LATENT STRUCTURE IN BAYESIAN MULTIVARIATE TIME SERIES MODELS

by

Omar Aguilar Chávez

Institute of Statistics and Decision Sciences
Duke University

Date: _____

Approved:

_____
Dr. Mike West, Supervisor

_____
Dr. James Berger

_____
Dr. Peter Mueller

_____
Dr. Gabriel Katul

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

1998

ABSTRACT

(Statistics)

LATENT STRUCTURE IN BAYESIAN MULTIVARIATE TIME SERIES MODELS

by

Omar Aguilar Chávez

Department of Institute of Statistics and Decision Sciences
Duke University

Date: _____

Approved:

_____
Dr. Mike West, Supervisor

_____
Dr. James Berger

_____
Dr. Peter Mueller

_____
Dr. Gabriel Katul

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

1998

Abstract

This dissertation introduces new classes of models and approaches to multivariate time series analysis and forecasting, with a focus on various problems in which time series structure is driven by underlying latent processes of key interest. The identification of latent structure and common features in multiple time series is first studied using wavelet based methods and Bayesian time series decompositions of certain classes of dynamic linear models. The results are applied to turbulence and geochemical time series data, the latter involving development of new time series models for latent time-varying autoregressions with heavy-tailed components for quite radically ill-behaved series. Natural extensions and generalizations of these models lead to novel developments of two key model classes, dynamic factor models for multivariate financial time series with stochastic volatility components, and multivariate dynamic generalized linear models for non-Gaussian longitudinal time series. These two model classes are related through common statistical structure, and the dissertation discusses issues of Bayesian model specification, model fitting and computation for posterior and predictive analysis that are common to the two model classes. Two motivating applications are discussed, one in each of the two model classes. The first concerns short term forecasting and dynamic portfolio allocation, illustrated in a study of the dynamic factor structure of daily spot exchange rates for a selection of international currencies. The second application involves analyses of time series of collections of many related binomial outcomes and arises in a project in health care quality monitoring with the Veterans Affairs (VA) hospital system.

Acknowledgements

I would like to express my appreciation and gratitude to my advisor Prof. Mike West for his encouragement, advice and support during my four years at ISDS. His valuable suggestions, constant input of creative ideas and enthusiastic guidance provided me with the great privilege of learning new attitudes towards research and life.

I greatly appreciate the useful discussions and contributions of Peter Mueller, Jim Berger, Brani Vidakovic and Giovanni Parmigiani. I also want to acknowledge the faculty, staff and fellow graduate students at ISDS for providing a stimulating and wonderful working environment. I am also grateful to the providers of the various data sets used in this thesis and especially to Gabriel Katul that kindly provided the turbulence data analyzed in Chapter 4.

I would especially like to dedicate this work to my parents, brother and sisters for their support and understanding that I have always received and to Viridiana for her love, constant support and dedication that are sources of inspiration and strength for being a better person everyday.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

A *latent* variable is simply a variable that cannot be measured directly. The origins of latent variable modeling date from the start of the twentieth century, specifically from the study of human abilities conducted by Spearman. The early studies were almost exclusively focused on the well known "factor analysis models" that included just a single factor. Further impetus came in the 1930s with the work of Thurstone and colleagues in multiple-factor models and simplest-structure solutions. Traditionally, one of the goals of the study of latent structure has been on the actual latent processes for example in measuring variables such as social class, personality or intelligence in the social and behavioral sciences. In order to obtain information on such variables, researchers are forced to consider other variables, which can be measured and which are related to the latent quantities of interest, but which may contain additional noise or error. However, in many other applications the concentration is more on the identification and interpretation of underlying processes that are driving the observations of certain types of phenomena. In particular, time series research focussed on latent structure analysis has been of great interest in isolating and interpreting the possible many components of sets of observed time series. The work presented here is mainly focussed on theoretical and computational developments in

Bayesian time series modeling for the identification of latent processes in line with existing theory of decomposition of a variety of Bayesian dynamic linear models. It is important to note though, that the practical utility of these models is to obtain reasonable interpretations of the inferred latent processes. In other words, there is no guarantee that the identified latent processes represent "real" phenomena and in some cases the results can only be seen as a convenient parsimonious description of complex structures.

The thesis starts with developments and discussion of non-parametric decompositions of time series using wavelet techniques. As a matter of fact, wavelets have been used widely for data de-noising and signal processing and hence the relevance of considering them as an important tool for identification of latent processes. In connection with this, Bayesian time series decompositions are developed in chapter 3 for certain classes of dynamic linear models. Specifically, decompositions of autoregressive models are illustrated in chapter 4 in studies of turbulence data and the results are compared with wavelet based techniques. The methodological developments in Bayesian time series of chapter 3 are further extended in chapter 5 to include latent time-varying autoregressions with heavy-tailed components that are used to model quite radically ill-behaved series. A posterior sampling algorithm via Markov Chain Monte Carlo (MCMC) is outlined in this chapter for these kind of general models and the results are presented in the analysis of eight chemical species obtained from the Greenland Ice Sheet covering the period $41-0$ Kyr BP. As a result of the analyses, multivariate models are explored and discussed as generalizations of the univariate models by including correlation structure between the multiple series in a factor model framework.

A second part of the thesis begins in chapter 6 with a full Bayesian approach for the basic factor model for multiple time series, considering identification issues and

describing the corresponding sampling algorithm for model fitting. In this context, novel dynamic factor models for multivariate financial time series and the incorporation of stochastic volatility components for the latent factor processes are proposed in chapter 7. These new models are of relevance in modeling multivariate processes with time-variations in the factor model parameters and represent direct generalizations of univariate stochastic volatility models. The methodology is applied in studies of dynamic factor structure of daily spot exchange rates for a selection of international currencies in chapter 8. In addition, model performance is compared with the much simpler method of dynamic variance discounting through different dynamic asset allocation strategies. Future potential developments and model extensions are also discussed in this chapter.

A third part of the thesis involves the development of new models for non-normal time series in a framework extending that of the dynamic generalized linear model. In part this was motivated by a collaborative project with the Veterans Affairs (VA) Management Science Group. The time series components of this research involves novel models and MCMC methods for collections of many time series of counts in a time-varying, longitudinal framework. Hierarchical non-Gaussian models for univariate and multivariate time series are developed chronologically in chapters 9 and 10. The goals are to monitor and evaluate patterns of change over time, and cross-sectional dependencies, in series of annual measures of health care quality in the VA hospital system. The framework combines cross-sectional, hierarchical models of the population of hospitals with time series structure to allow and measure time-variations in the associated hierarchical model parameters. Model assessment and residual analysis are discussed together with MCMC algorithms to fit these models. These models have real value and potential in various areas of application in business and socio-economics.

Chapter 2

Wavelet Decompositions

Wavelets have shown great potential and applicability in many fields in recent years, especially in signal processing, data compression and de-noising. The purpose of this chapter is to review some of the basic properties of wavelet methods and their interaction with statistical modeling approaches to find latent structure in time series data.

2.1 Introduction to Wavelets

Wavelets are fundamental building block functions used as basis in representing other functions, analogous to the trigonometric sine and cosine functions for the ordinary Fourier transformation. However, Fourier basis functions are localized in frequency but not in time, which means that small frequency changes in the Fourier transform will produce changes everywhere in the time domain. In contrast, wavelets are localized in both frequency and time, which is a key advantage over Fourier methods in dealing with a variety of functions, including messy signals with jumps and non-smooth features. In fact, wavelet approximations can compact the energy[1] of a

[1]Energy is an engineering term for the norm of $\mathcal{L}_2$ functions, $\|f\|^2 =< f, f >= \int f^2$. By definition a function f is in $\mathcal{L}_2(S)$ if $\int_S f^2$ is finite, in other words if f is square integrable over S.

signal into relatively small number of wavelet functions. Consequently, wavelets can represent many classes of functions in a compact and "faster[2]" way, needing fewer wavelet basis than sine-cosine basis functions for the approximations. In addition, wavelets can separate the signal into multiresolution components where the fine and coarse resolution components capture, respectively, the fine and coarse scale features in the signal. These properties make wavelet representations excellent tools for data compression and signal processing among other applications in a wide variety of scientific fields. In recent years, statistical wavelet modeling has become very popular in theoretical and applied areas, specially in non-parametric modeling, regression and density estimation studies.

There are several mathematical and not-too mathematical papers developing and explaining theory on wavelets. Key references are Daubechies (1988), Daubechies (1992), Strang (1989), Chui (1992), Walter (1994) and Vidakovic (1998b) among others. The fundamental theory on wavelets is connected with continuous decompositions of functions in $\mathcal{L}_2$. The definition of wavelets is traditionally done in terms of a family of functions with a "mother" and "father". The mother wavelets ψ, are good at representing the detail and high-frequency parts of a signal satisfying that $\int \psi(x)dx = 0$. The father wavelets ϕ on the other hand, are good at representing the smooth and low-frequency parts of a signal with $\int \phi(x)dx = 1$. A wavelet family of functions is then defined by translating and scaling the mother and the father $\psi_{a,b}(x)$ and $\phi_{a,b}(x)$. There are specific choices of a and b that can be used to find minimal orthogonal basis for functions in $\mathcal{L}_2$. In particular, a traditional choice is to set $a = 2^j$ and $b = k2^j$ for $j, k \in \mathcal{Z}$, defining the orthogonal wavelet basis,

$$\{\phi_{j,k}(x) = 2^{-j/2}\phi(2^{-j}x - k), j, k \in \mathcal{Z}\}, \tag{2.1}$$

$$\{\psi_{j,k}(x) = 2^{-j/2}\psi(2^{-j}x - k), j, k \in \mathcal{Z}\}, \tag{2.2}$$

[2]The computational complexity of the Fast Fourier Transformation is $O(n\log_2(n))$, whereas for the Fast Wavelet Transformation it goes down to $O(n)$.

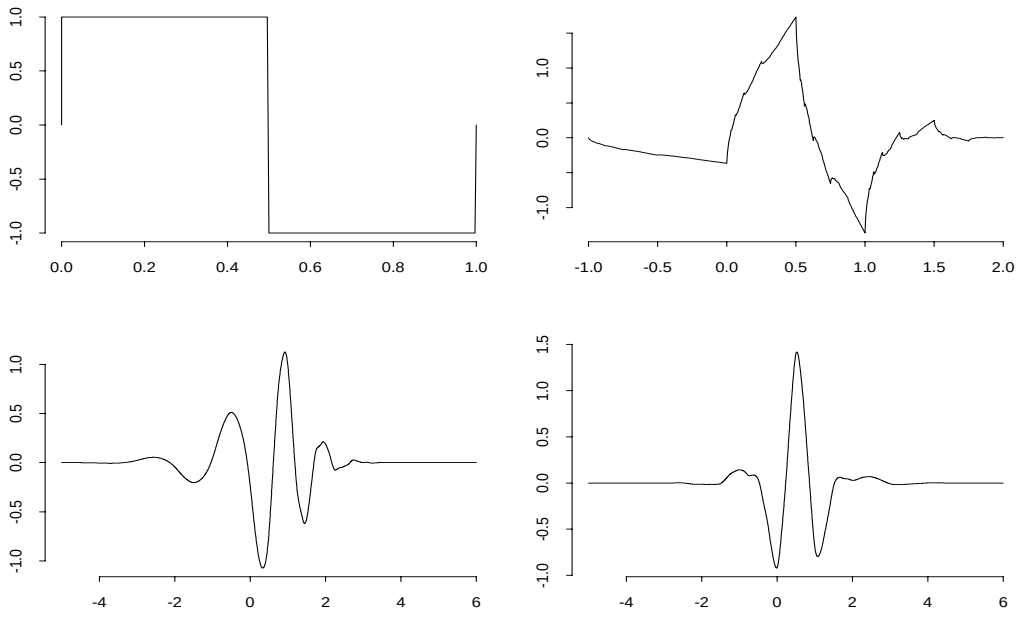

Figure 2.1: Four orthogonal mother wavelets $\psi_{0,0}$. Top: "Haar" (left) and "Daub4" (right). Bottom: "Daub12" (left) and "Symmlets12" (right).

where j is the *level* index associated with scale 2^j and k is the shift index associated with translation $k2^j$. There are many different wavelet families with different properties and features, being the most commonly used Haar's, Daubechies', Meyer's among others.

Figure 2.1 shows four different orthogonal mother wavelets $\psi_{0,0}$, that vary in width and smoothness. The top left frame displays the well-known "Haar" wavelet named after the mathematician Haar. This wavelet is not-continuous, orthogonal, symmetric and has compact support, meaning that it takes the value of zero outside a finite interval. The "Daub4" wavelet, plotted on the top right frame, was discovered by Ingrid Daubechies and it was the first type of continuous wavelet with compact support. The bottom left frame plots the "Daub12" mother wavelet, which is from the same family of the "Daub4" wavelet, but is relatively wider and smoother. The bottom right displays the "Symmlets12" wavelet constructed by Daubechies to be

as symmetrical as possible unlike the "Daub" wavelets. Traditionally, the number on the names indicates the width and smoothness of the wavelet, such that small numbers are narrower, less smooth and more localized as for example the "Daub4".

In any case, any element in $\mathcal{L}_2$ may be represented as a linear combination of the basis functions $\phi_{j,k}(x)$ and $\psi_{j,k}(x)$. Explicitly, the orthogonal wavelets series approximation to a continuous time signal $f(x)$ is given by:

$$f(x) \approx \sum_k s_{J,k}\phi_{J,k}(x) + \sum_k d_{J,k}\psi_{J,k}(x) + \sum_k d_{J-1,k}\psi_{J-1,k}(x) + \ldots + \sum_k d_{1,k}\psi_{1,k}(x) \quad (2.3)$$

where J is the number of levels and k ranges from 1 to the number of components at each level. The coefficients in the vector $\mathbf{d}' = (d_{1,k}, \ldots, d_{J,k}, s_{J,k})$ are called *wavelet coefficients* where the elements $\mathbf{s}'_J = \{s_{J,k}\}$, represent the underlying low frequency features of the signal at the coarse scale 2^J and are traditionally called "smooth" coefficients. The elements $\mathbf{d}_j = \{d_{j,k}\}$ or "detail" coefficients on the other hand represent fine scale deviations from the smooth behavior at scales 2^j for $j = 1, \ldots, J$. The number of coefficients at a certain scale is related to the width of the wavelet function. Typically, the wavelet coefficients at coarse scales, close to $j = J$, are larger in absolute value than the wavelet coefficients at fine scales, close to $j = 1$. This is a nice and convenient property for wavelet *Shrinkage* methods, as explained below.

In many applications of wavelets in statistical modeling, the observed signal is often a collection of discrete points or a function generated by a data set. In such cases, the theory should be reformulated into a single continuous to discrete conversion procedure bearing in mind that the discrete data is a sample from the continuous time signal with a certain sampling interval. The reformulated analysis for discrete signals is focused on discrete wavelet transformations that lead to wavelets approximations as (2.3) with resolution equal to the sampling interval.

2.2 Discrete Wavelet Transformation (DWT)

The DWT calculates the coefficients of the wavelet series approximation for a discrete signal $y_1, y_2, \ldots, y_n$, where n is a power of 2, say $n = 2^J$ for simplicity. The DWT maps the vector $\mathbf{y} = (y_1, y_2, \ldots, y_n)'$ to the set of n wavelet coefficients $\mathbf{d} = (\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_J, \mathbf{s}_J)'$. There are $n/2^1$ coefficients $d_{1,k}$ at the finest scale 2^1, $n/2^j$ coefficients $d_{j,k}$ at scales 2^j for $j = 1, 2, \ldots, J$ and $n/2^J$ smooth coefficients $s_{J,k}$ at coarsest scale 2^J. The mapping is performed by linearly transforming the signal by the orthogonal $n \times n$ wavelet matrix $\mathbf{W}$, namely

$$\mathbf{d} = \mathbf{W}\mathbf{y}.$$

In practice, the DWT is applied without computing the matrix multiplications explicitly, instead a fast "pyramid" algorithm is used which is, as discussed above, faster than the fast Fourier transform, (Mallat, 1989). This filtering procedure is based on an increasing sequence of closed and nested subspaces which approximate $\mathcal{L}^2$. The procedure is directly connected to the "quadrature mirror filters" in signal processing language. This algorithm involves a low-pass filter g, and high-pass filter h which are determined by the wavelet basis and are mutually related through $g_i = (-1)^i h_{1-i}$. The filters are applied to any sequence $\{a_n\}$ through the operators G and H with $k-$th elements given by,

$$(H\mathbf{a})_i = \sum_k h_{i-2k} a_i \quad \text{and} \tag{2.4}$$

$$(G\mathbf{a})_i = \sum_k g_{i-2k} a_i. \tag{2.5}$$

The algorithm starts by applying the filters to the data vector $\mathbf{y}$ and obtain the sub-vector of wavelet coefficients $\mathbf{d}_1 = G\mathbf{y}$ together with the corresponding smooth coefficients $\mathbf{s}_1 = H\mathbf{y}$ at level $j = 1$ or scale 2^1; see Figure 2.2. The procedure

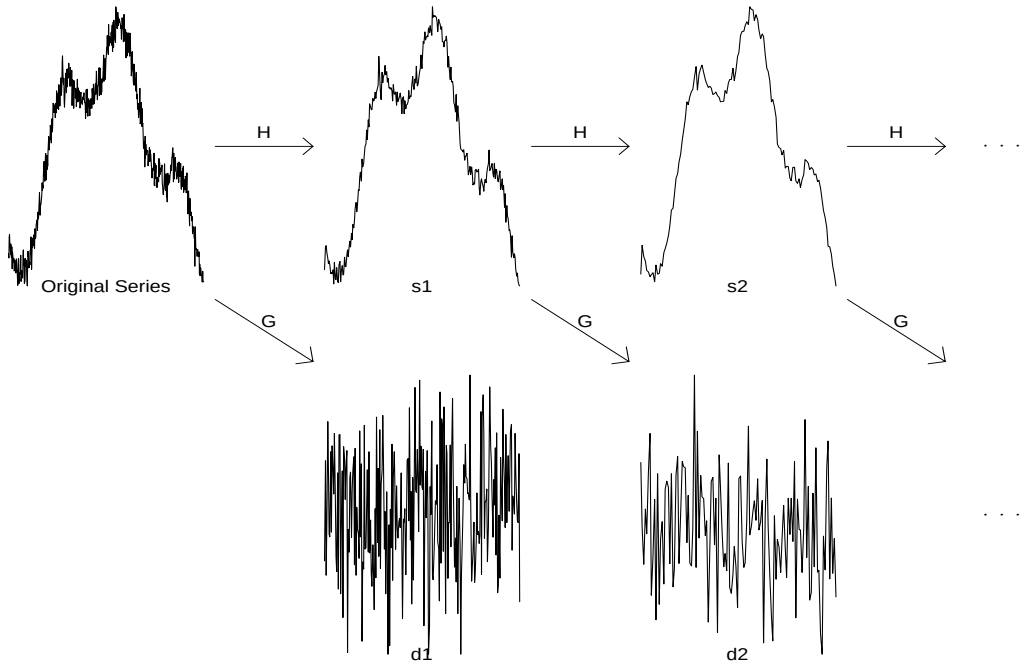

Figure 2.2: The "Pyramid" algorithm.

continues by applying the operators again to $\mathbf{s}_1$ to obtain $\mathbf{d}_2 = G\mathbf{s}_1 = GH\mathbf{y}$ and $s_2 = H^2\mathbf{y}$. Likewise, coefficients $\mathbf{d}_j$ and $\mathbf{s}_j$ are then computed by repeating this *decimation* procedure until level J, noting that $\mathbf{d}_J = GH^{J-1}$ and $\mathbf{s}_J = H^J\mathbf{y}$ contain only one coefficient. The complete vector of wavelet coefficients $\mathbf{d}$ can be rewritten in terms of filters as

$$\mathbf{d} = (G\mathbf{y}, GH\mathbf{y}, GH^2\mathbf{y}, \ldots, GH^{J-1}, H^J\mathbf{y})'. \tag{2.6}$$

In general, as stated before, the level j in the wavelet decomposition of $\mathbf{y}$ contains 2^j elements with filter representations,

$$\mathbf{d}_j = GH^{j-1}.$$

The interpretation of the coefficients at the first pass is that the signal $\mathbf{y}$ is represented with a coarser approximation $\mathbf{s}_1$ and $\mathbf{d}_1$ is the "detail" lost by approximating $\mathbf{y}$ by the linear combination in $\mathbf{s}_1 = H\mathbf{y}$ as displayed in the pyramid algorithm on Figure

2.2. The same interpretation applies for the next level where $\mathbf{d}_2$ is the detail lost by approximating $\mathbf{s}_1$ with a linear combination in $\mathbf{s}_2 = H\mathbf{s}_1$ and so forth.

The way of going back to time domain from the wavelet domain is by reconstructing the signal $\mathbf{y}$ with the inverse wavelet transformation **IWT**. In matrix form the IWT is performed trivially by $\mathbf{y} = \mathbf{W}'\mathbf{d}$ which is again equivalent to apply a fast reconstruction algorithm using mirror filters. Explicitly, apply the filters h and g to the sequence $\{a_n\}$ through the adjoint operators G^* and H^* with $k-$th coordinate defined by,

$$(H^*\mathbf{a})_i = \sum_i h_{i-2k}a_i \quad \text{and} \tag{2.7}$$

$$(G^*\mathbf{a})_i = \sum_i g_{i-2k}a_i. \tag{2.8}$$

The smooth coefficients $\mathbf{s}_{J-1}$ are recovered by applying the operators to the coefficients at scale 2^J through $\mathbf{s}_{J-1} = H^*\mathbf{s}_J + G^*\mathbf{d}_J$. Repeating the process again, the next smooth coefficients are obtained $\mathbf{s}_{J-2} = H^{*2}\mathbf{s}_J + H^*G^*\mathbf{d}_J + G^*\mathbf{d}_{J-1}$. The algorithm is performed iteratively through all the levels until the signal $\mathbf{y}$ is recovered.

Assume now, that all the detail coefficients in $\mathbf{d}$ are set to zero, keeping only the smooth coefficient at the coarsest level, namely $\mathbf{s}_J$. Apply now the reconstruction algorithm as described above and find the coarse scale smooth approximation, $H^{*J}\mathbf{s}_J$, for the signal $\mathbf{y}$. If this procedure is performed for each one of the wavelets coefficients, a decomposition in time domain is obtained in what is called *Multiresolution decomposition*.

2.2.1 Multiresolution Analysis

The wavelet filters, h and g described above, are constructed from the mother and father wavelets by multiresolution analysis **MRA** theory, (see Mallat, 1989 and Daubechies, 1992 for details). The basic idea comes from the decomposition in (2.3)

by defining

$$S_J(x) = \sum_k s_{J,k}\phi_{J,k}(x) \quad \text{and} \tag{2.9}$$

$$D_j(x) = \sum_k d_{j,k}\psi_{j,k}(x). \tag{2.10}$$

The orthogonal wavelet series approximation (2.3) to the continuous signal $f(x)$ is expressed as sum of orthogonal signal components representing different scales $f(x) \approx S_J(x) + D_J(x) + D_{J-1}(x) + \cdots + D_1(x)$. The coarsest scale signal $S_J(x)$ gives a coarse scale smooth approximation to the signal $f(x)$. A refinement to the $S_J(x)$ approximation is obtained by adding the detail signal at scale 2^{J-1}, namely $S_{J-1}(x) = S_J(x) + D_J(x)$. In general, $S_{j-1}(x) = S_J(x) + D_J(x) + \cdots + D_j(x)$ is a signal approximation at scale 2^j. The collection $\{S_J, S_{J-1}, \ldots, S_1\}$ provides a *multiresolution decomposition* of $f(x)$ at scales $2^J, 2^{J-1}, \ldots, 2$. These multiresolution approximation results apply directly to discrete wavelet transformations for discrete signals $\mathbf{y}$. The intuition behind MRA is that each one of the components of the decomposition, hence approximations, is obtained by setting to zero all but one subvector of wavelet coefficients and apply the IWT to this new set of coefficients. The fact that you can reconstruct the original signal after thresholding or shrinking some of coefficients has been recognized as a useful tool in non-parametric function estimation, signal recovery and many other areas.

2.3 Wavelet Shrinkage

The discrete wavelet transformation has the property of "disbalancing" the signal by concentrating and preserving the energy of the data in a small number of wavelets coefficients. Therefore, wavelets give parsimonious transformations ensuring that the high frequency features of a signal are described by a relative small number of wavelet coefficients. Based on this principle, Donoho and Johnston (1994) developed

the wavelet shrinkage technique with the idea that some detail coefficients might be omitted without affecting the important features of a signal. For instance, additional noise in the signal can be almost removed by shrinking some low frequency wavelets coefficients towards zero, despite the fact that this noise will indeed affect all wavelet coefficients. The procedure can be summarized in three steps:

1. Transform the observed signal $\mathbf{y}$ with the DWT and obtain a set of wavelet coefficients $\mathbf{d}$,

2. shrink some or all the wavelet coefficients with a shrinkage function, say $\delta_\lambda(x)$ to obtain a new vector of coefficients $\hat{\mathbf{d}}$ and

3. reconstruct the signal by applying the IWT to the shrunken coefficients.

There are different shrinkage functions $\delta_\lambda(x)$, from which the simplest case is to remove some of the wavelet coefficients (ie. thresholding). The most common thresholding rules are **hard** and **soft** shrinkage techniques which replace the coefficients in $\mathbf{d}$ that are smaller in absolute value than a fixed threshold λ,

$$\delta_\lambda(x) = \begin{cases} x & \mathbf{1}\{|x| > \lambda\} \quad \text{Hard} \\ \text{sign}(x)(|x| - \lambda) & \mathbf{1}\{|x| > \lambda\} \quad \text{Soft} \end{cases}$$

where $\mathbf{1}\{A\}$ is the indicator function on the set A. The hard shrinkage function presents a discontinuity, keeping untouched the coefficients that are above λ and it is usually used when the problem concerns about reducing the bias in estimation. On the other hand, soft shrinkage function is continuous and it is motivated from the principle that the noise affects all wavelet coefficients. Figure 2.3 shows hard and soft shrinkage rules applied to a linear function (dotted line) for a threshold parameter $\lambda = 2$. In either case, a crucial point is the way to select the threshold parameter λ. The most commonly used rule is the **universal** threshold parameter originally proposed by Donoho and Johnston (1994) which removes all the wavelet coefficients

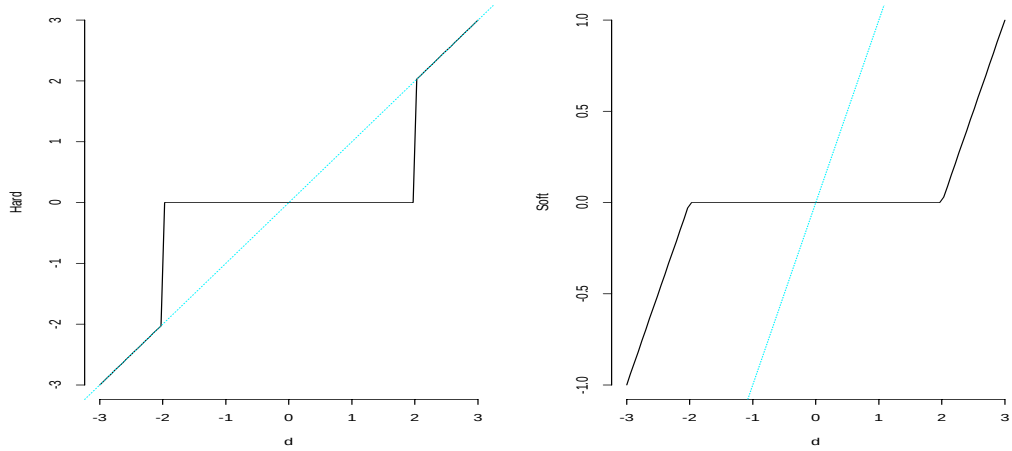

Figure 2.3: Hard and Soft shrinkage functions with $\lambda = 2$.

that are smaller than the expected maximum of an assumed uncorrelated Gaussian noise. The universal threshold is defined by $\lambda = \sqrt{2log(n)}\sigma$, where $n = 2^J$ is the length of the signal as before and σ is an estimate of the scale of the noise, traditionally computed as the sample standard deviation of the finest level of detail coefficients. Some other policies, like *minimax* and *adapt* thresholding rules are motivated on statistical theory and may even have different level dependent threshold parameters.

2.3.1 Bayesian Shrinkage Rules

One important application of wavelet shrinkage rules is in data de-noising as discussed before. For example, assume that the observed data y_i is the sum of an unobserved (latent) signal x_i and a noise component ν_i for $i = 1, 2, \ldots n$. Stacking all the equations and writing the model in vector form,

$$\mathbf{y} = \mathbf{x} + \boldsymbol{\nu} \qquad \text{with} \qquad \boldsymbol{\nu} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_n).^3$$

[3] $\mathbf{I}_n$ is the $n \times n$ identity matrix.

Wavelet coefficients are computed by linearly transforming the data with the discrete wavelet transformation through the matrix $\mathbf{W}$, namely,

$$\mathbf{d} = \boldsymbol{\theta} + \mathbf{W}\boldsymbol{\nu}, \tag{2.11}$$

where $\mathbf{d} = \mathbf{W}\mathbf{y}$ is the vector of observed coefficients and $\boldsymbol{\theta} = \mathbf{W}\mathbf{x}$ is the n dimensional vector of wavelet coefficients coming from the true signal $\mathbf{x}$. In addition, the new sequence of errors conserve the same distributional assumptions due to the orthogonal nature of the transformation, $\mathbf{W}\boldsymbol{\nu} \sim N(\mathbf{0}, \sigma^2\mathbf{I}_n)$. These facts clarify the principle that noise affects all wavelet coefficients and provide a nice way to estimate the latent process $\mathbf{x}$ by finding $\boldsymbol{\theta}$, and applying the inverse wavelet transformation $\mathbf{x} = \mathbf{W}'\boldsymbol{\theta}$.

It is well-known that Bayes rules are less ad-hoc than traditional methods and have desirable properties in estimation of the parameters. In particular, Bayesian rules shrink heavily small coefficients and only slightly larger ones. Consequently, Bayes based wavelet estimations are perfect candidates in estimating $\boldsymbol{\theta}$ in (2.11). The idea is then to find the posterior distribution of the wavelet coefficients $\boldsymbol{\theta}|\mathbf{d}$ by updating a "prior" distribution with the appropriate likelihood based on the observed wavelet coefficients $\mathbf{d}$, namely,

$$\mathbf{d}|\boldsymbol{\theta} \sim N(\boldsymbol{\theta}, \sigma^2\mathbf{I}_n). \tag{2.12}$$

See Vidakovic (1998a) for a variety of examples on this model with different prior distributions. One important issue to note here is the fact that the observed wavelet coefficients $\mathbf{d}$ are conditionally independent given $\boldsymbol{\theta}$ as presented in (2.12). However, the elements of $\boldsymbol{\theta}$ are correlated between and within levels due to the nature of the wavelet transformation.

There have been several papers proposing different prior distributions for the wavelet coefficients and the noise variance σ^2 for model (2.12). Vidakovic and Mueller (1995) suggested a hierarchical normal-inverse-gamma prior for the parameters assuming $\boldsymbol{\theta}|\sigma^2 \sim N(0, \sigma^2\Sigma)$ and $\sigma^2 \sim \text{IGamma}(\alpha, \beta)$ for fixed hyperparameters α and

β. The choice of Σ in this case was done under the assumption that the coefficients between levels are independent but correlated within levels. That is, Σ is block diagonal with sub-matrices Σ_j with different dimensions at each level and having entries $(\Sigma_j)_{i,k} = \rho^{|i-k|}$ for $|\rho| < 1$. The posterior distribution is again normal-inverse-gamma with posterior mean $\hat{\boldsymbol{\theta}} = \Sigma^* \mathbf{d}$ where $\Sigma^* = (\mathbf{I}_n + \Sigma^{-1})^{-1}$, which is difficult to handle directly for large sample sizes. The same problem of inverting $n \times n$ variances arises for different choices of priors that require the use of posterior simulation techniques like Markov Chain Monte Carlo. A different prior choice is the one proposed by Chipman *et al.* (1997) in what is called Adaptive Bayesian Wavelet Shrinkage. This technique was developed in a regression framework problem defining the prior as a mixture of two normals at each level of detail j,

$$\theta_j|\gamma_j \sim \gamma_j N(0, (c_j\sigma_j)^2) + (1-\gamma_j)N(0, \sigma_j^2),$$

where $\gamma_j \sim \text{Bernoulli}(p_j)$ for $j = 1, \ldots, J$. In the same line, Clyde *et al.* (1998), in a model selection framework, propose a mixture of a point mass at zero and a normal distribution as a prior for $\boldsymbol{\theta}$, namely $\theta_j|\gamma_j \sim N(0, (1-\gamma_j) + \gamma_j c_j \sigma^2)$. Many priors for the scale parameter σ have been proposed in the literature, some cases are based on approximations and in some other importance sampling and MCMC methods are used. In general, using unconditional mixture priors for $\boldsymbol{\theta}$ with a point mass at zero leads to *Bayesian thresholding rules*. Consequently, the resulting posterior mean for the wavelet coefficients $\boldsymbol{\theta}|\mathbf{d}$ gives smooth estimates on the edges of the signal. One even more interesting problem and the main focus of this thesis is to find latent time series structures and wavelets based methods seem to be a good starting point.

2.4 Wavelet Shrinkage in Time Series

The use of wavelet techniques for stationary processes is an emerging research that is already impacting theoretical and applied time series analyses. Some key references

are Chiann and Morettin (1994), Dijkerman and Mazumdar (1994) and McCoy and Walden (1996). One important issue in analyzing time series data with wavelets is the fact that the correlation structure of the signal interacts with the correlations inherent in the wavelet transformation. In other words, the "disbalance" property of the discrete wavelet transformations may affect the correlation structure of the time series process. An illustrative example is in the measurement error time series models, $y_t = x_t + \nu_t$, introduced above, where the latent time series process x_t is affected by additive noise ν_t representing outliers, truncations, measurement and sampling errors. In some applications, there may be strong evidence to support non-negligible additive observational errors, and they may impact severely on some inferences. Again, the use of wavelet shrinkage as a de-noising tool to get better estimates of the time series signal and the parameters involved in the model is appealing. In addition, the data can be decomposed into low frequency and high frequency components via multiresolution analysis resulting in a better understanding of the phenomena. However, the selection of the shrinkage function and the threshold parameter could add misleading information about the correlation structure of the signal x_t and hence bad estimates of the parameters could be obtained. Moreover, in almost all classical and non-classical thresholding rules applied to time series data the variances and covariances are not conserved after the reconstruction. The main problems are that, first, in time series data the high frequencies and the low frequencies are interacting by nature; second, a traditional assumption, for example in universal thresholding, is that the fine scales are modeled by uncorrelated Gaussian noise, which is obviously not satisfied in time series. In the next section, an alternative shrinkage method is analyzed with the idea of removing noise without affecting the structure of the data.

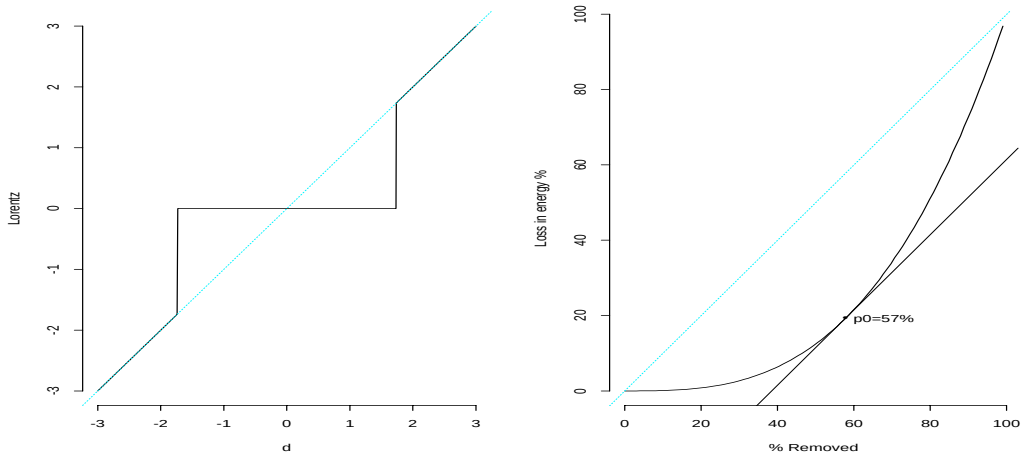

Figure 2.4: Lorentz shrinkage function.

2.4.1 Lorentz Thresholding Rule

An alternative distribution-free shrinkage method that takes into account the auto-correlation structure of the data is required, as stated before. In time series data, the distribution of energy in the wavelet domain is more disbalanced than the distribution of energy in the signal. Vidakovic (1995) proposed a thresholding method based on the Lorentz curve, which is a graphical representation of distribution inequality and a general measure of disbalance of energy. This method replaces the low energy coefficients by zero under the assumption that low energies should come from the noise component of the data. The idea is to replace the $100 \times p_0\%$ of the coefficients with the smallest energy with zero, where

$$p_0 = \frac{1}{n} \sum_i \mathbf{1}(d_i \leq \bar{d^2}),$$

and $\bar{d^2}$ is the mean of the energies $(d_1^2, d_2^2, \ldots, d_n^2)$.

The value p_0 represents the proportion at which the gain by thresholding an additional element will be smaller than the loss in the energy, both losses are measured on a 0-1 scale and are equally weighted. The left frame of Figure 2.4 shows the Lorentz thresholding rule applied to a linear function (dotted line) and the right

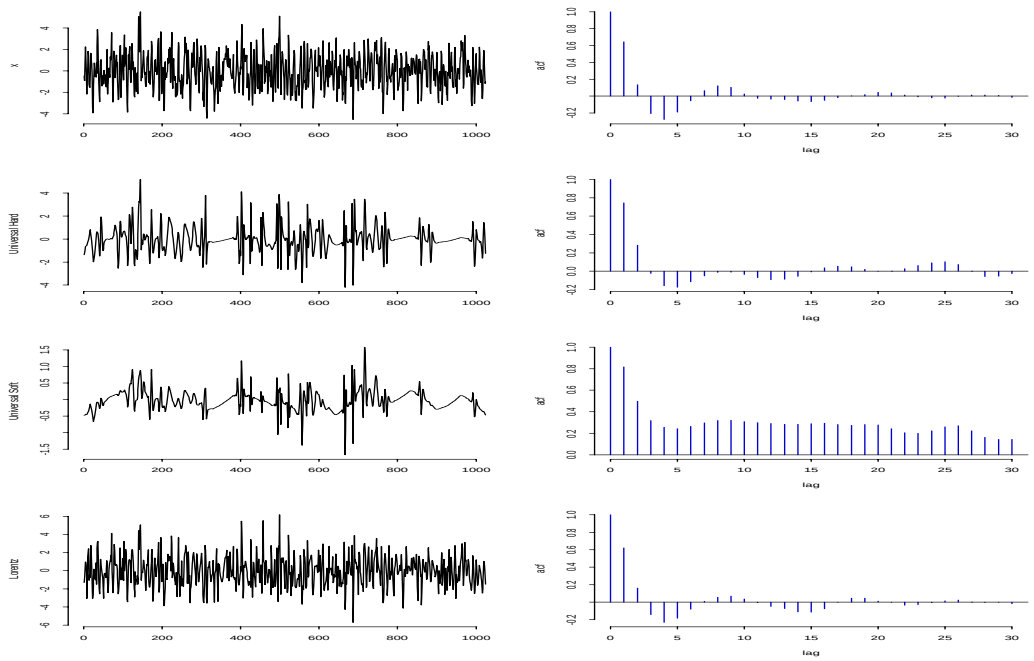

Figure 2.5: Simulated AR(2) Model.

frame shows the losss in energy (y-axis) by removing different percentages of wavelet coefficients (x-axis) with a tanget marked at $p_0 = 57\%$. As can seen in the graphs, the described thresholding procedure is equivalent to the hard thresholding for the particular threshold level of $\lambda = \sqrt{\sum_i \mathbf{d}_i^2/n}$. This choice of the threshold parameter will be more adequate for data where the low frequency terms and the high frequency terms interact a lot.

A simulated example to compare the effect of the choice of threshold parameters is now analyzed. Assume for instance that x_t is an autoregressive model of order two, namely $x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \epsilon_t$ with conditionally independent innovations $\epsilon_t \sim N(0,1)$ for $t = 1, \ldots, n$. Bayesian inferences and properties of the general AR(p) model are developed in the next chapter and are used here for illustration. A total of $n = 2^{10}$ observations were simulated from this AR(2) process for specific autoregressive parameters $\phi_1 = 1$ and $\phi_2 = -0.5$ and the resulting process is displayed in the first row of Figure 2.5 with its corresponding autocorrelation function. The

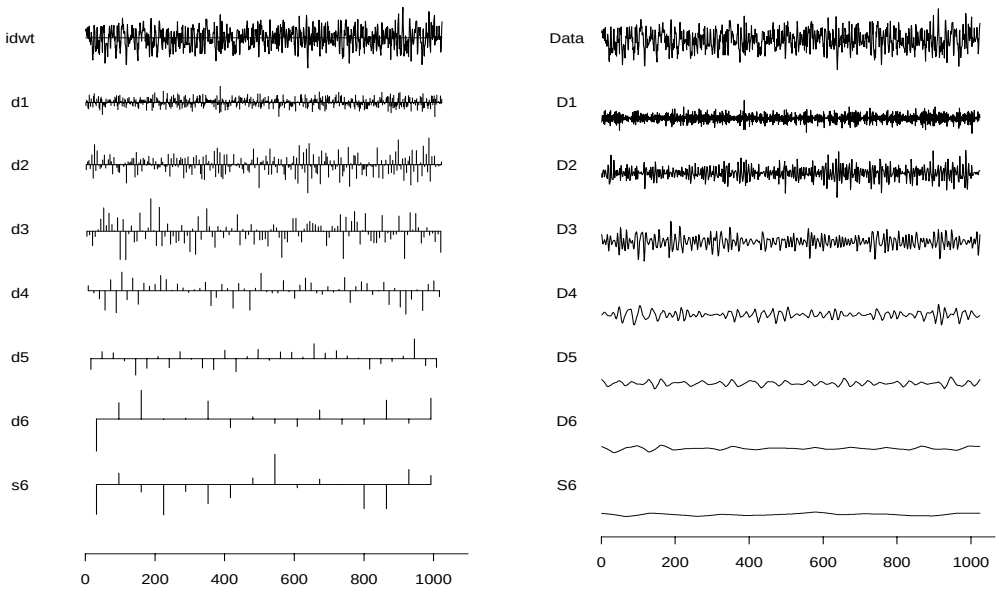

Figure 2.6: DWT and MRA of a simulated AR(2) process.

model generates quasi-cyclical behavior. The AR characteristic polynomial has a pair of complex roots with wavelength 8 and reciprocal modulus 0.7071. The signal was corrupted with additional $\nu_t \sim N(0, (0.5)^2)$ noise for a signal to noise ratio of 4. The observed data, $y_t = x_t + \nu_t$, was then decomposed using the "symmlets8" wavelet with resulting decomposition plotted in the left frame of Figure 2.6, where each one of the components represent a level in the wavelet decomposition as a result of the process of decimation. A multiresoution analysis was performed to explore the different frequency features of the data by decomposing it into additive components in time domain as stated above and plotted in the right frame of Figure 2.6. It is clear how the high frequency terms are concentrated in the finest levels of details and presumably most of the additive noise is captured in those levels too.

In order to estimate the latent process x_t, three different shrinkage functions were used to threshold some of the wavelet coefficients. Hard and Soft shrinkage function were applied first with universal threshold parameter $\lambda = \sqrt{2 * log(n)s} = 3.763$

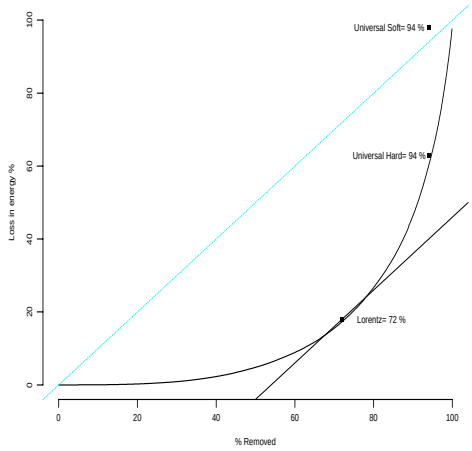

Figure 2.7: Energies of the AR(2) signal.

where $s = 0.987$ was estimated as the sample standard deviation of the first level of details $\mathbf{d}_1$, assumedly uncorrelated and normally distributed. The reconstructed processes and their autocorrelation functions are plotted in the second and third rows of Figure 2.5, for Hard and Soft shrinkage functions respectively. The effect of the wavelet decomposition in the reconstruction is very clear. In both cases, the estimated processes are over-smooth compared to the original signal and this is due to the large threshold parameter. Moreover, using the Soft shrinkage function the autocorrelation structure of the signal is completely changed providing misleading information about the nature of the latent process. The last row of Figure 2.5 shows the reconstructed series after applying the Lorentz thresholding criteria. No assumptions were required and the results are surprisingly good compared to the previous two methods. The energy of the latent process is really well preserved and the autocorrelation function exhibits practically the same features as the original signal x_t. To explain a little bit more the differences between these three methods, a comparison using Lorentz curve was performed and is displayed in Figure 2.7. The figure shows on the x-axis the percentage of coefficientes rejected and the loss in energy on the y-axis on a percentage scale too. The 45 degrees line represents the case when no thresholding is performed

and is plotted as a baseline. The Lorentz thresholding rule, by construction, is at the tangent of the curve with $p_0 = 72\%$ and energy loss of 18%, whereas the universal thresholding methods observe percentage of rejection of $p_0 = 94\%$ and energy losses of 63% and 98% for Hard and Soft respectively. This is one of the reasons of over-smoothing patterns in the reconstruction using universal thresholding methods. Bayesian inference on the autoregressive parameters was performed for each one of the reconstructed processes and for the observed noisy signal y_t. Posterior means, standard deviations (in parenthesis) and estimated moduli and wavelength are shown in Table 2.1.

	Noisy data	Hard	Soft	Lorentz
ϕ_1	0.684 (0.022)	1.210 (0.046)	1.234 (0.183)	0.853 (0.026)
ϕ_2	-0.259 (0.022)	-0.619 (0.046)	-0.505 (0.183)	-0.370 (0.026)
Wavelength	7.535	9.059	12.103	7.916
Moduli	0.509	0.787	0.711	0.608

Table 2.1: Estimations of the autoregressive parameters after thresholding.

As can be seen from the table, there is a lot of posterior uncertainty under universal thresholding methods. The resulting posterior estimates of the noisy data, that is with no thresholding at all, are improved by removing some of the wavelet coefficients according to the Lorentz criteria. The estimates of the moduli and wavelength show that the Lorentz thresholding criterion preserves most of the correlation structure of the data unlike universal methods. These results are verified in Table 2.2 showing the posterior means and standard deviations of the ϕ_1 parameter for different simulations of the AR(2) process under different signal-to-noise ratios. The posterior inferences on the autoregressive parameters under universal thresholding methods do not seem to change for different levels of noise. The reason is that the correlation structure of the reconstructed signal is very much influenced by the wavelet basis and hence is the

Signal/Noise	Noisy data	Hard	Soft	Lorentz
100	0.882(0.027)	1.159 (0.039)	1.304 (0.120)	1.089 (0.032)
5	0.676(0.021)	1.230 (0.041)	1.518 (0.145)	0.909 (0.026)
1	0.558(0.019)	1.247 (0.046)	1.476 (0.221)	0.684 (0.022)
0.45	0.491(0.017)	1.148 (0.041)	1.390 (0.176)	0.619 (0.020)
0.25	0.354(0.016)	1.248 (0.062)	1.599 (0.445)	0.457 (0.019)

Table 2.2: Estimations of the autoregressive parameters for five simulations of the AR(2) process under different signal to noise ratios.

same for all levels of noise. On the other hand, the posterior mean of ϕ_1 under the Lorentz criterion tends to shrink with more levels of noise as expected since the signal tends to be weaker. The results presented here are further explored and compared to other decompositions methods in specific applications in next chapters.

2.5 Multivariate Extensions

One possible extension of Bayesian wavelet thresholding methods involves applications with multiple time series observations. In theses cases, the correlation structure between the series could be used in the process of thresholding. That is, the information of all the series could be incorporated when selecting the wavelet coefficients to be removed. These algorithm can be then called **Ensemble Thresholding**. The basic setup up assumes observing q time series y_{it}, for $t = 1, \ldots, n$. Each one of the series will have a set of exchangeable wavelet coefficients d_{it}, coming from the same distribution. For example,

$$d_{it} \sim N(\delta_t, \sigma^2),$$

with $\delta_t \sim \pi \mathbf{1}\{\delta_t = 0\} + (1 - \pi)N(\delta_t|0, \psi^2)$ and standard priors for the other parameters. Of course there are many alternative ways to set the prior distribution for δ_t depending of the problem. In many cases for instance heavy-tailed assumptions on the observations may be desirable to model the possible differences in the series.

Chapter 3

Time Series Decomposition

In the previous chapter different non-parametric decompositions using wavelets theory were used to find hidden, or latent components that may have physical and relevant interpretations. In connection with this, some new results on time series decompositions are developed in this chapter for certain classes of dynamic models. This methodology is capable of isolating subcomponents of series with state-space representations using the theory of superposition and model structuring from West and Harrison (1997, chapters 5 and 6).

3.1 Component Dynamic Linear Models

The developments in latent structure analysis presented here begin with extensions of the measurement error models presented in section 2.4. Assume, again a scalar time series y_t, observed at equally spaced time points $t = 1, 2 \ldots$ and modeled as a dynamic linear model (DLM hereafter),

$$y_t = \mu_t + x_t + \nu_t, \tag{3.1}$$

$$x_t = \mathbf{F}_t' \boldsymbol{\theta}_t, \tag{3.2}$$

$$\boldsymbol{\theta}_t = \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \tag{3.3}$$

where x_t is the latent process combined with an additional component μ_t and the usual observational noise ν_t. The $p \times 1$ state vector is $\boldsymbol{\theta}_t$, $\mathbf{G}_t$ is the $p \times p$ evolution matrix, $\mathbf{F}_t$ is a $p-$dimensional vector of known constants and $\boldsymbol{\omega}_t$ is a $p-$dimensional vector of stochastic evolution noise or innovation. Traditionally, the error terms ν_t and $\boldsymbol{\omega}_t$ are assumed independent, mutually uncorrelated and normally distributed, $\nu_t \sim N(0, V)$ and $\boldsymbol{\omega}_t \sim \mathbf{N}(\mathbf{0}, \mathbf{W})$ for some variances V and $\mathbf{W}$ to be estimated. This general class of models naturally allows the decomposition of an observed signal y_t into additive components, μ_t, x_t and ν_t. Moreover, it assumes that the latent process x_t can be further decompose as a linear combination of subcomponents θ_{it} with weights given by $\mathbf{F}_t$. The analysis of this new class of models begins with a simple model where the process x_t is assumed to be actually observed, which is equivalent to suppose that $\mu_t = 0$ and $\nu_t = 0, \ \forall t$. In this chapter, some methodological issues on decompositions of signal processes arising from particular sub-models defined by (3.2) and (3.3) are developed in a way that the general model is naturally incorporated to the methodology.

3.1.1 Time Series Components

Consider the sub-model defined by (3.2) and (3.3) with the extra assumption that at each time point the evolution matrix $\mathbf{G}_t$ has distinct eigenvalues, denoted by α_{it} for $i = 1, \ldots, p$. The evolution matrix $\mathbf{G}_t$ can be decomposed by exploiting its eigenstructure and leading to the spectral representation $\mathbf{G}_t = \mathbf{B}_t \mathbf{A}_t \mathbf{B}_t^{-1}$, where $\mathbf{A}_t = \text{diag}(\alpha_{1t}, \ldots, \alpha_{pt})$ and $\mathbf{B}_t$ is a $p \times p$ matrix of eigenvectors corresponding to the eigenvalues in $\mathbf{A}_t$. One important point to note here, is the fact that the eigenvalues can be ordered arbitrarily and it is advisable to have the same ordering at all times to avoid identification problems.

The theory on chapter 5 of West and Harrison (1997) states that two models are

similar if and only if they have the same forecast function. Alternatively, two models are similar if and only if they have *similar* evolution matrices.[1] Following this theory, define the **similarity matrix** $\mathbf{H}_t = \mathbf{D}_t\mathbf{B}_t^{-1}$ at time t with $\mathbf{D}_t = \text{diag}(\mathbf{B}_t'\mathbf{F}_t)$ and transform the state vector $\boldsymbol{\theta}_t$ linearly by $\mathbf{c}_t = \mathbf{H}_t\boldsymbol{\theta}_t$. A new and similar DLM is created by reparametrizing (3.2) and (3.3) for the new state vector $\mathbf{c}_t = (c_{1,t}, c_{2,t}, \ldots, c_{p,t})'$, and noting that $\mathbf{F}_t'\mathbf{H}_t\mathbf{c}_t = \mathbf{1}'\mathbf{c}_t$ for $t = 1, 2, \ldots, n$. Consequently, the state space-representation of this new model is then,

$$x_t = \mathbf{1}'\mathbf{c}_t, \tag{3.4}$$

$$\mathbf{c}_t = \mathbf{G}_t^*\mathbf{c}_{t-1} + \boldsymbol{\omega}_t^*, \tag{3.5}$$

where $\mathbf{G}_t^* = \mathbf{A}_t\mathbf{D}_t\mathbf{B}_t^{-1}\mathbf{B}_{t-1}\mathbf{D}_{t-1}^{-1}$ and $\boldsymbol{\omega}_t^* = \mathbf{H}_t\boldsymbol{\omega}_t$ is a zero-mean Gaussian evolution noise. The main advantage of this reparametrization is that the process x_t in (3.4) is now rewritten as the sum of p latent processes of $\mathbf{c}_t$ related to the eigenvalues of $\mathbf{G}_t$. Note that some of the eigenvalues in $\mathbf{A}_t$ could be complex and in such cases the sum of the corresponding conjugate components c_{jt} will result in real-valued processes. Suppose for instance, that the p eigenvalues occur as c pairs of complex conjugates and $r = p - 2c$ distinct real values, then (3.4) can be expressed as

$$x_t = \sum_{j=1}^{c} z_{jt} + \sum_{j=1}^{r} y_{jt}, \tag{3.6}$$

with z_{jt} and y_{jt} corresponding to the complex roots and real roots respectively and have specific time series structure. This decomposition will be of practical importance as long as the components or a sum of them provide relevant scientific interpretations for the phenomena in study. Note that (3.6) depends only on the eigenstructure of the evolution matrix $\mathbf{G}_t$, which may depend on unknown parameters. A special case

[1] Two matrices $\mathbf{M}$ and $\mathbf{M}^*$ are similar if for some non-singular matrix $\mathbf{H}$, called **similarity matrix**, $\mathbf{M} = \mathbf{H}\mathbf{M}^*\mathbf{H}$.

of interest is that of autoregressive models, addressed in the next section and where the evolution matrix $\mathbf{G}_t$ has a specific and simple structure.

3.2 Autoregressive Models

A special and important case to consider is that of autoregressive signals. Explicitly, the process x_t will follow a traditional autoregressive model of order p if

$$x_t = \sum_{i=1}^{p} \phi_i x_{t-i} + \epsilon_t, \tag{3.7}$$

for $t = 1, 2, \ldots, n$, where ϕ_j are constant AR parameters and the innovations are assumed uncorrelated and normally distributed $\epsilon_t \sim N(0, \sigma^2)$. The process is called stationary if and only if the roots of the characteristic polynomial $\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p = \prod_{j=1}^{p}(1 - \alpha_j B) = 0$ lie outside of the unit circle. In other words, the AR(p) process is stationary if and only if the reciprocal roots α_j satisfy $|\alpha_j| < 1$, for $j = 1, 2, \ldots, p$. Consequently, the AR process in (3.7) can be expressed as an infinite sum of random shocks by inverting the equation $\phi(B)x_t = \epsilon_t$, where B is the usual backshift operator.

3.2.1 Reference Analysis

Given the first p observations, the sequential definition of model (3.7) and its Markovian properties define the joint distribution,

$$p(x_{p+1}, \ldots, x_n | \boldsymbol{\phi}, \sigma^2) = \prod_{t=p+1}^{n} p(x_t | x_{t-1}, \ldots, x_{t-p}, \boldsymbol{\phi}, \sigma^2),$$

where $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)'$. Under the traditional Gaussian innovation assumptions and writing the model in vector form, the likelihood function for the AR(p) model is obtained as $\mathbf{y}|\boldsymbol{\phi}, \sigma^2 \sim \mathbf{N}(\mathbf{F}\boldsymbol{\phi}, \sigma^2 \mathbf{I}_{n-p})$, where $\mathbf{y} = (x_{p+1}, x_{p+2} \ldots, x_n)'$ is a $n - p$

dimensional vector and $\mathbf{F}$ is a $(n-p) \times p$ matrix with t-th row $(x_{t-1}, \ldots, x_{t-p})'$. This has the form of a linear regression model and traditional methods apply. In particular, using the precision $\tau = 1/\sigma^2$ and assuming the traditional reference prior $p(\boldsymbol{\phi}, \tau) \propto 1/\tau$, the posterior distributions are given by,

$$\begin{aligned} \boldsymbol{\phi}|\sigma^2, \mathbf{y} &\sim \mathbf{N}\left(\hat{\boldsymbol{\phi}}, \sigma^2(\mathbf{F}'\mathbf{F})^{-1}\right), \\ \tau|y &\sim \text{Gamma}\left((n-2p)/2, s/2\right), \end{aligned}$$

where $\hat{\boldsymbol{\phi}} = (\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'\mathbf{y}$ and $s = (\mathbf{y} - \mathbf{F}\hat{\boldsymbol{\phi}})'(\mathbf{y} - \mathbf{F}\hat{\boldsymbol{\phi}})$ is the residual sum of squares. The marginal posterior distribution for $\boldsymbol{\phi}$ is a multivariate T with $n-2p$ degrees of freedom and has mode $\hat{\boldsymbol{\phi}}$. Monte Carlo techniques could be used to sample the posterior distribution of the reciprocal roots α_j and hence obtain posterior probabilities of stationarity of the process. Moreover, inferences on the dominant roots can be explored by ordering the posterior samples by moduli or wavelength, which is important to identify subcomponents in the decomposition described in the previous section.

3.2.2 Decomposition of AR(p) Models

The AR(p) is capable of exhibiting quasi-cyclical behavior as various distinct frequencies depending of the structure of the roots of its characteristic polynomial as explained in West (1995, 1997b). The decomposition presented in section 3.1.1 will be used to isolate sub-components coming from a state-space representation of the autoregressive model (3.7), namely,

$$x_t = \mathbf{F}'\boldsymbol{\theta}_t, \tag{3.8}$$

$$\boldsymbol{\theta}_t = \mathbf{G}\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \tag{3.9}$$

with a $p-$dimensional state vector $\boldsymbol{\theta}_t = (x_t, x_{t-1}, \ldots, x_{t-p+1})'$ and the $p-$dimensional vector $\mathbf{F}$, $p \times p$ constant and known evolution matrix $\mathbf{G}$ and the evolution noise

observing special forms,

$$\mathbf{F} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \mathbf{G} = \begin{pmatrix} \phi_1 & \phi_2 & \dots & \phi_{p-1} & \phi_p \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix}, \quad \text{and} \quad \boldsymbol{\omega}_t = \begin{pmatrix} \epsilon_t \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

The evolution noise $\boldsymbol{\omega}_t$ follows a singular normal distribution with zero mean and known constant variance $\mathbf{W}_t = \mathbf{W}$, with only one non-zero entry given by $\mathbf{W}_{1,1} = \sigma^2$. In this case, the eigenvalues of $\mathbf{G}$ are the reciprocal roots α_j of the characteristic polynomial of the AR(p) model described above. Following the result in (3.4), the time series can be decomposed into the sum of time-varying components corresponding to the autoregressive roots. Suppose as earlier that c pairs of complex conjugates $a_j \exp\{\pm i\omega_j\}$ are observed for $j = 1, \dots c$ with wavelengths or periods $\lambda_j = 2\pi/\omega_j$ representing quasi-periodic behavior in the time series. Correspondingly, write the real eigenvalues as a_j for $j = 2c+1, \dots, p$ and apply the decomposition in section 3.1.1 to reparametrize the state-space autoregression as in (3.4) and (3.5). That is, decompose the evolution matrix $\mathbf{G} = \mathbf{BAB}^{-1}$ and define a new state vector $\mathbf{c}_t = \mathbf{H}\boldsymbol{\theta}_t$ and evolution noise $\boldsymbol{\omega}_t^* = \mathbf{H}\boldsymbol{\omega}_t$, where $\mathbf{H} = \mathbf{DB}^{-1}$ and $\mathbf{D} = \text{diag}(\mathbf{B}'\mathbf{F})$. The evolution matrix in (3.5), $\mathbf{c}_t = \mathbf{G}^*\mathbf{c}_{t-1} + \boldsymbol{\omega}_t^*$ is now the diagonal matrix of reciprocal roots, $\mathbf{G}^* = \mathbf{A}$ ordered, say, according to the estimated periods or moduli. The decomposition of the AR(p) model as in (3.6) follows directly and the results is closely related to the standard partial fractions expansions of AR processes. Furthermore, the structure of the r components y_{jt}, corresponding to the real eigenvalues in the decomposition (3.6), is that of individual AR(1) processes with parameter a_j for $j = 1, \dots, r$,

$$y_{jt} = a_j y_{j,t-1} + \omega_{jt}^*$$

with errors ω_{jt}^* that are related across indices j. On the other hand, to find the structure of the z_{jt} real-valued components in (3.6), related to the sum of pairs of complex

conjugates further development is needed. Assume that the $j-$th component, z_{jt}, comes from the sum of the complex conjugates c_{ht} and c_{dt} in $\mathbf{c}_t$ for some indices d and h in $\{1, 2, \ldots, p\}$. For each j, z_{jt} can be modeled with a two dimensional DLM, namely

$$\begin{aligned} z_{jt} &= \mathbf{1}'\mathbf{c}^*_{jt}, \\ \mathbf{c}^*_{jt} &= \mathbf{G}_2\mathbf{c}_{j,t-1} + \delta_{jt} \qquad \text{with} \qquad \mathbf{G}_2 = a_j \begin{pmatrix} e^{i\omega_j} & 0 \\ 0 & e^{i\omega_j} \end{pmatrix}, \end{aligned}$$

with a two dimensional state vector $\mathbf{c}^*_{jt} = (c_{ht}, c_{dt})'$. Following the theory of canonical similar models from West and Harrison (1997, section 5.4.4), transform the state vector $\mathbf{c}^*_{jt}$ linearly by the non-singular similarity matrix $\mathbf{L} = \begin{pmatrix} 1 & 1 \\ i & -i \end{pmatrix}$ to obtain a new state vector $\mathbf{b}_{jt} = \mathbf{L}\mathbf{c}^*_{jt}$ in a similar dynamic linear model,

$$\begin{aligned} z_{jt} &= (1,0)'\mathbf{b}_{jt}, \\ \mathbf{b}_{jt} &= \mathbf{G}^*_2\mathbf{b}_{jt} + \delta^*_{jt} \qquad \text{with} \qquad \mathbf{G}^*_2 = a_j \begin{pmatrix} \cos(\omega_j) & \sin(\omega_j) \\ -\sin(\omega_j) & \cos(\omega_j) \end{pmatrix}, \end{aligned}$$

where $\delta^*_{j,t} = \mathbf{L}\delta_{j,t}$. Therefore, the components $z_{t,j}$ follow quasi-periodic ARMA(2,1) processes models with stochastically varying amplitudes and phases but with fixed periods and moduli. Explicitly,

$$z_{jt} = 2a_j\cos(\omega_j)z_{j,t-1} - a_j^2 z_{j,t-2} + \eta_{jt},$$

where the errors η_{jt} follow a zero-mean AR(1) and are correlated with ω^*_{jt}, $\forall j$. The physical meaning of the latent quasi-periodic components is the key issue here as the theory identifies and isolates those components.

