BAYESIAN MODELING AND ANALYSIS OF MULTIVARIATE TIME SERIES, WITH APPLICATIONS IN FINANCE AND HEALTH POLICY

by

Viridiana Alicia Lourdes de León

Institute of Statistics and Decision Sciences
Duke University

Date: ______________________

Approved:

____
Dr. Mike West, Supervisor

____
Dr. Brani Vidakovic

____
Dr. Dalene Stangl

____
Dr. James E. Smith

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

2000

<u>ABSTRACT</u>

(Statistics)

BAYESIAN MODELING AND ANALYSIS OF MULTIVARIATE TIME SERIES, WITH APPLICATIONS IN FINANCE AND HEALTH POLICY

by

Viridiana Alicia Lourdes de León

Department of Institute of Statistics and Decision Sciences
Duke University

Date: ______________________________
Approved:

____
Dr. Mike West, Supervisor

____
Dr. Brani Vidakovic

____
Dr. Dalene Stangl

____
Dr. James E. Smith

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

2000

Abstract

This dissertation develops Bayesian theory and computation to address important issues in two main socio-economic areas: financial modeling and institutional assessment.

The first part focusses on computational developments for model fitting and forecasting of multiple series of crude oil futures prices. The methodology is motivated by the central role that the stochastic behavior of commodity prices plays in the evaluation of commodity-related securities. A class of Bayesian multivariate dynamic linear models for oil future prices is developed based on a theoretical financial model that assumes two latent factor processes: a notional equilibrium price level and a process representing short-term deviations from equilibrium levels. A customized Markov Chain Monte Carlo (MCMC) sampling scheme is developed for inference and analysis of such model. In addition, several structures on the observational variance are explored including the challenging case of a singular variance matrix. Relevant and supporting theory of singular densities and DLMs under singular observational variance is reviewed and developed.

The second part involves the development of large-scale longitudinal models for institutional comparisons. Complex non-Gaussian hierarchical models are developed to profile providers in health-care delivery systems. The key motivating concern is to estimate health-care return-time distributions for individuals, and to evaluate differences due to year of care and hospital, in the context of a range of possible individual-level explanatory variables. Results indicate significant system-wide improvement in the health-care areas of study, in addition to large amounts of variation in this improvement across medical centers. Covariates such as age of patient, treatment priority, and diagnoses help to illustrate important potential new health policy

interventions and the outcomes of previous interventions. The study involves innovation in hierarchical/longitudinal models for the very large and complex data set, a range of exploratory data analytic developments, customized MCMC for Bayesian model fitting and some creativity in exploring the very high-dimensional posterior distributions and summarizing MCMC outputs.

Acknowledgements

I would like to express my sincere appreciation and deep gratitude to my advisor Professor Mike West. I greatly appreciate his enthusiasm, encouragement and endless support. His dedication of research and his determination to complete each task to the fullest has been truly inspiring to me; his insights, ideas and guidance were crucial to accomplish this dissertation.

Many thanks are due to Professors Peter Müller, Dalene Stangl and Brani Vidakovic from ISDS for their comments and assistance. I especially appreciate the generous help and the the numerous suggestions given to me by Professor James Smith from the Fuqua School of Business.

I am grateful to Herbie Lee for his proofreading and great companionship, to Hedibert Lopes and Alejandro Rios for their useful suggestions, and to Andrew Gallatin for all his help with computers. My thanks also go to my classmates John Kern and Lurdes Inoue for providing a cooperative environment of study and research and also my present and past officemates Jenise Swall, Heidi Ashish, Sining Chen and Ana Grohovac for their great companionship and for making my time in ISDS really enjoyable.

I am thankful to my sister, Mónica, for her constant encouragement and to Ingrid Pineda for many years of friendship and support. I reserve special thanks to my parents, Lupita and Julio, for their devoted love, enormous encouragement and limitless support. They have always given me the strength to stand up after bad times and always increased the joy in good times.

Finally, I give my most heartfelt appreciation to Omar. Without his love, understanding, support, and invaluable help, finishing this dissertation would have been impossible. Thank you!

Contents

Abstract **iv**

Acknowledgements **vi**

Dedication **xi**

List of Tables **xii**

List of Figures **xiii**

1 Introduction **1**

2 Crude Oil Futures Prices Data **4**

2.1 Crude Oil Markets and Volatility in Prices 4

2.2 Futures Market 6

2.3 Crude Oil Futures Data 7

2.4 Summary 10

3 Multivariate Dynamic Linear Model for Oil Futures Prices **11**

3.1 Latent Processes in Commodity Prices 12

3.1.1 Continuous-Time Model 13

3.1.2 Risk-Neutral Process 14

3.1.3 Discrete-Time Model 16

3.1.4 Interpretation of Parameters 18

3.1.5 State-Space Representation 19

3.1.6 Modified Discrete-Time Model 20

3.2 Bayesian Analysis 22

3.2.1 Prior Distributions 23

3.2.2 Implementation of the Gibbs Sampler 23

3.3 Summary 26

4 Structures for the Observational Variance Matrix 28

4.1 Diagonal Observational Variance 29

4.2 Non-diagonal Observational Variance 30

4.3 Structure in Function of Difference in Maturities 32

4.3.1 Bayesian Analysis and Implementation of the Gibbs Sampler . 34

4.3.2 Analysis Results 35

4.4 Factor Model 38

4.4.1 Bayesian Analysis and Implementation of the Gibbs Sampler . 40

4.4.2 Model Selection Using the BIC 42

4.4.3 Analysis Results 45

4.5 Summary 49

5 Singular Observational Variance Matrix 51

5.1 Singular Densities 51

5.1.1 Singular Multivariate Normal 52

5.1.2 Singular Wishart 54

5.2 Forward Filtering under Singular Observational Variance 55

5.3 Implementation of the Gibbs sampler 58

5.4 Analysis Results 60

5.5 Forecasting 70

5.5.1 Posterior Predictive Distributions 70

5.5.2 One-Step-Ahead Forecast 74

5.6 Summary 76

6 Non-Gaussian Hierarchical Models in Institutional Profiling 78

6.1 The VA Hospital System 79

6.2 Data Structure 80

6.3 The GLM 83

6.3.1 Discrete Duration Model 84

6.3.2 Discrete Proportional Odds Model 85

6.4 Single-year Model 85

6.4.1 Data Specification 86

6.4.2 Screening Covariates 88

6.4.3 Model Specification and Implementation 90

6.4.4 Analysis Results 93

6.5 Summary 103

7 Non-Gaussian Hierarchical Time Series Structure for Models in Institutional Profiling 105

7.1 Multi-year Model 106

7.1.1 Data Specification 106

7.1.2 Model Structure and Implementation 107

7.1.3 Analysis Results 110

7.2 Extensions 120

7.2.1 Model with Additional Residual Terms 120

7.2.2 Model Estimation for More Cut-off Values for Return-Times . 122

7.3 Summary 129

8 Summary and Future Directions 131

8.1 Multivariate DLMs for Futures Prices 131

8.1.1 General Results 131

8.1.2 Future Directions 133

8.2 Hierarchical Models in Institutional Profiling 134

8.2.1 General Results 135

8.2.2 Future Directions 135

A Some Results on Singular Variance Matrices **137**

A.1 A Practical Issue of the Rank of a Matrix 137

A.2 Generalized Inverse 137

A.3 Standardization of a Random Vector with Singular Variance 138

A.4 Conditional distribution of $\mathbf{y}_1$ given $\mathbf{y}_2$ when $\mathbf{y}_2$ is singular normally distributed 139

A.5 Singular Wishart Distribution 139

Bibliography **141**

Biography **148**

To my parents, Lupita and Julio
and
my husband Omar

List of Tables

4.1 BIC values assuming different number of factors. 44

5.1 Different Wishart distributions $\text{Wishart}_r\,(n, \mathbf{\Sigma})$ depending on rank($\mathbf{\Sigma}$)=ρ and degrees of freedom n. 54

6.1 Cumulative empirical frequency distributions of return-times at specific cut-offs, categorized by levels of several primary covariates. . . . 87

6.2 Cumulative empirical frequency distributions of return-times at specific cut-offs, categorized by four specific hospitals. 88

6.3 Selected covariates using the BIC based on uncensored data. 91

6.4 Selected covariates using the BIC based on all data. 91

List of Figures

2.1 Crude Oil futures prices (settlement) from 01/01/90 to 10/18/99. Vertical scale is US dollars. 8

2.2 Crude oil futures prices from 01/01/90 to 03/11/91. 8

2.3 Term structure of Crude Oil futures prices at 05/09/94. 9

4.1 Correlation structure of the estimated residuals in a particular iteration. 30

4.2 A sample from the posterior distribution of $\mathbf{V}$. 31

4.3 Autocorrelation function for the observed residuals $\hat{\boldsymbol{\epsilon}}_t$ estimated under the model assuming a non-diagonal $\mathbf{V}$. 32

4.4 Histogram of simulated sample from the posterior density of τ (left) and the posterior mean of the correlation function $\rho(d) = \exp\left\{\frac{d^{1/2}}{\tau}\right\}$ (right). 36

4.5 95% posterior intervals for $\boldsymbol{\Sigma}$. 37

4.6 Posterior mean of factors $\mathbf{f}_t$. 45

4.7 Posterior summaries for the factor loadings matrix $\mathbf{X}$. 47

4.8 95% posterior intervals for the specific variances ψ_i (left) and for the percentage of total variation of the specific variances $100\psi_i/\sigma_i^2$ (right). 48

4.9 Normal quantile plot for the posterior means of the standardized ordered observed residuals $\hat{\boldsymbol{\nu}}_t$. The vertical lines represent 95% posterior intervals. Each panel represents one dimension of $\hat{\boldsymbol{\nu}}_t$. 48

5.1 Singular bivariate normal density 52

5.2 Autocorrelation functions for the persistence parameter ϕ assuming $\rho = 1, \ldots, r$ where $r = 12$. 61

5.3 Posterior summaries of observation variance matrix **V**. The elements in the diagonal are standard deviations, the elements in the off-diagonal are correlations. 62

5.4 Posterior summaries for the persistence parameter ϕ (top left) κ in annual terms (top right) and the *half-life* in months of the short-term deviations (bottom). 63

5.5 Posterior summaries of the equilibrium drift rate μ_ξ, equilibrium risk-neutral drift rate μ_ξ^* and risk premiums λ_χ and $\lambda_\xi = \mu_\xi - \mu_\xi^*$. The prior distribution is indicated by the solid line. 64

5.6 Posterior summaries of the system variance matrix **W**. The standard deviations σ_χ and σ_ξ are in annual terms. 65

5.7 Posterior mean of the log of the heavy-tailed weights, $\log(\gamma_t^{-1})$. 66

5.8 95% posterior intervals for χ_t (upper frame) and ξ_t (lower frame) indicated by vertical segments. The dotted line in the lower frame represents the posterior mean of the log of the spot price. 67

5.9 Normal quantile plot of the posterior means of the ordered observation standardized residuals $\hat{\boldsymbol{\epsilon}}_t$ of dimension $\rho = 4$. The vertical lines represent 95% posterior intervals. Each panel represents one dimension of the standardized residuals $\hat{\boldsymbol{\epsilon}}_t$. The labeled points indicate the percentages that particular observations contributed in the computation of the posterior mean residual. 68

5.10 Futures prices data and 95% posterior intervals of the term structure (years) of oil futures prices for 02/13/95, 01/06/97 and 11/16/98 under short-term (first column) and long-term (second column) time span. . 69

5.11 Seven randomly selected samples from the predictive distribution of the next 23 values of the series. The first panel shows the actual data (log scale) from 01/01/98 to 03/27/00. Observed values, from 01/01/98 to 10/18/99, and simulated values, from 10/25/99 to 03/27/99, are separated by a dashed line. 72

5.12 Samples from the posterior predictive distribution for 10/25/99 for the twelve futures prices series in the log scale. The solid line indicates the observed price, each panel represents one of the twelve price series. 73

5.13 95% posterior intervals for the predictive distribution of the futures prices in the log scale. Each panel represents one futures prices series with certain maturity. 73

5.14 95% posterior intervals for the term structure of oil futures prices for 10/25/99 to 03/27/00. The solid line is the posterior mean, the points are the actual data. 75

6.1 Cumulative empirical proportions of return-times at specific "cut-offs": $1, 7, 14, 21, 30$ and 367 days. 82

6.2 Histograms of return-times in FY97. The first histogram includes all 35,368 individuals; the second includes only patients with return-time less or equal to 30 days. 86

6.3 95% posterior intervals for baseline return-time probabilities $p_0(t)$. . . 94

6.4 95% posterior intervals for *Age* effects $\delta_d(t)$ relative to *Age group* 1 (≤ 44 years). 95

6.5 95% posterior intervals for *DRG* effects $\gamma_g(t)$ relative to *DRG* 434. . . 96

6.6 95% posterior intervals for *Marital Status* effects $\kappa_k(t)$ relative to *Married* status. 96

6.7 95% posterior intervals for *Priority* effects $\eta_e(t)$ relative to *AN* (poor veterans). 97

6.8 95% posterior intervals for *Gender* effects $\chi_2(t)$, *Female* category relative to *Male*. 98

6.9 95% posterior intervals for *Racial* effects $\zeta_z(t)$ relative to *White*. . . . 99

6.10 95% posterior intervals for hospital specific random effects $\epsilon_j(t)$ for four selected hospitals indicated by station numbers. 100

6.11 95% posterior intervals for Hospital random effects $\epsilon_j(t)$ in analysis with cut-off $t = 30$ (top) and $t = 1$ (bottom). Hospitals are ordered by the posterior medians of the effects under the analysis with $t = 30$. 101

6.12 95% posterior intervals for the standard deviation $w(t)$ of the population distribution of hospital specific random effects ϵ_j. 102

7.1 95% posterior intervals for baseline return-time probabilities $p_{0,r}$ over years $r = 1, \ldots, 10$. The points correspond to the overall proportions returning within 30 days from the raw data. 111

7.2 95% posterior intervals for *Age group* effects $\delta_{2,r}$ over years $r = 1, \ldots, 10$ relative to *Age group* 1 ($\leq$ 65 years). 112

7.3 95% posterior intervals for *DRG* group effects $\gamma_{g,r}$ over years $r = 1, \ldots, 10$ relative to *DRG* 434. 113

7.4 95% posterior intervals for Priority group effects $\eta_{e,r}$ over years $r = 1, \ldots, 10$ relative to *AN* (poor veterans). 114

7.5 95% posterior intervals for *Diagnosis* group effects $\xi_{x,r}$ over years $r = 1, \ldots, 10$ relative to *Dependence* category. 115

7.6 95% posterior intervals for *Gender* effects $\chi_{2,r}$ over years $r = 1, \ldots, 10$, *Female* category relative to *Male*. 116

7.7 95% posterior intervals for hospital-specific random effects $\epsilon_{j,r}$ over years $r = 1, \ldots, 10$ in multi-year analysis with cut-off $t = 30$ days. Intervals are given for the four hospitals selected in the earlier, single year analysis, as labeled, plus one additional hospital (#139). 117

7.8 95% posterior intervals for the hospital-specific dependence parameters ϕ_j (top) and intervals for the hospital-specific innovation variance u_j (bottom) for all 136 hospitals. Hospitals ordered by posterior median of ϕ_j. The five hospitals selected earlier are highlighted. 118

7.9 95% posterior intervals for the hospital-specific random effects ϵ_j (top) and hospital-specific residual terms ν_j (bottom) of all 136 hospitals for FY97. The hospitals are ordered by posterior medians of the random effects. 120

7.10 95% posterior intervals for the hospital-specific residual terms ν_j of all 136 hospitals for FY97. The hospitals are ordered by posterior medians of the random effects. The five hospitals selected earlier are highlighted. 121

7.11 Histogram of a sample from the posterior distribution of v, the standard deviation of the hospital-specific residual/noise terms $\nu_{j,r}$. 122

7.12 95% posterior intervals for baseline return-time probabilities $p_{0,r}(t)$ over years $r = 1, \ldots, 10$ and cut-offs $t = 1, 7, 14, 21, 30, 367$. 123

7.13 95% posterior intervals for *Age group* effects $\delta_{2,r}(t)$ (upper frame) and the differences in implied return-time probabilities relative to the baseline (lower frame). 124

7.14 95% posterior intervals for *Gender* effects $\chi_{2,r}(t)$ (upper frame) and the differences in implied return-time probabilities relative to the baseline (lower frame). 126

7.15 95% posterior intervals for hospital-specific random effects $\epsilon_{j,r}(t)$ for hospital j =115. Graphs display the effects $\epsilon_{115,r}(t)$ (upper frame) and the differences in implied return-time probabilities relative to the baseline (lower frame). 128

7.16 95% posterior intervals for model hyperparameters for different return-time cut-offs $t = 1, 7, 14, 21, 30, 367$ days. 129

Chapter 1

Introduction

Bayesian methodology has played a central role in developing complex models of real world phenomena in many areas. This dissertation develops, under the Bayesian framework, two realistic and complex mathematical models to address specific issues in the social and economic sciences.

The dissertation is divided into two main parts. The first is focussed on computational developments for model fitting and forecasting of a financial model for crude oil futures prices. The second involves the development of a large-scale longitudinal model for institutional comparisons. The latter is exemplified by studies of quality-of-care measures to profile providers in health-care delivery systems.

In the first part the focus is on crude oil which is a natural resource that serves as an underlying asset of many financial instruments (Gibson and Schwartz, 1990). Its availability and cost play a central role in the economies of all countries. More than ever, statistical models are key to explaining the behavior of price movements to manage exposure to market fluctuations and reduce risk. Chapter 2 introduces general features of the crude oil related markets and describes the structure of oil futures prices data which is the major motivation for the models studied in subsequent chapters. Chapter 3 develops a class of Bayesian multivariate dynamic linear models for oil futures prices based on theoretical foundation in the financial literature. Customized Markov Chain Monte Carlo (MCMC) sampling schemes are developed for inference and computation on model parameters. The highly complicated model presented in this chapter is especially designed to follow standard financial theory for commodity pricing and represents important methodological contributions to both Bayesian statistics and financial science. To complete the model specification, Chapters 4 and 5 present different structures on the variance-covariance matrix of the observational errors. Chapter 4 starts by assuming a simple diagonal matrix structure followed by a traditional non-diagonal matrix and moving into a matrix where the correlations are functions of the difference of maturities in the futures contracts. Finally, a factor model for observation errors is assumed implying a decomposition of the covariance matrix into two sources of variability, one common across all series, and another specific of each one. Chapter 5 models the observational variance as non-diagonal and singular, representing a very challenging approach due to the complexity of the model. Important theory on singular densities and DLMs under singular observational variance is reviewed and developed. Results of the futures prices model under this singular structure are discussed for a particular data set including model diagnostics and forecasting.

In the second part of the dissertation, complex non-Gaussian hierarchical models are developed to profile providers in health-care delivery systems. The key motivating concern is to evaluate differences in return-time distribution by hospital and across years in the context of a range of possible individual-level explanatory variables. This study is part of a research project developed in collaboration with the Veterans Affairs (VA) Management Science Group. Chapter 6 introduces a class of hierarchical logistic models design to profile VA facilities within the fiscal year 1997. Chapter 7 extends the models to include a time series structure for the analyses of multiple years of data. The study involves innovation in hierarchical/longitudinal models for a very large and complex data set, a range of exploratory data analytic developments, customized MCMC for Bayesian model fitting and some creativity in exploring the very high-dimensional posterior distributions and summarizing MCMC outputs.

Finally, further potential developments and model extensions are discussed.

Chapter 2

Crude Oil Futures Prices Data

This chapter gives a general overview of crude oil markets and the kind of stochastic processes observed in commodity pricing. The main characteristics of the crude oil markets are addressed in the first part of the chapter emphasizing, among other things, the volatility structure present in oil-related products and general behavior of spot markets. The rest of the chapter describes a specific futures prices data set for different maturities together with a brief explanation on how *futures markets* are used to overcome spot market inefficiencies.

2.1 Crude Oil Markets and Volatility in Prices

Energy is perhaps the most strategic "material" in world commerce. The availability and cost of energy plays a central role in the economies of all countries. Among all energy products, crude oil dominates the energy market, being the largest cash commodity and a dominating influence on the rest of the products.

The worldwide oil market has five main components: spot, forward, futures, derivative and off-exchange (over-the-counter) markets. Crude oil and specific derivative products such as heating oil, gasoline, natural gas, among others, are traded in each market. The oil market has expanded enormously during the past twenty years.

In the early 1980s, the price most buyers paid was the official price set by the seller whereas today, almost all oil moving in international and domestic trade is sold at market price (Verleger, 1993).

The spot oil markets[1] are not very old and they have grown rapidly; until the early 1980s, the major oil companies used the spot markets only rarely. Now almost every oil related company has in some way a use of spot market services. They emerged when the supply chain from oil well to consumer was no longer in the hands of few entities and more participants appeared in the market.

Some important points to notice regarding oil markets are that there are many different spot prices due to the fact that crude oil is not uniquely definable. There are different grades and locations for this commodity and each spot contract may have its own unique characteristics. Furthermore, the market transactions are still not very well regulated (Verleger, 1993), and when the spot price is not very reliable, the corresponding futures contract[2] closest to maturity is used as an approximation for the spot price (Schwartz, 1997); especially for types of crude that match the subjacent commodity closely.

On the other hand, oil prices are generally more volatile than the prices of other commodities, stock market indices, interest rates or exchange rates. Moreover, the volatility in this market is naturally transferred into the prices of other energy products.

One of the major factors contributing to the volatility of oil prices is the disruption of supply and the drastic implications to the economy. For instance, the disruptions that followed the 1973 Arab embargo, the fall of the Shah of Iran in 1979, and the Iraqi invasion of Kuwait in 1990, each had roughly similar impacts on short-run supplies

[1] Market in which physical volumes of oil are exchanged for cash within two to four weeks of closing the deal.

[2] A futures contract is an agreement between two parties to buy or sell a specific amount of a commodity or financial instrument at a certain time in the future for a certain price (Hull, 1997).

and roughly identical impacts on both price levels and volatility (Verleger, 1993). High levels of volatility may lead to adverse impacts on producers or consumers. For example, credit institutions might reduce the amount of credit available to producers when prices are more volatile and the reduction in credit might depress long-run supply. On the other hand, consumers may be induced to make irrational investments to cut demand because prices have become more volatile.

These problems in the spot market and the levels of volatility in prices highlight the important role of *futures markets* and other financial derivative products in the oil industry.

2.2 Futures Market

The futures market for crude oil was established in March 1983 by the NYMEX after three main conditions for any futures market were satisfied (Dadkhah, 1992). These conditions are:

- the commodity is homogeneous, standardized and storable,
- the physical market is competitive with a large number of buyers and sellers, and
- the price of the commodity is uncertain, with no interventions and no government interferences.

The establishment of the futures market was one of the most important developments in the oil industry during the 80s. Since then, the futures market for oil has increased rapidly mainly due to the impressive influence of futures prices on the physical market. Futures provide a mechanism for the transfer of commodity risks and they make it possible for private information to be channeled into a forecast of commodity prices (Dadkhah, 1992). They provide a means by which oil producers can stabilize their incomes and reduce their exposures to financial risk associated with volatile prices (Verleger, 1993), and they make the inter-temporal allocation of supply and demand of the commodity more efficient. On the other hand, they also provide the energy industries with benchmark prices used as reference in numerous spot market transactions.

As a matter of fact, oil futures represent one of the most commonly used channels to combine investments on the energy sector with the use of financial instruments designed to diversify risk. See for instance Hull (1997), Stoll and Whaley (1993), Luenberger (1998) and NYMEX (1998b, 1998c), Clubley (1998), Verleger (1993), Dadkhah (1992), Rauscher (1989) for a more extensive review on futures and oil futures markets.

2.3 Crude Oil Futures Data

In order to understand, describe and eventually obtain reliable crude oil prices, twelve series of weekly observations of prices for crude oil futures contracts are analyzed. Figure 2.1 shows settlement prices[3] for crude oil futures contracts from January 1st, 1990 to October 18th, 1999, with a total of 512 data points in each series. These prices are publicly available and were obtained from data vendor DataStream. Each futures contract is on 1,000 U.S. barrels (42,000 gallons) of light, sweet crude oil[4] delivered at Cushing Oklahoma, see NYMEX (1998b, 1998c) for more detailed specifications on the contracts. In the plot, each series represents a different contract maturity from within a month to within seventeen months. To take a closer look at the data series

[3]The final price, established by exchange rule, for the prices prevailing during the closing period (NYMEX, 1998a).

[4]This includes six domestic and five foreign grades: West Texas Intermediate, Low Sweet Mix, New Mexican Sweet, North Texas Sweet, Oklahoma Sweet, South Texas Sweet, North Sea Brent and Forties Blend, Nigerian Bonny Light, Norwegian Oseberg Blend, and Colombian Cusiana. The foreign streams are delivered at a prespecified discount or premium (NYMEX, 1998b).

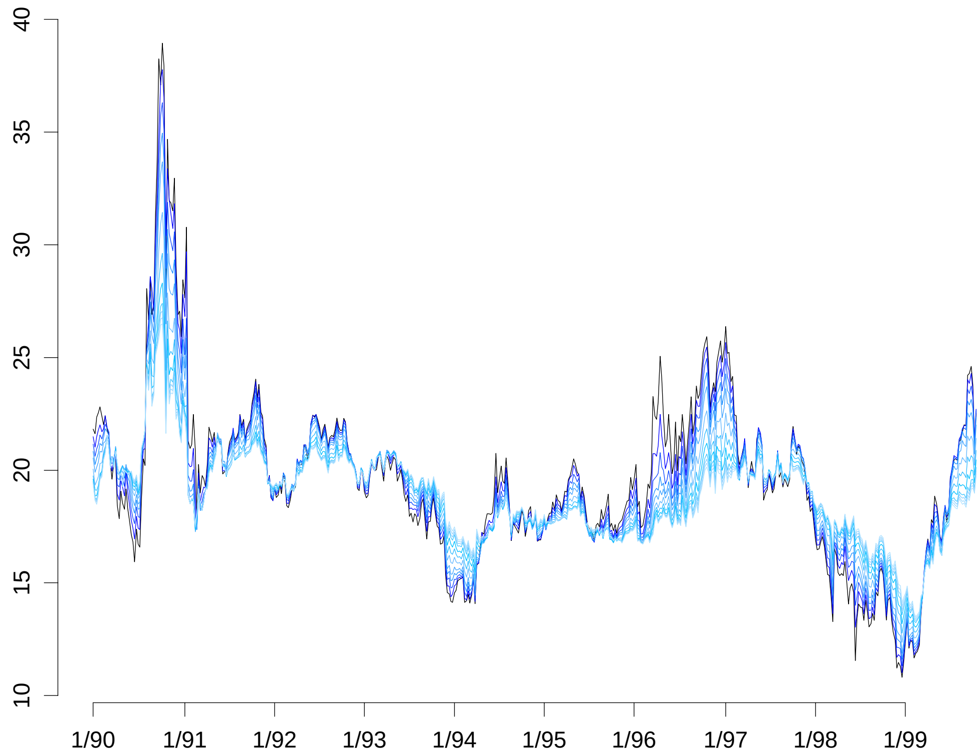

Figure 2.1: Crude Oil futures prices (settlement) from 01/01/90 to 10/18/99. Vertical scale is US dollars.

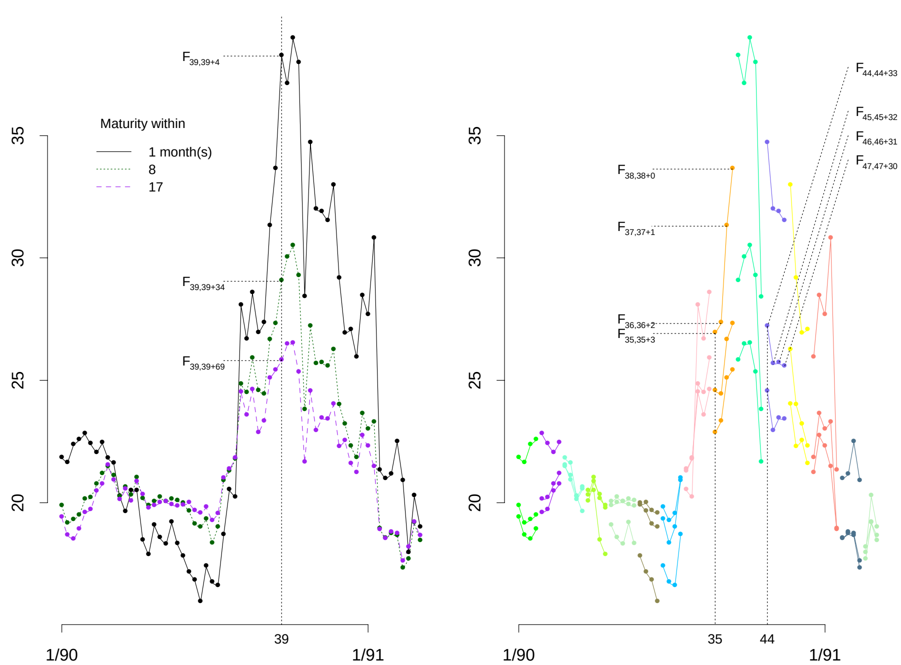

Figure 2.2: Crude oil futures prices from 01/01/90 to 03/11/91.

