, $m = 1, \ldots, M$. This strategy is feasible since we only have a small number, say 4 to 16, of competing models to consider. Suppose the current state of the Markov chain is $(m, \theta_1, \ldots, \theta_M)$, then the proposed algorithm proceeds as follows:

Step 1 Propose a visit to model $\mathcal{M}_{\widetilde{m}}$ with probability $J(m \to \widetilde{m})$. See the notation from Section 3.3.

Step 2 Generate $\tilde{\boldsymbol{\theta}}_m$ from $p(\cdot|\boldsymbol{y}, m)$, $\tilde{\boldsymbol{\theta}}_{\widetilde{m}}$ from $p(\cdot|\boldsymbol{y}, \widetilde{m})$ and make $\tilde{\boldsymbol{\theta}}_k = \boldsymbol{\theta}_k$ for all $k \neq m, \widetilde{m}$.

Step 3 The acceptance probability of the new model can be calculated as the mini-

mum of 1.0 and

$$\underbrace{\frac{p(\boldsymbol{y}|\tilde{\boldsymbol{\theta}}_{\widetilde{m}},\widetilde{m})p(\tilde{\boldsymbol{\theta}}_{\widetilde{m}}|\widetilde{m})p(\widetilde{m})}{p(\boldsymbol{y}|\boldsymbol{\theta}_m,m)p(\boldsymbol{\theta}_m|m)p(m)}}_{\text{model ratio}}\underbrace{\frac{J(\widetilde{m}\to m)}{J(m\to\widetilde{m})}\frac{\prod_{k=1}^{M}p(\boldsymbol{\theta}_k|\boldsymbol{y},k)}{\prod_{k=1}^{M}p(\tilde{\boldsymbol{\theta}}_k|\boldsymbol{y},k)}}_{\text{proposal ratio}}$$

which can be rewritten as

$$\frac{p(\boldsymbol{y}|\boldsymbol{\theta}_{\widetilde{m}},\widetilde{m})}{p(\boldsymbol{y}|\tilde{\boldsymbol{\theta}}_m,m)}\frac{p(\widetilde{m})}{p(m)}\frac{J(\widetilde{m}\to m)}{J(m\to\widetilde{m})}$$

Step 3 Cycle through 1 and 2 until convergence has been achieved.

In Step 2, we proposed a new model $\widetilde{m}$ and new parameters for both the current model m and the proposed model $\widetilde{m}$. Of course, we do not need to always keep parameter values for all competing models in memory. Only when and as a value $\boldsymbol{\theta}_k$ is required in Step 2 must we read a value from the appropriate file of posterior Monte Carlo simulations. This assumes that the samples from model $\mathcal{M}_k$ are approximately independent. This can be achieved, for example, by using sufficiently large batches when saving the MCMC simulations.

There are many alternative ways to propose the jumps between the models. The main motivation for using this particular algorithm is the easy access to posterior draws from each competing model and the lack for an easier way to find approximate proposals, especially when the size of the vector $\boldsymbol{\theta}$ varies considerably from model to model. Appendix B presents the RJMCMC algorithm in greater details.

In previous chapters, we have compared the performance of most of these estimators in the context of factor analysis. We have found that RJMCMC outperforms some of the alternative methods, and it is usually easier and faster to compute. The Schwartz Criterion, commonly known as Bayesian Information Criterion (BIC), was also found to perform considerably well. Recalling Section 3.5, the BIC for model m

is defined as

$$BIC(m) = log(Pr(D|\hat{\boldsymbol{\theta}}_i, m)) - 0.5 d_i log(N)$$

where N is the sample size, d_i is the number of parameters in model m and $\hat{\boldsymbol{\theta}}$ is the maximum likelihood estimator for $\boldsymbol{\theta}$. When $Pr(m) = Pr(\widetilde{m})$ for all m and $\widetilde{m}$ in $\mathcal{M}$, the BIC approximation for $Pr(m|D)$ is:

$$\hat{P}(m|D) = \frac{Pr(D|m)}{\sum Pr(D|\widetilde{m})} \approx \frac{e^{BIC(m)}}{\sum e^{BIC(\widetilde{m})}}$$

7.6 Results

We used the proposed model (7.2) with the hierarchical mixture of normal prior (7.5), (7.6) and (7.7) to analyze the data described in Section 7.2. Here $\boldsymbol{x}$ is composed, in this order, by the covariates CTX,GM-CSF and amifostine.

We considered maximum likelihood fits to the non-linear regression (7.2) for each patient separately to obtain initial starting values for the Markov chain Monte Carlo posterior simulation. The hyperparameters were set at:

$$\begin{aligned}
\alpha_0 &= 4.25 \quad \beta_0 = 1.125 \quad \nu_0 = 12 \quad \boldsymbol{S}_0^{-1} = \boldsymbol{I} \\
\tilde{\boldsymbol{\mu}} &= (2, -1.5, 4.5, 5, 8, -2, 0.5, -0.9, -2, -0.5)' \\
\tilde{\boldsymbol{d}}_l &= \tilde{\boldsymbol{d}}_{kl} = (0, -0.5, 0, 0, 0, 0, 0, -1, 0, 0) \\
\delta\tilde{\boldsymbol{V}} &= \tilde{\boldsymbol{V}}_k = \tilde{\tilde{\boldsymbol{V}}} = \boldsymbol{I}.
\end{aligned}$$

and $\delta = 5$, where α_0, β_0 were chosen such that $E(\sigma^2) = 0.5$ and $V(\sigma^2) = 2$. The Dirichlet parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\alpha}_k$ were set to the unity. For L_1 and L_2 we used uniform priors on the models presented in Table 7.1.

The reported inference is based on 10,000 MCMC iterations, beyond a burn-in of 100,000 iterations and saving only every tenth iteration. The credible intervals

for posterior predictive profiles are based on a subsample of size 500 taken from the final 10,000 MCMC samples. Also, we are entertaining four possible models with $1 \leq L_1, L_2 \leq 2$ (Table 7.1). In preliminary analysis we let L_1 (and L_2) go up to 4, but the change in the final inference was negligible. As starting values for the latent variables, r_{ki}, we have applied some recently developed techniques for model-based cluster analysis (Fraley and Raftery,1998, 1999) to the set of initial values for $\boldsymbol{\theta}_{ki}$, which in turn were obtained by maximum likelihood estimation per patient and per study. The final results are virtually the same when using either the smart initial values or any naive ones.

Table 7.1 compares the four competing models using the traditional Bayesian information criterion (BIC,Schwarz,1978) and the proposed RJMCMC algorithm. The model with $L_1 = L_2 = 2$ has the highest posterior probability based both the BIC and the RJMCMC criteria. Although the RJMCMC points to a particular model, it also reports significant posterior probabilities for other models. Those probabilities will later be used for model averaging. Figure 7.4 shows a trace plot of imputed model indicators when we ran the RJMCMC. The sampler is mixing well across competing models.

Model	L_1	L_2	BIC	RJMCMC
M_1	1	1	0.000	0.138
M_2	1	2	0.000	0.147
M_3	2	1	0.000	0.187
M_4	2	2	1.000	0.528

Table 7.1: Model's posterior probabilities based on the Bayesian information criterion and the proposed RJMCMC algorithm.

We now consider inference in the *a posteriori* most likely model, $L_1 = L_2 = 2$. Figure 7.5 shows the posterior means for the weights, i.e.

$$\{\varepsilon\pi_1, \varepsilon\pi_2, (1-\varepsilon)\pi_{11}, (1-\varepsilon)\pi_{12}, (1-\varepsilon)\pi_{21}, (1-\varepsilon)\pi_{22}\},$$

given to the mixture terms in each study. We show in Figure 7.6 profiles $f(\boldsymbol{\theta} = \boldsymbol{\mu} + \boldsymbol{d}_l, \cdot)$ and $f(\boldsymbol{\theta} = \boldsymbol{\mu} + \boldsymbol{d}_{kl}, \cdot)$ corresponding to simulated posterior draws for $(\boldsymbol{\mu}, \boldsymbol{d}_l, \boldsymbol{d}_{kl})$ to summarize the posterior distribution on $\boldsymbol{\mu}$, $\boldsymbol{d}_m$ and $\boldsymbol{d}_{kl}$. The predictive profile for the second idiosyncratic component ($l = 2$) from study $k = 1$ is considerably different from the other components, indicating the presence of some patients whose measurements are not in agreement with the rest of the patients in both studies. The lower the weight allocated to a component, the wider are the credible intervals, indicating that less information is present in such components.

Figure 7.7 shows scatterplots of univariate maximum likelihood estimators (used as initial values in the MCMC algorithm) and posterior means (with 90% credibility interval) for the elements of $\boldsymbol{\theta}_{ki}$. The patients' initial base lines $z_{1,ki}$ have posterior means that are virtually the same as the MLE ones. Conversely, the *borrowing strength* or shrinkage effect can be clearly seen in τ_1 and τ_2, the change points, and in the parameters that define the recovery logistic curve, since the posterior means appear more constant across individuals than do the MLEs.

Figure 7.8 shows the definition of two clinically meaningful summaries of the profiles: the number of days with WBC below a critical threshold (T), and the nadir WBC (N). Figure 7.9 shows posterior predictive inference for T and N under study $k = 1$ (CALGB 8881). For medium/high doses of GM-CSF, the higher the dose of CTX, the longer the time below threshold. The higher the dose of CTX, the lower the nadir WBC. In both cases, the higher the dose of GM-CSF, the less the patient will be at risk. From the posterior predictive plots, the relationships seem non-linear, confirming the need for extensions beyond single multivariate normal population models (implying linear regression).

Bayesian model averaging is presented in Table 7.2 and figure 7.10. CTX and GM-CSF doses were fixed at $3.0mg/m^2$ and $5.0\mu g/kg$, respectively, since our main

objective here is to illustrate how posterior inference varies when more uncertainty is included in the modeling. The entries are white blood cell counts for a new patient taken from study one (CALGB 8881), from study two (CALGB 9160) and from the population, respectively. The first four columns are based on the maximal model, which is, once again, the model with two components in p_c, p_1 and p_2, while the last four columns represents the Bayesian model averaging, when average across the four entertained models. The key feature in this picture are the wider credibility intervals obtained when allowing for model uncertainty in the number of mixture components. Finally, it seems that posterior inference for the base line parameter, z_1 is similar across models suggesting, as already discussed, that this parameter is roughly constant across studies.

We compared our methodology with alternatives currently used in the literature in order to see whether combining the studies in the way we propose improves predictive performance. Figure 7.11 compares posterior predictive profiles for a new patient from study CALGB 9160, based on two competing models, when the levels of CTX and GM-CSF where fixed at $3.0mg/m^2$ and $5\mu g/kg$, respectively. The alternative model is a particular case of our own model structure where the prior for ε is degenerated at zero, i.e. $Pr(\varepsilon = 0) = 1$ (or $L_1 = 0$). In such situation, information from related studies are combined in a higher level in the hierarchy through the hyperparameters. Notice that up to five days into the treatment both models give similar results. The reason for that, as mentioned earlier, is due to the fact that during such period of time the patients from both studies behave quite similarly. However, during the most critical period, between the fifth and the fifteenth days, the meta-analysis model presents profiles slightly shifted up and wider intervals, indicating the influence of study CALGB 8881 in the analysis (see the left upper corner of table 7.2). It is worth to mention that the competing models are easily entertained in our methodology by

simply allowing L_1 to be zero.

Best Model ($L_1 = L_2 = 2$)				Bayesian Model Averaging			
Number of Days	Percentiles			Percentiles			Number of Days
	5th	50th	95th	5th	50th	95th	
	CALGB 8881			CALGB 8881			
$\tau_1 = 3.00$	3176	7258	17889	2651	6883	16163	$\tau_1 = 3.75$
$\tau_2 = 9.25$	114	556	7284	39	523	7265	$\tau_2 = 8.75$
$\tau^* = 15.0$	152	3128	24959	35	1584	18334	$\tau^* = 15.0$
	CALGB 9160			CALGB 9160			
$\tau_1 = 2.75$	3934	7609	14697	3474	7183	14190	$\tau_1 = 3.25$
$\tau_2 = 9.25$	120	458	4221	41	419	3303	$\tau_2 = 9.00$
$\tau^* = 15.0$	168	3119	19118	47	1691	18531	$\tau^* = 15.0$
	Population			Population			
$\tau_1 = 3.75$	3841	7371	13643	3822	7680	14436	$\tau_1 = 3.25$
$\tau_2 = 8.75$	104	371	2535	85	416	4066	$\tau_2 = 8.75$
$\tau^* = 15.0$	174	3713	20228	114	2651	20963	$\tau^* = 15.0$

Table 7.2: Predictive profile for a new patient's white blood cells (WBC) counts. For each study/model τ_1 and τ_2 represent the time at which the patient's WBC starts to decline and recover, respectvely. Also, τ^* is the measurement after two weeks from chemotherapy. The first four columns are the results obtained by the best model, i.e. model 4 with a L_1=2-component mixture of normals for the common measure and a L_2=2-component mixture of normals for the specific measures. The last four columns represent the results from Bayesian model averaging (see Table 7.1).

7.7 Summary

We proposed an approach to meta-analysis over related random-effects models, allowing one to pool strength across similar studies. The mixture-of-normal random-effects models allow considerable flexibility while exploiting the computational simplicity of conjugate models. The proposed implementation includes Bayesian model averaging over mixtures of different size using a version of RJMCMC simulation. We propose a variation of RJMCMC that enables us to implement model averaging for the relatively high dimensional parameter vectors.