The theory of similar models could be also used to justify the decomposition of the AR(p) model (3.8) through the forecast function $f_t(k) = E(x_{t+k}|\boldsymbol{\theta}_t) = \mathbf{F}'G^k\boldsymbol{\theta}_t$ in

standard notation. Exploiting the eigenstructure of the constant evolution matrix as before $\mathbf{G}^k = \mathbf{B}\mathbf{A}^k\mathbf{B}^{-1}$,

$$f_t(k) = \sum_j^p c_{jt}\alpha_j^k, \tag{3.10}$$

where c_{jt} and α_j are defined as above. Note that the decomposition in (3.6) is obtained by evaluating the forecast function at $k = 0$ as $f_t(0) = x_t$. The forecast function (3.10) has contributions coming from real and complex components. In the real case, the contribution of each root is measured by the amplitude c_{jt} which varies over time depending on the random shocks. For stationary processes, $|\alpha_j| < 1$ for all j, the forecast function decays exponentially to zero. However, if for any j $|\alpha_j| > 1$ the process is non-stationary and the forecast function explodes. In the complex case, the frequencies ω_j determine the sinusoidal oscillation in the forecast function. If the process is stationary the oscillations decay exponentially through the damping factor a_j^k. In the non-stationary case the sinusoidal variation explodes in amplitude as $|a_j|^k$ increases. In general, the forecast function is a linear combination of exponentially decaying or exploding terms and decaying or exploding factors multiplying sinusoids of different periods. In any case the general result on decomposition holds allowing for stationary and non-stationary latent processes. However, when the series has time varying patterns of dependence or non-stationarities the assumption of constant AR(p) parameters could be weak and thus a need for models that include time-varying parameters.

3.3 Time Varying Autoregressive Models

An extension to the traditional AR(p) model in (3.7), is that of time varying autoregressions. These models, adapt better to non-stationary time series by including time variation in the autoregressive parameters. Moreover, the fact that the parameters

change over time induces the notion of instantaneous stationarity. Using the same notation as before, the process x_t follows a time-varying autoregressive model of order p or TVAR(p) if for $t = 1, 2, \ldots, n$,

$$x_t = \sum_{i=1}^{p} \phi_{it} x_{t-i} + \epsilon_t \tag{3.11}$$

where the innovations are uncorrelated and normally distributed $\epsilon_t \sim N(0, \sigma^2)$. The process will be stationary at time t if and only if the reciprocal roots, α_{jt}, of the characteristic polynomial $\phi_t(B) = 1 - \phi_{1t}B - \cdots - \phi_{pt}B^p$ satisfy $|\alpha_{jt}| < 1$ for all $j = 1, 2, \ldots, p$. The TVAR(p) model (3.11) is completely specified by incorporating an additional evolution model for the coefficients, often taken as a random walk $\boldsymbol{\phi}_t = \boldsymbol{\phi}_{t-1} + \boldsymbol{\omega}_t$ for some zero mean innovation $\boldsymbol{\omega}_t$. The extensions of the time series decompositions are mainly focussed in latent structure assessment as explained in the next section.

3.3.1 Decomposition of TVAR(p) Models

Following the theory from section 3.2.2, the TVAR(p) model (3.11) can be written in state space form as,

$$x_t = \mathbf{F}'\boldsymbol{\theta}_t \tag{3.12}$$

$$\boldsymbol{\theta}_t = \mathbf{G}_t\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t \tag{3.13}$$

where $\mathbf{F}$, the state vector $\boldsymbol{\theta}_t$ and the innovations sequence $\boldsymbol{\omega}_t$ are the same as in (3.8). The time varying structure of the model is introduced in the evolution variance,

$$\mathbf{G}_t = \begin{pmatrix} \phi_{1t} & \phi_{2t} & \ldots & \phi_{p-1,t} & \phi_{pt} \\ 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \ldots & 1 & 0 \end{pmatrix}.$$

With this form of the model and for any specified values of the ϕ_{jt} parameters, an instantaneous time series decomposition can be performed as described in section 3.1.1. In the same way as before, reparametrize the model and write it as in (3.4) and (3.5) by exploiting the eigenstructure of $\mathbf{G}_t$ at each time point. The decomposition (3.6) is obtained for c pairs of complex conjugates and $r = p - 2c$ distinct real eigenvalues $\boldsymbol{\alpha}_t = (\alpha_{1t}, \ldots, \alpha_{pt})'$ at time t. Note that since the estimate of the $\boldsymbol{\phi}_t = (\phi_{1t}, \ldots, \phi_{pt})'$ vector changes over time, the moduli a_{jt}, and the wavelengths λ_{jt} of the eigenvalues change too.[2]

In the constant AR(p) model, the evolution matrix $\mathbf{G}_t^*$ in (3.5) is simply the diagonal matrix of the reciprocal roots. However, in the TVAR(p) case such simplification of $\mathbf{G}_t^*$ is not possible and the structure of subcomponents is not as clear as before. For instance, if one assumes that $\mathbf{D}_t\mathbf{B}_t^{-1}\mathbf{B}_{t-1}\mathbf{D}_{t-1}^{-1} \approx \mathbf{I}_p$ as it is the case in many applications then $\mathbf{G}_t^* = \mathbf{A}_t$ and the results from the AR(p) model extend immediately; see Prado and West (1997) for further discussion and details. Explicitly, at each time point the r components y_{jt}, related to the real roots, follow *approximate* TVAR(1) models,

$$y_{jt} = a_{jt}y_{j,t-1} + \omega_{jt}^*,$$

with zero-mean innovations ω_{jt}^* related across j. The c components z_{jt} related to pairs of complex roots follow *approximate* quasi-cyclical TVARMA(2,1) models,

$$z_{jt} = 2a_{jt}\cos(\omega_{jt})z_{j,t-1} - a_{jt}^2 z_{j,t-2} + \eta_{jt}$$

with TVAR(1) errors η_{jt} correlated with ω_{jt}^*, $\forall j$. In the case where the components z_{jt} have physical interpretations, inferences on the changes of the periods and moduli are very important since they represent patterns of time variation in spectral characteristics of the signal.

[2]Same notation as before, a_{jt} for the real eigenvalues and $a_{jt}\exp\{\pm i2\pi/\lambda_{jt}\}$ for the complex conjugates pairs.

Once again note that the eigenvalues of $\mathbf{G}_t$ could be ordered in an arbitrary way, bearing in mind that the same order should be used at all the time points. In connection with this, a particular problem in the decomposition arises due to the fact that the assumption of r real eigenvalues and c complex may not be valid for all t. This identification problem may impact the estimation and interpretation of some of the components of the decomposition.

In general, the decomposition results for the AR(p) and TVAR(p) models derived above assume specific values of the AR parameters, known x_t process and different eigenvalues. A common choice for autoregressive parameters is the posterior mean which usually leads to different eigenvalues in any case. One obvious extension to the decomposition is to incorporate measurement error into this model by fitting higher order autoregressions in higher frequency terms see West (1995, 1997b) and West and Harrison (1997) for further details. More formal extensions for general models to include non-Gaussian observation errors ν_t, outliers and possibly non-stationary trend terms μ_t as in as in (3.1) from section 3.1 will be addressed in next chapters. This generalizations involve taking the x_t process as latent and unknown which is of key interest in many applications.

Chapter 4

Turbulence Data

In any kind of decomposition method and latent structure analysis the components and underlying processes should be relevant enough to help the scientist understand the phenomena in study. In this chapter an illustrative example using turbulence data from Katul and Vidakovic (1996) is explored and the two methodologies developed in the previous two chapters are applied. In this specific application the decomposition into low frequency and high frequency components is an important research topic in land-atmosphere interaction and atmospheric surface layer studies.

4.1 Turbulence

The structure of turbulent eddy motion in the atmospheric surface layer plays a central role in the transport mechanics of heat, mass, and momentum from ground into the atmosphere. Organized and coherent events, usually called **attached eddies**[1], are responsible for much of the heat and momentum transfer in boundary layer flows. These eddies are surrounded by a fluid that contains fine-scale eddies, called **detached eddies** , which do not contribute significantly to the production of turbulent

[1]The name comes from the theory that the mean-flow vorticity and the energy containing turbulent motions are caused by anisotropic coherent eddies attached to the wall, Townsend (1976).

fluxes and kinetic energy; see Figure 4.1 for a representation of eddies under different degrees of turbulence. These less-organized eddies are known to be statistically isotropic and follow Kolmogorov's theory (1941) **K41** hereafter. The quantification of the large-scale eddy motions from time series measurements of turbulent flow variables is of key interest to understand land-atmosphere interactions and thus a decomposition of turbulence data into attached and detached eddy motion becomes relevant.

The major problem in separating the attached eddy motion from time series measurements is the locality and non-periodicity of the organized events. The problem is addressed using the two different approaches for decompositions developed in chapters 1 and 2. First, the locality of wavelets in time is used to isolate the scale contributions of events in space. For this matter, Lorentz thresholding criterion, (Goel and Vidakovic, 1995; Vidakovic, 1995), is used to eliminate wavelet coefficients with small contributions to the total turbulent energy as described in section 2.4.1. Second, the time series decomposition in section 3.2.2 is applied to explore latent cyclical components in the time series after fitting higher order auto-regressions. The time series is decomposed into the sum of time-varying components corresponding to the auto-regressive roots. In this way it is possible to isolate sub-components of the series and their contribution to the decomposition. Scientific theory in this area, suggests that the results for both methods may be validated if they are consistent with the Townsend (1976) attached eddy hypothesis.

- The wavelet filtered series and the low frequency components in the time series decomposition will be the organized and attached eddies and should explain the majority of the variance of heat and momentum from the observed data.

- The difference between the original signal and the attached eddy motion should follow Kolmogorov's **K41** theory. That is, the detached and less organized

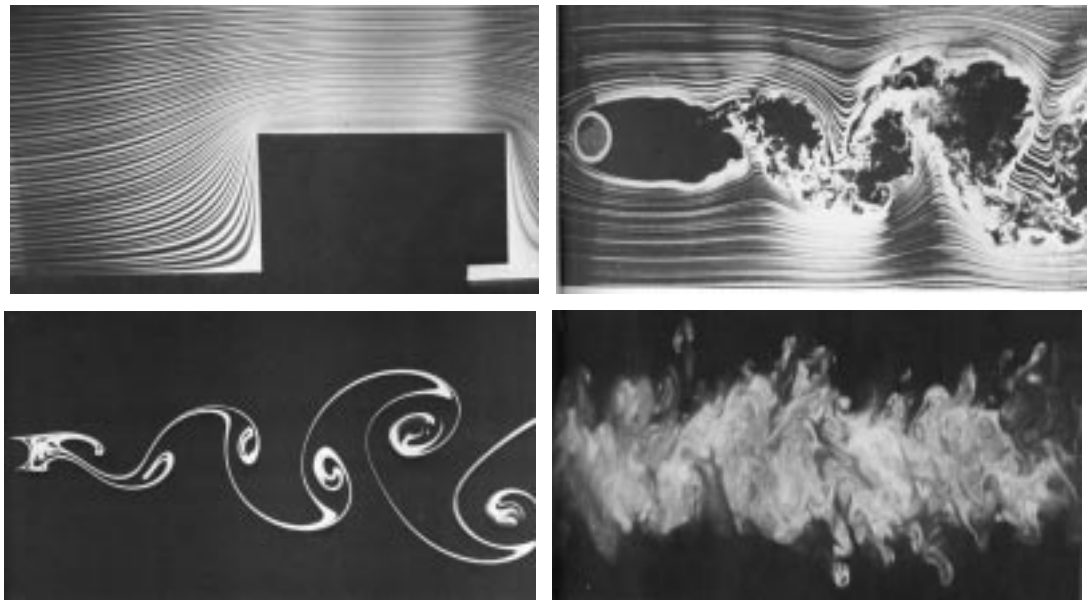

Figure 4.1: Eddies for different degrees of turbulence.

eddies should be close to the "-5/3 Power Law", $E_x(K) \propto K^{-5/3}$ where $E_x(K)$ is the Fourier power spectral density.

4.2 The Experiment

In 1993, a total of 50,000 measurements of three velocity components and air temperature were taken over a uniform dry lake bed in Owens valley, California on an elevation of $1,100m$. The momentum roughness length for this sandy surface was $0.13mm$; see Katul (1994) for details. The velocity components were measured in a range of $2m - 3.5m$ above the surface using a 56 Hz triaxal ultrasonic anemometer with a sampling period of 9.75 minutes. The velocity components were rotated to obtain measurements along longitudinal, lateral and vertical velocities. For the purpose of illustration in this chapter, an equally spaced sample of 10,000 was taken from the 50,000 observations on two of the velocity components. These samples are presented in Figure 4.2 for the longitudinal and vertical velocities denoted by U and W respectively.

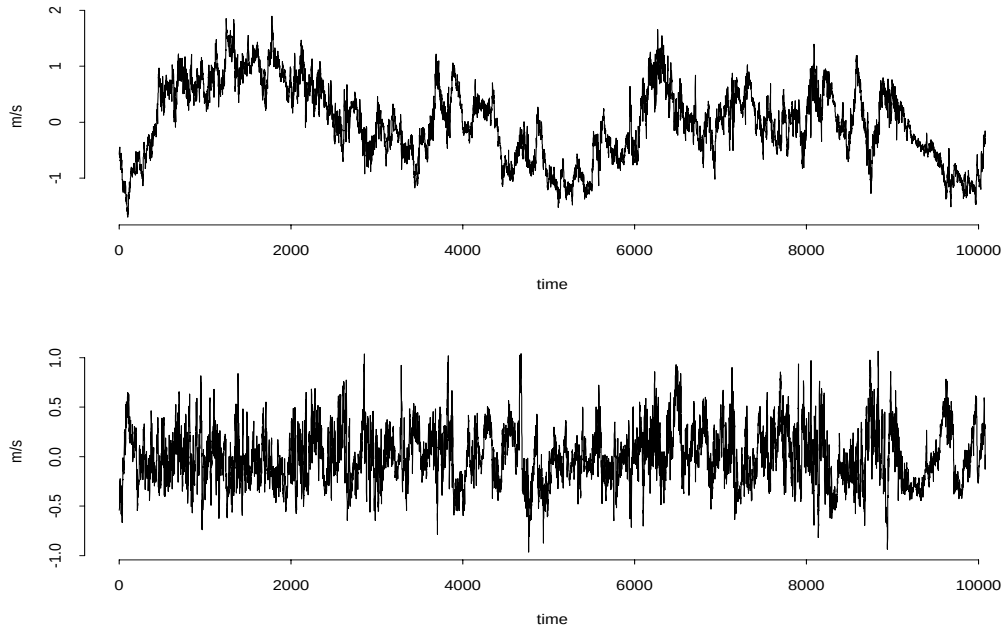

Figure 4.2: Turbulence Data: Velocity components measured over longitudinal direction U (top) and over vertical direction W (bottom).

4.2.1 Wavelet Analysis

The detailed structure of this two flow variables will be analyzed, as stated before, using two different methodologies in the context of Townsend's attached and detached eddy hypothesis. First, a wavelet decomposition of both signals were performed using the "symmlets8" wavelet basis. Thresholding criteria was then used to remove the wavelet coefficients with smaller contributions to the total energy, as detailed in section 2.4.1. Lorentz curve criteria was established when the gain in increasing the number of wavelet coefficients and the loss in energy were in balance. Approximately 87.42% and 86.39% wavelet coefficients below this criteria were removed, representing loss of energy of about 11.43% and 12.82% for U and W respectively. Figure 4.3 shows the Lorentz wavelet thresholding curves for U and W where the diagonal line represents a well balanced signal.

The point, p_0 is the tangent line parallel to the diagonal representing clearly the

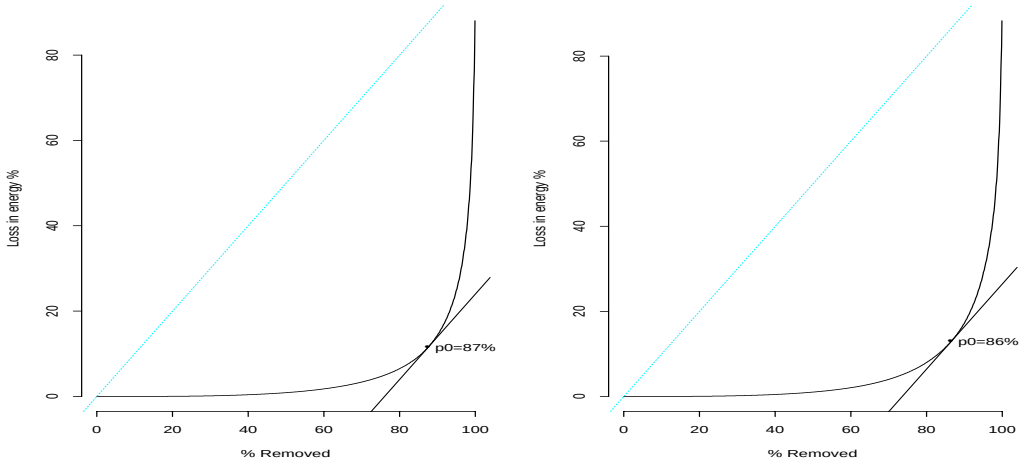

Figure 4.3: Lorentz wavelet thresholding of velocity components.

break-even point between percentage of coefficients rejected and loss in energy. The reconstruction was performed applying the inverse wavelet transformation to the new set of coefficients. The resulting series represents the attached eddies in each one of the velocity components. Furthermore, the difference between the original signal and the filtered series will be the detached eddy motion part of the series.

4.2.2 Time Series Decompositions

The second analysis of the series was done using the time series decomposition method described in chapter 2. For this kind of data a traditional way to start the analysis is by fitting higher order AR models to approximate what may be lower order ARMA models or non-linear features in the series. In this case, two constant AR(10) models were fitted to each one of the velocity components with reference priors as described in section 3.2. Figure 4.4 presents 95% posterior intervals for each one of the AR coefficients ϕ_j from model (3.7) and for both flow variables (U left frame and W right frame).

As can be seen from the picture, both series show similar patterns having the first coefficient around 0.9 and the rest of the coefficients close to zero. Monte Carlo

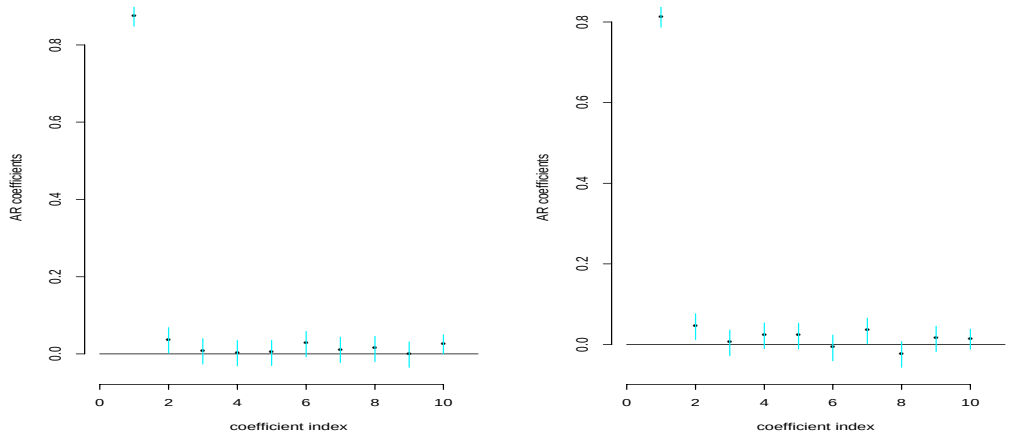

Figure 4.4: 95% posterior intervals for the AR(10) coefficients of the U (left) and W (right) velocity components

techniques were used to sample the posterior distribution of the reciprocal roots of the characteristic polynomial. That is, a random sample was drawn from the posterior distribution of the AR coefficients $\boldsymbol{\phi}$ for each one of the velocity components and the reciprocal roots were computed for each sample. In both cases, two real and four pairs of complex roots were observed. For the longitudinal velocity component U, posterior distributions of the wavelengths/periods λ_j and moduli a_j for the two dominant complex roots are displayed in Figure 4.5. The posterior means for the two largest periods were 12.06 and 5.57 with corresponding moduli of 0.73 and 0.65. The results were almost identical for the vertical component W and are not displayed here. Both series were then decomposed into the sum of time-varying components corresponding to the autoregressive roots, as developed in section 3.2.2 and both decompositions are displayed in Figure 4.6. The roots were ordered by wavelength but they could have been ordered by moduli or amplitude. The decomposition was just the sum of six real components, two corresponding to the real roots following AR(1) processes and four corresponding to the sum of the complex conjugates following quasi-cyclical ARMA(2,1) processes. Most of the implied latent components with lower moduli or

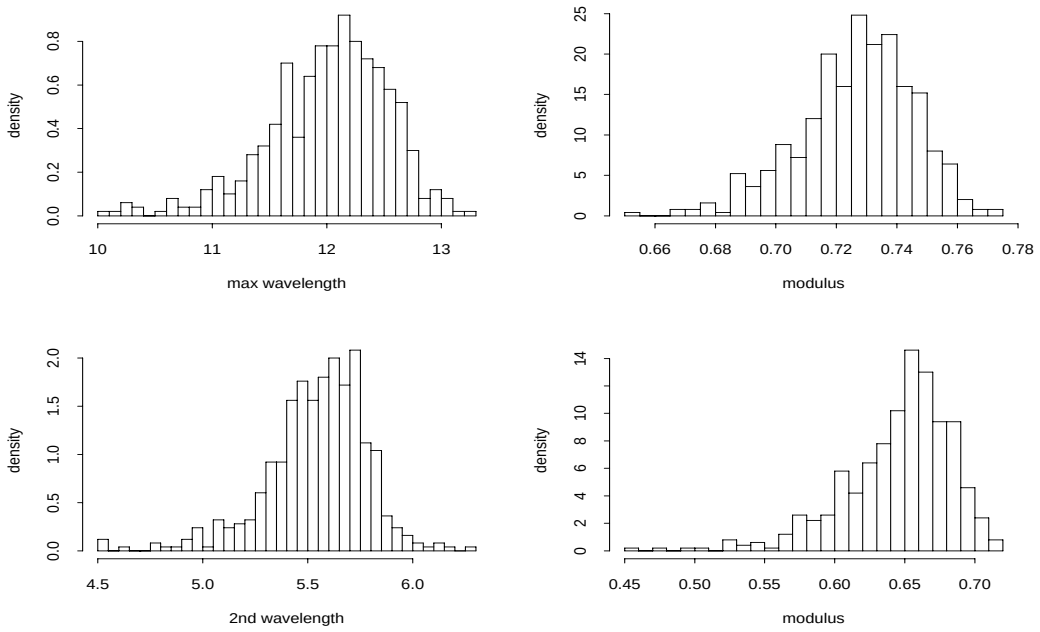

Figure 4.5: Posterior distributions of the wavelength and modulus of the two dominant complex roots for the longitudinal velocity component U.

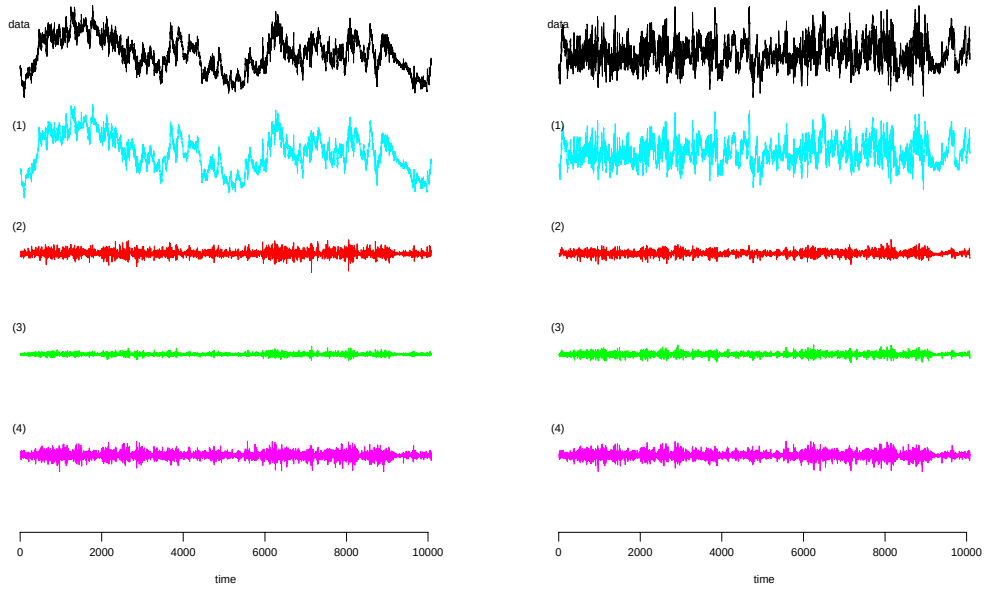

Figure 4.6: Time series decompositions of velocity components, U (left) and W (right).

very high frequencies are introduced to adequately capture the correlation structure in the series but do not represent physical meaningful components. For instance, in this particular application the interest lies only on two components corresponding to the attached and detached eddy motion. In each one of the decompositions of Figure 4.6, the original series, top row, is decomposed into three components. The dominant component, plotted in the second row, is related to a real root and can be interpreted as the trend of the series. The next component, in the third row, is related to the pair of complex roots with largest wavelength and the last row is the sum of the rest four components. In both cases the most dominant component was that first real root and the rest of the components have smaller periods, smaller moduli and are negligible in amplitude as expected. In the context of turbulence data, the first real component represents the attached eddy motion and the sum of rest of the components the detached eddy motion.

4.2.3 Results and Validations

The top two frames of Figure 4.7 display the first 250 points of the original signal (dotted line) for the longitudinal velocity component U on the left and the vertical velocity component W on the right. The estimated attached eddies for each flow variable series calculated with the time series decomposition is overlaid correspondingly. The bottom two frames show the same comparison using the estimated attached eddies computed using the Lorentz curve thresholding criteria. It is clear that large scale-eddies are well captured in both flow variables and using two different approaches. Note however, that the wavelet thresholding method results are smoother than those from the time series decomposition maybe because of the choice of the wavelet basis.

The decompositions found so far need to be validated with Townsend's theory.

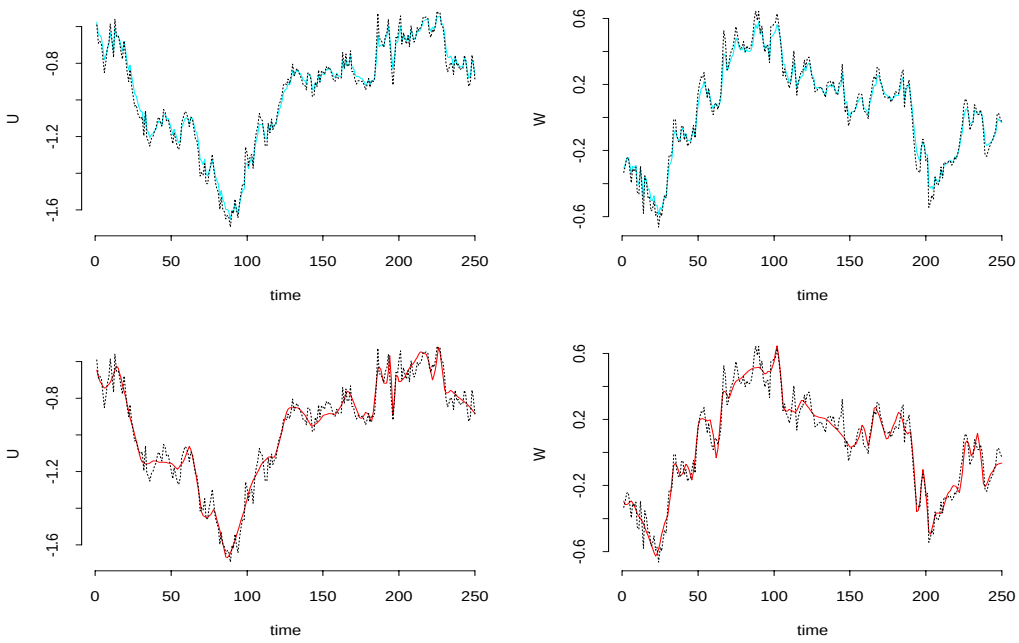

Figure 4.7: Original signal(dashed line) vs estimated attached eddy motion (solid line) calculated with time series decomposition (top frames) and with wavelet thresholding (bottom frames).

This theory claims that velocity components can be decomposed into

$$u_i = u_i^a + u_i^d \quad \text{and} \quad w_i = w_i^a + w_i^d,$$

where a and d represent attached and detached eddy motion respectively. Theory also suggests that for the attached eddy motion the following relations should hold:

- $\mathrm{E}(U^a) = \mathrm{E}(W^a) = 0$,
- $\sigma_U^2 = E((U^a)^2), \quad \sigma_W^2 = E((W^a)^2)$, variance conservation and
- $E(UW) = E(U^a W^a)$ covariance conservation.

Validations for Lorentz thresholding criteria and time series decomposition are presented in Table 4.1 for variance conservation and Table 4.2 for covariance conservation. As can be seen from Tables 4.1 and 4.2, in both flow variables about 99% for U

	U	Component	W	Component
Original	0.4420		0.0741	
Lorentz	0.4382	99.14 %	0.0692	93.39 %
TS Decomp.	0.4284	96.92 %	0.0689	92.98 %

Table 4.1: Variance Conservation

and 95% for W of the turbulence variance and fluxes are retained by the estimated attached eddy motion with Lorentz Thresholding criteria. On the other hand, with the proposed time series decomposition about 97% and 93% of the variance is conserved for U and W respectively. The covariance between U and W is preserved in more than 90% with both methodologies.

	Covariance	% Conserved
Original Series (UW)	-0.0230	
Lorentz Thresholding	-0.0224	97.39 %
Time Series Decomp.	-0.0210	91.31 %

Table 4.2: Covariance Conservation (UW)

A different validation procedure is presented in Figure 4.8. Each frame shows on the horizontal axis the resolution level number in a wavelet decomposition, 1 for the finest level and 11 for the coarsest level. On the vertical axis, $\log_2(\sum d_i^2)$, the logarithm of the energy is plotted for each level. This plot is useful to validate how the original series was reproduced by the estimated attached eddies at different frequencies. The top frames display the frequency composition of the attached eddies compared with the original series for U and W on the left and right frames respectively and for both methodologies. In any case, at low frequencies the differences between the original series and the estimations are negligible in both variables.

The detached eddies were validated according to Townsend's hypothesis as well. specifically, these fine-scale eddies should follow the frequently called "-5/3 Power

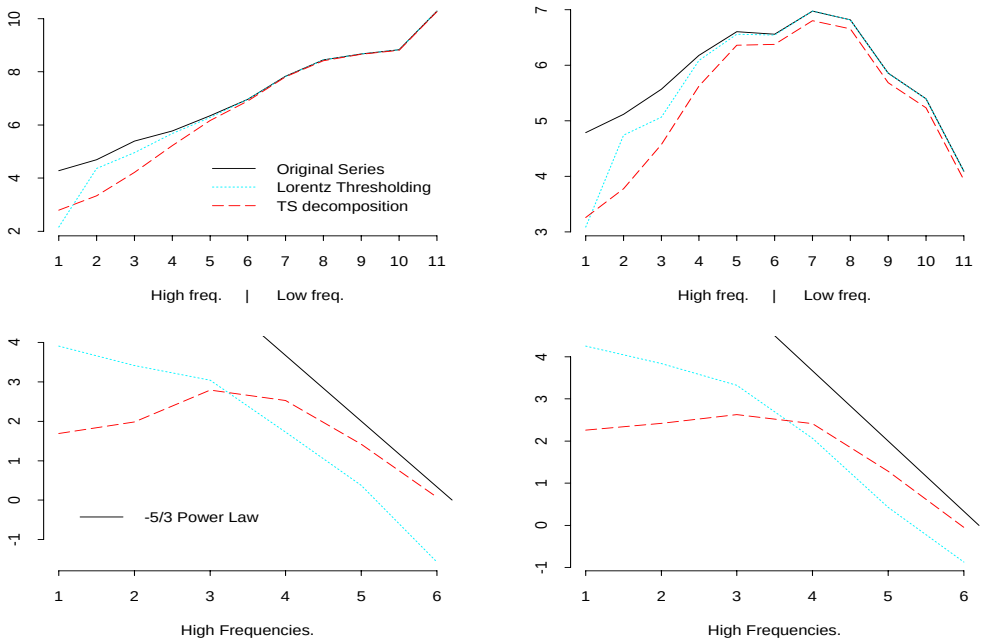

Figure 4.8: Energy at different frequencies (attached eddy motion top frames) and (dettached eddy motion bottom frames)

Law" suggested in Kolmogorov's **K41** scaling theory:

$$E_x(K) \propto K^{-5/3},$$

where $E_x(K)$ is the Fourier power spectral density and K is the wavenumber. The detached eddies were calculated taking the difference between the original signal and the filtered series for the Lorentz thresholding criteria. In the case of the time series decomposition, the fine-scale eddies were represented by the sum of all but the dominant real component. The bottom frames of Figure 4.8 illustrate a comparison of the energies of the estimated detached eddies with the $-5/3$ line at high frequencies (wavelet levels 1 through 6). It seems that the fine-scale eddies for both variables follow the $-5/3$ power law consistent with Kolmogorov's **K41** theory.

4.2.4 Conclusions

Some preliminary conclusions can be established from the results on this chapter. Two different methodologies have been used to make inferences on latent components that are present in time series data. In the wavelet decomposition, the reconstructed signal after thresholding is the estimation of a latent process and usually conserves most of the energy of the signal. Consequently, the wavelet based-solution is a good candidate for the attached eddy motion part of the velocity components in the turbulence example described above. In addition, the attached eddies are consistent with Townsends's theory and the detached eddies, calculated as the difference between the original signal and the attached eddies, follow Kolmogorov's **K41** theory. Nevertheless, the wavelet solution usually observes smooth features in the reconstruction of the latent process possibly due to the choice of the wavelet basis. In relation with this, Katul and Vidakovic (1996) developed an algorithm to find the optimal wavelet basis for turbulence data based on minimizing a relative entropy measure, which allows to maximize the discrimination procedure between organized and less organized eddies.

On the other hand, a time series decomposition as developed in section 3.2.2 is used to explore and isolate quasi-cyclical components of AR models at different frequencies. The time series is decomposed into time-varying components related to the roots of the characteristic polynomial of the model. The results in the turbulence example are also consistent with Townsend's attached eddy motion theory and the detached eddies follow Kolmogorov's theory as well. The two velocity components analyzed evidence a dominant real root interpreted as the attached eddies having a value very close to one in absolute value. This suggests the possibility of non-stationary models for the data and hence the need of decomposition results of models that allow for non-stationary components as in the case of time-varying autoregressions from

section 3.3.1.

The next step in the study of latent structure in time series is to consider the general time series component dynamic linear model presented in section 3.1. That is, assume that the process x_t is unobservable by including a trend and measurement error components in the models and estimate the process x_t and their sub-components together. These models are developed in the next chapter allowing for the possibility of having non-Gaussian error terms modeled by mixtures of normals.

Chapter 5

Heavy Tailed Components

The recent developments for analyzing univariate latent time series components, presented in chapters 2 and 3, are useful for identification and interpretation of underlying processes. This was clearly illustrated in the turbulence example of chapter 4 where the two different methodologies were applied giving similar and comparable results consistent with the scientific theories. The methodology is now extended to include additional components in the decompositions in the context of the component dynamic models from section 3.1. The generalizations here involve decomposition analyses of latent non-stationary time series models in the time-varying autoregressive framework described in section 3.3.1. In addition, the setup allows the inclusion of heavy-tailed distributions for the innovations helpful in modeling quite radically ill-behaved series. Moreover, the model accounts for possible stochastic changes in variance of the latent autoregressive processes modeled through discounting methods following West and Harrison (1997, chapter 8). This general class of models leads to important theoretical contributions and associated computational algorithms for model fitting and exploration. Consequently, a major component of this chapter is in developing a posterior sampling algorithm via MCMC to make inferences on model parameters and latent processes. These methodological developments in Bayesian

multivariate time series have potential in a wide variety of scientific fields and are presented here in the analysis of eight chemical species obtained from the Greenland Ice Sheet covering the period $41 - 0$ Kyr. BP (Marsh and Ditlevsen, 1996).

5.1 Heavy Tailed Components

The results developed in this section are natural extensions of the state-space autoregressions presented in West and Harrison (1997, section 15.3). Generalizations to the Component Dynamic Linear Model introduced in section 3.1 involve adding regression variables together with other latent processes representing different features of the signal. All these generalizations could be performed by modifying the basic model with two components and special evolution relations. Explicitly, consider a DLM of the form

$$\begin{aligned} y_t &= \mu_t + x_t + \nu_t, && (5.1) \\ \mu_t &= \mu_{t-1} + \omega_t, && (5.2) \\ x_t &= \sum_{j=1}^{p} \phi_{jt} x_{t-j} + \epsilon_t, && (5.3) \end{aligned}$$

for $t = 1, \ldots n$, where μ_t represents an underlying first order polynomial trend, x_t is a latent time-varying autoregressive model and ν_t is the observational noise. The three sequences of error components are assumed to be independent and mutually independent with conditionally normal distributions. That is,

$$\begin{aligned} \nu_t &\sim N(0, V), \\ \omega_t &\sim N(0, W/\tau_t) \text{ and} \\ \epsilon_t &\sim N(0, \sigma_t^2/\lambda_t). \end{aligned}$$

for $t = 1, \ldots n$. The additional weight parameters τ_t and λ_t are introduced to allow for errors with heavy-tailed distributions modeled by mixtures of normals. In particular,

scale mixtures of normals with prior mixing distributions are used to model non-normal behavior of the stochastic level changes ω_t and autoregressive innovations ϵ_t via,

$$\begin{aligned} p(\omega_t|W) &= \int p(\omega_t|0, W, \tau_t)p(\tau_t)d\tau_t \quad \text{and} \\ p(\epsilon_t|\sigma_t^2) &= \int p(\epsilon_t|0, \sigma_t^2, \lambda_t)p(\lambda_t)d\lambda_t. \end{aligned}$$

There are different choices of prior distributions that can address different features of the error distributions. For instance, the class of Gamma priors for the weights leads to marginal t-distributed error distributions. In particular, assuming $\tau_t \sim \text{Gamma}(m_1/2, m_1/2)$ and $\lambda_t \sim \text{Gamma}(m_2/2, m_2/2)$ for all t, the marginal distributions of ω_t and ϵ_t are Student-T with m_1 and m_2 degrees of freedom respectively. In cases where the changes in trend are expected to be bigger than the ones coming from the AR innovations, a fatter tailed distribution on the ω_t is needed and thus $m_1 \ll m_2$. Note that the traditional normal model is a particular case of this general setting when $m_1, m_2 \to \infty$, or equivalently $\tau_t = 1$ and $\lambda_t = 1$ for all t.

The setup of the model and the time-varying autoregressive structure of the latent process x_t allows for further decompositions in quasi-cyclical components as developed in section 3.3.1. However, the decomposition results are based on observing the x_t process itself and therefore the interest relies now in making inferences about the latent components μ_t and x_t, together with the autoregressive parameters $\boldsymbol{\phi}_t$, and variance components V, W, σ^2, τ_t and λ_t. The theoretical developments begin by defining a $(p+1)-$dimensional state vector $\boldsymbol{\theta}_t = (\mu_t, x_t, \ldots, x_{t-p+1})'$ and writing the model in the state-space form as before,

$$y_t = \mathbf{F}'\boldsymbol{\theta}_t + \nu_t \quad \text{and} \quad \boldsymbol{\theta}_t = \mathbf{G}_t\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t,$$

where the $(p+1)$ constant vector $\mathbf{F}$, the $(p+1) \times (p+1)$ evolution matrix $\mathbf{G}_t$ and

the $(p+1)-$dimensional evolution noise $\boldsymbol{\omega}_t$ are given by,

$$\mathbf{F} = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \mathbf{G}_t = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & \phi_{1,t} & \phi_{2,t} & \dots & \phi_{p-1,t} & \phi_{p,t} \\ 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 0 \end{pmatrix}, \quad \text{and} \quad \boldsymbol{\omega}_t = \begin{pmatrix} \omega_t \\ \epsilon_t \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

The evolution noise sequence $\boldsymbol{\omega}_t$, follows a conditional singular normal distribution with zero mean and variance,

$$\mathbf{W}_t = \begin{pmatrix} W/\tau_t & 0 & 0 & \dots & 0 \\ 0 & \sigma_t^2/\lambda_t & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix}.$$

A posterior analysis for this class of models is performed using Markov chain Monte Carlo (MCMC) simulation methods which are the standard for implementation of Bayesian inference in other than very simple models.

5.2 Implementation of the Gibbs Sampler

A Markov Chain Monte Monte Carlo algorithm specifies an irreducible and aperiodic Markov Chain with stationary distribution given by the desired joint posterior distribution. An implementation of the posterior sampling algorithm via Gibbs sampling is outlined here for the unknown parameters $\{V, \boldsymbol{\phi}_t, \sigma_t^2, W, \lambda_t, \tau_t, \mathbf{Z}_t; \forall t\}$ where $\mathbf{Z}_t = \{\boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_t\}$ is the set of state vectors up to time t. The sampling scheme is based on iterative updating using the full conditional densities of any subsets of the unknown parameters, call it ξ. The remaining variables combined with the full data set, $\mathbf{D}_n = \{y_0, y_1, \dots, y_n\}$, will be then denoted ξ^-.

Sampling the state vector, $\theta_t|\theta_t^-$

The conditional posterior distributions for the state vectors $\boldsymbol{\theta}_t|\boldsymbol{\theta}_t^-$ developed here follow the simulation method named **Forward Filtering, Backwards Sampling (FFBS)** originally developed by Carter and Kohn (1994), Frühwirth-Schnatter (1994) and detailed in West and Harrison (1997, section 15.2.3). The Forward Filtering step is performed by running a standard DLM analysis for $t = 0, 1, \ldots n$ to obtain forward updates of the state vector distributions $\boldsymbol{\theta}_t|\mathbf{D}_t \sim N(\mathbf{m}_t, \mathbf{C}_t)$ as explained in the Appendix A.1. The backwards sampling is performed by sequentially sampling $\boldsymbol{\theta}_n, \boldsymbol{\theta}_{n-1}, \ldots, \boldsymbol{\theta}_1$ from the backwards distributions $p(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t+1}, \mathbf{D}_t)$ for $t = n-1, n-2, \ldots, 1, 0$. The overall procedure involves three steps:

1. Compute the standard forward updates $\boldsymbol{\theta}_t|\mathbf{D}_t \sim N(\mathbf{m}_t, \mathbf{C}_t)$ for $t = 0, 1, \ldots n$,

2. sample the last state vector from $\boldsymbol{\theta}_n|D_n \sim N(\mathbf{m}_n, \mathbf{C}_n)$ and

3. sample backwards through time for $t = n-1, n-2, \ldots, 1, 0$, sequentially from $\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t+1}, \mathbf{D}_t \sim N(\mathbf{h}_t, \mathbf{H}_t)$ where $\boldsymbol{\theta}_{t+1}$ is the value sampled in the previous step.

In the special case of component dynamic linear models (5.1), with time-varying autoregressive components, the forward filtering step is performed by following the updating equations described in the Appendix A.1 to obtain the sequence of moments $\mathbf{m}_t$ and $\mathbf{C}_t$ together with the auxiliary quantities $\mathbf{a}_t, \mathbf{R}_t$ and $\mathbf{B}_t$ for $t = 0, 1, \ldots n$. At $t = n$ a sampled $\boldsymbol{\theta}_n$ from $N(\mathbf{m}_n, \mathbf{C}_n)$ is obtained to start the backwards sampling step. Note however, that the general backwards sampling algorithm described above degenerates in models with common components in consecutive state vectors as it is the case for autoregressive components. For instance, the conditional distributions $\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t+1}, \mathbf{D}_t$ are singular due to the fact that some of the elements in $\boldsymbol{\theta}_t$ and $\boldsymbol{\theta}_{t+1}$ are

common,

$$\begin{array}{rcccccccc} \boldsymbol{\theta}_t' & = & (\mu_t, & x_t, & x_{t-1}, & \ldots, & x_{t-p+2}, & x_{t-p+1}) \\ \boldsymbol{\theta}_{t+1}' & = & (\mu_{t+1}, & x_{t+1}, & x_t, & x_{t-1}, & \ldots, & x_{t-p+2}). \end{array}$$

Given $\boldsymbol{\theta}_{t+1}$ replace the common components on $\boldsymbol{\theta}_t$ on entries $2, 3, \ldots, p-1$, accordingly. The backwards sampling step at time t is therefore simplified to sample from bivariate distribution $p(\mu_t, x_{t-p+1}|\mu_{t+1}, x_{t+1}, \mathbf{x}_t, \mathbf{D}_t)$ where $\mathbf{x}_t' = (x_t, x_{t-1}, \ldots, x_{t-p+2})$ is the $(p-1)-$dimensional vector of common elements. The desired conditional distribution is calculated by Bayes theorem proportional to

$$p(\mu_t, x_{t-p+1}|\mathbf{x}_t, \mathbf{D}_t)p(\mu_{t+1}, x_{t+1}|\mu_t, \mathbf{x}_t, x_{t-p+1}, \mathbf{D}_t). \tag{5.4}$$

The first element, "the prior", can be calculated from the multivariate distributions $\boldsymbol{\theta}_t|\mathbf{D}_t \sim N(\mathbf{m}_t, \mathbf{C}_t)$. This could be done sequentially conditioning on the elements x_{t-i} for $i = 0, \ldots p-2$, reducing the dimension of the normal distribution by one at each stage. The second element, "the likelihood", is computed directly from the evolution equations (5.2) and (5.3), namely

$$\begin{aligned} \mu_{t+1} &\sim N(\mu_t, W/\tau_{t+1}) \quad \text{and} \\ x_{t+1} &\sim N\left(\sum_{j=1}^{p} \phi_{j,t+1} x_{t+1-j}, \sigma_{t+1}^2/\lambda_{t+1}\right). \end{aligned}$$

In other words, the desired likelihood function $p(\mu_{t+1}, x_{t+1}|\mu_t, \mathbf{x}_t, x_{t-p+1}, D_t)$ is simply the product of the normal distributions above coming from the pair of independent observations μ_{t+1} and x_{t+1} conditional on the two parameters (μ_t, x_{t-p+1}). The bivariate distribution (5.4) is now sampled to fill in the two missing elements of $\boldsymbol{\theta}_t$. To complete the backwards sampling step, the process is repeated sequentially for $t = n-1, n-2, \ldots, 1, 0$ to obtain a sample of $\mathbf{Z}_n$. One important thing to note here, is that the posterior distribution for the initial values $\boldsymbol{\theta}_0$ is sampled naturally in the algorithm. Nevertheless, a prior distribution is required for this purpose and it is set

to be normal with specific vague parameters; see West and Harrison (1997, section 15.3.1) for discussion in the constant AR(p) model.

Sampling the TVAR parameters, $\boldsymbol{\phi}_t|\boldsymbol{\phi}_t^-$

The estimation of the autoregressive parameters is performed by assuming a random walk on the evolution as was introduced in section 3.3. Given all the other parameters, the problem of finding posterior samples from the time varying autoregressive parameters can be written as a dynamic linear model with $\boldsymbol{\phi}_t$ as the state vector,

$$\begin{aligned} x_t &= \mathbf{F}_t'\boldsymbol{\phi}_t + \epsilon_t, \\ \boldsymbol{\phi}_t &= \boldsymbol{\phi}_{t-1} + \boldsymbol{\omega}_t, \end{aligned}$$

for $t = 1, 2, \ldots, n$, where $\mathbf{F}_t = (x_{t-1}, \ldots, x_{t-p})'$ is a $p-$dimensional vector, the innovations sequence is assumed $\epsilon_t \sim N(0, \sigma_t^2/\lambda_t)$ and the $p-$dimensional evolution error is normally distributed $\boldsymbol{\omega}_t \sim N(\mathbf{0}, W_t)$. The evolution variance $\mathbf{W}_t$ controls the variability of the state vector $\boldsymbol{\phi}_t$. A traditional approach used in West and Harrison (1997) is to specify $\mathbf{W}_t$ using a single discount factor δ with values in (0,1). Low discount factors are consistent with high variability in the $\boldsymbol{\phi}_t$ sequence whereas high discount factors between 0.9-0.99 lead to smoother estimates and are commonly used in practice.

Following the standard DLM's theory from West and Harrison (1997), and with the specific relationship for the evolution matrix $\mathbf{W}_t = \frac{(1-\delta)}{\delta}\mathbf{C}_{t-1}$ the updating relations in the Appendix A.1 are given by:

$$\begin{array}{lllllll} \mathbf{a}_t & = & \mathbf{m}_{t-1} & & \mathbf{R}_t & = & \mathbf{C}_{t-1}/\delta \\ f_t & = & \mathbf{F}_t'\mathbf{m}_{t-1} & & Q_t & = & \mathbf{F}_t'\mathbf{C}_{t-1}\mathbf{F}_t/\delta + \sigma_t^2/\lambda_t \\ \mathbf{A}_t & = & \mathbf{C}_{t-1}\mathbf{F}_t\mathbf{Q}_t^{-1}/\delta & & e_t & = & x_t - \mathbf{F}_t'\mathbf{m}_{t-1}. \end{array}$$

The Forward Filtering step is then completed by computing the sequence of posterior

distributions $\boldsymbol{\phi}_t|\mathbf{D}_t \sim N(\mathbf{m}_t, \mathbf{C}_t)$ where

$$\begin{aligned} \mathbf{m}_t &= \mathbf{m}_{t-1} + \mathbf{R}_t\mathbf{F}_t Q_t^{-1}(x_t - \mathbf{F}_t'\mathbf{m}_{t-1}) \quad \text{and} \\ \mathbf{C}_t &= \mathbf{R}_t(\mathbf{I} - \mathbf{F}_t Q_t^{-1}\mathbf{F}_t'\mathbf{R}_t). \end{aligned}$$

The Backwards Sampling algorithm described in the Appendix A.1 is performed by sampling the last element $\boldsymbol{\phi}_n|\mathbf{D}_n$ from $N(\mathbf{m}_n, \mathbf{C}_n)$ and then sampling sequentially $\boldsymbol{\phi}_t|\boldsymbol{\phi}_{t+1}, \mathbf{D}_t$ from $N(\mathbf{h}_t, \mathbf{H}_t)$ with $\mathbf{h}_t = (1-\delta)\mathbf{m}_t + \delta\boldsymbol{\phi}_{t+1}$ and $\mathbf{H}_t = (1-\delta)\mathbf{C}_t$ for $t = n-1, \ldots, 1$.

Sampling $V|V^-$

Inferences on the observational noise variance are given in terms of the precision parameter under non-informative Jeffrey's priors $p(V^{-1}) \propto 1/V^{-1}$. Therefore, the full conditional posterior distribution for V^{-1} is also Gamma with shape parameter $n/2$ and scale $\mathbf{e}'\mathbf{e}/2$, namely Gamma$(n/2, \mathbf{e}'\mathbf{e}/2)$, where $\mathbf{e} = (e_1, e_2, \ldots, e_n)'$ and $e_t = x_t - \mathbf{F}'\boldsymbol{\theta}_t, \ \forall t$.

Sampling $W|W^-$

Again under a diffuse Jeffrey's prior for the precision parameter of $p(W^{-1}) \propto 1/W^{-1}$, the full conditional posterior distribution is given by Gamma$(n/2, \mathbf{r}'\Upsilon\mathbf{r}/2)$, where $\Upsilon = \text{diag}(\tau_1, \ldots, \tau_n)$, $\mathbf{r} = (r_1, r_2, \ldots, r_n)'$ and $r_t = \mu_t - \mu_{t-1}, \ \forall t$.

Sampling the variance weights $\tau_t|\tau_t^-$ *and* $\lambda_t|\lambda_t^-$

The weights on the stochastic changes in trend have a common Gamma prior as discussed earlier, Gamma$(m_1/2, m_1/2)$, which yields to Gamma full conditional posteriors $\tau_t|\tau_t^- \sim \text{Gamma}\left((m_1+1)/2, m_1/2 + r_t^2/(2W)\right)$, with r_t as defined above and for each $t = 1, \ldots, n$. Likewise, the weights on the autoregressive innovations sequence have a common Gamma prior Gamma$(m_2/2, m_2/2)$ and therefore Gamma

full conditional posterior distributions $\lambda_t|\lambda_t^- \sim \text{Gamma}((m_2+1)/2, m_2/2+e_t^2/(2\sigma_t^2))$ after computing the residuals $e_t = x_t - \sum_{j=1}^{p} \phi_{jt} x_{t-j}$ for $t = 1, \ldots, n$. One extra point to note here is the fact that the degrees of freedom m_1 and m_2 are set a priori, traditional values are $m_1 = m_2 = 5$. Nevertheless, the uncertainty about these parameters could be included into the model assuming for example a discrete uniform prior distribution over the set $\{1, 2, 3, \ldots m_{max}\}$ and sampling each parameter according to a multinomial distribution with the corresponding posterior probabilities.

Sampling $\sigma_t^2|\sigma_t^{2-}$

There are different ways to model the variance parameter σ_t^2. In the extense literature of variance models for univariate time series, the most commonly used are the autoregressive conditional heteroscedasticity (ARCH) models developed by Engle (1982) and later Bollerslev *et al.* (1994). In such models, the conditional variance is a function of the squares of the previous observations and past variances. A more realistic alternative is to think that the variances follow some latent stochastic process. This is the basic idea behind the Stochastic Volatility Models (SVM hereafter) appearing in the theoretical literature on option pricing mainly to generalize the Black-Scholes formula to allow for stochastic volatility, Hull and White (1987). Bayesian inference on these latter models is an important research area that has been growing in the past five years. Some generalizations along this line will be addressed in next chapters in a multivariate framework together with multivariate discounting techniques.

For the purpose of univariate component dynamic linear model a discounted variance technique will be used as developed in West and Harrison (1997, section 10.8). The basic discounting methods follow foundational developments in Ameen and Harrison (1995) and Harrison and West (1987). The idea of discounting is to model the decay of information about the sequence of precision parameters $\gamma_t = 1/\sigma_t^2$ between

time points. In other words, the goal is to derive γ_t from γ_{t-1} by establishing a "random walk" model that represents the loss of information. The derivation follows as a special case of multivariate results in Uhlig (1994). Assume for instance that at time $t-1$, the posterior distribution for $\gamma_{t-1}|D_{t-1}$ is Gamma$(n_{t-1}/2, d_{t-1}/2)$ in a usual notation and set

$$\gamma_t = \beta_t \gamma_{t-1}/\delta, \tag{5.5}$$

where $1 < \delta < 0$ is the discount factor. Suppose now that the prior distribution for β_t at time t is $\beta_t|D_{t-1} \sim$ Beta$(\delta n_{t-1}/2, (1-\delta)n_{t-1}/2)$ and independent of γ_{t-1} noting that $\mathrm{E}(\beta_t|D_{t-1}) = \delta$. The prior distribution for γ_t at time t is then,

$$\gamma_t|D_{t-1} \sim \text{Gamma}(\delta n_{t-1}/2, \delta d_{t-1}/2).$$

Note that $E(\gamma_t|D_{t-1}) = E(\gamma_{t-1}|D_{t-1})$ but the dispersion has increased through the discounting degrees of freedom parameter $\delta n_{t-1} < n_{t-1}$. The posterior is conjugate after observing x_t, leading to the usual updating equations, $\gamma_t|D_t \sim \text{Gamma}(n_t/2, d_t/2)$, where $n_t = \delta n_{t-1} + 1$ and $d_t = \delta d_{t-1} + S_{t-1}e_t^2/Q_t$ following the notation in West and Harrison (1997). To include this theory in the Gibbs sampler framework, assume that $\gamma_0|\mathbf{D}_0 \sim \text{Gamma}(n_0/2, d_0/2)$ for some n_0 and d_0 and compute the moments of the forward distributions to complete the forward filtering.

The backwards sampling step includes sampling γ_n from Gamma$(n_n/2, d_n/2)$ and then sequentially $\gamma_t|\gamma_{t+1}, \mathbf{D}_t$ for $t = n-1, \ldots, 1$. To do that, note that

$$p(\gamma_t|\gamma_{t+1}, D_t) \propto p(\gamma_t|D_t)p(\gamma_{t+1}|\gamma_t, D_t),$$

where $p(\gamma_{t+1}|\gamma_t, D_t)$ is derived from the Beta evolution 5.5 leading to the following relation

$$\gamma_t = \eta_t + \delta\gamma_{t+1},$$

where $\eta_t \sim \text{Gamma}((1-\delta)n_t/2, d_t/2)$.

This new class of models is now applied in the analysis of eight chemical species obtained from the Greenland Ice Sheet, (Marsh and Ditlevsen, 1996). The class of latent time-varying autoregressions developed above are used to model each of the quite radically ill-behaved series. Further decompositions are then performed to find quasi-cyclical components in the series following the methodology from section 3.3.1.

5.3 Ice Cores Data

During the last glaciation, the North Atlantic region experienced major changes in climate relative to the Holocene period about 10,000 years in the past. The major ion series collected as part of the Greenland Ice Sheet Project two (GISP2) provide a particularly sensitive monitor of these events; see Mayewski and others (1994) and Marsh and Ditlevsen (1996) for more details on geological description of the measurements. The data consist of a multivariate set of time series that reveals a record of variability in the major soluble chemicals of the atmosphere over Greenland. Consequently, this information can be used to interpret climate changes and to identify potential influence of several major climate forcing agents. The analysis presented here is focused on the eight chemical species calcium, chloride, potassium, magnesium, sodium, sulphate, ammonium and nitrate originally analyzed in Marsh and Ditlevsen (1996). The values estimate relative estimate of ion species and are timed at equal spacings of almost 200 years covering the period $41 - 0$ Kyrs. BP for a total of 206 data points. Figure 5.1 plots the data in reverse order of time to reflect the clear changes and stability of these chemicals in more recent times. These eight series are believed to monitor terrestrial dusts and marine surface which are the two primary sources for chemical species transported to the Greenland atmosphere and hence the relevance of the analyses. Geological studies suggest the existance of five distinct regions, marked on the picture, related to major changes in climate in

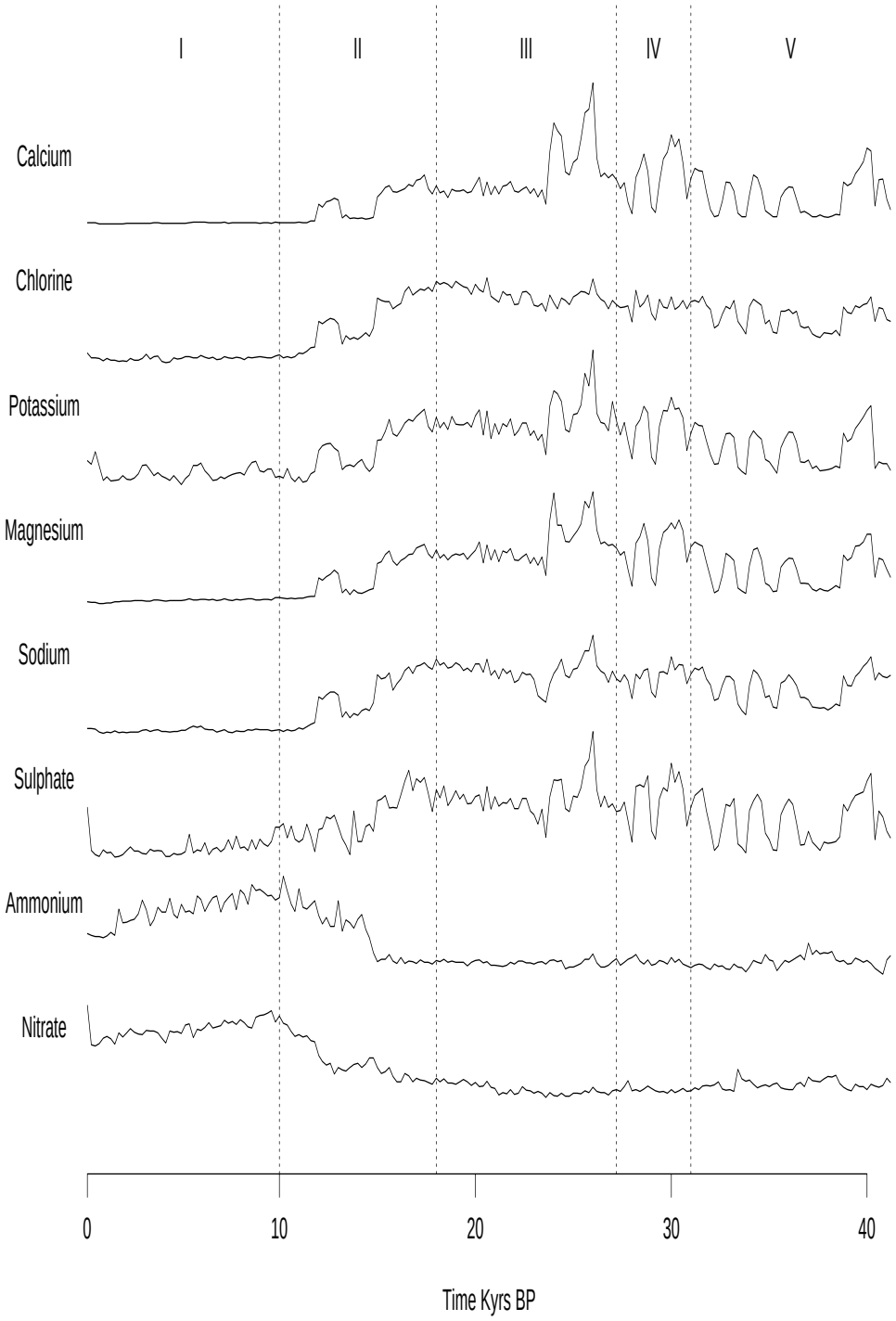

Figure 5.1: Geological chemical species, (x-axis) represents number of years before present

the North Atlantic region over time and differing in the contributed abundance of chemical species. In fact, the Holocene region (I), from 0-10,000 years in the past, observe relative minor changes in climate compared to the dramatic changes in the glacial (II-V) *"stadial"* (cold) and *"interstadial"* (mild) regions. The main interest here is in the analysis of the common features among the multivariate ion species and the contrasting divergence from that commonality by individual ion records. For instance, the first six series reflect similar patterns in their time-varying periodicities as can be seen in Figure 5.1. Conversely, the last two series, ammonium and nitrate, represent less than 8% of the total soluble ionic loading of the atmosphere during the pre-Holocene regions II-V and observe minimal variation in those portions of the series. The basic idea of the analysis is to find the univariate latent components driving the ion series and to explore common features in such processes that may help understanding the reasons for the changes in climate during this period. In addition, the model should consider the inherent measurement, sampling and laboratory errors that are naturally present in these kind of data. For these purposes the class latent autoregressions with heavy-tailed innovations discussed above are used to model each of the quite radically ill-behaved series.