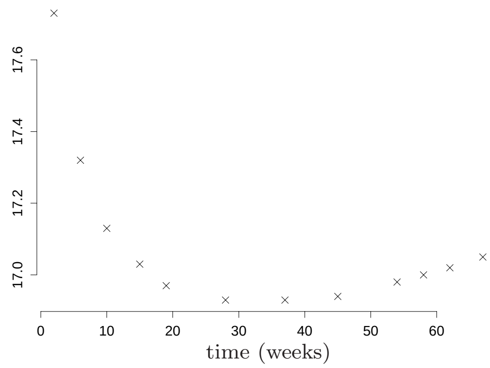

Figure 2.3: Term structure of Crude Oil futures prices at 05/09/94.

Figure 2.2 displays only three series: $\mathrm{F}_{t,t+k}$ denotes the price at time t of a futures contract that matures at time $t+k$; in the left panel, prices of futures contracts with three different maturities are displayed whereas in the right panel the "roll-over" of contracts is exemplified. That is, continuous line segments represent the prices of individual futures contracts rolling-over in time. For example, the price of one new contract that was first registered at week 35 is marked. In this case, the first data point in the first series (maturity within a month) is the price of a futures contract at week 35 that matures in 3 more weeks. The next observation in that series is from the same contract but one week later with maturity in 2 more weeks. In the same fashion, the price of the same contract is registered for four or five weeks until its maturity does not belong to the correspondent maturity of the series and the price of a different contract is recorded. For consistency, the changes in contracts are performed at the same time across all series.

Figure 2.3 provides a longitudinal view of the data at the beginning of the week of May 9th, 1994. This is one example of the behavior of futures prices at a given day over different maturities. It is clear that prices are not necessarily monotonic as a function of maturity. Actually, the figure displays a convex function with a downward trend in prices until they reach a minimum and then slowly growing price movements following a positive slope. Note that in this example, it takes approximately one year for the prices to "reach and follow" the long term line. This deviation from the line is usually associated with the effects on prices of stock-out and *convenience yield*[5] of the commodity. For some commodities, such as oil, inventories are small relative to consumption, and they can be quickly depleted if production is interrupted or if demand is suddenly increased. Therefore, prices can be dramatically affected in the short term by supply or demand shocks, inducing short term deviations from the equilibrium level as it is shown in Figure 2.3. Series with this characteristic are known as *mean-reverting* processes where prices tend to revert to the *mean* after a short term shock (Hull, 1997; Schwartz and Smith, 2000).

2.4 Summary

In this chapter, the importance of oil futures markets was highlighted together with a brief description of a data set of futures prices that will be used in subsequent chapters. As a matter of fact, oil futures markets play a central role in the physical market as instruments for hedging and also to help with pricing the products. An understanding of price risk management is key for the oil industry and related enterprises. The availability of insurance through futures and options markets allows buyers and sellers to use strategies to manage their exposure to market fluctuations and reduce their risk, making intervention less necessary. For these reasons, statistical analyses of futures prices are important in explaining commodity price movements and in making informed explanations available to market participants, observers and commentators.

[5]Benefits from owning an asset that are not obtained by the holder of a long futures contract on the asset (Hull, 1998).

Chapter 3

Multivariate Dynamic Linear Model for Oil Futures Prices

In order to understand and explain the sources of variability of oil futures prices and their impact on commodity pricing, a class of Bayesian multivariate dynamic linear models for oil futures prices is developed in this chapter. This new modeling framework represents direct generalizations of continuous-time models for commodity prices analyzed in Schwartz and Smith (2000) and reviewed in the first part of this chapter. A discrete-time version of this model and its state-space representation are then presented and discussed in a dynamic linear modeling framework. Due to the high complexity of the model, novel and customized Markov Chain Monte Carlo (MCMC) sampling schemes are developed for inference and computation representing a major component of the work presented here. These new highly complicated models are especially designed to follow standard financial theory for commodity pricing and hence represent important methodological contributions to both Bayesian statistics and financial science.

3.1 Latent Processes in Commodity Prices

In the previous chapter, the idea of *mean reversion* of oil futures prices was mentioned as one of the interesting characteristics of these financial instruments. It is said that prices are *mean-reverting* when they tend to revert to an equilibrium level after a short-term shock. In general, prices for some raw commodities tend to fluctuate randomly up and down in the short run in response to external shocks. In the particular case of crude oil, fluctuations are often attributed to reactions to political news and events such as wars, economic and diplomatic problems in oil-producing countries or strategic decisions made by the Organization of the Petroleum Exporting Countries (OPEC). However, "in the longer run [the price] ought to be drawn back towards the marginal cost of producing oil" (Dixit and Pindyck, 1994). Therefore, it is reasonable to assume that oil prices follow *mean-reverting processes*. For instance, Smith and Mccardle (1999) propose a model for commodity prices, including oil, where the log-prices follow an Ornstein-Uhlenbeck process, one of the simplest mean-reverting processes. On the other hand, Schwartz and Smith (2000) state that the log of the spot price is driven by two unobserved or latent processes where one of them follow an Ornstein-Uhlenbeck process. They claim that, under mean reversion, the log of the spot price can be decomposed into an equilibrium price level, and short-term deviations of the price from such level. The deviations might be caused by unexpected events affecting inventories due to the demand or the supply of the commodity, they are not expected to persist and should always be fluctuating around zero.

Although observed spot prices could be used to estimate the two latent processes mentioned above, some identifiability issues may emerge.[1] Therefore, information coming from futures prices will be key to explore the latent structure of crude oil

[1] In addition to the fact that the spot prices are not very reliable, as discussed in previous chapters.

markets. As a matter of fact, it is believed that changes in long maturity futures prices provide information about the equilibrium price level and changes in the price difference between near and long-term futures contracts provide information about the short-term deviations.

3.1.1 Continuous-Time Model

In this section, stochastic models for the two latent processes present in commodity prices are introduced and described following standard stochastic processes notation (Schwartz and Smith, 2000). Let S_t denote the spot price of a commodity at time t and define an instantaneous latent decomposition as

$$X_t = \xi_t + \chi_t,$$

where $X_t = \log(S_t)$, ξ_t represents the equilibrium level at time t and χ_t the corresponding short-term deviation in log prices at time t. The equilibrium level is then assumed to follow a Brownian motion process with drift μ_ξ, namely

$$d\xi_t = \mu_\xi dt + \sigma_\xi dz_\xi. \tag{3.1}$$

The short-term deviations are assumed to revert towards zero, following an Ornstein-Uhlenbeck process

$$d\chi_t = -\kappa\chi_t dt + \sigma_\chi dz_\chi, \tag{3.2}$$

where the mean reversion coefficient κ describes the rate at which the short-term deviations are expected to disappear. In these equations, dz_χ and dz_ξ are correlated increments of standard Brownian motion processes.

Equations (3.1) and (3.2) imply normal distributions for the χ_T and ξ_T processes with mean and covariance functions given by:

$$\mathrm{E}(\chi_T, \xi_T) = \left(e^{-\kappa T}\chi_0, \xi_0 + \mu_\xi T\right),$$

$$\mathrm{Cov}(\xi_T, \chi_T) = \begin{pmatrix} (1-e^{-2\kappa T})\sigma_\chi^2/2\kappa & (1-e^{-\kappa T})\rho_{\chi\xi}\sigma_\chi\sigma_\xi/\kappa \\ (1-e^{-\kappa T})\rho_{\chi\xi}\sigma_\chi\sigma_\xi/\kappa & \sigma_\xi^2 T \end{pmatrix},$$

where χ_0 and ξ_0 are initial values of the respective processes and $\rho_{\chi\xi}$ is the correlation of the Wiener processes z_χ and z_ξ. Therefore, given χ_0 and ξ_0, the log of the future spot price is normally distributed with mean and variance:

$$\begin{aligned} \mathrm{E}(X_T) &= e^{-\kappa T}\chi_0 + \xi_0 + \mu_\xi T, \\ \mathrm{Var}(X_T) &= (1-e^{-2\kappa T})\tfrac{\sigma_\chi^2}{2\kappa} + 2(1-e^{-\kappa T})\tfrac{\rho_{\chi\xi}\sigma_\chi\sigma_\xi}{\kappa} + \sigma_\xi^2 T. \end{aligned} \tag{3.3}$$

On one hand, the equations presented above give a neat decomposition of the spot price into latent processes that could be estimated if spot prices were reliable. On the other hand, theory suggests that futures prices are a legitimate source of information to estimate the short-term deviations and the equilibrium price level. Therefore, a link between futures prices and the two latent processes is needed and the traditional relationship between futures and expected spot prices arises as the best candidate. In the next section, possible modifications to the stochastic processes are proposed and analyzed in order to develop a natural connection between futures and spot prices.

3.1.2 Risk-Neutral Process

One of the oldest controversies in the theory of futures pricing is the relationship between futures price and the expected value of the spot price of the commodity at some future date. There are three hypotheses that propose a different relationship between futures and spot prices,

$$\mathrm{F}_{0,T} < \mathrm{E}_0(S_T), \tag{3.4}$$

$$\mathrm{F}_{0,T} > \mathrm{E}_0(S_T), \tag{3.5}$$

$$\mathrm{F}_{0,T} = \mathrm{E}_0(S_T). \tag{3.6}$$

where $F_{0,T}$ is the price today of a futures contract that matures in $t = T$, and $\mathrm{E}_0(S_T)$ is the expected value today of the spot price at $t = T$. Note that even in the case that either inequality (3.4) or (3.5) hold, a certain quantity could be added to one of the sides so that (3.6) will always be true. The latter equation relies on the notion of what is called "risk neutrality" where the expected profit to either position (buyer/seller) of a futures contract would be equal to zero; see Bodie *et al.* (1999) for more details on these hypotheses and futures pricing.

The *risk-neutral* or *risk-adjusted* assumption provides a natural link between the futures and spot prices and hence a necessary modification to estimate the two latent processes. Schwartz and Smith (2000) modified the continuous-time model presented in the previous section to include the risk-neutral assumption into the model. To support (3.6) under the risk-neutral assumption, two additional parameters are introduced, the risk premiums λ_χ and λ_ξ, which specify a constant reduction in the drifts of the processes (3.1) and (3.2). Namely,

$$d\xi_t^* = (\mu_\xi - \lambda_\xi)dt + \sigma_\xi dz_\xi^*, \tag{3.7}$$

$$d\chi_t^* = (-\kappa\chi_t^* - \lambda_\chi)dt + \sigma_\chi dz_\chi^*. \tag{3.8}$$

The risk-neutral long-term process (3.7) follows a Brownian motion process with drift $\mu_\xi^* = \mu_\xi - \lambda_\xi$ and the risk-neutral short-term process (3.8) follows an Ornstein-Uhlenbeck process reverting to $-\lambda_\chi/\kappa$.

As in (3.1) and (3.2) the processes χ_T^* and ξ_T^* are normally distributed with mean and covariance given by

$$\begin{aligned} \mathrm{E}(\chi_T^*, \xi_T^*) &= \left(e^{-\kappa T}\chi_0 - \left(1 - e^{-\kappa T}\right)\lambda_\chi/\kappa, \xi_0 + \mu_\xi^* T\right), \\ \mathrm{Cov}(\xi_T^*, \chi_T^*) &= \mathrm{Cov}(\xi_T, \chi_T), \end{aligned}$$

where $\chi_0^* = \chi_0$ and $\xi_0^* = \xi_0$. Therefore, under this risk-neutral assumption, the log of the future spot price is normally distributed with mean and variance:

$$\begin{aligned} \mathrm{E}(X_T^*) &= e^{-\kappa T}\chi_0 + \xi_0 - \left(1 - e^{-\kappa T}\right)\lambda_\chi/\kappa + \mu_\xi^* T, \\ \mathrm{Var}(X_T^*) &= \mathrm{Var}(X_T). \end{aligned}$$

At this point and under the modification induced by the risk-neutral hypothesis, a relationship between futures prices and the expected spot price is established and in theory, the two unobserved latent processes could be estimated using the information from futures commodity prices. However, continuous-time models assume instantaneous price movements and in practice only a finite amount of prices are recorded. Therefore, equivalent discrete-time process have to be developed in order to have realistic estimations of the latent processes and hence to understand the role that they play in commodity pricing.

3.1.3 Discrete-Time Model

The immediate discrete-time analogue of the continuous-time Ornstein-Uhlenbeck process is an AR(1), autoregressive process of order one. Hence the natural choice of discrete-time model for the short-term deviations is

$$\chi_t \;=\; \phi\chi_{t-1} + \omega_{1t}, \quad \omega_{1t} \sim N(0, \sigma_\chi^2),$$

where t now indexes discrete-time. Likewise, a Brownian motion process can be thought of as a continuous random walk, and therefore the natural discrete-time model for the equilibrium price level is

$$\xi_t \;=\; \mu_\xi + \xi_{t-1} + \omega_{2t}, \quad \omega_{2t} \sim N(0, \sigma_\xi^2),$$

ω_{1t} and ω_{2t} are correlated innovations at each point in time.

Under the risk-neutral hypothesis, two new parameters λ_χ and λ_ξ, representing the risk premiums and specifying a constant reduction in the drifts, are incorporated into the model as follows:

$$\begin{array}{rcll} \chi_t^* & = & \phi\chi_{t-1}^* - \lambda_\chi + \omega_{1t}^*, & \omega_{1t}^* \sim N(0, \sigma_\chi^2), \\ \xi_t^* & = & \mu_\xi - \lambda_\xi + \xi_{t-1}^* + \omega_{2t}^*, & \omega_{2t}^* \sim N(0, \sigma_\xi^2). \end{array} \tag{3.9}$$

Equations (3.9) imply that the processes χ_T^* and ξ_T^* are marginally normally distributed with mean and covariance functions:

$$\begin{aligned} \mathrm{E}(\chi_T^*, \xi_T^*) &= \left(\phi^T \chi_0 - \frac{1-\phi^T}{1-\phi} \lambda_\chi, \xi_0 + T\mu_\xi^* \right), \\ \mathrm{Cov}(\chi_T^*, \xi_T^*) &= \begin{pmatrix} \frac{1-\phi^{2T}}{1-\phi^2}\sigma_\chi^2 & \frac{1-\phi^T}{1-\phi}\sigma_{\chi\xi} \\ \frac{1-\phi^T}{1-\phi}\sigma_{\chi\xi} & T\sigma_\xi^2 \end{pmatrix}. \end{aligned}$$

Furthermore, given χ_0 and ξ_0, the log of the future spot price is marginally normally distributed with mean and variance:

$$\begin{aligned} \mathrm{E}(X_T^*) &= \phi^T \chi_0 - \tfrac{1-\phi^T}{1-\phi}\lambda_\chi + \xi_0 + T\mu_\xi^*, \\ \mathrm{Var}(X_T^*) &= \tfrac{1-\phi^{2T}}{1-\phi^2}\sigma_\chi^2 + 2\tfrac{1-\phi^T}{1-\phi}\sigma_{\chi\xi} + T\sigma_\xi^2, \end{aligned} \tag{3.10}$$

and finally, the spot price at time T is marginally lognormally distributed with mean and variance:

$$\begin{aligned} \mathrm{E}(S_T) &= \exp\left\{ \mathrm{E}(X_T^*) + \frac{1}{2}\mathrm{Var}(X_T^*) \right\}, \\ \mathrm{Var}(S_T) &= (\mathrm{E}(S_T))^2 \left(\exp\{\mathrm{Var}(X_T^*)\} - 1 \right). \end{aligned}$$

The equations above yield the desired relationships between futures prices, spot prices and the corresponding latent processes,

$$\log \mathrm{F}_{0,T} = \log \mathrm{E}(S_T) = \mathrm{E}(X_T^*) + \frac{1}{2}\mathrm{Var}(X_T^*), \tag{3.11}$$

which implies

$$\log \mathrm{F}_{0,T} = \left(\frac{1-\phi^{2T}}{1-\phi^2}, \frac{\phi^T - 1}{1-\phi}, T \right) \begin{pmatrix} \frac{1}{2}\sigma_\chi^2 \\ \lambda_\chi - \sigma_{\chi\xi} \\ \mu_\xi^* + \frac{1}{2}\sigma_\xi^2 \end{pmatrix} + (\phi^T, 1) \begin{pmatrix} \chi_0 \\ \xi_0 \end{pmatrix}, \tag{3.12}$$

where $F_{0,T}$ and T are recoded data and represent an observed price/maturity pair in one day.

3.1.4 Interpretation of Parameters

The expected value of the spot price on the log scale is equal to $\mathrm{E}(X_t) + \frac{1}{2}\mathrm{Var}(X_t)$. Given equation (3.3), this can be written as a function of time to maturity T as

$$g_S(T) = \left(\phi^T \chi_0 + \xi_0 + (1-\phi^{2T})\frac{\sigma_\chi^2}{4\kappa} + (1-\phi^T)\frac{\rho_{\chi\xi}\sigma_\chi\sigma_\xi}{\kappa}\right) + \left(\mu_\xi + \frac{1}{2}\sigma_\xi^2\right)T,$$

where $\phi = e^{-\kappa}$. The expression above will tend, for large[2] values of T, to

$$\lim_{T\to\infty} g_S(T) = \left(\xi_0 + \frac{\sigma_\chi^2}{4\kappa} + \frac{\rho_{\chi\xi}\sigma_\chi\sigma_\xi}{\kappa}\right) + \left(\mu_\xi + \frac{1}{2}\sigma_\xi^2\right)T.$$

A similar expression could be obtained for the futures prices from (3.10) and (3.11),

$$g_F(T) = \left(\phi^T \chi_0 + \xi_0 + \frac{\phi^T - 1}{1-\phi}\lambda_\chi + (1-\phi^{2T})\frac{\sigma_\chi^2}{4\kappa} + (1-\phi^T)\frac{\rho_{\chi\xi}\sigma_\chi\sigma_\xi}{\kappa}\right) + \left(\mu_\xi^* + \frac{1}{2}\sigma_\xi^2\right)T,$$

which for large values of T, will tend to

$$\lim_{T\to\infty} g_F(T) = \left(\xi_0 - \frac{\lambda_\chi}{1-\phi} + \frac{\sigma_\chi^2}{4\kappa} + \frac{\rho_{\chi\xi}\sigma_\chi\sigma_\xi}{\kappa}\right) + \left(\mu_\xi^* + \frac{1}{2}\sigma_\xi^2\right)T.$$

Note that μ_ξ and $\mu_\xi^* = \mu_\xi - \lambda_\xi$ play a central role on the slope of the long-term behavior of spot and futures prices, respectively, whereas λ_χ acts on the difference of these two price series. Consequently, spot and futures prices will be needed for inferences on these parameters. However, spot prices are never directly observed, as explained in Section 2.1, and there is not enough information in the futures prices and maturities to estimate μ_ξ. Consequently the precise location of the curve cannot be determined accurately (Schwartz and Smith, 2000). Furthermore, the meaning of "long-term" in this context is approximately six years, that is, the data would have to include prices with maturities in the range of about six years so that the long-term

[2]Note that for a value of $\phi = 0.98$ a "large" value of T would be at least 300 weeks ($\approx$ 6 years).

growth rate could be estimated. Futures prices with this long-term maturity are not presently publicly available.[3] Therefore, one has to bear in mind that inferences on μ_ξ, μ_ξ^* and λ_χ will not be easy using only futures prices data, and that informative priors may be needed.

3.1.5 State-Space Representation

It is well-known that Bayesian dynamic linear models provide a flexible framework to estimate linear models where certain parameters evolve stochastically over time. In this case, a state-space representation of the discrete-time model for commodity prices can be obtained by combining equations (3.9) and (3.12). Namely, if the data consist of r price/maturity pairs per day, then the r equations (3.12) combine in a multivariate model

$$\mathbf{y}_t \;=\; \mathbf{D}_t\boldsymbol{\mu}_0 + \mathbf{Z}_t\boldsymbol{\theta}_t, \tag{3.13}$$

where $\mathbf{y}_t$ is the $r \times 1$ vector of observations with $y_{it} = \log \mathrm{F}_{t,t+T_i}$, $\boldsymbol{\theta}_t' = (\chi_t, \xi_t)$ is the *state vector* of a DLM, $\boldsymbol{\mu}_0' = \left(\frac{1}{2}\sigma_\chi^2, \lambda_\chi - \sigma_{\chi\xi}, \mu_\xi^* + \frac{1}{2}\sigma_\xi^2\right)'$, $\mathbf{D}_t$ and $\mathbf{Z}_t$ are matrices with $(\frac{1-\phi^{2T_{it}}}{1-\phi^2}, -\frac{1-\phi^{T_{it}}}{1-\phi}, T_{it})$ and $(\phi^{T_{it}}, 1)$ as the *ith* rows, respectively, and $i = 1, ..., r$.

Moreover, following the suggestions of Schwartz (1997), an observational error with zero mean and variance $\mathbf{V}$ is included in (3.13). This error may take into account spreads, price limits, non-simultaneity of the observations and possible measurement errors in the data. Hence, the complete *observation equation* would be:

$$\mathbf{y}_t \;=\; \mathbf{D}_t\boldsymbol{\mu}_0 + \mathbf{Z}_t\boldsymbol{\theta}_t + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim \mathbf{N}(\mathbf{0}, \mathbf{V}), \tag{3.14}$$

On the other hand, equations (3.9) will form the *system equation*,

$$\boldsymbol{\theta}_t \;=\; \boldsymbol{\mu} + \mathbf{G}\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim \mathbf{N}(\mathbf{0}, \mathbf{W}), \tag{3.15}$$

[3] Futures on crude oil traded in a "formal" market are relatively new. As a result, the publicly available price history for long-term contracts is short. Moreover, futures with term to maturity greater than two years have almost no activity for days, therefore the "assigned" price may not reflect the future contracts supply and demand in that day.

where $\boldsymbol{\mu}' = (0, \mu_\xi)$, $\mathbf{W} = \begin{pmatrix} \sigma_\chi^2 & \sigma_{\chi\xi} \\ \sigma_{\chi\xi} & \sigma_\xi^2 \end{pmatrix}$ is the evolution variance and $\mathbf{G} = \begin{pmatrix} \phi & 0 \\ 0 & 1 \end{pmatrix}$. The *observation equation* (3.14) and *system equation* (3.15) together comprise the *state-space* representation of the discrete-time model. Finally, recall that the only observed quantities are the log futures prices $\mathbf{y}_t$ and their corresponding maturities $k_{1t}, ..., T_{rt}$.

3.1.6 Modified Discrete-Time Model

In this section, two modifications to the discrete-time model presented above are proposed. First, a change of measure of central tendency of the spot price is proposed when linking the futures prices to the state vector in order to add flexibility to the model and to simplify the estimation process. Second, heavy-tailed distributions are introduced into the system innovations to reflect the possibility of non-Gaussian errors.

Change of Measure of Central Tendency of the Spot Price

Under the risk-neutral assumption, the expected value of lognormally distributed spot prices proved to be the perfect link between the futures prices and the two latent processes: the equilibrium price level and short-term deviations. However, if a different measure of central tendency is taken instead of the mean, say the median, then equation (3.12) can be simplified to:

$$\log \mathrm{F}_{0,T} = \left(-\frac{1-\phi^T}{1-\phi}, T\right) \begin{pmatrix} \lambda_\chi \\ \tilde{\mu}_\xi^* \end{pmatrix} + (\phi^T, 1) \begin{pmatrix} \chi_0 \\ \xi_0 \end{pmatrix}. \tag{3.16}$$

This slight modification transforms the observation equation in (3.14) to have $\mathbf{D}_t$ as a matrix with *ith* row $(-\frac{1-\phi^{T_{it}}}{1-\phi}, T_{it})$ and $\boldsymbol{\mu}_0' = (\lambda_\chi, \tilde{\mu}_\xi^*)$. It is important to highlight that:

- The modified version of the discrete-time model simplifies the estimation process.

- If most of the variability on the log futures prices is explained by the state vector, then the first term in equation (3.12) will not be very important in the short-term.

- If the evolution variance is relatively small, differences between (3.12) and (3.16) will be negligible since the main difference in the two expressions lies in a term that is a function of the elements of the evolution variance.

- An even more important observation is that a linear, one-to-one, relationship between the parameters μ_ξ^* and $\tilde{\mu}_\xi^*$ can be easily established so that inferences on μ_ξ^* and $\tilde{\mu}_\xi^*$ are easily computed.

Heavy-tailed Distributions for System Innovations

A natural and critical extension to the already modified model is to consider heavy-tailed distributions for the system innovations. This new assumption enables the model to account for the volatile nature of crude oil futures prices that could be observed as outlying changes in the equilibrium price level and/or in the short-term deviations. Consequently, standard weight parameters γ_t are introduced to allow for innovation errors with heavy-tailed distributions constructed as a continuous scale mixture of normals with prior mixing distributions (West, 1984). In this way, non-normal behavior of the stochastic changes can be appropriately modeled with computational advantages due to the fact that these innovations are still conditionally normal.

The choice of the prior for the weight parameters will lead to different error distributions. For example, if a Gamma prior for the weights is selected, Gamma$(p/2, p/2)$, it implies that the innovations are marginally Student-T distributed with p degrees of freedom. The traditional normal model is a particular case of this general setting when $p \to \infty$ or equivalently $\gamma_t = 1$ for all t.

Thus the evolution equation (3.15) is now rewritten as

$$\boldsymbol{\theta}_t = \boldsymbol{\mu} + \mathbf{G}\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \qquad \omega_t \sim \mathbf{N}(\mathbf{0}, \mathbf{W}\gamma_t^{-1}),$$

with $\boldsymbol{\mu}'$, $\mathbf{W}$ and $\mathbf{G}$ as in (3.15).

Summarizing, a new class multivariate dynamic linear models for oil futures prices can be defined as

$$\begin{array}{rcll} \mathbf{y}_t & = & \mathbf{D}_t\boldsymbol{\mu}_0 + \mathbf{Z}_t\boldsymbol{\theta}_t + \boldsymbol{\epsilon}_t, & \boldsymbol{\epsilon}_t \sim \mathbf{N}(\mathbf{0}, \mathbf{V}), \\ \boldsymbol{\theta}_t & = & \boldsymbol{\mu} + \mathbf{G}\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, & \boldsymbol{\omega}_t \sim \mathbf{N}(\mathbf{0}, \mathbf{W}\gamma_t^{-1}), \end{array} \tag{3.17}$$

where $\mathbf{y}_t$ is a $r \times 1$ vector of observations with $y_{it} = \log \mathrm{F}_{t,t+T_i}$, $\boldsymbol{\theta}_t' = (\chi_t, \xi_t)$ is the *state vector*, $\boldsymbol{\mu}_0' = \left(\lambda_\chi, \tilde{\mu}_\xi^*\right)$, $\mathbf{V}$ is the observation variance, $\mathbf{D}_t$ and $\mathbf{Z}_t$ are matrices with $(-\frac{1-\phi^{T_{it}}}{1-\phi}, T_{it})$ and $(\phi^{T_{it}}, 1)$ as ith row respectively and $i = 1, ..., r$, $\boldsymbol{\mu}' = (0, \mu_\xi)$, $\mathbf{W} = \begin{pmatrix} \sigma_\chi^2 & \sigma_{\chi\xi} \\ \sigma_{\chi\xi} & \sigma_\xi^2 \end{pmatrix}$ is the evolution variance, $\mathbf{G} = \begin{pmatrix} \phi & 0 \\ 0 & 1 \end{pmatrix}$ and $\gamma_t \sim \mathrm{Gamma}(p/2, p/2)$.