The proposed approach has some practical and methodological limitations. Theoretically the number of terms in the mixture is treated as random. However, computational issues become relevant when more than a few models (number of terms) are entertained. From a more theoretical viewpoint, meta-analytic structures beyond exchangeable studies are ruled out, for instance, when we want to borrow strength only in some key features, and leave others specific to the respective study.

Although the discussion was in the context of meta-analysis over longitudinal data models, the proposed methods are more generally applicable. The approach is relevant whenever hierarchical modeling over related random probability distributions is required.

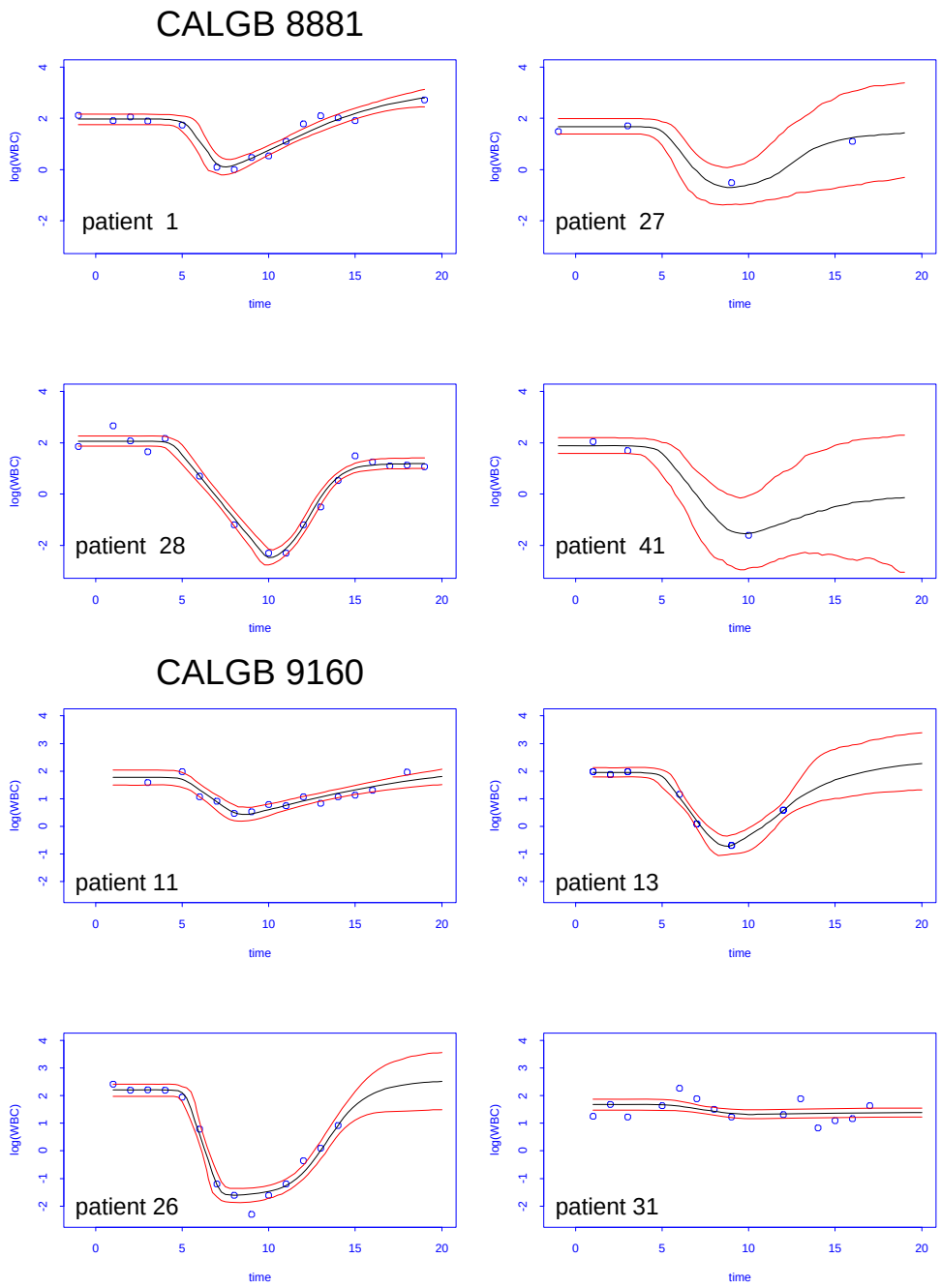


Figure 7.1: Observed white blood cell count (log(WBC/1000)) for some typical patients from studies CALGB 8881 and CALGB 9160. The solid lines represent posterior means, 5th and 95th percentiles for the patients' profiles when fitting the "best model" (Table 7.1 from Section 7.6).

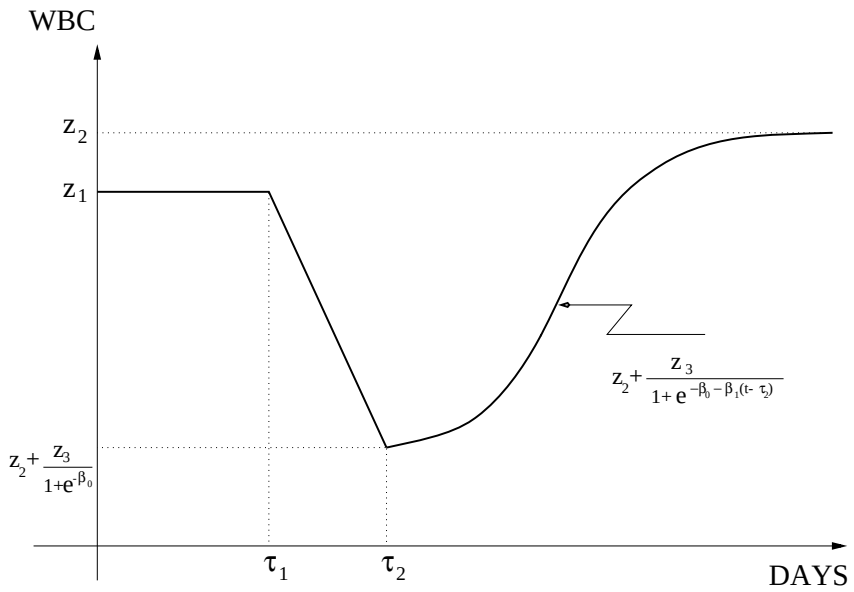


Figure 7.2: The profile for the a patient is a horizontal line between days 0 and τ_1, a logistic curve after day τ_2, a line between the two changing points. Day τ_1 is the day when blood cell counts start to drop from the initial level z_1 and τ_2 indicates the beginning of (logistic) recovery.

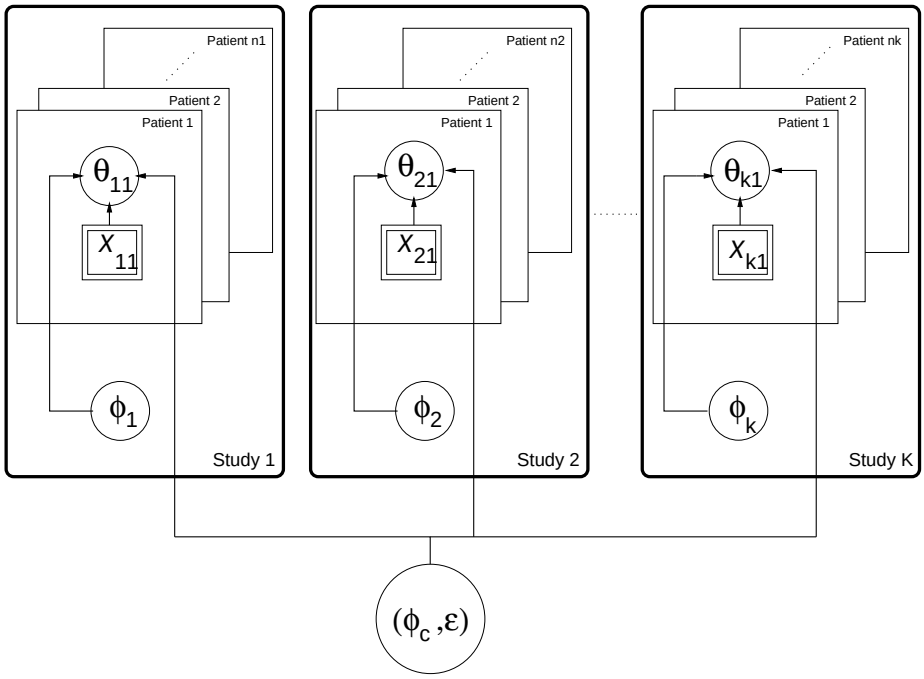


Figure 7.3: Graphical representation of the hierarchical model across related studies in expression (7.5). The vector of random-effects, θ_{ki}, parametrizes the mean profile for the i-th patient from the k-th study, while $\boldsymbol{x}_{ki}$ the patient's corresponding covariate measurements. $\boldsymbol{\phi}_k$ and $\boldsymbol{\phi}_c$ are study-specific and common hyperparameters (see equation 7.2).

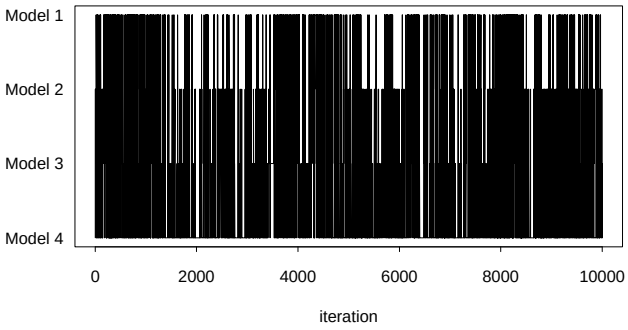


Figure 7.4: Trace plot for the RJMCMC algorithm. Model 1 through 4 represent, respectively, the following pairs (L_1, L_2): $(1,1), (1,2), (2,1)$ and $(2,2)$. See Table 7.1.

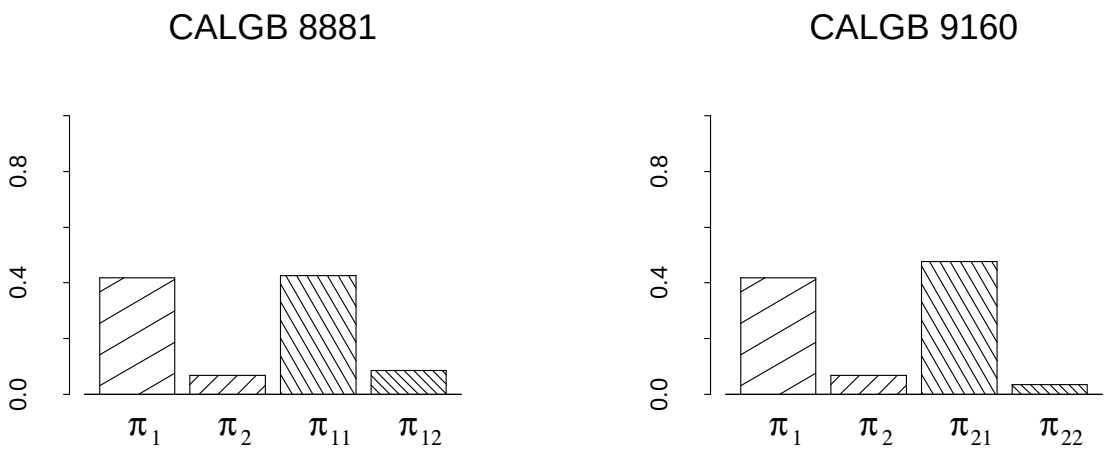


Figure 7.5: Weights of each component in the mixture for $\boldsymbol{\theta}_{ki}$, when fitting a model with $L_1 = L_2 = 2$. The posterior expectation for $(\pi_1, \pi_2, \pi_{11}, \pi_{12}, \pi_{21}, \pi_{22})$ is $(0.4185, 0.069, 0.4267, 0.0859, 0.4773, 0.0352)$.