5.3.1 Analysis and Decomposition

The models used for the inferential process follow the methodology described in section 5.1 analyzing each one of the univariate series separately. Explicitly, for each chemical series a model with a first order polynomial trend (5.2) superimposed on a latent TVAR(10) process x_t as in (5.3) and with additive measurement error ν_t was fitted, ignoring the chemical index $i = 1, \ldots, 8$ for simplicity. In addition, discount factors of $\delta = 0.99$ and $\beta = 0.95$ were set to model time variation in the AR parameters, ϕ_t and σ_t^2 respectively. In order to capture the big changes in

trend, prior distributions on the innovations variance weights were established as $\tau_t \sim \text{Gamma}(m_1/2, m_1/2)$ and $\lambda_t \sim \text{Gamma}(m_2/2, m_2/2)$ for the trend and time series components respectively with $m_1 = 1$ and $m_2 = 30$. As a result of this prior assumptions, $\lambda_t \approx 1$ and hence the marginal distribution on innovations of the TVAR(10) process $\epsilon_t|\sigma_t^2$ will be approximately normally distributed. On the other hand, the marginal distributions on the innovations of the first order polynomial trend $\omega_t|W$ will follow Cauchy distributions, which seems adequate to model the obvious and dramatic changes in the level of the ion series. The Gibbs sampler algorithm described above was used to draw 2,000 samples from the desire joint posterior distribution after discarding the first 5,000 iterations as "burn-in" period.

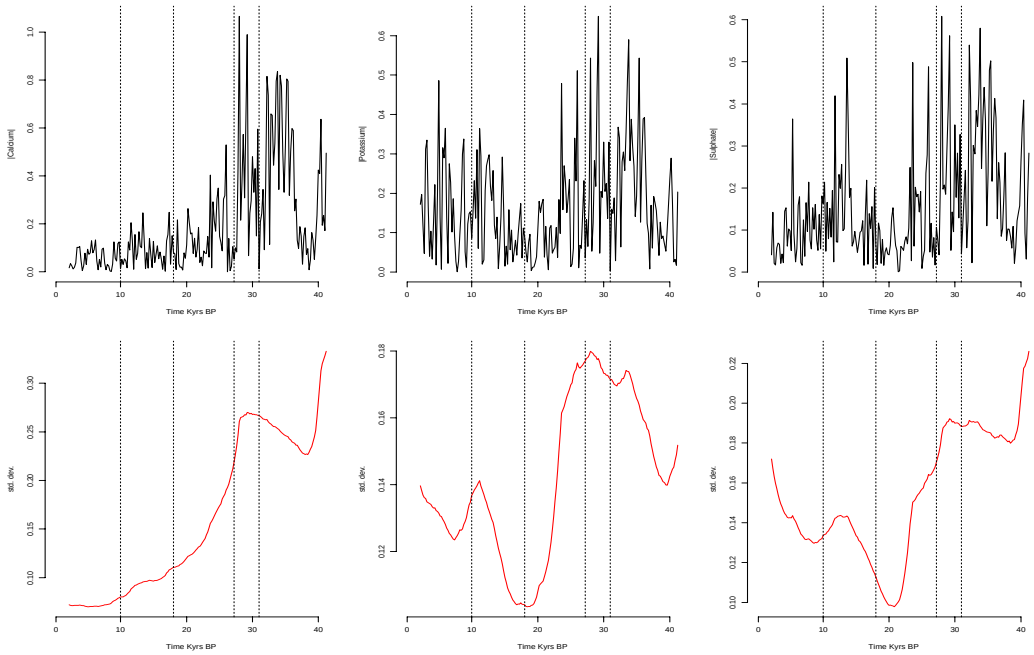

Figure 5.2: Absolute values $|x_t|$ and standard deviations σ_t of estimated TVAR(10) components of the calcium, potassium and sulphate ion series.

Figure 5.2 displays the posterior means of the time-varying innovations standard deviations σ_t on the second row, together with their corresponding posterior means of the latent time series components in absolute value $|x_t|$, in the first row. This

is illustrated for three ion series Calcium, Potassium and Sulphate which present relatively different patterns of variability in the five regions. For instance, the variability of these three chemicals is obviously bigger in the glaciation region than in the Holocene period, specially for Calcium which is understandable due to the fact that this chemical is usually associated with terrestrial sources Mayewski and others (1994).

For each series, the posterior means of the time-varying autoregressive processes, $\mathrm{E}(\phi_t|\mathbf{y})$, were estimated and the reciprocal roots α_{jt} computed by solving the AR polynomial at each time point. Consequently, an instantaneous decomposition of the estimated latent process was obtained as explained in section 3.3.1. In each case, the number of complex and real roots was consistent over time, generally indicating four and sometimes five pairs of complex eigenvalues. These complex roots lead to approximate TVARMA(2,1) processes z_{jt} in (3.6), with time-varying amplitudes, frequencies and moduli. Figure 5.3 displays the corresponding decompositions of the six more similar chemical series. Each one of the six frames displays, from top to bottom, the original data series y_t, the estimated posterior mean trajectory of the trend μ_t, the posterior mean of latent time series process x_t and then the quasi-cyclical components z_{jt} for $j = 1, 2, \ldots 4$. The quasi-cyclical components are ordered by decreasing wavelength/period and plotted in the same vertical scale of the data to appreciate individual contributions. The residuals are negligible by comparison, therefore at each time point, each one of the ion series is the sum of the trend plus four oscillatory components corresponding to the complex autoregressive roots and labeled (1)-(4).

The first thing to note from the picture is that the estimated trends capture the big changes in the original series, especially the radical events occurred in region II around 12-14 Kyrs. in the past. This effect is a consequence of the adequate inclusion

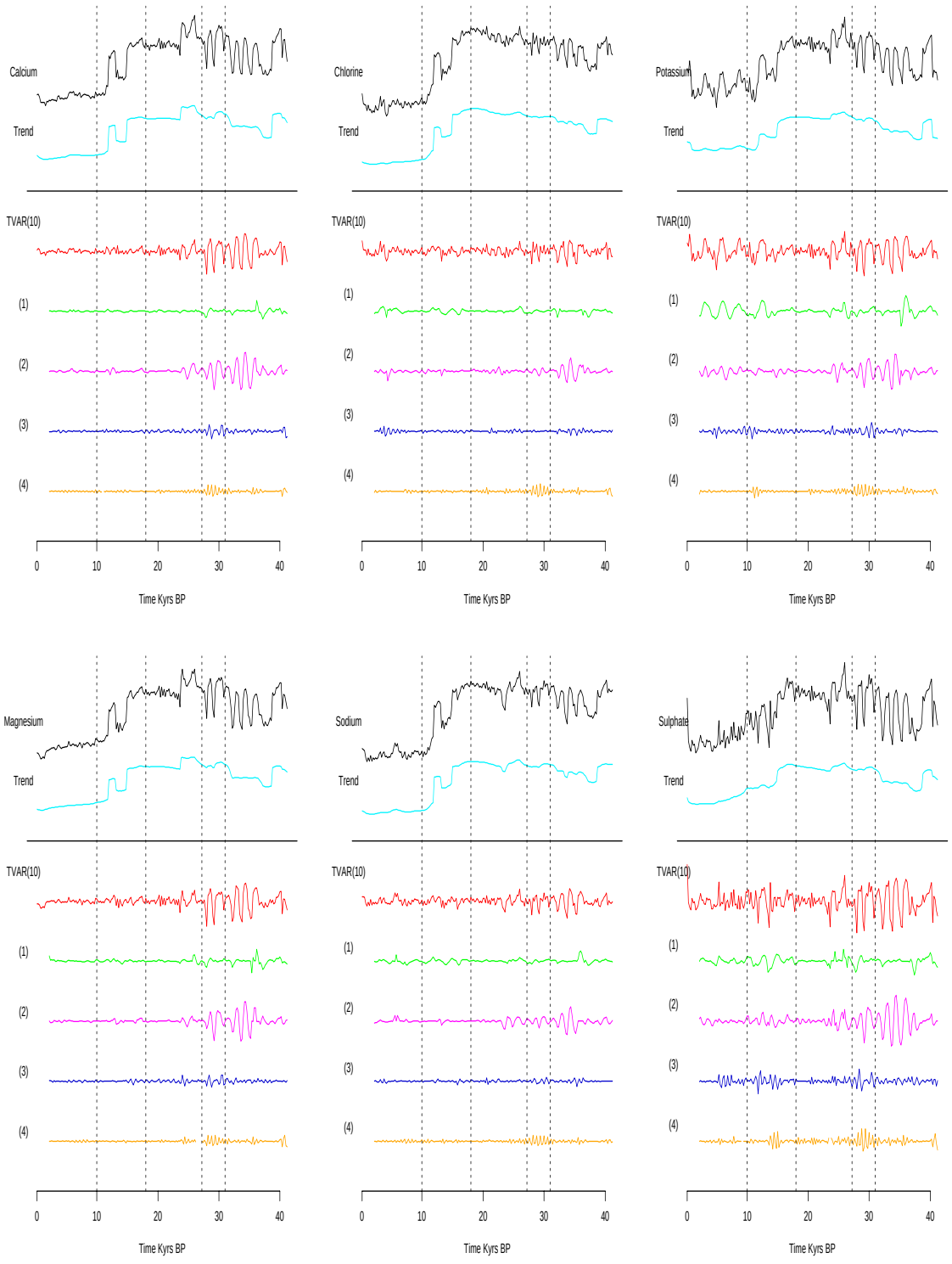

Figure 5.3: Time series decompositions of six ion series.

of heavy-tailed error distributions in the model. The latent quasi-cyclical components resulting of the decomposition of the estimated TVAR(10) processes present common features among these six chemical series and are clearly separated by the defined five regions explained above. In regions IV and V, the four oscillatory components of the six ion series present similar cyclical behavior in the decomposition. This is more evident in the last components in region IV and in the first two components in region V. It seems that these two regions represent the mixing of terrestrial and maritime air masses that are then transported as one to Greenland with a very specific cyclical behavior. The estimated characteristic time-varying frequencies ω_{jt} and moduli of the first two dominant components over the last two regions are plotted in Figure 5.4. As can be seen from this picture, there is a consistent pattern in the frequencies

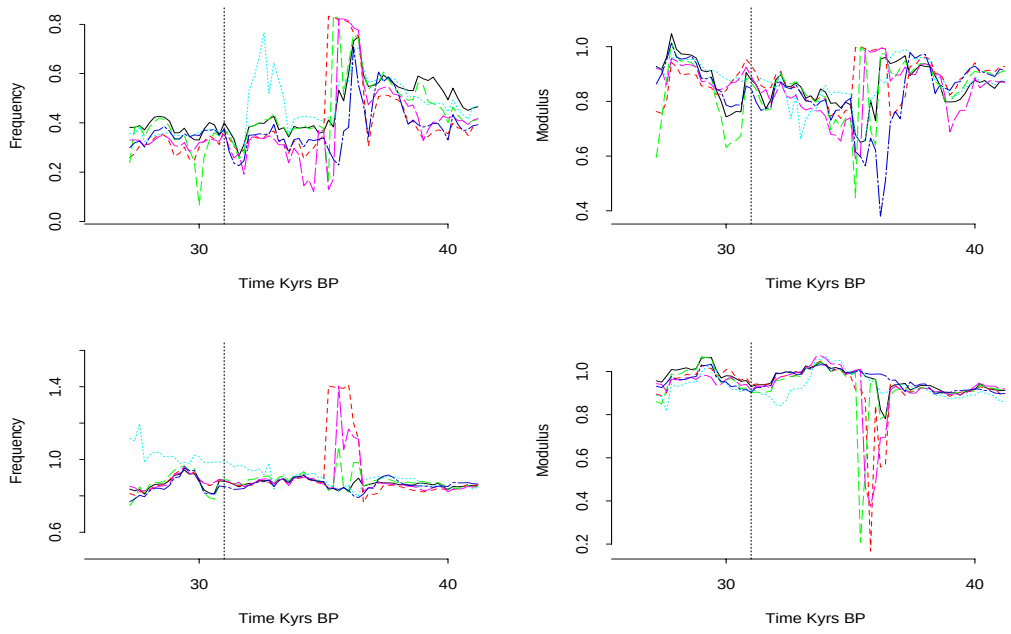

Figure 5.4: Frequencies and modulus of the two dominant components.

and moduli for the six series over time, noting that in region V, around 35 Kyrs., the two main components switch order. These two dominant components represent periods of 1.5-19 Kyrs. and 0.68-1.6 Kyrs. respectively over these two regions. There

is an extra need of further investigation for the adequate interpretation of these numbers in terms of physical events that may have occurred. In regions II and III all the series observe quasi-cyclical components with very small amplitudes and with the sulphate series having higher contributions in region II. This maybe because the system was reorganizing itself before moving into deglaciation and presenting some cyclical behavior in the last part of region III due to the mixing of terrestrial and maritime air masses together with cold temperatures, (Marsh and Ditlevsen, 1996). In the Holocene period, the six series have lower magnitudes due to the reduced transport efficiency from terrestrial and maritime sources.

Finally, the same analysis was performed for the last two components of the series ammonium and nitrate. These two series have similar decompositions in all the regions and are basically explained by the estimated trends as depicted in Figure 5.5. However, it is difficult to obtain relevant interpretations of the quasi-cyclical components due to the low amplitudes. One extra point to note here is the fact that

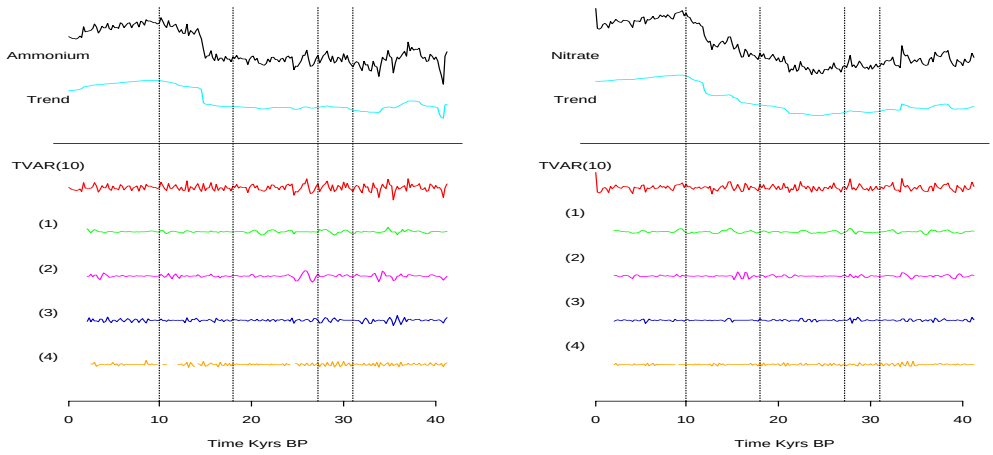

Figure 5.5: Time series decompositions of the two last ion series.

at some time points the components at lower frequencies tend to disappear. In these cases, instead of having a pair of complex roots, two real root components with very low amplitude are obtained and are usually unidentified from the noise terms. These

univariate analyses suggest the use of possible different discount factors δ_i or a change point model to represent the radical changes in the TVAR parameters. This is clear for example in the low amplitudes of the quasi-cyclical components observed in the Holocene region and in the changes in estimated periods from region III to region IV.

5.4 Multivariate Models

It is clear from the univariate analyses above that the set of chemical series have common features and in some cases the quasi-cyclical components have very similar structure. This is not surprising due to fact that these chemical species were not transported to Greenland as individual ion series, but rather as components of chemical compounds carried within air masses having their own chemical characteristic of source and transport histories.

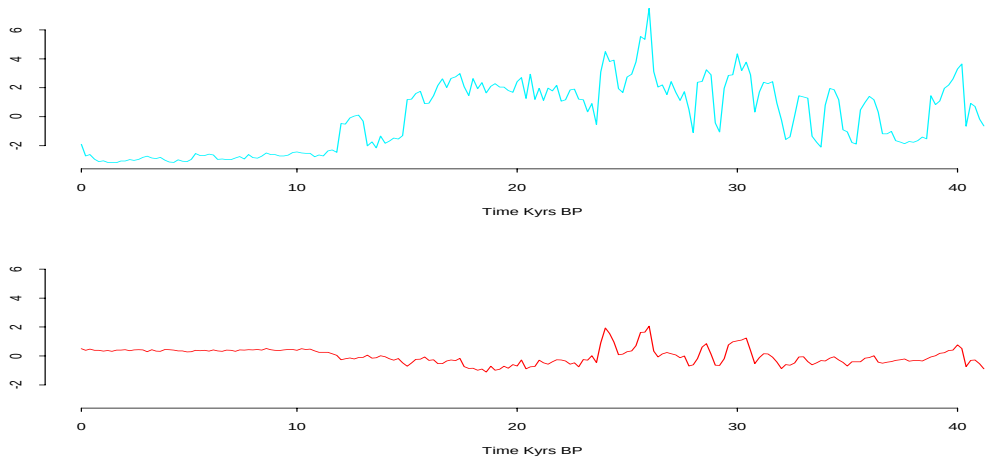

Figure 5.6: Latent orthogonal factors obtained by a simple principal components decomposition.

A common approach to investigate the joint behaviour of the chemical species is to consider a simple principal component analysis of the multivariate series. For example, a traditional principal component decomposition of the first 6 more similar series leads to latent orthogonal processes implied by the eigenstructure of the

covariance matrix, with eigenvectors given by

$$\begin{pmatrix} 0.402 & 0.541 & -0.350 & 0.278 & -0.162 & 0.565 \\ 0.396 & -0.655 & -0.065 & -0.070 & 0.436 & 0.464 \\ 0.414 & 0.183 & 0.050 & -0.872 & -0.171 & -0.059 \\ 0.419 & 0.218 & -0.280 & 0.158 & 0.564 & -0.596 \\ 0.410 & -0.427 & -0.217 & 0.244 & -0.660 & -0.326 \\ 0.407 & 0.123 & 0.863 & 0.271 & -0.006 & -0.007 \end{pmatrix}.$$

In this case, the first principal component displays the dominant multivariate chemical association among the series as can be seen in the top frame of Figure 5.6. In addition, the corresponding first eigenvector assigns almost equal weight to each of the six chemicals explaining about 92% of the total variance, see Figure 5.7. The

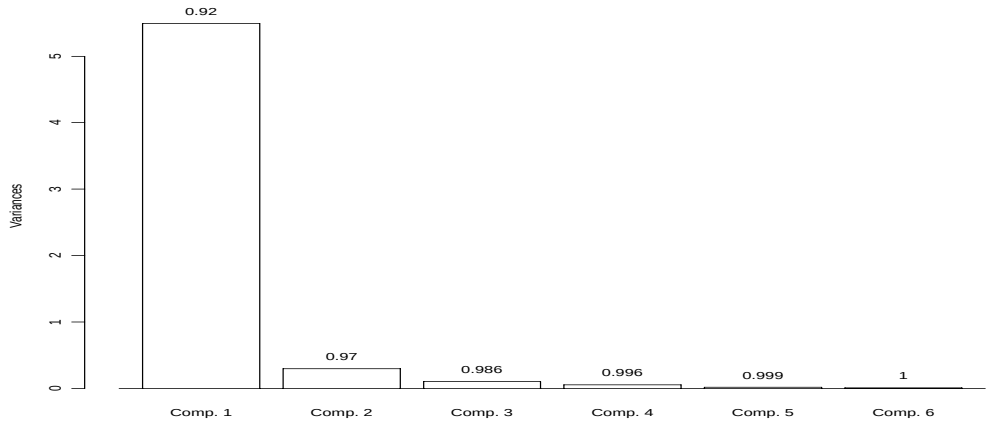

Figure 5.7: Explained variance with the principal components decomposition of six ion series.

interpretation of this component that combines so many chemical species must be that of a large-scale dominant feature of the atmospheric system in which all six species increase or decrease in the same proportions, (Mayewski and others, 1994). The second principal component is presented in the lower frame of Figure 5.6 and together with the first component explains 98% of the total variance. Unlike the first component, the predominant terrestrial calcium and the predominantly marine chloride species are oppositely loaded by the corresponding second eigen-vector con-

trasting these two main sources as clearly depicted in the additive decomposition of Figure 5.8. The principal component analysis used here is based on computing orthogonal components that may not be however appropriate in many applications. In fact, in this example the resulting components in any decomposition may not naturally be orthogonal due to the fact that the load from different source areas is always transported by the atmosphere. Moreover, the simple principal component analysis does not take into consideration the structure of the data overtime and the possible dynamic properties of the covariance matrix and the latent components that was suggested by the individual analyses above.

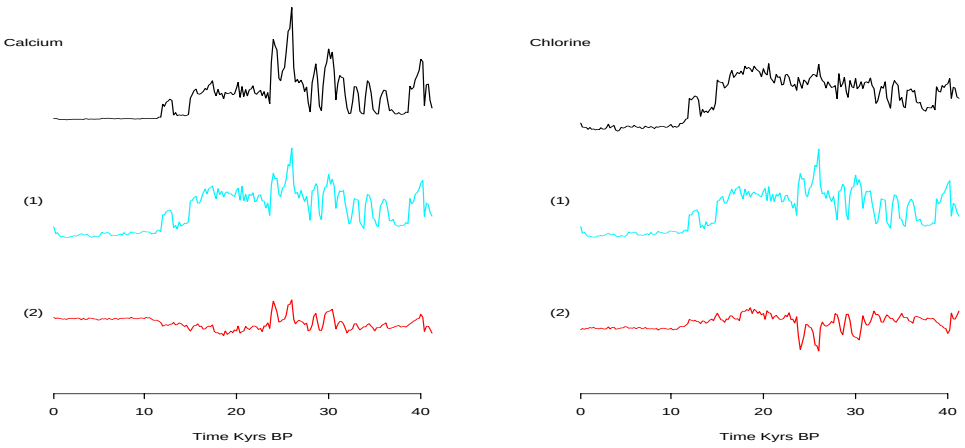

Figure 5.8: Principal components decomposition of the two dominant ion series, calcium and chlorine.

5.4.1 Further Extensions

General multivariate models that are capable of isolate common underlying latent process incorporating the covariance structure between the series and their possible correlation structure over time are obviously needed in this application and many others. For instance, it seems desirable to extend the univariate models from the previous section to consider all the ion series together in a multivariate model. This

first possible generalization is closely related to dynamic factor models and the theory on reduction of dimensionality developed by Pena and Box (1987) and Tiao and Tsay (1989) among others. This basic idea of developing dynamic factor models for the analysis of multivariate non-stationary time series has been widely studied in the econometrics and finance literature on co-integration and common component models as discussed in Escribano and Pena (1994).

The basic setup considers q parallel time series y_{it} that are driven by $k < q$ latent factor series f_{jt} in a time-varying dynamic linear model framework. The general model in a vector form is usually written as

$$\mathbf{y}_t = \boldsymbol{\theta}_t + \mathbf{X}_t \mathbf{f}_t + \boldsymbol{\epsilon}_t, \tag{5.6}$$

over $t = 1, 2, \ldots n$ where $\boldsymbol{\theta}_t$ is a vector of mean parameters, $\mathbf{X}_t$ are dynamic $q \times k$ regression matrices, $\mathbf{f}_t$ is a $k-$dimensional vector of latent factors and $\boldsymbol{\epsilon}_t$ is a zero-mean observation error. In the context of the ice cores data example, $\mathbf{f}_t$ can be thought of the common latent quasi-cyclical components that are present in the eight ion series and that are modeled by either time-varying VAR or VARMA models. In those cases direct extensions from the foundational results of Pena and Box (1987) suggest that the $\mathbf{y}_t$ vector itself follows a time-varying VARMA model. This is a general class of time-varying models that are relevant in many scientific areas in connection with generalizations of univariate decompositions of time-varying autoregressive processes. However, natural issues of parametrization and identification arise when dealing with factor models leading to questions of appropriate structure of the $\mathbf{f}_t$ processes and strict parametric constraints on the $\mathbf{X}_t$ matrices.

The first step in developing general multivariate models for analysis of latent factors/processes is to understand the theory behind the basic factor model in a complete Bayesian setup. This is one of the key points addressed in the next section, together with potential developments and applications in other scientific areas.

Chapter 6

Bayesian Factor Models

The general methodology developed in the previous chapter allows for the study of latent components of univariate series. It was clear from the ice-cores example that some of the estimated components have similar features and share common structure. In general, in many applications involving multiple time series, the latent structure is usually driven by a few sources. This is the basic idea behind dimension reduction techniques such as Principal Components Analysis and Factor Analysis. Furthermore, these techniques provide parsimonious descriptions and inferences on covariance and correlation matrices. A full Bayesian approach is developed in this chapter to make inferences on latent processes driving a set of multiple time series using factor analysis.

Like principal components, factor analysis is a mathematical model which attempts to explain the correlation between a larger set of variables in terms of a small number of unobservable or latent random variables called *factors*. These factors contain the information on the common features among the original variables and the complex relationships between them. In addition, the factors link together seemingly unrelated variables and consequently provide insight into the underlying structure of the data. In other words, factor analysis is primarily concerned with explaining the

covariance among the variables by identifying the sources of variation. Specifically, the factor model assumes that all the correlations are explained by the common factors and the residual variation comes from uncorrelated variable-specific sources. On the other hand, principal components analysis is concerned with finding the set of linear combinations of the original variables that accounts for most of the total variance making no distinction between the sources of variation. Nevertheless, principal components arises as a special case of factor modeling when the specific-sources of variability lead to negligible contributions and the resulting factors are orthogonal, (Press, 1985).

6.1 Historical Notes

Much of the foundational development of factor analysis was done in the early 1940s by psychologists seeking a better understanding of human intelligence in cross sectional studies. In finance and econometrics dynamic factor models have been developed and used widely in the area of asset pricing as an alternative to the Capital Asset Pricing Model (CAPM) since the early 1960s. The Asset Pricing Model developed by Sharpe (1964), Lintner (1965) and Ross (1976,1977) derived the Arbitrage Pricing Theory (APT) which characterizes the expected return on a security as an approximate linear function of the risk premiums on systematic factors in the economy; see Connor and Korajczyk (1996) for a review. There is an extensive literature in factor models based on maximum likelihood estimation and most of classical procedures to estimate the factor model are related to the Autoregressive Conditionally Heteroscedasticy (ARCH) models. Engle *et al.* (1990) suggest the Factor-ARCH model as a parsimonious structure for the conditional variance matrix of the asset excess returns. They apply one and two Factor-ARCH models to pricing Treasury bills. Some other works that use this approach in asset pricing include Bollerslev (1986,1987),

Bollerslev *et al.* (1988), Chou (1988), Diebold and Nerlove (1988), Domowitz and Hakkio (1985), Engle and Bollerslev (1986), Engle *et al.* (1987), French *et al.* (1986), McCurdy and Morgan (1988) and Milhoj (1987). These papers use univariate time series models to represent asset returns. However, Ross's (APT) draws its theoretical sharpness from the assumption that the number of assets approaches to infinity. Burmeister and McElroy (1991) tested linear restrictions of the APT in the factor model and estimated the parameters under nonlinear restrictions using a likelihood ratio test. A two-pass procedure has been proposed by Chen (1983), Connor and Korajczyk (1988), Lehman and Modest (1988), Roll and Ross (1980) among others. They estimate the factor loadings in the first pass and then the factors scores treating the loadings as the true values, so ignoring the uncertainty about parameters.

One key advantage of the Bayesian implementation is that the uncertainty of all the parameters is naturally incorporated and any function of the parameters may be estimated. A recent Bayesian paper on factor analysis, Press and Shigemasu (1989), gives a Bayesian analysis based on informative priors and computes the posterior distributions based on prior information that needs to be assessed. Martin and McDonald (1975) proposed finding joint modal estimates of the factor loading and disturbance covariance matrices with an implicit numerical solution. McCulloch and Rossi (1990) developed a Bayesian analysis of the APT using a two-pass procedure where the factors are extracted using Connor and Korajczyk's asymptotic principal components approach before the full Bayesian analysis is performed. Harvey and Zhou (1990) and Shanken (1987) proposed Bayesian tests for efficiency of a given portfolio. Recently, Geweke and Zhou (1996) analyze the static factor model with traditional assumptions in the context of the APT using a Gibbs sampling approach. They evaluate the posterior density for a proposed measure of the APT pricing errors. This chapter is focused on developing a full Bayesian analysis and implementation of

the factor model, addressing its practical importance in the study of latent structure. The basic model will be further developed to include dynamic properties where more complex features will be incorporated step-by-step.

6.2 The Orthogonal Factor Model

Assume that $\mathbf{y}_t$ is a $q-$dimensional random vector with mean $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_q)'$ and covariance matrix $\boldsymbol{\Sigma}$. The basic k-factor model for observations $t = 1, \ldots, n$ states that the q variables are represented by linear combinations of k common factors with $k \ll q$,

$$\mathbf{y}_t = \boldsymbol{\theta} + \mathbf{X}\mathbf{f}_t + \boldsymbol{\epsilon}_t, \tag{6.1}$$

where $\mathbf{X}$ is a $q \times k$ matrix of unknown constant coefficient parameters called the factor **loadings matrix**. The elements of the $k-$dimensional random vector $\mathbf{f}_t$ are the common factors or **factor scores** and $\boldsymbol{\epsilon}_t$ is a $q-$dimensional random vector of conditionally independent and series-specific quantities or **unique** factors. Traditional assumptions for this model are:

- $\boldsymbol{\epsilon}_t \sim N(\boldsymbol{\epsilon}_t|\mathbf{0}, \boldsymbol{\Psi})$ with $\boldsymbol{\Psi} = \text{diag}(\psi_1, \ldots, \psi_k)$,
- uncorrelated and standardized factors $\mathbf{f}_t \sim N(\mathbf{f}_t|\mathbf{0}, \mathbf{I}_k)$ and
- $\boldsymbol{\epsilon}_t$ and $\mathbf{f}_s$ are mutually independent for all t, s.

From these assumptions, the k-factor model can be expressed in terms of a simple condition on the covariance matrix $\boldsymbol{\Sigma}$,

$$\boldsymbol{\Sigma} = \mathbf{X}\mathbf{X}' + \boldsymbol{\Psi}, \tag{6.2}$$

where the elements on the diagonal of the factor covariance matrix $\mathbf{X}\mathbf{X}'$ are traditionally called **communalities**, $\mathbf{x}_i^2 = \sum_{j=1}^{k} x_{ij}^2$, for $i = 1, \ldots, q$ and the elements of

$\mathbf{\Psi}$ are called **specificities or uniquenesses**. The purpose of the study of the factor model as it is presented here is twofold. First, find the underlying factors that are common to all the series that may help in understanding the nature of the multivariate processes. Second, split the sources of variability into common components across the series and specific contributions of individual variables as in (6.2). These properties are very appealing in real applications; for instance, the basic APT model assumes that the returns on a vector of q assets $\mathbf{y}_t$ follow a k-factor model in line with Geweke and Zhou (1996).

6.2.1 Identification and Constraints

The $k-$factor model as it is presented in (6.1) is highly overparametrized and therefore unidentified by the data. Issues of identifiability and uniqueness of parameter estimates are usually difficult to answer in the context of factor-analytic techniques. As a matter of fact, the decomposition (6.2) is basically indeterminate without further restrictions.

I.- Rank of the loadings matrix

The issue of the rank of the loadings matrix $\mathbf{X}$ is usually not addressed by text books. The fact is that in almost all the cases there is an implicit assumption that the number of factors in the model k is chosen correctly. If this is not the case and $\mathbf{X}$ is not full rank then the model is not fully identified.

Assume for instance that $\text{rank}(\mathbf{X}) = r$ with $r < k$ and let $\mathbf{Q}$ be a $k \times (k-r)$ matrix such that $\mathbf{XQ} = \mathbf{0}$ and $\mathbf{Q}'\mathbf{Q} = \mathbf{I}_{k-r}$. If $\mathbf{M}$ is any $q \times (k-r)$ matrix where $\mathbf{MM}'$ is a diagonal matrix then,

$$\begin{aligned} \mathbf{XX}' + \mathbf{\Psi} &= (\mathbf{XX}' + \mathbf{MM}') + \mathbf{\Psi} - \mathbf{MM}' \\ &= (\mathbf{X} + \mathbf{MQ}')(\mathbf{X} + \mathbf{MQ}')' + \mathbf{\Psi} - \mathbf{MM}'. \end{aligned}$$

This implies that $\boldsymbol{\Sigma} = \hat{\mathbf{X}}\hat{\mathbf{X}}^{'} + \hat{\boldsymbol{\Psi}}$ where $\hat{\mathbf{X}} = \mathbf{X} + \mathbf{M}\mathbf{Q}'$ and $\hat{\boldsymbol{\Psi}} = \boldsymbol{\Psi} - \mathbf{M}\mathbf{M}'$ and hence the model is not identified.

Solution: The existence and uniqueness of the factor model is guaranteed if $\mathbf{X}$ and $\boldsymbol{\Psi}$ are such that $\boldsymbol{\Sigma} - \boldsymbol{\Psi} = \mathbf{X}\mathbf{X}'$ and $\text{rank}(\mathbf{X}\mathbf{X}') = k$.

II.- Orthogonal rotations

If the k-factor model holds, then it also holds if the factors are rotated. If $\mathbf{P}$ is any $k \times k$ orthogonal matrix, then $\mathbf{y}_t$ can also be written as.

$$\mathbf{y}_t = \boldsymbol{\theta} + \mathbf{X}^*\mathbf{f}_t^* + \boldsymbol{\epsilon}_t \tag{6.3}$$

where the rotated factors $\mathbf{f}_t^* = \mathbf{P}^{'}\mathbf{f}_t$ and corresponding factor loadings $\mathbf{X}^* = \mathbf{X}\mathbf{P}$ are valid for the k-factor model without affecting the distribution of $\mathbf{y}_t$. Moreover, the first two moments $\text{E}(\mathbf{f}_t^*) = \mathbf{0}$ and $\text{Var}(\mathbf{f}_t^*) = \mathbf{P}^{'}\mathbf{P} = \mathbf{I}_k$, which implies that $\boldsymbol{\Sigma} = \mathbf{X}^*\mathbf{X}^{*'} + \boldsymbol{\Psi}$. An infinite number of solutions are possible related through orthogonal transformations and the factor model is unidentified unless some restrictions are imposed. The problem concerns in general the invariance of the likelihood function under invertible linear transformations of the factor vectors. There are several approaches to constrain the model for identification, each raising its own questions of interpretation of the resulting factor structure, (Press, 1985; Press and Shigemasu, 1989).

Possible Solutions:

1. A traditional solution based on "hierarchical" constraints on the loadings matrix introduces flexibility into the model. From I above, the loadings matrix is assumed to have rank k, so one can assume without loss of generality that the first k rows of $\mathbf{X}$ are independent. Write the loadings matrix as $\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix}$

where $\mathbf{X}_1$ is the $k \times k$ matrix composed by the first k rows of $\mathbf{X}$ and $\mathbf{X}_2$ is the $(q-k) \times k$ matrix of the last rows. Since $\mathbf{X}_1$ is a nonsingular matrix there exists a unique orthogonal matrix $\mathbf{P}$ such that $\mathbf{X}_1\mathbf{P}^{'}$ is a lower triangular matrix with positive diagonal elements. Explicitly, define a symmetric and positive definite matrix $\mathbf{A} = \mathbf{X}_1\mathbf{X}_1^{'}$ and use the so called *LDU* decomposition. That is, $\mathbf{A} = \mathbf{LDU}$ where $\mathbf{L}$ is a lower triangular matrix with ones on the diagonal, $\mathbf{D}$ is a diagonal matrix with positive numbers and $\mathbf{U} = \mathbf{L}^{'}$ because $\mathbf{A}$ is symmetric. If $\mathbf{L}_1 = \mathbf{L}\mathbf{D}^{1/2}$ then $\mathbf{L}_1$ is a unique lower triangular matrix with positive diagonal elements corresponding to the Cholesky decomposition $\mathbf{A} = \mathbf{L}_1\mathbf{L}_1^{'}$. Therefore, $\mathbf{P} = \mathbf{L}_1^{-1}\mathbf{X}_1$ is a unique orthogonal matrix. Consequently, to guarantee identifiability of the factor model we assume that $\mathbf{X}$ is of the form,

$$\mathbf{X} = \begin{pmatrix} x_{11} & 0 & 0 & \cdots & 0 \\ x_{21} & x_{22} & 0 & \cdots & 0 \\ x_{31} & x_{32} & x_{33} & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ x_{k1} & x_{k2} & x_{k3} & \cdots & x_{kk} \\ x_{k+1,1} & x_{k+1,2} & x_{k+1,3} & \cdots & x_{k+1,k} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ x_{q1} & x_{q2} & x_{q3} & \cdots & x_{qk} \end{pmatrix}. \tag{6.4}$$

where $x_{i,i} > 0$ for $i = 1, \ldots k$ and $x_{i,j} = 0$ for $i < j, \quad i, j = 1, \ldots k$. This condition imposes $\frac{1}{2}k(k-1)$ constraints and uniquely identifies the loadings and associated factors. This solution is used by Geweke and Zhou (1996) and by construction gives a lot of weight to the first k series in determining the factors. In other words, the chosen order of the univariate time series in the $\mathbf{y}_t$ vector is viewed as defining the factors. The first series is a linear regression with the first factor, the second series is a linear regression with the two first factors and so forth. This focuses attention on the choice of ordering in model specification, and provides interpretation.

2. A different approach to solve this problem is by rotating the factor loadings to satisfy an arbitrary constraint such as

$$\mathbf{X}'\mathbf{D}^{-1}\mathbf{X} \text{ is diagonal} \tag{6.5}$$

where $\mathbf{D}$ is a diagonal matrix, it could be the identity or even $\mathbf{\Psi}$. In any case, the diagonal elements are written in decreasing order. This approach assumes that the columns of $\mathbf{X}$ are orthogonal with respect to a weighting function. This constraint is scale invariant and, except for possible changes of the sign of the columns, $\mathbf{X}$ is completely determined and a particular solution is always assured implying also $\frac{1}{2}k(k-1)$ constraints. This solution is more restrictive than the previous one from the interpretation point of view. The columns of the loadings matrix are forced to be orthogonal.[1] Because of that, the focus will be on constraints on the loadings matrix as in (6.4) and variations of it. In any case, the identification problem permits the examination of a variety of solutions for the purpose of selecting the most useful. For instance, it is always possible to rotate the factors after the estimation to obtain "reasonable" interpretations.

3. A different approach not pursued here is that taken by Press and Shigemasu (1989). They basically simplify the structure of the model. In their procedure, zeros appears in specific frequencies, in each of the rows and columns of the factor loading matrix $\mathbf{X}$. They propose a Bayesian solution based on informative priors on the frequencies of the known elements of the factor loadings matrix. That is, suppose that by some assessment scheme, s elements of $\mathbf{X}$ can be preassigned so that the identification problem is eliminated. Find s by solving $d = \frac{1}{2}q(q+1) - (qk+q) + s \geq 0$ and elicit the prior information for

[1] The principal components solution to the orthogonal factor model assumes that the observation errors are small enough, $\mathbf{\Psi} \to \mathbf{0}$ to be ignored, then $\mathbf{\Sigma} = \mathbf{X}\mathbf{X}'$. Write $\mathbf{X} = , \mathbf{D}^{1/2}$ where , is a $q \times k$ matrix whose columns are the normalized eigen-vectors corresponding to the k largest eigen-values and $\mathbf{D}$ is a diagonal matrix with the corresponding eigen-values.

such elements of $\mathbf{X}$. Let $\mathbf{x}_0$ be the $s \times 1$ vector of these elements and $\mathbf{x}_1$ the remaining $(qk - s) \times 1$ elements of $\mathbf{X}$. Assume that $\mathbf{x}_0 \sim N(0, \mathbf{D})$ and a diffuse prior for $\mathbf{\Psi}$ and $\mathbf{x}_1$. Finally, find the posterior distribution of the parameters to make inferences.

III.- Parsimony

Another identification problem in equation (6.2) concerns the number of parameters in the factor model. That is, there are $\frac{1}{2}q(q+1)$ distinct elements of $\mathbf{\Sigma}$, whereas the free parameters in the factor model are $qk + q$ from $\mathbf{X}$ and $\mathbf{\Psi}$ respectively minus $\frac{1}{2}k(k-1)$ from either of the conditions (6.4) or (6.5) in II above. So, in order to have a unique solution the difference d between the number of equations and the number of unknowns must be positive.

- If $d < 0$ then there are more parameters than equations and it is expected to find an infinity of exact solutions for $\mathbf{X}$ and $\mathbf{\Psi}$.
- If $d = 0$ then it is generally possible to find a solution. However, the model will have as many parameters as equations and hence the factor model (6.1) offers no simplification of the original assumptions and no gain in parsimony is obtained.
- If $d > 0$ then there will be more equations than parameters. In this case the factor model offers a simpler explanation of the behavior of $\mathbf{y}_t$ than the full covariance matrix.

Solution: Using either of the conditions (6.4) or (6.5) in II above, the conditions for identifiability are to select the number of factors such that $d \geq 0$ where

$$d = \frac{1}{2}q(q+1) - \left(qk + q - \frac{1}{2}k(k-1)\right).$$

q	k	q	k	q	k	q	k
1	0	8	4	15	10	22	15
2	0	9	5	16	10	23	16
3	1	10	6	17	11	24	17
4	1	11	6	18	12	25	18
5	2	12	7	19	13	26	19
6	3	13	8	20	14	27	20
7	3	14	9	21	15	28	21

Table 6.1: Maximum number of factors k for different number of series q.

Note that by solving this quadratic equation for k, an upper bound on the number of factors is obtained. Table 6.1 has the relationship between the number of series and the number of factors allowed. For realistic values of q this bound is unlikely to be problematic, as practical interest will be in models with smaller numbers of factors.

6.3 Bayesian Analysis

Classical methods for the k-factor model (6.1) are based on maximum likelihood estimators under the corresponding constraints to obtain unique solutions. The Bayesian framework allows for posterior estimates even when the likelihood function is intractable as it is in this case. The conditional distribution of observation t as described in (6.1) is given by

$$\mathbf{y}_t|\boldsymbol{\theta}, \mathbf{X}, \mathbf{f}_t, \boldsymbol{\Psi} \sim N(\boldsymbol{\theta} + \mathbf{X}\mathbf{f}_t, \boldsymbol{\Psi}). \tag{6.6}$$

Note that the distribution of the observations unconditional on the factors is easily calculated,

$$\mathbf{y}_t|\boldsymbol{\theta}, \mathbf{X}, \boldsymbol{\Psi} \sim N(\boldsymbol{\theta}, \mathbf{X}\mathbf{X}' + \boldsymbol{\Psi}). \tag{6.7}$$

To perform a full Bayesian analysis, the joint posterior distribution of all the unknown parameters, $\Theta = \{\boldsymbol{\theta}, \boldsymbol{\Psi}, \mathbf{X}, f_t; \forall t\}$ should be calculated by updating the prior distribution by Bayes' rule.

6.3.1 Prior Distributions

A convenient joint prior distribution for all the parameters is given by,

$$p(\Theta) = p(\boldsymbol{\theta}, \boldsymbol{\Psi}, \mathbf{X}, \mathbf{F}) = p(\boldsymbol{\theta})p(\mathbf{F})p(\mathbf{X}, \boldsymbol{\Psi}), \tag{6.8}$$

where $\mathbf{F} \sim N(\mathbf{0}, \mathbf{I}_n, \mathbf{I}_k)$[2] is the $n \times k$ factors matrix $\mathbf{F} = (\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_n)'$. Improper constant priors are used for $\boldsymbol{\theta}$ and the $qk - k(k-1)/2$ non-zero entries of the loadings matrix $\mathbf{X}$, namely $p(x_{i,j}) \propto c$. To complete the prior, independent Gamma priors are used for the precisions of the disturbance errors defined as $\tau_i = 1/\psi_i$, for $i = 1, \ldots, q$. Consequently, the joint prior is given by,

$$p(\Theta) \propto p(\mathbf{F}) \prod_{i=1}^{q} \tau_i^{\alpha_0 - 1} \exp(-\beta_0 \tau_i / 2). \tag{6.9}$$

Note that the non-informative Jeffrey's priors are easily obtained by setting $\alpha_0 = 0$ and $\beta_0 = 0$. However, Jeffrey's priors for the idyosincratic variances cannot be used here because the likelihood function presents a singularity at zero and hence yielding to improper posteriors. Therefore, proper priors for these parameters are necessary. More discussion and other alternatives for priors to follow.

Hierarchical Priors

In general, non-informative priors will be used for calculating posterior distributions throughout the thesis. Nevertheless, it is worth mentioning alternatives in special cases. As mentioned above, one traditional problem in orthogonal factor models is the so called *Heywood case* when the specific variances are very small $\psi_i \to 0$.[3] In such cases, Jeffreys' priors cannot be used and more structured proper priors are needed, like uniforms or inverse-Gammas.

[2] Standard notation for matrix Normal distributions, Dawid (1981); see Appendix A.2 for details.

[3] See Martin and McDonald (1975) for details from the classical point of view.

One other concern with non-informative priors is the fact that the loadings matrix parameters and the specific variances are not independent by construction. From (6.2) it is clear that $\mathbf{X}$ and $\mathbf{\Psi}$ are correlated and hierarchical priors may be appropriate. Press and Shigemasu (1989) proposed informative priors such as:

- $\mathbf{X}|\mathbf{\Psi} \sim N(\mathbf{X}_0, \mathbf{\Psi}, n_0\mathbf{I}_k)$. In this case n_0 and $\mathbf{X}_0$ should be assessed *a priori*. For example, the identifiability constraints can be imposed on $\mathbf{X}_0$ by setting some of its values to zero or forcing its columns to be orthogonal.

- The prior for the disturbance variance is an Inverse-Wishart distribution, namely $\mathbf{\Psi}^{-1} \sim \text{Wishart}(\nu_0, \mathbf{B}_0^{-1})$ where $\mathbf{B}_0 = \text{diag}(b_1, \ldots, b_q)$. Note that $\mathbf{\Psi}$ is not assumed to be diagonal but it is diagonal in expectation. The hyperparameters ν_0 and $\mathbf{B}_0$ have to be assessed.

6.3.2 Implementation of the Gibbs Sampler

A customized Markov Chain Monte Carlo is performed to sample the posterior distribution of all the parameters. As in previous chapters, the MCMC algorithm specifies an irreducible and aperiodic Markov Chain with stationary distribution given by the desired joint posterior distribution. An implementation of the posterior sampling algorithm is outlined here based on iterative updating using the full conditional densities of each unknown parameter. Using the same notation as before, any subsets of the unknown parameters Θ will be denoted ξ and ξ^- will represent the remaining variables combined with the full data set.

Sampling the factors $F|F^-$

From the unconditional distribution (6.7) and the prior assumption $\mathbf{f}_t \sim N(\mathbf{0}, I_k)$ in model (6.1), $\mathbf{y}_t$ and $\mathbf{f}_t$ are jointly normal,

$$\begin{pmatrix} \mathbf{f}_t \\ \mathbf{y}_t \end{pmatrix} \sim N\left[\begin{pmatrix} \mathbf{0} \\ \boldsymbol{\theta} \end{pmatrix}, \begin{pmatrix} \mathbf{I}_k & \mathbf{X}' \\ \mathbf{X} & \mathbf{Q} \end{pmatrix}\right],$$

where $\mathbf{Q} = \mathbf{X}\mathbf{X}' + \boldsymbol{\Psi}$. From properties of the normal distribution the conditional distribution of $\mathbf{f}_t$ is given by $\mathbf{f}_t|\mathbf{y}_t \sim N(\mathbf{A}(\mathbf{y}_t - \boldsymbol{\theta}), \mathbf{I}_k - \mathbf{A}\mathbf{Q}\mathbf{A}')$, where $\mathbf{A} = \mathbf{X}'\mathbf{Q}^{-1}$. This distribution can be written in matrix form allowing to sample all the factors at once and not one by one. After some algebra, the full conditional distribution for the factors,

$$\mathbf{F}|\mathbf{F}^- \sim N(\mathbf{Y}^*\boldsymbol{\Psi}^{-1}\mathbf{X}(I_k + \mathbf{X}'\boldsymbol{\Psi}^{-1}\mathbf{X})^{-1}), \mathbf{I}_n, (I_k + \mathbf{X}'\boldsymbol{\Psi}^{-1}\mathbf{X})^{-1}),$$

where $\mathbf{F} = (\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_n)'$ is the $n \times k$ factors matrix and $\mathbf{Y}^*$ is a $n \times q$ matrix with rows $\mathbf{y}_t - \boldsymbol{\theta}$ for $t = 1, \ldots, n$. The mean of this matrix normal distribution is usually called the Thompson's factor score. Note that by using the identification constraints on the loadings matrix as described in section 6.2.1 II, where $\mathbf{X}'\boldsymbol{\Psi}^{-1}\mathbf{X}$ is restricted to be diagonal, the latter matrix normal posterior distribution will have diagonal covariance matrices. See the appendix A.2 for details of how to sample matrix normal distributions.

Sampling the idiosyncratic variances $\Psi|\Psi^-$

Assuming independent Inverse-Gamma priors (6.9), the full conditional posterior distribution for the variances are conditionally independent Inverse-Gamma posteriors or equivalently for $i = 1, \ldots, q$, for the precisions $\tau_i|\tau_i^- \sim \text{Gamma}((\alpha_0+n)/2, (\beta_0+e_i)/2)$, where $e_i = \sum_{t=1}^{n}(y_{it} - \theta_i - \mathbf{x}_i'\mathbf{f}_t)^2$ and $\mathbf{x}_i$ is the i-th row of the loadings matrix $\mathbf{X}$.

Sampling the conditional mean and the loadings matrix $\boldsymbol{\theta}, \mathbf{X}|(\boldsymbol{\theta}, \mathbf{X})^-$

Under non-informative priors, the mean $\boldsymbol{\theta}$ and the elements of the loadings matrix (6.4) can be sampled in blocks. The idea is to sample independently each row of the loadings matrix with its corresponding mean parameter.

- From model (6.1), with the identification constraints on the loadings matrix (6.4), the first series is modeled as $y_{1t} = \theta_1 + x_{11} f_{1t} + \epsilon_{1t}$ for $t = 1, \ldots n$ and $x_{11} > 0$. The full conditional posterior distribution is a bivariate normal truncated in one dimension, namely,

$$\theta_1, x_{11}|(\theta_1, x_{11})^- \sim N\left((\mathbf{B}_1'\mathbf{B}_1)^{-1}(\mathbf{B}_1'\mathbf{Y}_1), \psi_1(\mathbf{B}_1'\mathbf{B}_1)^{-1}\right)\mathbf{I}_{\{x_{11}>0\}},$$

  where $\mathbf{Y}_1 = (y_{11}, \ldots, y_{1n})'$ is the first series and $\mathbf{B}_1$ is a $N \times 2$ matrix with rows $(1, f_{1t})'$ for $t = 1, \ldots n$. This distribution can be sampled directly using properties of the multivariate normal distribution without using rejection methods such as used in Geweke and Zhou (1996). The main problem of rejection methods is that the rejection probability increases when the values of the loadings are close to zero. This could happen when the first series is not very similar to the others and has weak linear relationship with the first factor. This issue affects convergence of the Markov Chain affecting the estimation of the specific variances too. The approach suggested here is to sample this distribution in two steps. First, sample easily the truncated component x_{11} from its marginal posterior distribution using the inverse cdf method and then sample the corresponding conditional posterior distribution $\theta_1|\theta_1^-$ using the value x_{11} just sampled. The procedure is similar for the first k rows of the loadings matrix.

- In general, to sample the $j-$th row, note that $y_{jt} = \theta_j + \sum_{l=1}^{j} x_{jl} f_{lt} + \epsilon_{jt}$, with $x_{jj} > 0$, for $j \leq k$ and $t = 1, \ldots n$. The full conditional posterior distribution

for the parameters is the truncated multivariate normal,

$$\theta_j, x_{j1}, \ldots x_{jj} | (\theta_j, x_{j1}, \ldots x_{jj})^- \sim N\left((\mathbf{B}_j'\mathbf{B}_j)^{-1}(\mathbf{B}_j'\mathbf{Y}_j), \psi_j(\mathbf{B}_j'\mathbf{B}_j)^{-1}\right)\mathbf{I}_{\{x_{jj}>0\}},$$

where $\mathbf{Y}_j$ is the $n \times 1$ vector of the first j series, $\mathbf{B}_j$ is a $n \times (j+1)$ matrix with rows $(1, f_{1t}, \ldots, f_{jt})'$ for $t = 1, \ldots n$. Again, sample first the truncated component x_{jj} from its univariate marginal normal distribution. Given the sampled value x_{jj} and the rest of the variables, sample $(\theta_j, x_{j1}, \ldots x_{j,j-1})$ from the corresponding conditional distribution.

- The remaining $q-k$ rows of the loadings matrix and mean components are sampled jointly using a matrix normal distribution. Let $\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}2 \end{pmatrix}$, where $\mathbf{X}_2$ represents the last $q-k$ rows of the loadings matrix and let also $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)'$ with $\boldsymbol{\theta}_2$ representing the last $q-k$ components. The full conditional distribution for the $(q-k) \times (k+1)$ matrix $\mathbf{Z} = [\boldsymbol{\theta}_2, \mathbf{X}_2]$ is,

$$\mathbf{Z}|\mathbf{Z}^- \sim N\left((\mathbf{B}'\mathbf{Y}_{(k+1):q})'(\mathbf{B}'\mathbf{B})^{-1}, \boldsymbol{\Psi}_{(k+1):q}, (\mathbf{B}'\mathbf{B})\right)$$

where $\boldsymbol{\Psi}_{(k+1):q} = \text{diag}(\psi_{k+1}, \ldots, \psi_q)$,$\mathbf{Y}_{(k+1):q}$ is the $n \times j$ matrix of series $k+1$ through q and $\mathbf{B}$ is a $n \times (k+1)$ matrix with rows $(1, f_{1t}, \ldots, f_{kt})'$ for $t = 1, \ldots n$.

Note that the inclusion of normal informative priors is straightforward. After implementing this model a sample from the posterior distribution of the parameters is obtained for the orthogonal static factor model (6.1).

6.4 The Oblique Factor Model

It is clear that the purpose of the restrictions imposed to the factor model in section 6.2.1 is to find one solution to the problem. That is, the constraints imposed on the factor loading matrix $\mathbf{X}$ as in (6.4) are a mathematical convenience to have a

unique solution and to make possible the inferential process. However, there will be situations in which the conditions may seem inappropriate and a solution in a rotated coordinate system may be preferable. As was explained in section 6.2.1 II, the loadings matrix $\mathbf{X}$ and hence the factors $\mathbf{f}$ may be rotated without affecting the distribution of the observations as in (6.3). Moreover, the rotation could be performed by transforming the model with any non-singular matrix. For example, assume that the $k-$factor model holds for n observations $\mathbf{y}_t$ as in (6.1) with loadings matrix $\mathbf{X}^*$ as (6.4) and factors $\mathbf{f}_t^* \sim N(\mathbf{f}_t^*|\mathbf{0}, \mathbf{I}_k)$. Let $\mathbf{H} \neq \mathbf{I}_k$ be any $k \times k$ positive definite matrix and decompose it as $\mathbf{H} = \mathbf{L}\mathbf{L}'$[4]. Define new factors as $\mathbf{f}_t = \mathbf{L}\mathbf{f}_t^*$ with corresponding loadings matrix $\mathbf{X} = \mathbf{X}^*\mathbf{L}^{-1}$ to obtain an equivalent model, for $t = 1, \ldots, n$,

$$\mathbf{y}_t = \boldsymbol{\theta} + \mathbf{X}\mathbf{f}_t + \boldsymbol{\epsilon}_t \quad \text{where,} \tag{6.10}$$

$$\mathbf{f}_t \sim N(\mathbf{f}_t|\mathbf{0}, \mathbf{H}) \quad \text{and} \tag{6.11}$$

$$\boldsymbol{\Sigma} = \mathbf{X}\mathbf{H}\mathbf{X}' + \boldsymbol{\Psi} = \mathbf{X}^*\mathbf{X}^{*\prime} + \boldsymbol{\Psi}. \tag{6.12}$$

This new setup is traditionally called **The Oblique Factor Model**. In the covariance matrix decomposition in (6.2) the factors contribute to the covariance matrix $\boldsymbol{\Sigma}$ only through the loadings matrix. On the other hand, the new set of factors may have better interpretations in real problems if they are correlated. Note that the decomposition of $\boldsymbol{\Sigma}$ is unaffected as is clear in (6.12). Furthermore, the common source of variability $\mathbf{X}\mathbf{H}\mathbf{X}'$ is decomposed in contributions from the factors variances $\mathbf{H}$ and contributions from the new loadings matrix.

A traditional approach to fit these models is to estimate the orthogonal factor model and afterwards transform the model by choosing the $\mathbf{H}$ matrix to make the factors as intuitively meaningful as possible. In order to eliminate the ambiguity in a factor analysis solution, some theories suggest to use a criteria to choose $\mathbf{H}$ that

[4] $\mathbf{L}$ could be the Cholesky decomposition of $\mathbf{H}$ or in the singular value decomposition $\mathbf{H} = \mathbf{E}\mathbf{D}\mathbf{E}'$, simply take $\mathbf{L} = \mathbf{E}\mathbf{D}^{1/2}$ with $\mathbf{D}$ the diagonal matrix of eigenvalues and $\mathbf{E}$ the matrix of the corresponding eigen-vectors.

depends on the variation among the communalities that result after an orthogonal transformation of the loadings matrix. A convenient choice of rotation is given by the *varimax method*, which selects an orthogonal matrix $\mathbf{H}$ to provide axes with a few large loadings and as many zero loadings as possible.

6.4.1 Similar Models

The idea of estimating the variances of the factors in the oblique model (6.11) is a step forward in understanding factor models. The ideal model though, will estimate the variances of the factors $\mathbf{H}$ together with the rest of the parameters unlike traditional approaches that rotate the factors and factors loadings after the estimation process. Note that the number of constraints will increase as more parameters are included into the model for identifiability reasons, as explained in section 6.2.1, III. The problem is that by including $\mathbf{H}$ a basic identification problem between $\mathbf{X}$ and $\mathbf{H}$ arises. In spite of that, the basic orthogonal model (6.1) can be reparametrized to include a diagonal variance $\mathbf{H}$ for the factors by changing the structure of the loadings matrix.

Consider an orthogonal factor model (6.1) with loadings matrix $\mathbf{X}^o$ of the form (6.4) with elements x^o_{ij} for $i = 1, \ldots, q$ $j = 1, \ldots, k$. Let $\mathbf{f}^o_t$ be the corresponding factors $\mathbf{f}^o_t \sim N(\mathbf{f}^o_t|\mathbf{0}, \mathbf{I}_k)$ and define a $k \times k$ diagonal matrix $\mathbf{H}$ where,

$$\mathbf{H}^{1/2} = \text{diag}(x^o_{1,1}, x^o_{2,2}, \ldots, x^o_{k,k}).$$

Define a new loadings matrix $\mathbf{X} = \mathbf{X}^o\mathbf{H}^{-1/2}$ with new factors given by $\mathbf{f}_t = \mathbf{H}^{1/2}\mathbf{f}^o_t$. The new factors are still uncorrelated but with unrestricted variances, $\mathbf{f}_t \sim N(\mathbf{f}_t|\mathbf{0}, \mathbf{H})$.