3.2 Bayesian Analysis

Classical estimation methods for model (3.17) are based on approximate maximum likelihood estimators. The Bayesian framework allows for posterior estimates even when the likelihood function is intractable as it is in this case. To perform a full Bayesian analysis, the joint posterior distribution of all the unknown parameters should be calculated by updating the prior distribution via Bayes' rule. Here, the unknown parameters are

$$\Omega = \{\phi, \boldsymbol{\mu}_0, \mu_\xi, \mathbf{V}, \mathbf{W}, \gamma_1, \ldots, \gamma_N\} \text{ and } \Theta = \{\boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_N\}.$$

From model (3.17) the following joint distribution can be derived:

$$\begin{aligned} p(\Omega, \Theta, \mathbf{y}) &= p(\Omega)p(\mathbf{y}, \Theta|\Omega) \\ &= p(\Omega)\prod_{t=1}^{N}\mathbf{N}(\mathbf{y}_t|\mathbf{D}_t\boldsymbol{\mu}_0 + \mathbf{Z}_t\boldsymbol{\theta}_t, \mathbf{V})\mathbf{N}(\boldsymbol{\theta}_t|\boldsymbol{\mu} + \mathbf{G}\boldsymbol{\theta}_{t-1}, \mathbf{W}\gamma_t^{-1}). \end{aligned}$$

3.2.1 Prior Distributions

To complete the model specification, the joint prior distribution is specified in terms of conditionally independent components where the marginal priors are proper and generally chosen to be conditionally conjugate

$$p(\Omega) = p(\phi)p(\boldsymbol{\mu}_0)p(\mu_\xi)p(\mathbf{V})p(\mathbf{W})p(\gamma_1)\dots p(\gamma_N).$$

For some parameters vague priors are appropriate, but for others, prior information is essential for the analysis. As explained in Section 3.1.4, inference on parameters μ_ξ, μ_ξ^* and λ_χ will be almost impossible using futures prices data, and so informative prior distributions will be needed for the analyses. For these three parameters, informative conditionally conjugate priors centered on values suggested by Schwartz and Smith (2000) were used. They consider reasonable values $\mu_\xi \approx 3\%$, $\mu_\xi^* \approx 1.61\%$ and $\lambda_\chi \approx 5\%$ (annual terms), given investor expectations during the time period they analyzed, 1/15/93-5/16/96. For the persistence parameter ϕ, a beta prior is used with mode at 0.95, and a concentration mass of approximately 96% between 0.85 and 1. For the rest of the parameters for now, vague conditionally conjugate priors are used.

3.2.2 Implementation of the Gibbs Sampler

A Markov Chain Monte Carlo algorithm (MCMC) specifies an irreducible and aperiodic Markov Chain with stationary distribution given by the desired joint posterior distribution (Gelfand and Smith, 1990). An implementation of the posterior sampling algorithm is outlined here for the unknown parameters Ω and Θ. The sampling scheme is based on iterative updating using the full conditional densities of any subsets of the parameters. A generic such subset will be denoted by ζ, and the remaining variables combined with the full data set will be represented by ζ^-.

Sampling the state vector $\Theta_t|\Theta_t^-$

Simulations from the conditional posterior distribution for the state vector $\Theta_t|\Theta_t^-$ are performed by implementing the algorithm *Forward Filtering, Backward Sampling* (FFBS) (Carter and Kohn (1994) and Frühwirth-Schnatter (1994)). See West and Harrison (1997) for a detailed explanation of this sampling scheme.

Sampling the evolution variance $\boldsymbol{W}|\boldsymbol{W}^-$

Assuming a conjugate prior for the evolution matrix $\mathbf{W}$, the full conditional posterior distribution is given by

$$\text{Inv-Wishart}_n\left(\mathbf{W}\middle|\nu+N,\left(\sum_{t=1}^{N}\boldsymbol{\omega}_t\boldsymbol{\omega}_t'\gamma_t+\mathbf{S}\right)^{-1}\right)$$

where ν and $\mathbf{S}$ are the degrees of freedom and the scale parameter respectively of the inverse-Wishart prior distribution.[4]

Sampling the evolution variance weights $\gamma_t|\gamma_t{}^-$

The weights on the stochastic changes in the state vector have a common Gamma prior, Gamma$(p/2, p/2)$, which yields a Gamma full conditional posterior $\gamma_t|\gamma_t{}^- \sim$ Gamma$\left(\gamma_t|\frac{p+2}{2}, \frac{1}{2}\left(p+\boldsymbol{\omega}_t'\mathbf{W}^{-1}\boldsymbol{\omega}_t\right)\right)$. The degrees of freedom parameter p is set a priori; the traditional value of $p = 5$ (West, 1984) is chosen for future analysis.

[4]The notation for the Wishart and inverse-Wishart is given by $\mathbf{W} \sim \text{Wishart}_n(\mathbf{W}|\nu, S)$ where $\mathbf{W}$ is a $n \times n$ matrix with $\text{E}(\mathbf{W}) = \nu S$ and $\mathbf{U} \sim \text{Inv-Wishart}_n(\mathbf{U}|\nu, S^{-1})$ where $\mathbf{U}$ is a $n \times n$ matrix with $\text{E}(\mathbf{U}) = S/(\nu - n - 1)$. In general, all the density distributions presented in this dissertation follow closely the notation provided by Gelman *et al.* (1995).

Sampling the trend $\mu_\xi|\mu_\xi^-$

Assuming a prior distribution $p(\mu_\xi)$, the full conditional posterior distribution is given by $p(\mu_\xi)N\left(\mu_\xi\left|m,c^2\right.\right)$, where

$$m=\left(\sum_{t=1}^{N}\gamma_t\right)^{-1}\left(\sum_{t=1}^{N}(\xi_t-\xi_{t-1})\gamma_t-\frac{\rho_{\chi\xi}\sigma_\xi}{\sigma_\chi}\sum_{t=1}^{N}(\chi_t-\phi\chi_{t-1})\gamma_t\right)$$

and $c^2=\left(\sum\limits_{t=1}^{N}\gamma_t\right)^{-1}\sigma_\xi^2(1-\rho_{\chi\xi}^2)$. If a conjugate normal prior $N\left(\mu_\xi\left|m_0,c_0^2\right.\right)$ is chosen, the posterior distribution is given by $N(\mu_\xi\left|m_1,c_1^2\right.)$ where $m_1=c_1^2\left(\frac{m}{c^2}+\frac{m_0}{c_0^2}\right)$ and $c_1^2=\frac{c^2c_0^2}{c^2+c_0^2}$.

Sampling the vector $\boldsymbol{\mu}_0|\boldsymbol{\mu}_0^-$

Assuming a prior distribution $p(\boldsymbol{\mu}_0)$, the full conditional posterior distribution is given by $p(\boldsymbol{\mu}_0)\mathbf{N}_2\left(\boldsymbol{\mu}_0\left|\mathbf{m},\mathbf{C}\right.\right)$, where $\mathbf{m}=\mathbf{C}\left(\sum\limits_{t=1}^{N}\mathbf{Z}_t'\mathbf{V}^{-1}(\mathbf{y}_t-\mathbf{F}_t\boldsymbol{\theta}_t)\right)$ and $\mathbf{C}^{-1}=\sum\limits_{t=1}^{N}\mathbf{Z}_t'\mathbf{V}^{-1}\mathbf{Z}_t$. If a conjugate normal prior $\mathbf{N}_2\left(\boldsymbol{\mu}_0\left|\mathbf{m}_0,\mathbf{C}_0\right.\right)$ is chosen, the posterior distribution is $N(\boldsymbol{\mu}_0\left|\mathbf{m}_1,\mathbf{C}_1\right.)$ where $\mathbf{m}_1=\mathbf{C}_1\left(\mathbf{C}^{-1}\mathbf{m}+\mathbf{C}_0^{-1}\mathbf{m}_0\right)$ and $\mathbf{C}_1=\left(\mathbf{C}^{-1}+\mathbf{C}_0^{-1}\right)^{-1}$.

Sampling the persistence parameter $\phi|\phi^-$

Conditional on ϕ^-, the posterior for the autoregressive parameter is proportional to

$$q(\phi)=p(\phi)\prod_{t=1}^{N}\mathbf{N}(\mathbf{y}_t|\mathbf{D}_t(\phi)\boldsymbol{\mu}_0+\mathbf{Z}_t(\phi)\boldsymbol{\theta}_t,\mathbf{V})\mathbf{N}(\boldsymbol{\theta}_t|\boldsymbol{\mu}+\mathbf{G}(\phi)\boldsymbol{\theta}_{t-1},\mathbf{W}\gamma_t^{-1}).$$

Since this is not of known form, a Metropolis-Hastings sampler is required (Tierney, 1994). A uniform distribution is used as a proposal distribution in the Metropolis algorithm (Chib and Greenberg, 1995). That is, given a *current* value of ϕ, a *candidate* value ϕ^* is sampled from

$$\mathbf{U}(\phi^*|\phi-\delta,\phi+\delta)$$

and accepted with probability $\min\{1, q(\phi^*)/q(\phi)\}$. A Beta distribution is used as a prior, as discussed in Section 3.2.1.

Sampling the observational variance matrix $\boldsymbol{V}|\boldsymbol{V}^-$

The sampling of the observational variance-covariance matrix will depend on how $\mathbf{V}$ is modeled. Section 3.1 detailed a model on commodity prices based on financial and economic principles where observational error is added to include bid-ask spreads, price limits, non-simultaneity of the observations, errors in the data (Schwartz, 1997) and everything else that is not explained by the theoretical model. So far, it is not clear which structure should be assumed for the variance of these errors. For instance, it is possible that they are correlated depending on a function of the different times to maturity, or that their variability is being generated by a common source for all the series. Inferences on $\mathbf{V}$ are of critical relevance here due to the structure of futures prices data as noted before. Therefore, an in-depth study of the structure on $\mathbf{V}$ is needed to ensure proper understanding of crude oil prices structures.

3.3 Summary

In this chapter a class of multivariate dynamic linear models for oil futures prices is introduced. These models are based on theoretical foundations of a continuous-time model developed by Schwartz and Smith (2000). The models presented here use information on the futures prices to make inferences on two latent processes: an equilibrium price level and corresponding short-term deviations. Proper estimation of these unobserved series will play a central role in valuing financial contingent claims on the crude oil and in the procedures for evaluating investments to extract or produce it. An implementation of a posterior sampling algorithm via Gibbs sampling was outlined here for all model parameters, except for the observation variance $\mathbf{V}$. The structure of $\mathbf{V}$ will be analyzed in the next two chapters.

Chapter 4

Structures for the Observational Variance Matrix

The main purpose of this chapter is to emphasize the important role that the observational variance matrix plays in the estimation of the latent processes inherent in futures pricing for some commodities. Therefore, the modified discrete model, introduced in the previous chapter, is now completed by exploring different structures of the observational variance. First, a simple diagonal variance matrix is proposed following traditional approaches in the literature. Second, a more appropriate non-diagonal matrix is assumed to include potential non-zero correlations between futures contracts. Third, a variance matrix where the correlations are functions of the differences between maturities of the futures contracts is analyzed. Finally, the variance matrix is assumed to have a factor structure and is then decomposed into common and specific sources of variability. In each case, full Bayesian analyses are performed by obtaining the corresponding full conditional distribution of the observational variance and including it into the overall sampling scheme described above.

4.1 Diagonal Observational Variance

The first structure to explore for the observational variance is a simple diagonal matrix $\mathbf{V} = \text{diag}(\sigma_1^2, \ldots, \sigma_r^2)$ where r is the number of observed series (in this case $r = 12$). This structure assumes uncorrelated errors and has been the standard approach used in the literature due to the complexity of the models and the difficulties that maximum likelihood estimation methods encounter. Under the Bayesian framework, common inverse-Gamma prior distributions are assumed for each variance, Inv-Gamma$(\sigma_j^2|\nu/2, \nu S^2/2)$, so that the conditional posterior distribution for σ_j^2, $j = 1, \ldots, r$ is then Inv-Gamma$\left(\sigma_j^2 \middle| \frac{\nu+N}{2}, \frac{1}{2}\left(\sum_{t=1}^{N} \epsilon_{jt}^2 + \nu S^2\right)\right)$.

The Gibbs sampler algorithm described in the previous chapter is then completed with the full conditional posterior for $\mathbf{V}$, and samples from the desired joint posterior distribution are drawn for all parameters. As it turns out, the persistence parameter ϕ experiences slow convergence rates.[1] At first sight, one could think that this could be caused by either a problem with the Metropolis step in sampling ϕ as a consequence of the flat likelihood function or that the model is not appropriate.

A simulation study showed that the Metropolis step is indeed sampling from the correct posterior distribution for the autoregressive parameter with a model where observational errors are assumed uncorrelated. In this study, the original/true values for the parameters were included in their corresponding posterior distributions and convergence was achieved. Therefore, it can be concluded that the problem may come from assuming the wrong structure for the observational variance. For a closer evaluation of the uncorrelated-errors assumption (i.e. diagonal $\mathbf{V}$), the Gibbs sampler

[1]For this MCMC sampling output and for the rest of the MCMC simulations studied in this dissertation, convergence was assessed generally by a visual inspection of the autocorrelation and trace plots from different starting values. For some cases, a more formal analysis was done using convergence diagnostic tests implemented in CODA (Best *et al.*, 1995) and BOA (Smith, 2000) softwares. See, for example, Cowles and Carlin (1996), Brooks (1998) and Mengersen *et al.* (1999) for a review of convergence diagnostics.

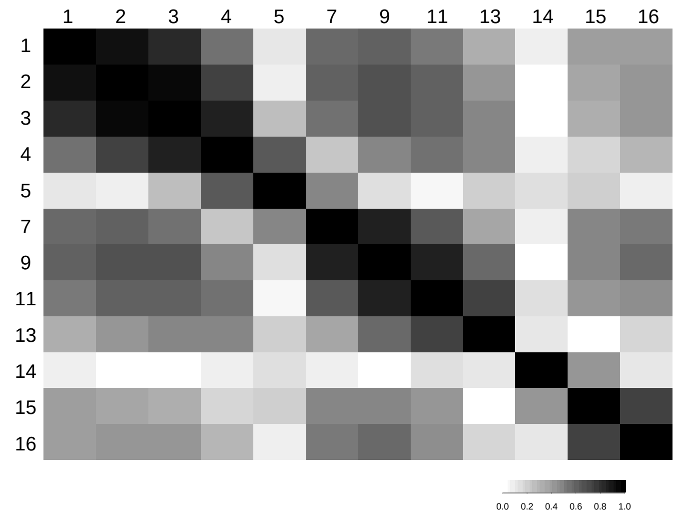

Figure 4.1: Correlation structure of the estimated residuals in a particular iteration.

was performed again with the persistence parameter held at a fixed value of 0.977 as suggested by (Schwartz and Smith, 2000). To further confirm our findings, Figure 4.1 presents a graphical representation of the correlation structure of the estimated residuals in a particular iteration selected at random. The range of color is (0,1),[2] where darker regions correspond to larger correlation values. As can be seen from the image, the empirical correlation structure of the residuals is quite inconsistent with the assumption of a diagonal observational variance matrix $\mathbf{V}$.

4.2 Non-diagonal Observational Variance

In this section, a more realistic, non-diagonal, full rank structure for $\mathbf{V}$ is considered to account for possible correlations between oil futures price errors at different maturities. In this case, the corresponding full conditional posterior distribution for $\mathbf{V}$

[2]The complete correlation range (-1,1) is not shown here, since all correlations are positive and it would limit the visual appreciation of the results.

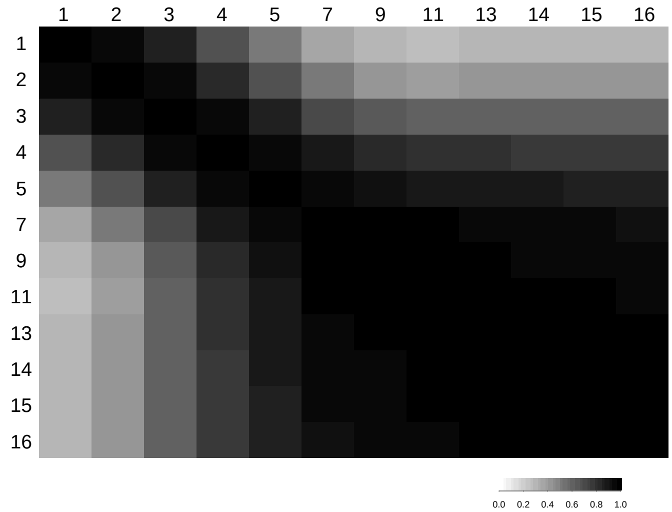

Figure 4.2: A sample from the posterior distribution of $\mathbf{V}$.

is easily computed as

$$\text{Inv-Wishart}_r \left(\mathbf{V} \middle| \nu + N, \left(\sum_{t=1}^{N} \boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t' + S \right)^{-1} \right)$$

when an inverse-Wishart prior distribution is assumed, namely $\text{Inv-Wishart}_r(\nu, S^{-1})$. As in the previous section, this full conditional will be incorporated into the Gibbs sampling scheme for the modified discrete time model as described in Section 3.2.2.

After a reasonable burn-in period, when the MCMC sampling output started to present convergence, the computer program experienced numerical instabilities when trying to invert a near singular matrix sampled from the posterior distribution of $\mathbf{V}$. This interesting behavior can be easily appreciated from the image representation of a sample from the posterior distribution of $\mathbf{V}$ as displayed in Figure 4.2. The graph suggests that the estimated residuals are highly collinear; the collinearity is especially marked between series with short-term maturities and between series with long-term maturities. These results indicate that either the observational variance $\mathbf{V}$ is close to singular or that the correlations implied by the covariance matrix are related through a function of the differences between maturities in the series. The latter conclusion

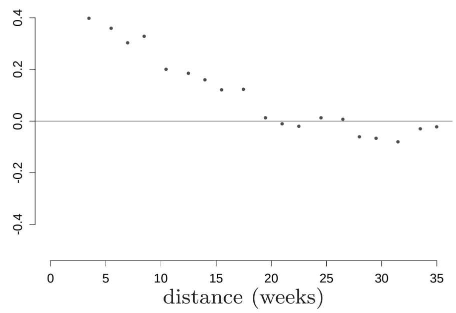

Figure 4.3: Autocorrelation function for the observed residuals $\hat{\boldsymbol{\epsilon}}_t$ estimated under the model assuming a non-diagonal $\mathbf{V}$.

will be explored in the next section by imposing a correlation structure of the series in function of their difference in maturities.

4.3 Structure in Function of Difference in Maturities

As with spatially correlated data, the observational errors corresponding to maturities within x months should be more highly correlated with errors corresponding to maturities within y months when $d = x - y$ is small. Therefore, a function of the difference in times to maturity between the series should be included in the correlation structure of the observational variance. A nice way to validate the hypothesis of "spatial structure" in the correlation matrix of the errors is to explore the autocorrelation function or correlogram of the estimated residuals as suggested in Cressie (1993) and Wackernagel (1998). Figure 4.3 displays the autocorrelation function of one subset of the posterior mean of the residuals estimated under the model with non-diagonal $\mathbf{V}$ above. The plot shows a decreasing trend in correlations as the difference in time to maturity increases supporting the "spatial" structure hypothesis.

Note that since maturities on the futures contracts change over time (see description of the data set in Section 2.3), the spatial structure yields to different correlation matrices at each point in time and therefore a time-varying structure for the observational variance is implicitly introduced into the model. Actually, write the observational variance $\mathbf{V}_t$ as

$$\mathbf{V}_t = \mathbf{\Sigma}^{1/2}\mathbf{R}_t\mathbf{\Sigma}^{1/2},$$

where $\mathbf{\Sigma} = \text{diag}(\sigma_1^2, \ldots, \sigma_r^2)$ and $\mathbf{R}_t$ is the correlation matrix. In this case, the $\{R_{ijt}\}$ element of $\mathbf{R}_t$ will be a function of d_{ijt}, the difference in time to maturity between series i and j at time t. The choice of the correlation function usually is a hard and complex task. The most commonly used class of correlation function is adopted here, it has a simple functional form and allows for some flexibility in modeling correlations:

$$R_{ijt} = \exp\left\{-\frac{|d_{ijt}|^a}{\tau}\right\} \tag{4.1}$$

where $0 < a \leq 2$ and $\tau > 0$ (Wackernagel, 1998). In the spatial statistics literature, a is usually assumed to determine commonly used correlation functions, for example $a = 1$ corresponds to *exponential* correlation structure and $a = 2$ corresponds to *Gaussian* correlation structure.

Now, the observational variance is defined in terms of the parameters a, τ, and $\mathbf{\Sigma}$. In this case, full Bayesian posterior inferences will be performed for a fixed value of a which implies sampling from the full conditional distributions for parameters $\Upsilon = \{\tau, \sigma_1^2, \ldots, \sigma_r^2\}$ as part of the overall Gibbs sampling algorithm presented in the previous chapter.

It is important to bear in mind that a key assumption for this model is that the correlation structure should depend only on the difference in maturities. That is, that series corresponding to maturities within sixteen and fifteen months should have very similar correlation than series with maturities within one and two months.

4.3.1 Bayesian Analysis and Implementation of the Gibbs Sampler

This section provides an outline of the iterative posterior sampling algorithm for the unknown parameter Υ as part of a sampling scheme to make inferences to the rest of the parameters. The full conditional distributions of the additional parameters $\mathbf{\Sigma}$ and τ, are presented in the two following sections, Metropolis-Hastings sampling is necessary for both of them. To start, given a prior density for $\mathbf{\Sigma}$ and τ and the joint distribution

$$\begin{aligned} p(\Omega, \Theta, \mathbf{y}) &= p(\Omega)p(\mathbf{y}, \Theta|\Omega) \\ &= p(\Omega)\prod_{t=1}^{N}\mathbf{N}(\mathbf{y}_t|\mathbf{D}_t\boldsymbol{\mu}_0 + \mathbf{Z}_t\boldsymbol{\theta}_t, \mathbf{V})\mathbf{N}(\boldsymbol{\theta}_t|\boldsymbol{\mu} + \mathbf{G}\boldsymbol{\theta}_{t-1}, \mathbf{W}\gamma_t^{-1}), \end{aligned}$$

where $\Omega = \{\phi, \boldsymbol{\mu}_0, \mu_\xi, \mathbf{\Sigma}, \tau, \mathbf{W}, \gamma_1, \ldots, \gamma_N\}$ and $\mathbf{\Sigma} = \text{diag}(\sigma_1^2, \ldots, \sigma_r^2)$, the full conditionals on $\mathbf{\Sigma}$ and τ can be derived, and therefore, a sample from their posterior distribution can be obtained. Again, prior independence is assumed for all the parameters:

$$p(\Omega) = p(\phi)p(\boldsymbol{\mu}_0)p(\mu_\xi)p(\mathbf{\Sigma})p(\tau)p(\mathbf{W})p(\gamma_1)\ldots p(\gamma_N).$$

Sampling the parameter $\tau|\tau^-$

With a Gamma prior distribution $p(\tau)$, the conditional posterior density for τ is proportional to

$$q(\tau) = p(\tau)\prod_{t=1}^{N}|\mathbf{R}_t|^{-1/2}\text{exp}\left\{-\frac{1}{2}\boldsymbol{\epsilon}_t^{*\prime}\mathbf{R}_t^{-1}\boldsymbol{\epsilon}_t^*\right\},$$

where $\boldsymbol{\epsilon}_t^* = \mathbf{\Sigma}^{-1/2}\boldsymbol{\epsilon}_t$, $\boldsymbol{\epsilon}_t = \mathbf{y}_t - (\mathbf{D}_t\boldsymbol{\mu}_0 + \mathbf{Z}_t\boldsymbol{\theta}_t)$ and $\mathbf{R}_t(\tau)$ as defined in (4.1). A Gamma distribution is adopted as the proposal density. The proposal distribution at the *current* value τ in the MCMC is $\pi(\tau^*|\tau) = \text{Gamma}(\tau^*|\alpha(\tau, b), \beta(\tau, b))$ where $\beta(\tau, b) =$

τ/b and $\alpha(\tau, b) = \tau\beta(\tau, b)$, so that the proposal mean is τ and the variance is $b = 0.1$. Therefore, the candidate τ^* will be accepted with probability $\min\left\{1, \frac{q(\tau^*)\pi(\tau|\tau^*)}{q(\tau)\pi(\tau^*|\tau)}\right\}$.

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

Assume independent prior densities for each of the variances, σ_j, $j = 1, \ldots, r$ so that $p(\boldsymbol{\Sigma}) = p(\sigma_1^2)\ldots p(\sigma_r^2)$. The conditional posterior distribution is proportional to

$$q(\boldsymbol{\Sigma}) = p(\boldsymbol{\Sigma}) \prod_{t=1}^{N} |\boldsymbol{\Sigma}|^{-1/2} \exp\left\{-\frac{1}{2}\boldsymbol{\epsilon}_t' \left(\boldsymbol{\Sigma}^{-1/2}\mathbf{R}_t^{-1}\boldsymbol{\Sigma}^{-1/2}\right)^{-1} \boldsymbol{\epsilon}_t\right\}.$$

The following proposal is used to approximate the conditional distribution of $\boldsymbol{\Sigma}$:

$$\pi(\boldsymbol{\Sigma}^*) \propto \prod_{t=1}^{N} |\boldsymbol{\Sigma}^*|^{-1/2} \exp\left\{-\frac{1}{2}\boldsymbol{\epsilon}_t'(\boldsymbol{\Sigma}^*)^{-1}\boldsymbol{\epsilon}_t\right\},$$

or equivalently inverse-Gamma distributions, Inv-Gamma$\left(\sigma_j^2 \middle| \frac{N}{2}, \frac{1}{2}\sum_{t=1}^{N} \boldsymbol{\epsilon}_{jt}^2\right)$, for each variance σ_j^2, $j = 1, \ldots, r$. The candidate $\boldsymbol{\Sigma}^*$ will be accepted with probability

$$\min\left\{1, \frac{q(\boldsymbol{\Sigma}^*)\pi(\boldsymbol{\Sigma})}{q(\boldsymbol{\Sigma})\pi(\boldsymbol{\Sigma}^*)}\right\}$$

to ensure that the samples are from the true posterior distribution. An important point to note here is that informative inverse-Gamma priors were used in this case, giving a negligible prior probability that the variances are close to zero; this is vital for the good mixing behavior of the parameters.

4.3.2 Analysis Results

The MCMC sampling scheme described in Section 3.2.2 was performed to draw 500,000 samples after discarding the first 100,000 iterations as the "burn-in" period. Of these, 10,000 equally spaced subsamples were extracted. Figure 4.4 displays

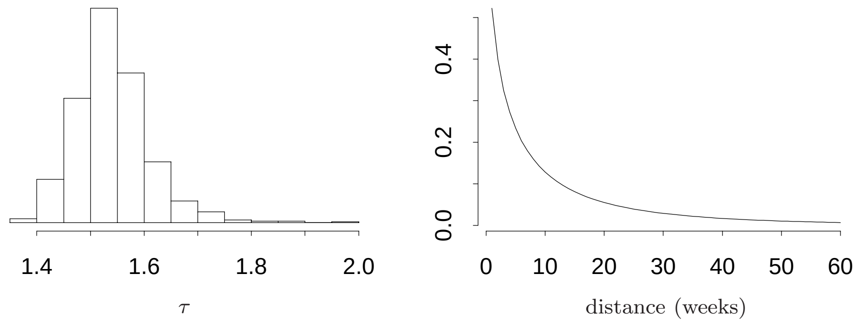

Figure 4.4: Histogram of simulated sample from the posterior density of τ (left) and the posterior mean of the correlation function $\rho(d) = \exp\left\{\frac{d^{1/2}}{\tau}\right\}$ (right).

posterior summaries of the parameter τ assuming a fixed value[3] of $a = 1/2$. The first panel shows a histogram of the simulated values of τ and the second displays the posterior mean of the correlation function. The second graph shows small values of correlations between observational errors corresponding to series with maturities more than 10 weeks apart. Recall that the data include maturities between zero and seventy weeks where traditional differences are 4, 8, 13, 17, 22, 26, 30, 34, 39, 44, 48, 52, 56, 60 and 65 weeks. Therefore, based on posterior inferences on τ, most of the resulting estimated correlations implied from the observational variances are small and close to zero. However, analyses from Section 4.1 suggested that errors are positively correlated and that a diagonal structure for the observational variance is not appropriate. Figure 4.5 displays 95% posterior intervals for $\mathbf{\Sigma}^{1/2} = \text{diag}(\sigma_1, \ldots, \sigma_r)$, with the posterior median indicated by the symbol $\times$. Note that the variances present a U-shaped pattern with the minimum value at five months to maturity. Moreover, the variances are small for all but the short maturity series. It is important to note that, although convergence of the MCMC was eventually achieved, many iterations were needed since the parameters ϕ and $\mathbf{\Sigma}$ presented slow mixing due to near singular variances implied by the small values of some elements in $\mathbf{\Sigma}$.

[3]The analysis was done for other values of $a = 0.1, 1, 2$. However, the mixing of τ and $\mathbf{\Sigma}$ was notoriously slower and convergence was barely achieved. Further study to infer a might be worthwhile where this model to be pursued in new applications.

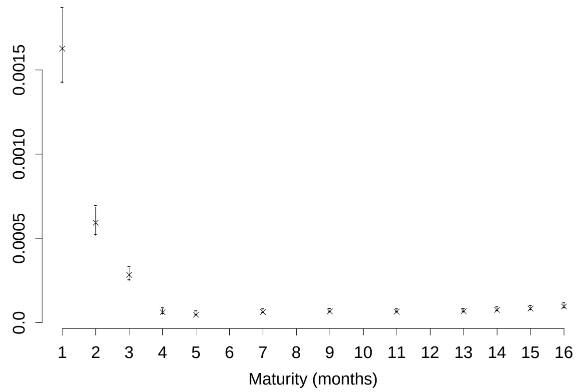

Figure 4.5: 95% posterior intervals for $\mathbf{\Sigma}$.