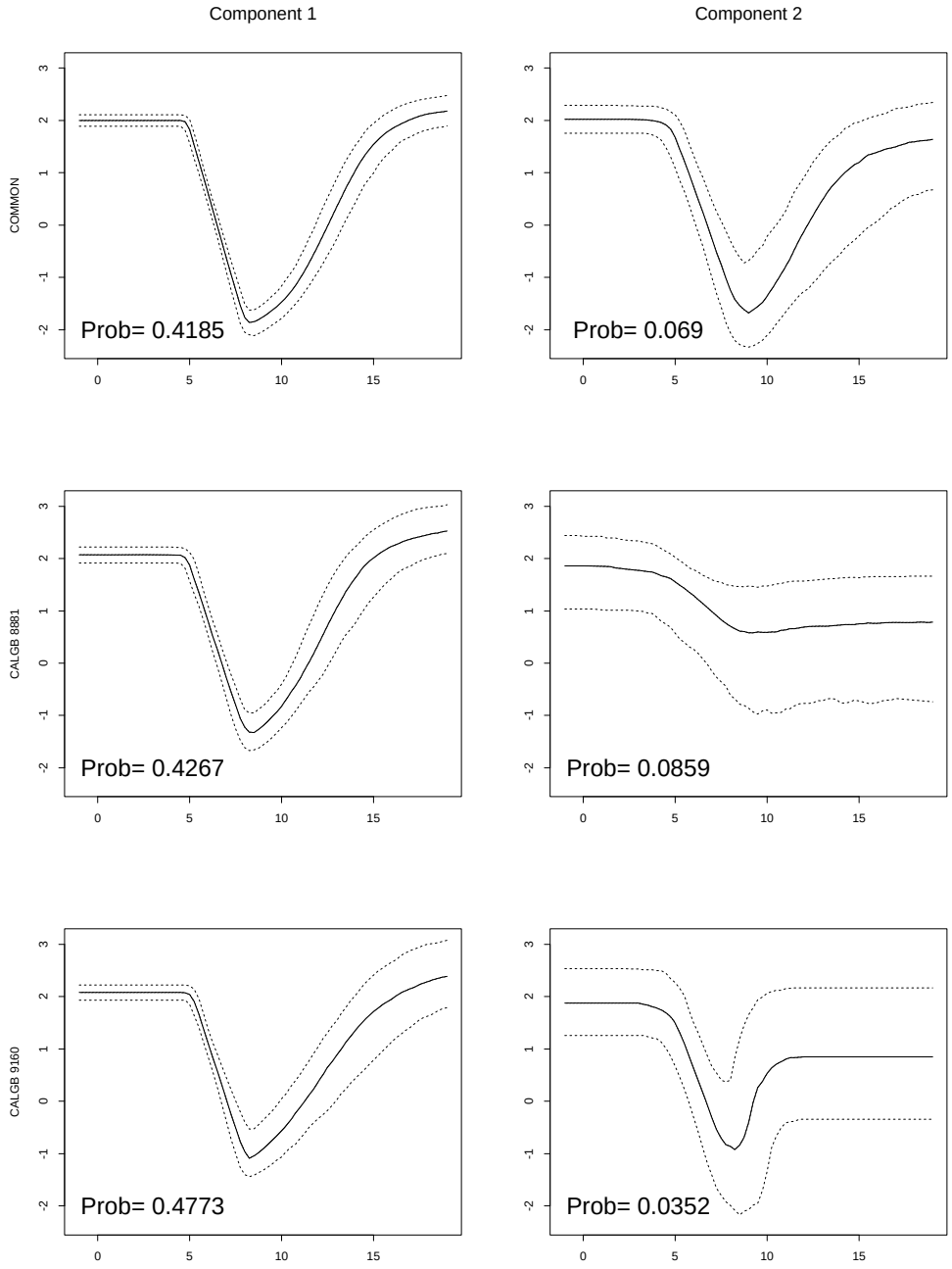


Figure 7.6: Posterior means and 90% credibility intervals for profiles as functions of the mixture locations, $\boldsymbol{\mu}, \boldsymbol{\mu} + \boldsymbol{d}_l$ and $\boldsymbol{\mu} + \boldsymbol{d}_{kl}$, for $l = 1, 2$ and $k = 1, 2$. The left column represents study 8881 ($k = 1$) and the right column represents study 9160 ($k = 2$).

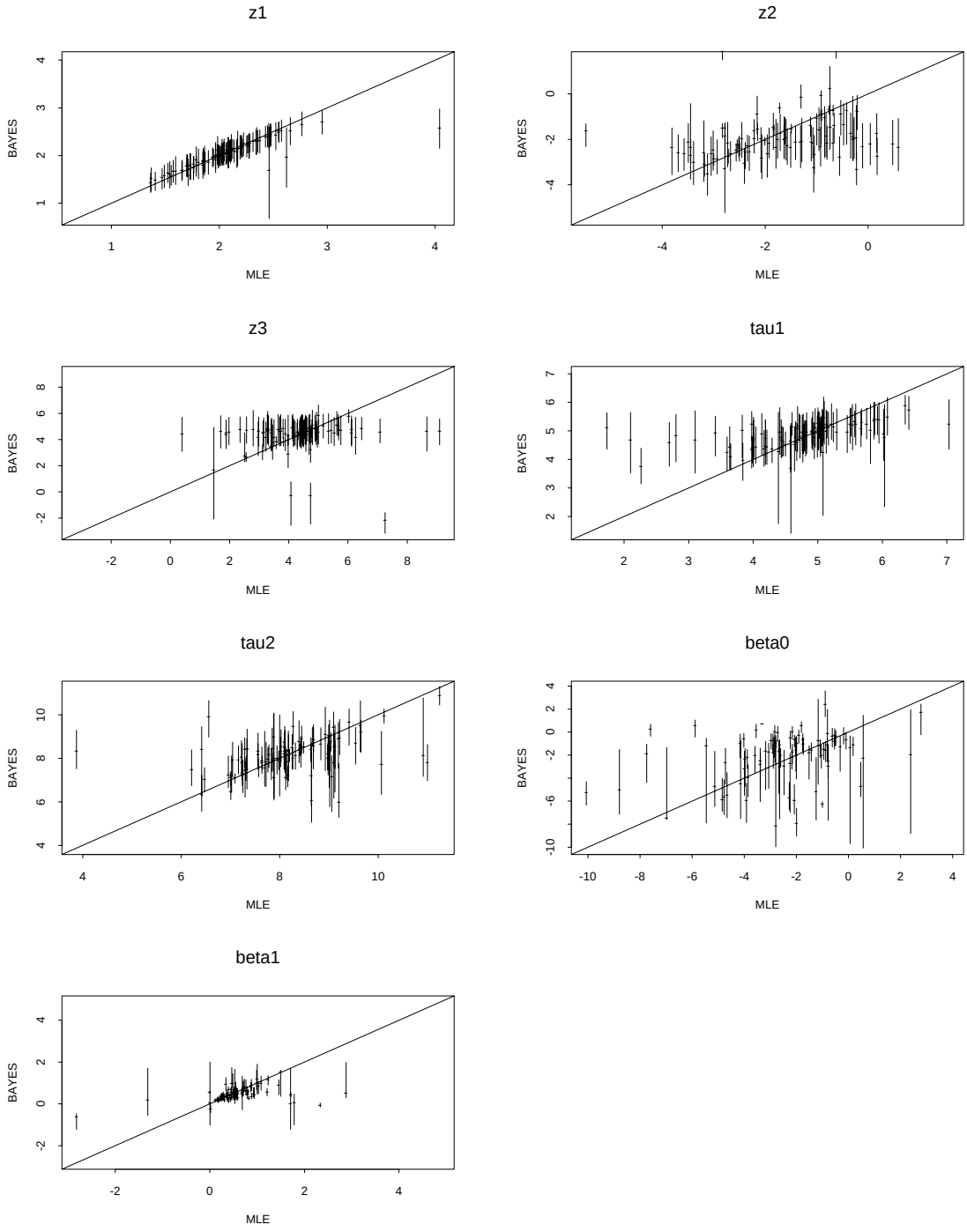


Figure 7.7: Plots of maximum likelihood estimators (obtained for each patient separately) against posterior means, 5th and 95th percentiles for the components of $\boldsymbol{\theta}$.

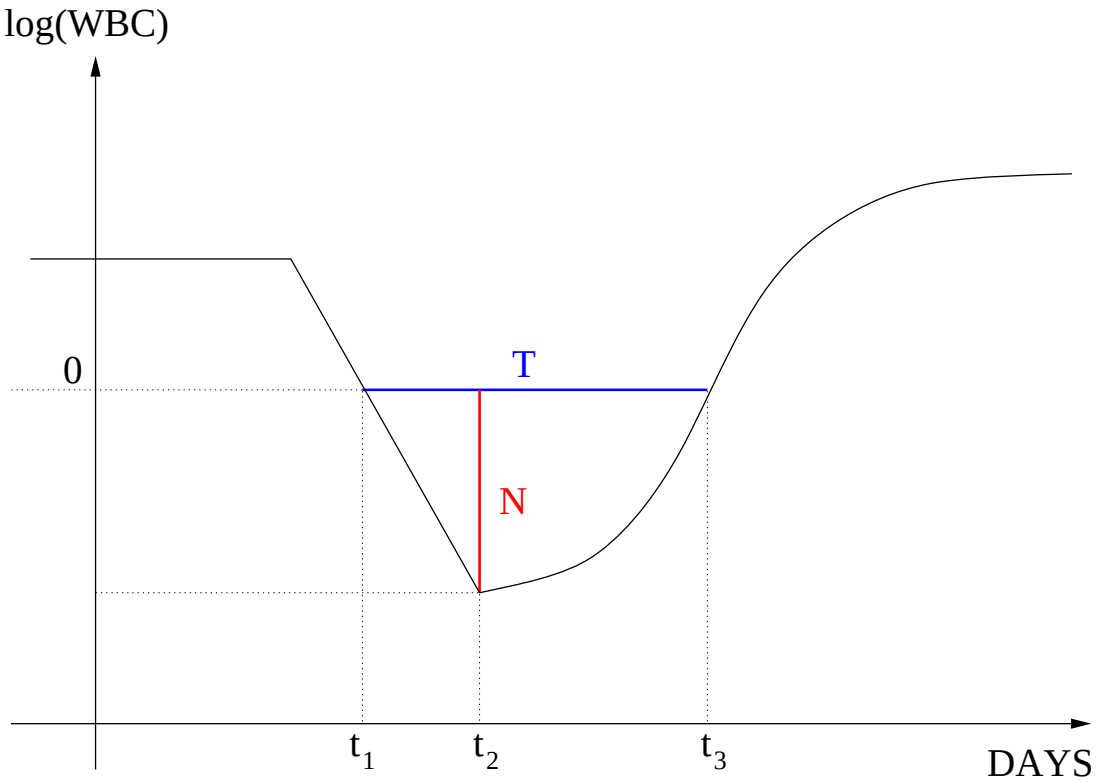
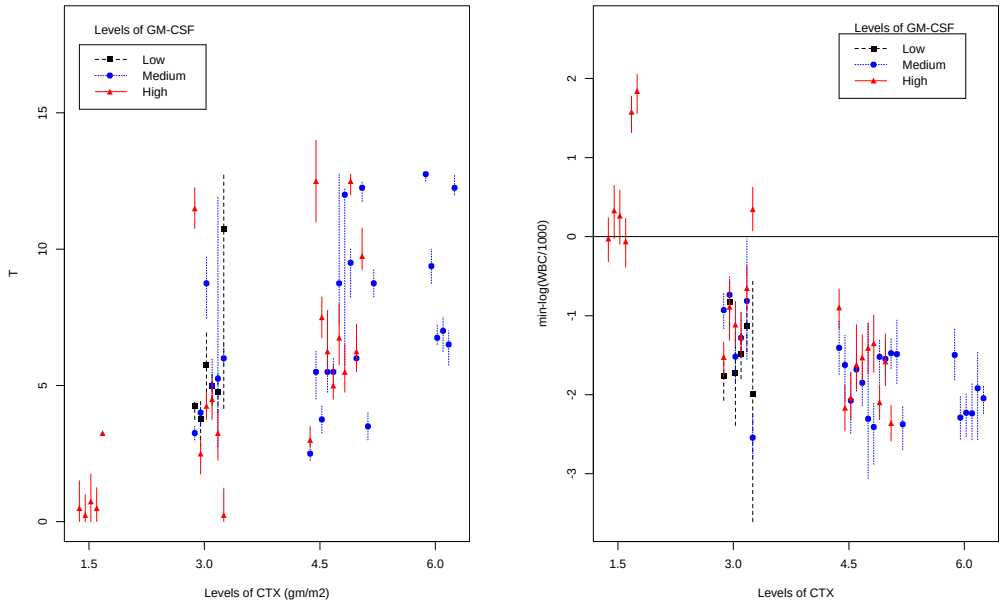


Figure 7.8: T represent the time length in which the patient has its white blood counts below 1000 (or $log(WBC/1000) < 0$). N measures how low one patients white blood counts can achieve. The events starting times, namely t_1 and t_2, are also of interest to decision makers.



(a) (b)

Figure 7.9: (a) Time length in which the patient's white blood counts was below 1000 (or $log(WBC/1000)$ was below zero). (b) Nadir: patient's minimum level of WBC (in log(WBC/1000) units). See Figure 7.8.