The new loadings matrix $\mathbf{X}$ will have a slightly different shape,

$$\mathbf{X} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ x_{21} & 1 & 0 & \cdots & 0 \\ x_{31} & x_{32} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ x_{k1} & x_{k2} & x_{k3} & \cdots & 1 \\ x_{k+1,1} & x_{k+1,2} & x_{k+1,3} & \cdots & x_{k+1,k} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ x_{q1} & x_{q2} & x_{q3} & \cdots & x_{qk} \end{pmatrix}. \tag{6.13}$$

where $x_{i,i} = 1$ for $i = 1, \ldots k$ and $x_{i,j} = 0$ for $i < j, \quad i, j = 1, \ldots k$. Note that the likelihood function of the observations is unaffected and the model is completely identified with the same number of constraints as in the orthogonal case. An easy way to understand this is to think that the positive elements of the diagonal of the loadings matrix $\mathbf{X}^o$ in the orthogonal model are the standard deviations of the factors in an equivalent model with unrestricted diagonal factors variance.

This setup also suggests an alternative way of sampling the full posterior distribution of the loadings matrix from section 6.3.2[5]. Moreover, this is a useful way to write the model that opens the possibility of introducing dynamic properties through the variances of the factors as it will be addressed in the next chapter. One financial example is the dynamic factor model proposed by Ross (1976) for APT with time-varying factor variances or equivalently time-varying loading matrices in which the covariances (the betas) of different assets with a particular factor change proportionally. The response of each asset's risk premium to the risk (own variance) of a particular factor is constant. Assets' risk premia change over time as the risk of a particular factors changes. In this setting, portfolios with excess returns processes that have constant conditional variances can always be constructed.

[5]This issue will be addressed in the next chapter in a more general framework.

Chapter 7

Dynamic Factor Models and Stochastic Volatility

The purpose of factor models, as described in previous chapter, is to find latent factors that are common across the multivariate time series. In addition, factor models implicitly decompose the covariance matrix of the observations in a convenient and interpretable way splitting the common and specific sources of variability. In this direction, one of the assumption of the models presented before is that the covariance matrix of the observations $\mathbf{\Sigma}$ is constant over time. In many applications though, the main interest is in explicitly describe changes through patterns of time-variation in parameters driving the underlying latent processes. Dynamic factor models are developed in this chapter to investigate possible time-varying latent processes and their implications in modeling changes in covariance matrices over time.

7.1 The Dynamic Factor Model

There are different ways to incorporate dynamic components to the orthogonal factor model (6.1). The basic idea is to think that there is an instantaneous factor decomposition of the covariance matrix of the observations. That is, assume that $\mathbf{y}_t$ is now

a $q-$dimensional random vector with mean $\boldsymbol{\theta}$ and time-varying covariance matrix $\boldsymbol{\Sigma}_t$ for $t = 1, \ldots n$. Now suppose that, at each time point, it is possible to decompose $\boldsymbol{\Sigma}_t$ in a common and specific variability sources as in (6.12) for the constant case. Consequently, there exist factors $\mathbf{f}_t$ and errors $\boldsymbol{\epsilon}_t$ such that the k-factor model holds at each time t. Note that the setup in (6.11) allows for the possibility of incorporating time-varying variances $\mathbf{H}_t$ to the latent factors in a natural way. That is, extend (6.11) by considering time-varying factors variances in the basic $k-$factor dynamic model, with $k < q$, for $t = 1, \ldots, n$,

$$\begin{aligned} \mathbf{y}_t &= \mathbf{X}\mathbf{f}_t + \boldsymbol{\epsilon}_t, \\ \mathbf{f}_t &\sim N(\mathbf{f}_t|\mathbf{0}, \mathbf{H}_t), \end{aligned} \tag{7.1}$$

where $\mathbf{H}_t = \text{diag}(h_{1t}, h_{2t}, \ldots, h_{kt})$ is the matrix of instantaneous factor variances, $\mathbf{X}$ is the loadings matrix of the form (6.13) with columns $\mathbf{x}_j$, the idyosincratic errors are normally distributed $\boldsymbol{\epsilon}_t \sim N(\boldsymbol{\epsilon}_t|\mathbf{0}, \boldsymbol{\Psi})$ and $\boldsymbol{\epsilon}_t$ and $\mathbf{f}_s$ are mutually independent for all t, s. This model enhances the interest in learning all the possible features of the latent factors and considers the factors variances as latent processes themselves. Moreover, this framework is similar to early work on this area by Harvey *et al.* (1994), Jacquier *et al.* (1994, 1995) and Kim *et al.* (1998). The dynamic factor model as presented in (7.1) incorporates the time-varying structure of the conditional variance of the observations $\boldsymbol{\Sigma}_t$ in an instantaneous decomposition,

$$\boldsymbol{\Sigma}_t = \mathbf{X}\mathbf{H}_t\mathbf{X}^{'} + \boldsymbol{\Psi} = \sum_{j=1}^{k} \mathbf{x}_j\mathbf{x}_j' h_{tj} + \boldsymbol{\Psi}_t, \tag{7.2}$$

for all t. Note however, that model (7.1) has implicitly defined a time varying structure in the loadings matrix by letting $\mathbf{X}_t^* = \mathbf{X}\mathbf{L}_t$ and $\mathbf{f}_t^* = \mathbf{L}_t^{-1}\mathbf{f}_t$ where $\mathbf{L}_t\mathbf{L}_t^{'} = \mathbf{H}_t$. Furthermore, the dynamic factor model as presented here is a natural generalization of univariate variance models. Modeling and forecasting changes in volatility

of the factor processes $\mathbf{H}_t$ and hence changes in the full covariance matrix of the observations is of key interest.

7.1.1 Initial Approach via Discounting

The model described in (7.1) assumes that the factors are uncorrelated a priori, namely $\mathbf{f}_t \sim N(\mathbf{f}_t, \mathbf{H}_t)$ with $\mathbf{H}_t = \text{diag}(h_{1t}, h_{2t}, \ldots, h_{kt})$, where for any pair d, l, the sequences h_{dt} and h_{lt} are mutually independent. In this case, the factors can be treated as univariate series and the individual factors variances h_{it} can be modeled independently. The natural idea would be to model the stochastic changes in variance using the univariate discounting technique from West and Harrison (1997, chapter 8) and explained in section 5.2. The Gibbs sampling setup from section 6.3.2 will include a new step to sample each of the individual sequences h_{it} as described in section 5.2. In addition, some changes in the full conditional distributions of the loadings matrix $\mathbf{X}$ in the form (6.13) will be needed.

In particular, if the discount factor is set to one, the factor variances will be constant over time $\mathbf{H}_t = \mathbf{H}$ implying a different way to sample the orthogonal factor model (6.1) from section 6.3.2. That is, instead of sampling the loadings matrix $\mathbf{X}$ using multivariate truncated normals, use model (7.1) with loadings matrix as in (6.13) and sample the factors variances from the corresponding Inverse-Gamma distributions.

The general results of this approach are related to the multivariate versions of discounting from West and Harrison (1997, chapter 16). However, general models that include correlated factors variances with forecasting properties are desirable and will be addressed in the next section.

7.2 Multivariate Stochastic Volatility Models

From the class of stochastic variance models, the so called Stochastic Volatility Model (SVM) has been recently used in practical problems. This model assumes a stationary process for volatilities or a function of them. In the univariate case, classical estimation of SVM is based on approximations due to the difficulties in handling the likelihood function. The most common implementations of such models are based on the Method of Moments and Quasi Maximum Likelihood, (Melino and Turnbull, 1990). Harvey *et al.* (1994) employed Kalman filtering to estimate the parameters by maximizing the quasi-likelihood. This estimator is consistent and asymptotically normal but is sub-optimal in finite samples because the log chi-square distribution that they used is poorly approximated by a normal distribution.

The use of MCMC techniques for the analysis of univariate SVM with conditional normally distributed observations was introduced by Jacquier *et al.* (1994) and Shephard (1993). Their *single move* sampler is based on a simple accept/reject procedure to sample the full conditional distribution of the log volatilities. The main problem with this method is that the log volatilities are highly correlated and therefore the movements in the samples are very small for each component of the full conditional producing slow converging rates. This method requires a huge amount of iterations to generate samples from the joint posterior distribution. In order to break correlations and improve this method, Shephard and Pitt (1997) propose to sample groups of consecutive log volatilities using a customized Metropolis algorithm.

An alternative method to sample the joint posterior distribution is by using the *multi-move* sampler based on state space models described in Kim *et al.* (1998). The idea is to apply a non linear transformation to the data and approximate a log chi-square distribution with a mixture of seven normals; see Kim *et al.* (1998) for details. These non-Gaussian or conditionally Gaussian state space models were introduced

by Shephard (1994b) and Carter and Kohn (1994) and are very well explained in West and Harrison (1997, chapter 15).

The extension of variance models to a multivariate framework is of special interest, especially in portfolio and asset pricing as will be addressed in the next chapter. Most of the work in this area is related to multivariate ARCH models and factor models as expected. Inferences issues for multivariate ARCH-style models are discussed in Müller and Pole (1998), whereas Bollerslev *et al.* (1988) estimate the multivariate version of the GARCH model. Engle (1982) proposed a k-factor GARCH model where the conditional variance matrix depends on the conditional variances of k orthogonal linear combinations of the data $\mathbf{y}_t$. The estimation was done by maximum likelihood and Engle *et al.* (1990) suggest a simple two-stage procedure. Bollerslev and Engle (1993) give conditions for covariance stationarity of k-factor GARCH models and show how the multivariate IGARCH models allow for the possibility of co-persistence in variance. Diebold and Nerlove (1988) and later King *et al.* (1994) proposed a relatively parsimonious latent factor model where the common changes in volatility are due to a single unobserved latent factor subject to ARCH effects. This model implies similar common movements in the levels and volatilities, that should be modeled separately.

Harvey and Stock (1988) and Harvey *et al.* (1994) incorporate common factors in multivariate SVM using the theory of unobserved components time series models described in Harvey (1989). They discuss the case where there are persistent movements in volatility, modeled by a multivariate random walk. A different generalization of SVM through dynamic factor models with non-Gaussian observation errors is related to the factor ARCH models of Diebold and Nerlove. It was mentioned as a possible multivariate model by Shephard (1996), Kim *et al.* (1998) and discussed by Jacquier *et al.* (1995). In the latter article, JPR proposed the MCMC *single move* sampler

method repeatedly to analyze the multivariate model. Their basic model specifies a stochastic factor structure for the variance covariance matrix as in (7.1) assuming that the univariate factor series f_{ti} follow standard univariate SV models and discussing possible extensions also mentioned in Kim *et al.* (1998). The extensions proposed in this thesis use dynamic factor models (7.1) and allow the log volatilities of the factors to follow a vector autoregression with correlated innovations.

7.2.1 Latent Volatility Processes

Suppose as usual that $\mathbf{y}_t$ is a $q-$dimensional random vector with mean $\boldsymbol{\theta}$ and covariance matrix $\boldsymbol{\Sigma}_t$ following a dynamic factor model (7.1) for $t = 1, \ldots, n$. In the factors variance $\mathbf{H}_t$, define $\lambda_{it} = \log(h_{it})$ for each $i = 1, \ldots k$ and write $\boldsymbol{\lambda}_t = (\lambda_{t1}, \ldots, \lambda_{tk})'$. Assume that the $\boldsymbol{\lambda}_t$ sequence follows a latent stationary vector autoregression of order one, VAR(1), centered around a mean $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k)'$. Namely,

$$\boldsymbol{\lambda}_t = \boldsymbol{\mu} + \boldsymbol{\Phi}(\boldsymbol{\lambda}_{t-1} - \boldsymbol{\mu}) + \boldsymbol{\omega}_t \quad \text{for} \quad t > 1, \tag{7.3}$$

$$\text{and} \qquad \boldsymbol{\lambda}_1 \sim N(\boldsymbol{\lambda}_1 | \boldsymbol{\mu}, \mathbf{W}). \tag{7.4}$$

where $\boldsymbol{\Phi} = \text{diag}(\phi_1, \ldots, \phi_k)$ is the diagonal matrix of factor individual autoregressive coefficients. This $\boldsymbol{\Phi}$ matrix plays the role of the persistence in volatility and its elements generally be close to one, lying in part of the stationary region $|\phi_i| < 1$. The scale parameter $\exp(\boldsymbol{\mu}/2)$ can be thought of as the modal instantaneous volatility. The vector innovations $\boldsymbol{\omega}_t$ are conditionally independent over time, and normally distributed with

$$\boldsymbol{\omega}_t \sim N(\boldsymbol{\omega}_t | \mathbf{0}, \mathbf{U}), \tag{7.5}$$

for some innovations variance $\mathbf{U}$ which is the volatility of the log-volatility. One of the consequences of this model is that each $\boldsymbol{\lambda}_t$ is drawn from the stationary distribution of this vector AR(1) model,

$$\boldsymbol{\lambda}_t \sim N(\boldsymbol{\lambda}_t | \boldsymbol{\mu}, \mathbf{W}), \tag{7.6}$$

where $\mathbf{W}$ satisfies $\mathbf{W} = \mathbf{\Phi}\mathbf{W}\mathbf{\Phi} + \mathbf{U}$. Consequently, the correlation patterns in $\mathbf{U}$ and $\mathbf{W}$, while similar, depend on the autoregressive parameters. In particular, for each factor pair j, h the covariance elements are $\mathbf{W}_{jh} = \mathbf{U}_{jh}/(1 - \phi_j\phi_h)$. The model allows dependencies across volatility series through non-zero off-diagonal entries in $\mathbf{U}$ and $\mathbf{W}$. Model (7.1) combined with equations (7.4) and (7.4) is a natural generalization of the univariate SVM, and thus can be called **Multivariate Stochastic Volatility Models**. In the case of $k = 1$ these models have been used as an approximation to the stochastic volatility diffusion by Hull and White (1987) and Chesney and Scott (1989).

7.3 Bayesian Inference and Computation

The model as specified so far comprises the basic dynamic factor structure (7.1) with supporting assumptions of conditional normality and independence, combined with the SV model (7.4) for the log-volatilities of the factor processes. Model completion for Bayesian analysis requires prior distributions for the full set of parameters, namely

$$\Theta = \{\boldsymbol{\theta}, \mathbf{X}, \mathbf{\Psi}, \boldsymbol{\mu}, \mathbf{\Phi}, \mathbf{U}, \mathbf{f}_t, \mathbf{H}_t; \forall t\}.$$

Bayesian inference for any specified prior distributions requires the computation and summarization of the implied posteriors for these model parameters. Despite the fact that there are too many parameters, the problem can be separated into two big parts. First, from the dynamic factor model setup in (7.1), it is clear that given the log volatilities $\boldsymbol{\lambda}_t$ and hence the factors variances $\mathbf{H}_t$, the loadings matrix $\mathbf{X}$, the factors $\mathbf{f}_t$, the mean $\boldsymbol{\theta}$ and the specific variances $\mathbf{\Psi}$ are conditionally independent of the rest of the parameters. Likewise, given the factors $\mathbf{f}_t$, the log-volatilities $\boldsymbol{\lambda}_t$, the scale parameter $\boldsymbol{\mu}$, the variance $\mathbf{U}$ and the VAR parameters $\mathbf{\Phi}$ are conditionally independent of the rest of the factor model parameters.

The priors are specified in terms of conditionally independent components similar to those in section 6.3.1 for the orthogonal factor model, namely

$$p(\Theta) = p(\mathbf{F})p(\boldsymbol{\theta})p(\mathbf{X})p(\boldsymbol{\Psi})p(\boldsymbol{\mu})p(\boldsymbol{\Phi})p(\mathbf{U}), \tag{7.7}$$

where the marginal priors are either standard reference priors or proper priors that are chosen to be conditionally conjugate, as discussed before. The outlook here is to explore the use of reference priors to the extent possible to provide an initial analysis framework. These prior specifications reflect this view, though these models do require the use of informative, proper priors for some model components due to identification issues. Further, specific applications may use alternative prior specifications, both in terms of informative priors on model components and in terms of prior dependencies between parameters as discussed in section 6.3.1. Assume, for instance, standard reference priors for the univariate entries in the conditional mean, the factor loading matrix and the idiosyncratic variance matrix, so that

$$p(\boldsymbol{\theta})p(\mathbf{X})p(\boldsymbol{\Psi}) \propto \prod_{j=1}^{q} \psi_j^{-1}.$$

Note that the prior for $\mathbf{X}$ is, of course, subject to the specified 0/1 constraints on values in the upper triangle and diagonal in (6.13), so the constant prior density applies only to the remaining, uncertain elements. In addition, use independent normal priors for the univariate elements of $\boldsymbol{\mu}$ and the diagonal elements of $\boldsymbol{\Phi}$. This allows for both reference priors, by setting the prior precisions to zero, and restriction of the values of each ϕ_j by adapting the prior to be truncated to $(0, 1)$. Finally, use an informative inverse Wishart prior for the VAR(1) innovations variance matrix $\mathbf{U}$. This will often be specified with hyper-parameters based on prior data analysis, as illustrated in the applications in next chapters. Notice that an improper reference prior on $\mathbf{U}$, together with that so specified for $\boldsymbol{\Psi}$, is simply inappropriate, as the two determine

separate sources of variability in the data that are confounded in the model. This point, rather critical to model implementation and resulting data analysis, is almost implicit in the prior work of Kim *et al.* (1998). These authors use informative proper priors for innovations variances that parallel our assumptions in their univariate SV models; though they present these priors without further discussion, the propriety is critical in overcoming otherwise potentially problematic confounding issues. Hence initial analysis of previous data, or some other prior elicitation activity, is needed. For instance, in the applied development of next chapter, Bayesian multivariate variance discounting analyses are used to provide preliminary analysis of a reserved initial section of data as input to this. For now, the key point is that the prior for $\mathbf{U}$ is both proper and has conditionally conjugate inverse Wishart form.

Iterative posterior simulation uses an MCMC strategy that extends those in existing SV models (Jacquier *et al.*, 1995; Kim *et al.*, 1998) to the multivariate case, introduces elements of MCMC algorithms for Bayesian factor analysis explained in section 6.3.2 and used in Geweke and Zhou (1996). Novel components derived from models with latent VAR components with correlated innovations are developed here following the work in (Aguilar and West, 1998; West and Aguilar, 1997) in a quite different context described in chapters 9 and 10 of this thesis. From the computation point of view, there are various possible extensions and alternative methods for components of the MCMC analysis below, such as in utilizing some of the ideas from Shephard and Pitt (1997), though have not been explored.

7.4 Implementation of the Gibbs Sampler

This section provide an outline of the iterative posterior sampling algorithm for the unknown parameters Θ for the multivariate SVM above. At some stages, it is easy to simulate directly from the full conditional distributions, at others it is required to in-

troduce novel Metropolis-Hastings accept/reject steps. As in previous sections, each of the following sections represent the full conditional distributions of the parameters and latent variables conditional on the previously simulated values of all other variables. To start, note that because of the conditionally independence relations, the full conditional for the idyosincratic variances $\mathbf{\Psi}$ is the same as for the orthogonal factor model and described in section (6.3.2).

Sampling the latent factors $\boldsymbol{f}_t|\boldsymbol{f}_t^-$

From model (7.1) $\mathbf{y}_t$ and $\mathbf{f}_t$ are jointly normal, namely

$$\begin{pmatrix} \mathbf{f}_t \\ \mathbf{y}_t \end{pmatrix} \sim N\left[\begin{pmatrix} \mathbf{0} \\ \boldsymbol{\theta} \end{pmatrix}, \begin{pmatrix} \mathbf{H}_t & \mathbf{A}_t\mathbf{Q}_t \\ \mathbf{Q}_t\mathbf{A}_t' & \mathbf{Q}_t \end{pmatrix}\right].$$

In consequence, the full conditional posterior distributions for the factors $\mathbf{f}_t$ are given by $\mathbf{f}_t|\mathbf{y}_t \sim N(\mathbf{A}_t(\mathbf{y}_t - \boldsymbol{\theta}), \mathbf{H}_t - \mathbf{A}_t\mathbf{Q}_t\mathbf{A}_t^{'})$, where $\mathbf{Q}_t = \mathbf{X}\mathbf{H}_t\mathbf{X}^{'} + \mathbf{\Psi}$ and $\mathbf{A}_t = \mathbf{H}_t\mathbf{X}^{'}\mathbf{Q}_t^{-1}$. The $\mathbf{f}_t$ are conditionally independent and so sample values are drawn independently from this set of normal distributions for $t = 1, 2 \ldots, n$.

Sampling the conditional mean and loadings matrix $\boldsymbol{\theta}, \boldsymbol{X}|(\boldsymbol{\theta}, \boldsymbol{X})^-$

The full conditional distribution for the mean $\boldsymbol{\theta}$ and the loadings matrix (6.13) is very similar to the orthogonal factor model from the previous chapter. The conditional likelihood function for these parameters is $\prod_{t=1}^{n} N(\mathbf{y}_t|\boldsymbol{\theta} + \mathbf{X}\mathbf{f}_t, \mathbf{\Psi})$, which is a log-quadratic form in $\boldsymbol{\theta}$ and the uncertain elements of $\mathbf{X}$. Thus, combined with a normal or uniform reference prior, implies a multivariate normal conditional posterior.

- From the dynamic model (7.1) and due to the shape of the loadings matrix (6.13), the first series is modeled by $y_{1t} = \theta_1 + f_{1t}\epsilon_{1t}$, for $t = 1, \ldots n$. Thus, the

full conditional distribution for θ_1 is straightforward,

$$\theta_1|\theta_1^- \sim N\left(\frac{\sum_{t=1}^n (y_{1t} - f_{1t})}{n}, \frac{\psi_1}{n}\right)$$

- To sample the $j-$th row note that $y_{jt} = \theta_j + \sum_{l=1}^{j-1} x_{jl} f_{lt} + f_{jt} + \epsilon_{jt}$, for $j \leq k$ and $t = 1, \ldots n$. The full conditional posterior distribution for the parameters is given by

$$\theta_j, x_{j1}, \ldots x_{j,j-1}|(\theta_j, x_{j1}, \ldots x_{j,j-1})^- \sim N\left((\mathbf{B}_j'\mathbf{B}_j)^{-1}(\mathbf{B}_j'\mathbf{Y}_j^*), \psi_j(\mathbf{B}_j'\mathbf{B}_j)^{-1}\right),$$

  where $\mathbf{Y}_j^*$ is the $n \times 1$ vector with elements $y_{jt} - f_{jt}$ and $\mathbf{B}_j$ is a $n \times j$ matrix with rows $(1, f_{1t}, \ldots, f_{j-1,t})'$ for $t = 1, \ldots n$.

- The remaining $q - k$ rows of the loadings matrix and mean components are sampled in the same way as in section 6.3.2 for the orthogonal factor model.

Sampling the latent log volatilities $\boldsymbol{\lambda}_t|\boldsymbol{\lambda}_t^-$

As it was established before, given the imputed values for the full factor process $\mathbf{f}_t$ over time t, the volatilities $\mathbf{H}_t$ are conditionally independent of $\boldsymbol{\theta}, \mathbf{X}$ and $\boldsymbol{\Psi}$. The approach taken here is to generalize the *Offset Mixture Method* developed in Kim *et al.* (1998) and combine it with recent developments in VAR models as in (Aguilar and West, 1998; West and Aguilar, 1997). The idea is to sample all the log-volatilities $\boldsymbol{\lambda}_t$ at once by producing an efficient algorithm based on DLM theory. To start, write the distributional assumptions on the factors $\mathbf{f}_t \sim N(\mathbf{0}, \mathbf{H}_t)$ as a non-Gaussian dynamic linear model. Specifically, for each factor i at time t write $f_{it} = \exp(\lambda_{it}/2)\zeta_{it}$, where $\zeta_{it} \sim N(0, 1)$ and $\lambda_{it} = \log(h_{it})$. Apply the traditional nonlinear transformation $z_{it} = \log(f_{it}^2)$ to get,

$$z_{it} = \lambda_{it} + \nu_{it},$$

for $i = 1, \ldots, k$ and $t = 1, \ldots n$. Note that the implied observation errors, defined by $\nu_{it} = \log(\zeta_{it}^2)$, are independent and follow $\log{-}\chi_1^2$ distributions. To deal with this, the *multi-move* sampler approach from the literature in univariate SVM will be used to approximate the $log(\chi^2)$ distribution with a discrete mixture of seven normals, namely

$$p(\nu_{it}) \approx \sum_{j=1}^{7} q_j N(\nu_{it} | m_j - 1.2704, v_j^2), \tag{7.8}$$

with mixing probabilities q_j, means m_j and variances v_j^2 selected to closely approximate the exact density and are given in Table 7.1 for $j = 1, \ldots 7$; see Kim *et al.* (1998) for details of the approximation. It should be noted that the mixture density can also be written in terms of indicator variables. Specifically, it is convenient to augment the parameter space and introduce indicator variables s_{it} over the set $\{1, 2, \ldots, 7\}$ to identify the normal mixture component for ν_{it}. Explicitly, ν_{it} is conditionally normal,

$$\begin{aligned} \nu_{it} | s_{it} = j &\sim N(\nu_{it} | m_j - 1.2704, v_j^2), \\ \text{with } \Pr(s_{it} &= q_j) \end{aligned}$$

for $j = 1, \ldots, 7$. Note that, this new variable s_{it} will be an extra parameter to include in the MCMC algorithm. Given the indicator variables s_{it}, the observation errors ν_{it} are independent over i as well as over t. This is a direct multivariate extension of the univariate approach taken by Kim *et al.* (1998). Explicitly, conditional on the indicators, a multivariate dynamic linear model for the sequence of log-volatility vectors is defined,

$$\begin{aligned} \mathbf{z}_t &= \boldsymbol{\lambda}_t + \boldsymbol{\nu}_t \\ \boldsymbol{\lambda}_t &= \boldsymbol{\mu} + \boldsymbol{\Phi}(\boldsymbol{\lambda}_{t-1} - \boldsymbol{\mu}) + \boldsymbol{\omega}_t \end{aligned} \tag{7.9}$$

where $\mathbf{z}_t = (z_{1t}, \ldots, z_{kt})'$, $\boldsymbol{\nu}_t = (\nu_{1t}, \ldots, \nu_{kt})'$ and with mutually independent errors $\boldsymbol{\nu}_t$ and $\boldsymbol{\omega}_t$. Equations (7.9) and (7.4) combined with the initial version for the $\boldsymbol{\lambda}_1$

j	Means m_j	Variances v_j^2	Weights q_j
1	-10.12999	5.79596	0.00730
2	-3.97281	2.61369	0.10556
3	-8.56686	5.17950	0.00002
4	2.77786	0.16735	0.04395
5	0.61942	0.64009	0.34001
6	1.79518	0.34023	0.24566
7	-1.08819	1.26261	0.25750

Table 7.1: Moments and weights of the mixing distributions to approximate a $\log(\chi^2)$*distribution*.

(7.4) define a conditionally Gaussian multivariate dynamic linear model with known variance matrices and state vector sequence $\boldsymbol{\lambda}_t$. To sample the log-volatilities apply DLMs theory as detailed in West and Harrison (1997, chapter 15). The forward-filtering, backwards-sampling algorithm is then used to sample the full set of vectors $\{\boldsymbol{\lambda}_t, t = 1, \ldots, n\}$ from the implied conditional posterior following formulas in the Appendix A.1. Note that, in more elaborate models for the volatility processes, the alternative sampling method using the simulation smoother of De Jong and Shephard (1995) may have computational advantages not realized in this simple VAR model setup.

One extra point to note is that ζ_{it} and ω_{it} are assumed to be uncorrelated. Conversely, suppose that these two errors are correlated with correlation parameter ρ_i for $i = 1, \ldots, k$. This assumption could be very important for stock returns producing what is called the *leverage effect*. That is, there is an asymmetrical relationship between the sign and magnitude of prices. If $\rho_i < 0$ then a decrease in price (negative return) is associated with a positive variance shock ω_{it}. However, information about the sign of this correlation is lost due to the transformation. Harvey and Shephard (1996) estimated the parameter by quasi-maximum likelihood method treating the

signs as an ancillary statistic.

Sampling the indicators $s_{it}|s_{it}^-$

Given the sampled values of $\boldsymbol{\lambda}_t$, sample each one of the indicator variables s_{it} independently from multinomial distributions with posterior probabilities,

$$\Pr(s_{it} = j|s_{it}^-) = \frac{q_j N(z_{it}|\lambda_{it} + m_j - 1.2704, v_j^2)}{\sum_{j=1}^7 q_j N(z_{it}|\lambda_{it} + m_j - 1.2704, v_j^2)} \quad j \in \{1, 2, \ldots, 7\}, \tag{7.10}$$

for $i = 1, \ldots, k$ and $t = 1, \ldots, n$. The indicators being conditionally independent and so easily sampled.

Sampling the VAR innovations variance matrix $\boldsymbol{U}|\boldsymbol{U}^-$

The structure of the full conditional posterior for the innovations variance $\mathbf{U}$, and the resulting Metropolis-Hastings strategy for simulation, is precisely as developed for component VAR models in a quite different context in West and Aguilar (1997) and Aguilar and West (1998). Define the volatilities deviation from the mean $\boldsymbol{\gamma}_t = \boldsymbol{\lambda}_t - \boldsymbol{\mu}$ and reparametrize the evolution equation (7.4) as,

$$\boldsymbol{\gamma}_t = \boldsymbol{\Phi}\boldsymbol{\gamma}_{t-1} + \boldsymbol{\omega}_t \tag{7.11}$$

and

$$\boldsymbol{\gamma}_1 \sim N(\boldsymbol{\gamma}_1|\mathbf{0}, \mathbf{W}). \tag{7.12}$$

The full conditional posterior for $\mathbf{U}$ is given, in terms of the inverse $\mathbf{U}^{-1}$, by

$$\begin{aligned} p(\mathbf{U}^{-1}|\{\boldsymbol{\gamma}_t\}, \boldsymbol{\Phi}) &\propto p(\mathbf{U}^{-1})p(\{\boldsymbol{\gamma}_t\}|\boldsymbol{\Phi}, \mathbf{U}) \\ &\propto p(\mathbf{U}^{-1})p(\boldsymbol{\gamma}_1|\boldsymbol{\Phi}, \mathbf{U}) \prod_{t=2}^{n} p(\boldsymbol{\gamma}_t|\boldsymbol{\gamma}_{t-1}, \boldsymbol{\Phi}, \mathbf{U}) \\ &\propto p(\mathbf{U}^{-1})a(\mathbf{U})|\mathbf{U}^{-1}|^{(n-1)/2}\exp(-\text{trace}(\mathbf{U}^{-1}\mathbf{G})) \end{aligned}$$

with

$$\mathbf{G} = \sum_{t=2}^{n} (\boldsymbol{\gamma}_t - \boldsymbol{\Phi}\boldsymbol{\gamma}_{t-1})(\boldsymbol{\gamma}_t - \boldsymbol{\Phi}\boldsymbol{\gamma}_{t-1})'$$

and

$$a(\mathbf{U}) = |\mathbf{W}|^{-1/2} \exp(-\text{trace}(\mathbf{W}^{-1}\boldsymbol{\gamma}_1\boldsymbol{\gamma}_1')/2)$$

with $\mathbf{W} = \boldsymbol{\Phi}\mathbf{W}\boldsymbol{\Phi} + \mathbf{U}$. Under a specified Wishart prior $W_{r_0}(\mathbf{U}^{-1}|\mathbf{R}_0)$, this posterior density is proportional to

$$a(\mathbf{U}) W_r(\mathbf{U}^{-1}|\mathbf{R})$$

where $r = r_0 + n - 1$ and $r\mathbf{R} = r_0\mathbf{R}_0 + (n-1)\mathbf{G}$. Use the Wishart distribution as a proposal distribution in a Metropolis-Hastings algorithm. That is, given a "current" value of $\mathbf{U}^{-1}$ and corresponding $\mathbf{W}$, sample a "candidate" value $\mathbf{U}^{-1*}$ from the Wishart distribution $W_r(\mathbf{U}^{-1}|\mathbf{R})$, and accept it with probability

$$\min\{1, a(\mathbf{U}^*)/a(\mathbf{U})\}$$

where $\mathbf{W}^* = \boldsymbol{\Phi}\mathbf{W}^*\boldsymbol{\Phi} + \mathbf{U}^*$.

Sampling the VAR coefficients $\boldsymbol{\Phi}|\boldsymbol{\Phi}^-$

The structure of the conditional posterior for $\boldsymbol{\Phi}$, depends only on the centered log volatilities $\boldsymbol{\gamma}_t = \boldsymbol{\lambda}_t - \boldsymbol{\mu}$ for each t, and the innovations variance $\mathbf{U}$,

$$\begin{aligned} p(\boldsymbol{\Phi}|\{\boldsymbol{\gamma}_t\}, \mathbf{U}) &\propto p(\boldsymbol{\Phi}) p(\{\boldsymbol{\gamma}_t\}|\boldsymbol{\Phi}, \mathbf{U}) \\ &\propto p(\boldsymbol{\Phi}) p(\boldsymbol{\gamma}_1|\boldsymbol{\Phi}) \prod_{t=2}^{N} p(\boldsymbol{\gamma}_t|\boldsymbol{\gamma}_{t-1}, \boldsymbol{\Phi}) \\ &\propto p(\boldsymbol{\Phi}) N(\boldsymbol{\gamma}_1|\mathbf{0}, \mathbf{W}) \prod_{t=2}^{N} N(\boldsymbol{\gamma}_t|\boldsymbol{\Phi}\boldsymbol{\gamma}_{t-1}, \mathbf{U}) \end{aligned}$$

where $\mathbf{W} = \boldsymbol{\Phi}\mathbf{W}\boldsymbol{\Phi} + \mathbf{U}$ is easily evaluated as a function of $\boldsymbol{\Phi}$ and $\mathbf{U}$. Write $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_k)'$ for the diagonal of $\boldsymbol{\Phi}$, and $\mathbf{E} = \text{diag}(\boldsymbol{\gamma}_{t-1})$. Then the conditional posterior

may be written as proportional to

$$p(\boldsymbol{\Phi})c(\boldsymbol{\Phi})N(\boldsymbol{\phi}|\mathbf{b},\mathbf{B})$$

where

$$\mathbf{b} = \mathbf{B}\sum_{t=2}^{n}\mathbf{E}'\mathbf{U}^{-1}\boldsymbol{\gamma}_t \quad \text{and} \quad \mathbf{B}^{-1} = \sum_{t=2}^{N}\mathbf{E}'\mathbf{U}^{-1}\mathbf{E},$$

and

$$c(\boldsymbol{\Phi}) = |\mathbf{W}|^{-I/2}\exp(-\text{trace}(\mathbf{W}^{-1}\boldsymbol{\gamma}_1\boldsymbol{\gamma}_1')/2).$$

Under independent uniform priors for each of the ϕ_j,[1] the full conditional posterior distribution for $\boldsymbol{\Phi}$ is the above multivariate normal $N(\boldsymbol{\phi}|\mathbf{b},\mathbf{B})$ truncated to the $(0,1)$ regions in each dimension, and then multiplied by the factor $c(\boldsymbol{\Phi})$. This may be sampled in several ways like using a Metropolis Hastings algorithm that takes the truncated multivariate normal component as a proposal distribution. Explicitly, given a "current" value of $\boldsymbol{\phi}$, with corresponding matrices $\boldsymbol{\Phi}$ and $\mathbf{W}$, sample a "candidate" vector $\boldsymbol{\phi}^*$ from this truncated normal, compute the corresponding diagonal matrix $\boldsymbol{\Phi}^*$ and variance matrix $\mathbf{W}^*$ such that $\mathbf{W}^* = \boldsymbol{\Phi}^*\mathbf{W}^*\boldsymbol{\Phi}^* + \mathbf{U}$, then accept this new $\boldsymbol{\phi}$ vector with probability

$$\min\{1, c(\boldsymbol{\Phi}^*)/c(\boldsymbol{\Phi})\}.$$

Sampling the mean log-volatilities $\boldsymbol{\mu}|\boldsymbol{\mu}^-$

The full conditional distribution for the mean of the log volatilities is given by,

$$\begin{aligned} p(\boldsymbol{\mu}|\boldsymbol{\mu}^-) &\propto p(\boldsymbol{\mu})p(\boldsymbol{\lambda}_1|\boldsymbol{\mu})\prod_{t=2}^{N}p(\boldsymbol{\lambda}_t|\boldsymbol{\lambda}_{t-1},\boldsymbol{\mu}) \\ &\propto p(\boldsymbol{\mu})N(\boldsymbol{\lambda}_1|\boldsymbol{\mu},\mathbf{W})\prod_{t=2}^{N}N(\boldsymbol{\lambda}_t|\boldsymbol{\mu}+\boldsymbol{\Phi}(\boldsymbol{\lambda}_{t-1}-\boldsymbol{\mu}),\mathbf{U}) \end{aligned}$$

[1] Kim *et al.* (1998) propose an informative prior to avoid problems when the data are close to be non-stationary, $\phi_j \approx 1$. The proposed prior involves letting $\phi_j = 2\phi_j^* - 1$ where $\phi_j^* \sim \text{Beta}(20, 1.5)$.

where $\mathbf{W} = \mathbf{\Phi}\mathbf{W}\mathbf{\Phi} + \mathbf{U}$ as explained above. Under a normal prior $N(\boldsymbol{\mu}|\mathbf{m}_0, \mathbf{M}_0)$, the full conditional posterior is normal $N(\boldsymbol{\mu}|\mathbf{m}, \mathbf{M})$ with

$$\mathbf{M}^{-1} = \mathbf{M}_0^{-1} + \mathbf{W}^{-1} + (n-1)(\mathbf{I} - \mathbf{\Phi})\mathbf{U}^{-1}(\mathbf{I} - \mathbf{\Phi})$$

and

$$\mathbf{m} = \mathbf{M}\left[\mathbf{M}_0^{-1}\mathbf{m}_0 + \mathbf{W}^{-1}\boldsymbol{\lambda}_1 + (\mathbf{I} - \mathbf{\Phi})\mathbf{U}^{-1}\sum_{t=2}^{n}(\boldsymbol{\lambda}_t - \mathbf{\Phi}\boldsymbol{\lambda}_{t-1})\right].$$

The case of a uniform reference prior is recovered by setting $\mathbf{M}_0^{-1} = \mathbf{0}$.

7.5 Forecasting

The goal of any forecasting exercise is to compute predictive densities. This is not done analytically here, but instead the posterior samples obtained from the Gibbs sampler algorithm are used to draw samples from the desire distributions. One of the main advantages of the stochastic volatility models is that by assuming a stationary process on the log-volatilities one can easily obtain reasonable forecasts of the actual volatilities and hence the factor processes themselves. Explicitly, the one-step-ahead prediction density for the factors is defined by

$$p(\mathbf{f}_{t+1}|\mathbf{D}_t) = \int p(\mathbf{f}_{t+1}|\mathbf{H}_{t+1}, \Xi, \mathbf{D}_t)p(\mathbf{H}_{t+1}|\mathbf{H}_t, \Xi, \mathbf{D}_t)p(\mathbf{H}_t, \Xi|\mathbf{D}_t)d\Xi d\mathbf{H}_t d\mathbf{H}_{t+1}$$

where $\Xi = \{\boldsymbol{\theta}, \mathbf{X}, \mathbf{\Psi}, \boldsymbol{\mu}, \mathbf{\Phi}, \mathbf{U}, \mathbf{f}_1, \ldots, \mathbf{f}_t\}$ and $\mathbf{D}_t = \{\mathbf{y}_1, \ldots, \mathbf{y}_t\}$. This distribution is sampled by the method of composition as follows. Use the posterior simulation values of $p(\mathbf{H}_t, \Xi|\mathbf{D}_t)$ obtained from the Gibbs sampler algorithm to sample $p(\mathbf{H}_{t+1}|\mathbf{H}_t, \Xi, \mathbf{D}_t)$. Namely, for each set of samples $\mathbf{H}_t^{(j)}$ and $\Xi^{(j)}$(j=1,2,...,M) use the VAR(1) evolution equation, $\boldsymbol{\lambda}_{t+1} = \boldsymbol{\mu} + \mathbf{\Phi}(\boldsymbol{\lambda}_t - \boldsymbol{\mu}) + \boldsymbol{\omega}_{t+1}$, to draw values from,

$$\boldsymbol{\lambda}_{t+1}^{(j)}|\boldsymbol{\lambda}_t^{(j)} \sim N(\boldsymbol{\mu}^{(j)} + \mathbf{\Phi}^{(j)}(\boldsymbol{\lambda}_t^{(j)} - \boldsymbol{\mu}^{(j)}), \mathbf{U}^{(j)}).$$

This will directly give samples of the factors volatility processes $\mathbf{H}_{t+1}^{(j)} = \exp\{\boldsymbol{\lambda}_{t+1}^{(j)}\}$. Then, if required, draws from the predictive distribution of the actual factors can be obtained by sampling $\mathbf{f}_{t+1}^{(j)} \sim N(\mathbf{0}, \mathbf{H}_{t+1}^{(j)})$. However, in many applications, the interest lies more on forecasting the actual time series $\mathbf{y}_{t+1}$ rather than the latent processes. The methodology is easily extended to pursue these goals by computing the one-step-ahead predictive distribution of the actual observations,

$$p(\mathbf{y}_{t+1}|\mathbf{D}_t) = \int p(\mathbf{y}_{t+1}|\boldsymbol{\Sigma}_{t+1}, \Theta, \mathbf{D}_t) p(\boldsymbol{\Sigma}_{t+1}|\boldsymbol{\Sigma}_t, \mathbf{D}_t) p(\boldsymbol{\Sigma}_t|\mathbf{D}_t) p(\Theta|\mathbf{D}_t) d\Theta d\boldsymbol{\Sigma}_t \boldsymbol{\Sigma}_{t+1},$$

where $\Theta = \{\boldsymbol{\theta}, \mathbf{X}, \boldsymbol{\Psi}, \boldsymbol{\mu}, \boldsymbol{\Phi}, \mathbf{U}, \mathbf{f}_t, \mathbf{H}_t; \forall t\}$. This distribution can be sampled by the method of composition as well by using the computed samples of the factors volatility processes $\mathbf{H}_{t+1}^{(j)}$ to find direct samples of the observations variance,

$$\boldsymbol{\Sigma}_{t+1}^{(j)} = \mathbf{X}^{(j)} \mathbf{H}_{t+1}^{(j)} \mathbf{X}^{\prime(j)} + \boldsymbol{\Psi}^{(j)},$$

for $j = 1, 2, \ldots, M$. At this point, and without further sampling, one can easily obtain the moments of the predictive distribution $\mathbf{y}_{t+1}|\mathbf{D}_t$ using the following facts:

$$\begin{aligned}
\mathrm{E}(\mathbf{y}_{t+1}|\mathbf{D}_t) &= \mathrm{E}\left[\mathrm{E}(\mathbf{y}_{t+1}|\boldsymbol{\theta}, \boldsymbol{\Sigma}_{t+1})|\mathbf{D}_t)\right] && (7.13) \\
&= \mathrm{E}(\boldsymbol{\theta}|\mathbf{D}_t) \quad \text{and} && (7.14) \\
\mathrm{Var}(\mathbf{y}_{t+1}|\mathbf{D}_t) &= \mathrm{Var}\left[\mathrm{E}(\mathbf{y}_{t+1}|\boldsymbol{\theta}, \boldsymbol{\Sigma}_{t+1})|\mathbf{D}_t\right] + \mathrm{E}\left[\mathrm{Var}(\mathbf{y}_{t+1}|\theta, \boldsymbol{\Sigma}_{t+1})|\mathbf{D}_t\right] \\
&= \mathrm{Var}(\boldsymbol{\theta}|\mathbf{D}_t) + \mathrm{E}(\boldsymbol{\Sigma}_{t+1}|\mathbf{D}_t),
\end{aligned}$$

which can be evaluated using Monte Carlo methods. Moreover, samples from the actual predictive distribution of the actual returns are obtained by drawing values from $\mathbf{y}_{t+1} \sim N(\boldsymbol{\theta}^{(j)}, \boldsymbol{\Sigma}_{t+1}^{(j)})$. These samples can further used for diagnostic checking by calculating, for instance, probabilities that $\mathbf{y}_{t+1}^2$ is less than the observed value $\mathbf{z}_{t+1}^2$,

$$\Pr(\mathbf{y}_{t+1}^2 < \mathbf{z}_{t+1}^2) = \frac{1}{M} \sum_{j=1}^{M} \Pr(\mathbf{y}_{t+1}^2 < \mathbf{z}_{t+1}^2 | \boldsymbol{\theta}^{(j)}, \boldsymbol{\Sigma}_{t+1}^{(j)}).$$

These probabilities are independent and identically distributed uniform random variables which can be mapped using the inverse cdf method to standard normal samples and used for normality tests as described in Kim *et al.* (1998).

In the illustrations in next chapter the dynamic factor model with stochastic volatility components is used to estimate and forecast time-varying covariance matrices and applied in dynamic asset allocation and portfolio construction.

Chapter 8

Dynamic Asset Allocation

In the last chapter a new class of dynamic factor models for multivariate time series and the incorporation of stochastic volatility components for latent factor processes were developed. The models are direct generalizations of univariate stochastic volatility models, and represent specific varieties of models recently discussed in the growing multivariate stochastic volatility literature. The predicting ability of these models, through modeling dependencies in volatility processes, makes them suitable for a wide variety of applications. In particular, potential improvements, relative to standard methods, in short-term forecasting of multiple series are of great interest in the financial and econometric areas. In this chapter, the dynamic factor stochastic volatility models are used in forecasting and portfolio construction in studies of multiple international exchange rates series. The results are compared and connected with the much simpler method of dynamic variance discounting that are reviewed in the next section and used to obtain prior information on parameters for the dynamic factor model as discussed above. Discounting methods have been, for over a decade, a standard approach in applied financial econometrics in the Bayesian forecasting world. This standard method simply "reacts" to volatility changes but do not anticipate the forms of those changes as explained below.

8.1 Bayesian Discount Estimation

The univariate discounting methods developed in West and Harrison (1997) and described in section 5.2 is now extended to the general multivariate context. This approach leads to the embedding of exponentially smoothed estimates of "local" variance/covariance structure within a Bayesian modeling framework, and so provides for adaptation to stochastic changes as time series data are processed. Modifications to allow for changes in discount rates in order to adapt to varying degrees of change, including marked/abrupt changes in volatility patterns, extend the basic approach. The resulting update equations for sequences of estimated volatility matrices have univariate components that relate closely to variants of ARCH and GARCH models, and so it is not surprising that they have proven useful in many applications.

The matrix discounting methods were introduced in Quintana and West (1987) and Quintana and West (1988) and have been used as a component of applied Bayesian forecasting models in financial econometric settings for over a decade (Quintana, 1992; Putnam and Quintana, 1994a; Putnam and Quintana, 1994b; Putnam and Quintana, 1995; Quintana and Putnam, 1996; Quintana *et al.*, 1995). Variants of the basic method of variance matrix discounting, described next, have formal theoretical bases in matrix-variate "random walks" Uhlig (1994, 1997); see West and Harrison (1997, section 16.4.5) for details.

8.1.1 Variance-Covariance Discounting

Consider as usual a $q-$variate conditionally independent time series $\mathbf{y}_t$, $(t = 1, 2, \ldots, n)$ normally distributed with stochastically time-varying variance matrices $\mathbf{\Sigma}_t$, namely, $N(\mathbf{y}_t|\mathbf{0}, \mathbf{\Sigma}_t)$ for each t. Bayesian discounting methods arise from a matrix-variate random walk for the $\mathbf{\Sigma}_t$ process, resulting in simple sequential updating of inverse Wishart posteriors for inference on $\mathbf{\Sigma}_t$ as time evolves. Using the notation of West

and Harrison (1997), the time t posterior is of the form $p(\mathbf{\Sigma}_t|\mathbf{D}_t) = W_{n_t}^{-1}(\mathbf{\Sigma}_t|\mathbf{S}_t)$ where $\mathbf{D}_t = \{\mathbf{D}_0, \mathbf{y}_1, \ldots, \mathbf{y}_t\} = \{\mathbf{D}_{t-1}, \mathbf{y}_t\}$ is the sequentially updated information set at time t. Here n_t is the degrees of freedom and $\mathbf{S}_t$ a posterior estimate of $\mathbf{\Sigma}_t$, the posterior harmonic mean. The notation $W_r^{-1}(\cdot|\mathbf{S})$ indicates the inverse Wishart distribution with r degrees of freedom and scale matrix $\mathbf{S}$ (see West and Harrison, as referenced). The sequence of estimates $\mathbf{S}_t$ is trivially updated sequentially in time by the forward exponential moving average formula

$$\mathbf{S}_t = (1 - a_t)\mathbf{S}_{t-1} + a_t \mathbf{y}_t \mathbf{y}_t' \tag{8.1}$$

with weight $a_t = 1/(1 + \delta n_{t-1})$ based on a discount factor δ. This discount factor lies in $(0, 1)$, is typically between 0.9 and 1 and will be very close to unity for data at high sampling rates. Having analyzed a fixed stretch of data $t = 1, \ldots, n$, the sequence of estimates $\mathbf{S}_t$ is revised by the related backward smoothing formula to incorporate the data at times $t+1, \ldots, n$ in inference on $\mathbf{\Sigma}_t$. Denoting the revised estimate of $\mathbf{\Sigma}_t$ by $\mathbf{S}_{t,n}$, the formula is given in terms of inverse variance matrices by the backward recursion

$$\mathbf{S}_{t,n}^{-1} = (1 - \delta)\mathbf{S}_t^{-1} + \delta \mathbf{S}_{t+1,n}^{-1} \tag{8.2}$$

for each $t = n-1, n-2, \ldots, 1$, and starting with $\mathbf{S}_{n,n} = \mathbf{S}_n$. See West and Harrison (1997, pp608-609) for further details, and the various references by Quintana and coauthors listed above for development and application in econometric finance.

Despite the fact that there are ranges of possible models for a non-zero, stochastic mean function that may be of interest, the most basic model of an assumedly constant mean $\boldsymbol{\theta}$ is considered here, so that

$$\mathbf{y}_t \sim N(\boldsymbol{\theta}, \mathbf{\Sigma}_t), \tag{8.3}$$

consistent with the dynamic factor model (7.1). In this case the forward updating

formula (8.1) is modified by replacing the observation $\mathbf{y}_t$ by the appropriate standardized forecast error, following West and Harrison (1997, pp608-609).

As it was mentioned earlier, one important issue in any model for variance-covariance matrices is understanding the nature of changes in covariance patterns, and the underlying latent mechanisms driving such changes. Because of that, in some applications of variance matrix discounting methods, a principal component or factor structure of the variance matrices $\mathbf{\Sigma}_t$ has been explored. See, for example, the studies of monthly exchange rate time series in Quintana and West (1987) also reported in West and Harrison (1997, pp608-609). In this particular application, a standard principal component decomposition of the estimates $\mathbf{S}_{t,n}$ provides insight into the related component structure of $\mathbf{\Sigma}_t$ and the relevant numbers of "important" factors k. Usually, this will yield a small number of dominant components representing latent orthogonal factors contributing measurably to both total variability in the series and the covariance structure, together with additional residual components. This is a nice connection with the dynamic factor decomposition from previous chapters. However, dynamic factor models does not necessarily involve orthogonal factor structure as in principal component methods. Furthermore, dynamic factor models explicitly represent the latent structure and make direct inferences of the factor processes and their parameters. In addition, discounting methods do not have real predictive capabilities, simply allowing for and estimating changes rather than anticipating them. The dynamic factor model (7.1), as stated above, allows for short-term forecasting in a natural way. This will come clear in the following comparisons of standard discounting methods and dynamic factor models in an application of international exchange rates series together with the resulting dynamic portfolio allocations. A final note on these models is that with finance data the mean vector $\boldsymbol{\theta}$ will have very small elements having small impact in resulting inferences on the $\mathbf{\Sigma}_t$ sequence. However,

the impact on portfolio allocations that are mean-dependent are not necessarily small and hence important. Extensions to models in which the time series has a non-zero mean, possibly modeled via a dynamic regression model are of great interest in the applied models of Quintana and Putnam (1996), Quintana *et al.* (1995). These extensions will be discussed later in the thesis.

8.2 Studies of International Exchange Rates

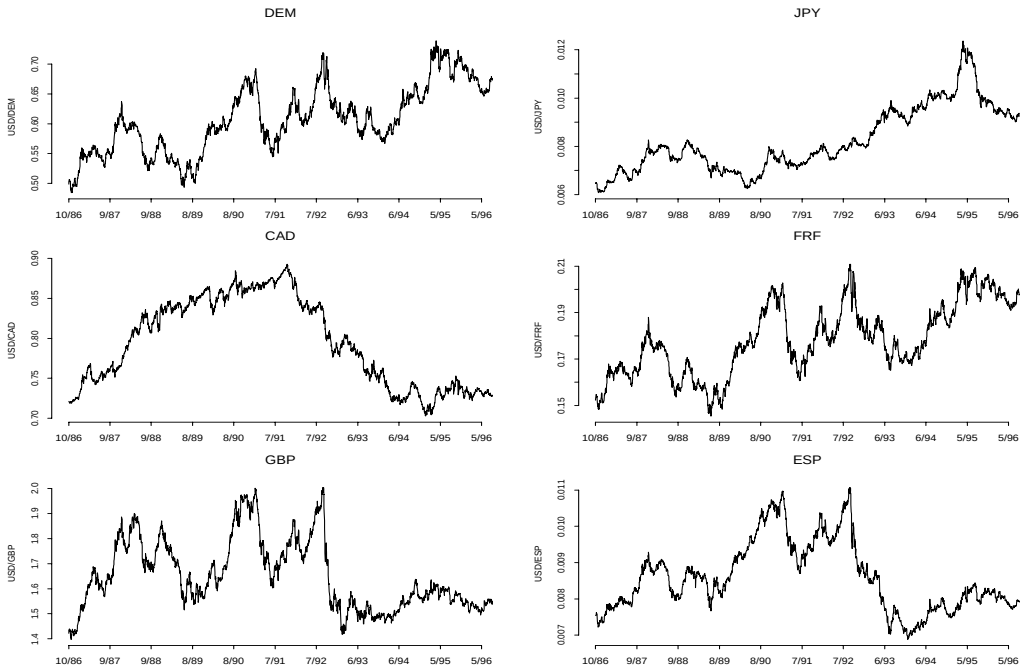

Figure 8.1: Daily exchange rate time series from 10/09/86 to 08/09/96.

8.2.1 Data and Initial Discounting Analyses

Figure 8.1 displays time series graphs of the weekday closing spot exchange rates of several currencies quoted in US dollars (USD) covering the period from 10/09/86 to 08/09/96, for a total of 2,567 observations on each series. The currencies are, in order, the Deutschmark/Mark (DEM), Japanese Yen (JPY), Canadian Dollar

(CAD), French Franc (FRF), British Pound (GBP) and Spanish Peseta (ESP). From an economic point of view it is better to study the returns on the different currencies rather than the prices. Therefore, the analyses presented here are based on the one-day-ahead returns calculated as $y_{it} = s_{it}/s_{i,t-1} - 1$ and graphed in Figure 8.2 for $i = 1, \ldots, q = 6$.

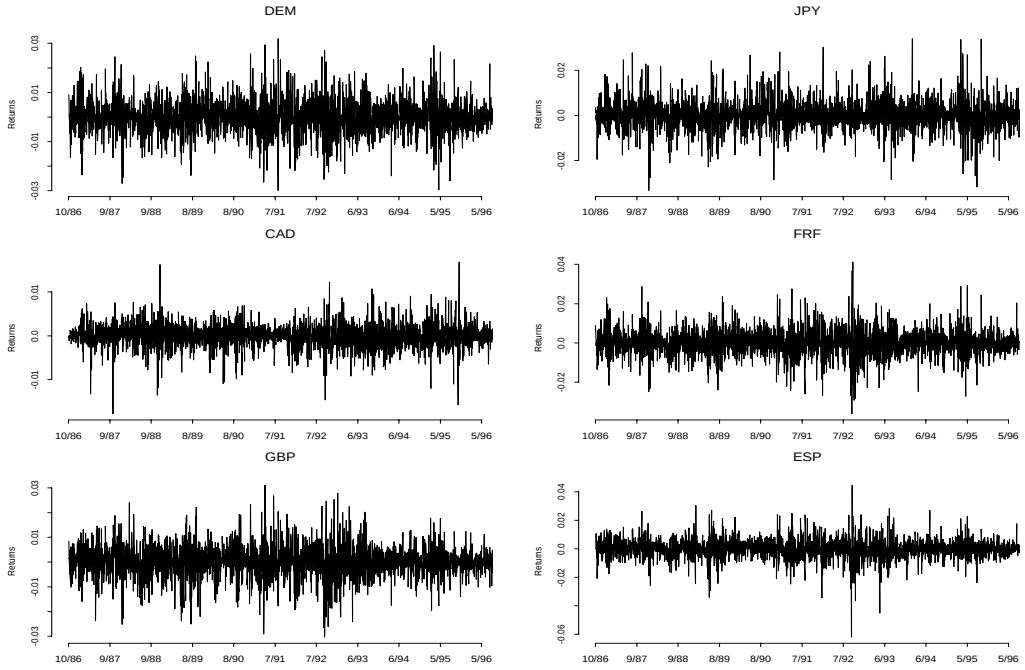

Figure 8.2: Daily exchange rate returns.

Initial analyses were run using the variance matrix discounting method described in section 8.1.1 and are summarized in Figures 8.3, 8.4 and 8.5. Three separate analyses were run, differing only through the value of the discount factor, specified as $\delta = 0.9, 0.95, 0.99$ for the three cases. The initial prior distribution in each case is very vague, namely $W_1^{-1}(\mathbf{\Sigma}|\mathbf{I})$. Figure 8.3 displays the time trajectories of the diagonal elements of $S_{t,n}$, i.e., the sequence of posterior point estimates of the conditional variances of the six currencies. In addition, Figure 8.4 shows the estimated time-varying correlation structure between the returns series. In each of the three analyses, principal components decompositions were made of each of the posterior estimates

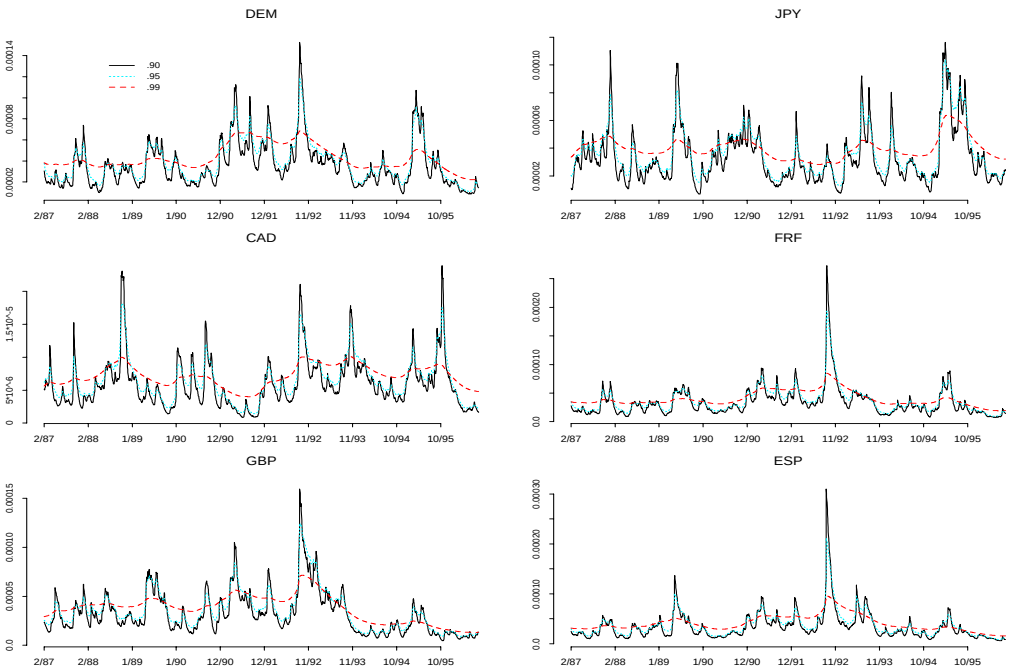

Figure 8.3: Conditional variances from discount analyses.

$\mathbf{S}_{t,n}$ over $t = 1, \ldots, n = 2567$. In each analysis and essentially uniformly over the time period, this yields three dominant components with fairly stable time trajectories for the corresponding eigenvectors representing the dynamic factor loadings. Figure 8.5 displays the corresponding related trajectories of the estimates of the volatilities of the underlying orthogonal latent factors, namely the eigenvalues of the $\mathbf{S}_{t,n}$ matrices over time. The greater adaptivity induced by lower discount factors is apparent in these graphs. Explicitly, the very low $\delta = 0.9$ is over-adaptive, responding very markedly to small changes in realized volatilities. In contrast, the higher discount factor $\delta = 0.99$ induces a much greater degree of smoothing of the volatility trajectories, and is likely under-adaptive in time of really marked change, such as towards the end of 1992 when Britain withdrew from the EU exchange rate agreement, resulting in marked swings and increased volatility in the European currencies across the board. The impact of this event is evident in the estimated trajectories of both the marginal variances of currencies and in the corresponding variances of the factors arising from the direct

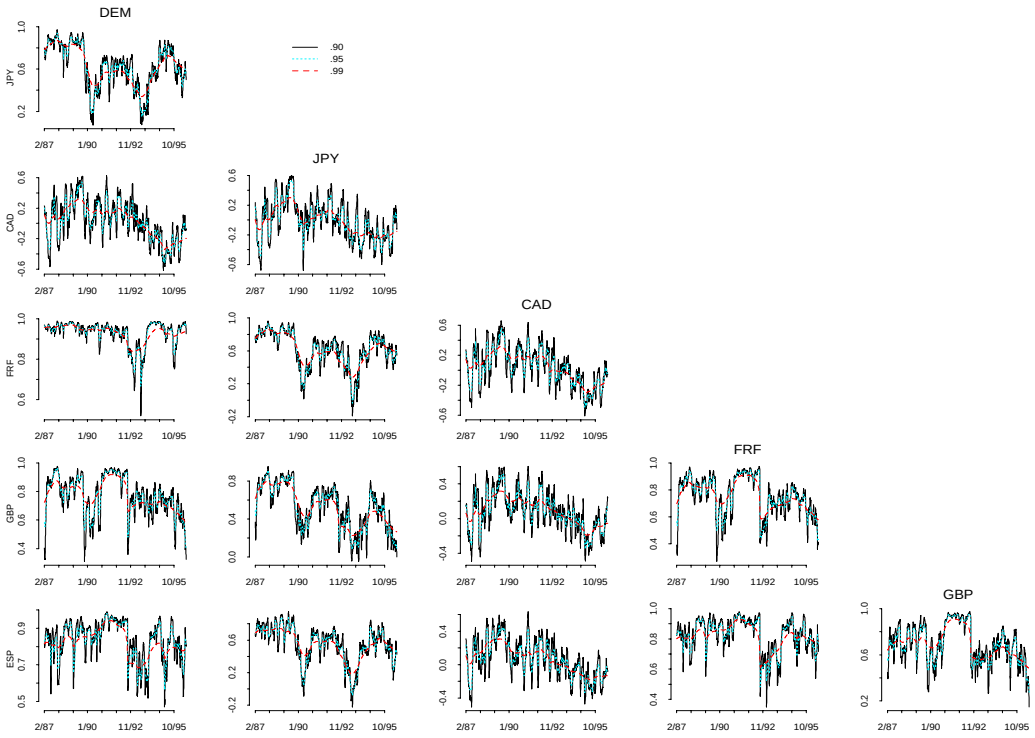

Figure 8.4: Estimated correlations from discount analyses.

principal components decompositions in Figure 8.5.