To further study the structure of the observational variance, a principal components analysis (PCA) was performed on the observed residuals[4] to explain the possibility of common latent sources of variability. As it turns out, approximately 85% of the overall variability is explained by the first component, and around 97% of the variability is due to the first three components. The first component has important weights associated with the first four residual series, which correspond to futures contracts with short-term maturities. The second and third components have higher weights for series with mid-term and long-term maturities respectively. These interesting findings suggest the existence of common latent processes driving the variability of the model residuals. Therefore, a factor model representation seems to be a reasonable tool to decompose the observational variance $\mathbf{V}$ into two sources of variability: one representing common sources of variability across the series and another one representing series-specific variations.

[4]Particular iterations selected at random.

4.4 Factor Model

Factor analysis is a mathematical model which attempts to explain the correlation between a large set of variables in terms of a small number of underlying *factors* (Mardia *et al.*, 1979). Factor analysis is primarily concerned with explaining the covariance between variables by identifying the sources of variation. The factor model assumes that all the correlations are explained by the common factors and the residual variation comes from uncorrelated variable-specific sources (Press, 1985). In this section, a factor model representation is assumed for the observational errors with the purpose of identifying common sources of variability across series and specific contributions of individual series (maturities).

A basic k-factor model for the covariance matrix $\mathbf{V}$ is defined by:

$$\mathbf{V} = \mathbf{X}\mathbf{X}' + \mathbf{\Psi} \tag{4.2}$$

where $\mathbf{X}$ is the $r \times k$ factor loadings matrix with columns $\mathbf{x}_j$ and $\mathbf{\Psi} = \mathrm{diag}(\psi_1, \ldots, \psi_r)$ is the diagonal matrix of instantaneous, series-specific or "idiosyncratic" variances. The elements on the diagonal of the factor covariance matrix $\mathbf{X}\mathbf{X}'$ are sometimes called *commonalities* $\mathbf{x}_i = \sum_{j=1}^{k} x_{ij}^2$ for $i = 1, \ldots, r$, and the elements of $\mathbf{\Psi}$ are sometimes called *specificities* or *uniquenesses*. In terms of the residual series $\boldsymbol{\epsilon}_t$, this is equivalent to the representation:

$$\boldsymbol{\epsilon}_t = \mathbf{X}\mathbf{f}_t + \boldsymbol{\nu}_t \tag{4.3}$$

where $\mathbf{f}_t$ is a k-dimensional random vector of common factors, or *factor scores*, and $\boldsymbol{\nu}_t$ is a r-dimensional random vector of conditionally independent and series-specific quantities, or unique factors. Traditional assumptions for this model are:

- $\boldsymbol{\nu}_t \sim \mathbf{N}(\boldsymbol{\nu}_t | \mathbf{0}, \mathbf{\Psi})$,
- uncorrelated and standardized factors $\mathbf{f}_t \sim \mathbf{N}(\mathbf{f}_t | \mathbf{0}, \mathbf{I}_k)$ and

- $\boldsymbol{\nu}_t$ and $\mathbf{f}_s$ mutually independent for all t, s.

Therefore, the observation equation of the futures prices in model (3.17) can be rewritten as

$$\mathbf{y}_t = \mathbf{D}_t\boldsymbol{\mu}_0 + \mathbf{Z}_t\boldsymbol{\theta}_t + \mathbf{X}\mathbf{f}_t + \boldsymbol{\nu}_t. \tag{4.4}$$

As it is well known in the literature, the k-factor model must be constrained to guarantee identifiability and uniqueness (Aguilar and West, 2000). To ensure this, the loadings matrix $\mathbf{X}$ is traditionally constrained to have a lower triangular shape, see Aguilar (1998), Chapter 6 and references therein for details. That is,

$$\mathbf{X} = \begin{pmatrix} x_{11} & 0 & 0 & \cdots & 0 \\ x_{21} & x_{22} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ x_{k1} & x_{k2} & x_{k3} & \cdots & x_{kk} \\ x_{k+1,1} & x_{k+1,2} & x_{k+1,3} & \cdots & x_{k+1,k} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ x_{r1} & x_{r2} & x_{r3} & \cdots & x_{rk} \end{pmatrix} \tag{4.5}$$

where $x_{ii} > 0$ for $i = 1, \cdots, k$, and $x_{ij} = 0$ for $i < j$, $i, j = 1, \cdots, k$. This form of the loadings matrix ensures invariance of the model under invertible linear transformations of the factor vectors, and induces a full rank loadings matrix. On the other hand, the conditions imposed on the loading matrix yield nice interpretations of the factor scores. In addition, it induces an upper bound on the number of factors k, given by the integer part of

$$r + 1/2 - \sqrt{1 + 8r}/2. \tag{4.6}$$

The main idea is to impose enough conditions to ensure that the number of free parameters in the factor representation does not exceed the $r(r+1)/2$ parameters in the unrestricted $\mathbf{V}$; see Aguilar and West (2000) and Aguilar (1998) for further details.

4.4.1 Bayesian Analysis and Implementation of the Gibbs Sampler

The previous Bayesian/MCMC analysis is now extended to incorporate the factor representation on the observational variance $\mathbf{V}$. Full conditional distributions are computed for the unknown parameters $\Upsilon = \{\mathbf{\Psi}, \mathbf{X}, \mathbf{f}_t; \forall t\}$ along the lines of Geweke and Zhou (1996), Polasek (1999), Aguilar (1998), Pitt and Shephard (1999) and Aguilar and West (2000).

Traditional prior distributions for factor model parameters are assumed to calculate the full conditional distributions. Namely,

$$p(\Upsilon) = p(\mathbf{\Psi}, \mathbf{X}, \mathbf{F}) = p(\mathbf{F})p(\mathbf{X})p(\mathbf{\Psi}), \tag{4.7}$$

where $\mathbf{F} \sim \mathbf{N}(\mathbf{0}, \mathbf{I}_N, \mathbf{I}_k)$[5] is the $N \times k$ factors matrix $\mathbf{F} = (\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_N)'$, $p(x_{ij}) \propto c$ for the non-zero entries of the loadings matrix $\mathbf{X}$, and independent priors $p(\tau_i) = \text{Gamma}(\tau_i|\alpha_0, \beta_0)$, $i = 1, \ldots, r$, for the precisions $\tau_i = 1/\psi_i$.

Now the conditional distribution of the residuals $\boldsymbol{\epsilon}_t = \mathbf{y}_t - (\mathbf{D}_t\boldsymbol{\mu}_0 + \mathbf{Z}_t\boldsymbol{\theta}_t)$, as described in (4.3), is given by

$$\boldsymbol{\epsilon}_t|\mathbf{X}, \mathbf{f}_t, \mathbf{\Psi} \sim \mathbf{N}(\boldsymbol{\epsilon}_t|\mathbf{X}\mathbf{f}_t, \mathbf{\Psi}), \tag{4.8}$$

and the corresponding distribution unconditional on the factors is

$$\boldsymbol{\epsilon}_t|\mathbf{X}, \mathbf{\Psi} \sim \mathbf{N}(\boldsymbol{\epsilon}_t|\mathbf{X}\mathbf{X}' + \mathbf{\Psi}). \tag{4.9}$$

Samples from the posterior distribution for Υ can be obtained by sampling iteratively from the full conditional densities of each unknown parameter. Using the same notation as before, any subsets of the unknown parameters Υ will be denoted ζ and ζ^- will represent the remaining variables combined with the full data set.[6]

[5] Notation for a matrix normal distribution; Dawid (1981) and West and Harrison (1997).

[6] The author thanks Omar Aguilar for provision of the code to sample the factor model parameters.

Sampling the factors $\boldsymbol{F}|\boldsymbol{F}^{-}$

Assuming a prior $\mathbf{f}_t \sim \mathbf{N}(\mathbf{f}_t|\mathbf{0}, \mathbf{I}_k)$, the full conditional distribution for the factors can be written in matrix normal form allowing to sample all the factors at once,

$$\mathbf{F}|\mathbf{F}^{-} \sim \mathbf{N}\left(\mathbf{F}\left|\boldsymbol{\epsilon}^{*}\boldsymbol{\Psi}^{-1}\mathbf{X}(\mathbf{I}_k + \mathbf{X}'\boldsymbol{\Psi}^{-1}\mathbf{X})^{-1}, \mathbf{I}_n, (\mathbf{I}_k + \mathbf{X}'\boldsymbol{\Psi}^{-1}\mathbf{X})^{-1}\right.\right),$$

where $\mathbf{F} = (\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_n)'$ is the $N \times k$ factors matrix and $\boldsymbol{\epsilon}^{*}$ is an $N \times r$ matrix with rows $\boldsymbol{\epsilon}_t$ for $t = 1, ..., N$.

Sampling the idiosyncratic variances $\boldsymbol{\Psi}|\boldsymbol{\Psi}^{-}$

Assuming independent inverse-Gamma priors, the full conditional posterior distribution for the variances are conditionally independent inverse-Gamma posteriors. For the precisions $\tau_i|\tau_i^{-} \sim \text{Gamma}(\tau_i|(\alpha_0+N)/2, (\beta_0+e_i)/2)$, where $e_i = \sum_{t=1}^{n}(\boldsymbol{\epsilon}_{it}-\mathbf{x}_i'\mathbf{f}_t)^2$ and $\mathbf{x}_i$ is the i-th row of the loadings matrix $\mathbf{X}$ for each $i = 1, \ldots, r$. Informative priors are used for the specific variances to give a small prior probability that the variances are close to zero.

Sampling the loadings matrix $\boldsymbol{X}|\boldsymbol{X}^{-}$

Under non-informative priors, the elements in each row of the loadings matrix (4.5) can be sampled independently:

1. The full conditional posterior distribution for the first element x_{11} is a truncated normal,

$$x_{11}|(x_{11})^{-} \sim N\left(x_{11}\left|(f_{1t}'f_{1t})^{-1}(f_{1t}'\boldsymbol{\epsilon}_1), \psi_1(f_{1t}'f_{1t})^{-1}\right.\right)\mathbf{I}_{\{x_{11}>0\}},$$

   where $\boldsymbol{\epsilon}_1 = (\boldsymbol{\epsilon}_{11}, \ldots, \boldsymbol{\epsilon}_{1N})'$ is the first series with $\boldsymbol{\epsilon}_{1t} = x_{11}f_{1t} + \boldsymbol{\nu}_{1t}$ for $t = 1, \ldots N$.

2. The full conditional posterior distribution for the j-th row $\mathbf{x}_{j\cdot} = (x_{j1}, \ldots, x_{jj})$ is the following truncated multivariate normal,

$$\mathbf{x}_{j\cdot}|(\mathbf{x}_{j\cdot})^{-} \sim N\left(\mathbf{x}_{j\cdot}\left|(\mathbf{B}_j'\mathbf{B}_j)^{-1}(\mathbf{B}_j'\boldsymbol{\epsilon}_j), \psi_j(\mathbf{B}_j'\mathbf{B}_j)^{-1}\right.\right)\mathbf{I}_{\{x_{jj}>0\}},$$

   where $\boldsymbol{\epsilon}_j$ is the $N \times 1$ vector of the first j series and $\mathbf{B}_j$ is a $N \times j$ matrix with rows $(f_{1t}, \ldots, f_{jt})'$ for $t = 1, \ldots N$.

3. The remaining $r - k$ 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 $r - k$ rows of the loadings matrix. The full conditional distribution for the $(r - k) \times k$ matrix $\mathbf{Z} = \mathbf{X}_2$ is,

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

   where $\boldsymbol{\Psi}_{(k+1):r} = \mathrm{diag}(\psi_{k+1}, \ldots, \psi_r)$,$\boldsymbol{\epsilon}_{(k+1):r}$ is the $N \times j$ matrix of series $k+1$ through r and $\mathbf{B}$ is a $N \times k$ matrix with rows $(f_{1t}, \ldots, f_{kt})'$ for $t = 1, \ldots N$.

Note that in the factor model, the number of factors k is assumed to be known and in this case the PCA from previous section pointed out the possibility of two or three common factors. Model selection on the number of factors is addressed in the next section.

4.4.2 Model Selection Using the BIC

In this section, an approximate Bayesian model selection method will be used to guide choice of the appropriate number of factors in model (4.4).

The first issue to address is how many models have to be considered. As mentioned above, in order to have full model identification, the number of factors k is subject to an upper bound given by the integer part of (4.6). In this case, the maximum number of factors to account is seven, therefore the models to compare are $M_1, \ldots, M_7$ where M_j assumes j factors. Denote the prior probability of M_j by $p(M_j)$. Bayes' theorem implies that the posterior probability of M_j is given by

$$p(M_j|D) = \frac{p(D|M_j)p(M_j)}{\sum_{i=1}^{7} p(D|M_i)p(M_i)}$$

where D is the observed data and $p(D|M_j)$ represents the marginal data density under model M_j.

Using a standard asymptotic approximation based on the Laplace method for integrals, these posterior probabilities can be approximated using the *Bayesian Information Criterion* (BIC) (Schwarz, 1978),

$$p(M_j|D) \approx \frac{\exp\left(-\frac{1}{2}BIC_j\right)}{\sum_{i=1}^{7} \exp\left(-\frac{1}{2}BIC_i\right)},$$

where, for each integer k,

$$BIC_k = -2l_k + d(k)\log(N), \tag{4.10}$$

$d(k)$ is the number of parameters in model M_k and l_k is the maximum of the log-likelihood. Specifically, $l_k = \log\, p(\mathbf{Y}|\hat{\Omega}_k, k)$ where

$$p(\mathbf{Y}|\hat{\Omega}_k, k) = \prod_{t=1}^{N} p(\mathbf{y}_t|D_{t-1}, \hat{\Omega}_k, k), \tag{4.11}$$

$\mathbf{Y} = (\mathbf{y}_1, \ldots, \mathbf{y}_N)$, $\hat{\Omega}$ is MLE of $\Omega = \{\phi, \boldsymbol{\mu}_0, \mu_\xi, \mathbf{X}_k, \boldsymbol{\Psi}, \mathbf{W}, \gamma_t; \forall t\}$ and

$$p(\mathbf{y}_t|D_{t-1}, \hat{\Omega}_k, k)$$

is the one-step forecast distribution obtained by *Forward Filtering* (West and Harrison, 1997).

No. Factors	1	2	3	4	5	6	7
BIC	-34663	***-34836***	-34763	-34699	-34627	-34573	-34527

Table 4.1: BIC values assuming different number of factors.

On the other hand, the BIC is an approximation of the log of the *Bayes factor*. The Bayes factor, B_{ij} for comparing two models M_i and M_j, is the ratio of the posterior odds in favor of M_i to the prior odds in favor of M_i, this is, the posterior odds in favor of model i are

$$\frac{p(M_i|D)}{p(M_j|D)} = \frac{p(D|M_i)}{p(D|M_j)}\frac{p(M_i)}{p(M_j)},$$

the Bayes factor is defined as

$$B_{ij} = \frac{p(D|M_i)}{p(D|M_j)}$$

and

$$2\log(B_{ij}) \approx BIC_j - BIC_i.$$

Thus two models can be compared by taking the difference of their BIC values, with the model having the smaller BIC value being preferred. See Kass and Raftery (1995) for modern applications of Bayes factors in model selection, Raftery (1995), Wasserman (1997), Lee (1998) for Bayesian model selection and a summary of other model selection methods, and Lopes and West (1999) for an in-depth study on model selection addressing the uncertainty on the number of factors.

Table 4.1 displays the BIC values for each model M_j, $j = 1, \ldots, 7$ where the two-factor model seems to be more appropriate for the residuals $\boldsymbol{\epsilon}_t$. It is important to bear in mind that the BIC is only an approximation to the log of the Bayes factor where the usual error is of order $O(1)$ (Kass and Wasserman, 1995; Raftery, 1995).[7] In certain special cases, the error goes to zero asymptotically, but in general,

[7] $O(b_n)$ represents any quantity such that $\frac{O(b_n)}{b_n} \to k$ as $n \to \infty$, where k is a constant.

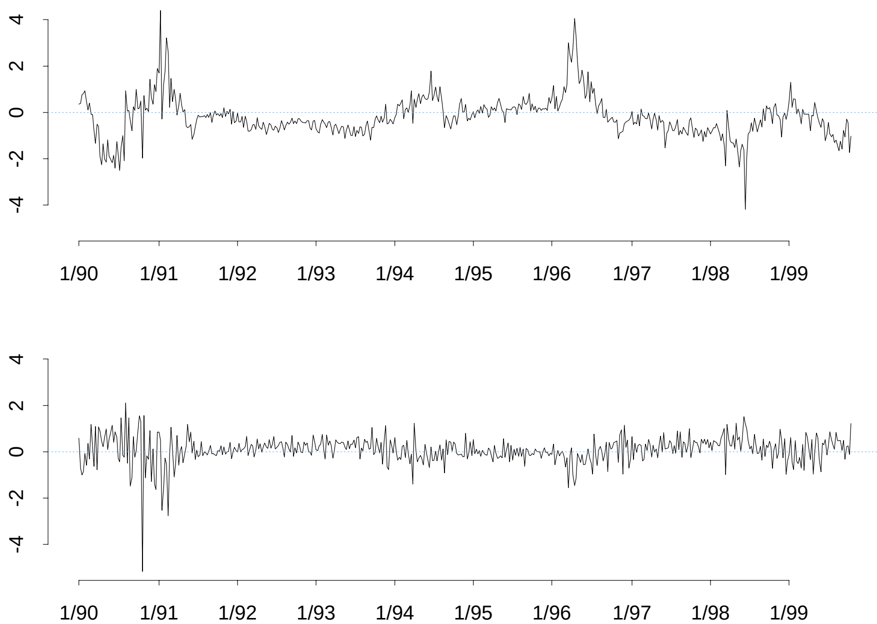

Figure 4.6: Posterior mean of factors $\mathbf{f}_t$.

it does not vanish even with an infinite amount of data. Actually, under certain situations, the BIC can be inconsistent also when the dimension of the parameter goes to infinity (Berger *et al.*, 2000).

In order to confirm results from the BIC, approximate posterior means for the k factor processes and the correspondent loadings matrix $\mathbf{X}$ were explored for $k > 2$. In all cases, the loadings for the third factor were negligible supporting the results given by the BIC.

4.4.3 Analysis Results

After a "burn-in" period of 10,000 iterations, 10,000 samples were drawn out of 100,000 simulations using the Gibbs sampler algorithm described in Section 3.2.2 complemented with the full conditionals for the factor model parameters above. The graphs in Figure 4.6 display the trajectories of approximate posterior means for the two factor processes. As can be seen in the picture, there is a clear correlation between the two factor processes especially in high volatility periods (e.g., the early part of the Gulf war). Recall that the factors are assumed uncorrelated apriori and with zero prior mean whereas the posterior mean of the factors show high correlation over time and significant departures from zero. Figure 4.7 provides histogram approximations to the marginal posterior distributions for the elements of the loadings matrix $\mathbf{X}$. The graphs indicate that the first factor can be attributed to series with short-term maturities and the second factor can be attributed to series with long-term maturities.

Figure 4.8 displays posterior inferences for the idiosyncratic variances ψ_i. The first frame shows 95% posterior intervals for the ψ_i variances observing a U shape pattern reaching the lowest value approximately at the series with thirteen months to maturity. The second frame displays margins for $100\psi_i/\sigma_i^2$ where σ_i^2 is the i-th diagonal element of $\mathbf{V}$. These ratios measure the percentage of total variation in each of the series explained by the idiosyncratic terms. In general, the idiosyncratic variances are relatively small suggesting that the common factors take care of more than the 80% of the variability across the board being even more dramatic for series with short-term maturities. In fact, informative priors have to be used to avoid having zero idiosyncratic variances.

Figure 4.9 displays a normal quantile plot of the posterior means of the standardized ordered estimated residuals $\hat{\boldsymbol{\nu}}_t$. Each plot includes vertical lines representing approximate 95% intervals showing the uncertainty in the marginal posteriors. These graphs suggest that when the factor representation is included in the model and the factors correctly estimated the series-specific factors, $\hat{\boldsymbol{\nu}}_t$ evidence reasonable normal behavior. However, there are still some outlying residuals corresponding to periods of high volatility in the futures prices.

Figure 4.7: Posterior summaries for the factor loadings matrix **X**.

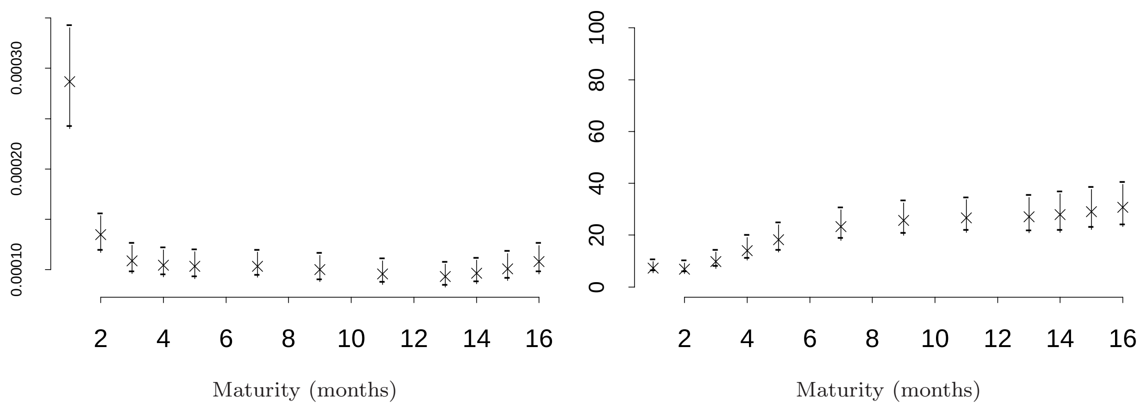

Figure 4.8: 95% posterior intervals for the specific variances ψ_i (left) and for the percentage of total variation of the specific variances $100\psi_i/\sigma_i^2$ (right).

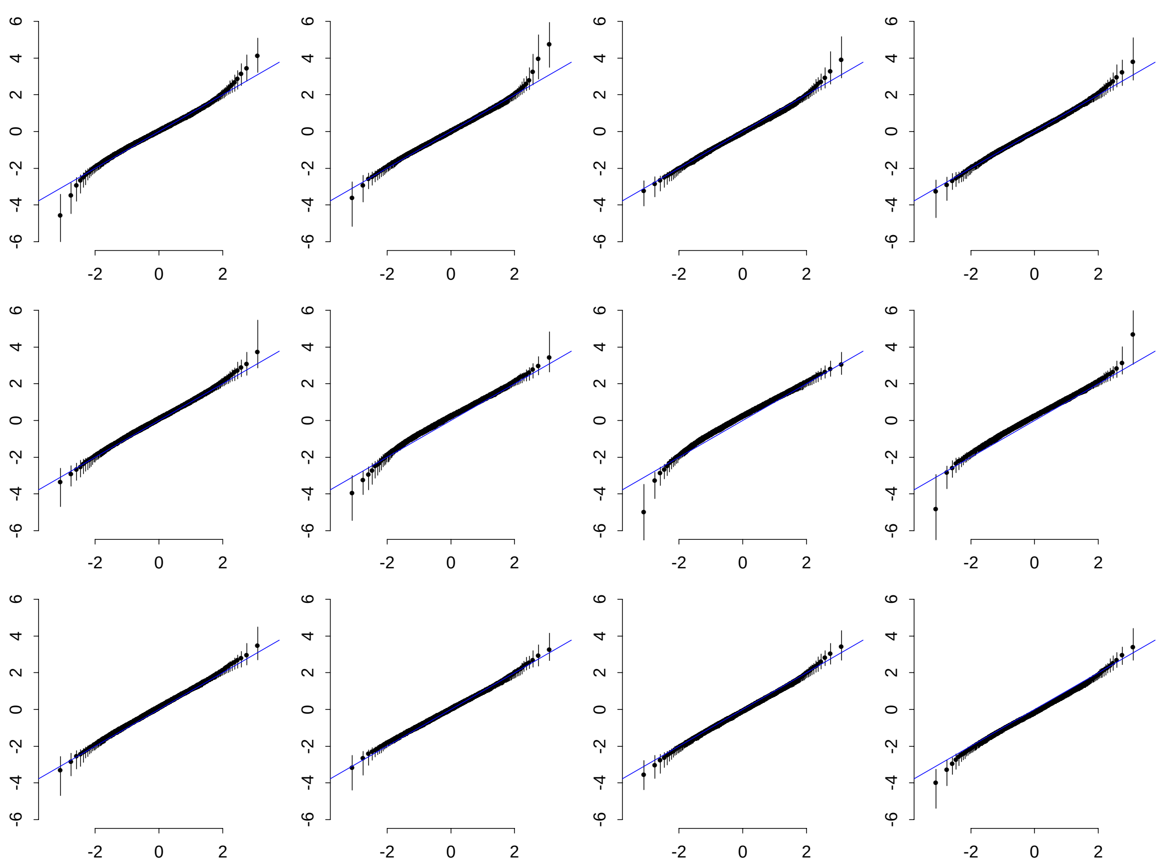

Figure 4.9: Normal quantile plot for the posterior means of the standardized ordered observed residuals $\hat{\boldsymbol{\nu}}_t$. The vertical lines represent 95% posterior intervals. Each panel represents one dimension of $\hat{\boldsymbol{\nu}}_t$.

4.5 Summary

In this chapter, different structures for the observational variance matrix of the discrete time model for futures prices are proposed and analyzed.

- First, the traditional diagonal variance assumption yielded poor estimates of the parameters. This indicates that there may be common response of the futures prices at external shocks besides those characterized in the state vector.
- Second, a non-diagonal variance matrix $\mathbf{V}$ was assumed. This structure takes into account for potential correlations in observational errors. The results showed nearly singular covariance estimates and the possibility of a correlation structure in function of the difference in maturities of the series.
- Third, a "spatial" structure for $\mathbf{V}$ was assumed under the notion that the difference in maturities are driving the behavior of the correlation between errors. Once again, the parameter estimation procedure presented numerical instabilities due to the smaller values of the eigenvalues of the estimated observational variance. Moreover, principal components decompositions suggested possible common latent factors as the main sources of variability.
- Finally, a factor representation of the observational variance was established to identify common and series-specific sources of variability. Results show two main latent factors driving the correlations between the residual terms. Moreover, the loadings matrix estimates suggested that these factors correspond to short-term and long-term maturity contracts respectively. A critical result from the factor model representation was the fact that some of the idiosyncratic variance estimates were very small suggesting that most of the variability is explained by the two main factors.

The fact that the posterior mean of the factors present some dependence over time may suggest a time varying structure for the persistence parameter. On the other hand, it may also suggest the addition of potential stochastic volatility structure present in the factors variance.

Furthermore, the heavy tailed distributions of the standardized observed residuals $\hat{\boldsymbol{\nu}}_t$ could be handled with a scale mixture of normals as performed on the evolution variance or by incorporating independent stochastic volatility models for the idiosyncrasies.

A key point to note here is that in all the assumed structures for the observational variance $\mathbf{V}$, the results evidenced nearly singular behaviors. This was even more clear with the factor model representation where informative priors had to be used to avoid having zero idiosyncratic variances. This is an extremely important issue that will be addressed in the next chapter and that eventually will have a relevant impact on the analyses results for the futures prices data. Moreover, this application opens the possibility for developing new theory and methods in the dynamic linear models arena where the observational variances could be singular or nearly singular.

Chapter 5

Singular Observational Variance Matrix

Extensive analyses of the modified discrete time model for oil futures prices were performed in the previous chapter under different structures of the observational variance $\mathbf{V}$. Results have indicated that some components of the observation errors a have linear relationship within each other implying singular covariance matrices. In this chapter, Bayesian dynamic linear models theory is extended to include the possibility for singular observational covariance matrices. First, key results on singular distributions are presented. Second, filtering algorithms and updating mechanisms are outlined for the general dynamic linear model when the observational variance is singular. Finally, a Gibbs sampling algorithm is developed to sample from the joint posterior distribution for model parameters.

5.1 Singular Densities

This section reviews two densities that play a key role in the Bayesian modeling of the futures prices database and in general when the observational variances $\mathbf{V}$ are singular. These densities are the *Singular Multivariate Normal* and the *Singular Wishart* distributions. The former involves a singular variance matrix and the latter a singular scale matrix.

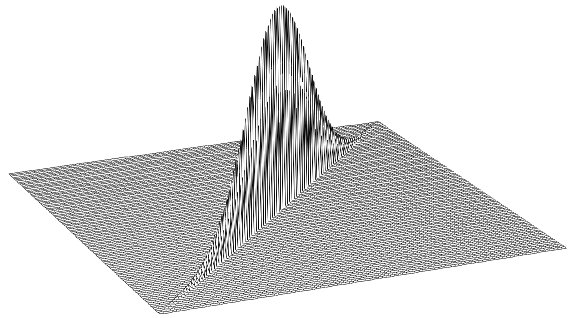

Figure 5.1: Singular bivariate normal density

The singular multivariate normal distribution has been around for the past forty years, Radhakrishna (1962), Marsaglia (1964), Khatri (1968), Mitra and Rao (1968), Styan (1970), Rao (1978), Siotani *et al.* (1985), Bhimasankaram and Sengupta (1991) have developed theory and applications around it. Studies on singular Wishart are more recent, see for example Siotani *et al.* (1985), Bhimasankaram and Sengupta (1991), Oktaba and Kieloch (1993) and Uhlig (1994). In the following subsections the singular multivariate normal and the singular Wishart distributions are introduced.