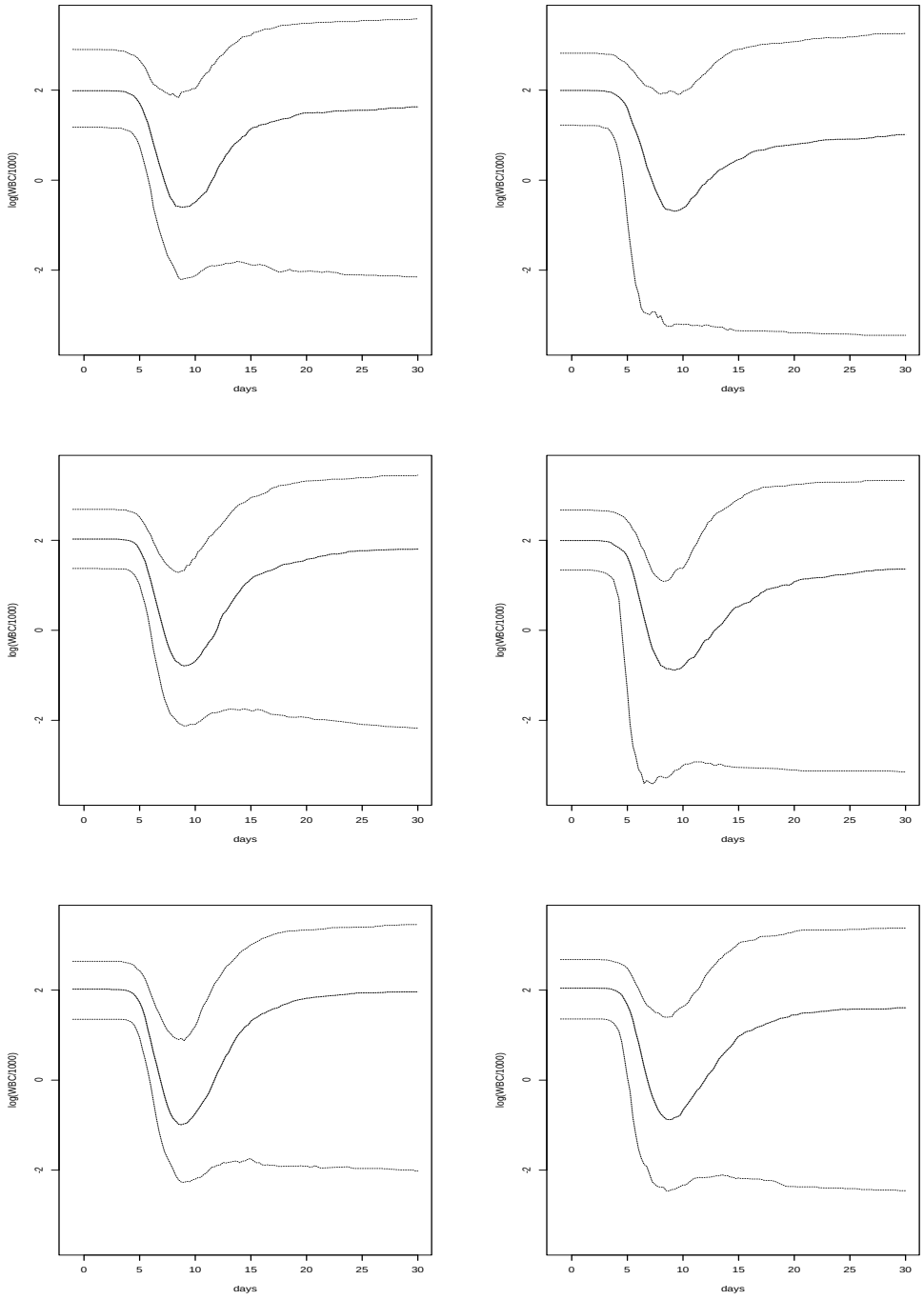


Figure 7.10: Predictive profile for new patients. Left column represents the best single model (Model 4: $L_1 = L_2 = 2$). The right column is Bayesian model averaging across the four competing models (see Figure 7.4). Rows are profiles from patients taken from studies CALGB 8881 (top row), CALGB 9160 (middle row) and from the population (bottom row), respectively. CTX=3.0mg/m^2 and medium level of GM-CSF.

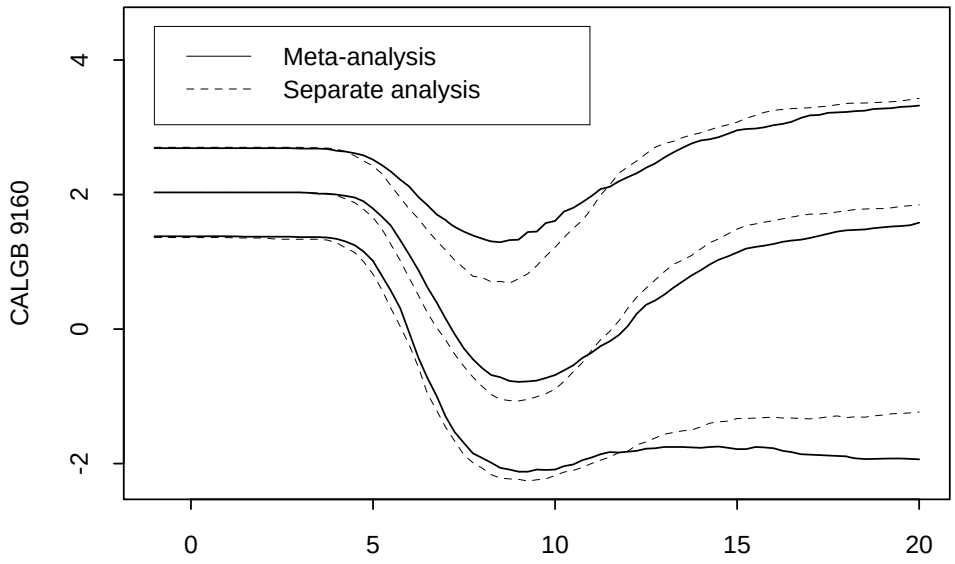


Figure 7.11: Predictive profile for new patients from study CALGB 9160. The solid lines represent posterior means, 5th and 95th percentiles based on our meta-analysis model ($L_1 = L_2 = 2$). The dashed lines represent the same percentiles for the alternative model where $Pr(\varepsilon = 0) = 1$ (see the last paragraph of Section 7.6 for further details). CTX=3.0mg/m^2 and GM-CSF=5$\mu g/kg$.

Chapter 8

Final comments and perspectives

As mentioned in the abstract and in the introduction, this thesis combines theoretical, methodological and empirical findings in multivariate Bayesian statistics. More specifically, the work is divided in three major parts: (*i*) modeling uncertainty in factor analysis, (*ii*) modeling and sequentially analysing factor models with stochastic variance components, and (*iii*) modeling longitudinal data models using multivariate mixture priors.

Summary sections are provided in each chapter. They are mainly designed to highlight and elucidate the theoretical and empirical nuances of the problem at hand. The following general comments indicate possible generalizations and open issues in each of the three sections.

8.1 Model uncertainty in factor models

Chapters 2, 3 and 4 dealt basically with factor analysis' oldest problem of choosing the number of common factors. Markov chain Monte Carlo is relatively simple when fixing the number of factors. We reviewed the literature on the subject and we proposed a RJMCMC algorithm. We compared our methodology with alternative approaches, most of which tackled the problem of computing normalizing constants

or ratios of normalizing constants. Factor models are used, routinely and most of the time automatically, by scientific researchers and practitioners in many fields. However, issues concerning prior information and its influence in the analysis certainly provides a barrier to the wider use of the Bayesian paradigm more often. Hence, in Chapter 4, we assess model uncertainty to the number of common factors when prior information is lacking. Additional comments are given below:

- As we have said, inference is straightforward when the number of common factors is kept fixed. However, posterior inference becomes extremely difficult when the number of factor is larger than necessary, i.e. when the model is overparametrized. Posterior multimodality is one of the problems we have encountered in our empirical analysis. Sections 2.3 presents some theoretical arguments that help to explain this problem, which is basically due to the apparent rank deficiency of the factor loading matrix.

- Also related to the previous discussion is the question of how to incorporate this inadequacy in the prior for the factor loadings and idiosyncratic variances. In our applications we have studied sensitivity of the posterior to the choice of the prior hyperparameters. In most cases, and with datasets, as small as fifty observations, the effect of the hyperparameters were negligible or immaterial. However, we believe that further research should be undertaken assess prior information in factor models to more intelligently.

- Except in the case of posterior multimodality caused by likelihood unidentifiability, prior issues are not the major problems in factor models with fixed numbers of factors. The same cannot be said when it comes to models of different orders. In Chapter 4 we have learned, from simulation studies, that model choice is not sensitive to how vague the prior distributions for the factor loadings and idiosyncratic variances are. Contrarily, the expected posterior prior

approach seemed to be much more robust and the results relatively satisfactory. A question that remains unclear and unanswered is how to choose the training measure m^* and what is the final shape of the prior. We anticipate more interest in those questions in the near future, for factor models are extensively used and software for automatic Bayesian procedures will be developed.

- Our RJMCMC algorithm worked relatively well in analyses of all simulation and real datasets. For small numbers of alternative models, we have also found, as expected, that the BIC generally provided robust and reliable initial guides to the choice of the number of factors. We attribute the success of the RJMCMC to the fact that we carefully choose the proposal densities to be as close as possible to the posterior distribution under each competing model. As stressed in the conclusions of Chapter 3, we have found that the other "standard" methods of approximation are not consistently accurate in identifying the correct models in ranges of simulation studies, except for the Newton and Raftery, the bridge sampling and the Laplace-Metropolis estimators. Among the alternative methods the Laplace-Metropolis is relatively simple to compute.

8.2 Factor models with stochastic variance components

Chapters 5 and 6 presented new developments in factor analysis with covariance structures in certain classes of multivariate stochastic volatility models. In Chapter 5 we extended existing works in some theoretical and practical directions and our main contributions consisted of (*i*) allowing the factor loadings to evolve in time, and (*ii*) estimating sequentially the parameters of the model, as opposed to fixing them. In order to perform estimation and forecasting in the generalized models, new MCMC algorithms were proposed or existing ones adapted.

- One of the most important empirical results, we believe, is the relationship between the number of factors, and therefore, common stochastic variances, and whether or not the idiosyncratic variances are modeled by stochastic volatility models. We have found the rather intuitive result that more factors are necessary when the idiosyncratic variances does not follow stochastic volatility models. Those extra factors basically influenced just one of the time series involved in the analysis, therefore using the factor model without its most appealing characteristic, which is to explain co-movements between the time series.

- Also related to the previous comment is the fact that an indeterminacy may occur when one of the common factors influences just one of the time series and the idiosyncratic variances follow stochastic volatility processes. In this case it is impossible to identify in the likelihood which term is the common factor and which is the idiosyncratic factor and the posterior distribution of the parameters related to those processes are bound to exhibit multimodality. Possible solutions are to restrict the prior distributions to avoid such indeterminacy and assign higher prior probability to models with lower number of common factors.

- In our applications of the simulation-based sequential analysis, we found the corroborative fact that models with lower numbers of factors have higher one-step ahead posterior probabilities and build portfolios with higher one-step ahead cumulative returns then models with higher number of common factors. Even though we found this only in our applications, we believe that it can be generalized to other financial time series datasets.

- More complex structures for the factor loadings, such as higher order autoregressions, are possible extensions of the models we have studied. Also, the loadings discount factor, which plays an important role in the estimation pro-

cess, deserves more attention. For instance, replacing the discount factor by a constant variance matrix for the factor loadings would be an alternative solution, even though the discount factor has a natural appeal in expressing the amount of information retained by the model through time.

- When treating the parameters in the sequential analysis, the tuning quantities δ, a and h, that govern the posterior-to-prior evolution (section 6.4) play important roles. Another open question is how to transform nonnegative parameters in the mixture of normal used in the kernel approximation for the posterior of the parameters.

8.3 Longitudinal data models

In chapter 7, we propose a new class of longitudinal data models where random effects are replaced by random measures represented by a flexible mixture of multivariate normals in a way that allows us to implement meta-analysis over related studies. The number of terms in the mixture models is random and a novel RJMCMC sampler is proposed. We apply our methodology to two cancer studies. Some of our main findings and open issues are as follows:

- The idea of splitting the source of variability into common and specific is not new. However, our formulation is innovative in the sense that instead of random effects we have random measures.

- Since we are modeling the random measure through a multivariate mixtures of normals, we face problems related to identifiability issues. Unfortunately, general solutions and guidance do not exist, as far as we know. Additional identification problems appear in our context, where there is a mixture of two mixtures: the first representing the measure common across studies and the

second representing the measure specific to each particular study. In section 7.3.1 we fully discuss these identification problems. An alternative approach would be to build the hierarchical structure on the weights instead of the normal locations. We believe that in this way we would be better able to control identifiability. This is current research.

- We consider Bayesian model averaging. The averaging is across a rather small number of competing models and a RJMCMC sampler is designed to jump across those models. The RJMCMC sampler mixes well and the estimated posterior model probabilities are used for Bayesian model averaging. As expected, the profiles for new patients have larger posterior bounds in averaged model than in the model with the highest posterior probability.

Appendix A

Univariate and multivariate distributions

In this appendix we present some of the main univariate and multivariate distributions used within the main body of this thesis. Even though all of them can be found in basic/intermediate statistical textbook, we believe that restating then will be profitable for the sake of uniformity since there are many available parametrizations out there. References include Zellner (1971),Johnson and Kotz (1972), Dawid (1981),Press (1982), Drèze and Richard (1983) and West and Harrison (1997), among many others. We tried to follow very closely the notation provided by Gelman *et al.* (1995).

A.1 Univariate distributions

In this section define the univariate distributions we have used throughout the thesis.

[**Normal**]

A random variable y is said to follow a normal distribution with parameters μ and σ^2 if and only if its probability density function is,

$$p(y|\mu,\sigma^2) = (2\pi\sigma^2)^{-1/2} exp\left\{-\frac{(y-\mu)^2}{2\sigma^2}\right\} \qquad y \in \Re \tag{A.1}$$

- $E(y) = \mu \in \Re$;
- $V(y) = \sigma^2 > 0$;
- Notation : $y \sim N(\mu, \sigma^2)$

[**Truncated-normal**]

A positive random variable y is said to follow a truncated-normal distribution at zero with parameters μ and σ^2 if and only if its probability density function is,

$$p(y|\mu, \sigma^2) = (2\pi\sigma^2)^{-1/2} exp\left\{-\frac{(y-\mu)^2}{2\sigma^2}\right\} / \Phi(\mu/\sigma) \qquad y > 0 \tag{A.2}$$

where $\Phi(z) = Pr(Z < z)$ and $Z \sim N(0, 1)$.