Notice that the end-1992 volatility changes impact across all factors, highlighting the apparent dependencies in factor trajectories across the entire time period. This indicates the need for dependence structure in modeling latent volatility processes in dynamic factor analyses, as is allowed in the theoretical framework described in chapter 7 and investigated in factor model data analyses below. Such dynamic principal component analyses can be seen as providing informal, exploratory views of possible latent factor structure, albeit conditioned on the mathematically convenient but practically questionable orthogonality constraints. It appears that at most three factors are necessary, which is anticipated as the currencies represent three distinct trading blocs: Canada, Japan and the EU. The trajectories of the three minor eigenvalues remain at consistently negligible levels across the time frame here, so that a model with three factors plus currency-specific random effects is indicated.

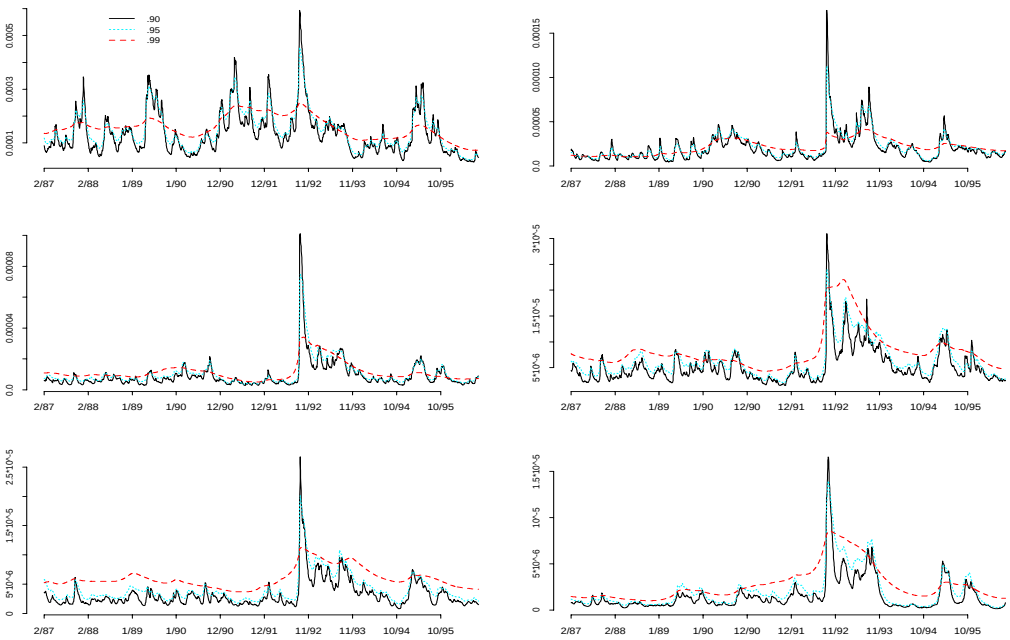

Figure 8.5: Factor variances from discount analyses.

8.2.2 Dynamic Factor Analysis

A dynamic model as defined in (7.1) was used for the $q = 6$ series and with $k = 3$ factors. Note that in order to have a model that provides a maximal specification under the assumed structure of the factor loadings matrix (6.13), and assuming each of the ψ_j to be non-zero, the number of factors must necessarily be no greater than three. Hence, in addition to being suggested by the discounting analyses, this serves here as an encompassing model; if fewer than three factors are supported by the data, that fact will be reflected in posterior inferences about factor loadings and variances.

An appropriate, informative prior for the key matrix $\mathbf{U}$ in the volatility model was based on a pre-analysis of the initial 200 observations on the time series, reserving these few observations for this alone and then analyzing the remaining data with the factor model. From the Bayesian discount model with a discount factor of 0.9, the point estimates of the three dominant eigenvalues of each $\mathbf{S}_{t,n}$ were extracted and used as ad-hoc estimates of the factor volatilities h_{jt}, for each j, t. Three separate

AR(1) models where then fitted to the log-volatilities so computed, using standard reference Bayesian analyses. This provided posteriors for the AR parameters and innovations variances, in each volatility series marginally, that are taken as "ball-park" initial estimates to be used to specify an informative prior for $\mathbf{U}$ prior to analysis of the remaining data. This preliminary analysis gave approximate prior means of the three innovations variances around 0.001–0.002. With this in mind, the prior for $\mathbf{U}$ in the factor model analysis was chosen to be inverse Wishart $W_{r_0}^{-1}(\mathbf{U}|\mathbf{R}_0)$ with $r_0 = 100$ degrees of freedom (half the prior sample size in the ad-hoc analysis) and $\mathbf{R}_0 = 0.0015\mathbf{I}$, appropriately "centering" the prior for $\mathbf{U}$. Note that the prior does not anticipate correlations across volatility processes, though this could easily be done. Independent diffuse uniform priors for the idyosincratic variances $\mathbf{\Psi}$ were also used to obtain proper posteriors as explained in chapter 6.

The MCMC analysis of this factor model involved a range of experiments with Monte Carlo sample sizes and starting values, and MCMC diagnostics. The summary numerical and graphical inferences are based on over 20,000 simulations of posteriors, generated following a 5,000 burn-in period. A set of 1,000 spaced 20 apart was sub-sampled so as to break correlations and record resulting samples for graphical display purposes. Summary graphs appear in Figures 8.6 to 8.15 inclusive. First, Figure 8.6 graphs estimated trajectories of conditional variances of the currencies – the posterior means of the diagonal elements of $\mathbf{\Sigma}_t = \mathbf{X}\mathbf{H}_t\mathbf{X} + \mathbf{\Psi}$. Note the similarity with the trajectories from the more adaptive of the discount analyses in Figure 8.3, as is to be expected. Analogously, Figure 8.7 plots the corresponding estimated correlations between the series, showing again similar patterns to the low discount model in Figure 8.4.

Figure 8.8 provides histogram approximations to marginal posteriors for the elements of the loadings matrix $\mathbf{X}$. Note that the first column (apart from the fixed

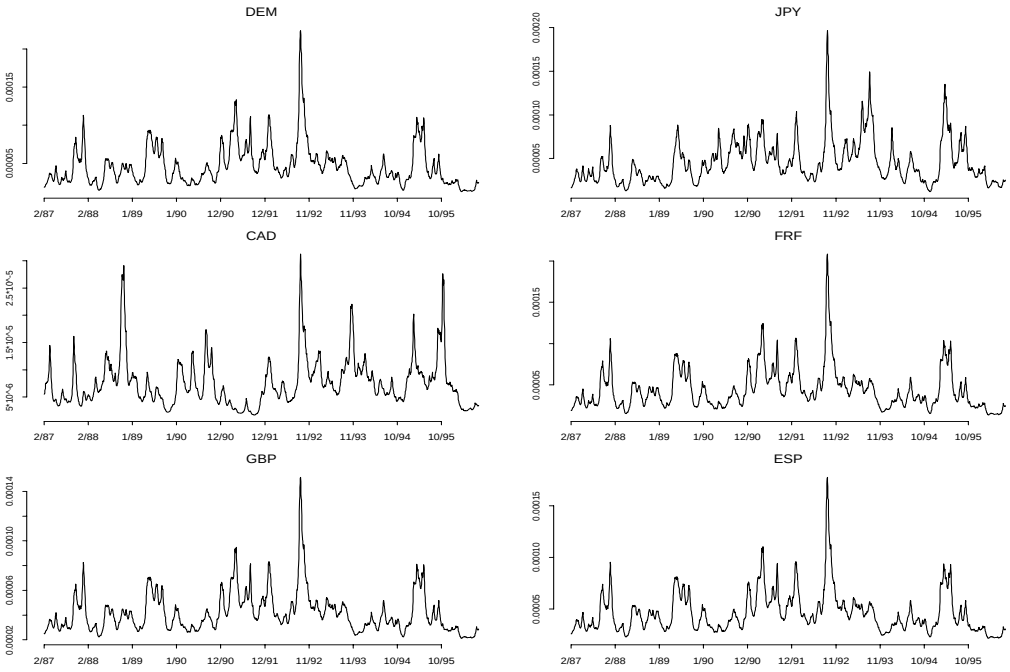

Figure 8.6: Conditional variances from dynamic factor model.

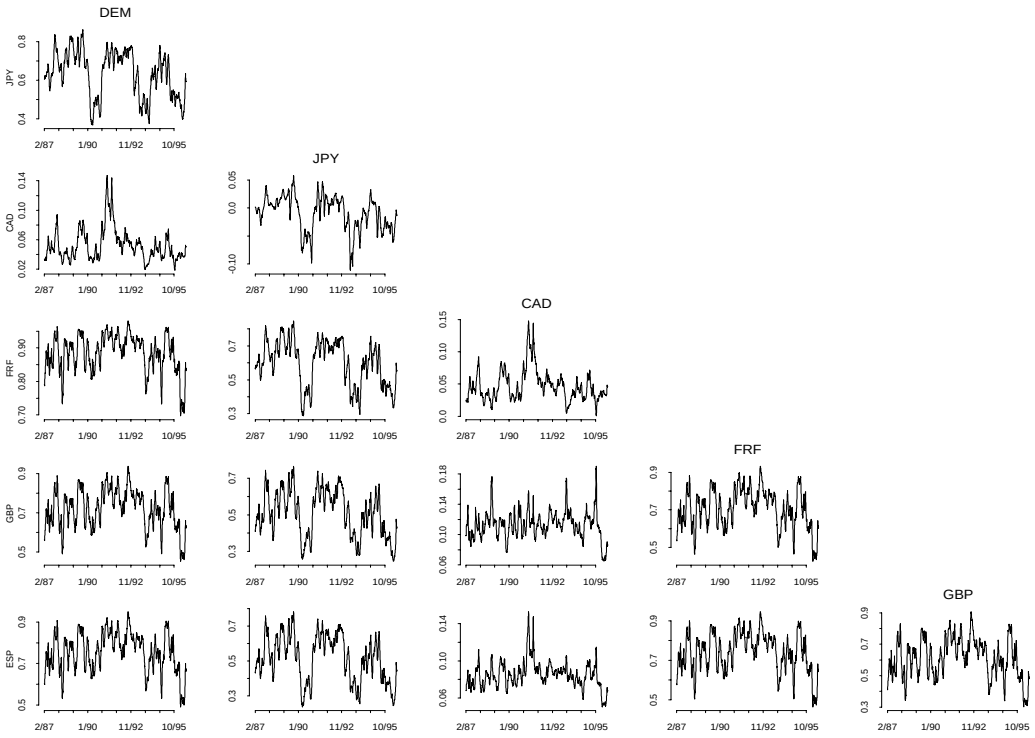

Figure 8.7: Estimated correlations from dynamic factor model.

lead element) gives positive weight to all but CAD, representing the relative strength of the US dollar to the currencies of the EU and Japan. The CAD has almost no weight here, as to be expected as its value in international markets is most strongly determined by the US dollar alone, and the relative values of the weights on the EU countries naturally reflect their relative strengths. The loadings on the second factor are very small and, if non-negligible, negative, apart from Japan with the fixed unit weight. This is therefore largely the Japan:US factor, with some residual contrast between Japan and the rest of the currencies. Similar comments applies to the third factor which essentially represents the Canadian:US rates, and in which the main residual contrast is that reflecting the differential status of Britain to the rest of the EU, presumably driven in part by the departure of Britain from the exchange rate control system.

The graphs in Figure 8.9 display the trajectories of approximate posterior means for the three factor processes f_{tj} on the first column and their conditional standard deviations $\sqrt{h_{tj}}$ on the second column. The main points to note here are the appearance of peaks in the volatility processes consistent with positive correlations in volatility across the three factors, and the relative scales of volatility: all three factors are evidently contributing measurably to overall variability in the multiple series, though the factors appear to be roughly ordered in terms of decreasing overall levels. Again, this is consistent with expectations from a substantive viewpoint.

Figures 8.10 and 8.11 summarize marginal posterior inferences for key fixed model parameters, all in boxplot form. Figures 8.10 displays boxplots of posterior margins as follows. The upper left frame displays margins for the elements of the conditional mean $\boldsymbol{\theta}$. The upper right frame displays margins for the diagonal elements $\sqrt{\psi_j}$ of the residual variance matrix $\boldsymbol{\Psi}$ namely the standard deviations of the specific factors. The lower frame displays margins for $100\psi_j/\sigma_{jt}^2$ where σ_{tj}^2 is the $j-$th diagonal element

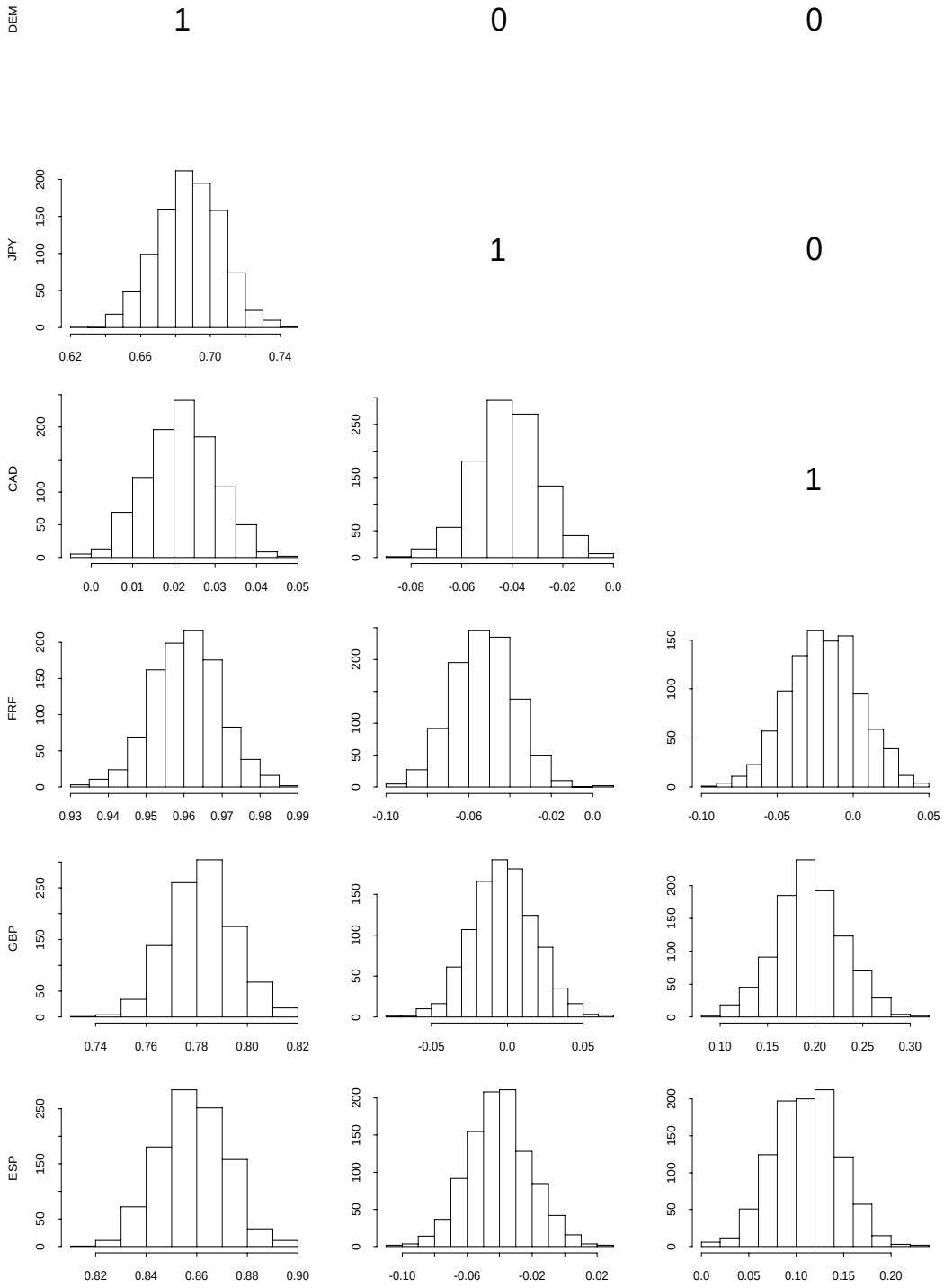

Figure 8.8: Posterior summaries for the factor loadings matrix $\mathbf{X}$.

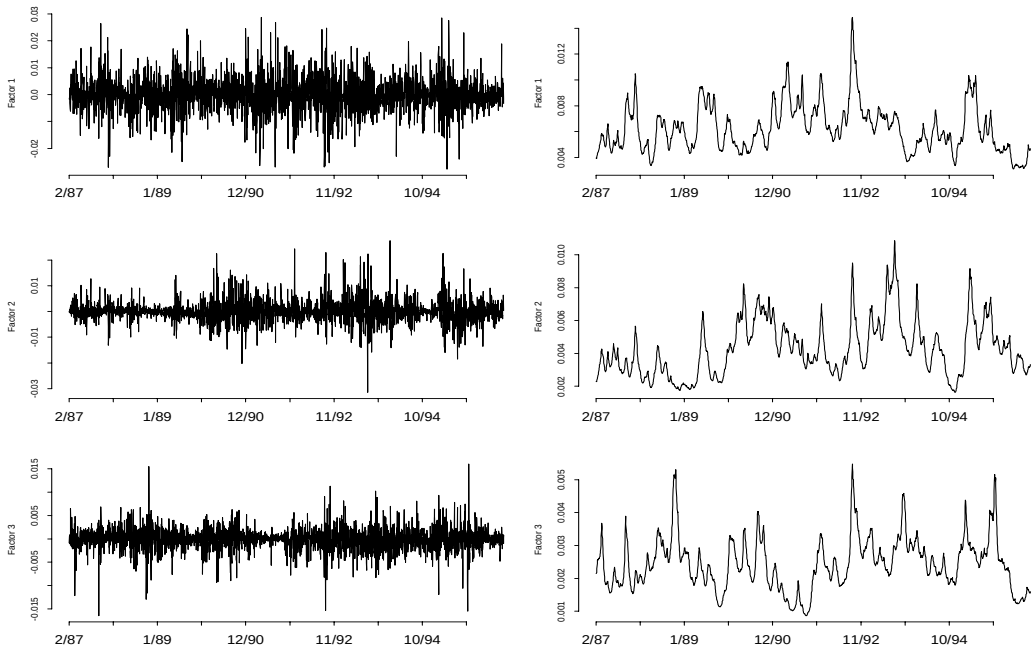

Figure 8.9: Dynamic factor processes and factor standard deviations.

of $\mathbf{\Sigma}_t$. These ratios measure percent total variation in each of the currency series that is contributed by the idiosyncratic terms – generally non-negligible, and appreciable for both GBP and ESP.

Similar displays appear in Figure 8.11 for elements of the parameters $\boldsymbol{\mu}, \mathbf{\Phi}$ and $\mathbf{U}$ in the VAR(1) volatility model, (7.4). Specifically, the upper left frame displays margins for the three scale parameters $\exp(\mu_j/2)$ where the μ_j are the entries of the stationary mean $\boldsymbol{\mu}$ of the VAR(1) model. Converting from the log-volatility to volatility scales, these scale factors $\exp(\mu_j/2)$ represent the standard deviations of the implied stationary distribution, i.e., base levels of conditional variation in the three factor processes. The rough ordering of factors according to marginal variability is clear here. The upper right frame displays margins for the AR parameters ϕ_j in $\boldsymbol{\phi}$, indicating that all three are obviously very close to, but less than, unity, and so representing high persistence in the volatility processes. The approximate posterior means for the three ϕ_j are 0.97, 0.98 and 0.98, respectively. The lower frame displays

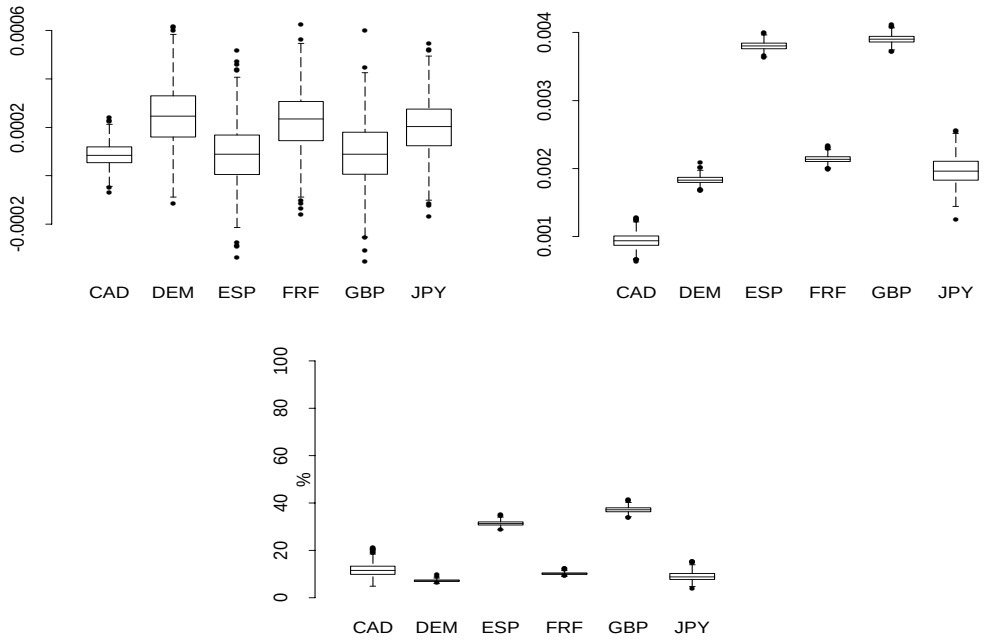

Figure 8.10: Posterior summaries for $\boldsymbol{\theta}$ and $\boldsymbol{\Psi}$.

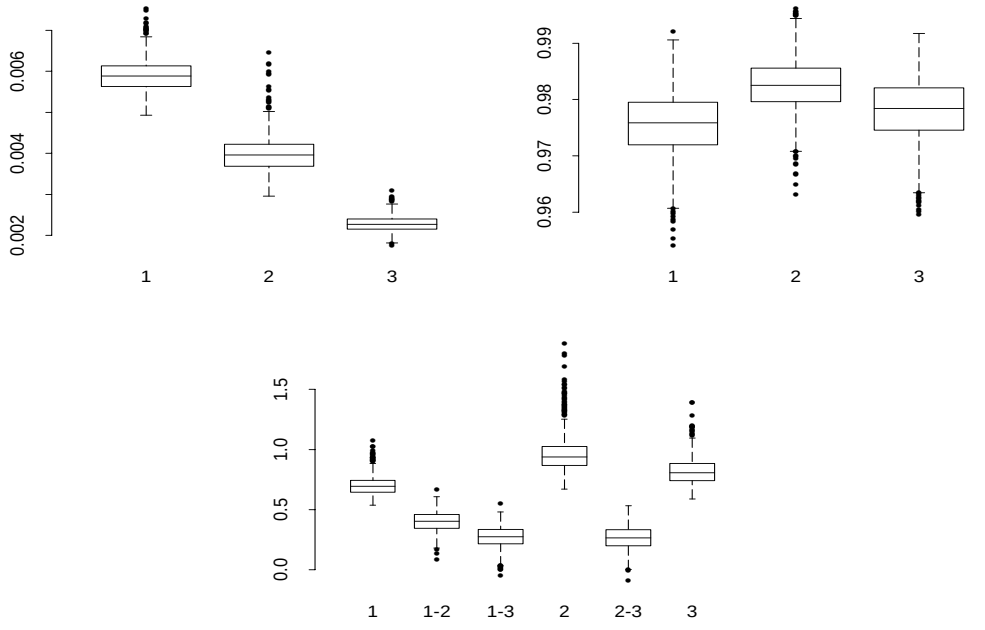

Figure 8.11: Posterior summaries for $\boldsymbol{\mu}$, $\boldsymbol{\Phi}$ and $\mathbf{U}$.

margins for the standard deviations and correlations of matrix $\mathbf{W}$, the marginal variance matrix in the VAR(1) volatility model. The earlier noted positive correlations between factor processes are indicated here. Related numerical summaries provide approximate posterior means of the variances in $\mathbf{W}$ as 0.50, 0.86 and 0.82, respectively, while the corresponding posterior means for the variances in the innovations matrix $\mathbf{U}$ are 0.027, 0.034 and 0.025, respectively.

Figures 8.12, 8.13, 8.14 and 8.15 display trace plots, autocorrelation functions and histograms of marginal distributions of selected parameters to asses convergence of the Markov Chain. For instance, each one of the columns of Figure 8.12 provides, in order, information on the autoregressive parameters ϕ_j, the scale factor $\exp(\mu_j/2)$ and the marginal variances $\mathbf{W}$. The plot is divided in three groups of frames corresponding to the three latent factors. Next, Figure 8.13 plots the trace plots of the loadings matrix $\mathbf{X}$ and Figures 8.14 and 8.15 graph the corresponding plots for the conditional mean vector $\boldsymbol{\alpha}$ and the specific variances $\boldsymbol{\Psi}$ respectively. As can be seen from these plots, it seems that the Markov Chain converged to the true posterior distribution.

Reanalyses under ranges of diffuse but proper priors were performed to assess sensitivity. Across several choices of seemingly uninformative though proper priors, the posterior results confirm those from the analyses above.

The results from discount analyses and the dynamic factor model can be further used in the decision making process, for instance in deciding on a fair option price and/or the allocation of investments within a portfolio. As a matter of fact, the time-varying conditional variances of each one of the variances could be used to price derivate products on individual currencies with stochastic volatility following theory suggested in Hull and White (1987). Moreover, the changes in the variance-covariance matrix can have important consequences in the process of allocating assets dynamically. That is, ideally at each time point an investor will allocate resources such that

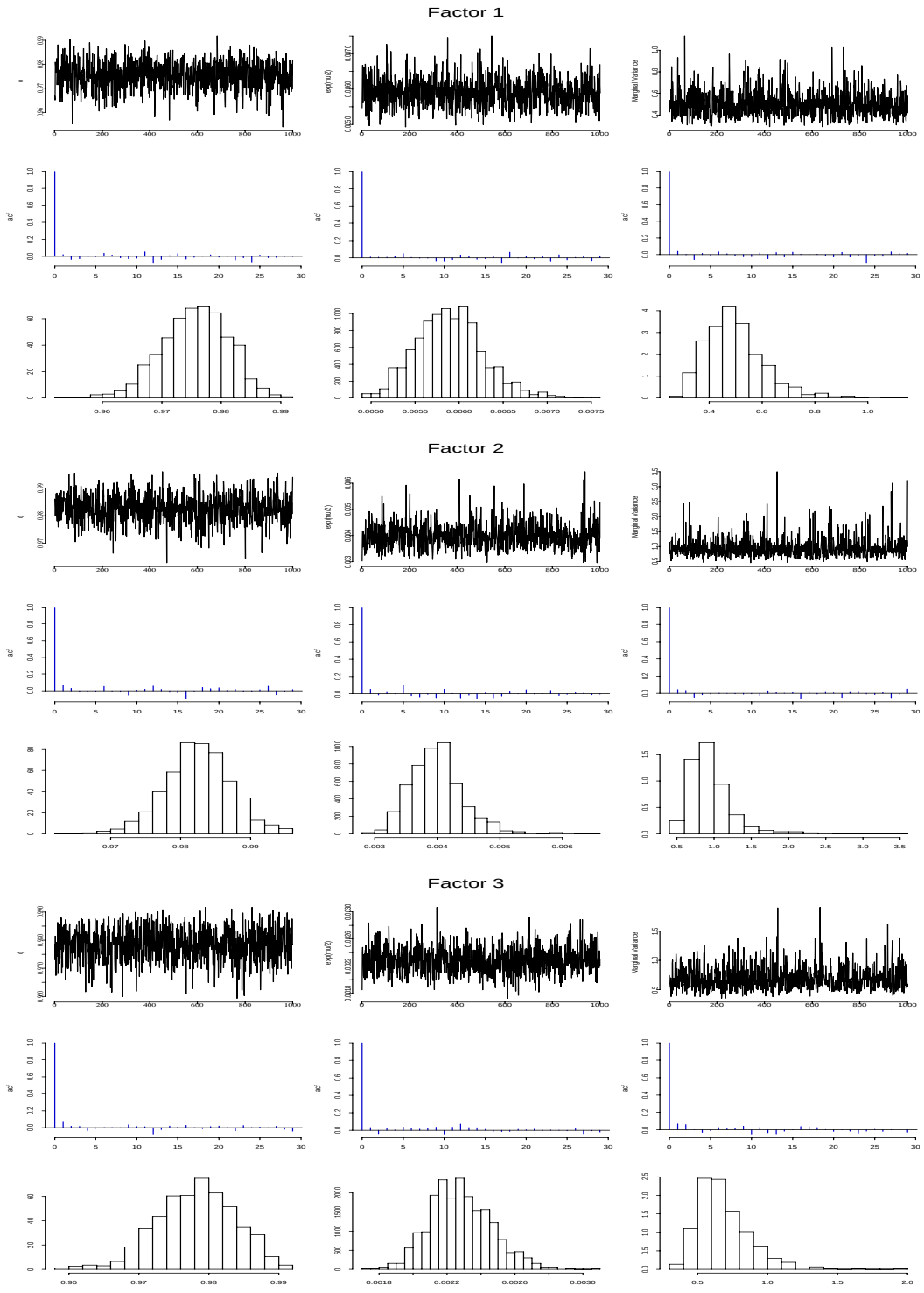

Figure 8.12: Trace plots, acf and histograms of some stochastic volatility model parameters: ϕ_j, $\exp(\mu_j/2)$ and $\mathbf{W}$.

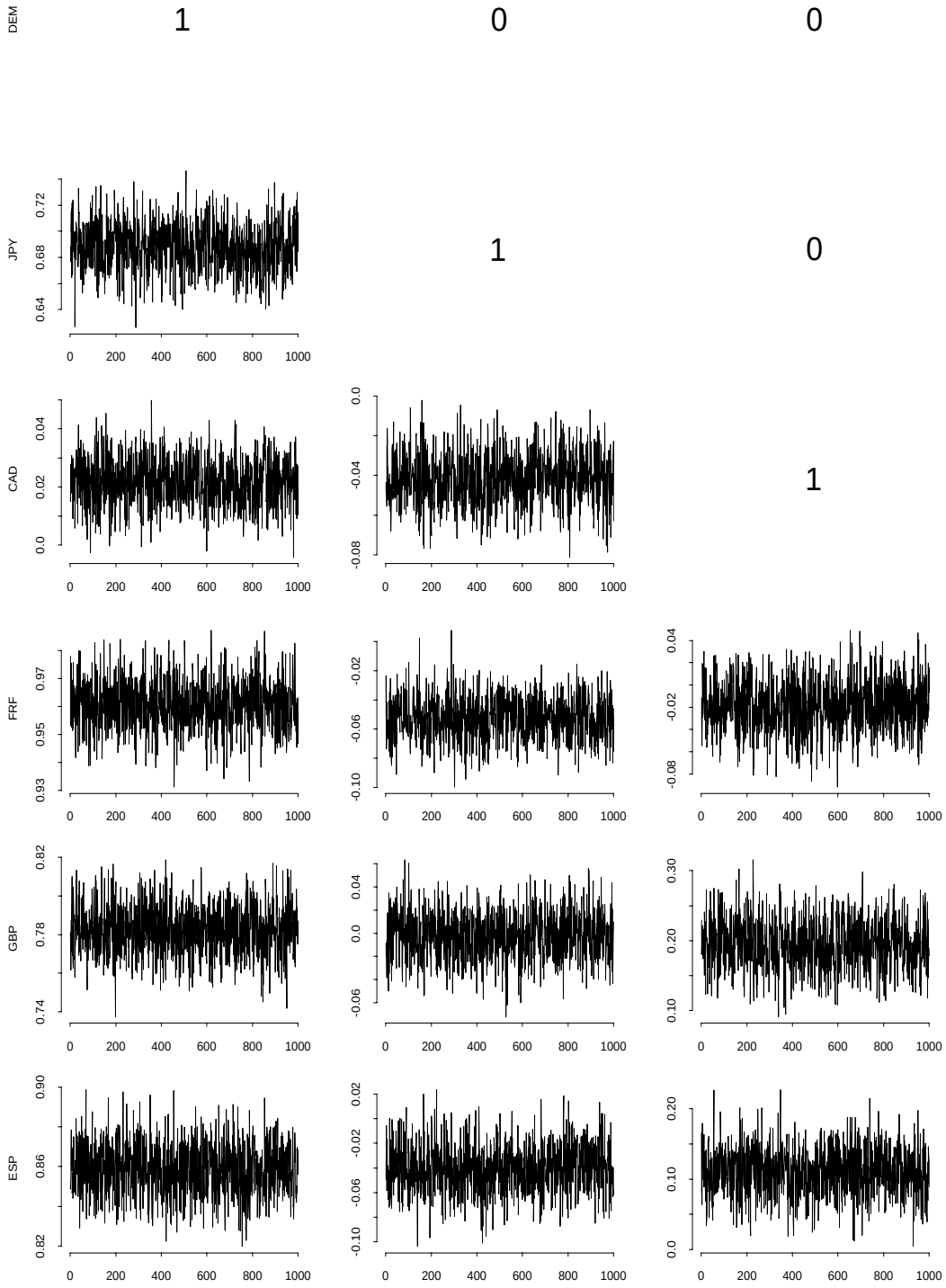

Figure 8.13: Trace plots for the factor loadings matrix $\mathbf{X}$.

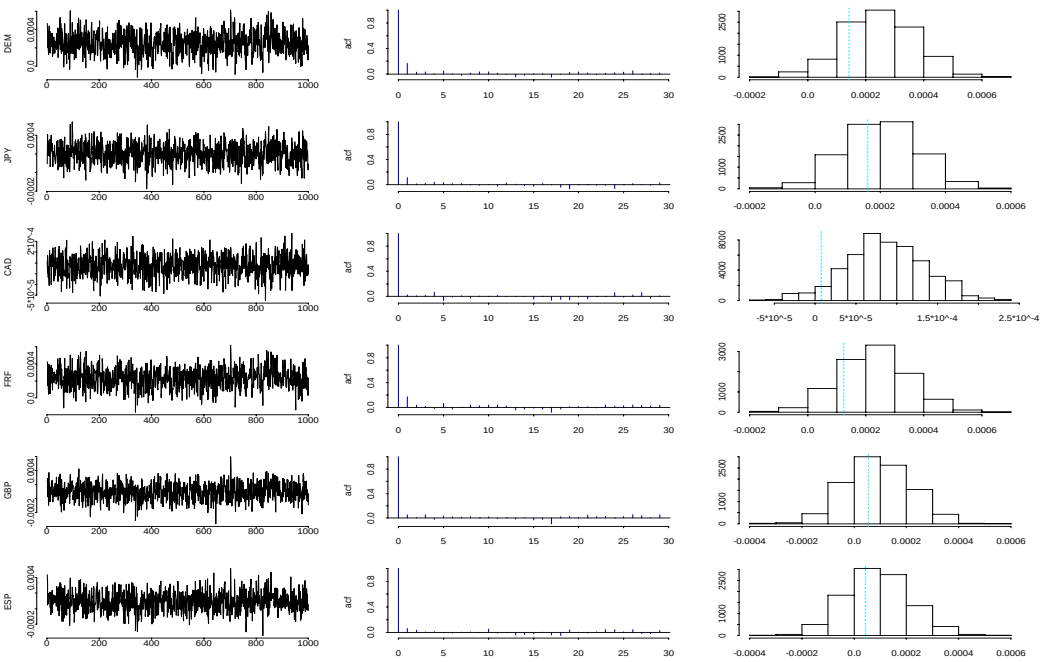

Figure 8.14: Posteriors for the mean parameter vector $\boldsymbol{\theta}$.

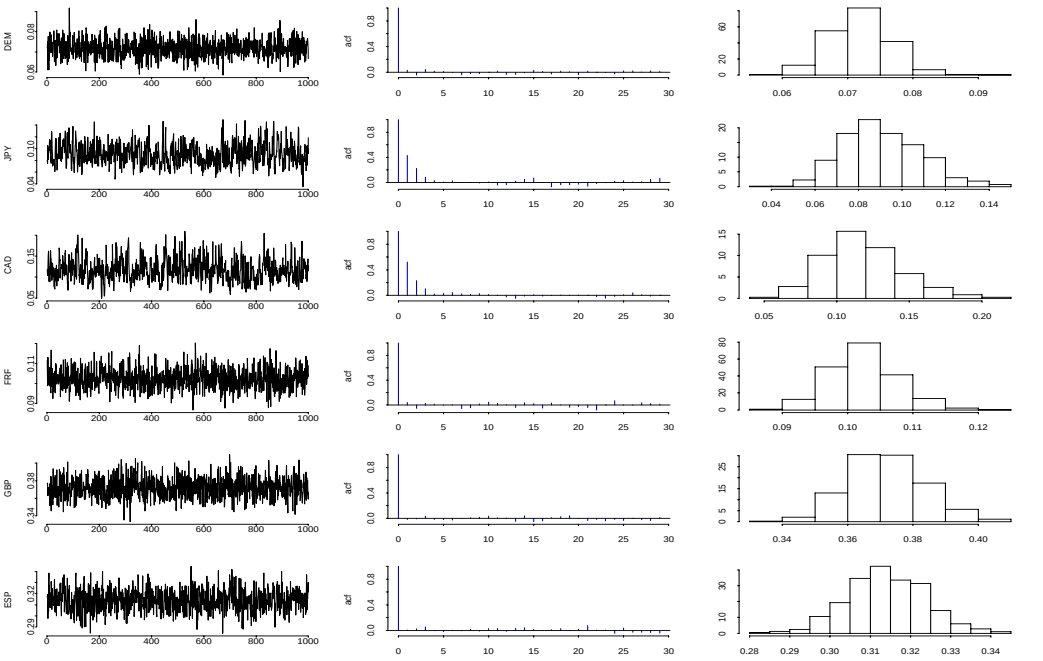

Figure 8.15: Posteriors for the specific variances ψ_i.

the expected return will be maximum with the minimum risk. In addition, the resulting portfolios and their corresponding returns can be used in a model comparison framework between the discount analyses and the dynamic factor solution.

8.3 Bayesian Portfolio Construction

Optimal portfolio selection has been of interest for academics and practitioners for a long time. The standard approach is to find the optimal distribution of resources by minimizing the risk for a given level of expected return. The major theoretical work to generate optimal portfolios through a quadratic programming procedure is due to Markowitz (1959). This theory involves generally a mean parameter and a covariance matrix of the returns which are usually estimated based on historical data. From the Bayesian point of view, Quintana (1992), Putnam and Quintana (1994a) and Quintana and Putnam (1996) implemented portfolio rules to problems with fixed income securities and futures contracts in currency and equity index markets respectively. They use the predictive variance-covariance matrices calculated with discount methods to find the optimal portfolios. The methodology here is based on the existing literature and is developed in the context of the international exchange rates application.

Let $\mathbf{y}_t$ be the $q-$dimensional vector of rates of returns on q currencies at time t. The dynamic asset allocation theory assumes that at each time point $t-1$, an existing investment in the various currencies under study may be reallocated according to a portfolio $\mathbf{a}_t$ for the next time point. The elements of $\mathbf{a}_t$ are the \$US amounts invested in the corresponding currency. For the purpose of this thesis, no transaction costs are assumed and the dollars may be reallocated freely and instantaneously to arbitrary long or short positions across the currencies, subject initially only to $\mathbf{a}_t'\mathbf{1} = 1$. The realized portfolio return at time t is the \$US amount $r_t = \mathbf{a}_t'\mathbf{y}_t$, and models may

be compared on the basis of cumulative returns over chosen time intervals. The portfolio allocation decision problem involves the general Markowitz mean-variance optimization, and this is applied at each time point one-step ahead. In all models, the time t situation is summarized through posterior one-step ahead means and variance matrices for $\mathbf{y}_t$, denoted here by $\mathbf{g}_t$ and $\mathbf{G}_t$. These quantities can be calculated as the mean and variance of the one-step-ahead predictive density of the returns at time t as described in equations (7.14) and (7.15) from section 7.5.

The decision context targets minimization of the one-step ahead variance of returns subject to specified target means. For a specified target mean return m, the one-step-ahead portfolio $\mathbf{a}_t$ will be chosen to minimize the one-step ahead variance of returns $\mathbf{a}_t'\mathbf{G}_t\mathbf{a}_t$ subject to constraints $\mathbf{a}_t'\mathbf{1} = 1$ and $\mathbf{a}_t'\mathbf{g}_t = m$. The so called *efficient* frontier is the set of portfolios that have maximal expected return given an upper bound on the variance. It can be shown that the *efficient* frontier is the solution $\mathbf{a}_t^{(m)}$ that minimizes $\frac{1}{2}\mathbf{a}_t'\mathbf{G}_t\mathbf{a}_t$ over the constraint set, (Markowitz, 1959). The well-known solution to this quadratic programming problem through Lagrange multipliers is given the **mean-variance efficient** portfolio,

$$\mathbf{a}_t^{(m)} = \mathbf{G}_t^{-1}(a\mathbf{g}_t + b\mathbf{1})$$

where

$$a = \mathbf{1}'\mathbf{G}_t^{-1}\mathbf{e} \quad \text{and} \quad b = -\mathbf{g}_t'\mathbf{G}_t^{-1}\mathbf{e}$$

with

$$\mathbf{e} = (\mathbf{1}m - \mathbf{g}_t)/d \quad \text{and} \quad d = (\mathbf{1}'\mathbf{G}_t^{-1}\mathbf{1})(\mathbf{g}_t'\mathbf{G}_t^{-1}\mathbf{g}_t) - (\mathbf{1}'\mathbf{G}_t^{-1}\mathbf{g}_t)^2.$$

Two other standard portfolio allocations can be derived in the same way. First, the so called **target-independent** allocation derived at the boundary of the mean-variance efficient frontier leads to weights

$$\mathbf{a}_t^{(me)} = (\mathbf{1}'\mathbf{G}_t^{-1}\mathbf{g}_t)^{-1}\mathbf{G}_t^{-1}\mathbf{g}_t,$$

and second, the strictly risk-averse **minimum variance** portfolio which just depends only on the estimation of the variance-covariance matrix, namely,

$$\mathbf{a}_t^{(mv)} = (\mathbf{1}'\mathbf{G}_t^{-1}\mathbf{1})^{-1}\mathbf{G}_t^{-1}\mathbf{1}.$$

8.3.1 Special Portfolio Strategies

There are of course different varieties an extensions of the three basic strategies presented above as a result of including or removing constraints. The focus of this is more on comparison of portfolio strategies rather than on comparison of models. For instance, the restriction that $\mathbf{a}_t'\mathbf{1} = 1$ forces the decision maker to constrain his short or long positions in the specified market. However, there are many situations when the optimal decision would be to take your money to the bank in, for example, high-volatility situations. Consider then allocations in which the portfolios are completely unconstrained; that is, remove the unit sum constraint on the allocation vector $\mathbf{a}_t$. This means that the allocation may be chosen without regard to resources, permitting arbitrary long or short positions across the currencies. This typifies the practical working context in the global investments in large financial institutions and is in line with recent work with discount models, (Quintana and Putnam, 1996). The mean-efficient portfolio with expected return target m, under an unconstrained strategy is given by

$$\mathbf{a}_t^{(*m)} = \lambda\mathbf{G}_t^{-1}\mathbf{g}_t$$

where

$$\lambda = m/(\mathbf{g}_t'\mathbf{G}_t^{-1}\mathbf{g}_t).$$

A different situation arises in the implementation of portfolios in higher dimension which usually result in extreme weights on particular assets. A traditional strategy in these cases is to introduce upper and lower constraints in the optimization problem, $l_{it} < a_{it} < u_{it}$. Typical choices of the bounds are $l_{it} = 0$ reflecting the fact that

no short sales are allowed and a constant upper bound $u_{it} = u_0$. Consequently, the higher level of u_0 the more aggressive the portfolio in the sense of the few number of securities held and the higher tracking error of the portfolio. The focus in this thesis is on the three basic portfolio strategies, mean-variance efficient, target independent and minimum variance with traditional assumptions and restrictions. These strategies will be used in the next section mainly for model comparisons in the international exchange rates example.

8.3.2 Dynamic Portfolio Comparisons for Exchange Rates

Model comparisons are made with explicit focus on one-step forecast accuracy in the context of dynamic portfolio allocations, essentially following the perspective of Quintana (1992), Putnam and Quintana (1994a) and Quintana and Putnam (1996). A similar perspective is adopted in Polson and Tew (1997) though with very different models. The comparisons here are based on posterior distributions from the models fitted to the entire time series, so that they do not represent real-time, sequential forecasts, but nevertheless do provide a coherent basis for model comparisons with a utility function directly measuring real-world performance in terms of cumulative financial return.

For comparisons, Figure 8.16 graphs the trajectories of the constrained optimal mean-variance efficient portfolio weights $\mathbf{a}_t^{(me)}$ from the three Bayesian discount analyses and the dynamic factor model analysis reported above. The trajectories of the weights $\mathbf{a}_t^{(*m)}$ for $m = 0.00016$ under the factor model alone are displayed in Figure 8.17, for clarity.

The weigths are graphed as percentages, i.e., simply 100 times their actual values, as they are constrained to sum to unity. Figure 8.18 graphs the trajectories of cumulative returns over time based on these four models in the frames in the first two rows.

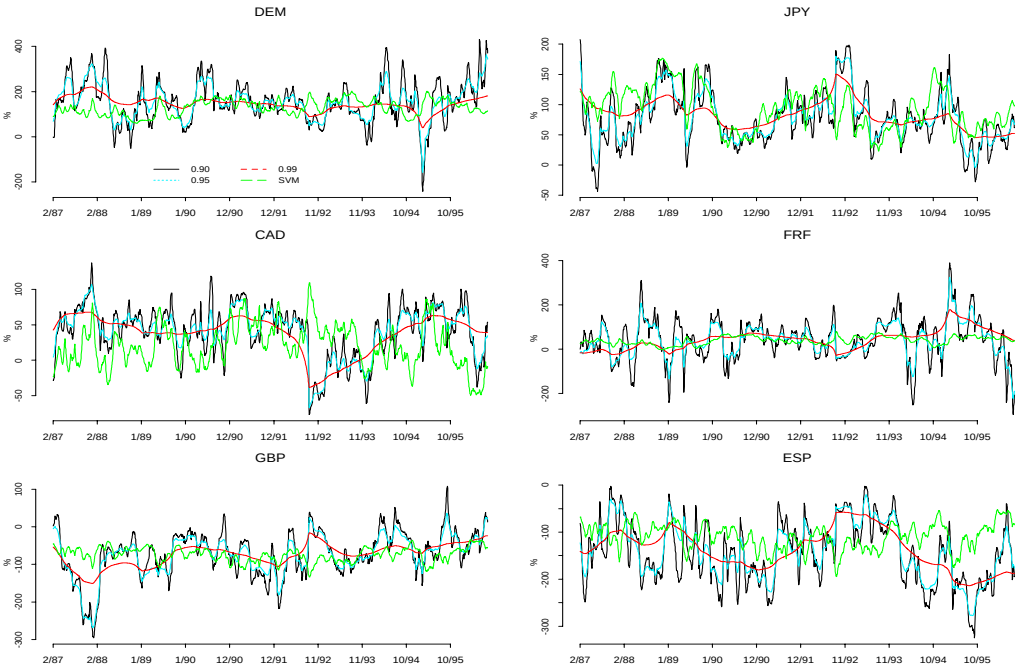

Figure 8.16: Dynamic weights $\mathbf{a}_t^{(me)}$ for the mean-efficient portfolio: all models.

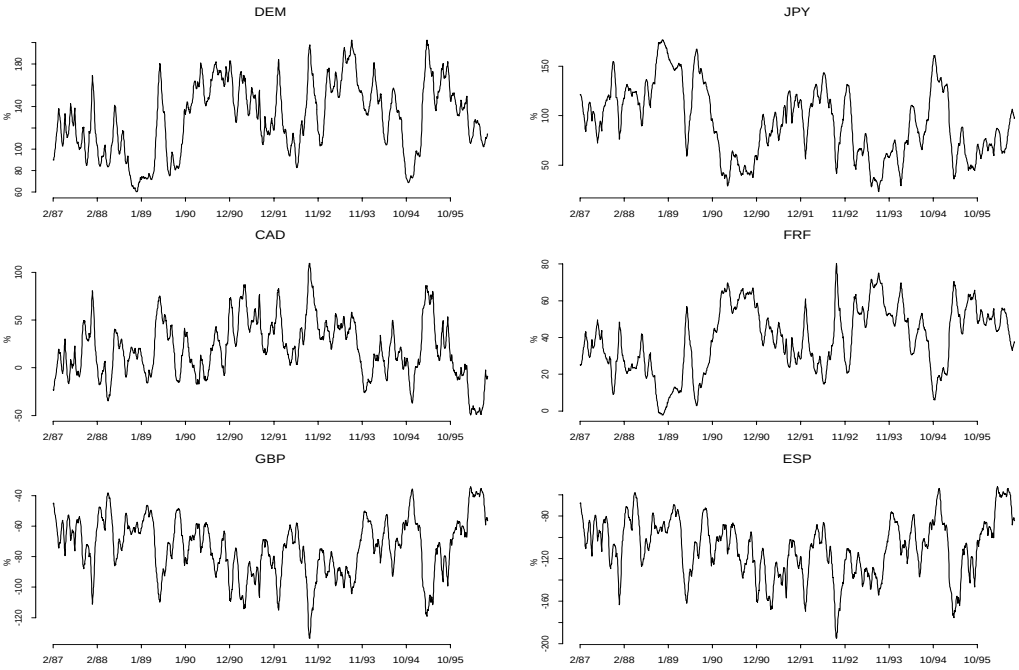

Figure 8.17: Dynamic weights $\mathbf{a}_t^{(*m)}$ for $m = 0.00016$ in the unit-sum, contrained portfolios: factor model.

These four frames correspond to four different fixed investment strategies: the two risk-averse strategies $\mathbf{a}_t^{(mv)}$ and $\mathbf{a}_t^{(me)}$, and then two strategies $\mathbf{a}_t^{(m)}$ with daily target returns $m = 0.00016$ and $m = 0.00028$ respectively. These appear in the order (top left) $\mathbf{a}_t^{(mv)}$, (top right) $\mathbf{a}_t^{(me)}$, (center left) $\mathbf{a}_t^{(0.00016)}$ and (center right) $\mathbf{a}_t^{(0.00028)}$. After a period in which the model behave very similarly in the determination of portfolio allocations, the changes in volatility in 1992 are more appropriately captured by the most adaptive discount model ($\delta = 0.9$) and the dynamic factor model. These two models proceed to clearly dominate the others in terms of cumulative returns under the practically relevant strategies. It is interesting to note that, though the very adaptive discount model produces cumulative returns that closely shadow the factor model, the weights in the factor model are relatively much more stable over time. The discount model adapts the weights quite widely as it permits very marked patterns of change in the full variance matrix of returns, whereas the factor model assigns changes in observed volatilities to appropriate model components and so induces more stability in weight trajectories.

In order to compare portfolio strategies, unconstrained mean-variance efficient portfolios were also considered. The third row of graphs in Figure 8.18 displays the cumulative return trajectories using the optimal portfolio weights $\mathbf{a}_t^{(*m)}$ from this strategy, to be compared to the earlier figures using constrained portfolios. The two graphs provides display under portfolios (lower left) $\mathbf{a}_t^{(*0.00016)}$ and (lower right) $\mathbf{a}_t^{(*0.00028)}$. As can be seen now, the dynamic portfolio model is very clearly dominant, achieving cumulative returns that are about twice as large as those of the most competitive discount model, and also exceeding those of the constrained allocations. The response following the major structural changes in volatility at the end of 1992 leads to a marked swing in portfolio structure that the unconstrained allocations capitalize in a major way in the factor model, with a persistent effect on cumulative

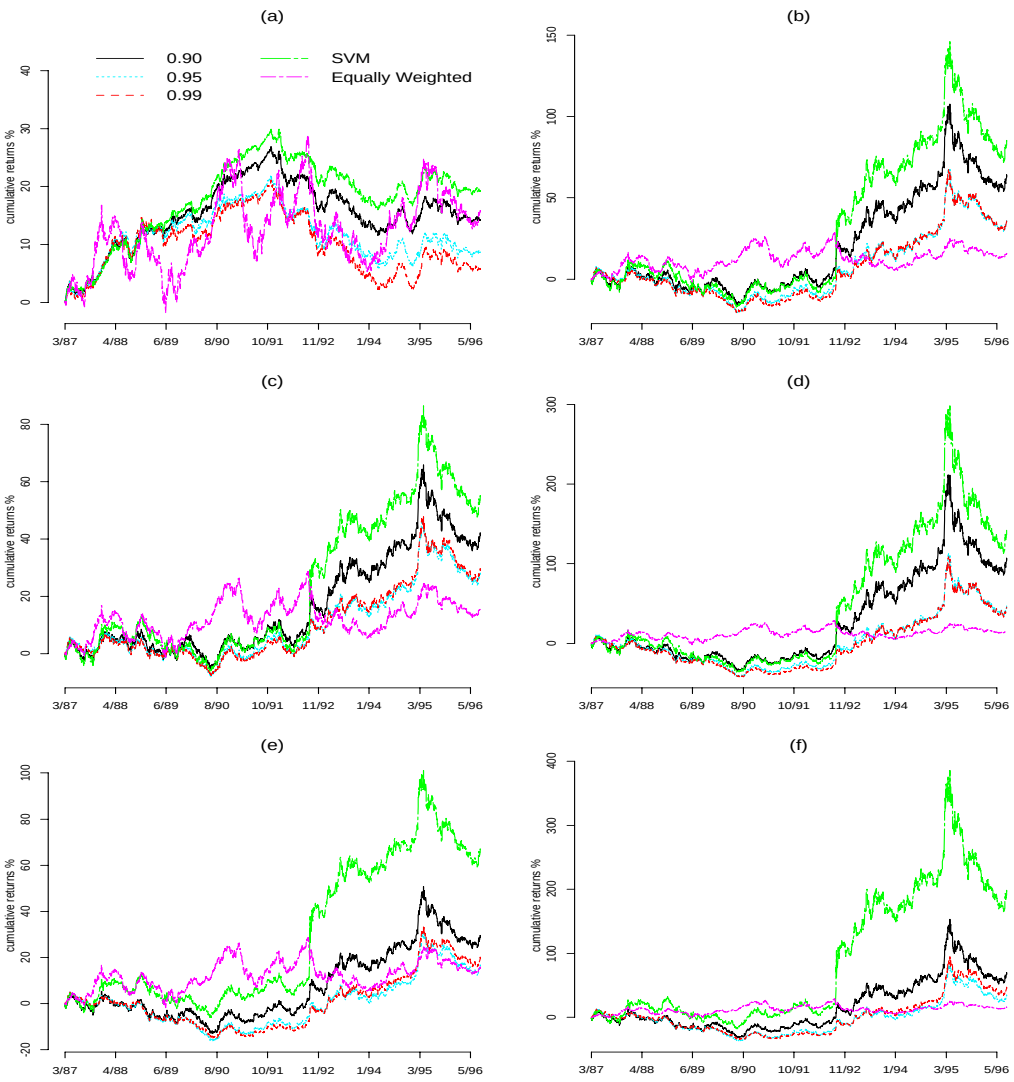

Figure 8.18: Cumulative returns under different dynamic portfolios.

returns thereafter.

The trajectories of optimal, uncostrained portfolio weights for the case $m = 0.00016$ under the dynamic factor model are graphed in Figure 8.19. The values plotted are 100 times the actual weights divided by the total $\mathbf{1}'\mathbf{a}_t^{(*m)}$ at each time point, indicating the relative weight of each currency in the portfolio at each time. This provides a direct comparison with corresponding trajectories of weights from the unit-sum constrained allocation appearing in Figure 8.17, where the actual and

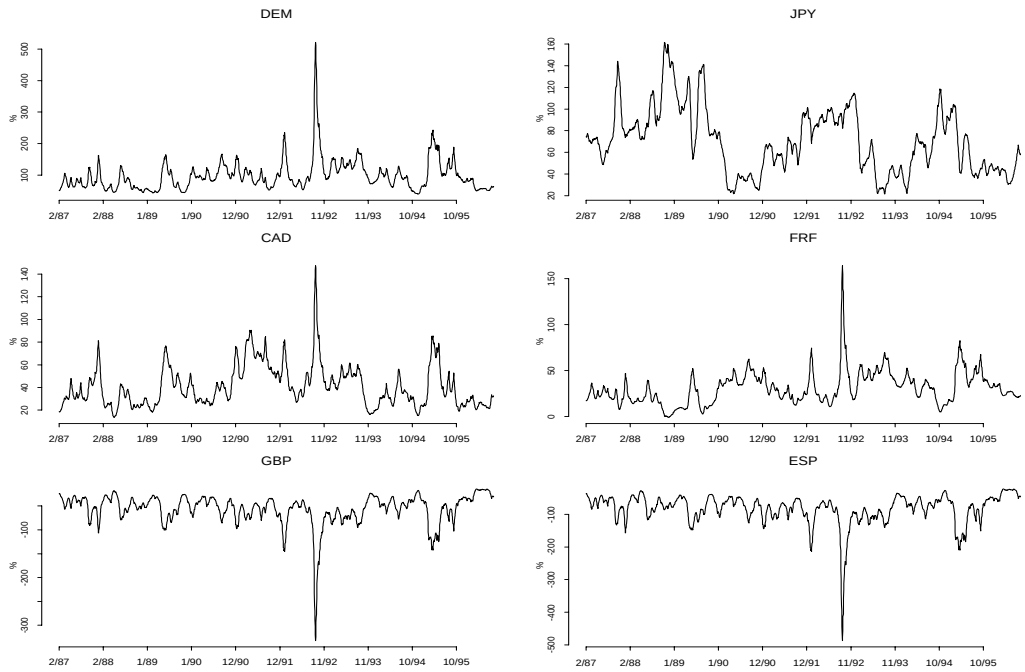

Figure 8.19: Dynamic weights $\mathbf{a}_t^{(*m)}$ with $m = 0.00016$ in the unconstrained portfolios: factor model. The weights are graphed here as percentages of the totals $\mathbf{1}'\mathbf{a}_t^{(*m)}$ at each time.

relative values coincide. The impact of the Britain's withdrawal from the EU exchange rate agreement, in late 1992, and the resulting marked increase in volatility and the portfolio's response are very clear in Figures 8.18 and 8.19. The allocations shift swiftly and quite radically in extent to short positions on Sterling and the strongly associate Peseta, while simultaneously adopting radically long positions on the strong Mark and Yen. At the same time, the total investment dropped markedly; the major changes in volatility led to the anticipation of high levels of increased risk, and the total $\mathbf{1}'\mathbf{a}_t^{(*m)}$ invested in the marked decreased radically as a result. Figure 8.20 graphs the time trajectory of the total $\mathbf{1}'\mathbf{a}_t^{(*m)}$, indicating relative stability in the fluctuations around a level of unity, but with marked swings up and down in periods of low and high volatility, respectively.

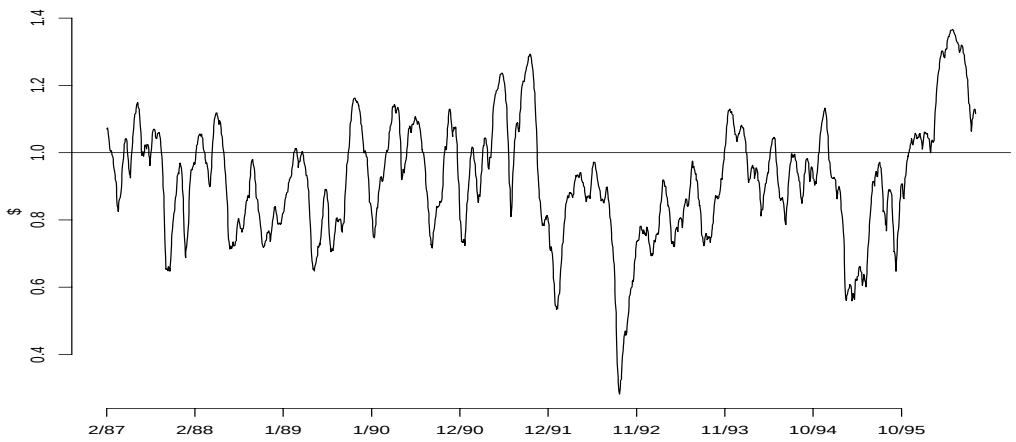

Figure 8.20: Total investments $\mathbf{1}'\mathbf{a}_t^{(*m)}$ with $m = 0.00016$ in the unconstrained potfolios: factor model.