5.1.1 Singular Multivariate Normal

For starters, a general definition for a multivariate normal distribution that includes both the non-singular and the singular variance matrix cases is needed. Following studies described in Siotani *et al.* (1985), a random vector $\mathbf{x}$ in the ρ-dimensional space $\Re^{\rho}$ is said to have a r-variate normal distribution of rank[1] ρ with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ if $\mathbf{x}$ has the same distribution as $\mathbf{B}\mathbf{u} + \boldsymbol{\mu}$ where $\boldsymbol{\Sigma} = \mathbf{B}\mathbf{B}'$, $\mathbf{B} : r \times \rho$, $\text{rank}(\mathbf{B}) = \rho \leq r$ and $\mathbf{u} \sim \mathbf{N}_{\rho}(\mathbf{u}|\mathbf{0}, \mathbf{I}_{\rho})$. Note that $\mathbf{x}$ is simply defined as a linear transformation of a non-singular normal random vector that induces non-

[1]For a practical issue on the rank of a matrix see Appendix A.1.

singular distributions when $\rho = r$ and singular distributions when $\rho < r$. Observe also that the matrix $\mathbf{B}$ is not unique[2] and can be obtained in many ways such as the spectral matrix decomposition. For instance, $\mathbf{\Sigma}$ can be decomposed as $\mathbf{\Sigma} = \mathbf{HDH}'$ where $\mathbf{H}'\mathbf{H} = \mathbf{HH}' = \mathbf{I}_r$, $\mathbf{D} = \mathrm{diag}(l_1, \ldots, l_r)$, $l_1, \ldots, l_r$ eigenvalues of $\mathbf{\Sigma}$; therefore $\mathbf{\Sigma} = \mathbf{BB}'$ for $\mathbf{B} = \mathbf{H}_1\mathbf{D}_0^{1/2}$ where $\mathbf{H}_1 : r \times \rho$ in $\mathbf{H} = (\mathbf{H}_1, \mathbf{H}_2)$ and $\mathbf{D}_0 = \mathrm{diag}(l_1, \ldots, l_\rho)$ with $l_i > 0$.

Singular Multivariate Normal Density

As stated above, if $\mathbf{Bu} + \boldsymbol{\mu}$ and $\mathbf{x}$ have the same distribution, then the characteristic function of $\mathbf{x}$ can be written as $\phi_{\mathbf{X}}(\mathbf{t}) = \mathrm{E}\left(\exp\{i\mathbf{t}'\mathbf{x}\}\right)$. On the other hand, consider the following linear transformation on $\mathbf{y} = \mathbf{H}'\mathbf{x} = (\mathbf{y}_1, \mathbf{y}_2)$, where $\mathbf{H} : r \times r$ is an orthogonal matrix as defined above, $\mathbf{y}_1 = \mathbf{H}_1'\mathbf{x}$ and $\mathbf{y}_2 = \mathbf{H}_2'\mathbf{x}$. The characteristic function of $\mathbf{y}$ is now given by,

$$\phi_{\mathbf{Y}}(\mathbf{s}) = \mathrm{E}\left(\exp\{i\mathbf{s}'\mathbf{y}\}\right) = \exp\left\{i\mathbf{s}_1'\boldsymbol{\delta}_1 - \frac{1}{2}\mathbf{s}_1'\mathbf{D}_0\mathbf{s}_1 + i\mathbf{s}_2'\boldsymbol{\delta}_2\right\},$$

where $\mathbf{s}' = \mathbf{t}'\mathbf{H} = (\mathbf{s}_1, \mathbf{s}_2)$ and $\boldsymbol{\delta}' = (\boldsymbol{\delta}_1', \boldsymbol{\delta}_2') = \boldsymbol{\mu}'\mathbf{H}$. Therefore, $\mathbf{y}_1$ and $\mathbf{y}_2$ are independently distributed random variables with $\mathbf{y}_1 \sim \mathbf{N}_\rho(\mathbf{y}_1|\boldsymbol{\delta}_1, \mathbf{D}_0)$ a non-singular density and $\mathbf{y}_2 = \boldsymbol{\delta}_2$ with probability one. Since the Jacobian of the linear transformation of $\mathbf{x}$ to $\mathbf{y}$ is one, the density of $\mathbf{x}$ can be written as,

$$p(\mathbf{x}|\boldsymbol{\mu}, \mathbf{\Sigma}) = \frac{1}{(2\pi)^{\rho/2}|\mathbf{D}_0|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})'\mathbf{\Sigma}^+(\mathbf{x} - \boldsymbol{\mu})\right\}, \tag{5.1}$$

where $\mathbf{x}$ lies on the ρ-dimensional linear subspace defined by $\mathbf{y}_2 - \boldsymbol{\delta}_2 = \mathbf{H}_2'(\mathbf{x} - \boldsymbol{\mu}) = 0$ and $\mathbf{\Sigma}^+ = \mathbf{H}_1\mathbf{D}_0^{-1/2}\mathbf{H}_1'$ is the generalized inverse of $\mathbf{\Sigma}$ (see Appendix A for a general review on generalized inverse methods). Figure 5.1 shows an example of a singular bivariate normal distribution where the probability mass is concentrated in a linear set of dimension one.

[2] $\mathbf{\Sigma}$ can be rewritten as $\mathbf{\Sigma} = \bar{\mathbf{B}}\bar{\mathbf{B}}'$ where $\bar{\mathbf{B}} = \mathbf{BL}$ and $\mathbf{L}$ is any $\rho \times \rho$ orthogonal matrix.

	proper $n \geq \rho$	improper $n < \rho$
$\mathbf{\Sigma}$ positive definite $\rho = p$	*Regular*	*Pseudo*
$\mathbf{\Sigma}$ singular $\rho < p$	*Singular*	*Singular pseudo*

Table 5.1: Different Wishart distributions $\text{Wishart}_r\,(n, \mathbf{\Sigma})$ depending on rank($\mathbf{\Sigma}$)=ρ and degrees of freedom n.

5.1.2 Singular Wishart

Siotani *et al.* (1985) provides a general definition for the Wishart that includes the non-singular and the singular cases. A random matrix $\mathbf{W} : r \times r$ is said to have a Wishart distribution with scale matrix $\mathbf{\Sigma}$ and n degrees of freedom if $\mathbf{W} = \mathbf{Y}\mathbf{Y}' = \mathbf{B}\mathbf{V}\mathbf{B}'$ where $\mathbf{Y} = (\mathbf{y}_1, \ldots, \mathbf{y}_N)$, $\mathbf{y}_i = \mathbf{B}\mathbf{u}_i$, $\mathbf{u}_i \sim \mathbf{N}_\rho(\mathbf{u}_i|\mathbf{0}, \mathbf{I}_\rho)$ i.i.d. for $i = 1, \ldots, N$, $\mathbf{U} = (\mathbf{u}_1, \ldots, \mathbf{u}_N)$, $\rho \times \rho$ full rank matrix, $\mathbf{V} = \mathbf{U}\mathbf{U}'$, $\rho \times \rho$ full rank matrix, $\mathbf{B} : r \times \rho$ and $\mathbf{B} = \mathbf{H}_1\mathbf{D}_0^{1/2}$ as defined in the previous section. In this case, whenever $\rho = r$ the distribution is the standard (non-singular) Wishart distribution and when $\rho < r$ the distribution is called singular Wishart.

Singular Wishart Density

Bhimasankaram and Sengupta (1991) present the Wishart p.d.f., denoted by $\text{Wishart}_r\,(n, \mathbf{\Sigma})$, in three main cases; when (a) $\mathbf{\Sigma}$ is positive definite, $n \geq r$, (b) $\mathbf{\Sigma}$ is positive semidefinite of rank $\rho < r$ and $n \geq \rho$, and (c) $\mathbf{\Sigma}$ is positive semidefinite with rank $\rho < r$, $n < \rho$ and n integer. Table 5.1 introduces the so called traditional names of the Wishart distributions for the (a)-(c) cases. The first case has been studied extensively in the past decade; see for example Uhlig (1994) and references therein.

In general, a Wishart distribution, $\mathbf{W} \sim \text{Wishart}_r\,(n, \mathbf{\Sigma})$, is called singular if its density function is of the form,

$$p(\mathbf{W}|\mathbf{\Sigma}) = \frac{1}{2^{\rho n/2}\Gamma_\rho\left(\frac{n}{2}\right)}|\mathbf{\Sigma}^{+}|^{n/2}|\mathbf{D}_0|^{(n-r-1)/2}\exp\left\{-\frac{1}{2}\mathrm{tr}\left(\mathbf{\Sigma}^{+}\mathbf{W}\right)\right\} \tag{5.2}$$

and $(\mathbf{I}-\mathbf{\Sigma}\mathbf{\Sigma}^{+})(\mathbf{W}) = 0$, where $n \geq r$, $\mathbf{\Sigma}$ is singular with rank $\rho < r$, $\mathbf{W} = \mathbf{H}_1\mathbf{D}_0\mathbf{H}_1'$ and $\mathbf{D}_0 = \mathrm{diag}(l_1, \ldots, l_\rho)$ for $l_i > 0$ the ρ non-zero eigenvalues of $\mathbf{W}$. Here $\mathbf{\Sigma}^{+}$ denotes the generalized inverse of $\mathbf{\Sigma}$ and $\mathrm{E}(\mathbf{W}) = n\mathbf{\Sigma}$. A sketch of the derivation of this p.d.f. is given in Appendix A.5 for integer values of n.

At this point, most of the necessary tools to include singular distributions into the dynamic linear models methodology have been introduced. The next step is then to apply the results presented above to the standard DLM theory which will be done in the following section.

5.2 Forward Filtering under Singular Observational Variance

In this section, the *Forward Filtering* mechanism for the DLMs will be extended to include singular (or near-singular) observational variances $\mathbf{V}$.

Let $\mathbf{y}_t$ be a time series vector recorded at equally spaced time points $t = 1, 2, \ldots$ following a standard dynamic linear model, as presented in West and Harrison (1997). Assume now that the observational variance is singular with the usual state-space representation, namely

$$\begin{aligned}\mathbf{y}_t &= \mathbf{F}_t'\boldsymbol{\theta}_t + \boldsymbol{\nu}_t,\\ \boldsymbol{\theta}_t &= \mathbf{G}_t\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t,\end{aligned}$$

where $\boldsymbol{\theta}_t$ is the $n \times 1$ state vector, $\mathbf{G}_t$ is the $n \times n$ evolution matrix, $\mathbf{F}_t$ is a $n \times r$ matrix and $\boldsymbol{\nu}_t$ and $\boldsymbol{\omega}_t$ are error terms assumed independent, mutually uncorrelated and normally distributed, $\boldsymbol{\nu}_t \sim N(\boldsymbol{\nu}_t|\mathbf{0}, \mathbf{V}_t)$, $\boldsymbol{\omega}_t \sim \mathbf{N}(\boldsymbol{\omega}_t|\mathbf{0}, \mathbf{W}_t)$. In this case, since $\mathbf{V}_t$ is singular with rank $\rho < p$, $\boldsymbol{\nu}_t$ is then distributed as a singular normal with density given by function (5.1).

The goal here is to generalize the "updating equations" from Theorem 4.1 in West and Harrison (1997) by defining the one-step forecast for the data and the posterior distribution of the state vector for a multivariate DLM under singular distributions. In this case and following standard notation in West and Harrison (1997), the Forward Filtering algorithm of a DLM with singular observational variance will be:

(a) Posterior at $t-1$: $(\boldsymbol{\theta}_{t-1}|D_{t-1}) \sim \mathbf{N}(\boldsymbol{\theta}_{t-1}|\mathbf{m}_{t-1}, \mathbf{C}_{t-1})$.

(b) Prior at t: $(\boldsymbol{\theta}_t|D_{t-1}) \sim \mathbf{N}(\boldsymbol{\theta}_t|\mathbf{a}_t, \mathbf{R}_t)$
where $\mathbf{a}_t = \mathbf{G}_t\mathbf{m}_{t-1}$ and $\mathbf{R}_t = \mathbf{G}_t\mathbf{C}_{t-1}\mathbf{G}_t' + \mathbf{W}_t$.

(c) One-step forecast: $(\mathbf{y}_t|D_{t-1}) \sim \mathbf{N}(\mathbf{y}_t|\mathbf{f}_t, \mathbf{Q}_t)$
where $\mathbf{f}_t = \mathbf{F}_t'\mathbf{a}_t$ and $\mathbf{Q}_t = \mathbf{F}_t'\mathbf{R}_t\mathbf{F}_t + \mathbf{V}_t$.

(d) Posterior at t: $(\boldsymbol{\theta}_t|D_t) \sim N(\boldsymbol{\theta}_t|\mathbf{m}_t, \mathbf{C}_t)$
with $\mathbf{m}_t = \mathbf{a}_t + \mathbf{R}_t\mathbf{F}_t\mathbf{Q}_t^+\mathbf{e}_t$ and $\mathbf{C}_t = \mathbf{R}_t - \mathbf{R}_t\mathbf{F}_t\mathbf{Q}_t^+\mathbf{F}_t'\mathbf{R}_t$, where $\mathbf{e}_t = \mathbf{y}_t - \mathbf{f}_t$.

where $\mathbf{Q}_t^+$ denotes the generalized inverse of $\mathbf{Q}_t$. Note that the key modification here is in relation to the singularity of the distributions. Whether a normal distribution is singular or not depends on whether the corresponding variance-covariance matrix is singular or not. This is usually determined by the rank of such matrices,

$$
\begin{aligned}
&\text{rank}(\mathbf{Q}_t) \le \text{rank}(\mathbf{F}_t'\mathbf{R}_t\mathbf{F}_t) + \text{rank}(\mathbf{V}_t) = \min(n, \text{rank}(\mathbf{R}_t)) + \rho,\\
&\text{rank}(\mathbf{R}_t) \le \text{rank}(\mathbf{G}_t\mathbf{C}_{t-1}\mathbf{G}_t') + \text{rank}(\mathbf{W}_t) = \text{rank}(\mathbf{C}_{t-1}) + \text{rank}(\mathbf{W}_t),\\
&\text{rank}(\mathbf{C}_t) \le \text{rank}(\mathbf{R}_t) + \min(\text{rank}(\mathbf{R}_t), n, \text{rank}(\mathbf{Q}_t)),
\end{aligned}
$$

observe that $\mathbf{Q}_t$ may be full rank if $n \ge \text{rank}(\mathbf{R}_t) \ge r - \rho$.

The proof of (b)-(d) is by induction and mimics the proof of Theorem 4.1 in West and Harrison (1997) with a small change to manage the singular observational variance $\mathbf{V}$. It proceeds as follows,

(a) Suppose (a) holds.

(b) $(\boldsymbol{\theta}_t|D_{t-1})$ follows a normal distribution since it is a linear function of $\boldsymbol{\theta}_{t-1}$, and $\boldsymbol{\omega}_t$ which are independent normally distributed random variables. The mean $\mathbf{a}_t$ and the variance $\mathbf{R}_t$ are easily deduced.

(c) Although $(\mathbf{y}_t|\boldsymbol{\theta}_t)$ has a singular normal distribution, $\mathbf{y}_t$ is a linear transformation of a normal distribution, and therefore $(\mathbf{y}_t|D_{t-1})$ is normally distributed with mean $\mathrm{E}(\mathbf{y}_t|D_{t-1}) = \mathbf{F}_t'\mathbf{a}_t$ and variance $\mathrm{Var}(\mathbf{y}_t|D_{t-1}) = \mathbf{V}_t + \mathbf{F}_t'\mathbf{R}_t\mathbf{F}_t$.

(d) Finally, $\mathbf{y}_t$ and $\boldsymbol{\theta}_t$ are jointly normally distributed conditional on D_{t-1} with covariance $\mathrm{Cov}(\mathbf{y}_t, \boldsymbol{\theta}_t|D_{t-1}) = \mathrm{Var}(\boldsymbol{\theta}_t|D_{t-1}) = \mathbf{F}_t'\mathbf{R}_t$. Then

$$\left(\begin{array}{c} \boldsymbol{\theta}_t \\ \mathbf{y}_t \end{array} \middle| \; D_{t-1} \right) \sim N \left(\begin{array}{c} \boldsymbol{\theta}_t \\ \mathbf{y}_t \end{array} \middle| \; \left(\begin{array}{c} \mathbf{a}_t \\ \mathbf{f}_t \end{array} \right), \left(\begin{array}{cc} \mathbf{R}_t & \mathbf{R}_t\mathbf{F}_t \\ (\mathbf{R}_t\mathbf{F}_t)' & \mathbf{Q}_t \end{array} \right) \right).$$

Therefore, to obtain the posterior distribution of the state vector at time t, the conditional distribution of $\boldsymbol{\theta}_t|\mathbf{y}_t, D_{t-1}$ is needed. In this case, since $\mathbf{y}_t|D_{t-1}$ may be singular, standard normal theory cannot be applied to obtain the desired conditional distribution. However, Muirhead (1982) proofs that under these circumstances, the conditional distribution exists and $\boldsymbol{\theta}_t|D_t \sim \mathbf{N}(\boldsymbol{\theta}_t|\mathbf{m}_t, \mathbf{C}_t)$, with $\mathbf{m}_t = \mathbf{a}_t + \mathbf{R}_t\mathbf{F}_t\mathbf{Q}_t^+(\mathbf{y}_t - \mathbf{f}_t)$ and $\mathbf{C}_t = \mathbf{R}_t - \mathbf{R}_t\mathbf{F}_t\mathbf{Q}_t^+\mathbf{F}_t'\mathbf{R}_t$. See Appendix A.4 for the theorem on the conditional distributions when the joint distribution is singular as developed in Muirhead (1982).

Once the theory for DLMs under singular distributions has been established, the final step consists on finding the full conditional distributions for the model parameters in an MCMC framework.

5.3 Implementation of the Gibbs sampler

In this section, a Gibbs sampler algorithm is outlined to sample from the joint posterior distribution of the modified discrete time model described in Chapter 3 with the main difference that now $\mathbf{V}$ is assumed to be singular:

$$\begin{aligned} \mathbf{y}_t &= \mathbf{D}_t\boldsymbol{\mu}_0 + \mathbf{Z}_t\boldsymbol{\theta}_t + \boldsymbol{\epsilon}_t, \qquad \boldsymbol{\epsilon}_t \sim \mathbf{SN}(\mathbf{0}, \mathbf{V}), && (5.3)\\ \boldsymbol{\theta}_t &= \boldsymbol{\mu} + \mathbf{G}\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \qquad \boldsymbol{\omega}_t \sim \mathbf{N}(\mathbf{0}, \mathbf{W}\gamma_t^{-1}), && (5.4)\end{aligned}$$

where $\mathbf{y}_t$ is a $r \times 1$ vector of observations with $y_{it} = \log \mathrm{F}_{t,t+k_i}$, $\boldsymbol{\theta}_t' = (\chi_t, \xi_t)$ is the *state vector*, $\boldsymbol{\mu}_0' = \left(\lambda_\chi, \tilde{\mu}_\xi^*\right)$, $\mathbf{V}$ is the singular observation variance, $\mathbf{D}_t$ and $\mathbf{Z}_t$ are matrices with $(-\frac{1-\phi^{k_{it}}}{1-\phi}, k_{it})$ and $(\phi^{k_{it}}, 1)$ as ith row respectively and $i = 1, ..., r$, $\boldsymbol{\mu}' = (0, \mu_\xi)$, $\mathbf{W} = \begin{pmatrix} \sigma_\chi^2 & \sigma_{\chi\xi} \\ \sigma_{\chi\xi} & \sigma_\xi^2 \end{pmatrix}$ is the evolution variance, $\mathbf{G} = \begin{pmatrix} \phi & 0 \\ 0 & 1 \end{pmatrix}$ and $\gamma_t \sim \mathrm{Gamma}(\gamma_t | p/2, p/2)$.

The joint posterior distribution of the unknown parameters

$$\Omega = \{\phi, \boldsymbol{\mu}_0, \mu_\xi, \mathbf{V}, \mathbf{W}, \gamma_1, \ldots, \gamma_N\} \text{ and } \Theta = \{\boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_N\}$$

is given by

$$\begin{aligned} p(\Omega, \Theta, \mathbf{y}) &= p(\Omega)p(\mathbf{y}, \Theta|\Omega) \\ &= p(\Omega)\prod_{t=1}^{N} \mathbf{SN}(\mathbf{y}_t|\mathbf{D}_t\boldsymbol{\mu}_0 + \mathbf{Z}_t\boldsymbol{\theta}_t, \mathbf{V})\mathbf{N}(\boldsymbol{\theta}_t|\boldsymbol{\mu} + \mathbf{G}\boldsymbol{\theta}_{t-1}, \mathbf{W}\gamma_t^{-1}). \end{aligned} \quad (5.5)$$

Sampling the state vector $\Theta_t|\Theta_t^-$

Simulations from the full conditional posterior distribution for the state vector $\Theta_t|\Theta_t^-$ can be performed by implementing the *Forward Filtering, Backward Sampling* algorithm (FFBS) where the *Forward Filtering* part is described in Section 5.2 and for the *Backward Sampling* part standard DLMs theory applies.

Sampling the vector $\boldsymbol{\mu}_0|\boldsymbol{\mu}_0^-$

Basically, the full conditional distribution for $\boldsymbol{\mu}_0$ remains unchanged as the one presented in Section 3.2.2. The only modification is to consider the generalized inverse $\mathbf{V}^+$ instead of $\mathbf{V}^{-1}$ since the observational errors $\boldsymbol{\epsilon}_t$ are now singular normally distributed.

Sampling the persistence parameter $\phi|\phi^-$

Given the joint distribution (5.5), the posterior distribution for the autoregressive parameter is proportional to

$$q(\phi) = p(\phi) \prod_{t=1}^{N} \mathbf{SN}(\mathbf{y}_t|\mathbf{D}_t(\phi)\boldsymbol{\mu}_0 + \mathbf{Z}_t(\phi)\boldsymbol{\theta}_t, \mathbf{V})\mathbf{N}(\boldsymbol{\theta}_t|\boldsymbol{\mu} + \mathbf{G}(\phi)\boldsymbol{\theta}_{t-1}, \mathbf{W}\gamma_t^{-1})$$

which requires similar iteratively resampling mechanisms, as the Metropolis step described in Section 3.2.2, to approximate the full conditional posterior distribution.

Sampling the singular observational variance matrix $\boldsymbol{V}|\boldsymbol{V}^-$

Assuming a singular Wishart prior distribution for the generalized inverse of $\mathbf{V}$, $\text{SWishart}_r\left(\mathbf{V}^+|\boldsymbol{\nu}, \mathbf{S}\right)$ as in (5.2), the full conditional distribution is given by

$$\text{SWishart}_r\left(\mathbf{V}^+ \middle| \boldsymbol{\nu} + N, \sum_{t=1}^{N} \boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t' + \mathbf{S}\right).$$

To simulate from this singular distribution, a simple procedure based on multivariate normal simulations could be applied in the case of integer degrees of freedom. In other words, in order to obtain a sample from $\text{SWishart}_r\left(n, \boldsymbol{\Sigma}\right)$ where $n \geq r$ and $\text{rank}(\boldsymbol{\Sigma}) = \rho < r$,

(a) generate $\mathbf{y}_1, \ldots, \mathbf{y}_n$, n independent samples from a ρ-dimensional multivariate $\mathbf{N}(\mathbf{0}, \mathbf{I})$ distribution;

(b) compute $\mathbf{Z}_i = \mathbf{B}\mathbf{y}_i$ where $\mathbf{\Sigma} = \mathbf{B}\mathbf{B}'$ and $\mathbf{B}$ is a $r \times \rho$ matrix. For example $\mathbf{B} = \mathbf{H}_1\mathbf{D}_0^{1/2}$ with $\mathbf{H}_1$ and $\mathbf{D}_0$ as detailed in Section 5.1.1; and

(c) finally, $\mathbf{W} = \sum_{i=1}^{n} \mathbf{Z}_i\mathbf{Z}_i'$ will be a sample from $\text{SWishart}_r(n, \mathbf{\Sigma})$.

In the previous sections, new developments in the dynamic linear model theory were presented when singular observational variances were assumed. Moreover, the sampling mechanisms for this new class of models were outlined in an MCMC framework. The theory presented here represent the final "piece of the puzzle" to implement the modified discrete time model for the futures prices data set.

5.4 Analysis Results

In this section, results from the modified discrete time model for the oil futures data set are presented assuming that the observational variance is singular. The complete Gibbs sampler algorithm was implemented by including the singular DLM full conditional distributions, presented in previous sections. The remaining MCMC components are developed earlier in Section 3.2.2.

Discussion of model fitting and analyses below are based on specific rank ρ for the observation variance matrix. Even though inference on ρ is not formally address here, some discussion on empirical experiences with different ranks are illustrated.

The sampler was run separately assuming different values of the rank of $\mathbf{V}$, $\rho = 1, \ldots, r$. For each of the ranks assumed, a total of 10,000 samples from the posterior distributions for all model parameters were obtained by subsampling from 500,000 iterations after a burn-in period of 5,000. The main points to highlight from analyses under different values of ρ are:

- For these models MCMC algorithms tend to experience convergence difficulties in models with large ρ (i.e. close to full rank matrices). On the other hand, in

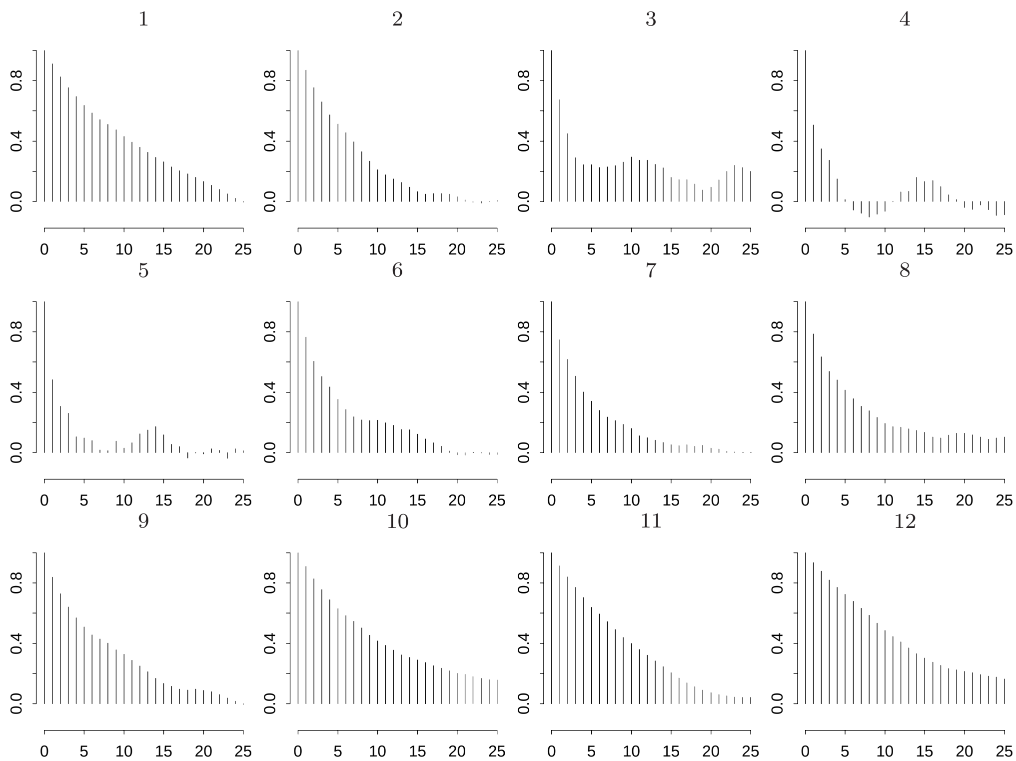

Figure 5.2: Autocorrelation functions for the persistence parameter ϕ assuming $\rho = 1, \ldots, r$ where $r = 12$.

a model in which the rank appears to be appropriate for the data under study, empirical evidence shows that MCMC algorithms converge rapidly. Figure 5.2 displays autocorrelation functions of the sampled values of the persistence parameter ϕ which was the parameter with slowest convergence rates. The plots show that ϕ converges faster with ranks $\rho = 4$ and 5.[3]

- Poor model fitting and forecasts occur when a large or very small value of ρ is assumed.

- Following earlier studies with this kind of data, a fixed value for rank($\mathbf{V}$) = 4 was chosen for the analyses presented here; although results from posterior distributions of model parameters are very similar for rank $\rho = 3, 4, 5$ and 6.