- $E(y) = \mu + \frac{\sigma^2}{\sqrt{2\pi\sigma^2}} exp\left\{-\mu^2/(2\sigma^2)\right\} / \Phi(\mu/\sigma)$;
- Notation : $y \sim TN(\mu, \sigma^2)$

[**Gamma**]

A random variable y is said to follow a gamma distribution with parameters α and β if and only if its probability density function is,

$$p(y|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} exp\{-\beta y\} \qquad 0 < y, \alpha, \beta \tag{A.3}$$

- $E(y) = \frac{\alpha}{\beta}$;
- $V(y) = \frac{\alpha}{\beta^2}$;
- Notation : $y \sim G(\alpha, \beta)$

[**Inverse-gamma**]

A random variable y is said to follow an inverse-gamma distribution with parameters α and β if and only if its probability density function is,

$$p(y|\alpha,\beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} y^{-(\alpha+1)} exp\{-\beta/y\} \qquad 0 < y, \alpha, \beta \tag{A.4}$$

- $E(y) = \frac{\beta}{\alpha-1}$, for $\alpha > 1$;
- $V(y) = \frac{\beta^2}{(\alpha-1)^2(\alpha-2)}$, for $\alpha > 2$;
- Notation : $y \sim IG(\alpha, \beta)$

[**Beta**]

A random variable y is said to follow a beta distribution with parameters α and β if and only if its probability density function is,

$$p(y|\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} y^{\alpha-1}(1-y)^{\beta-1} \qquad 0 \leq y \leq 1, \alpha > 0, \beta > 0 \tag{A.5}$$

- $E(y) = \frac{\alpha}{\alpha+\beta}$;
- $V(y) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$;
- Notation : $y \sim Be(\alpha, \beta)$

A.2 Multivariate distributions

In this section define the multivariate and matricvariate distributions we have used throughout the thesis.

[**Multivarite normal**]

A n-dimensional random vector $\boldsymbol{y}$ is said to follow a multivariate normal distribution with parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ if and only if its probability density function is,

$$p(\boldsymbol{y}|\boldsymbol{\mu},\boldsymbol{\Sigma}) = k \mid \boldsymbol{\Sigma} \mid^{-1/2} exp\left\{-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\} \qquad \boldsymbol{y} \in \Re^n \tag{A.6}$$

where $k = (2\pi)^{-n/2}$.

- $E(\boldsymbol{y}) = \boldsymbol{\mu} \in \Re^n$;
- $V(\boldsymbol{y}) = \boldsymbol{\Sigma}$ is symmetric and positive definite;
- Notation : $\boldsymbol{y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

[**Multinomial**]

A n-dimensional random vector $\boldsymbol{y}$ is said to follow a multinomial distribution with parameters t and $\boldsymbol{\pi}$ if and only if its probability function is,

$$p(y|t,\boldsymbol{\pi}) = \frac{t!}{\prod_{i=1}^{n} y_i!} \prod_{i=1}^{n} \pi_i^{y_i} \tag{A.7}$$

for $\boldsymbol{y} = (y_1,\ldots,y_n)'$ in $\mathcal{Y} = \{\boldsymbol{y} : \boldsymbol{y}'\mathbf{1} = t,\ 0 \leq y_i \leq n\ i = 1,\ldots,n\}$ and $\boldsymbol{\pi} = (\pi_1,\ldots,\pi_n)'$ in $\Pi = \{\boldsymbol{\pi} : \boldsymbol{\pi}'\mathbf{1} = 1,\ 0 \leq \pi_i \leq n\ i = 1,\ldots,n\}$.

- $E(y_i) = n\pi_i$;
- $V(y_i) = n\pi_i(1-\pi_i)$;
- $Cov(y_i, y_j) = -n\pi_i\pi_j$;
- Notation : $\boldsymbol{y} \sim M(n, \boldsymbol{\pi})$

[**Dirichlet**]

A n-dimensional random vector $\boldsymbol{y}$ is said to follow a Dirichlet distribution with parameter $\boldsymbol{\alpha}$ if and only if its probability density function is,

$$p(\boldsymbol{y}|\boldsymbol{\alpha}) = \frac{\Gamma(\alpha_0)}{\prod_{i=1}^{n}\Gamma(\alpha_i)}\prod_{i=1}^{n} y_i^{\alpha_i - 1} \tag{A.8}$$

for $\boldsymbol{y} = (y_1, \ldots, y_n)'$ in $\mathcal{Y} = \{\boldsymbol{y} : \boldsymbol{y}'\mathbf{1} = 1, y_i \geq 0 \ i = 1, \ldots, n\}$ and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)'$ in $\{\boldsymbol{\alpha} : \boldsymbol{\alpha}'\mathbf{1} = \alpha_0, \ \alpha_i > 0 \ i = 1, \ldots, n\}$.

- $E(y_i) = \alpha_i/\alpha_0$;
- $V(y_i) = \frac{\alpha_i(\alpha_0 - \alpha_i)}{\alpha_0^2(\alpha_0+1)}$;
- $Cov(y_i, y_j) = -\frac{\alpha_i\alpha_j}{\alpha_0^2(\alpha_0+1)}$;
- Notation : $\boldsymbol{y} \sim Dir(\boldsymbol{\alpha})$

[**Matrix variate normal**]

A $T \times n$ random matrix $\boldsymbol{y}$ is said to follow a matricvariate normal distribution with parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Omega}$ and $\boldsymbol{\Sigma}$ if and only if its probability density function is,

$$p(\boldsymbol{y} \mid \boldsymbol{\mu}, \boldsymbol{\Omega}, \boldsymbol{\Sigma}) = [(2\pi)^{Tn} \mid \boldsymbol{\Sigma} \mid^{T} \mid \boldsymbol{\Omega} \mid^{n}]^{-1/2} exp\{-\frac{1}{2}\mathrm{tr}\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{\mu})'\boldsymbol{\Omega}^{-1}(\boldsymbol{y} - \boldsymbol{\mu})\} \tag{A.9}$$

where $\boldsymbol{\mu}$ is a $(T \times n)$ matrix and $\boldsymbol{\Omega}$ and $\boldsymbol{\Sigma}$ are $(T \times T)$ and $(n \times n)$ symmetric and definite positive matrices, respectively.

- $E(\boldsymbol{y}) = \boldsymbol{\mu} \in \Re^{T \times n}$;
- $V(vec(\boldsymbol{y})) = \boldsymbol{\Sigma} \otimes \boldsymbol{\Omega}$, where $vec(\boldsymbol{y})$ is the TN-dimensional vector obtained after stacking the columns of $\boldsymbol{y}$;

- $\boldsymbol{\Omega}$ is the matrix that describes the covariance among the columns of $\boldsymbol{y}$, while $\boldsymbol{\Sigma}$ is the matrix that describes the covariance among its rows;
- Defining $\boldsymbol{y} = (\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_n) = (\boldsymbol{y}_{(1)}, \boldsymbol{y}_{(2)}, \cdots, \boldsymbol{y}_{(T)})'$, where y_i's are T-dimensional vectors and $y_{(i)}$'s are n-dimensional vectors, it can be shonw that $E(\boldsymbol{y}_i \boldsymbol{y}_j') = \sigma_{ij}\boldsymbol{\Omega}$ and $E(\boldsymbol{y}_{(i)} \boldsymbol{y}_{(j)}') = \omega_{ij}\boldsymbol{\Sigma}$;
- Notation: $\boldsymbol{y} \sim N(\boldsymbol{\Theta}, \boldsymbol{\Omega}, \boldsymbol{\Sigma})$.

[**Wishart**]

A $n \times n$ symmetric and positive definite random matrix $\boldsymbol{\Omega}$ is said to follow Wishart distribution with parameters ν and $\boldsymbol{S}$ if and only if its probability density function is,

$$p(\boldsymbol{\Omega}|\nu, \boldsymbol{S}) = k \mid \boldsymbol{\Omega} \mid^{(\nu-n-1)/2} exp\left\{-\frac{1}{2} tr \boldsymbol{S}^{-1}\boldsymbol{\Omega}\right\} \tag{A.10}$$

where $k^{-1} = 2^{\nu n/2}\pi^{n(n-1)/4}\prod_{i=1}^{n}\Gamma[(\nu+1-i)/2] \mid \boldsymbol{S} \mid^{\nu/2}$.

- $\nu \geq n$ is the number of degrees of freedom;
- $\boldsymbol{S}$ is a $n \times n$ symmetric and positive definite scale matrix;
- Notation: $\boldsymbol{\Omega} \sim W(\nu, \boldsymbol{S})$;
- $E(\boldsymbol{\Omega}) = \nu\boldsymbol{S}$;
- $V(\omega_{ij}) = \nu(s_{ij}^2 + s_{ii}s_{jj})$;
- $COV(\omega_{ij}, \omega_{kl}) = \nu(s_{ik}s_{jl} + s_{il}s_{jk})$
- If $n = 1$, then $\omega \sim G(\nu/2, S^{-1}/2)$.

[**Inverse-Wishart**]

A $n \times n$ symmetric and positive definite random matrix $\mathbf{\Sigma}$ is said to follow Wishart distribution with parameters ν and $\boldsymbol{Q}$ if and only if its probability density function is,

$$p(\mathbf{\Sigma}|\nu, \boldsymbol{Q}) = k \mid \mathbf{\Sigma} \mid^{(\nu+n+1)/2} exp\left\{-\frac{1}{2} tr \mathbf{\Sigma}^{-1} \boldsymbol{Q}\right\} \tag{A.11}$$

where $k^{-1} = 2^{\nu n/2} \pi^{n(n-1)/4} \prod_{i=1}^{n} \Gamma[(\nu + 1 - i)/2] \mid \boldsymbol{Q} \mid^{-\nu/2}$.

- $\nu \geq n$ is the number of degrees of freedom;
- $\boldsymbol{Q}$ is a $n \times n$ symmetric and positive definite scale matrix;
- Notation: $\mathbf{\Sigma} \sim IW(\nu, \boldsymbol{Q})$;
- $E(\mathbf{\Sigma}) = \frac{1}{\nu-(n+1)} \boldsymbol{Q}$, for $\nu > n + 1$;
- If $n = 1$, then $\sigma^2 \sim IG(\nu/2, Q/2)$.

Appendix B

Reversible jump MCMC and metropolized Carlin and Chib

In this appendix we present the Reversible Jump algorithm as introduced in Green (1995). Among many others, Richardson and Green (1997), Dellaportas *et al.* (1998), Denison *et al.* (1997), Troughton and Godsill (1997), Insua and Müller (1998), Barbieri and O'Hagan (1996) and Huerta and West (1999) applied the reversible jump sampler to mixture models, variable selection, curve fitting, autoregressive models, neural networks, ARMA models and component structure in AR models, respectively.

We also explore its relationship to Carlin and Chib's (1995) pseudo-prior method. Particular attention is given to the *Metropolized Carlin-Chib* algorithm simultaneously introduced by Dellaportas *et al.* (1998) and Godsill (1998).

The results presented here are mainly based on the developments from Dellaportas *et al.* (1998) and Godsill (1998). Additional overview and/or further extensions can be found in Chen *et al.* (2000), Section 9.5, and Gamerman (1997), Section 7.3.

B.1 The RJMCMC algorithm

Suppose that the competing models can be enumerable and are represented by the set $\mathcal{M} = \{\mathcal{M}_1, \mathcal{M}_2, \ldots\}$. Under model $\mathcal{M}_k$, the posterior distribution is

$$p(\boldsymbol{\theta}_k|\boldsymbol{y}, k) \propto p(\boldsymbol{y}|\boldsymbol{\theta}_k, k)p(\boldsymbol{\theta}_k|k) \tag{B.1}$$

where $p(\boldsymbol{y}|\boldsymbol{\theta}_k, k)$ and $p(\boldsymbol{\theta}_k|k)$ represent the probability model and the prior distribution of the parameters of model $\mathcal{M}_k$, respectively. Then,

$$p(\boldsymbol{\theta}_k, k|\boldsymbol{y}) \propto p(k)p(\boldsymbol{\theta}_k|k, \boldsymbol{y}) \tag{B.2}$$

The RJMCMC methods involve Metropolis-Hastings type algorithms that move a simulation analysis between models defined by $(k, \boldsymbol{\theta}_k)$ to $(k', \boldsymbol{\theta}_{k'})$ with different defining dimensions k and k'. The resulting Markov chain simulations jump between such distinct models and form samples from the joint distribution $p(\boldsymbol{\theta}_k, k)$. The algorithm are designed to be reversible so as to maintain detailed balance of a irreducible and aperiodic chain that converges to the correct target measure. Further details of the general methodology and ideas can be found in Green (1995).

Here we present the algorithm in a schematic form. If the current state of the Markov chain is $(k, \boldsymbol{\theta}_k)$, then one possible version of the RJMCMC algorithm is as follows:

Step 1. Propose a visit to model $\mathcal{M}_{k'}$ with probability $J(k \to k')$.

Step 2. Sample $\boldsymbol{u}$ from a proposal density $q(\boldsymbol{u}|\boldsymbol{\theta}_k, k, k')$.

Step 3. Set $(\boldsymbol{\theta}_{k'}, \boldsymbol{u}') = g_{k,k'}(\boldsymbol{\theta}_k, \boldsymbol{u})$, where $g_{k,k'}(\cdot)$ is a bijection between $(\boldsymbol{\theta}_k, \boldsymbol{u})$ and $(\boldsymbol{\theta}_{k'}, \boldsymbol{u}')$, where $\boldsymbol{u}$ and $\boldsymbol{u}'$ play the role of matching the dimensions of both vectors.

Step 4. The acceptance probability of the new model, $(\boldsymbol{\theta}_{k'}, k')$ can be calculated as the minimum between one and

$$\underbrace{\frac{p(\boldsymbol{y}|\boldsymbol{\theta}_{k'}, k')p(\boldsymbol{\theta}_{k'})p(k')}{p(\boldsymbol{y}|\boldsymbol{\theta}_k, k)p(\boldsymbol{\theta}_k)p(k)}}_{\text{model ratio}} \underbrace{\frac{J(k' \to k)q(\boldsymbol{u}'|\boldsymbol{\theta}_{k'}, k', k)}{J(k \to k')q(\boldsymbol{u}|\boldsymbol{\theta}_k, k, k')} \left|\frac{\partial g_{k,k'}(\boldsymbol{\theta}_k, \boldsymbol{u})}{\partial(\boldsymbol{\theta}_k, \boldsymbol{u})}\right|}_{\text{proposal ratio}} \tag{B.3}$$

Looping through steps 1-4 generates a sample $\{k_l, l = 1, \ldots, L\}$ for the model indicators and $Pr(k|\boldsymbol{y})$ can be estimated by

$$\widehat{Pr}(k|\boldsymbol{y}) = \frac{1}{L}\sum_{l=1}^{L} \mathbf{1}_k(k_l) \tag{B.4}$$

where $\mathbf{1}_k(k_l) = 1$ if $k = k_l$ and zero otherwise. The choice of the model proposal probabilities, $J(k \to k')$, and the proposal densities, $q(\boldsymbol{u}|k, \boldsymbol{\theta}_k, k')$, must be cautiously made, especially in highly parameterized problems.