8.4 Conclusions

The investigations indicate the feasibility of formal Bayesian analysis of structured dynamic factor models. The analysis is accessible computationally with nowadays moderate computational resources, and the empirical studies suggest that the analysis will be manageable with 20-30 dimensional time series and several factors. More extensive applications in short-term forecasting and on-line portfolio allocations with higher dimensional models for longer-term exchange rate futures are currently investigated. The example here is suggestive of potential benefits, and supportive of the view that exploiting systematic volatility patterns via factor structuring may yield meaningful improvements in short-term forecasting and decision making in dynamic portfolio allocation, especially in the unconstrained portfolio optimizations as illustrated in the final row of graphs in Figure 8.18. In the case of constrained portfolio optimizations, the over-adaptive discounting method almost match the factor model analysis in terms of portfolio allocation performance in some cases, though is clearly eventually dominated in terms of cumulative return trajectories by the factor model. The conjecture is that, in studies of forecasting and portfolio allocation with

longer term horizons, such as 30-day exchange rate futures, and in extended models that incorporate dynamic regression components, the factor modeling approach will clearly dominate discounting methods. This is the subject of current and near-future research.

The dynamic factor models illustrated are amenable to direct implementation using our customized MCMC methods with the minimal/reference prior specifications used here. The use of variance discounting methods on a reserved initial section of the data to provide input to informative priors is important in identifying "ball-park" scales for the $\mathbf{U}$ matrix of the VAR(1) SV model. Though not pursued here, other aspects of such preliminary analyses may be used to determine informative priors for other elements of the factor model. The established discounting methods are, relative to dynamic factor models, trivial to implement in the current context, a fact that is important in using discount methods to specify partial prior structure in the dynamic models. The empirical findings here indicate that, not surprisingly with this kind of data, moderately adaptive discount methods fare well in time of slow change in volatility levels and patterns, but are relatively uncompetitive in cases of more marked structural change. This is to be expected. Looking ahead, models and approaches that attempt to simplify the process of factor modeling, perhaps somehow integrating elements and concepts of variance matrix discounting into a specified factor structure, may be attractive from a computational/implementation viewpoint.

8.5 Further Extensions and Current Research

In the factor model context as developed in chapter 7, there are several relevant technical and modeling issues to be explored. For instance, further study and empirical assessments on time series with larger numbers of univariate components and larger

numbers of factors are under investigation. Specifically, extending the exchange rates example above, a data set with spot prices from $q = 12$ currencies is currently analyzed using 5 latent factors. The currencies are, in order, the Deutschmark Mark (DEM), Japanese Yen (JPY), Canadian Dollar (CAD), New Zealand Dollar (NZD), British Pound (GBP), French Franc (FRF), Spanish Peseta (ESP), Australian Dollar (AUD), Belgium Franc (BEF), Netherlands Guilder (NLG), Sweden Krone (SEK) and Switzerland Franc (CHF). The first five series were chosen to represent economies of five different blocs. The latent factors and their corresponding estimated standard deviations are plotted for illustration in Figure 8.21. One important point to mention here is the fact that by increasing the number of factors in the model, the effects from the CAD and JPY factors disappear. This is natural given the fact that the EU countries have higher correlations between them and hence the factor model identifies factors that are more helpful in explaining the covariances in the observations. These facts are interesting from the model uncertainty and model choice points of view by addressing questions of choices about the numbers of factors, and about the ordering of time series in the context of the specific structure adopted for factor loadings.

Heavy-Tailed Distributions

There has been plenty of studies that reinforce the idea that the conditional distribution of the returns is non-normal for almost all financial time series. For example, in model (7.1) the disturbance errors $\boldsymbol{\epsilon}_t$ follow a normal distribution. Following the approach in chapter 5, section 5.1 a scale mixture of normals for $\boldsymbol{\epsilon}_t$ could be assumed by introducing λ_t weights,

$$p(\boldsymbol{\epsilon}_t|\Psi) = \int N(\boldsymbol{\epsilon}_t|\mathbf{0}, \Psi/\lambda_t)p(\lambda_t)d(\lambda_t),$$

where $\lambda_t \sim \text{Gamma}(m/2, m/2)$ giving marginal t-distributions for $\boldsymbol{\epsilon}_t$ with m degrees of freedom. This way of modeling the disturbance errors allows us to make posterior

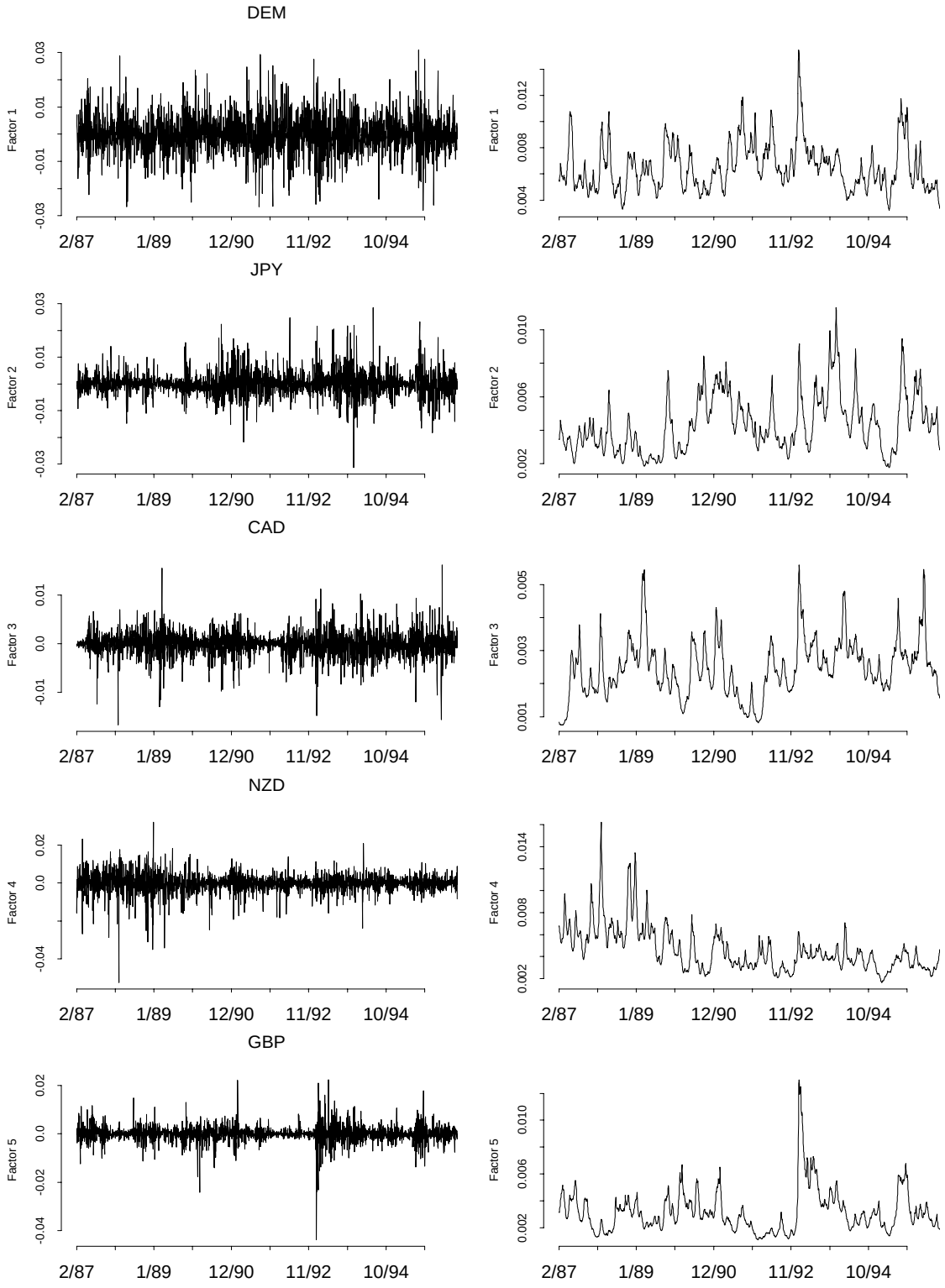

Figure 8.21: Factors and standard deviations in a 12 currencies analysis.

inferences on outliers due to the specific factors $\boldsymbol{\epsilon}_t$. Moreover, this idea can be applied to the distribution of the factors in model (7.1) and/or on the innovations of the log volatility processes in (7.4).

Time Varying Loadings Matrices

Further model extensions under investigation relax the assumptions of constancy of the factor loadings. In this line, Molenaar *et al.* (1992) proposed a dynamic factor model for the analysis of multivariate non stationary time series introducing the idea of lagged factor loadings.

Modeling the Factors

Following the discussion on chapter 5, there has been a lot of studies of the reduction of dimensionality in vector time series. The idea is to identify unobserved common factors with correlation structure that generate an observable vector of time series. The extensions should include factors $\mathbf{f}_t$ with dynamic properties and/or time series models. There are many options of how to model the underlying factors; some examples are:

- $\mathbf{f}_t$ follows a k-dimensional ARMA(p_f,q_f). This assumption will imply that $\mathbf{y}_t$ will also follow a k-dimensional ARMA(p_y,q_y), (Pena and Box, 1987). They assume that the columns of the loadings matrix $\mathbf{X}$ are orthogonal.

- $\mathbf{f}_t$ is a random walk plus noise. This model is called **The common trends** model and is closely related to the **S**eemingly **U**nrelated **T**ime **S**eries **E**quations where the loadings matrix $\mathbf{X}$ has only one non-zero entry per row. Harvey *et al.* (1994) used this model after applying the non-linear transformation to the actual data and taking the volatilities as the common factors. This model has

the property that $q - k$ linear combinations are stationary, even though all the elements of $\mathbf{y}_t$ are only stationary in first differences.

- $\mathbf{f}_t$ follows a k-dimensional time-varying autoregressive model.

In all these cases the DLMs forward filtering backwards sampling algorithm can be used to sample the posterior distribution of the factor scores $\mathbf{f}_t$.

Models for the Conditional Mean

The dynamic factor model (7.1) with stochastic volatility components as presented in chapter 7, could be extended to allow for dynamic regression models on the conditional mean $\boldsymbol{\theta}$. This extension is likely very necessary for serious practical applications, where the dynamic factor structure is imposed on the errors of the model. Particularly, extending the work by Quintana *et al.* (1995) and Quintana and Putnam (1996) a shrinkage technique could be used in a DLM, such as,

$$\begin{aligned}
\mathbf{y}_t &= \mathbf{F}_t'\boldsymbol{\beta}_t + \boldsymbol{\epsilon}_t, & \boldsymbol{\epsilon}_t &\sim N(\mathbf{0}, \boldsymbol{\Sigma}_t), \\
\boldsymbol{\beta}_t &= \mathbf{B}\boldsymbol{\theta}_t + \boldsymbol{\nu}_t, & \boldsymbol{\nu}_t &\sim N(\mathbf{0}, \mathbf{U}_t), \\
\boldsymbol{\theta}_t &= \boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, & \boldsymbol{\omega}_t &\sim N(\mathbf{0}, \mathbf{W}_t),
\end{aligned}$$

where $\mathbf{F}_t$ is a matrix of observed explanatory variables, $\boldsymbol{\beta}_t$ is a vector of system regression parameters and more important $\boldsymbol{\epsilon}_t$ is an error vector following a multivariate stochastic volatility model as presented above, namely $\boldsymbol{\epsilon}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}_t)$. In this model, the matrix $\mathbf{B}$ represent the *link* matrix that controls the shrinkage assumptions relating the beta coefficients and the theta hyperparameters that follow a random walk, (Quintana *et al.*, 1995).

The experience to date lead us to believe that all these extensions will be very fruitful and support the preliminary conclusions reached in this thesis about the potential utility of factor models.

Chapter 9

Latent Structure in Non-Gaussian Longitudinal Time Series

The latent time series methodology developed in previous chapters is now extended to data coming from non-normal distributions. The basic assumption is that the parameters of such distributions are related through time and hierarchically across possible several multivariate series. This is a new class of hierarchical multivariate time series models developed for longitudinal data and that may have important impacts in many socio-economic areas. For instance, the general methods developed here can be used to find "performance indicators" as part of an attempt to introduce accountability into public sector activities such as education, health and social services, where the focus is on the development of quantitative comparisons between institutions. The idea is to find a summary statistical measurement on an institution or system which is intended to be related to the "quality" of its functioning, (Goldstein and Spiegelhalter, 1996). In particular, the contributions and methodology in the following two chapters are motivated by studies of health care "quality monitors" in the nationwide VA Hospital System. This section of the thesis represents, in part, research that was performed in consultation and collaboration with the Veterans Affairs Management Science Group, Bedford MA.

9.1 The VA Hospital System

The performance monitoring system of the US Department of Veterans Affairs collects, reports and analyses data from across the system of over 170 hospitals as discussed in Burgess *et al.* (1996), sections 3 and 4. The VA Management Science Group is interested in assessing and comparing clinical and health care process performance between facilities using analytic strategies. Therefore, the main objectives in a statistical analysis are those of adequately defining and accurately estimating underlying measures of hospital-level performance in the monitor-specific areas. In addition, the methodology should be able to address the special interest in inferences about relative performance over time and across hospitals. To be more specific, the statistical contributions in exploration and modeling of longitudinal data in the context of this application are developed to:

- understand patterns of variability over time, in periods of single years, in hospital-level and monitor area-specific performance measures across a selection of quality monitors, and
- understand patterns of dependencies between sets of monitors, in addition to and in combination with assessment of time-variations.

These goals are motivated by policy interests in accurately estimating measures of hospital-level performance in key areas of health care provision, in assessing changes over time in such measures to monitor impact of internal policy changes (or the lack thereof), and ultimately in connection with the development of management and economic incentives designed to encourage and promote care provision at sustained and acceptable levels.

9.1.1 Data Structure

The data provided by the VA Management Science group are based on a uniform process of data collection and consolidation across national databases, and encompass a range of in-patient, outpatient and long term care activities at each of the VA medical centers. The data is collected annually for each hospital providing care in areas covered by specific monitors. The hospital-area records consist of the total numbers of individuals who were exposed to a specific and well-defined outcome in that area, and for how many such individuals that defined outcome occurred. In addition, the VA provides information on one independent variable, which is also included in each record. In this case this variable is , referred as the **DRG predictor**, and represents a predicted or expected number of outcomes out of the total. This quantity is based on assumedly exogenous prediction of the outcome proportion and is designed to provide some degree of correction for hospital/monitor specific case-mix and characteristics of the patient population profile. This thesis is focused on exclusively on three quality monitors, coded $\mathcal{M}1, \mathcal{M}2$ and $\mathcal{M}3$, concerning discharges from the hospital in three areas,

- $\mathcal{M}1$: *General Psychiatric*,
- $\mathcal{M}2$: *Substance Abuse Psychiatric* and
- $\mathcal{M}3$: *Basic Medical.*

The outcomes/responses in these monitor areas represent annual numbers of individuals under a binary classification in an area of basic medical or psychiatric health care. Specifically, the response recorded is the number of individuals who failed to return for an out-patient visit (appropriately well-defined) within 30 days of discharge. Low return rates are indicative of low "quality" in these specific care areas. The analysis presented here is based on records for a subset of 152 hospitals in the system

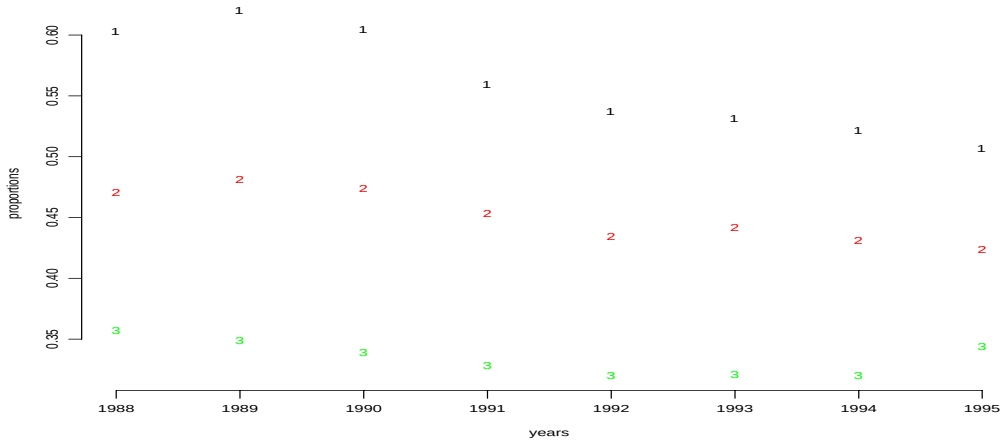

Figure 9.1: Raw averages of observed response proportions over the eigth years for monitors $\mathcal{M}1, \mathcal{M}2$ and $\mathcal{M}3$.

that have complete records. Each hospital reports information on the three monitors for eight years, from 1988 to 1995. There is an overall suggestion of decreasing levels of observed responses across the eight years and in all three monitors, most marked in Monitors $\mathcal{M}1, \mathcal{M}2$ and, to a lesser extent Monitor $\mathcal{M}3$, as can be seen in Figure 9.1, which graphs the simple average responses across years. The average DRG values (not shown) do not show similar decreasing patterns indicating that this is very likely a hospital system-wide feature, perhaps due to VA policy and/or general improvements in care provision over the years. Figure 9.2 displays aspects of data on the three monitors (columns) separately but combined over all eight years (1988-1995). The upper row displays the observed proportions of successes in each monitor against corresponding total numbers of patients in each case. The superimposed lines denote pointwise 99% intervals based on assumed common binomial distributions for the outcomes in each monitor (more discussion on this in the next section). The second row of graphs displays the observed proportions versus the DRG-based predicted proportions. This display is redrawn in the lower row, but now with the DRG predictor on the probability scale.

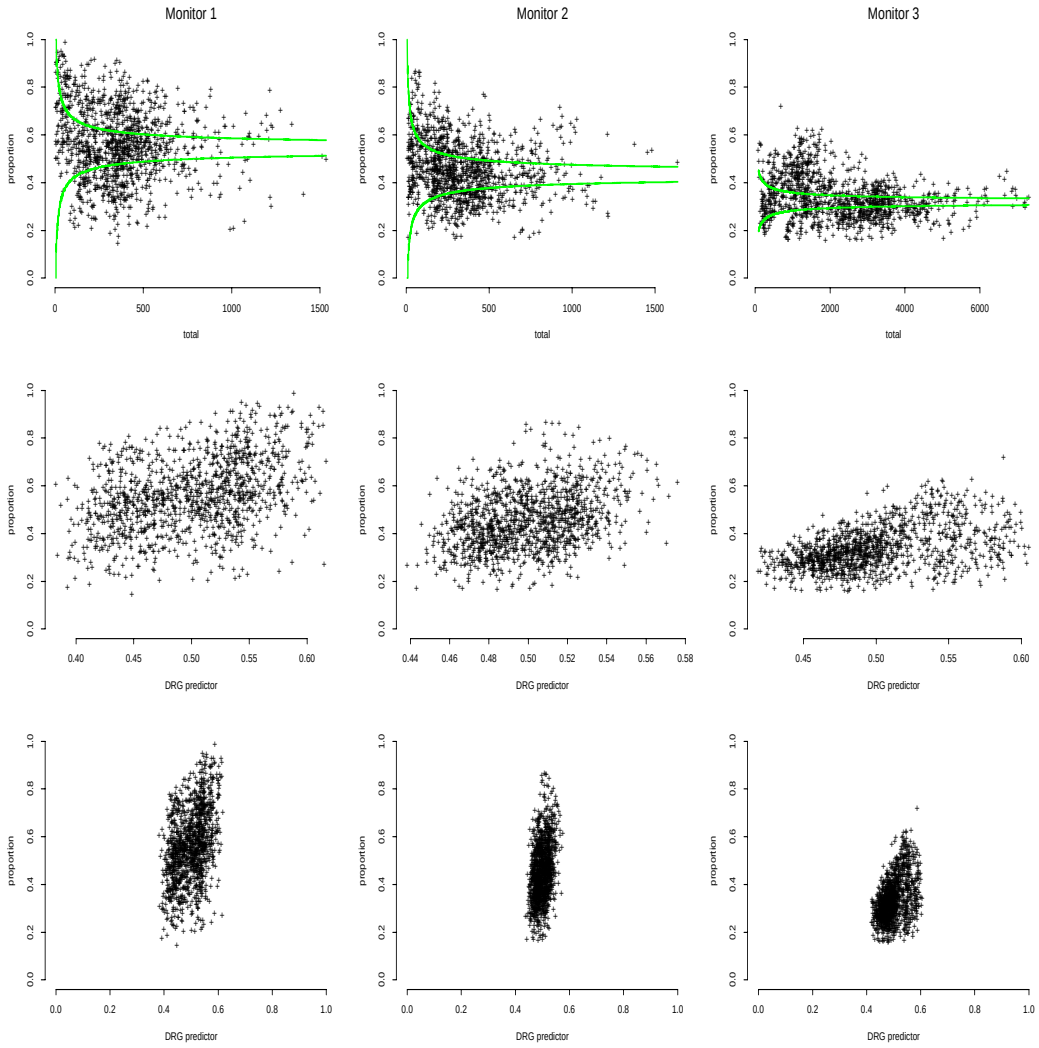

Figure 9.2: Observed proportions in all three monitors combined over all years 1988-1995 for a total of 1216 points in each frame.

Figure 9.3 provides related graphs for each year of data separately, each point corresponding to one hospital in each frame. These pictures will be discussed in more detail in the next section as part of the preliminary data exploration and basic modeling perspectives together with some discussion on long-term work by CN Morris and CL Christiansen (hereafter M&C). These authors have provided a firm basis for addressing key questions of definition and appropriate modeling frameworks for estimation of performance measures (Burgess *et al.*, 1996).

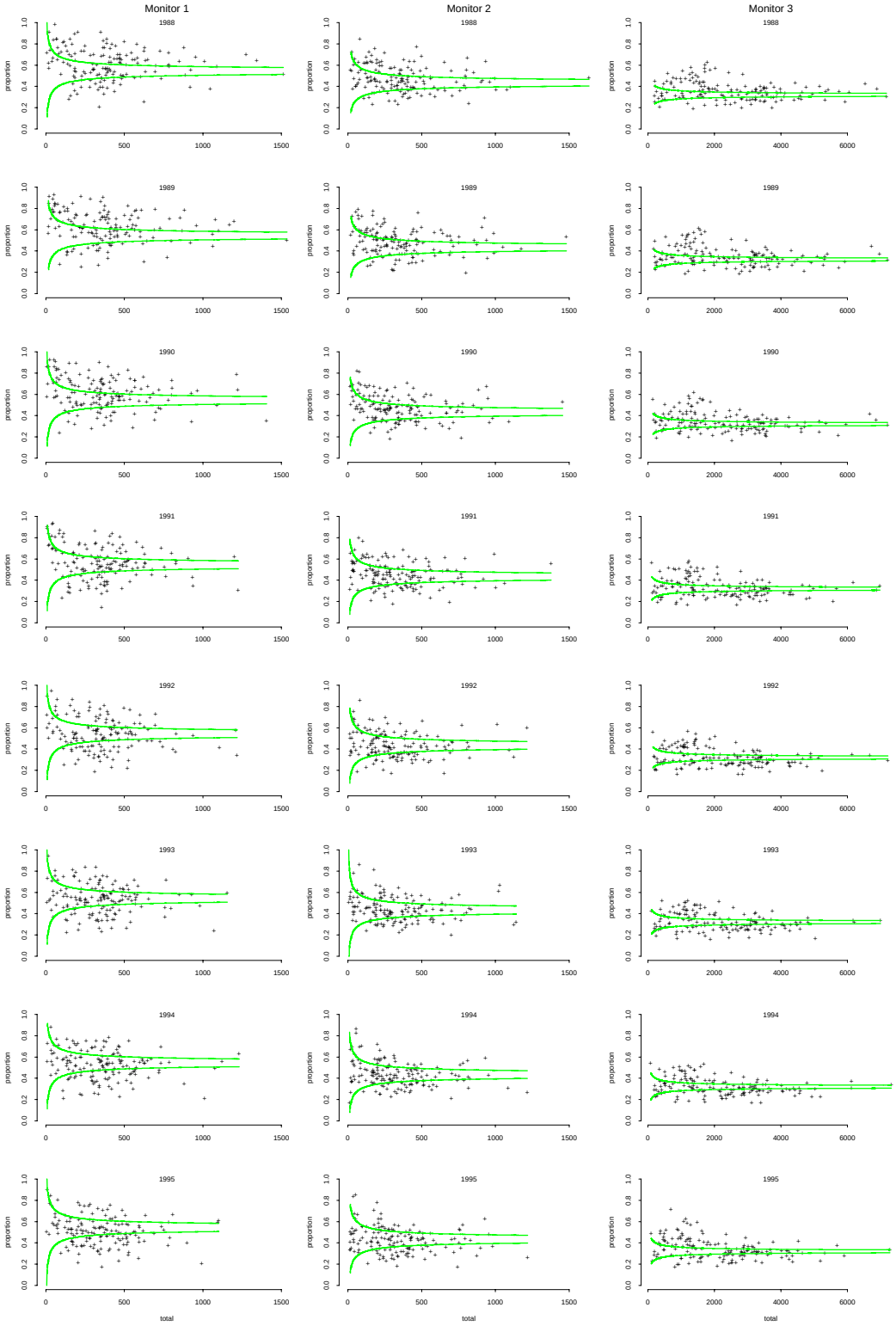

Figure 9.3: Observed proportions in all three monitors (columns) for all years 1988-1995 (rows).

9.2 Initial Perspectives and Notation

The nature of the data, counts out of totals, suggests the use of traditional binomial models as a starting point. For instance, consider data only on ONE monitor collected in ONE year across all hospitals $i = 1, \ldots, I$. Let z_i be the number of patients, out of the total n_i, who failed to return for an out-patient visit within 30 days of discharge from hospital i. Assume conditionally independent binomial models,

$$z_i | n_i, p_i \sim \text{Bin}(z_i | n_i, p_i) \tag{9.1}$$

for each hospital i. Note that with this assumptions the unknown "success probabilities" p_i are hospital-specific and the totals n_i are assumed uninformative about p_i. As it was mentioned before, Figure 9.2 displays the proportions z_i/n_i for the $I = 152$ hospitals, combined over all eight years for a total of $8 \times 152 = 1216$ observations in each frame, representing each of the three monitors. Super-imposed on the first row of graphs are approximate 99% intervals under the binomial distribution (as n_i varies along the x-axis) assuming $p_i = p$ is fixed at the overall average proportion. Many observations lie outside these bands indicating the very high degree of over-dispersion relative to a single binomial model. This extra-binomial variation is evident in all three monitors and for the eight years of study as can be seen in Figure 9.3. This variation is to be explained by models that describe how the individual p_i vary across hospitals (and, later, across years), using a combination of regression on the DRG predictor and random effects. However, in this case the values of the DRG predictor variable lie in a very narrow range on the probability scale across the three monitors. This is very clear in the third row of graphs in Figure 9.2. Though there is an apparent positive association between the DRG variable and the observed outcomes, evident in the second row of graphs, it is clear that the lack of variation will lead to a high degree of uncertainty about any regression model on the DRG. The predictor

simply does not vary substantially across the "design" space. For Monitor $\mathcal{M}1$, for example, the DRG variable is very close to 0.4 across all hospitals in all years. This low variability will result in regression coefficients with flat likelihood functions.

Let d_i be the predicted proportion of "successes" for hospital i based on the independent DRG variable. The d_i is supposed to predict p_i on the basis of system-wide studies of patient case-mix profiles and historical data. Adopting the standard logistic regression framework involves modeling $\mu_i = \log(p_i/(1-p_i))$ as a linear regression on some function of d_i, the natural choice being the centered observed logit predictor $x_i = l_i - \bar{l}$, where $l_i = \log(d_i/(1-d_i))$ and this is assumed here. Write

$$\mu_i = \alpha_i + \beta_1 x_i \tag{9.2}$$

for $i = 1, \ldots, I$. This gives a basic linear regression on the DRG-based predictor with hospital population slope β_1, allowing for hospital-specific intercept parameters α_i. These α_i terms stand for *all* sources of extra-binomial variation not adequately captured in the DRG predictor and all the differences and variations in success rates due to differences in patient profiles across hospitals. These differences are confounded with differences due to policies and practices in the area of care and therefore there is no way of "unconfounding" these issues without additional covariate information.

Under the assumption of exchangeability for the α_i parameters (**random effects**), drawn from a hospital-population prior delivers a class of Bayesian hierarchical models related to those developed in Burgess *et al.* (1996), Christiansen and Morris (1997) and Goldstein and Spiegelhalter (1996). Some other variations and generalizations of the basic model are part of the standard Bayesian random effects generalized linear models that have been used for institutional comparisons. Some of these variations include models of M&C, as follows:

- *Poisson approximations.* In cases where n_i is large and p_i is small, appeal to Poisson approximations to the binomial sampling model leads M&C to the class

of PRIMM regression models in which z_i is conditionally Poisson with mean $n\lambda_i$ (Christiansen and Morris, 1997). In such cases the logistic regression is approximated by a log-linear regression with $\lambda_i = x_i^{\beta_1}\eta_i$ where $\eta_i = \exp(\alpha_i)$. The PRIMM models of M&C adopt gamma distributions for the hospital population priors of the random effects η_i. In the case of monitors $\mathcal{M}1$, $\mathcal{M}2$ and $\mathcal{M}3$ the conditions, of large n_i and low p_i, for the Poisson approximation are not satisfied. As can be seen in the pictures, the outcome proportions across are concentrated in 0.15-0.85 for a wide variety of sample sizes.

- *Normal approximations.* In cases where the n_i are reasonably large and the p_i are not close to zero or one, the binomial likelihood function $p(z_i|n_i, \mu_i)$ can be adequately approximated as a function of μ_i by a function proportional to $\exp(-(y_i - \mu_i)^2/s_i)$ where $y_i = \log(z_i/(n - z_i))$ and $s_i = n_i\hat{p}_i(1 - \hat{p}_i)$ with $\hat{p}_i = z_i/n_i$. This leads to the normal hierarchical model as developed in Burgess *et al.* (1996). These models are certainly more adequate for the VA data and hence a good choice for initial data exploration and a nice connection with technical methods of posterior simulation in binomial models as will be seen.

In the next section basic random effects hierarchical models are developed and applied separately for each one of the three monitors and for each year of data. These models, as stated before, are close in spirit and form to those based on normal approximations developed by Jim Burgess, Cindy Christiansen, Carl Morris and Ted Stefos. However, the approach taken here is based on the exact conditional binomial sampling distributions rather than approximations.

9.3 `MODEL-1`: Basic Random Effects Model

The first approach to model data with binomial structure (9.1) is to assume normal priors for the population distribution of the random effects α_i on the linear regression

model describe in (9.2), namely

$$\alpha_i \sim N(\alpha_i|\beta_0, w^2), \tag{9.3}$$

for some mean β_0 and standard deviation w. An equivalent model can be found by defining the zero-mean random effects

$$\epsilon_i = \alpha_i - \beta_0, \tag{9.4}$$

hence model (9.2) is reparametrized to give

$$\mu_i = \beta_0 + \beta_1 x_i + \epsilon_i \quad \text{and} \quad \epsilon_i \sim N(\epsilon_i|0, w^2). \tag{9.5}$$

This new "hierarchically centered" representation gives nice interpretations of the parameters and hence is preferred. In this model, the level β_0 represents the hospital system-wide average in corrected responses on this logit scale. The ϵ_i terms represent hospital-specific departures from the system-wide underlying level β_0. It is clear that posterior inferences for both the absolute random effects α_i and the relative effects ϵ_i are of interest. In addition, it is worth mentioning that the unexplained variability in the logits μ_i, and hence in the probabilities p_i, is related to the prior variance w^2. Large values of such variance represent high levels of heterogeneity in outcomes that are not captured via the DRG predictor. For example, if w is very small, $\beta_1 = 1$ and $\beta_0 = 0$ then $E(\mu_i) = x_i$ and hence $p_i \approx d_i$.

Individual analyses, for each monitor and year, are considered to investigate the validity of this basic binomial regression model. The studies assume binomial structure (9.1) with normal population distribution of random effects (9.5) and diffuse reference priors for the additional parameters (β_0, β_1, w).

Posterior summaries for this basic single-year, single-monitor model, referred as `MODEL-1`, are displayed in Figures 9.4 to 9.9 inclusive. The graphs display boxplots of posterior distributions for selected model parameters. The boxes are centered at

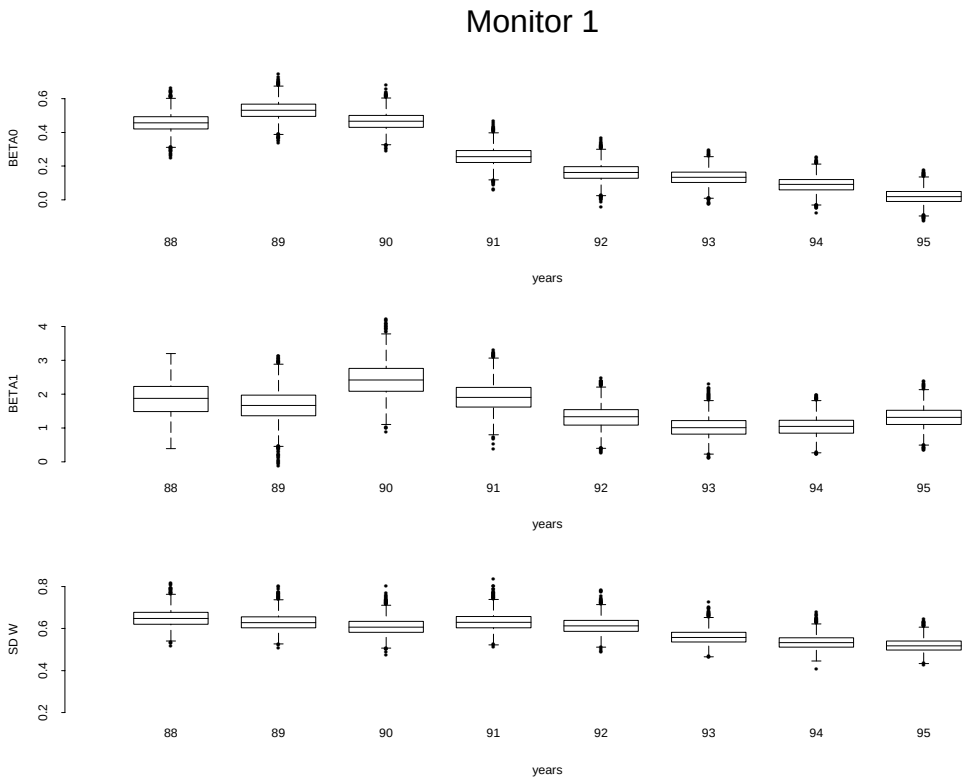

Figure 9.4: Summary interval estimates of β_{0t}, β_{1t} and ω_t using the single-year, single-monitor `MODEL-1`.

posterior medians, have boxes drawn out to posterior upper and lower quartiles, and lines from the quartiles to notches at points no further than 1.5 times the interquartile range from the quartiles. Points are plotted outside these final notches to represent the very extreme tails of the posteriors–most of these can be simply ignored, as almost all the posterior probability lies between the notches (well over 99% under a normal posterior). In the next subsections some comments on the posterior graphs are made together with initial conclusions for each of the three monitors over the eight years.

Hospital population levels

The parameter β_0 in (9.5) stands for the hospital population level in the VA system. That is, at a mean DRG variate $x_i = 0$, parameter β_0 represents the average response probability, on the logit scale, across the hospital system. Differences across years

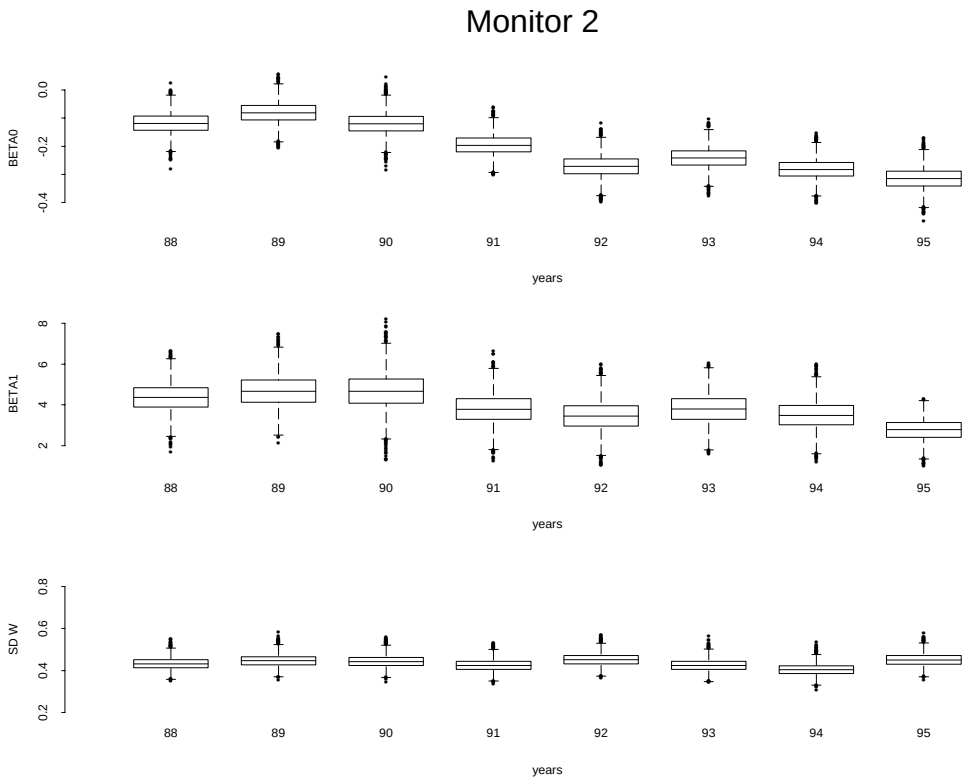

Figure 9.5: Summary interval estimates of β_{0t}, β_{1t} and ω_t using the single-year, single-monitor `MODEL-1`.

for a single monitor require explanations, possibly in terms of changes in system-wide policies and practices, or in terms of otherwise uncontrolled "drifts" in levels of quality. Posterior distributions of these parameters are plotted in Figures 9.4, 9.5 and 9.6 for monitors $\mathcal{M}1, \mathcal{M}2$ and $\mathcal{M}3$ respectively. As can be seen from these graphs there are meaningful differences in the β_0 parameters across the eight years in each of the three monitors. The main feature is a general decreasing trend in β_0 over the years for all three monitors, more markedly for Monitors $\mathcal{M}1$ and $\mathcal{M}2$. This corresponds to generally decreased probabilities of return for out-patient visits within 30 days of discharge, and the apparent similarities between Monitors $\mathcal{M}1$ and $\mathcal{M}2$ are consistent with the two being related areas of care, each related to psychiatric discharges. Monitor $\mathcal{M}3$, General Medical Discharges, exhibits a quite abrupt increase in β_0 in 1995, after decreasing and leveling off in 1993-4; this requires

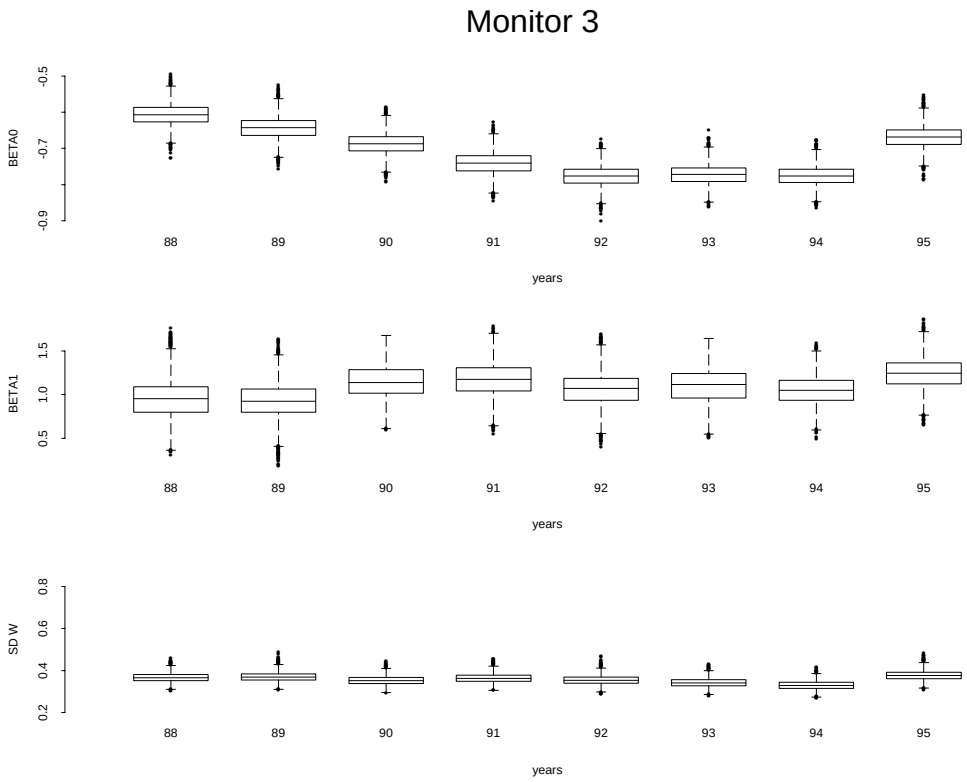

Figure 9.6: Summary interval estimates of β_{0t}, β_{1t} and ω_t using the single-year, single-monitor `MODEL-1`.

further consideration and interpretation from VA personnel.

DRG regression effect

Now consider the β_1 regression parameters, referring to the second graph in each of Figures 9.4, 9.5 and 9.6. Again there are apparent differences over the years within each monitor, although the values are relatively stable compared with β_0 and have a lesser impact on overall conclusions.

Dispersion in hospital population of random effects

Now consider the standard deviation w that determines the dispersion in the random effects distribution on the population of hospitals, lower graph in each of Figures 9.4, 9.5 and 9.5. Within each monitor, w appears essentially constant over the years

and the range of values indicated supports appreciable variation consistent with the high degree of extra binomial variation apparent in the data. This variability is highest for Monitor $\mathcal{M}1$, at s.d. levels of $0.6 - 0.7$ on this logit scale, then Monitor $\mathcal{M}2$, at levels around $0.4 - 0.5$, and with levels around $0.3 - 0.4$ for Monitor $\mathcal{M}3$. As a benchmark for interpretation, consider a case in which the DRG predicted proportion is 50% (so zero on the logit scale) and assume $\beta_0 = 0$ so that baseline population probabilities are around 0.5. A standard deviation of $w = 0.5$ leads to an 95% interval for the outcome probability of roughly $0.27 - 0.73$. Thus, even ignoring the additional binomial variation about the outcome probability, the random effects distribution covers much of the observed range of the data.

Random effects for three example hospitals

Figures 9.7, 9.8 and 9.9 display inferences on hospital-specific random effects for three arbitrarily selected hospitals, those with station numbers 41, 92 and 2. In each year, posterior boxplots for the corresponding hospital-specific relative levels ϵ_i are graphed for Monitors $\mathcal{M}1, \mathcal{M}2$ and $\mathcal{M}3$. These plots provide examples of the kinds of patterns of variations exhibited by the random effects within individual hospitals. One notable feature is that of evident dependence over the years in the ϵ_i within specific hospitals, especially in terms of sustained signs. Though the series of length eight are very short, this is supportive of systematic dependence structure over time that is naturally expected: a hospital that has tended to be below the population norm in terms of its proportions of outcomes in recent years will be expected to maintain its below average position this year, so that the ϵ parameters of this hospital will tend to be of the same sign. Therefore, there is a clear need of some form of time series structure to describe and incorporate such dependencies, and in order to explore and assess their effects on estimation of all model parameters.

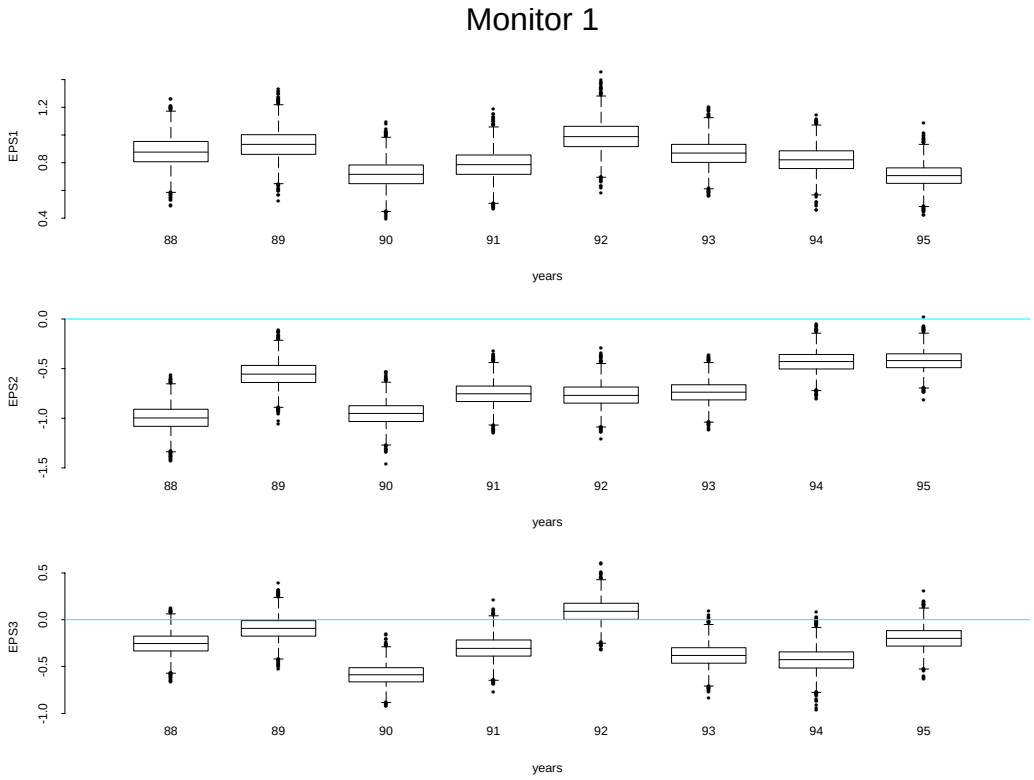

Figure 9.7: Summary interval estimates of hospital-specific random effects ϵ_{it} for three hospitals using the single-year, single-monitor `MODEL-1`.

For the hospital population parameters (β_0, β_1, w), reanalyses under ranges of diffuse but proper priors were performed to assess sensitivity. Across several choices of seemingly uninformative though proper priors, the posterior results confirm those from the reference prior based analyses above. Model adequacy has been explored and verified through post-hoc residual analyses, examining plots and features of posterior samples of well-defined residuals. This is not illustrated here, but it will described in the more elaborate models of the next chapter. Additional ranges of analyses were performed using the normal approximation to the binomial likelihood functions. In general, very similar posterior results for (β_0, β_1, w) were delivered, indicating that the normal approximations are indeed very good for inferring these parameters and with these data sets. However, a small number of hospital/monitor pairs have very low sample sizes n_i in some years, and there are some in which the observed outcome

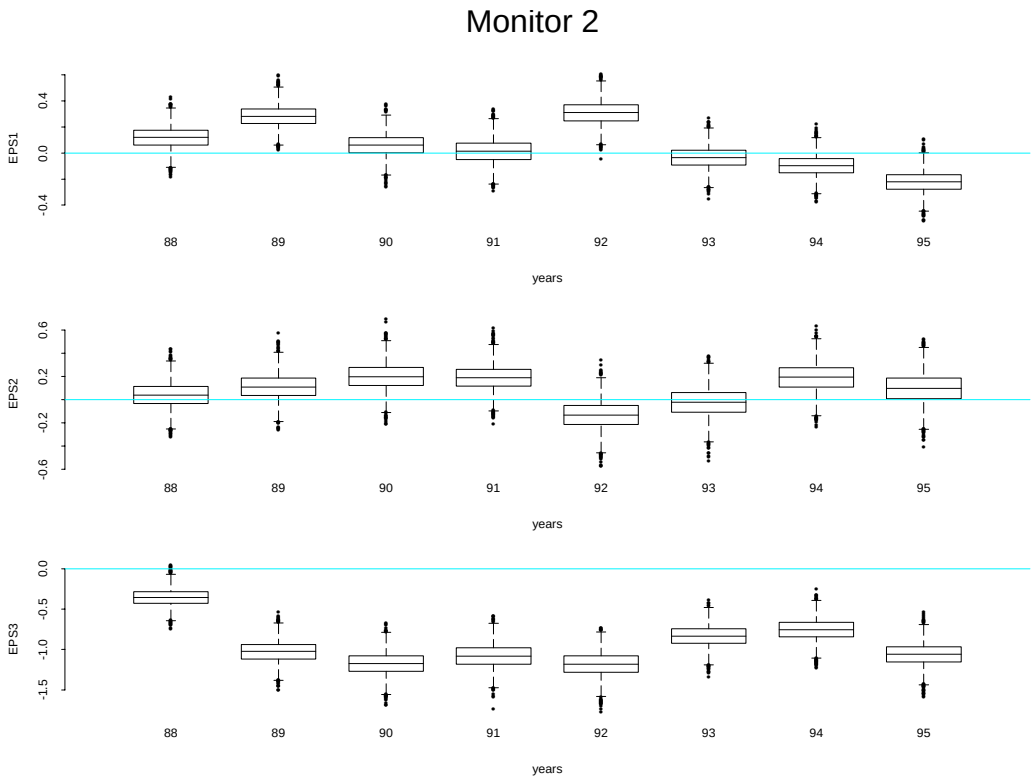

Figure 9.8: Summary interval estimates of hospital-specific random effects ϵ_{it} for three hospitals using the single-year, single-monitor `MODEL-1`.

proportions are close to zero or one. In such cases, the posteriors for the corresponding α_i parameters cannot be expected to be necessarily reliable approximations, and any errors in approximation must be expected to impact on inferences about other parameters. Hence it is preferable to remain in the theoretically sound binomial framework rather than to adopt the (technically and computationally more manageable) normal approximations directly. However, for much of the data the normal model represents a good approximation.

An important conclusion at this point is the apparent need to consider models in which the overall levels β_0 of outcome responses across all hospitals vary year to year, and in which the regression effects of the DRG variable (however weakly identified they may be) may also vary. In other words, management policies that operate across all hospitals, and general improvements (or otherwise) in care provision that

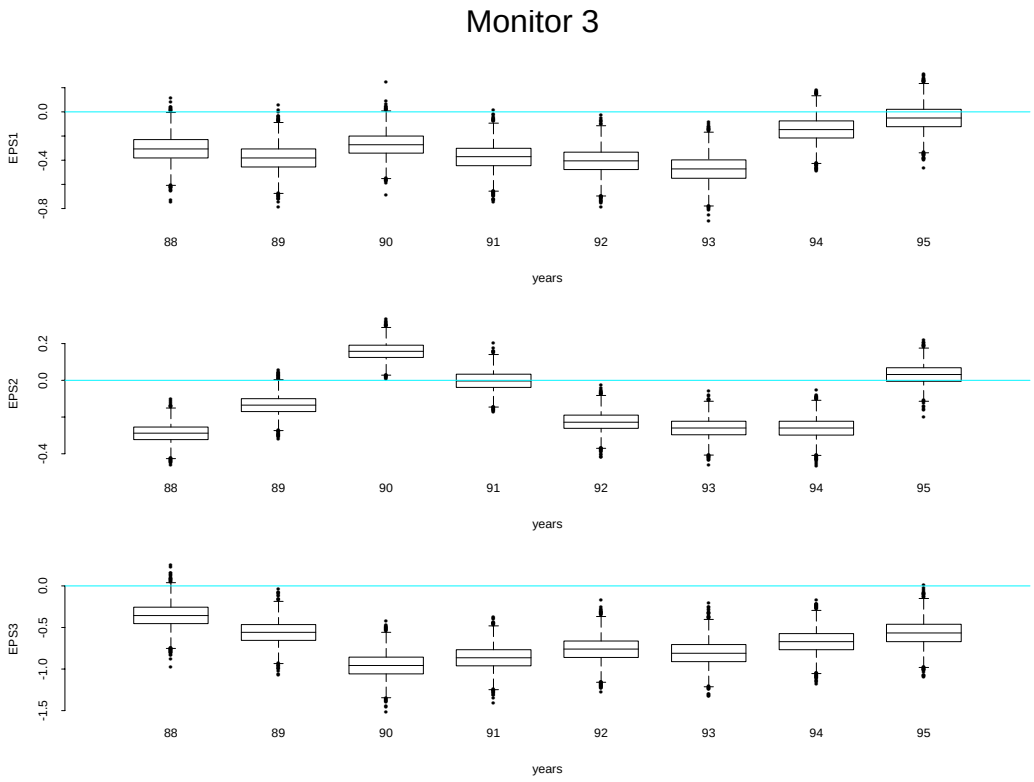

Figure 9.9: Summary interval estimates of hospital-specific random effects ϵ_{it} for three hospitals using the single-year, single-monitor `MODEL-1`.

impact all hospitals in similar should represented by changes in β_0 from year to year. Furthermore, its is desirable to introduce time series concepts to address the notion of systematic patterns of variation over the years in hospital-specific effects on a given monitor. In the first part of next chapter, a class of time series random effects models that necessarily include additional components of unpredictable variability in outcome probabilities as well as time series components, is discussed. All this is developed in the single monitor context, and summary inferences from our analyses of the three monitors separately are discussed and compared to these initial studies.

Chapter 10

Multivariate Non-Gaussian Hierarchical Time Series Models

This chapter continues the study of the VA hospital monitor time series, extending the modeling approaches with new Bayesian models for discrete, multivariate time series. In addition, the class of multivariate models developed here is related to the dynamic factor models from chapters 6 and 7 through common statistical structure enhancing the latent time series structure in the data. For instance, the initial modeling described in the previous chapter indicates apparent variability over time in the hospital/monitor population parameters β_0 and β_1, as well as the random effects α_i that explain much of the observed extra-binomial variability. Nevertheless, there are indications of relative stability of the ϵ_i across years in some hospitals, which is consistent with expectation. Unless policies and protocols in the monitor care areas are radically changed from one year to the next, there should be stability in these quantities as representing true quality levels; any changes beyond this reflect random variations due to the characteristics of the patient sample presenting at the hospital. The next stage of investigation concerns the structure of single monitor series over time. Specifically, models that incorporate the view that the ϵ_i are expected to remain stable within each hospital over such a short number of years, but that

unexplained sources of variability at the hospital level are expected to induce random changes year to year.

10.1 MODEL-2: Models for Single Monitor Time Series

The first approach towards a more general model is to introduce one pair of hospital population parameters, β_0 and β_1, for each year separately, and explore analyses that do not impose any further structure, simply allowing for the estimation of these parameters. For example, the apparent decrease over time in the overall levels of response on Monitor $\mathcal{M}1$ will then be largely accounted for by estimating separate β_0 parameters for each of the eight years, but it is not anticipated in the model. This model does not therefore provide a structure to explore or model dependence over time in the β parameters, and so is not useful for prediction to future years without modification or intervention. This is not a criticism; since the goals of the study are to evaluate patterns over time in the eight years of data and to explore inferences about changes in hospital-specific effects over the years. Development of predictive models for the β parameters would require the evaluation of expert opinion about the reasons behind any inferred time evolution and the use of this in phrasing appropriate predictive models.

For now and keeping the same focus on a single monitor as in the previous chapter, write α_{it} for the logistic random effects parameter of hospital i in year t, with $t = 1, \ldots, 8$ representing years 1988 to 1995 inclusive. The corresponding relative effects are now $\epsilon_{it} = \alpha_{it} - \beta_{0t}$ where β_{0t} and β_{1t} are the population parameters of the logistic regression in each year $t = 1, \ldots, 8$. The first logical linear departure from the random models of the initial study is to use autoregressions. Specifically, AR(1) models are used for the series α_{it} over time t, considered conditionally independently

over hospitals. With such a short time span (8 years) more complex models are largely untenable. Moreover, the AR(1) model is a natural, interpretable model that describes this year's α and "close to" last year's but with a degree of "noise" added. It also turns out to be adequate as a model of dependence through time for the three monitors under study, and, as mentioned below, has a desirable consequence in that the annual marginal distributions of the hospital-specific effects are the same across the short span of eight years.

Modeling any form of time series dependence over the years in the α_{it} quantities introduces partial stochastic constraints so that the α_{it} are no longer as free to vary as in the single year models `MODEL-1` of section 9.3 where they are essentially unconstrained, viewed simply as independent across years. Hence, it is necessary to consider that some of the evident variation in the logit parameters μ_{it} will be unexplained with this time series structure. Because of that, additional terms, representing residual, unexplained variation, are included in the linear model for the logit parameters. Specifically, the extension of the single monitor model in equations (9.1), (9.2) and (9.5) is as follows. Independently across hospitals $i = 1, \ldots, I$, and over all years $t = 1, \ldots, 8$, the data are assumed to arise from the set of $8I$ binomial models

$$(z_{it}|n_{it}, p_{it}) \sim \text{Bin}(z_{it}|n_{it}, p_{it}) \tag{10.1}$$

with logistic regression on the DRG predictors and random effects,

$$\mu_{it} \equiv \log(p_{it}/(1-p_{it})) = \alpha_{it} + \beta_{1t}x_{it} + \nu_{it}. \tag{10.2}$$

This regression is similar to equation (9.2) but now the subscripting makes explicit the year and hospital-specific parameters. In addition, the residual terms,

$$\nu_{it} \sim N(\nu_{it}|0, v^2) \tag{10.3}$$

are included to model residual variation not explained by the regression and hospital-

specific random effects α_{it}. The time series components assume that, again independently across hospitals i, the α_{it} follow first-order autoregressions centered around the current population level β_{0t} in year t. For $t > 1$,

$$\alpha_{it} = \beta_{0t} + \phi(\alpha_{i,t-1} - \beta_{0,t-1}) + \omega_{it}, \tag{10.4}$$

where $\omega_{it} \sim N(\omega_{it}|0, u^2)$ independently over hospitals i and years t. At $t = 1$,

$$\alpha_{i1} \sim N(\alpha_{i1}|\beta_{01}, w^2). \tag{10.5}$$

Here ϕ is the autoregressive parameter and will generally be close to one, lying in part of stationary region $0 < \phi < 1$, and following the relationship $w^2 = u^2/(1-\phi^2)$. Equivalently, in terms of the annual deviations of the hospital-specific effects from the annual levels, $\epsilon_{it} = \alpha_{it} - \beta_{0t}$, related over time by

$$\epsilon_{it} = \phi\epsilon_{i,t-1} + \omega_{it}. \tag{10.6}$$

Note that ϕ and u are assumed constant in the time series components. This is supported on the basis of exploratory and confirmatory analyses, though could be relaxed to allow for differing variances across hospitals and/or years as may be desirable for other applications. This class of multiple-year, single-monitor, hierarchical random effects time series models has some important characteristics, as follows.

- Unconstrained year-to-year variation in the population level of logit-probabilities, β_{0t} and in the regression coefficient on the DRG-based predictor, β_{1t}.

- The hospital-specific relative effects ϵ_{it} are now structurally related over time within each hospital according to the AR model. With an appropriately large value of ϕ, this implies high positive correlations between the ϵ_{it} in a given hospital over the years. This is consistent with the view that, for example, a hospital that is been generally "good" in a specific monitor/care in one year

will have a high probability of remaining "good" the next year, and vice versa. This is true irrespective of the system-wide changes in quality levels; they are captured in changes in the β_{0t} sequence, and the ϵ_{it} represent departures above or below the system levels β_{0t}.

- The constancy of ϕ and u over years is sufficient (though not necessary) to induce (desirable) stationarity into the model for the relative effects ϵ_{it}. Thus, for example, the implied marginal distribution in any chosen year t is given by

$$\epsilon_{it} \sim N(\epsilon_{it}|0, w^2) \tag{10.7}$$

  with $w^2 = u^2/(1-\phi^2)$, independently over i. Equivalently, $\alpha_{it} \sim N(\alpha_{it}|\beta_{0t}, w^2)$ for each year t. So, in each year, the relative random effects are a random sample from a zero-mean normal distribution. This is consistent with a view of no global changes in the hospital population makeup, variability in expected levels being essentially constant over the short period of years once the DRG predictor and any system-wide changes are accounted for through β_{1t} and β_{0t}, respectively. Changes in relative performance of hospitals can therefore be assessed across years.

- Residual variation in the logit probabilities not explained by either the DRG-based regression or the correlated random effects is contributed by the ν_{it} in equation (10.2). Estimation of the residual variance parameter v^2 together with all other parameters will provide indications of the extent of this "unexplained" variability. With respect to the exploratory, single year model with no time series structure in section 9.3, the original α parameters of equation (9.2) are decomposed into two components: the new α_{it}, still hospital-specific random effects but now structurally related over time, and the purely residual and unpredictable components ν_{it}. The model degenerates to the earlier non-time

series model if $\phi = 0$, in which case the α_{it} and ν_{it} terms are not distinguished. The analyses to follow indicate relatively high time series dependencies with appropriate values of ϕ that are positive and significantly large. As a result, the "systematic" (though random) variation explained by the α_{it} is isolated and identified separately from the unexplained variation in the ν_{it}.

One final feature to note, concerns the time series structure of the **combined hospital-specific random effects** $\epsilon_{it} + \nu_{it}$ above. This combined term represents a time series whose structure is autoregressive, moving-average of order one, ARMA(1,1). The addition of the residual/noise terms ν_{it} to the AR(1) process ϵ_{it} acts to modify the correlation structure. Of course if v is small compared to w the modification is small.

Summaries of analysis of each of the three monitors separately appear in Figures 10.1 to 10.6 inclusive, in formats similar to those from the preliminary analysis under `MODEL-1`.

Hospital population parameters

Across years, inferences for the β_{0t} and β_{1t} quantities are essentially the same as under `MODEL-1`, as is to be expected as these quantities are not constrained; the displays of yearly boxplots for these quantities in each of the three single monitor analysis appear in the upper two rows of Figures 10.1, 10.2 and 10.3 for monitors $\mathcal{M}1$, $\mathcal{M}2$ and $\mathcal{M}3$ respectively.

For each of the monitor analyses, boxplots of posteriors for the monitor-specific w appear in the lower row of Figures 10.1, 10.2 and 10.3. The supported ranges are lower that those of the `MODEL-1` analyses in each case. This is to be expected if the variances v^2 of the "unpredictable" components of variation in the random effects model are non-negligible.