[3]Previous work on oil futures prices, under similar ranges of maturities, used an observational matrix of rank four (Schwartz, 1997; Schwartz and Smith, 2000).

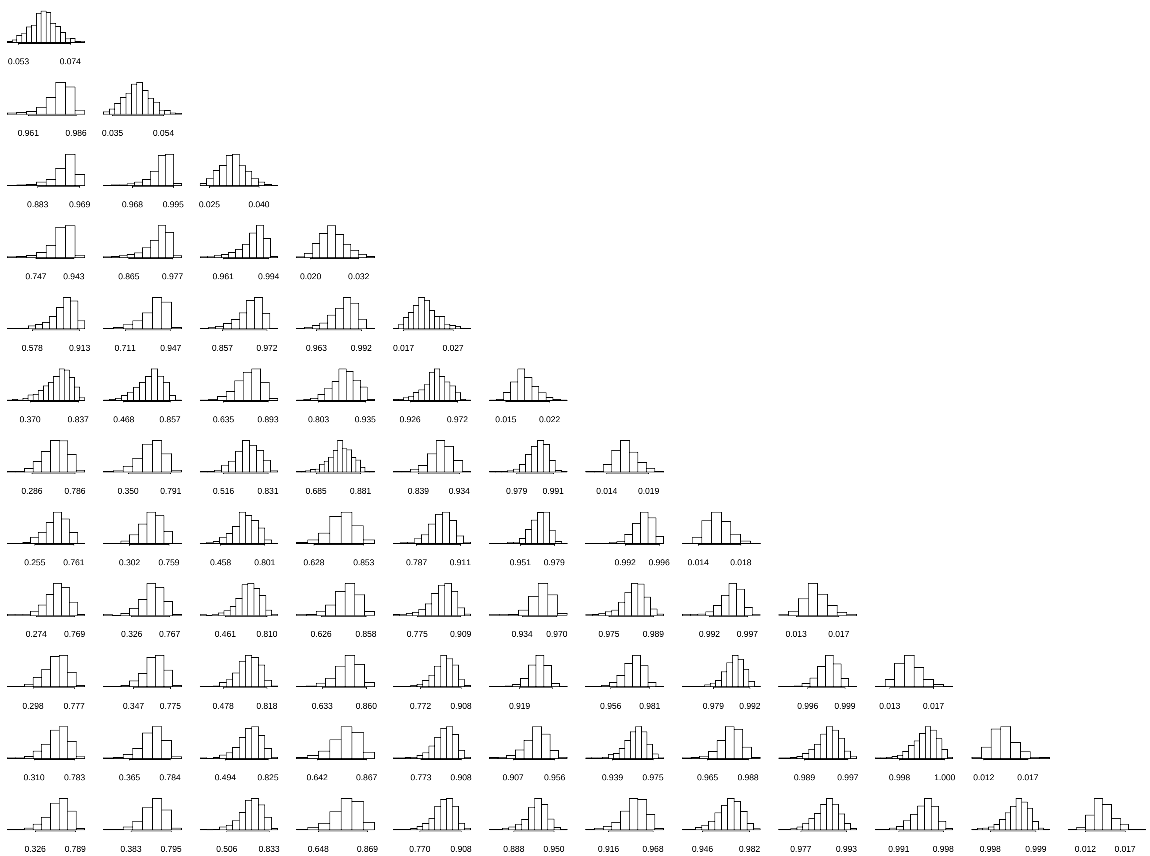

Figure 5.3: Posterior summaries of observation variance matrix $\mathbf{V}$. The elements in the diagonal are standard deviations, the elements in the off-diagonal are correlations.

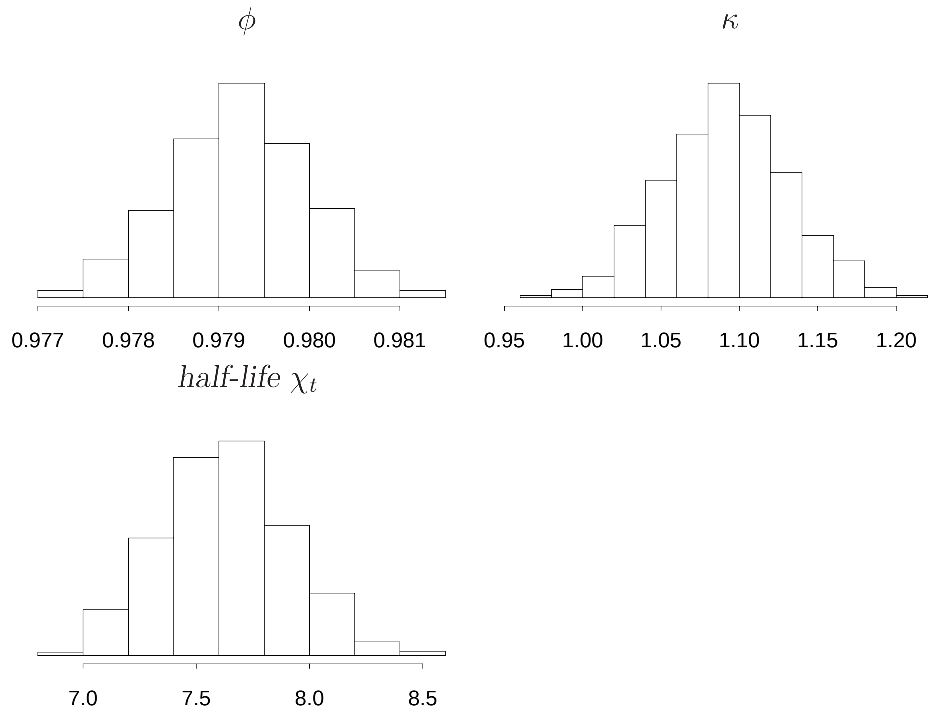

Figure 5.4: Posterior summaries for the persistence parameter ϕ (top left) κ in annual terms (top right) and the *half-life* in months of the short-term deviations (bottom).

Figure 5.3 to 5.6 inclusive provide histogram approximations to marginal posteriors for the parameters $\{\phi, \boldsymbol{\mu}_0, \mu_\xi, \mathbf{V}, \mathbf{W}\}$ and some functions of them. Figure 5.3 displays posterior summaries of the estimated observation variance $\mathbf{V}$. The diagonal elements represent the standard deviations and the off-diagonal histograms correspond to correlations between the twelve observational errors. Note that, as expected, there are many correlations close to one and the estimated posterior standard deviations decrease as the corresponding maturity increases. In other words, the errors for futures prices with short-term maturities have higher volatility than the ones with long-term maturities as one might expect.

Figure 5.4 provides posterior summaries for parameters that represent the rate at which the short-term deviations, χ_t, are expected to disappear. The top frames display summaries of the persistence parameter ϕ and its continuous version in annual

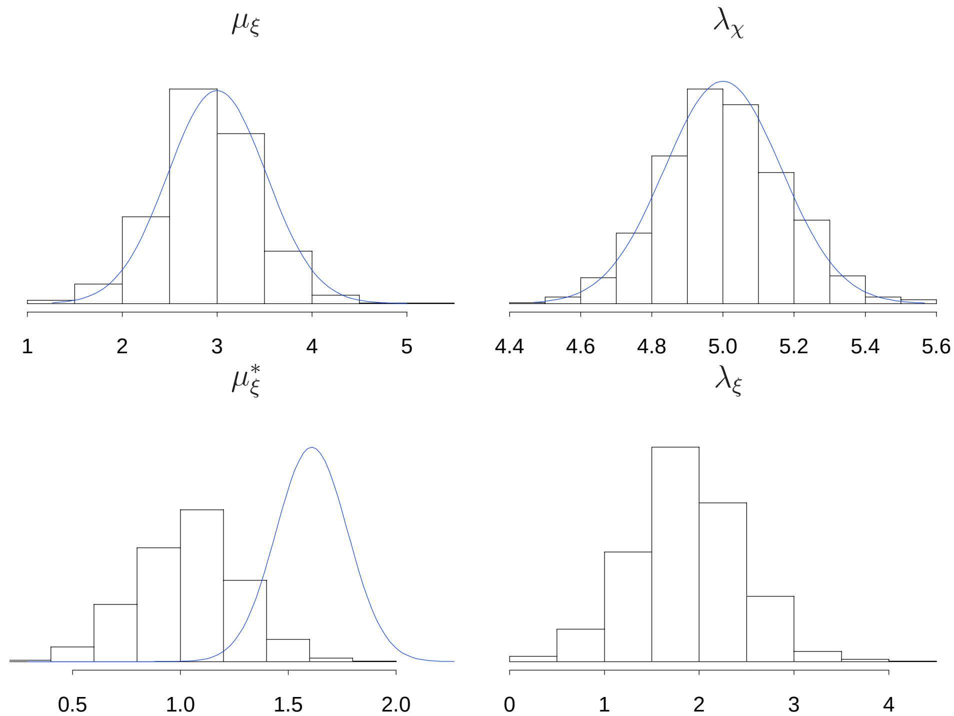

Figure 5.5: Posterior summaries of the equilibrium drift rate μ_ξ, equilibrium risk-neutral drift rate μ_ξ^* and risk premiums λ_χ and $\lambda_\xi = \mu_\xi - \mu_\xi^*$. The prior distribution is indicated by the solid line.

terms, $\kappa = -52\log(\phi)$. The bottom frame displays summaries of the *half-life* of the deviations defined as the time h in which the deviation χ_{t+h} is expected to be halve the value of χ_t. That is, the value of h where $\text{E}(\chi_{t+h}) = \frac{1}{2}\text{E}(\chi_t)$ which in months is equivalent to $h = \frac{12}{52}\frac{\log(0.5)}{\log(\phi)}$. These graphs indicate that the persistence on the short-term deviations are expected to halve in about seven and a half months and it will take at least 300 weeks[4] ($\approx$ 6 years) to reach the long-term trend (see also Figure 5.10).

Figure 5.5 shows posterior summaries for the equilibrium drift rate μ_ξ, equilibrium risk-neutral drift rate μ_ξ^* and risk premiums λ_χ and $\lambda_\xi = \mu_\xi - \mu_\xi^*$, where the solid line indicates the prior distribution. Recall that these parameters needed strong priors, since there is not enough information in the data to correctly estimate them,

[4]This is the number of weeks k for which ϕ^k is "small", see Section 3.1.4.

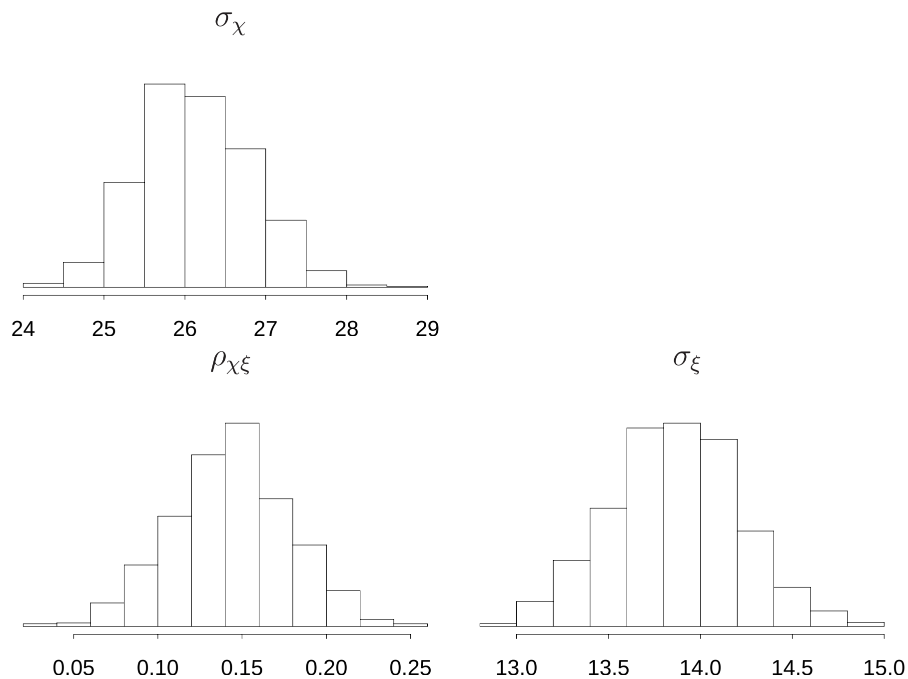

Figure 5.6: Posterior summaries of the system variance matrix $\mathbf{W}$. The standard deviations σ_χ and σ_ξ are in annual terms.

especially for λ_χ and μ_ξ as explained in Sections 3.1.4 and 3.2.1. The first two graphs suggest that the posterior distributions of λ_χ and μ_ξ are slightly different from their corresponding priors, as expected. The bottom graph, shows that the posterior estimates of the equilibrium risk-neutral drift rate μ_ξ^* are smaller than the one assumed apriori. Recall that the long-term slope of the log of the futures prices in model (3.13) is a function of μ_ξ^* and even though the data have information about it, the range of maturities is not enough to make precise inference about the long-term slope without prior information.

Figure 5.6 shows a histogram of samples from the posterior distribution of the evolution variance $\mathbf{W}$ where the standard deviations σ_χ and σ_ξ are in annual terms. Figure 5.7 shows the trajectories of approximate posterior means for the log of the heavy-tailed weights, $\log(\gamma_t^{-1})$. These weight parameters were introduced to allow for

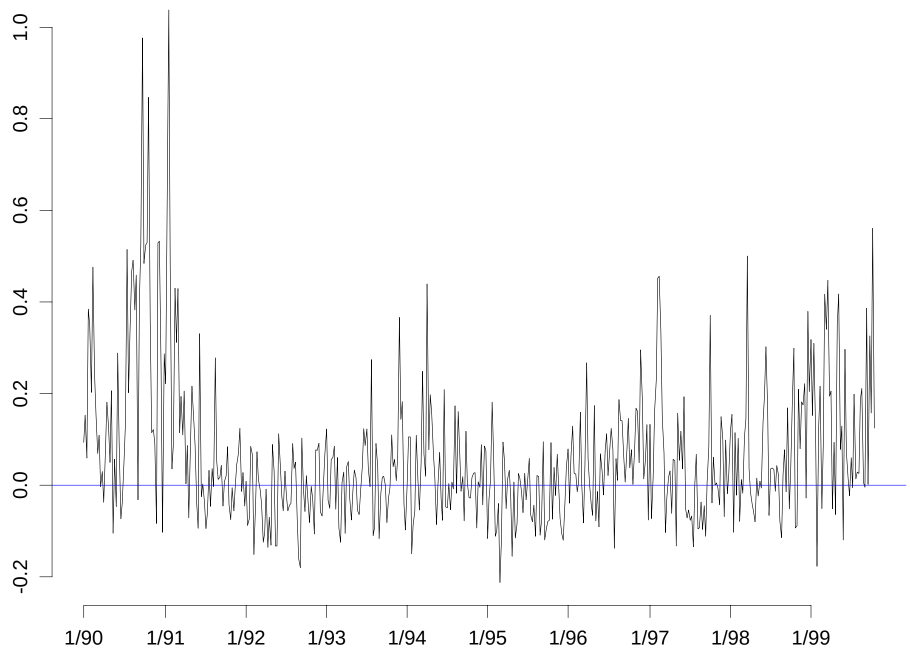

Figure 5.7: Posterior mean of the log of the heavy-tailed weights, $\log(\gamma_t^{-1})$.

the system errors to have heavy-tailed distributions in the evolution equation (5.4):

$$\boldsymbol{\theta}_t = \boldsymbol{\mu} + \mathbf{G}\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t, \qquad \boldsymbol{\omega}_t \sim \mathbf{N}(\mathbf{0}, \mathbf{W}\gamma_t^{-1}),$$

$t = 1, \ldots, N$, where γ_t is assumed apriori Gamma$(\gamma_t|p/2, p/2)$, $p = 5$.[5] The graph clearly illustrates that in periods of high crude oil price volatility, the innovations for the short-term deviations and equilibrium price level observed heavy-tailed behavior. For example, the most notorious volatile period is during the Gulf War in the summer and fall of 1990; another highly volatile period is during the end of 1993 and beginning of 1994 when high production levels by OPEC and other countries were more than sufficient to satisfy global demand, leading to an increase in the worldwide crude oil inventory levels (Schwartz and Smith, 2000), and finally there is a general increased in volatility starting in 1997. Moreover, the graph suggests serial autocorrelations which may point to the addition of stochastic volatility structures to the system

[5]Some other values $p = 3, 4, 6, 7, 10$ for degrees of freedom were used and main results were basically the same.

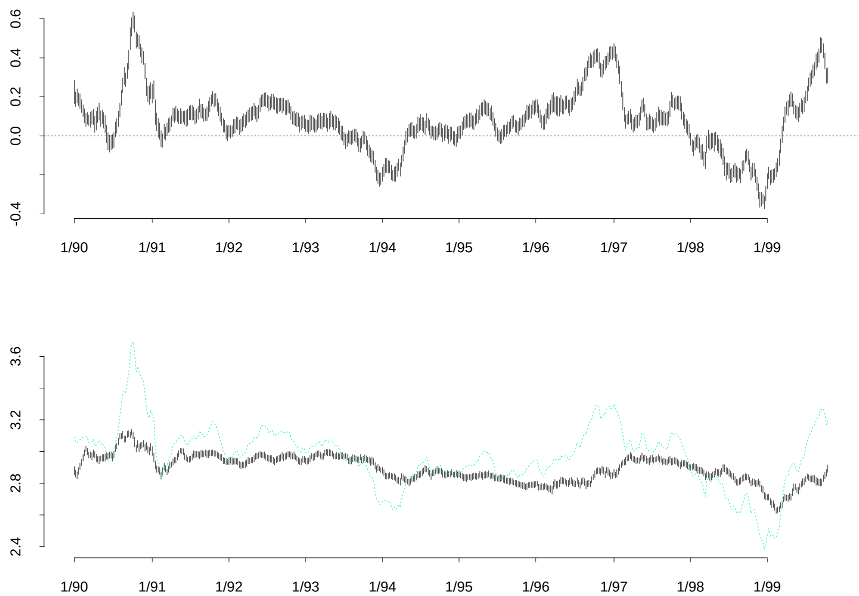

Figure 5.8: 95% posterior intervals for χ_t (upper frame) and ξ_t (lower frame) indicated by vertical segments. The dotted line in the lower frame represents the posterior mean of the log of the spot price.

variance matrix.

Figure 5.8 shows 95% posterior intervals for the two main latent processes present in crude oil prices; namely the short-term deviations, χ_t (upper frame) and the equilibrium price level, ξ_t (lower frame). The dotted line in the lower frame represents the posterior mean of the spot price on the log scale. Observe that, as expected, the χ_t process evidences substantial short-term volatility that is then transmitted to the spot prices. On the other hand, equilibrium price levels ξ_t evidence smoother patterns due to their long-term horizon.

Figure 5.9 shows normal quantile plots of the posterior means of the ordered standardized observed residuals $\hat{\boldsymbol{\epsilon}}_t$ of dimension $\rho = 4$.[6] As can be seen, they evidence heavier than normal tails. Actually, outlying observations 32, 43, 39, 50 and 55 correspond to dates lying between 08/06/90 and 01/14/91, which is the volatile period

[6]See Appendix A.3 for details on standardization of a random vector with singular variance.

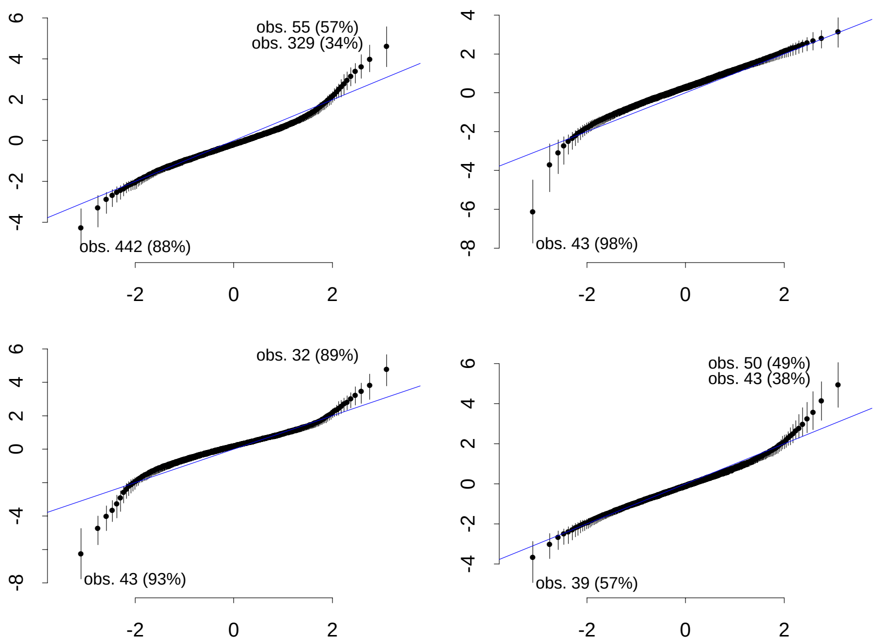

Figure 5.9: Normal quantile plot of the posterior means of the ordered observation standardized residuals $\hat{\boldsymbol{\epsilon}}_t$ of dimension $\rho = 4$. The vertical lines represent 95% posterior intervals. Each panel represents one dimension of the standardized residuals $\hat{\boldsymbol{\epsilon}}_t$. The labeled points indicate the percentages that particular observations contributed in the computation of the posterior mean residual.

of the Gulf War. This suggests the addition of either heavy-tailed distributions for $\boldsymbol{\epsilon}_t$ or potential stochastic volatility structure in the variance of such errors.

Figure 5.10 presents actual futures prices data and 95% posterior intervals of the estimated term structure in years for three different dates. The first column shows plots with the intervals of short-term maturities of the futures prices in the data and the second column plots the same intervals but in a longer time frame. The term structure for observation t is the pattern of futures prices under various maturities at day t, given by

$$\operatorname{tst}(k) = \left(-\frac{1-\phi^k}{1-\phi}, k \right) \begin{pmatrix} \lambda_\chi \\ \tilde{\mu}_\xi^* \end{pmatrix} + (\phi^k, 1) \begin{pmatrix} \chi_t \\ \xi_t \end{pmatrix}.$$

Note that the graphs show that the short span of maturities in the data makes it difficult to estimate the long-term slope, $\tilde{\mu}_\xi^*$, with only the data as mentioned in

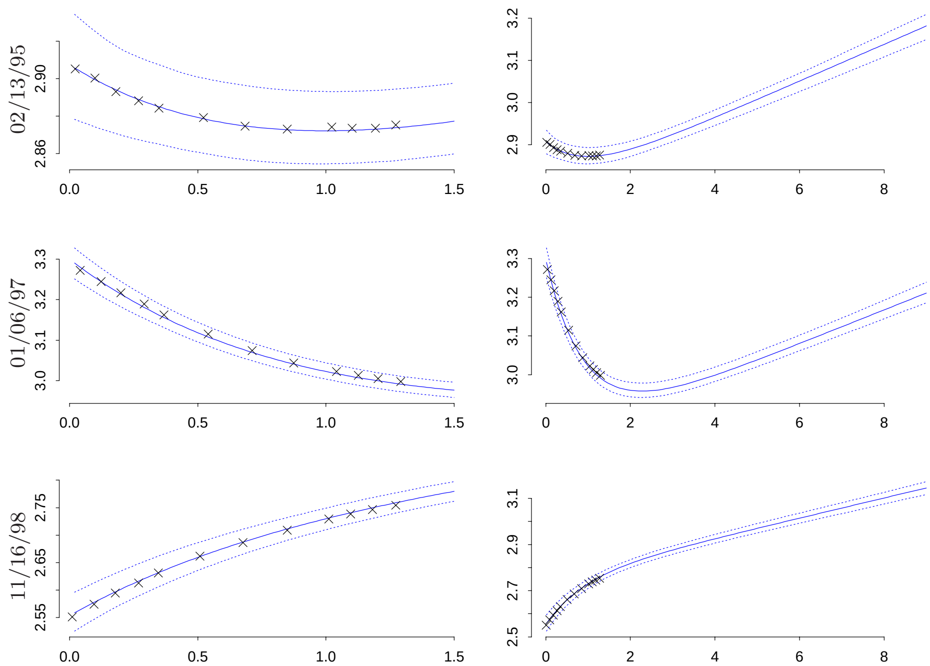

Figure 5.10: Futures prices data and 95% posterior intervals of the term structure (years) of oil futures prices for 02/13/95, 01/06/97 and 11/16/98 under short-term (first column) and long-term (second column) time span.

Section 3.1.4. Another point to observe here is the narrow posterior intervals for the term structure, which reflects quite precise posterior inferences on ϕ, χ_t, ξ_t and μ_ξ^*, combined with a fairly informative prior for λ_χ.

Non-informative priors for λ_χ, μ_ξ and μ_ξ^*

Similar analyses were performed assuming less informative priors for the parameters λ_χ, μ_ξ and μ_ξ^*. Results show that posterior inferences for λ_χ and μ_ξ evidence high levels of uncertainty as expected. The approximate posterior median and end-points of a 95% interval for μ_ξ^* (annual %) are 0.1686-0.3769-0.5784, this is, that the estimated long-term slope is flatter when using only the information in the futures prices with short-term maturities.

For the rest of the parameters, the posterior inferences remained basically unchanged, except for the persistence parameter ϕ that observed slight modifications.

The approximate posterior median and end-points of a 95% interval for ϕ are 0.9705-0.9721-0.9737. The distribution observed lower values than the one with informative priors for λ_χ, μ_ξ and μ_ξ^*. This implies a slightly shorter period for the short-term deviations to revert to the mean, which is mainly caused by the change in the long-term slope. One important point to note here is that under these non-informative priors, parameters ϕ and μ_ξ^* show slower convergence rates. This supports the idea of the lack of data-based information for λ_χ, μ_ξ and μ_ξ^*.

5.5 Forecasting

Finally, to validate the model, *out-of-sample* forecasts will be generated for the next 23 values (about six months) of the series. That is, random samples from the posterior predictive distributions and one-step-ahead forecast distributions are drawn for the period 10/25/99 through 03/27/00 and are compared with the actual observed values.

5.5.1 Posterior Predictive Distributions

A basic technique for model validation is to draw simulated values from the posterior predictive distribution of future outcomes $p(\tilde{\mathbf{y}}|\mathbf{y})$ where $\mathbf{y} = (\mathbf{y}_1, \ldots, \mathbf{y}_N)$ and $\tilde{\mathbf{y}} = (\mathbf{y}_{N+1}, \ldots)$. These samples are then compared to the actual observed data points. See for example Berger (1985) and Bernardo and Smith (1994) for details on posterior predictive distributions and Gelman *et al.* (1995) for similar examples.

In this case, given the j-th simulated value from the posterior distribution of the parameters: $\phi^{(j)}, \boldsymbol{\mu}_0^{(j)}, \mu_\xi^{(j)}, \mathbf{V}^{(j)}, \mathbf{W}^{(j)}, \boldsymbol{\theta}_N^{(j)}$ for $j = 1, \ldots, 10000$, the next 23 state vectors $\boldsymbol{\theta}_{N+1}^{(j)}, \ldots, \boldsymbol{\theta}_{N+23}^{(j)}$, hence latent processes, were generated from the system equation (5.4).[7] In addition, the next 23 simulated values $\mathbf{y}_{N+1}^{(j)}, \ldots, \mathbf{y}_{N+23}^{(j)}$ were

[7]Using variance weights $\gamma_{N+1}, \ldots, \gamma_{N+23}$ generated from the prior Gamma$(p/2, p/2)$.

generated from the observation equation (5.3). Figure 5.11 shows seven randomly selected samples (path) from the predictive distribution of the next 23 values of the series. The first part of each plot represents the last 94 points from the observed series, futures prices from 01/01/98 to 10/18/99. Simulated values, from 10/25/99 to 03/27/00, and observed values are separated by a dashed line. The first panel shows the actual futures prices (log scale) from 01/01/98 to 03/27/00.

The plots give an idea of the potential path of the series in the next six months, according to the model. As a matter of fact, the actual data and the simulations are undistinguishable (first frame versus the rest) exhibiting the same general patterns and structure as the twelve time series at hand. Figure 5.12 shows marginal histograms of samples from the posterior predictive distribution of the first out-of-sample data point one week ahead (i.e. futures prices, in log scale, for 10/25/99 for twelve different maturities). The observed value is always in the middle of the histogram showing that the model is generally adequate. Figure 5.13 displays 95% posterior intervals of the predictive distribution of the futures prices in the log scale. The first two or three new observations are in the middle of the predictive interval, however, the last eleven observations, from 01/17/00 to 03/27/00, are systematically biased towards the upper part of the distribution. This is a common behavior in most financial data series were accurate forecast for such wide horizons are difficult to obtain.

To complement this analysis and to explore the potential advantages of the model in short-term forecasting, one-step-ahead forecasts for the same 23 values will be generated sequentially in the next section.