Independent sampler: If all parameters of the proposed model are generated from the proposal distribution, then $(\boldsymbol{\theta}_{k'}, \boldsymbol{u}') = (\boldsymbol{u}, \boldsymbol{\theta}_k)$ and the Jacobian in (B.3) is one. We implement an independent RJMCMC sampler for the factor model from Chapter 3. There, the proposal density $q(\boldsymbol{u}|\boldsymbol{\theta}_k, k, k')$ is reduced to suitably chosen $q_{k'}(\boldsymbol{\theta}_{k'})$ (equation 3.1 from Section 3.2).

Standard Metropolis-Hastings: When the proposed model k' equals the current model k, the loop through steps 1-4 corresponds to the traditional Metropolis-Hastings algorithm (Metropolis *et al.*, 1995; Hastings, 1970; Peskun, 1973; Chib and Greenberg, 1995).

Posterior densities as proposal densities: If $p(\boldsymbol{\theta}_k|\boldsymbol{y}, k)$ is available in close form for each model $\mathcal{M}_k$, then $q(\boldsymbol{u}'|\boldsymbol{\theta}_{k'}, k', k) = p(\boldsymbol{\theta}_k|\boldsymbol{y}, k)$ and the acceptance probability (equation B.3) reduces to the minimum between one and

$$\frac{p(k')p(\boldsymbol{y}|k')}{p(k)p(\boldsymbol{y}|k)} \frac{J(k' \to k)}{J(k \to k')} \tag{B.5}$$

using the fact that $p(\boldsymbol{y}|\boldsymbol{\theta}_k, k)p(\boldsymbol{\theta}_k)p(k) = p(\boldsymbol{\theta}_k, k|\boldsymbol{y})p(\boldsymbol{y}|k)$. Again, the Jacobian equals one. The predictive density or normalizing constant, $p(\boldsymbol{y}|k)$, is also available in close form. Moreover, if $J(k' \to k) = J(k \to k')$, the acceptance probability is the minimum between one and the posterior odds ratio from model $\mathcal{M}_{k'}$ to model $\mathcal{M}_k$, that is the move is automatically accepted when model $\mathcal{M}_{k'}$ has higher posterior probability than model $\mathcal{M}_k$; otherwise the posterior odds ratio determines how likely is to move to a lower posterior probability model.

B.2 Metropolized Carlin and Chib's algorithm

Let $\boldsymbol{\Theta} = (\boldsymbol{\theta}_k, \boldsymbol{\theta}_{-k})$ be the vector containing the parameters of all competing models. Then the joint posterior of $(\boldsymbol{\Theta}, k)$ is

$$p(\boldsymbol{\Theta}, k|\boldsymbol{y}) \propto p(k)p(\boldsymbol{y}|\boldsymbol{\theta}_k, k)p(\boldsymbol{\theta}_k|k)p(\boldsymbol{\theta}_{-k}|\boldsymbol{\theta}_k, k) \tag{B.6}$$

where $p(\boldsymbol{\theta}_{-k}|\boldsymbol{\theta}_k, k)$ are *pseudo-prior* densities (Carlin and Chib, 1995). Carlin and Chib propose a Gibbs sampler where the full posterior conditional distributions are

$$p(\boldsymbol{\theta}_k|\boldsymbol{y}, k, \boldsymbol{\theta}_{-k}) \propto \begin{cases} p(\boldsymbol{y}|\boldsymbol{\theta}_k, k)p(\boldsymbol{\theta}_k|k) & \text{if } k = k' \\ p(\boldsymbol{\theta}_k|k') & \text{if } k = k' \end{cases} \tag{B.7}$$

and

$$p(k|\boldsymbol{\Theta}, \boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{\theta}_k, k)p(k) \prod_{m \in \mathcal{M}} p(\boldsymbol{\theta}_m|k) \tag{B.8}$$

Notice that the pseudo-prior densities and the RJMCMC's proposal densities have similar functions. As a matter of fact, Carlin and Chib suggest using pseudo-prior distributions that are close to the posterior distributions within each competing model. This is basically what we do when implementing our RJMCMC sampler for the factor model (see Chapter 3).

The main problem with Carlin and Chib's Gibbs sampler is the need of evaluating and drawing from the pseudo-prior distributions at each iteration of the MCMC scheme. This problem can be overwhelmingly exacerbated in large situations where the number of competing models is relatively large (See Clyde, 1999, for applications and discussions in variable selection in regression models).

To overcome this last problem Dellaportas *et al.* and Godsill (1998) proposes "Metropolizing" Carlin and Chib's Gibbs sampler. If the current state of the Markov chain is at $(\boldsymbol{\theta}_k, k)$, then they suggest proposing and accepting/rejecting a move to a new model in the following way:

Step 1. Propose a new model $\mathcal{M}_{k'}$ with probability $J(k \to k')$.

Step 2. Generate $\boldsymbol{\theta}_{k'}$ from the pseudo-prior $p(\boldsymbol{\theta}_{k'}|k)$.

Step 3. The acceptance probability of the new model, k' can be calculated as the minimum between one and

$$\frac{p(\boldsymbol{y}|\boldsymbol{\theta}_{k'}, k')p(k')J(k' \to k)\prod_{m\in\mathcal{M}} p(\boldsymbol{\theta}_m|k')}{p(\boldsymbol{y}|\boldsymbol{\theta}_k, k)p(k)J(k \to k')\prod_{m\in\mathcal{M}} p(\boldsymbol{\theta}_m|k)}$$

which can be simplified to

$$\frac{p(\boldsymbol{y}|\boldsymbol{\theta}_{k'}, k')p(k')J(k' \to k)p(\boldsymbol{\theta}_{k'}|k')p(\boldsymbol{\theta}_k|k')}{p(\boldsymbol{y}|\boldsymbol{\theta}_k, k)p(k)J(k \to k')p(\boldsymbol{\theta}_k|k)p(\boldsymbol{\theta}_{k'}|k)} \tag{B.9}$$

since the other pseudo-prior densities cancel out.

Once again, if $p(\boldsymbol{\theta}_k|\boldsymbol{y}, k)$ is available in close form for each model $\mathcal{M}_k$, and $p(\boldsymbol{\theta}_k|k') = p(\boldsymbol{\theta}_k|\boldsymbol{y}, k)$, then the acceptance probability in (B.9) reduces to (B.5). As we have mentioned earlier the pseudo-prior densities and the RJMCMC's proposal densities have similar functions and the closer their are to the competing models' posterior probabilities the better the sampler mixing.

Appendix C

Mixture of normals

In Chapter 7 latent indicators where introduced in order to break the nonlinearity in the multivariate mixture prior (see equations 7.6, 7.7 and 7.8). Such simplication comes with a cost, i.e. it is likely that the sampled Z_{ki}'s will be highly correlated to the sampled $\boldsymbol{\mu}$'s causing poor mixing of the Markov chain and slowing down the convergence to the posterior distribution. In this section we elaborate more on the mixture of normal model and show to overcome such problem.

C.1 Model and prior setting

Initially, let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ be a random sample from a (multivariate) mixture of normals, i.e.

$$p(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\pi}, \boldsymbol{S}) = \sum_{k=1}^{K} \pi_k dN(\boldsymbol{x}|\boldsymbol{\mu}_k, \boldsymbol{S}) \tag{C.1}$$

where $dN(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{S})$ denotes the probability density function of a (multivariate) normal[1] with mean (vector) $\boldsymbol{\mu}$ and variance(-covariance matrix) $\boldsymbol{S}$, evaluate at the point $\boldsymbol{x}$. Also, $\boldsymbol{\mu} = (\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_K)$ and $\boldsymbol{\pi} = (\boldsymbol{\pi}_1, \ldots, \boldsymbol{\pi}_K)$. We will assume throughout this appendix that $\boldsymbol{\pi}$ and $\boldsymbol{S}$ are known quantities.

[1]See Section of Appendix A for further details about the multivariate normal distribution.

Traditionally, latent indicator variables z_i are included in the model for clarity in such a way that

$$Pr(z_i = j) = \pi_j \tag{C.2}$$

for all $i = 1, \ldots, n$ and $j = 1, \ldots, K$. Therefore, the likelihood in (C.1) can be rewritten as,

$$p(\boldsymbol{x}|\boldsymbol{z}, \boldsymbol{\mu}, \boldsymbol{S}) = \prod_{j=1}^{K} \prod_{i \in I_j} dN(\boldsymbol{x}_i|\boldsymbol{\mu}_j, \boldsymbol{S}) \tag{C.3}$$

where $\boldsymbol{x} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$, $\boldsymbol{z} = (z_1, \ldots, z_n)$, $I_j = \{i : z_i = j, i = 1, \ldots, n\}$ and $n_j = card(I_j)$. Consequently, when combining the likelihood in (C.3) with prior distributions for $\boldsymbol{z}$ and $\boldsymbol{\mu}_k$ ($\boldsymbol{z}|\boldsymbol{\pi} \sim M(K, \boldsymbol{\pi})$ and $\boldsymbol{\mu}_k \sim N(\boldsymbol{\mu}_{k0}, \boldsymbol{V}_{k0})$, respectively), the joint posterior for $\boldsymbol{\mu}$ and $\boldsymbol{z}$ is,

$$p(\boldsymbol{\mu}, \boldsymbol{z}|\boldsymbol{x}, \boldsymbol{\pi}, \boldsymbol{S}) \propto \left[\prod_{j=1}^{K} \prod_{i \in I_j} dN(\boldsymbol{x}_i|\boldsymbol{\mu}_j, \boldsymbol{S})\right] \left[\prod_{j=1}^{K} dN(\boldsymbol{\mu}_j|\boldsymbol{\mu}_{j0}, \boldsymbol{V}_{j0})\right] \tag{C.4}$$

Obviously, analytically tratable posterior inference is impossible. Next, we present the full conditional distributions $\boldsymbol{z}$ and $\boldsymbol{\mu}$ to be used in the Gibbs sampler algorithm.

C.2 Full conditionals

Full conditionals for $\boldsymbol{\mu}$ and $\boldsymbol{z}$ are given below. Notice that the full conditional for the elements of $\boldsymbol{z}$ are not conditional on $\boldsymbol{\mu}$. We believe that this strategy improves the Markov chain mixing since it eliminates, at least partially, high dependence of $\boldsymbol{\mu}$ and $\boldsymbol{z}$, usually observed in practical applications.

C.2.1 Full conditionals of the locations

Sampling $\boldsymbol{\mu}$ given $\boldsymbol{z}$ and $\boldsymbol{x}$ is straightforward. It is easily shown that

$$p(\boldsymbol{\mu}|\boldsymbol{z}, \boldsymbol{x}) \propto \prod_{j=1}^{K} \left\{ \left[\prod_{i \in I_j} dN(\boldsymbol{x}_i|\boldsymbol{\mu}_j, \boldsymbol{S})\right] dN(\boldsymbol{\mu}_j|\boldsymbol{\mu}_{i0}, \boldsymbol{V}_{j0}) \right\} \tag{C.5}$$

which has the kernel of a multivariate normal distribution with mean vector and covariance matrix given by

$$\boldsymbol{\mu}_{ji} = \boldsymbol{V}_{j1}\left(n_j\boldsymbol{S}^{-1}\tilde{\boldsymbol{x}}_j + \boldsymbol{V}_{j0}^{-1}\boldsymbol{\mu}_{j0}\right) \quad \text{and} \quad \boldsymbol{V}_{j1} = \left(n_j\boldsymbol{S}^{-1} + \boldsymbol{V}_{j0}^{-1}\right)^{-1}$$

respectively, for $n_j\tilde{\boldsymbol{x}}_j = \sum_{i\in I_j}\boldsymbol{x}_{(i)}$.