For the monitor-specific autoregressive parameters ϕ, posterior boxplots appear

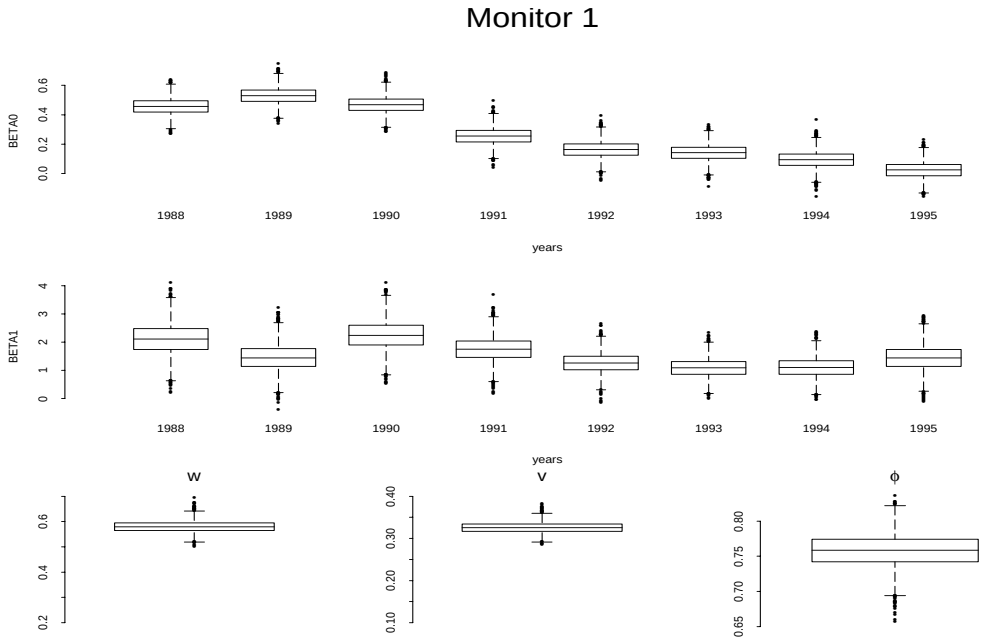

Figure 10.1: Summary interval estimates of β_{0t}, β_{1t}, standard deviations w, v and AR(1) parameter ϕ using the multi-year, single-monitor `MODEL-2`.

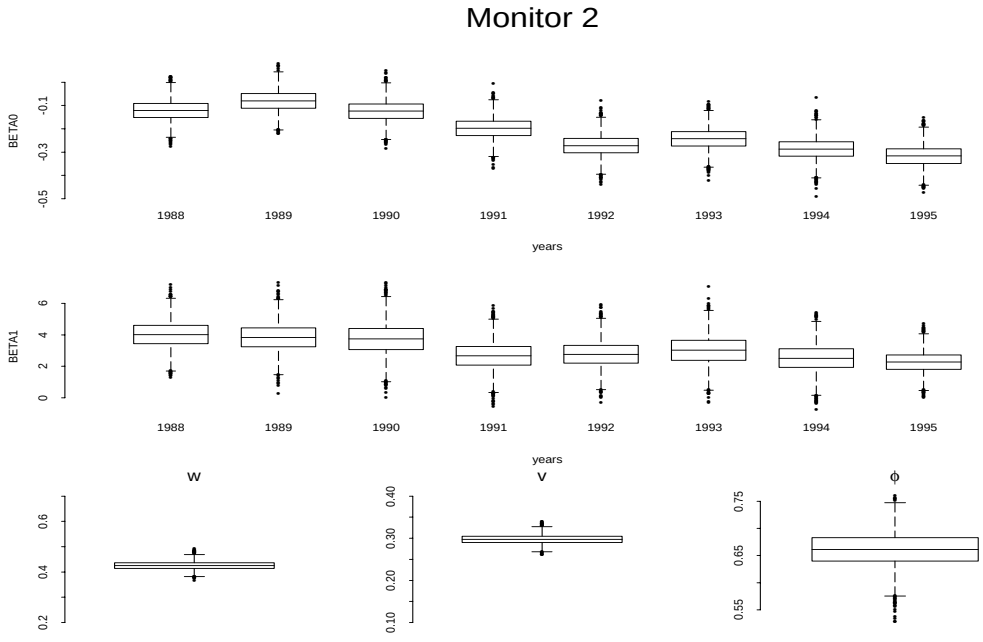

Figure 10.2: Summary interval estimates of β_{0t}, β_{1t}, standard deviations w, v and AR(1) parameter ϕ using the multi-year, single-monitor `MODEL-2`.

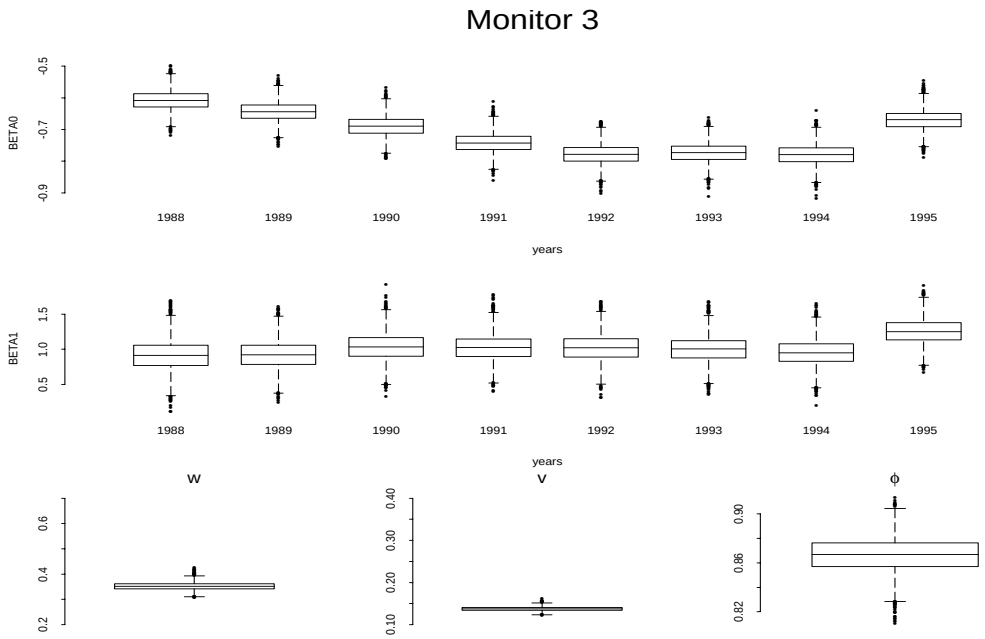

Figure 10.3: Summary interval estimates of β_{0t}, β_{1t}, standard deviations w, v and AR(1) parameter ϕ using the multi-year, single-monitor `MODEL-2`.

in the lower row of Figures 10.1, 10.2 and 10.3. These indicate highly significant dependence structures in each case, with inferred values of ϕ in the ranges $0.7-0.8$ for $\mathcal{M}1$, $0.6-0.75$ for $\mathcal{M}2$ and $0.8-0.9$ for $\mathcal{M}3$. The posteriors have some overlap though do suggest that the ϕ parameters may be different across monitors. The dependence in the random effects time series is high in each case, but perhaps higher for $\mathcal{M}3$ than $\mathcal{M}1$ or $\mathcal{M}2$, the latter two being more comparable. There are thus apparent differences here between $\mathcal{M}3$ and the other two monitors, perhaps associated with the fact that latter two are more closely related health care areas, both involving psychiatric discharges. These ϕ values lead to smoothing in the estimation of the individual random effects series for each hospital within each analysis, noted below.

For the monitor-specific standard deviation v of the unpredictable/residual components of variation, posterior boxplots appear in the lower row of Figures 10.1, 10.2 and 10.3. These indicate non-negligible values, in the ranges of $0.3 - 0.37$ for $\mathcal{M}1$,

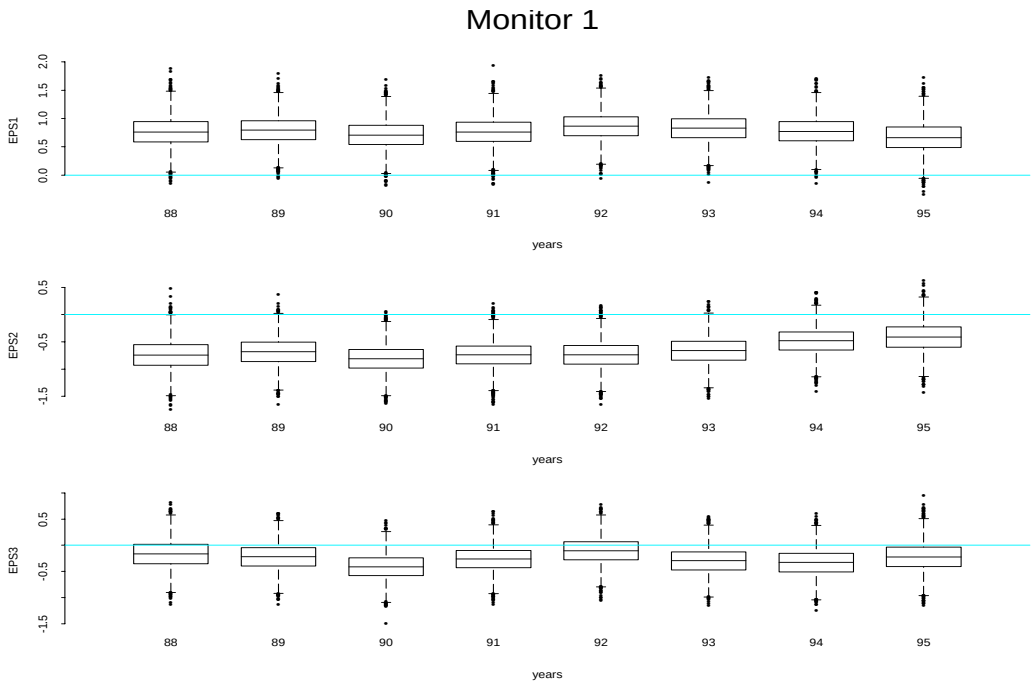

Figure 10.4: Summary interval estimates of hospital-specific random effects ϵ_{it} for three hospitals using the multi-year, single-monitor `MODEL-2`.

$0.27 - 0.33$ for $\mathcal{M}2$ and $0.12 - 0.16$ for $\mathcal{M}3$. By comparison with the posteriors for the w parameters this indicates that the unpredictable/residual components of the model contribute significantly to the variability in outcome probabilities. In terms of the variance ratio $v^2/(v^2 + w^2)$, the ϵ_{it} residuals contribute, very roughly, about $20 - 25\%$ variation for $\mathcal{M}1$, about 30-35% for $\mathcal{M}2$, but only about 15% for $\mathcal{M}3$. Again there are quantitative differences between $\mathcal{M}3$ and the other two monitors, perhaps related to the underlying nature of the differences in health care area.

Random effects for three example hospitals

Summary inferences for the relative random effects ϵ_{it} for the three selected hospitals are again given in terms of posterior boxplots for each year in each of the single monitor analyses. These appear in Figures 10.4, 10.5 and 10.6 for $\mathcal{M}1$, $\mathcal{M}2$, and $\mathcal{M}3$ respectively.

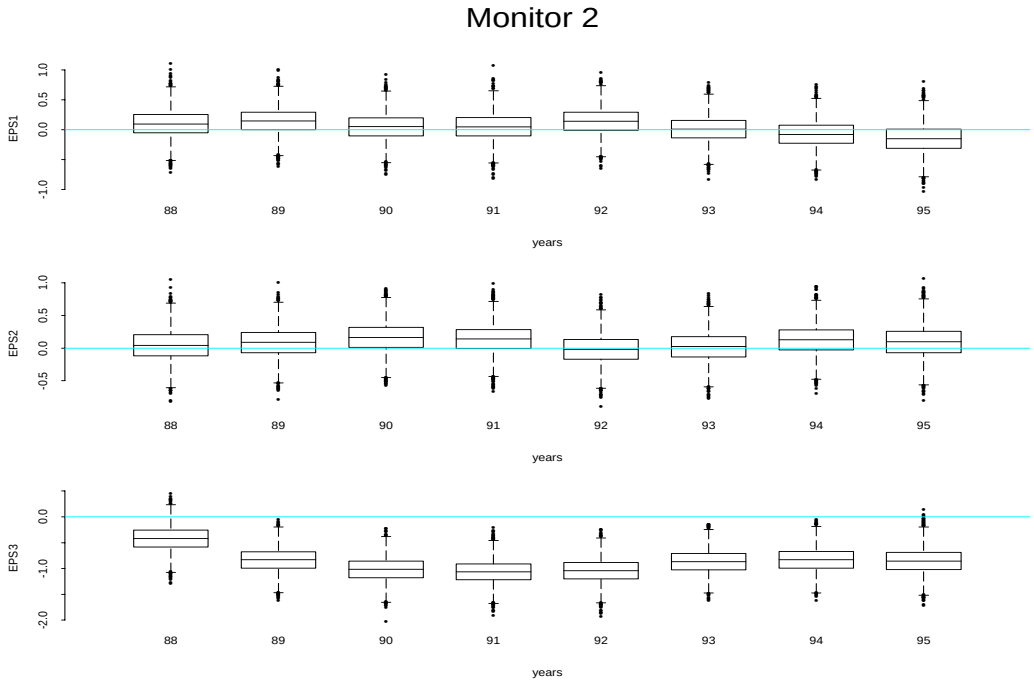

Figure 10.5: Summary interval estimates of hospital-specific random effects ϵ_{it} for three hospitals using the multi-year, single-monitor `MODEL-2`.

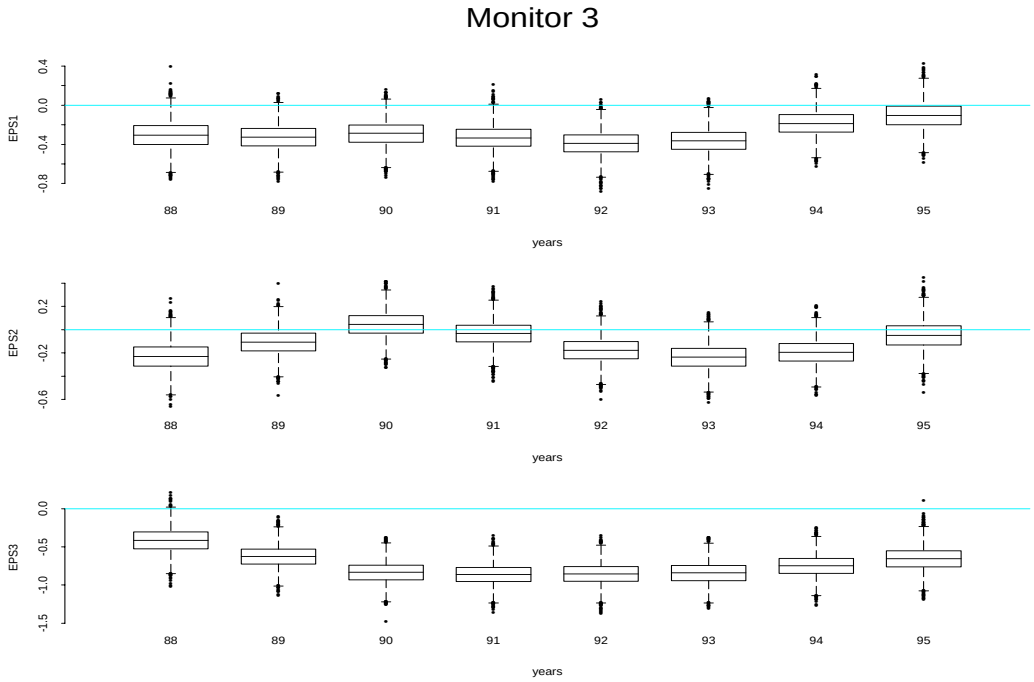

Figure 10.6: Summary interval estimates of hospital-specific random effects ϵ_{it} for three hospitals using the multi-year, single-monitor `MODEL-3`.

The most immediate features relate to the apparent smoothing over the years in the patterns of the relative random effects ϵ_{it} within each monitor. Relative to the corresponding figures from the exploratory analyses under `MODEL-1`, consecutive values of the ϵ_{it} within each hospital are linked by the autoregressive model and, as fairly high positive values of the autoregressive parameters are inferred, this leads to a reasonable degree of shrinkage as each ϵ_{it} is now estimated partly by the data in neighbouring years as well as by that in year t. The resulting smoothing is also evident in the patterns of the absolute levels α_{it} not shown here.

The next section ties three single monitor models together in a general multiple monitor model: this is a binomial/logit model in which hospital-specific random effects are related through time via a multivariate time series model representing the relationships across monitors. The univariate submodels are precisely those already developed in this section; these multivariate models explore and assess aspects of the dependencies among the three monitor series.

10.2 `MODEL-3`: Multivariate Random Effects Time Series Model

The `MODEL-2` class appears to provide satisfactory description of the dependence structure over time in hospital-specific quality measures. The second key goal of this study is to explore possible dependencies in quality monitor outcomes across monitor areas. Hence the interest relies in possible dependencies among the hospital-specific relative effects ϵ_{it} across the three monitors. For instance, $\mathcal{M}1$ and $\mathcal{M}2$ are related areas of psychiatric care, so it is expected that the quality levels are positively correlated between these two monitors. $\mathcal{M}3$ relates to general medical care and so also might be expected to be positively related as representing some form of "overall" quality at the specific hospital. The next step therefore is to tie the three

sets of monitor data together into a model that allows for the investigation of the structure of the joint distribution of sets of relative effects across the three monitors, and of the nature of dependencies in changes over time in these effects.

This new class of models, referred to as `MODEL-3`, considers simultaneously monitors $j = 1, 2, 3$ in each year $t = 1, \ldots, 8$ and for each hospital $i = 1, \ldots, I$. This development is obviously general and can be applied to any numbers of monitors and years. On the three monitors, the observed outcomes $\mathbf{z}_{it} = (z_{i1t}, z_{i2t}, z_{i3t})'$, represent three conditionally independent binomial responses out of totals $\mathbf{n}_{it} = (n_{i1t}, n_{i2t}, n_{i3t})'$ and with "success" probabilities $\mathbf{p}_{it} = (p_{i1t}, p_{i2t}, p_{i3t})'$, respectively. The joint density of these data is

$$p(\mathbf{z}_{it}|\mathbf{n}_{it}, \mathbf{p}_{it}) = \prod_{j=1}^{3} \text{Bin}(z_{ijt}|n_{ijt}, p_{ijt})$$

conditionally independently over hospitals i and years t. The general multiple-monitor hierarchical/random effects model structure of `MODEL-3` is as follows.

Regression and hierarchical/random effects structure

On the logit scale, $\boldsymbol{\mu}_{it} = (\mu_{i1t}, \mu_{i2t}, \mu_{i3t})'$ with

$$\mu_{ijt} \equiv \log(p_{ijt}/(1 - p_{ijt})) = \beta_{0jt} + \beta_{1jt}x_{ijt} + \epsilon_{ijt} + \nu_{ijt} \tag{10.8}$$

for each monitor $j = 1, 2, 3$. Here x_{ijt} is the centered logit transform of the corresponding DRG predicted proportion, i.e., with the mean for year t across all hospitals i within monitor j subtracted. The β parameters are collected in vectors as $\boldsymbol{\beta}_{0t} = (\beta_{01t}, \beta_{02t}, \beta_{03t})'$ and $\boldsymbol{\beta}_{1t} = (\beta_{11t}, \beta_{12t}, \beta_{13t})'$. With the 3×3 design matrices $\mathbf{X}_{it} = \text{diag}(x_{i1t}, x_{i2t}, x_{i3t})$, the model is rewritten in vector form as,

$$\boldsymbol{\mu}_{it} = \boldsymbol{\beta}_{0t} + \mathbf{X}_{it}\boldsymbol{\beta}_{1t} + \boldsymbol{\epsilon}_{it} + \boldsymbol{\nu}_{it},$$

where $\boldsymbol{\epsilon}_{it} = (\epsilon_{i1t}, \epsilon_{i2t}, \epsilon_{i3t})'$ and $\boldsymbol{\nu}_{it} = (\nu_{i1t}, \nu_{i2t}, \nu_{i3t})'$. In each year, marginally, the hospital-specific random effect vectors $\boldsymbol{\epsilon}_{it}$ are conditionally independent over hospitals

i and

$$\boldsymbol{\epsilon}_{it} \sim N(\boldsymbol{\epsilon}_{it}|\mathbf{0}, \mathbf{W})$$

for some constant 3×3 matrix $\mathbf{W}$. This matrix represents the variability in the systematic components of corrected quality levels across the entire hospital population, the related variability in changes in relative quality levels year-to-year, and the dependencies between such quality measures across the three monitors. Notice that essentially arbitrary correlations are admitted across monitors, and inference on the elements of $\mathbf{W}$ will provide for assessment of such cross-monitor dependencies.

This model simply collects together the three single monitor models of the previous section and extends that framework to allow for possible correlation structure across monitors via the multivariate normal distributions for the vectors of relative effects $\boldsymbol{\epsilon}_{it}$. The covariance matrix $\mathbf{W}$, defines and measures the joint variability and dependence structure not only in each year, but also a modified form of the year to year *changes* in the effects.

Time series structure of random effects

The random effects vectors $\boldsymbol{\epsilon}_{it}$ are correlated over years t. The same basic AR(1) models for monitor-specific quantities is maintained, so that the autoregressive parameters are now indexed by monitor index j too. Then the three univariate AR(1) models combine in the vector autoregression

$$\boldsymbol{\epsilon}_{it} = \Phi\boldsymbol{\epsilon}_{i,t-1} + \boldsymbol{\omega}_{it}$$

where $\Phi = \text{diag}(\phi_1, \phi_2, \phi_3)$ is the diagonal matrix of monitor-specific autoregressive coefficients. The vector innovations $\boldsymbol{\omega}_{it}$ are conditionally independent over time, and normally distributed with

$$\boldsymbol{\omega}_{it} \sim N(\boldsymbol{\omega}_{it}|\mathbf{0}, \mathbf{U}) \tag{10.9}$$

for some variance matrix $\mathbf{U}$ to be estimated. Now the normal distribution $\boldsymbol{\epsilon}_{it} \sim N(\boldsymbol{\epsilon}_{it}|\mathbf{0}, \mathbf{W})$ is the yearly margin under this vector AR(1) model. This implies that $\mathbf{W}$ satisfies $\mathbf{W} = \Phi\mathbf{W}\Phi + \mathbf{U}$, so that correlation patterns in $\mathbf{U}$ and $\mathbf{W}$, while similar, depend on the autoregressive parameters. In particular, for each monitor pair j, h the covariance elements are $\mathbf{W}_{jh} = \mathbf{U}_{jh}/(1 - \phi_j\phi_h)$. The DRG-corrected hospital-specific level parameters are now given by

$$\boldsymbol{\alpha}_{it} = \boldsymbol{\beta}_{0t} + \boldsymbol{\epsilon}_{it}$$

for each i, and t. These follow the centered VAR(1) model

$$\boldsymbol{\alpha}_{it} = \boldsymbol{\beta}_{0t} + \Phi(\boldsymbol{\alpha}_{i,t-1} - \boldsymbol{\beta}_{0,t-1}) + \boldsymbol{\omega}_{it} \tag{10.10}$$

for $t > 1$, and have yearly margins $N(\boldsymbol{\alpha}_{it}|\boldsymbol{\beta}_{0t}, \mathbf{W})$. Note again that the single monitor `MODEL-2` structures are embedded here with the addition of the cross-monitor structure through $\mathbf{W}$ (equivalently $\mathbf{U}$). In addition, note that (10.10) has the same structure used in (7.4) to model the log volatilities in the dynamic factor model, VAR(1) with correlated innovations.

Residual components

Residual, unpredictable variations in the binomial probabilities across hospitals and years is described by the residual terms

$$\boldsymbol{\nu}_{it} \sim N(\boldsymbol{\nu}_{it}|\mathbf{0}, \mathbf{V}) \tag{10.11}$$

with monitor-specific variances v_1^2, v_2^2 and v_3^2 on the diagonal of the matrix $\mathbf{V}$, and now admitting cross-monitor dependencies through the covariance elements of $\mathbf{V}$. The idea is to simply collect the monitor-specific residual parameters together in a vector for the multiple monitor model, and add covariance structure that permits the estimation of cross-monitor dependencies.

In this case, the addition of the residual/noise terms $\boldsymbol{\nu}_{it}$ to the VAR(1) process $\boldsymbol{\epsilon}_{it}$ modifies the correlation structure giving a VARMA(1,1) model with $N(\boldsymbol{\epsilon}_{it}+\boldsymbol{\nu}_{it}|\mathbf{0}, \mathbf{W}+\mathbf{V})$ yearly margins. It is important to note that the overall levels of random effects variability, and the associated overall measures of cross-monitor dependencies, are represented through the yearly marginal variance matrix $\mathbf{W}+\mathbf{V}$. In summaries of analyses below, posterior inferences on elements of $\mathbf{W}+\mathbf{V}$, as well as the component matrices separately are explored. It may be suggested that correlation patterns in $\mathbf{V}$ might be expected to be similar to those in the time series random effects modelled through $\mathbf{W}$ (and/or $\mathbf{U}$). The current model does not anticipate this, leaving $\mathbf{V}$ and $\mathbf{W}$ unrelated *a priorí,* but the framework obviously permits the assessment of potential similarities in posterior inferences.

Prior distributions

Inference is based on posterior distributions for all model parameters and random effects under essentially standard reference/uninformative priors for: (a) the annual population parameters $\boldsymbol{\beta}_{0t}$ and $\boldsymbol{\beta}_{1t}$, (b) the population residual variance matrix $\mathbf{V}$, and (c) the variance-covariance matrix $\mathbf{U}$; the prior is completed with independent uniform priors for the autoregressive parameters ϕ_j on (0,1). The implementation of the model is performed using a customized Gibbs sampling algorithm detailed in section 10.6 of this thesis.

10.3 Results for the VA

Summaries of analysis of the three monitor series combined in this multivariate, multi-year model appear in the Figures 10.7 to 10.13 inclusive. The formats of figures follow those of the earlier models, with additional inferences about the correlation patterns in $\mathbf{W}$ displayed as boxplots of posteriors as usual. The results relative to

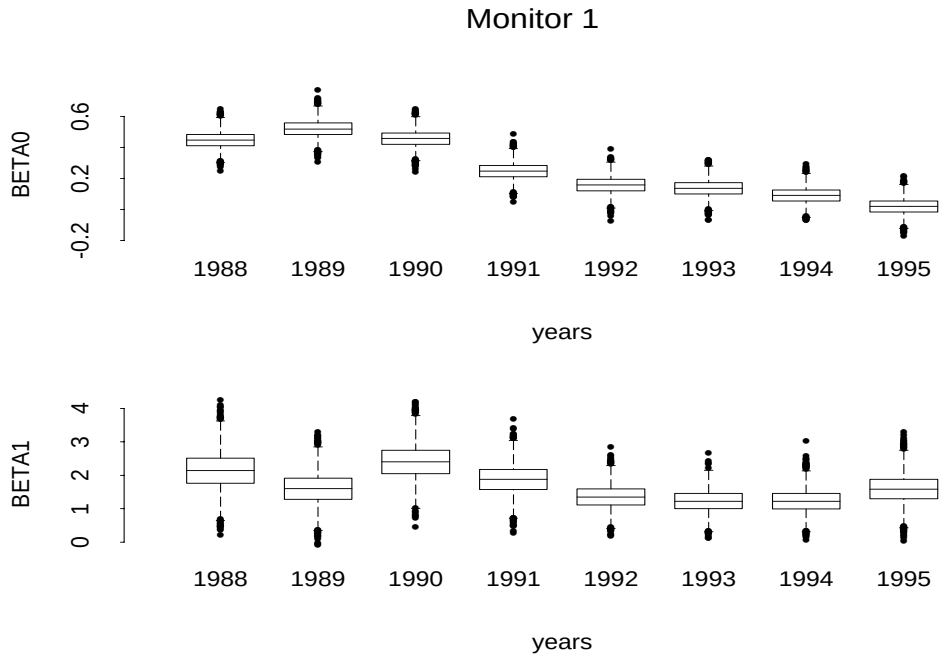

Figure 10.7: Summary interval estimates of β_{0t} and β_{1t} using the multi-year, multiple-monitor `MODEL-3`.

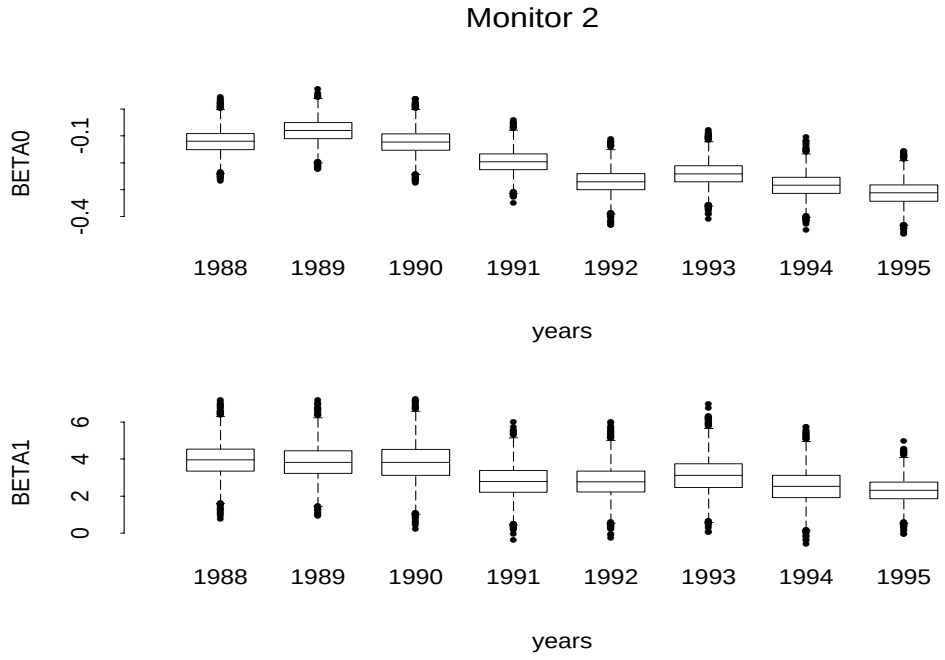

Figure 10.8: Summary interval estimates of β_{0t} and β_{1t} using the multi-year, multiple-monitor `MODEL-3`.

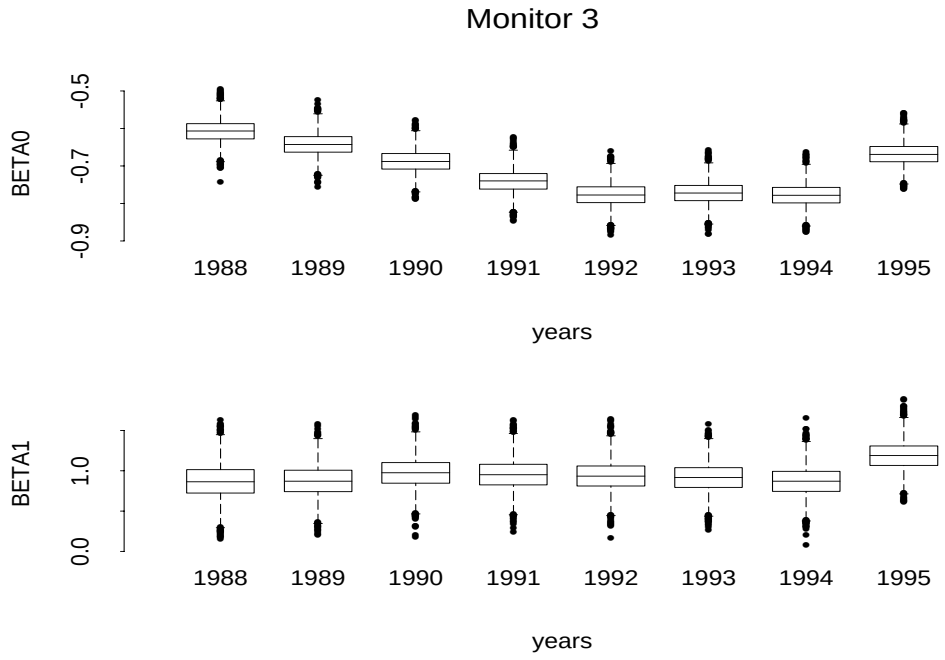

Figure 10.9: Summary interval estimates of β_{0t} and β_{1t} using the multi-year, multiple-monitor `MODEL-3`.

the inferences deduced from the `MODEL-2` analysis show the following features.

- There are only very minor changes evident in the posterior distributions for population parameters β_{0t} and β_{1t} across years t. See Figures 10.7, 10.8 and 10.9.

- There are similarly almost no changes evident in the posterior distributions for the variances v_j^2 of the residual noise terms ν_{ijt}. The posterior distributions of the standard deviations and correlations are displayed in the second frame of Figure 10.13.

- There are similarly only very minor changes evident in the posterior distributions for random effects ϵ_{ijt} over the years for the three selected hospitals, 41, 92 and 2. See Figures 10.10, 10.11 and 10.12 for $\mathcal{M}1, \mathcal{M}2$ and $\mathcal{M}3$ respectively.

- Figure 10.13 provide summaries of the marginal posteriors for the three autore-

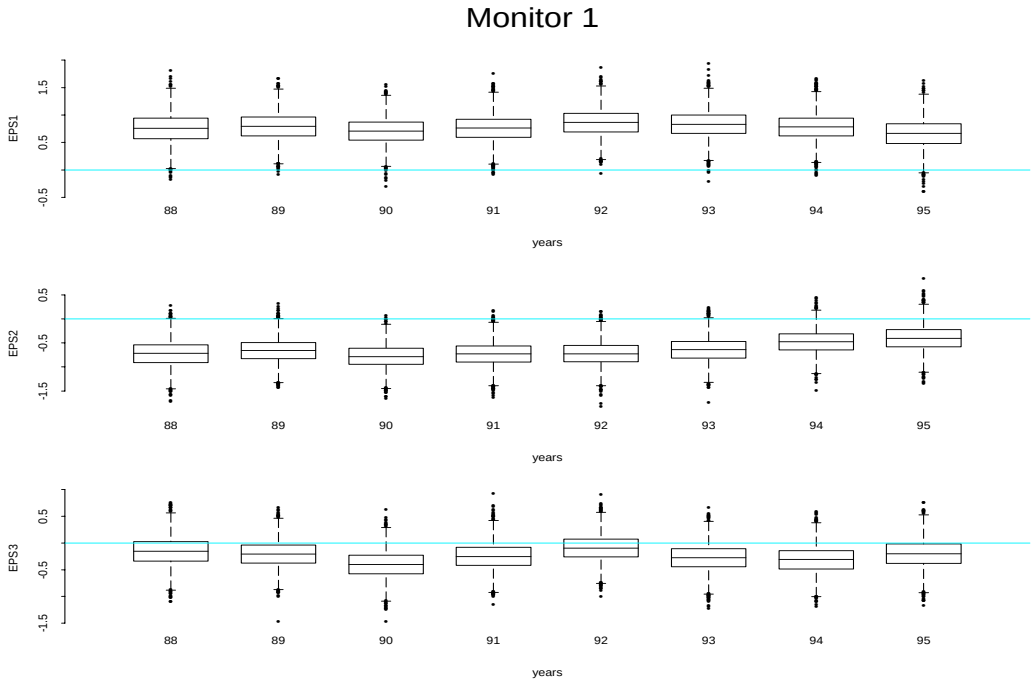

Figure 10.10: Summary interval estimates of hospital-specific random effects ϵ_{it} for three hospitals using the multi-year, multiple-monitor `MODEL-3`.

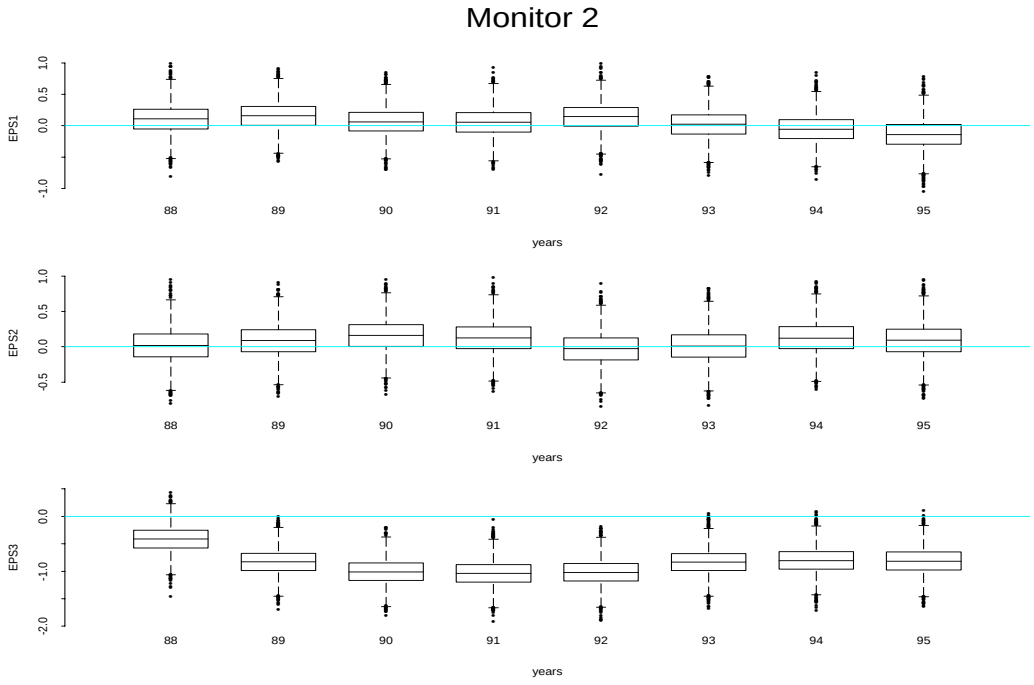

Figure 10.11: Summary interval estimates of hospital-specific random effects ϵ_{it} for three hospitals using the multi-year, multiple-monitor `MODEL-3`.

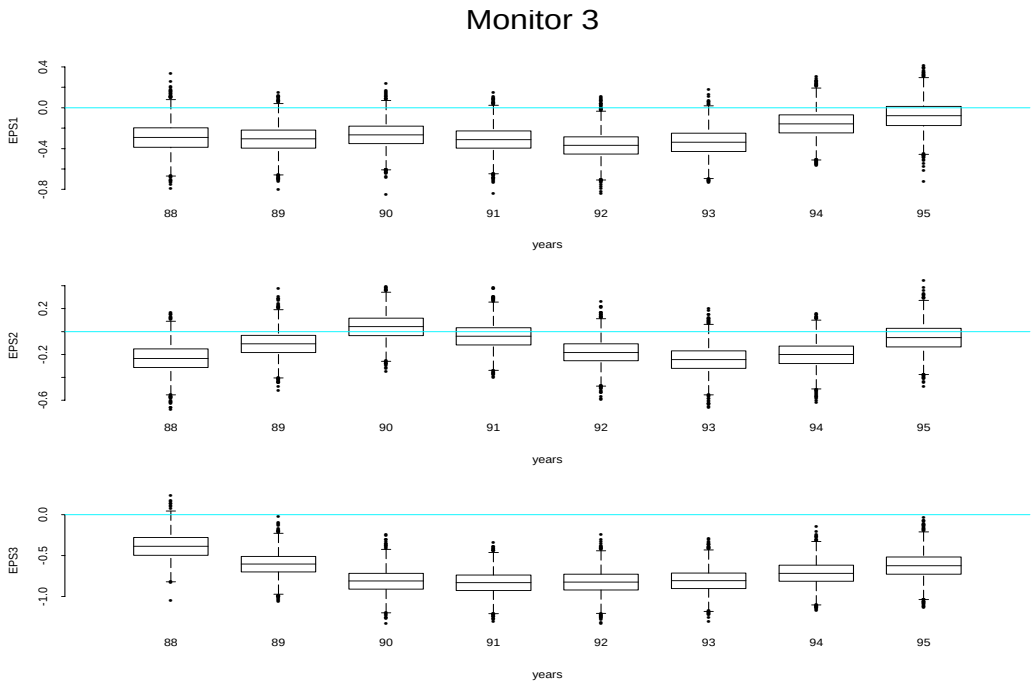

Figure 10.12: Summary interval estimates of hospital-specific random effects ϵ_{it} for three hospitals using the multi-year, multiple-monitor `MODEL-3`.

gressive parameters in Φ. These are essentially similar to those delivered in the individual `MODEL-2` analyses, as might be expected.

- Figure 10.13 also provide summaries of the marginal posteriors for elements and functions of the variance matrices $\mathbf{W}$ and $\mathbf{V}$. First, note that the posteriors displayed for the standard deviations under both $\mathbf{W}$ and $\mathbf{V}$ are essentially as in the separate `MODEL-2` analyses. The new parameters here are the correlations between the monitor effects, both in the systematic component $\mathbf{W}$ and in the residual component $\mathbf{V}$. Posterior boxplots are displayed here for the correlations in both $\mathbf{W}$ and $\mathbf{V}$, and, most importantly, in the combined variance $\mathbf{W}+\mathbf{V}$. As mentioned above, it is this latter matrix that most fully described cross-monitor structure within each year. The lower left frame of Figure 10.13 summarizes inferences on the standard deviation and correlation elements of $\mathbf{W}+\mathbf{V}$. This indicates fairly low overall correlations, consistent with the generally low cor-

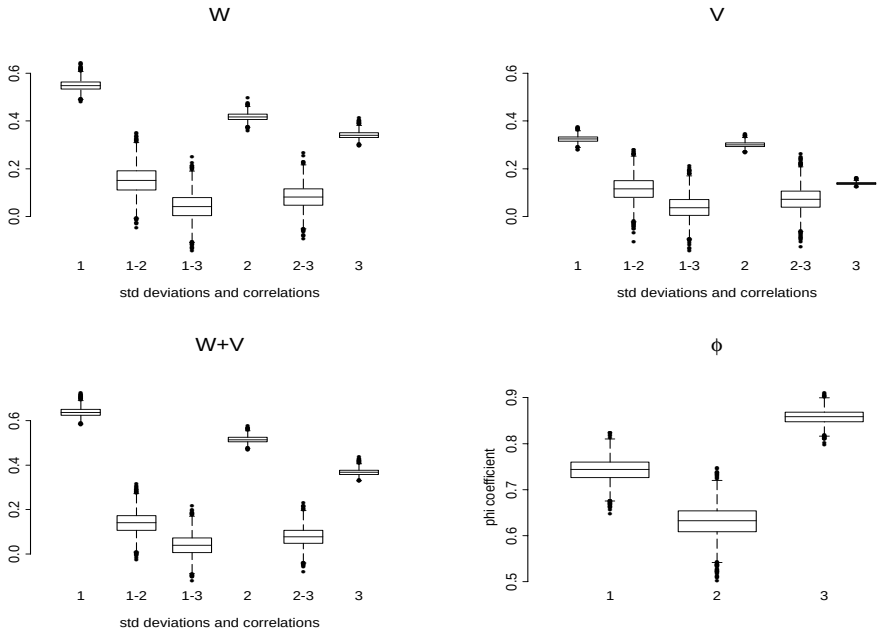

Figure 10.13: Summary interval estimates of population parameters for `MODEL-3`, including standard deviations and correlations of the variance matrices $\mathbf{W}, \mathbf{V}$ and the overall $\mathbf{W} + \mathbf{V}$ and the AR coefficients $\mathbf{\Phi}$.

relations exhibited by each of $\mathbf{W}$ and $\mathbf{V}$ separately. Denoting posterior means by "hats" and writing $\mathbf{E}$ for the column eigen vector matrix of $\hat{\mathbf{W}} + \hat{\mathbf{V}}$, then

$$\hat{\mathbf{W}} + \hat{\mathbf{V}} = \begin{pmatrix} 0.417 & 0.044 & 0.009 \\ 0.044 & 0.271 & 0.014 \\ 0.009 & 0.014 & 0.136 \end{pmatrix}, \ \mathbf{E} = \begin{pmatrix} 0.961 & -0.275 & -0.015 \\ 0.273 & 0.957 & -0.097 \\ 0.041 & 0.089 & 0.995 \end{pmatrix}.$$

This indicates correlations between $\mathcal{M}1$ and $\mathcal{M}2$ of around 0.132, between $\mathcal{M}1$ and $\mathcal{M}3$ of around 0.036; and between $\mathcal{M}2$ and $\mathcal{M}3$ of 0.073. These are quite low correlations, though they do support the earlier suggestion that the correlation between $\mathcal{M}1$ and $\mathcal{M}2$ might be higher than that between either $\mathcal{M}1$ and $\mathcal{M}3$, or between $\mathcal{M}2$ and $\mathcal{M}3$, in view of the care areas of origination. The eigenvalues of $\hat{\mathbf{W}} + \hat{\mathbf{V}}$ are roughly $0.43, 0.26$ and 0.13; so the principal components explain roughly 52%, 32% and 16% of variation described by this estimated variance matrix, indicating that all three vary appreciably and that

no reduction to two or fewer seems appropriate. Posterior uncertainty about the variance matrices, and the eigenstructure, does not materially impact these qualitative conclusions. To exemplify this, the full posterior sample produces the following approximate posterior means and 95% intervals for the three eigenvalues of $\mathbf{W}+\mathbf{V}$: 0.42 (0.38-0.48), 0.25 (0.22-0.29) and 0.13 (0.11-0.16), closely comparable to the estimates quoted above. Evidently, the eigenvector matrix is dominated by the diagonal terms, and all three are close to unity; note that the eigenvector matrix would be the identity were the monitors uncorrelated. The first column represents an average monitor effect, dominated by $\mathcal{M}1$ and $\mathcal{M}2$ and essentially excluding $\mathcal{M}3$; this may be viewed as a psychiatric care component alone. The second column represents a contrast between $\mathcal{M}1$ and $\mathcal{M}2$, again essentially excluding $\mathcal{M}3$; the final column almost wholly represents $\mathcal{M}3$ alone, and to the extent that the coefficients for $\mathcal{M}1$ and $\mathcal{M}2$ are non-ignorable, contrasts these two psychiatric care monitors with the general medical monitor $\mathcal{M}3$.

Investigating conditional distributions provides additional insights into the levels and nature of dependencies. For example, consider the relevance of dependencies by looking at questions like *"how much might the $\mathcal{M}2$ effect change if the $\mathcal{M}1$ effects changes by e?"*. Various such exercises have been undertaken, and simply confirm that the levels of correlations inferred in our analyses are so low as to really limit their impact on predictive questions, especially in the context of the realistic levels of posterior uncertainties about the population parameters, the $\beta_{0t}, \epsilon_{ij,t-1}$ and so forth. Hence, this level of correlation structure between monitors, though certainly non-negligible, is relatively minor. The story may, however, be quite different with other collections of monitors.

- Finally, Figures 10.14, 10.15 and 10.16 display twenty randomly chosen sets of

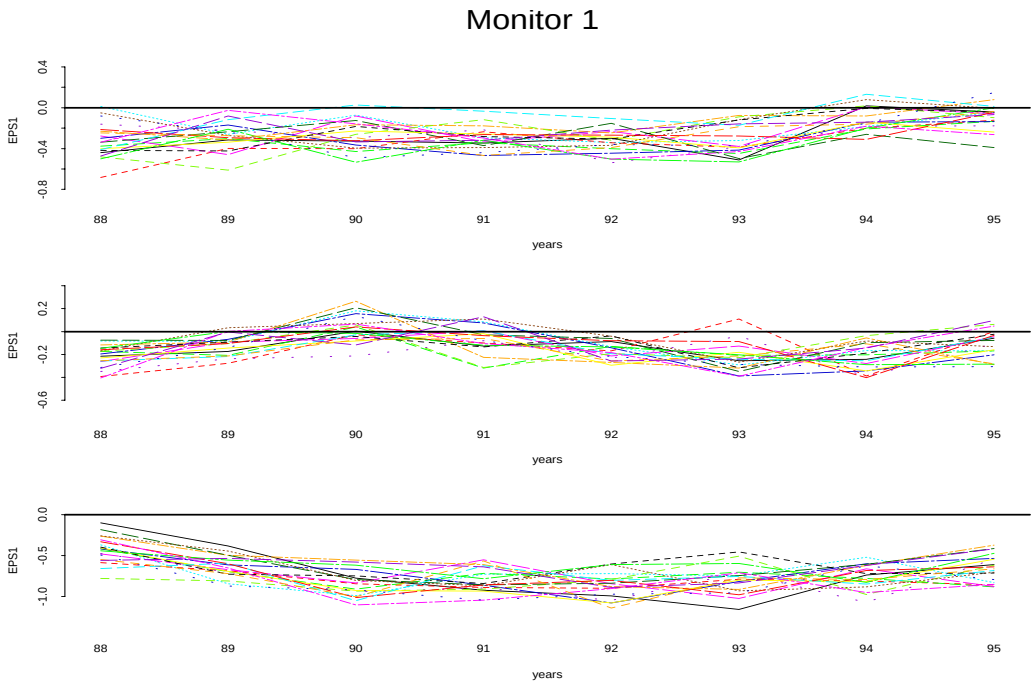

Figure 10.14: Selected posterior samples for hospital-specific random effects ϵ_{ijt} on monitor $\mathcal{M}1$ in hospitals 41, 92 and 2.

posterior sampled values for the relative random effects $\boldsymbol{\epsilon}_{it}$ for hospitals 41, 92 and 2. These summaries and examples the kinds of patterns of variation exhibited by the random effects individual hospitals. In addition these graphs highlight the smooth, systematic dependence structure over time discussed above.

10.4 Model Assesment

Residual structure analysis

Additional graphs in Figures 10.17 to 10.18 inclusive are illustrative of graphical assessments of model adequacy via more-or-less standard Bayesian "residual analysis". The primary concern is to identify and investigate any apparent residual structure in the data following model fitting. Referring to the approximate logistic normal version of the binomial model in section 9.3, note that the model implies approximate

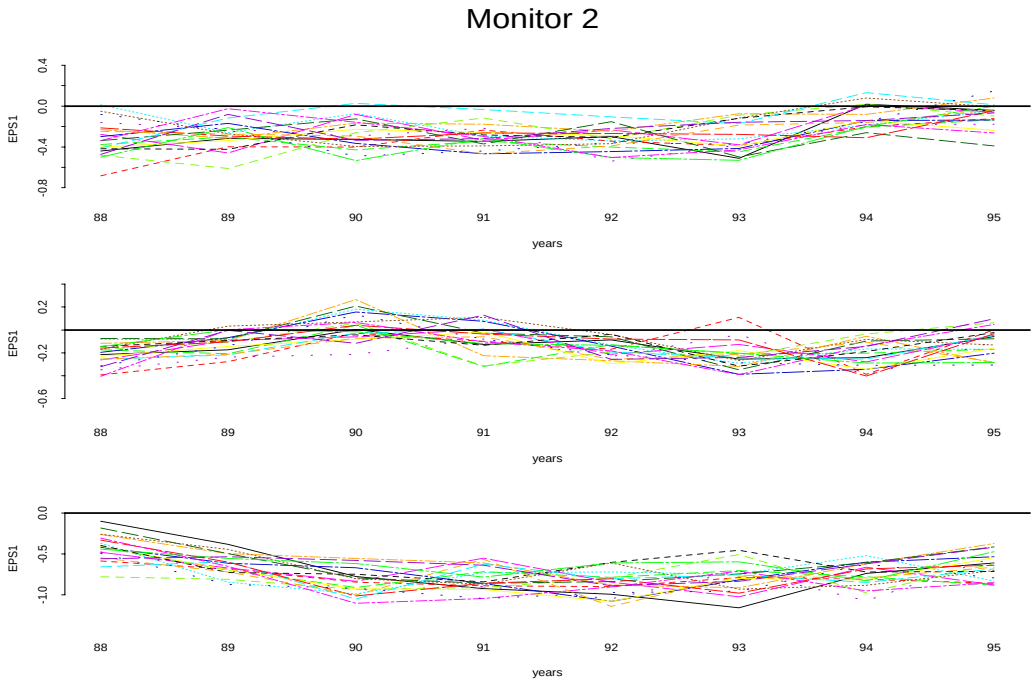

Figure 10.15: Selected posterior samples for hospital specific-random effects ϵ_{ijt} on monitor $\mathcal{M}2$ in hospitals 41, 92 and 2.

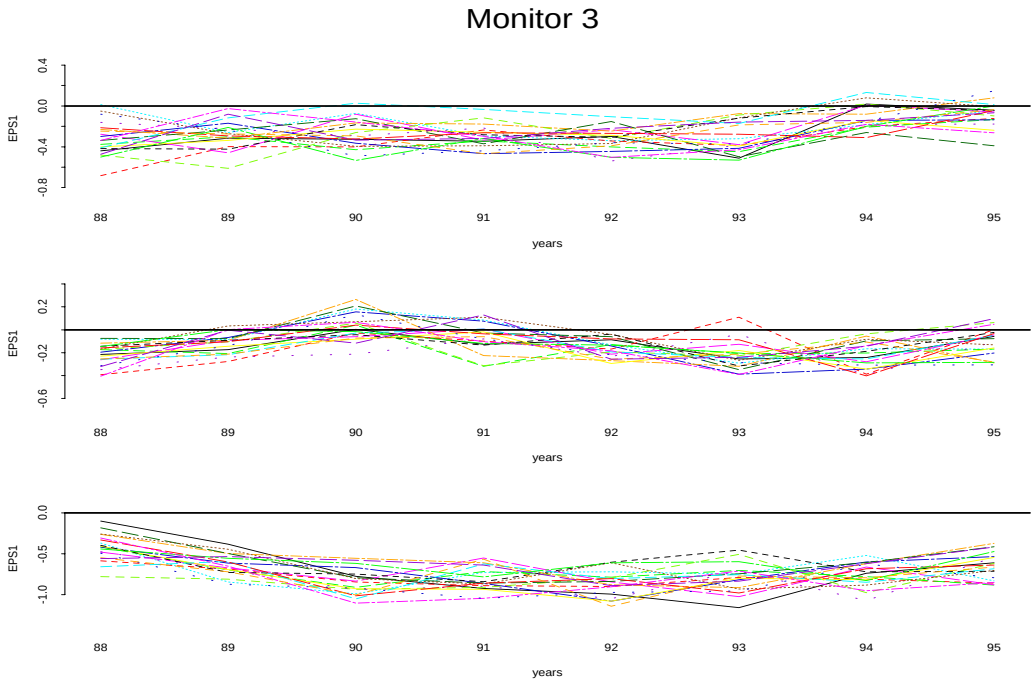

Figure 10.16: Selected posterior samples for hospital specific-random effects ϵ_{ijt} on monitor $\mathcal{M}3$ in hospitals 41, 92 and 2.

(standard) normality and conditional independence of the standardized data residuals $e_{it} \equiv (y_{it} - \mu_{it})/s_{it}$ where, extending the subscripts to the current model, y_{it} is the logit transform of the observed outcome proportion z_{it}/n_{it} and s_{it} the corresponding approximate standard deviation, for each i, t. In the simulation-based analysis of the exact binomial model here, the full joint posterior distribution for all model parameters and random effects are repeatedly sampled. This means that a sequence of samples of the set of μ_{it} is also sampled, which represent draws from their posterior distribution. These lead to trivially computed values of the e_{it} which similarly represent samples from the posterior distribution of the standardized data residuals. Using any one of these samples of residuals numerical and graphical measures of concordance with, or departures from, the theoretical normal distribution are computed, and so model fit is assessed. The graphs in Figures 10.17 and 10.18 display some graphs using just two sets of sampled residuals on monitor $\mathcal{M}2$ in year $t = 1995$, the final year of the data. These are representative of all monitor/year residuals and repeated sampling from the posteriors produces very similar graphs. For these samples, the plots include a simple graph against hospital number, a graph against the corresponding sample sizes n_{it}, a normal quantile plot and a simple histogram. None of these graphs indicate any kind of meaningful departures from normality, and as this is maintained across residual samples, it provides support for model adequacy.

More formally, the graphs in Figure 10.19 summarize the posterior distributions of the actual data residuals e_{it} for monitor $\mathcal{M}2$ in year $t = 1995$. The first row of Figure 10.19 displays graphical summaries of this posterior distribution in terms of marginal posterior 95% intervals for each hospital, with posterior means marked. The hospitals are ordered here according to the posterior medians of the underlying outcome probabilities p_{it} on this monitor in this year. There are no outlying hospitals or other worrisome features evident in this display. This comforting conclusion is

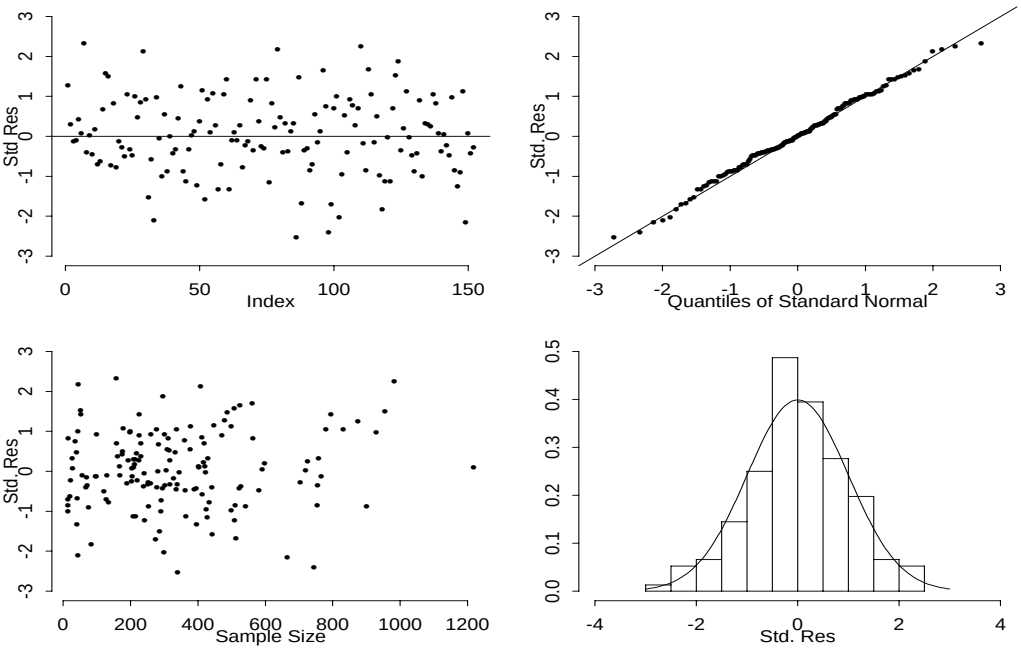

Figure 10.17: Summaries of a randomly chosen draw from the posterior distribution of the observed residuals e_{it} for all hospitals on monitor $\mathcal{M}2$ in 1995.

supported by additional displays of aspects of the posterior distribution of the *ordered* observation residuals. The bottom pictures of Figure 10.19 display a normal quantile plot and a histogram of the posterior means of the *ordered* observation residuals for all hospitals i. The curve superimposed on the histogram is the standard normal density function. The quantile plot includes vertical lines representing approximate posterior 95% intervals showing the uncertainty in the marginal posteriors for the ordered residuals. The conformity with normality here is excellent.

The next section describes some summary inferences for all hospitals on one monitor in one year to illustrate additional possible uses of the models and further aspects of model fit.

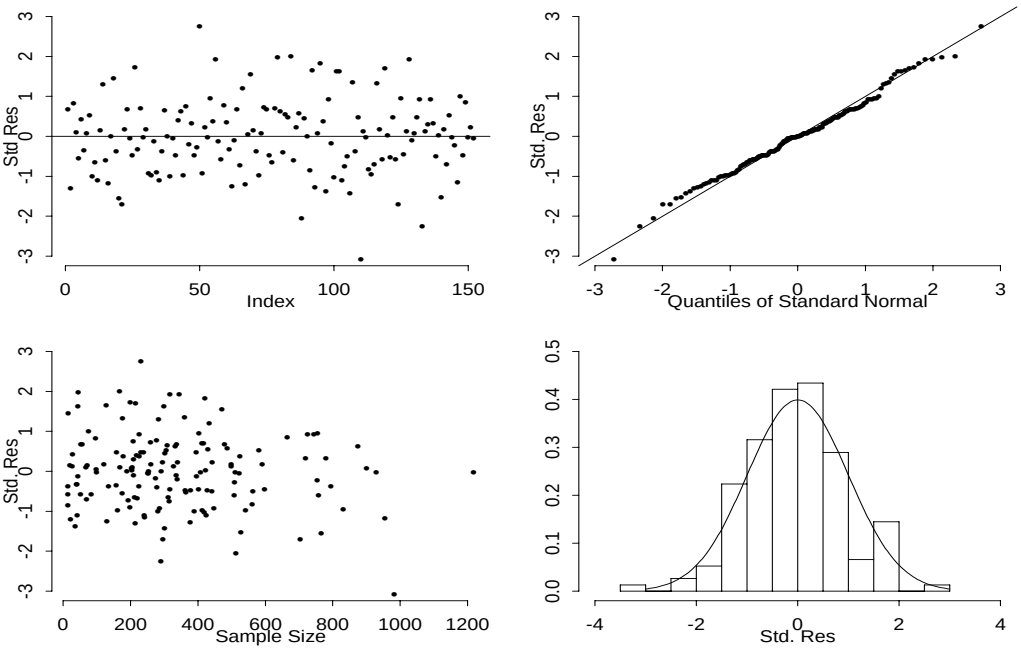

Figure 10.18: Summaries of a second randomly chosen draw from the posterior distribution of the observed residuals e_{it} for all hospitals on monitor $\mathcal{M}2$ in 1995.