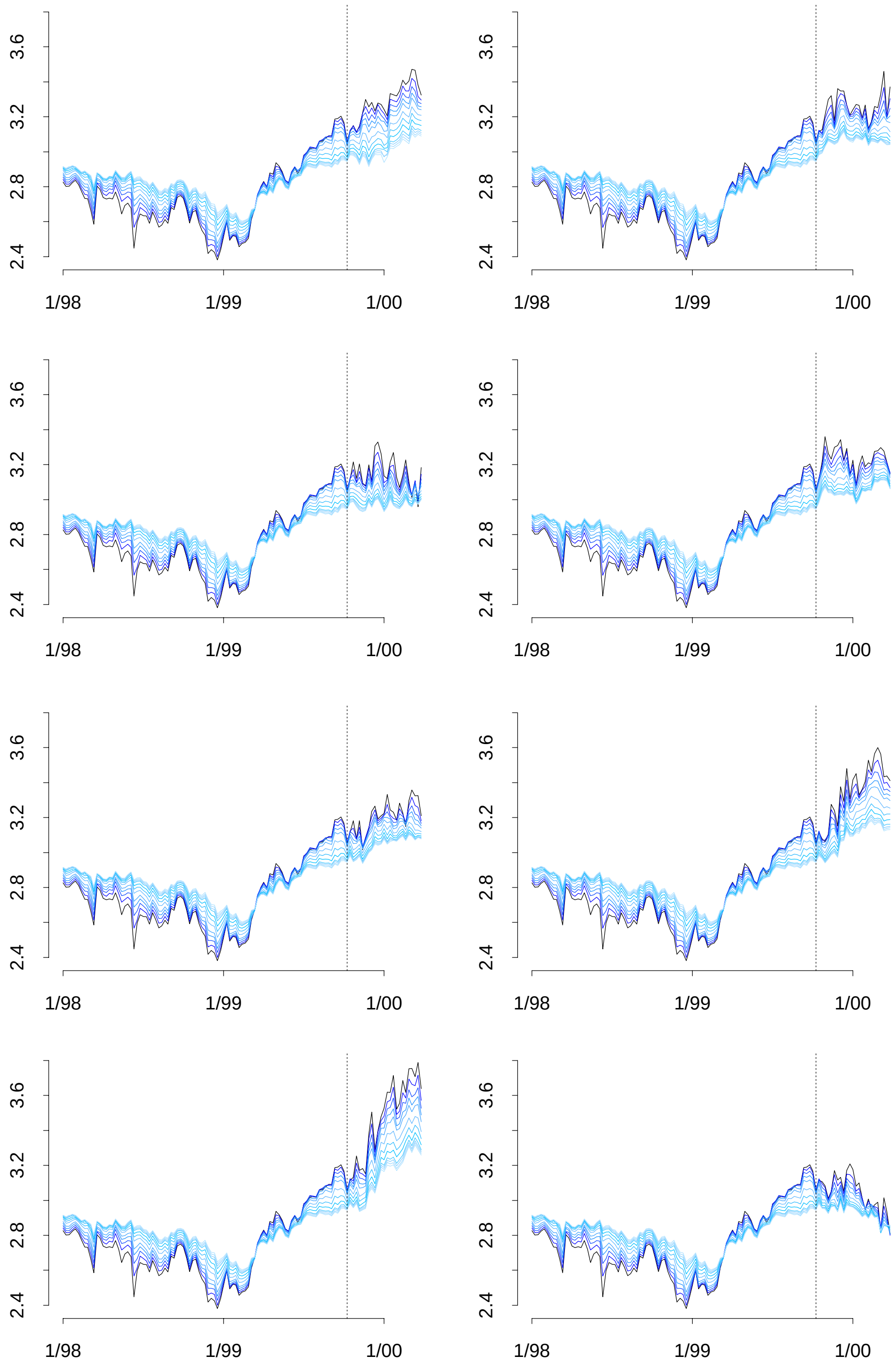

Figure 5.11: Seven randomly selected samples from the predictive distribution of the next 23 values of the series. The first panel shows the actual data (log scale) from 01/01/98 to 03/27/00. Observed values, from 01/01/98 to 10/18/99, and simulated values, from 10/25/99 to 03/27/99, are separated by a dashed line.

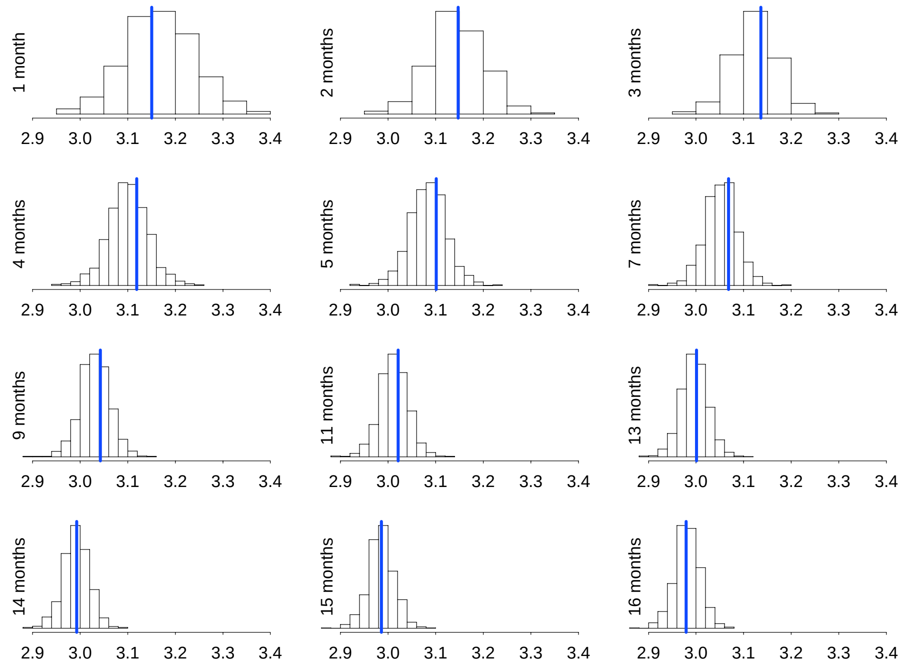

Figure 5.12: Samples from the posterior predictive distribution for 10/25/99 for the twelve futures prices series in the log scale. The solid line indicates the observed price, each panel represents one of the twelve price series.

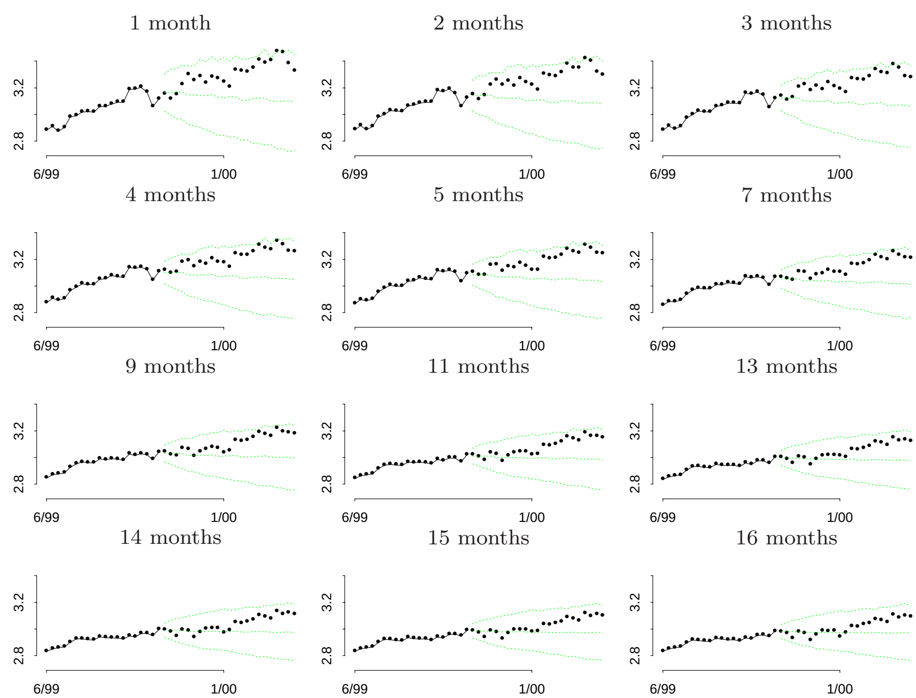

Figure 5.13: 95% posterior intervals for the predictive distribution of the futures prices in the log scale. Each panel represents one futures prices series with certain maturity.

5.5.2 One-Step-Ahead Forecast

The exact one-step-ahead forecast distribution $p(\mathbf{y}_{l+1}|D_l)$, where $l = N, \ldots, N+22$ and $D_l = (\mathbf{y}_1, \ldots, \mathbf{y}_l)$ for the modified discrete time model with singular observational variances can be obtained using the forward filtering algorithm for DLMs (West and Harrison, 1997),

$$p(\mathbf{y}_{l+1}|D_l) \sim \mathbf{N}(\mathbf{y}_{l+1}|\mathbf{f}_{l+1}, \mathbf{Q}_{l+1})$$

where

$$\begin{aligned}
\mathbf{f}_{l+1} &= \mathbf{Z}_{l+1}\mathbf{a}_{l+1} + \mathbf{D}_{l+1}\boldsymbol{\mu}_0, \\
\mathbf{Q}_{l+1} &= \mathbf{Z}_{l+1}\mathbf{R}_t\mathbf{Z}'_{l+1} + \mathbf{V}_{l+1}, \\
\mathbf{a}_{l+1} &= \mathbf{G}\mathbf{m}_l + \mu, \\
\mathbf{R}_{l+1} &= \mathbf{G}\mathbf{C}_l\mathbf{G}' + \mathbf{W}_{l+1}/\gamma_{l+1}, \\
\mathbf{m}_{l+1} &= \mathbf{a}_{l+1} + \mathbf{R}_{l+1}\mathbf{Z}'_{l+1}\mathbf{Q}^{+}_{l+1}\mathbf{e}_t, \\
\mathbf{C}_{l+1} &= \mathbf{R}_{l+1} - \mathbf{R}_{l+1}\mathbf{Z}'_{l+1}\mathbf{Q}^{+}_{l+1}\mathbf{Z}_{l+1}\mathbf{R}_{l+1}, \\
\mathbf{e}_{l+1} &= \mathbf{y}_{l+1} - \mathbf{f}_{l+1}.
\end{aligned}$$

In this case, parameters $\{\phi, \boldsymbol{\mu}_0, \mu_\xi, \mathbf{V}, \mathbf{W}\}$ were fixed at their posterior means and $\gamma_{l+1} = 1$ for each $l = N, \ldots, N+22$. Note that the one-step forecast distribution, $p(\mathbf{y}_{l+1}|D_l)$, considers information of the data until $t = N$ on inference of the parameters $\{\phi, \boldsymbol{\mu}_0, \mu_\xi, \mathbf{V}, \mathbf{W}\}$ and considers data until $t = l$ only in computing the forecast error $\mathbf{e}_{l+1} = \mathbf{y}_{l+1} - \mathbf{f}_{l+1}$. Figure 5.14 shows 95% posterior intervals for the term structure of oil futures prices for the next 23 observations, the solid line is the posterior mean, m_{l+1}, $l = N, \ldots, N+22$ and the points are the actual observed data. Basically, the posterior mean delineates future outcomes well, except in some days of abnormal increments in price, for example in 1/17/00 and 3/6/00. In addition, each graph displays at maturity equal to zero, the posterior mean of the log of the spot price defined as the sum of the equilibrium price level and the short-term deviations (the sum of the two estimated latent processes). Observe that when the futures contract is close to maturity (for example 12/20/99, 1/17/00, 2/21/00, 3/20/00), the spot price and the futures price at short-term maturity are extremely close.

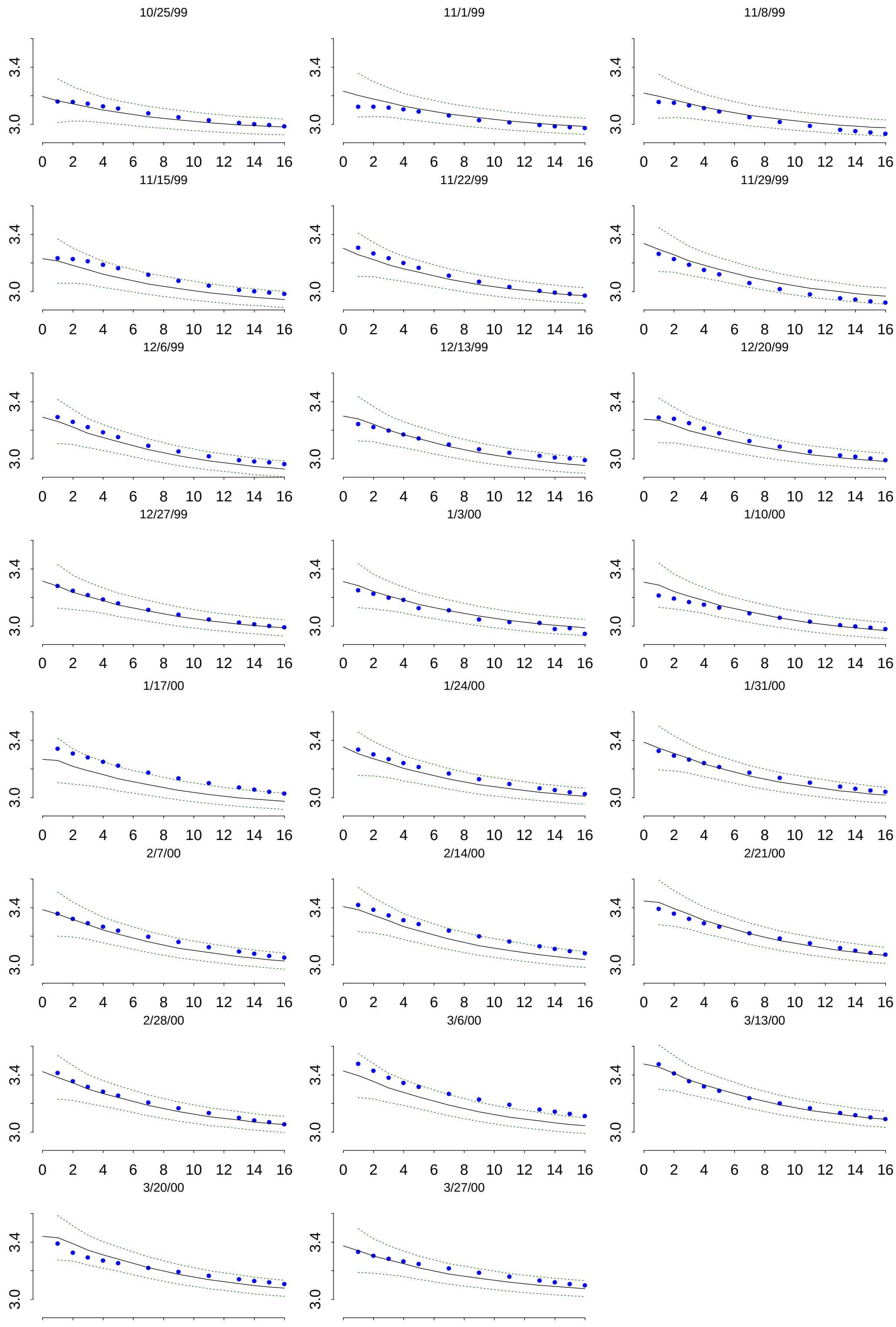

Figure 5.14: 95% posterior intervals for the term structure of oil futures prices for 10/25/99 to 03/27/00. The solid line is the posterior mean, the points are the actual data.

5.6 Summary

In previous chapters, analyses suggested the singular structure of the observational variance in the modified discrete time model. This challenging assumption was described, developed and analyzed in this chapter using the futures prices data set. In general, although singular normal theory has been developed for many years, its use in practical problems has always been avoided.

In the first part of the chapter, key statistical theory is reviewed and developed to complete the sampling scheme for the parameters described in Section 3.2.2 assuming singular observational covariance matrices. The *singular normal* and *singular Wishart* distributions were presented as simple linear transformations of a non-singular normally distributed random vector. Furthermore, the standard dynamic linear model mechanisms were extended to include singular distributions. In other words, the *forward filtering* algorithm under singular observational variance is defined completing the pieces for a sampling algorithm for $\mathbf{V}$.

In the second part of the chapter, the MCMC sampling scheme is implemented and results are analyzed for all model parameters assuming a fixed value of $\text{rank}(\mathbf{V}) = 4$. Finally, out-of-sample forecasts are generated and compared to the actual observed values for model validation and general assessment. Some important findings are summarized as follows,

- Prices reach the long-term trend at approximately six years.
- Informative priors are necessary to estimate the risk premium λ_χ, equilibrium drift rate μ_ξ and equilibrium risk-neutral drift rate μ_ξ^* since there is not enough information in the data.
- Volatile periods commonly observed in commodity pricing are difficult to capture without the "heavier than normal tails" assumptions. Outlying residuals

are observed especially in the Gulf War period. This could be resolved by extending the theory to incorporate either time-varying singular observational variances in the model or singular T-student distributions.

- Results under singular distribution assumptions, and with a factor decomposition of the observational variance, yield similar conclusions for most model parameters. Especially when some of the idiosyncratic variances are extremely close to zero.

- Simulations from posterior predictive distributions generate sample paths for "future futures" prices that exhibit the same general patterns and structure evident in the twelve time series at hand, especially it seems that there is some potential for this model to improve short-term forecast. Apart from possible stochastic volatility extension, the model is generally adequate.

Chapter 6

Non-Gaussian Hierarchical Models in Institutional Profiling

In previous chapters, a dynamic linear model for crude oil futures prices was developed where latent processes played a central role in modeling these challenging multiple series over time. In the next two chapters, a latent time series structure for non-Gaussian processes is assumed, analyzed and discussed under a hierarchical framework where some of the parameters are related over time. Hierarchical logistic regression models are then developed to attempt to profile providers in health-care delivery systems.

Profiling is the process of comparing quality of care, use of services or cost of providers (care-providing networks, hospitals, physicians, schools) against normative or community standards. The use of statistics to assess institutional performance started before 1840 when the Statistical Society of London set up a Committee on *Hospital Statistics* (Macfarlane, 1996). However, the use of comparative performance measures of health-care became widespread only recently (Goldstein and Spiegelhalter, 1996). Provider profiling is currently being conducted by a wide range of institutions for a variety of purposes. These applications range from quality improvement through provider feedback, provider selection, provider compensation, provider marketing and sometimes punitive sanctions (e.g., McNeil, 1992; Normand *et al.*, 1997; Stangl and Huerta, 2000).

In the work presented here, complex non-Gaussian hierarchical models are developed to profile US Department of Veterans Affairs (VA) facilities. The main question to be addressed is whether or not there are relevant differences in performance in *out-patient substance abuse* programs at the hospital level across the entire VA hospital system. This chapter introduces a class of hierarchical models suitable for the problem at hand including a preliminary analysis of a single year of data.

The next chapter extends the model to include a time series structure to analyze ten years worth of data. Contributions and methodology in this work represent research performed in consultation and collaboration with the Veterans Affair Management Science Group, Bedford MA.

6.1 The VA Hospital System

The VA operates over 170 hospitals as one of the largest U.S. health-care delivery systems. The system provides a broad spectrum of medical, surgical, and rehabilitative care to US military veterans with a priority on poor veterans and those with service connected injuries. Intense interest has arisen in the development of *performance indicators* to make quantitative comparisons between VA facilities. These indicators can be used also by the hospitals to identify variations in practice patterns or potential quality problems and can be compared to evolving standards being developed outside the VA.

In particular, the VA Management Science Group is interested in assessing and comparing *clinical and health-care* process performance between facilities. Primary goals of statistical analysis are to help define and estimate measures of hospital level performance over time and to include information on individual patients in the study.

The VA has collected and analyzed data from across the system (Burgess *et al.*, 2000). Time to follow-up care of individuals for selected mental-health disorders has been tracked as a measure of quality of care in the medical center facilities.

The study of VA quality over time began with a series of analyses in West and Aguilar (1997), Aguilar and West (1998), and West *et al.* (1998) that focused on annual quality measures over the years 1988-1995. These studies developed statistical models for patient "return for follow-up" data at a highly aggregated level. The data studied annual quality measures at hospital-level. Similar issues are explored here, but now focusing on patterns of variability in more disaggregated data, namely the *individual patient return-times* categorized by hospital and several other socio-demographic and medical history covariates. As with the earlier studies, the key motivating concern is to evaluate differences in return-time distributions by -hospital and across years- in the context of a range of possible individual-level explanatory variables.

6.2 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. The focus here is entirely on the *Substance Abuse Psychiatric* care area, of key current interest to VA policy makers, and an example of over eighty VA care areas. The data consist of 463,015 records of individual visits in substance abuse psychiatric care across the system in the ten fiscal years 1988 to 1997, inclusive. Each record consists of (a) the return-time in days, measured from the day of discharge from initial visit to the day of return to follow-up care; (b) hospital/station number where the patient is receiving treatment; and (c) particular information on the individual which will be used as covariates in the models described later in the chapter. Following initial consultation with VA personnel, the following definitions and groupings of the covariates are adopted:

- ***Age factor***: classifying cases as in *Age group* 1 (age$\leq$ 44 years), 2 (45 $\leq$age$\leq$ 64 years), and 3 (age$\geq$ 65 years).
- ***Diagnosis related group*** (*DRG*): classifying cases in four groups by diagnosis. Patients in the same *DRG* will use roughly the same amount of health-care resources: (1) *DRG 434*, alcohol/drug abuse with complications or comorbidities, (2) *DRG 435*, alcohol/drug abuse without complications or comorbidities, (3) *DRG 436*, alcohol/drug dependency with rehabilitation therapy and (4) *DRG 437*, alcohol/drug dependency with rehabilitation and detoxification.
- ***Marital status***: classifying cases as of *Marital status* 1 (*Married*), 2 (*SDW*, Separated - Divorced - Widowed), and 3 (*UN*, Unknown - Never married).
- ***Priority code status***: based on a means test indicator that defines eligibility priority codes for use of VA services. Classifying cases in as 1 (*AN*, poor veterans), 2 (*AS*, veterans with service connected injuries) and 3 (*Others*).
- ***Gender***: classifying cases as 1 (*Male*) and 2 (*Female*).
- ***Race*** or national origin: classifying cases as 1 (*White*), 2 (*Black*, not Hispanic), 3 (*Hispanic*) and 4 (*Other*, including Asian, American Indian and unknown).
- ***Diagnosis***: classifying cases into one of 11 groups, labeled $1-11$ and associated with the principal medical diagnosis code of the case from *The International Classification of Diseases, Ninth Revision, Clinical Modification* (ICD-9-CM), as follows: (1) Chronic alcohol dependency, (2) Other drug dependency, (3) Acute alcohol dependency, (4) Alcoholic psychoses, (5) Opiate dependency and combinations, (6) Drug psychoses, (7) Alcohol abuse, (8) Drug abuse, (9)

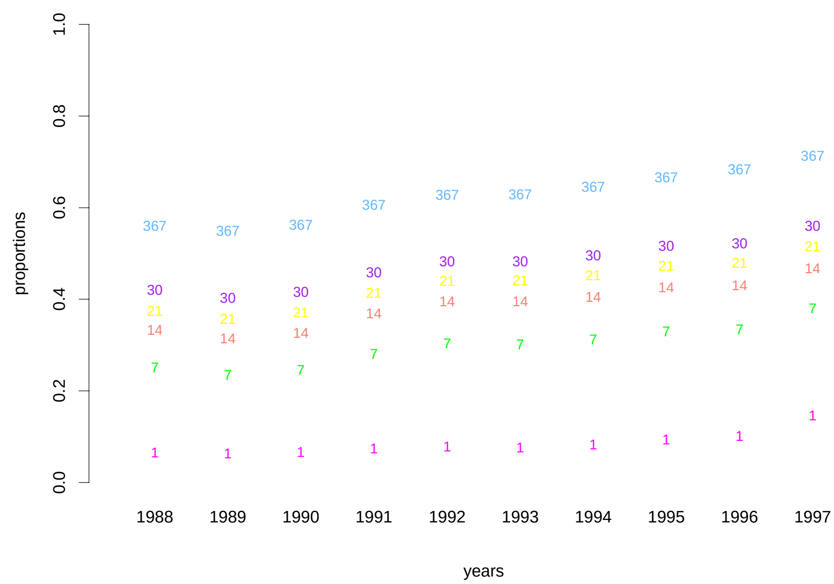

Figure 6.1: Cumulative empirical proportions of return-times at specific "cut-offs": $1, 7, 14, 21, 30$ and 367 days.

All non-mental health diagnoses, (10) Other disorders and (11) Non-substance abuse psychoses.

- ***Other variables***: a few additional categorical covariates relating to socio-economic and military service history of individuals, and region of country in which the facility is located. These variables will not be used in this work.

It is important to bear in mind that the measure of "quality" for this study is the reported time of return to follow-up care where a high probability of "long" return represents "poor quality" and a high probability of "rapid" return represents "high quality".

Figure 6.1 summarizes proportions of individuals classified according to a certain "cut-off" on the return-time scale based on information based on the VA. For every year, the proportion of individuals having return-time less than or equal to the cut-off is displayed. The plot illustrates an increasing trend on time of the proportions for all cut-offs, consistent with a general quality improvement in the system. Return-times are truncated at one year (367 days) and patients with return-times greater than 367 days are considered as uninformative censored cases, i.e essentially as non-returners.[1] The rational here is that after some period of time, veterans that did not return can be assumed to be censored and possibly receiving follow-up care in another health-care setting. With no way of linking data across systems to prove this, a one year follow-up is used as a potential censored cut-off for leaving the VA health-care system.

6.3 The GLM

Consider data only collected on one year for all individuals visiting VA facilities in the substance abuse psychiatric care area. Let z_i be 1 if patient i returned for an out-patient visit within t days of discharge and 0 otherwise for $i = 1, ..., N$. Here the cut-off t is specified in advance. Assume conditional independent Bernoulli models,

$$z_i | p_i \sim \text{Bernoulli}(p_i), \tag{6.1}$$

for each individual $i = 1, .., N$, where p_i is the success probability of the event $T_i \leq t$ and T_i is the time to return of patient i. Covariates described in Section 6.2 are used to explain the variability of p_i, that is, following the standard logistic regression framework, a linear relationship between $\mu_i = \text{logit}(p_i) = \log(p_i/(1-p_i))$ and the set of explanatory variables implies

$$\mu_i = \beta_0 + \mathbf{x}_i' \boldsymbol{\theta}$$

for $i = 1, ..., N$, where β_0 represents a baseline return logit probability, $\mathbf{x}_i$ is the column vector of values of the covariates for patient i, and $\boldsymbol{\theta}$ is a regression parameter

[1]VA services are not provided based on strict health system enrollment, as with a managed care system.

column vector relative to the specified covariate structure in $\mathbf{x}_i$. Note that since all of the candidate covariates are categorical, $\mathbf{x}_i$ is a vector of binary dummy variables. By convention, the base level of each categorical covariate is taken as level 1 and the regression parameter is set to zero at the base level, so that the effects of the covariates are fixed and need to be estimated. This applies to all covariates except the variable *Hospital*, which is treated as a random effect.

Note that positive values of effects correspond to increases in return probabilities. In case of the fixed effects this increment is relative to the base level of the covariate; in case of the random effects the increment is relative to an "average" hospital. Larger positive effects correspond to higher return probabilities, hence to shorter return-times and higher quality of care. On the other hand, larger negative effects correspond to lower return probabilities, hence to longer return-times and lower quality of care.

It is important to mention that there are substantive and empirical reasons to think that the model should include main effects only, that is, with no terms representing interactions between covariates. This issue is discussed later in the chapter.

6.3.1 Discrete Duration Model

Now consider model (6.1) for different cut-offs t. For any individual, the dependence of the return-time probability on cut-off is made explicit via

$$p(t) = \mathrm{P}(T \leq t)$$

with

$$\mathrm{logit}(p(t)) = \beta_0(t) + \mathbf{x}'\boldsymbol{\theta}(t). \tag{6.2}$$

This is known as a *duration model* (e.g., McCullagh and Nelder, 1989 and Lindsey, 1997) and in particular as a *discrete duration model* when t is discrete, as it will be for this work. Different values of cut-offs to consider will be $t = 1, 7, 14, 21, 30$ and 367.

6.3.2 Discrete Proportional Odds Model

The duration model on the odds scale implies that, for any $t > s$ and for an individual with regression vector $\mathbf{x}$,

$$\frac{p(t)}{1-p(t)} = \pi(t,s)\frac{p(s)}{1-p(s)}$$

where

$$\log(\pi(t,s)) = \beta_0(t) - \beta_0(s) + \mathbf{x}'(\boldsymbol{\theta}(t) - \boldsymbol{\theta}(s)),$$

and the *odds ratio*, $\pi(t,s)$, represents the ratio of the odds on a return-time no greater than t to the odds on a return-time no greater than s.