C.2.2 Full conditionals of the latent indicators

Sampling z_i given $\boldsymbol{z}_{(i)} = (z_1, \ldots, z_{i-1}, z_{i+1}, \ldots, z_n)$ and $\boldsymbol{x}$ is less clear. Initially,

$$\begin{aligned} Pr(z_i = j|\boldsymbol{z}_{(i)}, \boldsymbol{x}) &\propto Pr(z_i = j|\boldsymbol{z}_{(i)}, \boldsymbol{x}_{(i)})p(\boldsymbol{x}_i|z_i = j, \boldsymbol{z}_{(i)}, \boldsymbol{x}_{(i)}) \\ &\propto Pr(z_i = j)p(\boldsymbol{x}_i|z_i = j, \boldsymbol{z}_{(i)}, \boldsymbol{x}_{(i)}) \end{aligned}$$

where $\boldsymbol{x}_{(i)} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{i-1}, \boldsymbol{x}_{i+1}, \ldots, \boldsymbol{x}_n)$ and $Pr(z_i = j) = \pi_j$. Also,

$$\begin{aligned} p(\boldsymbol{x}_i|z_i = j, \boldsymbol{z}_{(i)}, \boldsymbol{x}_{(i)}) &= \int p(\boldsymbol{x}_i|z_i = j, \boldsymbol{z}_{(i)}, \boldsymbol{x}_{(i)}, \boldsymbol{\mu})p(\boldsymbol{\mu}|\boldsymbol{z}_{(i)}, \boldsymbol{x}_{(i)})d\boldsymbol{\mu} \\ &= \int dN(\boldsymbol{x}_i|\boldsymbol{\mu}_j, \boldsymbol{S})p(\boldsymbol{\mu}_j|\boldsymbol{z}_{(i)}, \boldsymbol{x}_{(i)})d\boldsymbol{\mu}_j \\ &\times \prod_{l\neq j}\int p(\boldsymbol{\mu}_l|\boldsymbol{z}_{(i)}, \boldsymbol{x}_{(i)})d\boldsymbol{\mu}_l \end{aligned}$$

where the last product of integrals is equal to one, following the conditional independence of $\boldsymbol{\mu}$'s given $\boldsymbol{x}$ and $\boldsymbol{z}$ shown above. It is easy to show that

$$p(\boldsymbol{\mu}_j|\boldsymbol{z}_{(i)}, \boldsymbol{x}_{(i)}) = dN(\boldsymbol{\mu}_j|\boldsymbol{\mu}_{j1,i}, \boldsymbol{V}_{j1,i})$$

with $\boldsymbol{V}_{j1,i} = \left(n_{j,i}\boldsymbol{S}^{-1} + \boldsymbol{V}_{j0}^{-1}\right)^{-1}$ and $\boldsymbol{\mu}_{j1,i} = \boldsymbol{V}_{j1,i}\left(n_{j,i}\boldsymbol{S}^{-1}\tilde{\boldsymbol{x}}_{j,i} + \boldsymbol{V}_{j0}^{-1}\boldsymbol{\mu}_{j0}\right)$. Analogously, $I_{j,i} = \{l : z_l = j, l = 1, \ldots, n \text{ and } l \neq i\}$, $n_{j,i} = card(I_{j,i})$ and $n_{j,i}\tilde{\boldsymbol{x}}_{j,i} = \sum_{l\in I_{j,i}}\boldsymbol{x}_l$. Therefore,

$$\begin{aligned} p(\boldsymbol{x}_i|z_i = j, \boldsymbol{z}_{(i)}, \boldsymbol{x}_{(i)}) &\propto \int dN(\boldsymbol{x}_i|\boldsymbol{\mu}_j, \boldsymbol{S})dN(\boldsymbol{\mu}_j|\boldsymbol{\mu}_{j1,i}, \boldsymbol{V}_{j1,i})d\boldsymbol{\mu}_j \\ &\propto dN(\boldsymbol{x}_i|\boldsymbol{\mu}_{j1,i}, \boldsymbol{V}_{j1,i} + \boldsymbol{S}) \end{aligned}$$

and

$$Pr(z_i = j|\boldsymbol{z}_{(i)}, \boldsymbol{x}) \propto \pi_j dN(\boldsymbol{x}_j|\boldsymbol{\mu}_{j1,i}, \boldsymbol{S} + \boldsymbol{V}_{i1,j}) \tag{C.6}$$

which can be sampled from straighforwardly.

Bibliography

Aguilar, O. and West, M. (2000) Bayesian dynamic factor models and variance matrix discounting for portfolio allocation. *Journal of Business and Economic Statistics (June 2000 Issue)*.

Akaike, H. (1987) Factor analysis and AIC. *Psychometrika*, **52,** 317–332.

Arminger, G. and Muthén, B.O. (1998) A Bayesian approach to nonlinear latent variable models using the Gibbs sampler and the Metropolis-Hastings algorithm. *Psychometrika*, **63,** 271–300.

Barbieri, M.M. and O'Hagan, A. (1996) A reversible jump MCMC sampler for Bayesian analysis of ARMA time series. Technical report. Dipartimento di Statistica,Probabilità e Statistiche Applicate, Università La Sapienza,Roma.

Bartholomew, D.J. (1981) Posterior analysis of the factor model. *The British Journal of Mathematical and Statistical Psychology*, **34,** 93–99.

Bartholomew, D.J. (1987) *Latent Variable Models and Factor Analysis*. London: Charles Griffin.

Bartholomew, D.J. (1995) Spearman and the origin and development of factor models. *British Journal of Mathematical and Statistical Psychology*, **48,** 211–220.

Berger, J.O. and Pericchi, L. R. (1996) The intrinsic Bayes factor for model selection and prediction. *Journal of the American Statistical Association*, **91,** 109–122.

Berzuini, C., Best, N.G., Gilks, W.R. and Larizza, C (1997) Dynamic conditional independence models and Markov chain Monte Carlo methods. *Journal of the American Statistical Association*, **92,** 1403–1412.

Besag, J. (1989) A candidate's formula: A curious result in Bayesian prediction. *Biometrika*, **76,** 183–183.

Box, G.E.P. and Tiao, G.C. (1973) *Bayesian Inference in Statistical Analysis*. Massachusetts: Addison-Wesley.

Bozdogan, H. and Ramirez, D.E. (1987) An expert model selection approach to determine the "best" pattern structure in factor analysis models. In *Multivariate Statistical Modelling and Data Analysis* (eds H. Bozdogan and A.K. Gupta).

Bozdogan, H. and Shigemasu, K. (1998) Bayesian factor analysis model and choosing the number of factors using a new informational complexity criterion. Technical Report. Department of Statistics, University of Tennessee.

Broemeling, L.D. (1985) *Bayesian Analysis of Linear Models.* New York: Marcel Dekker.

Carlin, B.P. and Chib, S. (1995) Bayesian model choice via Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society,* B, **57,** 473–484.

Carter, C. and Kohn, R. (1994) On Gibbs sampling for state space models. *Biometrika*, **81,** 541–553.

Celeux, G., Hurn, M. and Robert, C.P. (1999) Computational and inferential difficulties with mixture posterior distributions. Technical Report. Laboratorie de Statistique, CREST-ENSAE.

Chen, M.-H., Shao, Q.-M. and Ibrahim, J.G. (2000) *Monte Carlo methods in Bayesian computation.* New York: Springer-Verlag.

Chib, S. (1995) Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association*, **90,** 1313–1321.

Chib, S. and Greenberg, E. (1995) Understanding the Metropolis-Hastings algorithm. *The American Statistician*, **49,** 327–335.

Clyde, M. (1999) Bayesian model averaging and model search strategies (with discussion). In *Bayesian Statistics 6* (eds J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith), pp. 157–185. John Wiley.

Cowles, M.K. and Carlin, B.P. (1996) Markov chain Monte Carlo convergence diagnostics: A comparative review. *Journal of the American Statistical Association*, **91,** 883–904.

Dalal, S. and Hall, W. (1983) Approximating priors by mixtures of natural conjugate priors. *Journal of the Royal Statistical Society, Ser.* B, **45,** 278–286.

Davidian, M. and Gallant, A. R. (1993) The nonlinear mixed effects model with a smooth random effects density. *Biometrika*, **80,** 475–488.

Davidian, M. and Gallant, A.R. (1992) Smooth nonparametric maximum likelihood estimation for population pharmacokinetics, with application to Quinidine. *Journal of Pharmacokinetics and Biopharmaceutics*, **20,** 529–556.

Dawid, A.P. (1981) Some matrix-variate distribution theory: Notational considerations and a Bayesian application. *Biometrika*, **68,** 265–274.

Dellaportas, P., Forster, J.J. and Ntzoufras, I. (1998) On Bayesian model and variable selection using MCMC. Working Paper. Athens University of Economics and Business.

Dellaportas, P. and Smith, A.F.M. (1993) Bayesian inference for generalized linear and proportional hazards models via Gibbs sampling. *Applied Statistics*, **42,** 443–459.

Denison, D.G.T., Mallick, B.K. and Smith, A.F.M. (1997) Automatic Bayesian curve fitting. Working Paper. Imperial College.

Diaconis, P. and Ylvisaker, D. (1985) Quantifying prior opinion. In *Bayesian Statistics 2* (eds J.M. Bernardo, M.H. DeGroot, D.V. Lindley and A.F.M. Smith), pp. 163–175. John Wiley.

DiCiccio, T.J., Kass, R.A., Raftery, A.E. and Wasserman, L. (1997) Computing Bayes' factors by combining simulation and asymptotic approximations. *Journal of the American Statistical Association*, **92,** 903–915.

Diebold, F. X. and Nerlove, M. (1989) The dynamics of exchange rate volatility: A multivariate latent factor ARCH model. *Journal of Applied Econometrics*, **4,** 1–21.

Diebolt, J. and Robert, C.P. (1994) Estimation of finite mixture distributions through Bayesian sampling. *Journal of the Royal Statistical Society, Ser.* B, **56,** 163–175.

Doucet, A. (1998) On sequential simulation-based for Bayesian filtering. Technical Report CUED/F-INFENG/TR 310. Department of Engineering, Cambridge University.

Drèze, J. H. and Richard, J. F. (1983) Bayesian analysis of simultaneous equation systems. In *Handbook of Econometrics* (eds Z. Griliches and M. Intriligator). Amsterdan: North-Holland.

Engle, R. (1987) Multivarite ARCH with factor structures - cointegration in variance. Technical Report. University of California at San Diego.

Fraley, C. and Raftery, A.E. (1998) How many clusters? Which clustering method? Answers via model-based cluster analysis. Technical Report 329. Department of Statistics, University of Washington.

Fraley, C. and Raftery, A.E. (1999) MCLUST: software for model-based cluster and discriminant analysis. Technical Report 342. Department of Statistics, University of Washington.

Frühwirth-Schnatter, S. (1994) Data augmentation and dynamic linear models. *Journal of Time Series Analysis*, **15,** 183–202.

Gamerman, D. (1997) *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Chapman and Hall.

Gelfand, A.E. and Dey, D. (1994) Bayesian model choice: Asymptotics and exact calculations. *Journal of the Royal Statistical Society, Ser.* B, **56,** 501–514.

Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B. (1995) *Bayesian Data Analysis*. Chapman and Hall.

Geweke, J. (1989) Bayesian inference in econometric models using Monte Carlo integration. *Econometrica*, **57,** 1317–1339.

Geweke, J.F. and Singleton, K.J. (1980) Interpreting the likelihood ratio statistic in factor models when sample size is small. *Journal of the American Statistical Association*, **75,** 133–137.

Geweke, J.F. and Zhou, G. (1996) Measuring the pricing error of the arbitrage pricing theory. *The Review of Financial Studies*, **9,** 557–587.

Giakoumatos, S.G., Dellaportas, P. and D.D., Politis (1999) Bayesian analysis of some multivariate time-varying volatility models. Technical Report. Department of Statistics, Athens University of Economics and Business, Greece.

Gilks, W. R., Richardson, S. and Spiegelhalter, D.J. (1996) *Markov Chain Monte Carlo in Practice*. Chapman and Hall.

Godsill, S.J. (1998) On the relationship between MCMC model uncertainty methods. Technical Report CUED/F-INFENG/TR.305. Cambridge University Engineering Department.

Godsill, S.J., Doucet, A. and West, M. (2000) Methodology for Monte Carlo smoothing with application to time-varying autoregressions. In *In Proceedings International Symposium on Frontiers of Time Series Modelling.* Institute of Statistical Mathematics, Tokyo.

Gordon, N.J., Salmond, D.J. and Smith, A.F.M. (1995) Sequential Monte Carlo methods for dynamic systems. *IEE Proceedings-F*, **140,** 107–113.

Green, P. (1995) Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, **82,** 711–732.

Harvey, A.C. (1989) *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press.

Harvey, A.C., Ruiz, E. and Shephard, N. (1994) Multivariate stochastic variance models. *Review of Economic Studies*, **61,** 247–264.

Hastings, W.K. (1970) Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, **57,** 97–109.

Hoeting, J.A., Madigan, D., Raftery, A.E. and Volinsky, C.T. (1999) Bayesian model averaging: a tutorial. Technical Report # 9814. Department of Statistics, Colorado State University.

Huerta, G. and West, M. (1999) Priors and component structures in autoregressive time series models. *Journal of the Royal Statistical Society-Series* B, **61,** 881–899.

Ihara, M. and Kano, Y. (1995) Identifiability of full, marginal, and conditional factor analysis model. *Statistics and Probability Letters*, **23,** 343–350.

Insua, D.R. and Müller, P. (1998) Feedforward neural networks for nonpametric regression. Discussion Paper 98–02. Institute of Statistics and Decision Sciences, Duke University.

Jacquier, E., Polson, N.G. and Rossi, P. (1995) Models and priors for multivariate stochastic volatility. Discussion Paper. Graduate School of Business, University of Chicago.

Jefferys, W.H. and Berger, J.O. (1992) Ockham's razor and Bayesian analysis. *American Scientist*, **80,** 64–72.

Johnson, N. L. and Kotz, S. (1972) *Distributions in Statistics: Continuous Multivariate Distributions*. New York: John Wiley.

Joreskog, K.G. (1967) Some contributions to maximum likelihood factor analysis. *Psychometrika*, **32,** 443–482.

Kass, R.A. and Raftery, A.E. (1995) Bayes' factor. *Journal of the American Statistical Association*, **90,** 773–795.

Kim, S.N., Shephard, N. and Chib, S. (1998) Stochastic volatility: likelihood inference and comparison with ARCH models. *Review of Economic Studies*, **65,** 361–393.

Kitagawa, G. (1998a) Monte Carlo filter and smoother for non-Gaussian non-linear state space models. *Journal of Computational and Graphical Statistics*, **5,** 1–25.