10.4.1 Summary Inferences for Monitor $\mathcal{M}2$ in 1995

To illustrate the possible additional uses of these models in exploring the variability in outcomes across the hospital system, we have considered only outcomes on only Monitor $\mathcal{M}2$ in the last year of the data, 1995. The analysis produces simulation-based descriptions of the joint posterior distribution for all $\mathcal{M}2/1995$ parameters, namely the full set of quantities

$$\beta_0, \beta_1, \{\epsilon_i, \nu_i\}$$

for $i = 1, \ldots, I$ and where the subscripts for monitor ($j = 2$) and year ($t = 1995$) are dropped for clarity here. So β_0 and β_1 are the hospital population level and DRG-regression coefficient in 1995, and the ϵ_i and ν_i are the systematic and residual random effects for hospital i on $\mathcal{M}2$ in this year. Refer to equation (10.8) and suppress the subscripts j and t. The particular analysis summarized is based on a posterior sample with Monte Carlo sample size 5,000 for all these quantities; see section 10.6

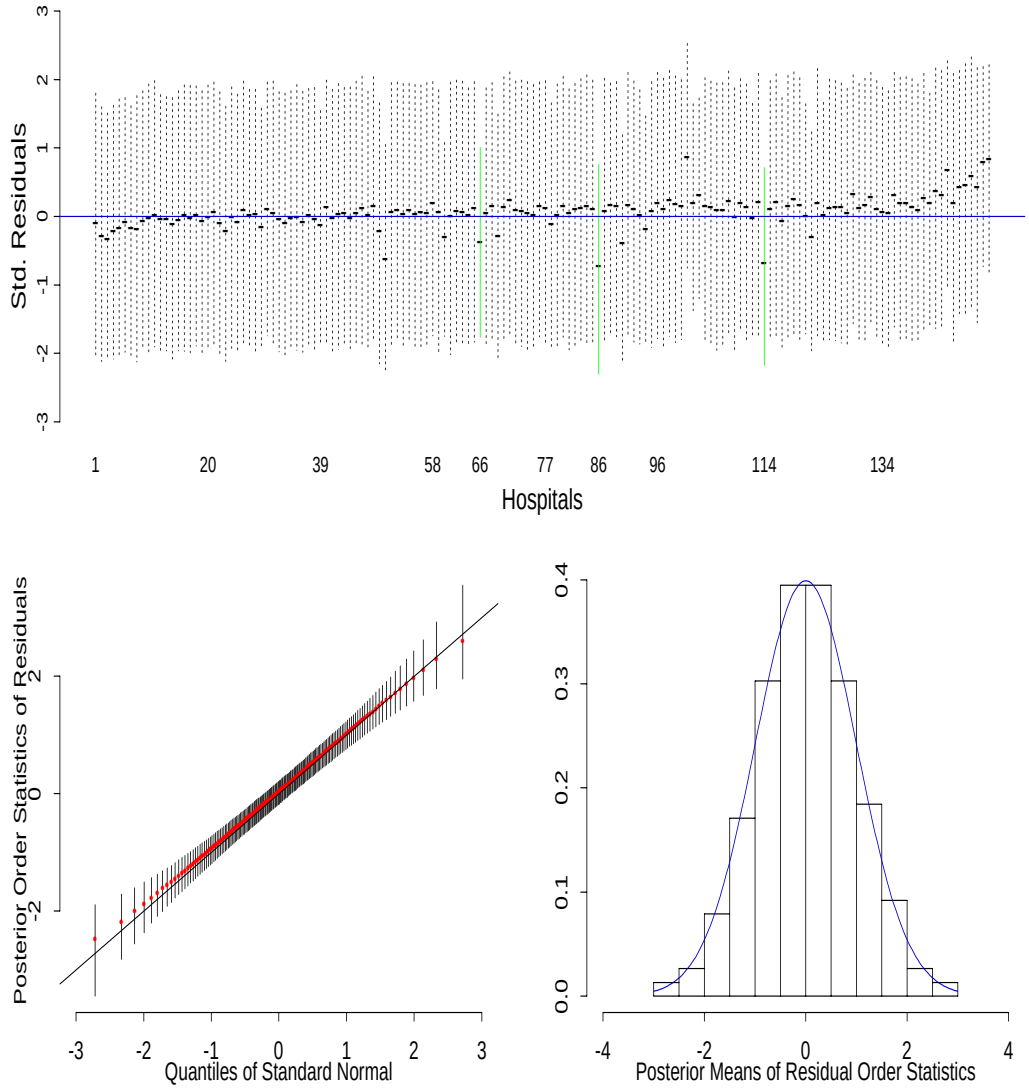

Figure 10.19: Top: Posterior 95% intervals with means marked for the observation residuals e_{it} for all hospitals on monitor $\mathcal{M}2$ in 1995. Bottom: Normal quantile plot with 95% intervals (left) and histogram of the posterior means of the ordered observation residuals (right).

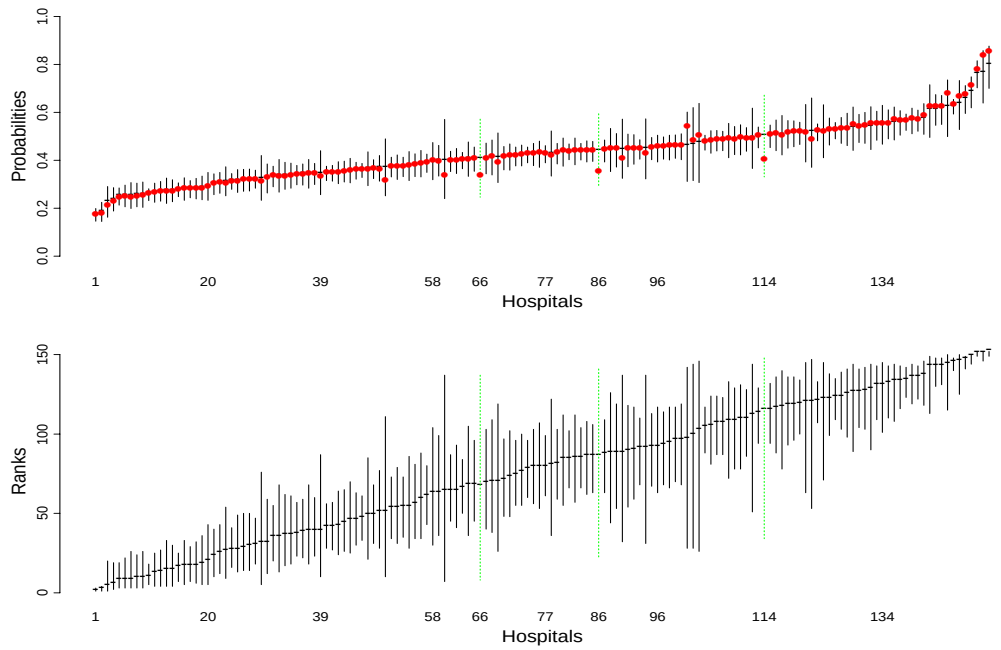

Figure 10.20: Posterior 95% intervals for outcome probabilities p_i (top) and corresponding ranks (bottom) on monitor $\mathcal{M}2$ in 1995 for all hospitals ordered by posterior medians of p_i.

for details of the implementation.

Absolute outcome levels and comparisons across hospitals

For each hospital i, the corresponding posterior sample for the actual quality outcome probability are computed, $p_i = 1/(1 + \exp(-\theta_i))$ where $\theta_i = \beta_0 + \beta_1 x_i + \epsilon_i + \nu_i$. This was done, followed by calculation of posterior medians and approximate 95% (equal tails) intervals for each p_i. These intervals, with medians marked, appear in the top frame of Figure 10.20. The hospitals are ordered by the posterior medians so computed, and labeled by actual station numbers. The small number of hospitals with low sample sizes in $\mathcal{M}2$ are identifiable here as they have wider intervals than the majority; the width of the interval reflects the spread of the marginal posterior which is a decreasing function of sample size.

If interest lies in questions about thresholds for the p_i, they can be directly ad-

dressed using these posterior distributions. The probability that any p_i exceeds or lies below any specified threshold can be immediately deduced from the posterior for p_i. A crude version of this across hospitals would involve simply drawing a horizontal line at the specified thresholds on the graphs in top frame of Figure 10.20. Superimposed on the intervals are the values of the observed outcome proportions z_i/n_i, for comparison. Note that most of the observed outcomes are well within the interval and often very close to the median of the corresponding p_i. This partly reflects the large sample sizes n_i in most cases that lead to very influential observations and so fitted values that will be close to observed outcomes. The model is not, however, over-fitting, as is confirmed by the earlier discussions of residual structure analyses. Note, however, that the picture needs modification to more adequately represent the fit; these intervals are summary estimates of the true outcome levels p_i, and so the variability expressed by the width of the intervals does not incorporate the natural binomial variability in the outcomes around the levels. To do this, additional samples of binomial outcomes are computed by drawing conditionally binomial counts $z_i^* \sim Bin(z_i^*|n_i, p_i^*)$ for each i, and where p_i^* represent the set of posterior draws for p_i. This produces the posterior predictive distribution for actual outcomes at each hospital. These can be summarised in the same way as the p_i, in terms of interval estimates for the actual outcomes; the resulting 95% intervals with medians marked appear in the top frame of Figure 10.21. Again, the actual data are superimposed, and it is now clear that the additional binomial variability broadens the intervals so as to very adequately cover the observed data.

This latter graph helps to immediately identify outcomes that are rather extreme compared to their posterior predictive distributions. These tend to be at hospitals with low sample sizes, and some further investigation of specific hospitals flagged in this graph, and also noted in the following graphs, may be of interest. For example,

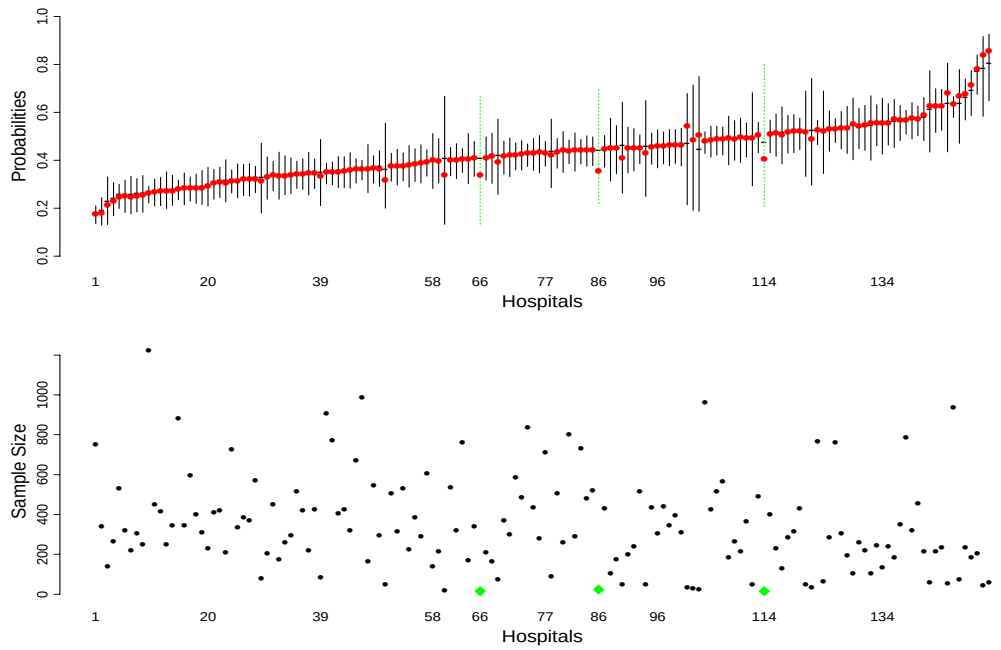

Figure 10.21: Posterior 95% intervals for actual outcome proportions (top) and observed sample sizes n_i (bottom) on monitor $\mathcal{M}2$ in 1995 for all hospitals.

stations numbered 66, 86 and 114 are the three hospitals with observed outcome proportions most over-predicted, and cases 101, 152 and 151 those most under-predicted. These are all cases for which the sample sizes n_i are relatively small, as is seen from the bottom plot of Figure 10.21. Of the three over-predicted, the DRG-predicted proportions are very large relative to the majority of the hospitals, as is seen in the plots of Figure 10.22. The model naturally adapts to these lower observed outcomes for such cases, as is evident in the top graph in Figure 10.23. This displays approximate 95% posterior intervals for the combined random effects $\epsilon_i + \nu_i$ for each hospital i. These are plotted on this logit scale in a format similar to the top picture in Figure 10.20 and with hospitals plotted in the same order for comparison. Note that these "low" hospitals have random effects lower than average, indicating that the model has indeed adapted to the extreme observations. The intervals are relatively wider for these cases, indicative of the smaller sample sizes. This adaptation is, however,

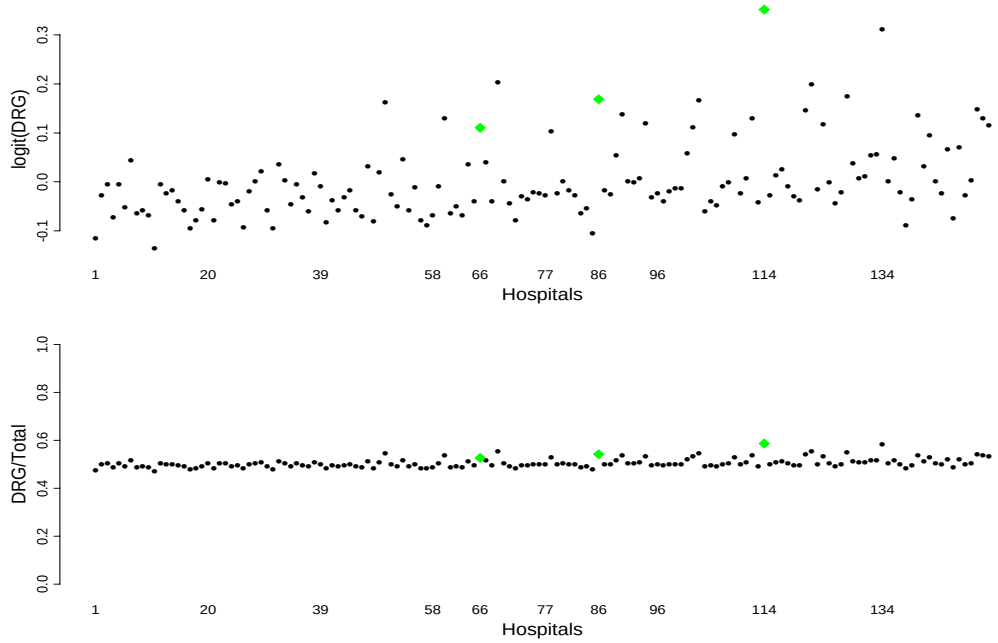

Figure 10.22: Observed value of DRG predicted proportion (top) and corresponding logits (bottom) on monitor $\mathcal{M}2$ in 1995 for all hospitals.

constrained by the model form and also by the the high, and hence influential, values of the corresponding DRG predictor. The basic issue is that the DRG predictor for these cases seems too high, a suggestion that might be the subject for further investigation and study. Nevertheless, the fit of the model to even these very extreme cases is very good, consistent with earlier subjective assessment of graphs of sampled data residuals.

Relative outcome levels and comparisons across hospitals

To further investigate the relative performance of hospitals, the hospital-specific departures from population levels are examined, i.e., the combined random effects $\epsilon_i + \nu_i$ for each i. As noted above, the top picture of Figure 10.23 displays approximate 95% posterior intervals for each of these quantities for all hospitals, with the hospitals in the order chosen for the top picture of Figure 10.20 for comparison. There is a

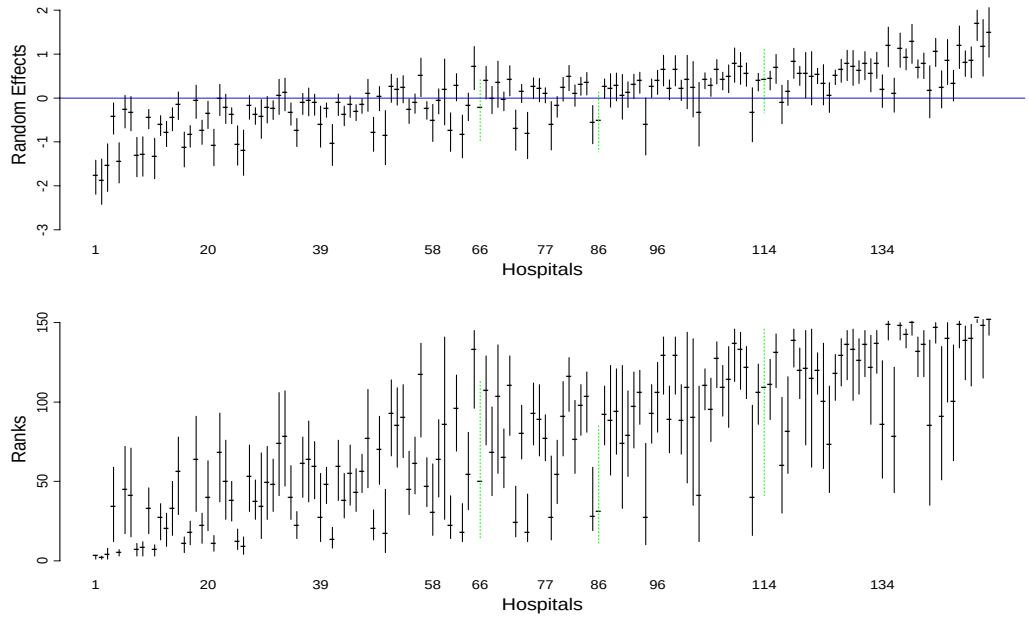

Figure 10.23: Posterior 95% intervals for combined random effects $\epsilon_i + \nu_i$ (top) and corresponding ranks (bottom) on monitor $\mathcal{M}2$ in 1995 for all hospitals.

general increasing trend in the random effects through the hospitals in this order, consistent with the ordering by outcome probabilities. The pattern is not monotonic, however; the probabilities include the effects of the DRG-based predictor, and the current graph focuses exclusively on the relative performance levels free from the regression effects. The bottom frame in Figure 10.23 displays related posterior intervals, in the same hospital order evaluating 95% posterior intervals for the *ranks* of the hospitals in terms of increasing levels of $\epsilon_i + \nu_i$. Evidently, the four or five hospitals with the highest *estimated* outcome probabilities have very high ranks, indicating that their *true* outcome probabilities are indeed very likely to be among the largest few across the system. Similar comments apply to those with the lowest estimated probabilities. Among the majority of the hospitals, however, there are much higher uncertainties about rankings, with posterior intervals spanning fairly wide ranges. This way of investigating relative performance should be contrasted with that based

on the outcome probabilities themselves; the bottom frame of Figure 10.20 displays the plot of posterior intervals for rankings of the p_i displayed in the top picture of that same Figure, which gives a very different picture and resulting inferences than that based on the rankings of effects $\epsilon_i + \nu_i$. The differences are important as the two ranking graphs address very different questions: Figure 10.20 summarises absolute performance, impacted by patient-mix and other confounding factors; Figure 10.23 represents relative quality levels once these factors are accounted for through the model, and a firmer basis for assessing relative performance due to hospital-specific policies and practices. This is evident in the cases of hospitals 66, 86 and 114 noted above, for which appropriately lower rankings are indicated in the lower frame of Figure 10.23 than in the "unadjusted" rankings in the lower frame of Figure 10.20. Even then, there is high uncertaintry about rankings for most hospitals, not only those with small sample sizes, reflecting the inherent difficulties in ranking now well understood in this and other areas.

The posterior distributions summarised here can be explored in various other ways. Subsets of hospitals can be selected for further such summary and deeper investigation, for example. Similarly, questions about changes in hospital-specific effects can be addressed by looking at interval estimates of the parameters measuring changes year to year. Much can be extracted from posterior analysis that has not been illustrated here.

10.5 Conclusions

The key summary conclusions arisisng from this new class of multiple monitor, hierarchical random effects time series model are as follows.

- The inferred patterns of change over time in the key population levels β_{0jt} and DRG regression coefficients β_{1jt} are essentially the same in all models explored.

There are evident changes in β_{0jt} that require consideration and interpretation by VA personnel. There are differences over the years in the regression coefficient β_{1jt} too, though these changes are less marked.

- The multiple monitor time series models isolate changes over time and dependencies among such changes in the hospital-specific random effects across the three monitors. Though dependencies across monitors exist, they are apparently quite small. There are negligible differences in inferences about the key population quantities/parameters in extending from from the single monitor `MODEL-2` framework to the linked multiple monitor `MODEL-3` framework, at least for these three monitors. With other sets of monitors or other contexts, dependence patterns may be stronger and then the `MODEL-3` framework of more interest. There will be bigger and possible meaningful differences in inferences about hospital-specific random effects in cases of very high correlations between monitors, and in cases when the sample sizes n_i are much smaller than is typical with these three monitors.

- A critical feature of this work has been the identification of several components of variability underlying within-year variation and across-year changes in observed quality levels. The important part of this is the partitioning the hospital-specific variation in outcome probabilities into two components: one, a partially systematic and positively dependent AR component, represented by the ϵ_{ijt}; and two, the purely unpredictable component represented by the ν_{ijt}. In each monitor separately, and with confirmation from the multi-monitor studies, the latter are very significant, and contribute between 15-30% of the total random effects variance on the logit scale. The extent of the contribution is monitor-specific, with that for the *general medical discharge* monitor $\mathcal{M}3$ being significantly lower than either of the (more comparable) psychiatric monitors

$\mathcal{M}1$ and $\mathcal{M}2$. Thus hospital-specific levels of $\mathcal{M}3$ are more stable over time and hence more predictable.

- Summary graphs of posterior inferences for specific monitor:year choices can provide useful insight into the distribution of outcome probabilities across the hospital system, about relative levels of performance, and about changes over time in such levels. Issues of how to most effectively summarise and use such information require consideration. Specific hospital/monitor/year outcomes that show up as evident extremes in these analyses may require further investigation.

- It should be clear that the models and computational methods (detailed in section 10.6 below) may be applied in other contexts, and that the basic binomial sampling model may be replaced by other non-Gaussian forms as context demands. The expectation is that the work will be developed in such ways and that the models will find use in various other applications in the socio-economic arena.

10.5.1 Further Extensions

There are various modelling assumptions that could be explored and tested by further development of the existing framework. These include extensions to allow for non-normal error distributions for the components of random effects ϵ_{ijt} and ν_{ijt}, as discussed in chapter 5, section refHTC. It is relatively straightforward to replace the normality assumptions with alternative, heavy-tailed error distributions, such as Student-T, in order to explore questions of robustness and sensitivity. The conjecture is that on the basis of the residual analysis and exploration to date, such extensions would not materially impact on the analysis or substantially change the summary inferences, although this may not be the case in studies of other monitor series.

One further addition planned is to extend the analysis to incorporate arbitrary

patterns of missing data. The current findings are restricted to analyses of 152 hospitals on which data is available on each of the three monitors in each of the eight years of study. Seven more hospitals in the data base have outcomes reported only on subsets of monitor/year pairings, so were omitted from the study. The models can be extended to incorporate these, and, more generally, to allow for any patterns of missing data.

A related development also planned for the near future involves extension to within-year predictions for model validation. In conjunction with the developments for missing data, minor changes to the analysis may be made to allow for reanalysis with selected hospital/monitor cases removed from the data set, but with the model still including such cases in analysis. This leads to posterior distributions for the parameters at these hospitals that are updated from the priors based on information from the other hospitals, but which are not now based on the actual outcomes at the omitted hospitals. Exploration of the concordance between these true outcomes and the posterior predictions provides us with out-of-sample, or cross-validatory, assessment of model adequacy.

10.6 Methodology and Computation

The fitting of the binomial/logit random effects time series models, both single monitor and multiple monitors, involves new statistical methodology and computational techniques. The models presented share much of the structure and conceptual basis of models recently published in Cargnoni *et al.* (1997). In addition, the multivariate model developed in this chapter is closely related to the dynamic factor model from chapters 6 and 7 through common statistical structure. Therefore, some of the components of the computational methods developed for the factor models are incorporated in this generalised linear models context. The methodology is novel and

will apply to other hierarchical generalised linear models with time series structure, and be of relevance to researchers in many socio-economic areas.

Implementation of the Gibbs Sampler

In each model class the analysis is performed using customised Markov chain Monte Carlo (MCMC) simulation methods to evaluate, explore and summarise posterior distributions. Technical details for the most general `MODEL-3` class are illustrated here. Details for the other models are routine simplifications. All summarised results are based on over 100,000 simulations of posteriors, which are generated following 10,000 "burn-in" simulations that are discarded. Of the total 100,000, a set of 5,000 is subsampled spaced 20 apart so as to break correlations and lead to essentially independent samples for summary analysis.

For the `MODEL-3` class ξ denotes, as before, any subset of the relevant prameters $\{\boldsymbol{\beta}_{0t}, \boldsymbol{\beta}_{1t}, \mathbf{V}, \mathbf{U}, \Phi, \boldsymbol{\alpha}_{i,t}, \boldsymbol{\mu}_{it}\}$. For the remaining variables together with the full data set write ξ^-. The Gibbs/Metropolis-Hastings framework then involves iteratively resampling from exact conditional posterior distributions $p(\xi|\xi^-)$ for some subsets of parameters, and from proposal distributions that approximate conditionals for other parameters, the latter then subject to accept/reject tests to ensure the accepted samples are from the true posterior. In turn the algorithm goes as follows.

- $\boldsymbol{\beta}_{0t}$ given $\boldsymbol{\beta}_{0t}^-$
- $\boldsymbol{\beta}_{1t}$ given $\boldsymbol{\beta}_{1t}^-$
- $\mathbf{V}$ given $\mathbf{V}^-$
- $\mathbf{U}$ given $\mathbf{U}^-$
- Φ given Φ^-

- $(\boldsymbol{\alpha}_{it}, \boldsymbol{\mu}_{it})$ given $(\boldsymbol{\alpha}_{it}, \boldsymbol{\mu}_{it})^-$.

The relevant conditional posteriors and simulation structures are now detailed for each of these steps. Note that the conditional distributions for $\mathbf{U}$ and Φ are exactly the same as the ones developed for the dynamic factor model in chapter 7 and are rewritten here just to make the chapter self-contained.

Sampling $\boldsymbol{\beta}_{0t}|\boldsymbol{\beta}_{0t}^-$

The $\boldsymbol{\beta}_{0t}$ are conditionally independent with full conditional posterior distributions, under reference priors, given by

$$p(\boldsymbol{\beta}_{01}|\boldsymbol{\beta}_{01}^-) \propto \prod_{i=1}^{I} p(\boldsymbol{\alpha}_{i1}|\boldsymbol{\beta}_{01})$$

and, for $t > 1$,

$$p(\boldsymbol{\beta}_{0t}|\boldsymbol{\beta}_{0t}^-) \propto \prod_{i=1}^{I} p(\boldsymbol{\alpha}_{it}|\boldsymbol{\alpha}_{i,t-1}, \boldsymbol{\beta}_{0t}, \boldsymbol{\beta}_{0,t-1}).$$

The posteriors are then multivariate normals,

$$\boldsymbol{\beta}_{01}|\boldsymbol{\beta}_{01}^- \sim N(\boldsymbol{\beta}_{01}|\mathbf{b}_{01}, \mathbf{W}/I) \quad \text{and} \quad \boldsymbol{\beta}_{0t}|\boldsymbol{\beta}_{0t}^- \sim N(\boldsymbol{\beta}_{0t}|\mathbf{b}_{0t}, \mathbf{U}/I),$$

where

$$\mathbf{b}_{01} = \sum_{i=1}^{I} \boldsymbol{\alpha}_{i1}/I$$

and, for $t > 1$,

$$\mathbf{b}_{0t} = \sum_{i=1}^{I} \{\boldsymbol{\alpha}_{it} - \Phi(\boldsymbol{\alpha}_{i,t-1} - \boldsymbol{\beta}_{0,t-1})\}/I.$$

Sampling $\boldsymbol{\beta}_{1t}|\boldsymbol{\beta}_{1t}^{-}$

Under reference priors, the $\boldsymbol{\beta}_{1t}$ vectors are conditionally independent over t, with full conditional posterior distributions $N(\boldsymbol{\beta}_{1t}|\mathbf{b}_{1t}, \mathbf{B}_{1t})$ for each t, and where

$$\boldsymbol{\beta}_{1t} = \mathbf{B}_{1t} \sum_{i=1}^{I} \mathbf{X}_{it}' \mathbf{V}^{-1} (\boldsymbol{\mu}_{it} - \boldsymbol{\alpha}_{it}) \quad \text{and} \quad \mathbf{B}_{1t}^{-1} = \sum_{i=1}^{I} \mathbf{X}_{it}' \mathbf{V}^{-1} \mathbf{X}_{it}.$$

Sampling $\mathbf{V}|\boldsymbol{V}^{-}$

Assuming a reference prior for the variance-covariance matrix $\mathbf{V}$, conditional posterior is the inverse Wishart distribution $Wi(\mathbf{V}^{-1}|8I, \mathbf{H})$ where

$$\mathbf{H} = \sum_{i=1}^{I} \sum_{t=1}^{8} \boldsymbol{\nu}_{it} \boldsymbol{\nu}_{it}'.$$

Sampling $\boldsymbol{U}|\boldsymbol{U}^{-}$

Assuming a reference prior for the variance-covariance matrix $\mathbf{U}$, the conditional posterior is given, in terms of the inverse $\mathbf{U}^{-1}$, by

$$\begin{aligned} p(\mathbf{U}^{-1}|\{\boldsymbol{\epsilon}_{it}\}, \Phi) &\propto p(\mathbf{U}^{-1}) p(\{\boldsymbol{\epsilon}_{it}\}|\Phi, \mathbf{U}) \\ &\propto p(\mathbf{U}^{-1}) \prod_{i=1}^{I} p(\boldsymbol{\epsilon}_{i1}|\Phi, \mathbf{U}) \prod_{t=2}^{8} p(\boldsymbol{\epsilon}_{it}|\boldsymbol{\epsilon}_{i,t-1}, \Phi, \mathbf{U}) \\ &\propto a(\mathbf{U}) Wi(\mathbf{U}^{-1}|7I, \mathbf{G}) \end{aligned}$$

with

$$\mathbf{G} = \sum_{i=1}^{I} \sum_{t=2}^{8} (\boldsymbol{\epsilon}_{it} - \Phi \boldsymbol{\epsilon}_{i,t-1})(\boldsymbol{\epsilon}_{it} - \Phi \boldsymbol{\epsilon}_{i,t-1})'$$

and

$$a(\mathbf{U}) = |\mathbf{W}|^{-I/2}\exp(-\text{trace}(\mathbf{W}^{-1}\mathbf{A})/2)$$

where $\mathbf{A} = \sum_{i=1}^{I} \boldsymbol{\epsilon}_{i1}\boldsymbol{\epsilon}_{i1}'$ and $\mathbf{W} = \Phi\mathbf{W}\Phi + \mathbf{U}$. The inverse Wishart distribution is used as a proposal distribution in the Metropolis-Hastings algorithm. That is, given a "current" value of $\mathbf{U}$ and corresponding $\mathbf{W}$, a "candidate" value $\mathbf{U}^*$ from the inverse Wishart distribution is sampled, and accept it with probability

$$\min\{1, a(\mathbf{U}^*)/a(\mathbf{U})\}$$

where $\mathbf{W}^* = \Phi\mathbf{W}^*\Phi + \mathbf{U}^*$.

Sampling $\Phi|\Phi^-$

Conditional on Φ^- the posterior for Φ depends only on the random effects $\boldsymbol{\epsilon}_{it}$ and $\mathbf{U}$ via

$$\begin{aligned} p(\Phi|\{\boldsymbol{\epsilon}_{it}\}, \mathbf{U}) &\propto p(\Phi)p(\{\boldsymbol{\epsilon}_{it}\}|\Phi, \mathbf{U}) \\ &\propto p(\Phi)\prod_{i=1}^{I} p(\boldsymbol{\epsilon}_{i1}|\Phi)\prod_{t=2}^{8} p(\boldsymbol{\epsilon}_{it}|\boldsymbol{\epsilon}_{i,t-1}, \Phi) \\ &\propto p(\Phi)N(\boldsymbol{\epsilon}_{i1}|\mathbf{0}, \mathbf{W})\prod_{t=2}^{8} N(\boldsymbol{\epsilon}_{it}|\Phi\boldsymbol{\epsilon}_{i,t-1}, \mathbf{U}) \end{aligned}$$

where $\mathbf{W} = \Phi\mathbf{W}\Phi + \mathbf{U}$ is easily evaluated as a function of Φ and $\mathbf{U}$. Write $\boldsymbol{\phi} = (\phi_1, \phi_2, \phi_3)'$ for the diagonal of Φ, and $\mathbf{E} = \text{diag}(\boldsymbol{\epsilon}_{i,t-1})$. Then the conditional posterior may be written as proportional to

$$p(\Phi)c(\Phi)N(\boldsymbol{\phi}|\mathbf{f}, \mathbf{F})$$

where

$$\mathbf{f} = \mathbf{F}\sum_{i=1}^{I}\sum_{t=2}^{8}\mathbf{E}'\mathbf{U}^{-1}\boldsymbol{\epsilon}_{it} \quad \text{and} \quad \mathbf{F}^{-1} = \sum_{i=1}^{I}\sum_{t=2}^{8}\mathbf{E}'\mathbf{U}^{-1}\mathbf{E},$$

and

$$c(\Phi) = |\mathbf{W}|^{-I/2}\exp(-\text{trace}(\mathbf{W}^{-1}\mathbf{A})/2)$$

where $\mathbf{A} = \sum_{i=1}^{I} \boldsymbol{\epsilon}_{i1}\boldsymbol{\epsilon}_{i1}'$ and $\mathbf{W} = \Phi\mathbf{W}\Phi + \mathbf{U}$. Under independent uniform priors for the ϕ_j, the full conditional posterior distribution for Φ is the above multivariate normal form truncated to the $(0,1)$ regions in each dimension, and then multiplied by the factor $c(\Phi)$. This may be sampled in several ways; in particular a Metropolis Hastings algorithm is used taking the truncated multivariate normal component as a proposal distribution. That is, given a "current" value of $\boldsymbol{\phi}$, with corresponding matrices Φ and $\mathbf{W}$, a "candidate" vector $\boldsymbol{\phi}^*$ is sampled from this truncated normal, compute the corresponding diagonal matrix Φ^* and variance matrix $\mathbf{W}^*$ such that $\mathbf{W}^* = \Phi^*\mathbf{W}^*\Phi^* + \mathbf{U}$, then accept this new $\boldsymbol{\phi}$ vector with probability

$$\min\{1, c(\Phi^*)/c(\Phi)\}.$$

Sampling $\boldsymbol{\alpha}_{it}|\boldsymbol{\alpha}_{it}^-$

Simulations are based on proposal distributions derived from the normal-logit approximations to the data model. Write

$$\tilde{\mathbf{y}}_{it} = \mathbf{y}_{it} - \mathbf{X}_{it}\boldsymbol{\beta}_{1t} = \boldsymbol{\alpha}_{it} + \boldsymbol{\eta}_{it}$$

where $\mathbf{y}_{it}$ is the vector of logit transforms of the observed outcome proportions. Under the model structure and assumptions, $\boldsymbol{\eta}_{it} \sim N(\boldsymbol{\eta}_{it}|\mathbf{0}, \mathbf{V} + \mathbf{S}_{it})$ where

$$\mathbf{S}_{it} = \text{diag}(s_{i1t}, s_{i2t}, s_{i3t})$$

is the diagonal matrix of approximate data variances in the normal-logit model. Combined with the model equations

$$\boldsymbol{\alpha}_{it} \;=\; \boldsymbol{\beta}_{0t} + \Phi(\boldsymbol{\alpha}_{i,t-1} - \boldsymbol{\beta}_{0,t-1}) + \boldsymbol{\omega}_{it}$$

and the initial version for the $\boldsymbol{\alpha}_{i1}$, this gives us a multivariate dynamic linear model with known variance matrices and state vector sequence $\boldsymbol{\alpha}_{it}$. Standard results for DLMs now apply, as in West and Harrison (1997). To sample from the full conditional posterior distribution a multivariate versions of the forward-filtering, backwards-sampling algorithm were implemented as detailed in the Appendix A.1 and used in this thesis in several places in chapter 5 and 7.

Sampling $\boldsymbol{\mu}_{it}|\boldsymbol{\mu}_{it}^{-}$

Given the just-sampled values of all $\boldsymbol{\alpha}_{it}$, we again use the normal-logit data model to generate candidate values of the $\boldsymbol{\mu}_{it}$ that are then tested for acceptance based on the exact conditional posteriors. As above, the approximate normal-logit data model serves to provide a very useful candidate generating model, as follows. For each i and t the exact conditional posteriors are

$$p(\boldsymbol{\mu}_{it}|\boldsymbol{\mu}_{it}^{-}, \mathbf{z}_{it}) \propto p(\mathbf{z}_{it}|\mathbf{n}_{it}, \boldsymbol{\mu}_{it})p(\boldsymbol{\mu}_{it}|\boldsymbol{\mu}_{it}^{-})$$

where the likelihood function $p(\mathbf{z}_{it}|\mathbf{n}_{it}, \boldsymbol{\mu}_{it})$ is the product of the three binomial-logit functions, and the conditional prior $p(\boldsymbol{\mu}_{it}|\boldsymbol{\mu}_{it}^{-})$ is the trivariate normal

$$\boldsymbol{\mu}_{it}|\boldsymbol{\mu}_{it}^{-} \sim N(\boldsymbol{\mu}_{it}|\boldsymbol{\alpha}_{it} + \mathbf{X}_{it}\boldsymbol{\beta}_{1t}, \mathbf{V}).$$

Again using the normal-logit approximation to the binomial data models, the approximate, and conditionally independent, normal posteriors are obtained

$$\boldsymbol{\mu}_{it}|y_{it} \approx N(\boldsymbol{\mu}_{it}|\mathbf{m}_{it}, \mathbf{Q}_{it})$$

where

$$\mathbf{Q}_{it} = (\mathbf{V}^{-1} + \mathbf{S}_{it}^{-1})^{-1} \quad \text{and} \quad \mathbf{m}_{it} = \mathbf{Q}_{it}(\mathbf{V}^{-1}(\boldsymbol{\alpha}_{it} + \mathbf{X}_{it}\boldsymbol{\beta}_{1t}) + \mathbf{S}_{it}^{-1}y_{it})$$

for each i and t.

Use these latter normal distributions as proposal distributions: generate candidate $\boldsymbol{\mu}_{it}$ values from each of this set of approximate posteriors, and accept/reject them according to a Metropolis-Hastings test. It is easily seen that this is a simple test, based on the ratio of the exact binomial to the approximate normal-logit likelihood functions. Specifically, if $\boldsymbol{\mu}_{it}$ is the current, "old" value of $\boldsymbol{\mu}_{it}$ from the previous MCMC iteration, a new value $\boldsymbol{\mu}_{it}^{*}$ from the probing distribution $N(\boldsymbol{\mu}_{it}|\mathbf{m}_{it}, \mathbf{Q}_{it})$ is accepted with probability

$$\min\{1, a(\boldsymbol{\mu}_{it}^{*})/a(\boldsymbol{\mu}_{it})\}$$

where $a(\cdot)$ is the ratio

$$a(\boldsymbol{\mu}_{it}) = p(\mathbf{z}_{it}|\mathbf{n}_{it}, \boldsymbol{\mu}_{it})/N(y_{it}|\boldsymbol{\mu}_{it}, \mathbf{S}_{it}),$$

i.e., the ratio of the product of the three exact binomial likelihood components to the product of the three approximate normal-logit components. Note the very close similarities of this method to that in multinomial time series modeling in Cargnoni *et al.* (1997)

Appendix A

Useful Results

A.1 MCMC in Dynamic Linear Models

To obtain a sample of a full set of state vectors $\boldsymbol{\theta}_t$ in normals or conditionally normal DLMs, Carter and Kohn (1994) and Frühwirth-Schnatter (1994) introduced the prototype simulation method named *Forward Filtering, Backwards Sampling*. The algorithm is defined in three basic steps:

1. Calculate the standard forward updates $\boldsymbol{\theta}_t|\mathbf{D}_t \sim N(\mathbf{m}_t, \mathbf{C}_t) \ \ \forall t$.

2. Sample the last state vector from $\boldsymbol{\theta}_n|D_n \sim N(\mathbf{m}_n, \mathbf{C}_n)$.

3. sample backwards through time for $t = n-1, n-2, \ldots, 1, 0$ sequentially from $\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t+1}, \mathbf{D}_t \sim N(\mathbf{h}_t, \mathbf{H}_t)$ where $\boldsymbol{\theta}_{t+1}$ is the value sampled in the previous step.

A.1.1 Forward Filtering

Detailed description of theory behind the equations and models described below can be found in West and Harrison (1997), sections 4.3 and 16.2. Consider the Dynamic Linear model defined by the quadruple $\{\mathbf{F}_t, \mathbf{G}_t, \mathbf{V}_t, \mathbf{W}_t\}$ and written in state space

form for each t,

$$\begin{aligned} \mathbf{y}_t &= \boldsymbol{\alpha}_t + \mathbf{F}_t'\boldsymbol{\theta}_t + \boldsymbol{\nu}_t & \boldsymbol{\nu}_t &\sim N(0, \mathbf{V}_t) \\ \boldsymbol{\theta}_t &= \boldsymbol{\beta}_t + \mathbf{G}_t\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t & \boldsymbol{\omega}_t &\sim N(0, \mathbf{W}_t) \\ \boldsymbol{\theta}_0 &\sim N(\mathbf{m}_0, \mathbf{C}_0) \end{aligned}$$

for some prior moments $\mathbf{m}_0$ and $\mathbf{C}_0$. Assuming that $\boldsymbol{\nu}_t$ and $\boldsymbol{\omega}_t$ are independent and mutually uncorrelated DLMs theory leads to the following updating equations:

(a) Posterior at $t-1$:

$$\boldsymbol{\theta}_{t-1}|\mathbf{D}_{t-1} \sim N(\mathbf{m}_{t-1}, \mathbf{C}_{t-1}).$$

(b) Prior at t:

$$\boldsymbol{\theta}_t|\mathbf{D}_{t-1} \sim N(\mathbf{a}_t, \mathbf{R}_t),$$

where

$$\mathbf{a}_t = \boldsymbol{\beta}_t + \mathbf{G}_t\mathbf{m}_{t-1} \qquad \text{and} \qquad \mathbf{R}_t = \mathbf{G}_t\mathbf{C}_{t-1}\mathbf{G}_t' + \mathbf{W}_t.$$

(c) One-step forecast:

$$\mathbf{y}_t|\mathbf{D}_{t-1} \sim N(\mathbf{f}_t, \mathbf{Q}_t),$$

where

$$\mathbf{f}_t = \boldsymbol{\alpha}_t + \mathbf{F}_t'\mathbf{a}_t \qquad \text{and} \qquad \mathbf{Q}_t = \mathbf{F}_t\mathbf{R}_t\mathbf{F}_t' + \mathbf{V}_t.$$

(d) Posterior at t:

$$\boldsymbol{\theta}_t|\mathbf{D}_t \sim N(\mathbf{m}_t, \mathbf{C}_t),$$

with

$$\mathbf{m}_t = \mathbf{a}_t + \mathbf{A}_t\mathbf{e}_t \qquad \text{and} \qquad \mathbf{C}_t = \mathbf{R}_t - \mathbf{A}_t\mathbf{Q}_t\mathbf{A}_t\prime,$$

where

$$\mathbf{A}_t = \mathbf{R}_t\mathbf{F}_t\mathbf{Q}_t^{-1} \qquad \text{and} \qquad \mathbf{e}_t = \mathbf{y}_t - \mathbf{f}_t.$$

A.1.2 Backwards Sampling

The backwards sampling algorithm as described in West and Harrison (1997) section 15.2.3 requires drawing samples from the conditional distributions $\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t+1}$. This is done from the filtering recurrences in West and Harrison (1997) section 4.7. Explicitly,

$$\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t+1} \sim N(\mathbf{h}_t, \mathbf{H}_t),$$

where

$$\mathbf{h}_t = \mathbf{m}_t + \mathbf{B}_t(\boldsymbol{\theta}_{t+1} - \mathbf{a}_{t+1})$$

and

$$\mathbf{H}_t = \mathbf{C}_t - \mathbf{B}_t\mathbf{R}_{t+1}\mathbf{B}_t',$$

with

$$\mathbf{B}_t = \mathbf{C}_t\mathbf{G}_{t+1}'\mathbf{R}_{t+1}^{-1}.$$

A.2 Matrix Normal Distributions

Dawid (1981) and West and Harrison (1997) define the class of matrix-variate normal distributions for matrices of jointly normal quantities. The $n \times p$ matrix $\mathbf{X}$ has a matrix normal distribution with mean $\mathbf{M}_{n\times p}$, left variance (columns) $\Delta_{n\times n}$ and right variance (rows) $\Sigma_{p\times p}$ if it has a density function given by

$$p(\mathbf{X}) = (2\pi)^{-np/2}|\Delta|^{p/2}|\Sigma|^{n/2}\exp\{-0.5 * \text{trace}(\mathbf{X}'\Delta^{-1}\mathbf{X}\Sigma^{-1})\}.$$

One important property of this distribution is that all marginal and conditional distributions of the elements of $\mathbf{X}$ and linear combinations of them are normally distributed. See above references for more details on the properties of this distributions which is traditionally denoted as

$$\mathbf{X} \sim N(\mathbf{M}, \Delta, \Sigma).$$

Random number generation from this distribution can be performed easily by generating a matrix $\mathbf{Z}$ of np independent standard normal samples and computing,

$$\mathbf{X} = \mathbf{M} + \mathbf{LZH}',$$

where $\Delta = \mathbf{LL}'$ and $\Sigma = \mathbf{HH}'$.

Bibliography

Aguilar, O. and West, M. (1998) Analysis of hospital quality monitors using hierarchical time series models. In *Bayesian Statistics in Science and Technology: Case Studies 4* (ed. C. Gatsonis et al). New York: Springer-Verlag. Forthcoming.

Ameen, J.R. and Harrison, P.J. (1995) Normal discount Bayesian models. In *Bayesian Statistics 2* (eds D.V. Lindley J.M. Bernardo M.H. De Groot and A.F.M. Smith). Amsterdam: North Holland and Valencia University Press.

Bollerslev, T. (1986) Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, **31,** 307–327.

Bollerslev, T. (1987) A conditional heteroskedastic time series model for speculative prices and rates of return. *Review of Economics and Statistics*, **69,** 542–546.

Bollerslev, T. and Engle, R.F. (1993) Common persistence in conditional variances. *Econometrica*, **61,** 167–186.

Bollerslev, T., Engle, R.F. and Nelson, D.B. (1994) ARCH models. In *The Handbook of Econometrics* (eds R.F. Engle and D. McFadden), pp. 263–274. North-Holland.

Bollerslev, T., Engle, R.F. and Wooldridge, J.M. (1988) A capital asset pricing model with time varying covariances. *Journal of Political Economy*, **96,** 116–131.

Burgess, J.F., Christiansen, C.L., Michalak, S.E. and Morris, C.N. (1996) Risk adjustment and economic incentives in identifying extremes using hierarchical models: A profiling application using hospital monitors. Manuscript, Management Science Group US Department of Veterans Affairs, Bedford MA.

Burmeister, F. and McElroy, M.B. (1991) The residual market factor, the APT and the mean variance efficiency ARCH models. *Review of Quantitative Finance and Accounting*, **1,** 27–49.

Cargnoni, C., Müller, P. and West, M. (1997) Bayesian forecasting of multinomial time series through conditionally Gaussian dynamic models. *Journal of the American Statistical Association*, **92,** 640–647.

Carter, C.K. and Kohn, R. (1994) On Gibbs sampling for state space models. *Biometrika*, **81,** 541–553.

Chen, N. (1983) Some empirical tests of the theory of arbitrage pricing. *Journal of Finance*, **38,** 1393–1414.

Chesney, M. and Scott, O. (1989) Pricing european options: A comparison of the modified Black Scholes model and a random variance model. *Journal of Financial and Quantitative Analysis*, **24,** 267–284.

Chiann, C. and Morettin, P. (1994) Wavelet analysis for stationary processes. Unpublished manuscript, University of Sao Paulo, Sao Paulo, Brazil.

Chipman, H., McCulloch, R. and Kolaczyk, E. (1997) Adaptive Bayesian wavelet shrinkage. *Journal of the American Statistical Association*, **92,** 1413–1421.

Chou, R.Y. (1988) Volatility persistence and stock valuations: Some empirical evidence using GARCH. *Journal of Applied Econometrics*, **3,** 279–294.

Christiansen, C.L. and Morris, C.N. (1997) Hierarchical Poisson regression modeling. *Journal of the American Statistical Association*, **92,** 618–632.

Chui, C.K. (1992) *An Introduction to Wavelets.* San Diego, CA: Academic Press, Inc.

Clyde, M., Parmigianni, G. and Vidakovic, B. (1998) Multiple shrinkage and subset selection in wavelets. *Biometrika*; Forthcoming.

Connor, G. and Korajczyk, A. (1988) Risk and return in an equilibrium apt: An application of a new methodology. *Journal of Financial Economics*, **21,** 255–289.

Connor, G. and Korajczyk, A. (1996) The arbitrage pricing theory and multi-factor models of assets returns. Technical Report. Northwestern University.

Daubechies, I. (1988) Orthonormal bases of compactly supported wavelets. *Commun. Pure Appl. Math.*, **41,** 919–996.

Daubechies, I. (1992) Ten lectures on wavelets. Society for industrial and applied mathematics. Philadelphia, PA.

Dawid, P. (1981) Some matrix-variate distribution theory: Notational considerations and a Bayesian application. *Biometrika*, **68,** 265–274.

De Jong, P. and Shephard, N. (1995) The simulation smoother for time series models. *Biometrika*, **82,** 339–350.

Diebold, F.S. and Nerlove, M. (1988) The dynamics of exchange rate volatility: A multivariate latent factor ARCH model. *Journal of Applied Econometrics*, **4,** 1–22.

Dijkerman, R.W. and Mazumdar, R.R. (1994) Wavelet reperesentations of stochastic processes and multiresolution stochastic models. *IEEE Transaction on signal processing*, **42,** 1640–1652.

Domowitz, I. and Hakkio, C.S. (1985) Conditional variance and the risk premium in the foreign exchange market. *Journal of International Economics*, **19,** 47–66.

Donoho, D. and Johnston, T. (1994) Bank of wavelet-related reports. Technical Report. Stanford University.

Engle, R.F. (1982) Autoregressive conditional heteroskedasticity with estimates of the variance of the united kingdom inflation. *Econometrica*, **50,** 987–1007.

Engle, R.F. and Bollerslev, T. (1986) Modeling the persistence of conditional variances. *Econometric Reviews*, **5,** 1–50.

Engle, R.F., Lilien, D. and Robins, R. (1987) Estimating time varying risk premia in the term structure: The ARCH-M model. *Econometrica*, **55,** 391–408.

Engle, R.F., Ng, V.K. and Rothschild, M. (1990) Asset pricing with a factor-ARCH covariance structure, empirical estimates for treasury bills. *Journal of Econometrics*, **45,** 213–237.

Escribano, A. and Pena, D. (1994) Cointegration and common factors. *Journal of Time Series Analysis*, **15,** 577–586.

French, K.R., Schwert, G.W. and Stambaugh, R.F. (1986) Expected stock returns and volatility. *Journal of Financial Economics*, **19,** 36–129.

Frühwirth-Schnatter, S. (1994) Data augmentation and dynamic linear models. *Journal of Time Series Analysis*, **15,** 183–202.

Geweke, J. and Zhou, G. (1996) Measuring the pricing error of the arbitrage pricing theory. *The Review of financial Studies*, **9,** no. 2, 557–587.

Goel, P. and Vidakovic, B. (1995) Wavelet transformations as diversity enhancers. Technical Report 95–08. ISDS, Duke University.

Goldstein, H. and Spiegelhalter, D.J. (1996) League tables and their limitations:

Statistical issues in comparisons of institutional performance. *Journal of the Royal Statistical Society*, **159,** 385–443.

Harrison, P.J. and West, M. (1987) Practical Bayesian forecasting. *The Statistician*, **36,** 115–125.

Harvey, A.C. (1989) *Forecasting structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.

Harvey, A.C., Ruiz, E. and Shepard, N. (1994) Multivariate stochastic variance models. *Review of Economic Studies*, **61,** 247–264.

Harvey, A.C. and Shephard, N. (1996) Estimation of an asymmetric stochastic volatility model for asset returns. *Journal of Business and Economic Statistics*, **14,** 429–434.

Harvey, A.C. and Stock, J.H. (1988) Continuous time autoregressive models with common stochastic trends. *Journal of Economic Dynamics and Control*, **12,** 265–284.

Harvey, C.R. and Zhou, G. (1990) Bayesian inference in asset pricing tests. *Journal of Financial Economics*, **26,** 221–254.

Hull, J. and White, A. (1987) The pricing options on assets with stochastic volatility. *Journal of Finance*, **42,** 281–300.

Jacquier, E., Polson, N.G. and Rossi, P. (1994) Bayesian analysis of stochastic volatility models (with discussion). *Journal of Business and Economic Statistics*, **12,** 371–388.

Jacquier, E., Polson, N.G. and Rossi, P. (1995) Models and priors for multivariate stochastic volatility. Technical Report. Graduate School of Business, University of Chicago.

Katul, G. (1994) A model for sensible heat flux probability density function for near-neutral and slightly stable atmospheric flows. *Boundary-Layer Meteorol.*, **71,** 1–20.

Katul, G. and Vidakovic, B. (1996) The partitioning of attached and detached eddy motion in the atmospheric surface layer using lorentz wavelet filtering. *Boundary-Layer Meteorol*, **77,** 153–172.

Kim, S., Shephard, N. and Chib, S. (1998) Stochastic volatility: Likelihood inference and comparison with ARCH models. *Review of Economic Studies*; Forthcoming.

King, M., Sentana, E. and Wadhwani, S. (1994) Volatility and links between national stock markets. *Econometrica*, **62,** 901–933.

Kolmorogorov, A.N. (1941) The local structure of turbulence in incompressible viscous fluid for very large reynolds numbers. *Dokl. Akad. Nauk. SSSR*, **30,** 301–305.

Lehman, B.N. and Modest, D.M. (1988) The empirical foundations of the arbitrage pricing theory. *Journal of Financial Economics*, **21,** 213–254.

Lintner, J. (1965) The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *Review of Economics and Statistics*, **47,** 13–37.

Mallat, S. (1989) A theory of multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **11,** 674–693.

Markowitz, H.M. (1959) *Portfolio Selection: Efficient Diversification of Investments.* New York: John Wiley and Sons.

Marsh, N.D. and Ditlevsen, P.D. (1996) Dynamics of glaciation and deglaciation identified through chemical tracers in ice-cores. Unpublished manuscript: Department of Geophysics, University of Conhagen, Denmark.

Martin and McDonald, R.P. (1975) Bayesian estimation in unrestricted factor analysis: A treatment for Heywood cases. *Psychometrika*, **40,** 505–517.

Mayewski, P. *et al.* (1994) Changes in atmospheric circulation and ocean ice cover over the north atlantic during the last 41000 years. *Science*, **263,** 1747–1751.

McCoy, E.J. and Walden, A.T. (1996) Wavelet analysis and synthesis of stationary long-memory processes. *Journal of Compuational and Graphical Statistics*, **5,** 26–56.

McCulloch, R. and Rossi, P.E. (1990) Posterior predictive and utility based approaches to testing arbitrage pricing theory. *Journal of Financial Economics*, **28,** 7–38.

McCurdy, T.H. and Morgan, I.G. (1988) Testing the martingale hypothesis in Deutschmark futures with models specifying the form of heteroskedasticity. *Journal of Applied Econometrics*, **3,** 187–202.

Melino, A. and Turnbull, S.M. (1990) Pricing options with stochastic volatility. *Journal of Econometrics*, **45,** 239–265.

Milhoj, A. (1987) A conditional variance model for daily deviations of an exchange rate. *Journal of Business and Economic Statistics*, **5,** 99–103.

Molenaar, P.C.M., Goodier, J.G.D. and Schmitz, B. (1992) Dynamic factor analysis of nonstationary multivariate time series. *Psychometrika*, **57,** 333–349.

Müller, P. and Pole, A. (1998) Monte Carlo posterior integration in GARCH models. *Sankhya, (Ser B)*; Forthcoming.

Pena, D. and Box, G.E.P. (1987) Identifying a simplifying structure in time series. *Journal of the American Statistical Association*, **82,** 836–843.

Polson, N.G. and Tew, B.V. (1997) Bayesian portfolio selection: An empirical analysis of the SP500 index 1970-1996. Technical Report. Graduate School of Business, University of Chicago.

Prado, R. and West, M. (1997) Exploratory modelling of multiple non-stationary time series: Latent process structure and decompositions. In *Modelling Longitudinal and Spatially Correlated Data: Methods, Applications and Future Directions* (ed. T.G. Gregoire), Lecture Notes in Statistics 12. New York: Springer Verlag.

Press, S.J. (1985) *Applied Multivariate Analysis: Using Bayesian and Frequentist Methods of Inference*. California: Krieger.

Press, S.J. and Shigemasu, K. (1989) Bayesian inference in factor analysis. In *Contributions to Probability and Statistics: Essays in Honor of Ingram Olkin* (eds S.J. Press L.J Gleser M.D. Perlman and A.R. Sampson), pp. 271–287. New York: Springer Verlag.

Putnam, B.H. and Quintana, J.M. (1994a) New Bayesian statistical approaches to estimating and evaluating models of exchange rates determination. In *Proceedings of the ASA Section on Bayesian Statistical Science, Joint Statistical Meetings*. American Statistical Association.

Putnam, B.H. and Quintana, J.M. (1994b) Shrink wrap for currencies. In *Asset Al-*

location: Applying Quantitative Discipline to Asset Allocation (ed. B.H. Putnam), pp. 175–181. London: Global Investor, Euromoney Publications.

Putnam, B.H. and Quintana, J.M. (1995) The evolution of Bayesian forecasting models. In *Asset Allocation: Applying Quantitative Discipline to Asset Allocation* (ed. B.H. Putnam), pp. 139–146. London: Global Investor, Euromoney Publications.

Quintana, J.M. (1992) Optimal portfolios of forward currency contracts. In *Bayesian Statistics 4* (eds D.V. Lindley J.M. Bernardo M.H. De Groot and A.F.M. Smith). Oxford University Press.

Quintana, J.M., Chopra, V.K. and Putnam, B.H. (1995) Global asset allocation: Stretching returns by shrinking forecasts. In *Proceedings of the ASA Section on Bayesian Statistical Science, Joint Statistical Meetings*. American Statistical Association.

Quintana, J.M. and Putnam, B.H. (1996) Debating currency markets efficiency using multiple-factor models. In *Proceedings of the ASA Section on Bayesian Statistical Science, Joint Statistical Meetings*. American Statistical Association.

Quintana, J.M. and West, M. (1987) An analysis of international exchange rates using multivariate DLMS. *The Statistician*, **36,** 275–281.

Quintana, J.M. and West, M. (1988) Time series analysis of compositional data. In *Bayesian Statistics 3* (eds D.V. Lindley J.M. Bernardo M.H. De Groot and A.F.M. Smith). Oxford University Press.

Roll, R.W. and Ross, S.A. (1980) An empirical investigation of the arbitrage pricing theory. *Journal of Finance*, **35,** 1073–1103.

Ross, S.A. (1976) The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, **8,** 341–360.

Ross, S.A. (1977) Risk return and arbitrage. In *Risk and Return in Finance* (eds I. Friend and J.L. Bicksler), vol. 1. Cambridge: Ballinger.

Shanken, J. (1987) A Bayesian approach to testing portfolio efficiency. *Journal of Financial Economics*, **19,** 195–216.

Sharpe, W.F. (1964) Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, **19,** 425–442.

Shephard, N. (1993) Fitting non-linear time series models, with applications to stochastic volatility. *Journal of Applied Econometrics*, **8,** 135–152.

Shephard, N. (1994a) Local scale models: State space alternatives to integrated GARCH processes. *Journal of Econometrics*, **6,** 333–364.

Shephard, N. (1994b) Partial non-Gaussian state space. *Biometrika*, **81,** 115–152.

Shephard, N. (1996) Statistical aspects of ARCH and stochastic volatility. In *Time Series Models in Econometrics, finance and Other Fields* (eds O.E. Barndorff-Nielson D.R. Cox and D.V. Hinkley), pp. 1–67. London: Chapman & Hall.

Shephard, N. and Pitt, M.K. (1997) Likelihood analysis of non-Gaussian measurements time series. *Biometrika*, **84**.

Strang, G. (1989) Wavelets and dilation equations. *SIAM Review*, **31,** 614–627.

Tiao, G.C. and Tsay, R.S. (1989) Model specification in multivariate time series (with discussion). *Journal of the Royal Statistical Society*, **51,** 195–213.

Townsend, A. (1976) *The Structure of Turbulent Shear Flow*. Cambridge University Press.

Uhlig, H. (1994) On singular Wishart and singular multivariate Beta distributions. *Annals of Statistics*, **22,** 395–405.

Uhlig, H. (1997) Bayesian vector-autoregressions with stochastic volatility. *Econometrica*, **65,** 59–73.

Vidakovic, B. (1995) Unbalancing data with wavelet transformations. *SPIEE*, **2569,** 845–857.

Vidakovic, B. (1998a) Nonlinear wavelet shrinkage with Bayes rules and Bayes factors. *Journal of the American Statistical Association*, **93,** 173–179.

Vidakovic, B. (1998b) Statistical modeling by wavelets. Lecture Notes, ISDS.

Vidakovic, B. and Mueller, P. (1995) Wavelet shrinkage with affine Bayes rules with applications. Technical Report 95–34. ISDS, Duke University.

Walter, G.G. (1994) *Wavelets and Other Orthogonal Systems with Applications*. Boca Raton, FL: CRC Press.

West, M. (1995) Bayesian inference in cyclical components of dynamic linear models. *Journal of the American Statistical Association*, **90,** 845–857.

West, M. (1997a) Modelling and robustness issues in Bayesian time series analysis (with discussion). In *Bayesian Robustness 2*. IMS Monographs.

West, M. (1997b) Time series decomposition. *Biometrika*, **84,** 489–494.

West, M. and Aguilar, O. (1997) Studies of quality monitor time series: The V.A. hospital system. Technical Report 96–02. ISDS Duke University, Bedford, MA. Report for the VA Management Science Group.

West, M. and Harrison, P.J. (1997) *Bayesian Forecasting and Dynamic Models*, Second edn. New York: Springer Verlag.

Biography

I was born in Mexico City, Mexico in July 25, 1970. I obtained a B.S degree in Actuarial Sciences and a post-graduate diploma in Applied Statistics from the Autonomous Technological Institute of Mexico (ITAM) in May 92 and September 93 respectively. I arrived at Duke University in the Fall of 1994 and became a Ph.D candidate in the Spring of 1996.