Kitagawa, G. (1998b) Self-organising state space model. *Journal of the American Statistical Association*, **93,** 1203–1215.

Kleinman, K.P. and Ibrahim, J.G. (1998a) A semi-parametric Bayesian approach to generalized linear mixed models. *Statistics in Medicine*, **17,** 2579–2596.

Kleinman, K.P. and Ibrahim, J.G. (1998b) A semi-parametric Bayesian approach to the random effects model. *Biometrics*, **54,** 921–938.

Lawley, D.N. and Maxwell, A.E. (1963) *Factor Analysis as a Statistical Method.* London: Butterworth.

Lee, S.Y. (1981) A Bayesian approach to confirmatory factor analysis. *Psychometrika*, **46,** 153–160.

Lewis, S.M. and Raftery, A.E. (1997) Estimating Bayes' factors via posterior simulation with the Laplace-Metropolis estimator. *Journal of the American Statistical Association*, **92,** 648–655.

Lichtman, S.M., Ratain, M.J., Van Echo, D.A., Rosner, G., Egorin, M.J., Budman, D.R., Vogelzang, N.J., Norton, L. and Schilsky, R.L. (1993) Phase I trial of granulocyte-macrophage colony-stimulating factor plus high-dose cyclophosphamide given every 2 weeks: a Cancer and Leukemia Group B study. *Journal of the National Cancer Institute*, **85,** 1319–1325.

Liu, J. and West, M. (2000) Combined parameter and state estimation in simulation-based filtering. In *Sequential Monte Carlo Methods in Practice* (eds A. Doucet, N. de Freitas and N.J. Gordon). Springer-Verlag: New York.

Liu, J.S. and Chen, R. (1995) Sequential Monte Carlo methods for dynamic systems. *Journal of the American Statistical Association*, **90,** 567–576.

Magder, L.S. and Zeger, S.L. (1996) A smooth nonparametric estimate of a mixing distribution using mixtures of Gaussians. *Journal of the American Statistical Association*, **91,** 1141–1151.

Mallet, A. (1986) A maximum likelihood estimation method for random coefficient regression models. *Biometrika*, **73,** 645–656.

Mallet, A., Mentré, F., Steimer, J.-L. and Lokiec, F. (1988) Handling covariates in population pharmacokinetics with an application to gentamicin. *Biomedical measurement informatics and control*, **2,** 138–146.

Markowitz, H.M. (1959) *Portfolio Selection: Efficient Diversification of Investiments.* New York: John Wiley and Sons.

Martin, J.L. and McDonald, R.P. (1981) Bayesian estimation in unrestricted factor analysis: A treatment for Heywood cases. *Psychometrika*, **40,** 505–517.

Meng, X.L. and Wong, W.H. (1996) Simulating ratios of normalizing constants via a simple identity: a theoretical exploration. *Statistica Sinica*, **6,** 831–860.

Mengersen, K. and Robert, C.P. (1995) Testing for mixtures: a Bayesian entropic approach. In *Bayesian Statistics 5* (eds J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith). John Wiley.

Mengersen, K.L., Robert, C.P. and Guihenneuc-Jouyaux, Ch. (1999) MCMC convergence diagnostics: A reviewww. In *Bayesian Statistics 6* (eds J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith). John Wiley.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N. and Teller, A.H. (1995) Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, **21,** 1087–1092.

Migon, H.S. and Gamerman, D. (1999) *Statistical Inference: An Integrated Approach.* Arnold.

Müller, P., Quintana, F. and Rosner, G. (1999) Hierarchical meta-analysis over related non-parametric Bayesian. Discussion Paper # 99-22. Institute of Statistics and Decision Sciences, Duke University.

Müller, P. and Rosner, G. (1997) A semiparametric Bayesian population model with hierarchical mixture priors. *Journal of the American Statistical Association*, **92,** 1279–1292.

Müller, P. and Rosner, G. (1998) Semiparametric PK/PD models. In *Practical Nonparametric and Semiparametric Bayesian Statistics* (eds D. Dey, P. Müller and D. Sinha). Springer-Verlag.

Neal, R.M. (1998) Markov chain sampling methods for Dirichlet process mixture models. Technical Report 9815. Department of Statistics and Department of Computer Science, University of Toronto.

Newton, M.A. and Raftery, A.E. (1994) Approximate Bayesian inference with the weighted likelihood bootstrap. *Journal of the Royal Statistical Society, Ser.* B, **56,** 3–48.

O'Hagan, A. (1995) Fractional Bayes factor for model comparison. *Journal of the Royal Statistical Society, Ser.* B, **57,** 99–138.

Pérez, J. M. (1998) Development of expected posterior prior distributions for model comparison. Ph.D. thesis. Purdue University.

Pérez, J. M. and Berger, J. M. (2000) Expected posterior prior distributions for model selection. Discussion Paper 00-08. Institute of Statistics and Decision Sciences, Duke University.

Peskun, P.H. (1973) Optimum Monte-Carlo sampling using Markov chains. *Biometrika*, **60,** 607–612.

Pitt, M.K. and Shephard, N. (1999a) Filtering via simulation: auxiliary particle filters. *Journal of the American Statistical Association*, **94,** 590–599.

Pitt, M.K. and Shephard, N. (1999b) Time varying covariances: A factor stochastic volatility approach (with discussion). In *Bayesian statistics 6* (eds J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith), pp. 547–570. London: Oxford University Press.

Polasek, W. (1997) Factor analysis and outliers: A Bayesian approach. Discussion Paper, University of Basel.

Polson, N.G. and Tew, B.V. (1997) Bayesian portfolio selection: An empirical analysis of the SP500 index 1970-1996. Unpublished Working Paper. University of Chicago.

Press, S.J. (1982) *Applied Multivariate Analysis: Using Bayesian and Frequentist Methods of Inference (2nd edition).* New York: Krieger.

Press, S.J. and Shigemasu, K. (1989) Bayesian inference in factor analysis. In *ASA Proceedings of Social Statistics Section*, pp. 292–294.

Press, S.J. and Shigemasu, K. (1994) Posterior distribution for the number of factors. In *ASA Proceedings of Bayesian Statistical Science*, pp. 75–77.

Raftery, A.E. (1996) Hypothesis testing and model selection. In *Markov Chain Monte Carlo in Practice* (eds W.R. Gilks, S. Richardson and D.J. Spiegelhalter). Chapman and Hall.

Raftery, A.E., Madigan, D. and Hoeting, J.A. (1997) Bayesian model averaging for linear regression models. *Journal of the American Statistical Association*, **92,** 179–191.

Richardson, S. and Green, P. (1997) Reversible jump Markov chain Monte Carlo computation and Bayesian model determination (with discussion). *Journal of the Royal Statistical Society - Series* B, **59,** 731–758.

Roeder, K. and Wasserman, L. (1997) Practical Bayesian density estimation using mixtures of normals. *Journal of the American Statistical Association*, **92,** 894–902.

Rosner, G. and Müller, P. (1997) Bayesian population pharmacokinetics and pharmacodynamic analyses using mixture models. *Journal of Pharmacokinetics and Biopharmaceutics*, **25,** 209–233.

Schwarz, G. (1978) Estimating the dimension of a model. *The Annals of Statistics*, **6,** 461–464.

Seber, G.A.F. (1984) *Multivariate Observations*. New York: Wiley.

Sentana, E. (1998) The relation between conditionally heteroskedastic factor models and factor ARCH models. *Econometrics Journal*, **1,** 1–9.

Shephard, N. (1996) Statistical aspects of ARCH and stochastic volatility. In *Time Series Models in Econometrics, Finance and Other Fields* (eds D.R. Cox, D.V. Hinkley and O.E. Barndorff-Nielsen). London: Chapman and Hall.

Spencer, C. M. and Goa, K. L. (1995) Amifostine. a review of its pharmacodynamic and pharmacokinetic properties, and therapeutic potential as a radioprotector and cytotoxic chemoprotector. *Drugs*, **91,** 1001–1031.

Spiegelhalter, D. J. and Smith, A. F. M. (1982) Bayes factor for linear and log-linear

models with vague prior information. *Journal of the Royal Statistical Society, Ser.* B, **44,** 377–387.

Tierney, L. and Kadane, J.B. (1986) Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Society*, **81,** 82–86.

Troughton, P.T. and Godsill, S.J. (1997) Reversible jump sampler for autoregressive time series, employing full conditionals to achieve efficient model space moves. Technical Report CUED/F-INFENG/TR.304. Cambridge University Engineering Department.

Wakefield, J.C. (1996) The Bayesian analysis of population pharmacokinetic models. *Journal of the American Statistical Association*, **91,** 62–75.

Wakefield, J.C., Aarons, L. and Racine-Poon, A. (1998) The Bayesian approach to population pharmacokinetic/pharmacodynamic modelling. In *Case Studies in Bayesian Statistics* (eds B.P. Carlin, A.L. Carriquiry, C. Gatsonis, A. Gelman, R.E. Kass, I. Verdinelli and M. West). Springer-Verlag.

Wakefield, J.C., Smith, A.F.M., Racine-Poon, A. and Gelfand, A. E. (1994) Bayesian analysis of linear and non-linear population models by using the Gibbs sampler. *Applied Statistics*, **43,** 201–221.

Walker, S. and Wakefield, J.C. (1994) Population models with a nonparameric random coefficient distribution. Technical Report. Imperial College London.

Wasserman, L. (1997) Bayesian model selection and model averaging. Technical Report 666. Department of Statistics, Carnegie Mellon University.

West, M. (1993a) Approximating posterior distributions by mixtures. *Journal of the Royal Statistical Society, Ser.* B, **55,** 409–422.

West, M. (1993b) Mixture models, Monte Carlo, Bayesian updating and dynamic models. In *Computing Science and Statistics: Proceedings of the 24th Symposium of the Interface* (ed. J.H. Newton), pp. 325–333. Interface Foundation of North America, Fairfax Station, Virginia.

West, M. and Harrison, P.J. (1997) *Bayesian Forecasting and Dynamic Model.* New York: Springer-Verlag.

Zeger, S. L. and Karim, M. R. (1991) Generalized linear models with random effects: A Gibbs sampling approach. *Journal of the American Statistical Association*, **86,** 79–86.

Zellner, A. (1971) *An Introduction to Bayesian Inference in Econometrics.* New York: Wiley.

Biography

I was born in Aquidauana, Brazil in August 8, 1968. I obtained B.S. and M.S. degrees in Statistics from the Federal University of Rio de Janeiro (UFRJ) in September 1991 and November 1994 respectively. I arrived at Duke University in the Fall of 1996 and became a Ph.D candidate in the Spring of 1998. I have held an assistant professor position at the Department of Statistical Methods, Federal University of Rio de Janeiro, Brazil, since May, 1996. Prior to that, from April 1992 to April 1996, I held a lecturer position at the Fluminense Federal University, also in Brazil. From March 1991 to July 1996 I worked as a research assistant with the Macroeconometric group at the Brazilian Research Institute of Applied Economy (IPEA). In the past ten years, I have attended more than twenty conferences, nationally (such as the Brazilian Symposia of Probability and Statistics, Brazilian Schools of Time Series and Econometrics, Brazilian Meetings of Econometrics), and internationally (such as the VI Valencia Meeting on Bayesian Statistics, Case Studies in Bayesian Statistics and the European Conferences on Highly Structured Stochastic Systems), most of which with contributed posters and contributed/invited talks. I have co-authored the following scientific articles:

1. Lopes, H.F. and Müller, P. and Rosner, G. (2000) Meta-analysis for longitudinal data models using multivariate mixture priors. ISDS Discussion Paper 00-13.

2. Lopes, H.F. and West, M. (2000) Model uncertainty in factor analysis. ISDS Discussion Paper 98-38.

3. Lopes, H.F., Moreira, A.R.B., and Schmidt, A.M. (1999) Hyperparameter estimation in forecasting models *Computational Statistics and Data Analysis*, 29, pp. 387-410.

4. Huerta, G. and Lopes, H.F. (1999) Bayesian Forecasting and inference in latent structure for the Brazilian GDP and industrial production index. ISDS Discussion Paper 99-08.

5. Moreira, A.R.B., Fiorencio, A. and Lopes, H.F. (1997) A Multivariate model to forecast GNP, inflation and liquidity. (In Portuguese). *The Brazilian Review of Econometrics*, 17, pp. 67-111.

6. Lopes, H. F., and Ehlers, R. S.(1997). Bayesian forecasting the levels of vector autoregressions for log-transformed time series. Technical Report. Department of Statistical Methods, Federal University of Rio de Janeiro, Brazil.

7. Moreira, A.R.B., Fiorencio, A. and Lopes, H.F. (1996) Identification of the common trends of GNP, inflation and liquidity. (In Portuguese). *Perspectiva da Economia Brasileira*, IPEA.

8. Lopes, H.F., and Migon, H.S. (1996). Impulse response functions in generalized Bayesian vector autoregressive models. Technical Report. Department of Statistical Methods, Federal University of Rio de Janeiro, Brazil.

9. Lima, E.C.R., Lopes, H.F., Moreira, A.R.B., and Pereira, P.L.V. (1995) Stochastic trends of the Brazilian GNP. (In Portuguese) *Pesquisa e Planejamento Econômico*, 25, pp. 249-278.

10. Issler, J.V., Moreira, A.R.B, and Lopes, H.F. (1994). Common cycles in structural identification of multivariate systems. *Annals of the XVI Brazilian Meeting of Econometrics.*

11. Migon, H.S., Lima, E.C.R. and Lopes, H.F. (1993). Dynamic effects of aggregate demand and supply disturbances: the Brazilian case. (In Portuguese). *Revista Brasileira de Economia*, 47, 3, 1993.