A very special case arises if all the covariate effects are constant, $\boldsymbol{\theta}(t) = \boldsymbol{\theta}$ for all t, which implies that $\log(\pi(t,s)) = \beta_0(t) - \beta_0(s)$ for any $t > s$. Hence, the odds on the individual returning before time t is that of returning before time $s < t$ adjusted by the multiplicative factor $\pi(t,s)$ which in this special case, depends only on the difference between the baseline effects $\beta_0(t) - \beta_0(s)$. That is, the odds ratio does not depend at all on the covariates. This model is known as a discrete *proportional odds model* (e.g., McCullagh, 1980 and McCullagh and Nelder, 1989) and will be mentioned again later. The models will be introduced with an in-depth analysis of the data in the most recent year, 1997.

6.4 Single-year Model

The study begins with an in-depth analysis of just one year's worth of data, 1997. Independent analyses are done for each cut-off under the duration model (6.2) to address the key aspects of the study. Namely, to find (1) patterns of hospital specific random effects compared across hospitals for a given cut-off and across cut-offs, (2) patterns of covariate (fixed) effects, and (3) how such patterns may vary with cut-off.

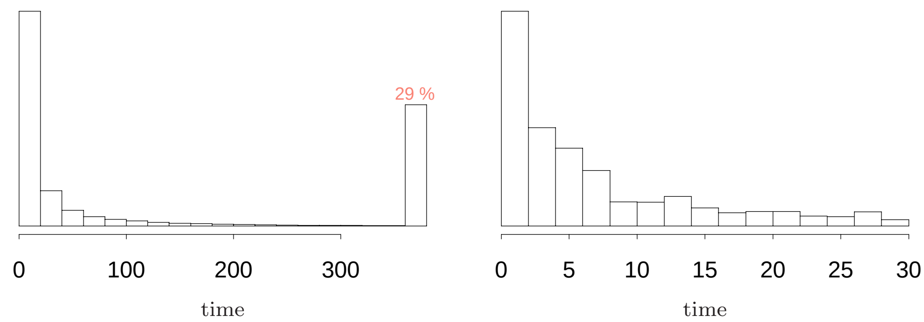

Figure 6.2: Histograms of return-times in FY97. The first histogram includes all 35,368 individuals; the second includes only patients with return-time less or equal to 30 days.

6.4.1 Data Specification

The fiscal year 1997 data set (FY97) includes a total of 35,368 individuals treated at 140 hospitals. This represents a decline from over 50,000 inpatients treated in FY88. Figure 6.2 shows histograms of return-times for these individuals. The first panel includes all observations showing significative numbers of censored cases in this year. The second panel includes only patients with return-times within 30 days, revealing high returns at 14, 21 and 28 days. Both panels indicate that many patients return within a week.

Tables 6.1 and 6.2 provide summary frequencies of individuals classified according to a moving cut-off on the return-time scale. Each row of the table is an estimate of a discretized version of the marginal distribution function of return-times in the specific covariate group of that row. In addition, the total numbers of individuals in each sub-category are indicated by the entries *n:* $\cdot$ in each row. Table 6.2 includes such frequencies for four hospitals selected from the full 140 (hospital numbering serves only to provide labels for hospitals).

From these tables, note that some 14.67% of the total of 35,368 individuals had a return-time of exactly 1 day, 38.03% returned within 7 days, 46.78% returned within 14 days, and so forth. The return-times are truncated at 367 days, with 28.71% of all

Return-time cut-off:	$t = 1$	$t \leq 7$	$t \leq 14$	$t \leq 21$	$t \leq 30$	$t \leq 367$
All data *(n: 35,368)*	14.7	38.0	46.8	51.6	56.0	71.3
***Age group*:**						
≤44 *(n: 16,255)*	15.9	40.1	48.5	52.9	57.3	71.9
45-64 *(n: 16,963)*	14.4	37.9	47.0	52.2	56.8	72.5
65+ *(n: 2,150)*	7.6	23.3	32.0	36.3	40.5	57.1
***DRG*:**						
434 *(n: 7,404)*	13.6	35.0	43.7	48.9	54.2	71.2
435 *(n: 15,407)*	16.5	38.0	45.9	50.4	54.8	70.5
436 *(n: 9,422)*	13.4	40.4	50.3	55.2	59.4	72.5
437 *(n: 3,135)*	12.3	38.3	48.1	52.6	56.6	71.9
***Marital Status*:**						
M *(n: 7,463)*	12.4	35.2	44.7	49.8	54.5	69.2
SDW *(n: 20,527)*	15.3	39.0	47.5	52.2	56.5	72.0
UN *(n: 7,378)*	15.2	38.2	46.9	51.6	56.2	71.5
***Priority*:**						
AN *(n: 23,617)*	15.5	38.5	46.9	51.4	55.6	69.9
AS *(n: 10,078)*	13.6	38.5	48.7	54.4	59.9	78.3
Other *(n: 1,673)*	9.9	28.4	34.2	37.1	39.2	48.2
***Diagnosis*:**						
1. *(n: 17,787)*	14.5	38.4	47.3	52.0	56.3	70.8
2. *(n: 5,829)*	17.0	42.0	51.0	55.8	60.1	74.4
3. *(n: 2,937)*	16.3	37.6	45.6	51.0	55.7	69.7
4. *(n: 2,425)*	12.4	33.0	40.6	45.2	48.9	64.3
5. *(n: 2,309)*	15.9	38.4	46.3	50.1	54.7	72.4
6. *(n: 1,505)*	10.7	33.9	44.5	50.4	54.8	75.0
7. *(n: 1,112)*	11.6	31.9	40.7	46.0	52.5	70.5
8. *(n: 1,006)*	14.6	39.2	47.2	51.9	57.4	77.3
9. *(n: 194)*	8.8	25.3	32.0	36.6	41.2	61.3
10. *(n: 196)*	9.7	33.2	45.4	52.6	60.7	80.1
11. *(n: 68)*	2.9	26.5	35.3	44.1	52.9	73.5
***Gender*:**						
Male *(n: 34,566)*	14.7	38.0	46.7	51.5	56.0	71.3
Female *(n: 802)*	13.2	39.3	48.0	53.2	58.9	71.3
***Race*:**						
White *(n: 20,432)*	14.2	37.3	46.1	51.1	55.6	70.6
Black *(n: 12,095)*	16.1	40.0	48.4	52.9	57.1	72.9
Hispanic *(n: 1,662)*	11.3	34.4	44.7	50.4	54.5	69.8
Other *(n: 1,179)*	13.3	36.1	44.5	49.1	53.9	68.7

Table 6.1: Cumulative empirical frequency distributions of return-times at specific cut-offs, categorized by levels of several primary covariates.

Return-time cut-off:	$t = 1$	$t \leq 7$	$t \leq 14$	$t \leq 21$	$t \leq 30$	$t \leq 367$
***Hospital*:**						
135*(n: 314)*	46.8	62.1	66.2	69.4	72.3	80.3
115*(n: 400)*	16.3	49.5	60.0	64.8	69.3	82.8
1*(n: 373)*	3.0	11.8	18.5	23.3	26.3	39.4
126*(n: 271)*	30.6	59.4	66.8	69.0	70.1	84.5

Table 6.2: Cumulative empirical frequency distributions of return-times at specific cut-offs, categorized by four specific hospitals.

individuals returning at a time greater than 367 days or not at all; these are regarded as uninformatively censored cases and essentially as non-returners as discussed above. In the case of the covariates, for example, 15.90% of all the 16,255 individuals in *Age group* 1 (less than 45 years of age) have return-times of exactly 1 day, compared to only 7.63% in *Age group* 3 (greater than 65 years of age).

6.4.2 Screening Covariates

This section describes initial work of exploratory Bayesian modeling as a screening analysis to identify potentially relevant covariates from the set of key variables discussed above. Approximate Bayesian model selection methods were used based on approximate Bayes' factor and corresponding model probabilities as discussed in Section 4.4.2. The study here used a slight modification of code by Volinsky (1996). The logistic regression in (6.2) with a chosen set of covariates is identified as a *model* and suppose the set of such models is indexed by $k = 1, \ldots, H$. Therefore models, labeled M_k, $k = 1, \ldots, H$, will differ in the specific covariates they select as "in" the model. Theoretically, assuming a uniform prior probability of $1/H$ on each possible model, Bayes' theorem implies that the posterior probability of model k is given by

$$p(M_k|D) = \frac{p(D|M_k)}{\sum_{i=1}^{H} p(D|M_i)}, \tag{6.3}$$

where D is the observed data and $p(D|M_k)$ represents the marginal data density under model M_k. A standard asymptotic approximation to these posterior probabilities provides for simplified computations that are useful in preliminary screening for covariates. This approximation, based on the Laplace method for integrals commonly used in Bayesian asymptotics, delivers the standard Bayesian information criterion (BIC)-based posterior probabilities

$$p(M_k|D) \approx \frac{\exp\left(-\frac{1}{2}BIC_k\right)}{\sum_{i=1}^{H} \exp\left(-\frac{1}{2}BIC_i\right)},$$

where the BIC measure for model k is $BIC_k = Dev_k - df_k log(N)$; here Dev_k is the classical deviance of model M_k, N is the total number of observations, and df_k is the degrees of freedom associated with M_k (i.e., N minus the number of regression parameters in the model).

The approximate posterior model probabilities are computed in this way for all possible combinations of covariates. This preliminary screening analysis assumed that the *Hospital* effects are fixed effects, rather than random, to simplify the computations. Tables 6.3 and 6.4 provide summaries of the results, identifying key covariates and giving summaries of the approximate posterior probabilities and BIC measures for several of the most relevant models. The full screening analysis was performed twice for each possible return-time cut-off: once using only uncensored data (Table 6.3), and secondly using all the data (Table 6.4). The tables indicate, by the entries $\times$ in each column, which covariates are selected in a set of chosen models and analyses, represented by the rows. For example, the first three rows of Table 6.3 refer to the analyses of the uncensored data alone with models having return-time cut-off $t = 1$. It turns out that three models were selected. The first model has an approximate posterior probability of 0.97 for covariates *Hospital, DRG, Age, Priority* and *Diagnosis*, excluding *Marital Status, Gender* and *Race*.

In the analyses based on all the data, relevant selected models include the covariates *Hospital, DRG, Age, Priority* and *Diagnosis*, across all possible return-time cut-offs. In analyses using only the uncensored data, the covariates *Hospital, DRG* and *Diagnosis* are always selected, and *Age* and *Priority* are only excluded in a few models with low probability. *Gender* and *Marital Status* are present in some models with low posterior probability. The most important case for *Gender* is at return-time cut-off $t = 30$, when there is a small but non-negligible posterior probability to include this variable in the model. *Marital Status* is included in some lower probability models in the analyses using all data, but only at return-time cut-offs $t = 1, 7$. Based on this exploratory screening analysis, covariates *Hospital, DRG, Age, Priority* and *Diagnosis* will be included in models for more formal study. Even though *Gender, Marital Status* and *Race* factors appear to be only of marginal relevance, they will be included in formal analysis as specific interests from the VA Management Science Group exist in quantifying their effects.

Full Bayesian and maximum likelihood analyses were performed on the primary *Hospital* effects for the models selected as displayed in the tables. It is of real note that, across various models with different subsets of covariates, these point estimates of Hospital effects are very stable indeed, varying negligibly across models. This is most reassuring, as it indicates that inferences about these primary effects will be robust to the issue of whether or not to include marginally interesting covariates, and also that unmodelled interaction effects are likely small and may be safely ignored.

6.4.3 Model Specification and Implementation

The analysis strategy involved fitting the full logistic model of equation (6.2) to the data, and repeating the analysis in separate studies based on cut-offs at $t = 1, 7, 14, 21, 30$ and 367 respectively. Including the covariates displayed in Table 6.1

Return-time	Hospital	DRG	Age	Marital Status	Priority	Diagnosis	Gender	Race	Post. Prob (%)	BIC
	×	×	×		×	×			96.50	-124812.93
$t = 1$	×	×	×			×			2.39	-124805.53
	×	×	×		×	×	×		1.11	-124804.01
$t \leq 7$	×	×	×		×	×			100.00	-119399.98
$t \leq 14$	×	×	×		×	×			100.00	-119985.11
	×	×	×		×	×			94.86	-121003.07
$t \leq 21$	×	×			×	×			2.78	-120996.02
	×	×	×			×			1.38	-120994.61
	×	×	×		×	×	×		0.98	-120993.93
	×	×	×			×			79.94	-122626.23
	×	×	×		×	×			13.97	-122622.74
$t \leq 30$	×	×	×			×	×		4.53	-122620.49
	×	×	×		×	×	×		0.89	-122617.23
	×	×			×	×			0.68	-122616.68

Table 6.3: Selected covariates using the BIC based on uncensored data.

Return-time	Hospital	DRG	Age	Marital Status	Priority	Diagnosis	Gender	Race	Post. Prob (%)	BIC
	×	×	×		×	×			97.65	-167137.68
$t = 1$	×	×	×	×	×	×			1.28	-167129.01
	×	×	×		×	×	×		1.07	-167128.65
$t \leq 7$	×	×	×		×	×			98.60	-158916.50
	×	×	×	×	×	×			1.40	-158907.99
$t \leq 14$	×	×	×		×	×			100.00	-157769.12
$t \leq 21$	×	×	×		×	×			99.09	-157536.78
	×	×	×		×	×	×		0.91	-157527.40
$t \leq 30$	×	×	×		×	×			96.71	-157713.02
	×	×	×		×	×	×		3.29	-157706.26
$t \leq 367$	×	×	×		×	×			100.00	-160494.80

Table 6.4: Selected covariates using the BIC based on all data.

and the random effects, equation (6.2) is:

$$\operatorname{logit}(p(t)) = \beta_0(t) + \epsilon_j(t) + \begin{cases} \delta_d(t) & \text{for } \textit{Age group } d = 2, 3, \\ \gamma_g(t), & \text{for } \textit{DRG level } g = 2, 3, 4, \\ \kappa_k(t), & \text{for } \textit{Marital status group } k = 2, 3, \\ \eta_e(t), & \text{for } \textit{Priority level } e = 2, 3, \\ \xi_x(t), & \text{for } \textit{Diagnosis group } x = 2, \ldots, 11, \\ \chi_c(t), & \text{for } \textit{Gender group } c = 2(\text{women}), \\ \zeta_z(t), & \text{for } \textit{Racial group } z = 2, 3, 4, \end{cases} \tag{6.4}$$

where all parameters are cut-off t specific, β_0 is the baseline return logit probability, ϵ_j is the random effect associated to hospital j and δ_d, γ_g, κ_k, η_e, ξ_x, χ_c and ζ_z are fixed effects associated to each level d, g, k, e, x, c, z of the covariates.

West and Aguilar (1997), Aguilar and West (1998), and West *et al.* (1998) considered similar models at the single cut-off of $t = 30$ and using highly aggregated data at the hospital level involving no individual-level covariates. On the other hand, the models presented here expand the scope to several cut-offs on the return-time scale, so providing access to information about hospital and covariate effects at fine levels of detail that may be relevant in assessment and interpretation in connection with possible VA policy questions and appropriate profiling of VA psychiatric care units. In addition, these generalizations of previous models imply significant complications technically and computationally in terms of dealing with a vastly larger and more complex database.

Each specific analysis involved completing the model specification with prior distributions for covariate effects and hyperparameters, and then computation of posterior distributions. Within each analysis, the hospital specific random effects were assumed drawn from a normal population model $\epsilon_i(t) \sim N(0, w(t)^2)$ where the cut-off dependent standard deviation $w(t)$ represents the dispersion in effects across the VA system. On the other hand, very vague but proper priors are adopted for all fixed effects parameters, the effects of all other covariates. Specifically, the elements of $\boldsymbol{\theta}(t)$

have independent, zero-mean normal priors with variances of 1000, and the hospital precision parameter $1/w(t)^2$ has a gamma prior with shape and scale parameters both equal to 0.001.

Posterior analysis uses Markov chain Monte Carlo (MCMC) methods to iteratively simulate from the full joint posterior distribution of $\boldsymbol{\theta}(t)$ and $w(t)$ producing large Monte Carlo samples for summary inferences on all parameters and effects. Separated and uncorrelated analyses were performed for each chosen cut-off t. The sampler was implemented using WinBUGS software (Gilks *et al.*, 1996; Spiegelhalter *et al.*, 1999).

6.4.4 Analysis Results

Posterior summaries for the single-year model are displayed in Figures 6.3 to 6.12 inclusive. The graphs show, for selected model parameters, approximate 95% posterior intervals with posterior medians and quartiles. These intervals are all presented as vertical lines in frames designed so that it is straightforward to make comparisons. In addition, each frame displays intervals from each of the independent analyses corresponding to different cut-offs t. All such analyses were performed twice, one set of analyses used only individuals whose return-times were less than 367 days and a second set of analyses used all the data, including those censored. The analyses are distinguished using dotted lines for intervals based on the full data set and solid lines for intervals based on the data set excluding censored cases. In the next sections, comments on the posterior graphs are made.

Baseline Duration Model Parameters

The baseline parameters β_0 in (6.2) represents the response probability on the logit scale at an "average" hospital and for individuals in the base levels of all other covariates.

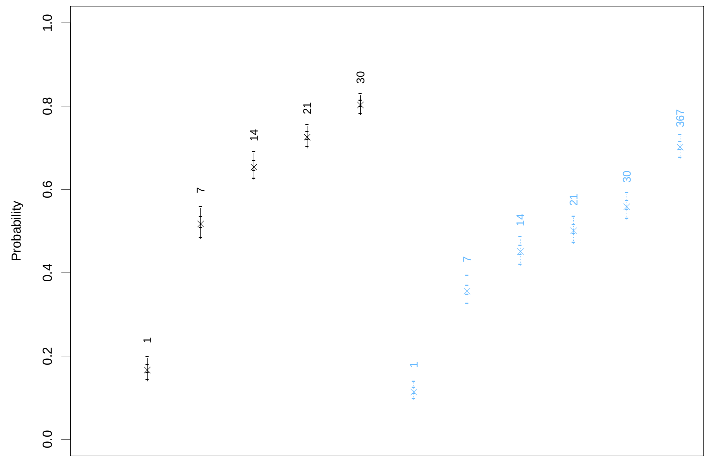

Figure 6.3: 95% posterior intervals for baseline return-time probabilities $p_0(t)$.

Posterior distributions of these parameters are plotted in Figure 6.3 in the probability scale. The graph displays intervals for $p_0(t) = 1/(1 + \exp(-\beta_0(t))$ in each of the independent analyses with different cut-offs in the restricted data set ignoring the censored return-times (solid) and with all the data (dotted). The $\times$ symbols indicate posterior medians. Note that $p_0(t)$ increases as a function of t, as expected, although the analyses were not linked to enforce monotonicity.

Analysis of Covariates

Considering the regression parameters, Figures 6.4 to 6.9 present intervals and estimates of the fixed effects; recall that the effects are referenced to the effect of zero in the base level of each covariate. Figure 6.4 displays *Age* effects $\delta_j(t)$. The plot suggests some evidence of non-proportional odds behavior in the *Age* covariate, with group 2 (45-64) effects being less than zero for thresholds $t = 1, 7$ and 14, but not obviously different from zero for higher values of t. This corresponds to initially decreased quality, in terms of lower return probabilities, at early times for *Age group* 2 relative to the younger group, but that these differences disappear after 3 weeks. The older individuals have uniformly lower probabilities of return at all stages, and the

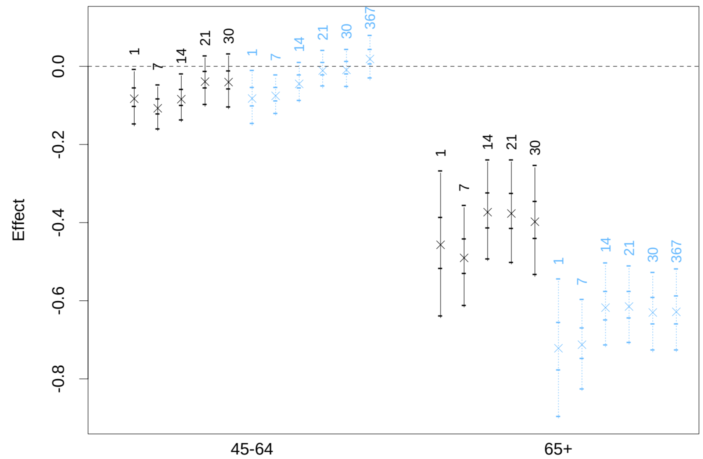

Figure 6.4: 95% posterior intervals for *Age* effects $\delta_d(t)$ relative to *Age group* 1 (≤ 44 years).

effect seems to be roughly constant with respect to cut-off. The results are consistent between the two data sets (with and without censored cases) in *Age group* 2, but clearly not in *Age group* 3. In this elderly category, the effects are apparently lower in the analysis of the full data set, consistent with the raw frequencies that indicate larger numbers of older individuals as non-returners. The generally negative effects in group 3 may reflect the possibility that elderly veterans who are also eligible for Medicare may find it easier to have out-patient psychiatric follow-up care covered in the private sector rather than in VA hospitals.

Figure 6.5 plots *DRG* effects $\gamma_g(t)$. The graph illustrates that there is a decreasing trend with t in the posterior median for *DRG 435*. Though there is considerable uncertainty, this is suggestive of a non-proportional odds effect and would imply a persistent decrease in return probability with threshold at later times. In *DRG 436* and *437*, there are clear differences between estimated effects for $t = 1$ relative to later cut-offs. These *DRG* categories have generally higher return probabilities than the rest, with the very significant exception of the immediate returns. At $t = 1$, the return probabilities in these two groups are essentially consistent with the base level

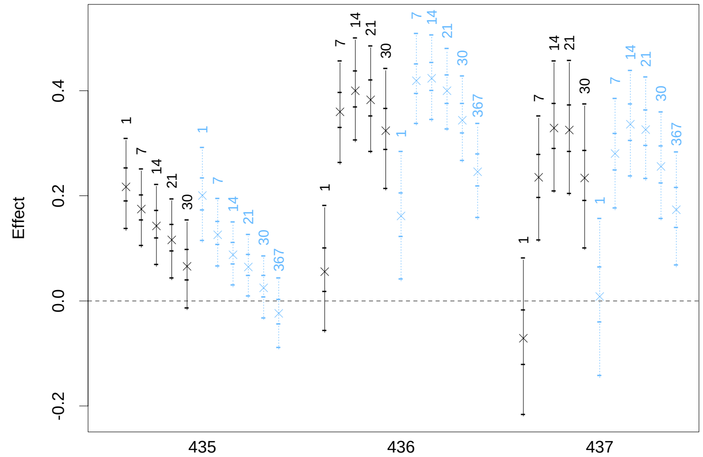

Figure 6.5: 95% posterior intervals for *DRG* effects $\gamma_g(t)$ relative to *DRG* 434.

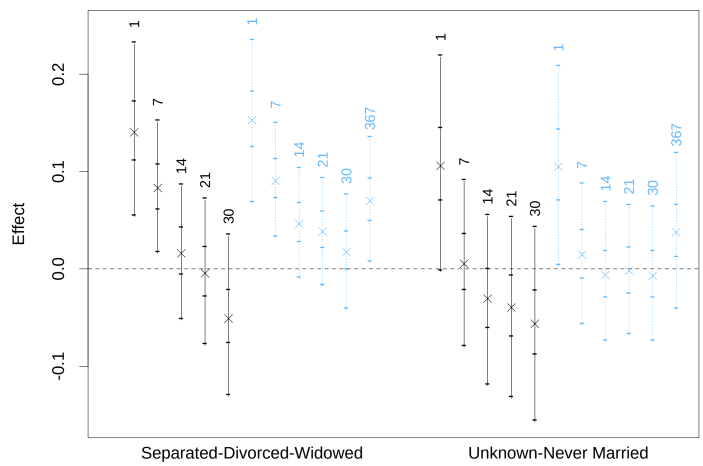

Figure 6.6: 95% posterior intervals for *Marital Status* effects $\kappa_k(t)$ relative to *Married* status.

DRG 434; by comparison, *DRG 435* is clearly above the base level at $t = 1$.

Figure 6.6 shows *Marital Status* effects $\kappa_k(t)$. Although it is believed that marital status would affect the efficacy of the support network where married patients might have different pattern in follow-up appointments than the rest, the plot suggests that there are no major differences between the groups except perhaps at the early return-times $t = 1, 7$. The overall effects are small in terms of their impact on return probabilities. There is some evidence of increased probabilities of return immediately,

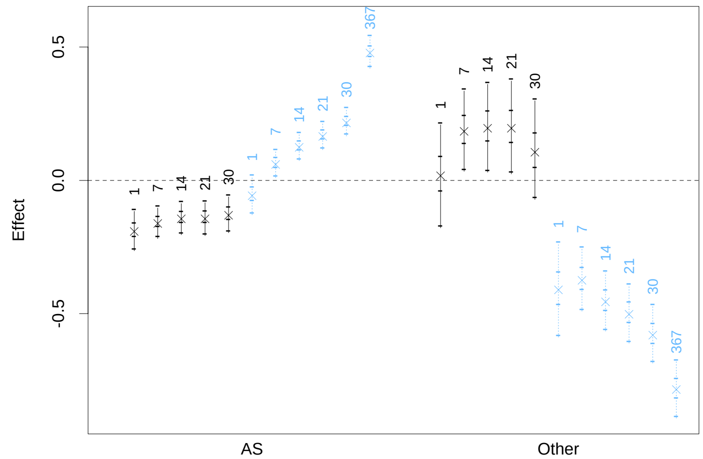

Figure 6.7: 95% posterior intervals for *Priority* effects $\eta_e(t)$ relative to *AN* (poor veterans).

$t = 1$, in both groups 2 and 3 relative to that among *Married* individuals, and perhaps a minor increase up to $t = 7$ among *Separated - Divorced - Widowed* individuals.

Figure 6.7 displays *Priority* effects $\eta_e(t)$. The graph presents different results for the analysis of the data excluding the censored cases and the analysis of the full data set. When excluding the censored cases, *Priority* group *AS* (veterans with service connected injuries) has consistently lower return probabilities than the rest, and the *Other* group is consistently higher except at $t = 1$ and 30. The estimated effects generally appear constant, consistent with a proportional odds structure, with that one exception. On the other hand, when including all data, *Priority* group *AS* evidences increasing return probabilities that are higher than average apart from at $t = 1$, and the *Other* group has dramatically lower return probabilities which appear to decrease at higher return-times. VA *Priority* levels are hypothesized to have two possible effects which are important in interpreting these results. First, higher priority veterans (*AN*,*AS*) may be given better access to the system and priority in getting favorable appointments, although once veterans are accepted for care they are supposed to be treated the same. Second, *Priority* level also affects the likelihood of leaving the

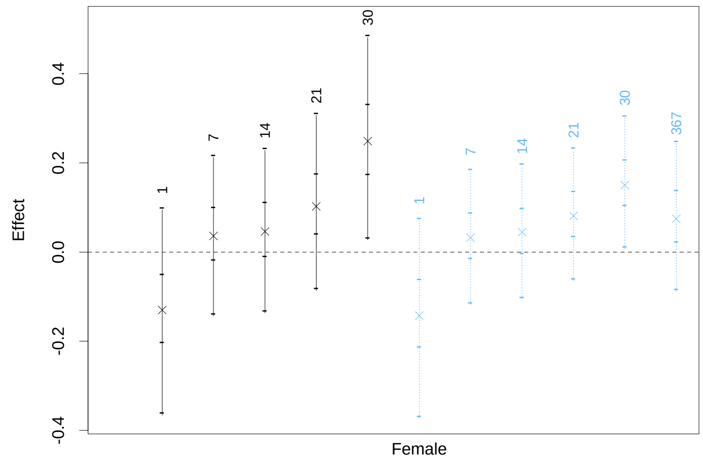

Figure 6.8: 95% posterior intervals for *Gender* effects $\chi_2(t)$, *Female* category relative to *Male*.

system to receive private sector care, specifically, the *Other* category has a higher probability of leaving and hence a lower probability of returning. Therefore, once the censored cases are excluded, this effect should be reduced. Censored individuals may well be receiving their follow-up psychiatric care in the private sector, as mentioned earlier.

Diagnosis covariate (not plotted), as explained in Section 6.2, has eleven levels. Levels 9, 10 and 11 effects are much less precisely estimated than the rest due to smaller sample size in these groups (see Table 6.1). There is some evidence of non-proportional odds behavior in several of the *Diagnosis* categories, though this is neither highly significant nor uniform across categories. The categories with higher labels ($j = 5, 6, \ldots$), tend to have generally negative effects, and hence return probabilities lower than the earlier categories relative to the chronic alcohol dependency reference group. The small sample sizes in many diagnosis groups suggests the merging of similar diagnosis groups.

Figure 6.8 displays *Gender* effects $\chi_2(t)$. The plot suggests that there is a generally increased effect with longer return-times, consistent with small increases in return probabilities for women relative to men at later times. The effects are rather

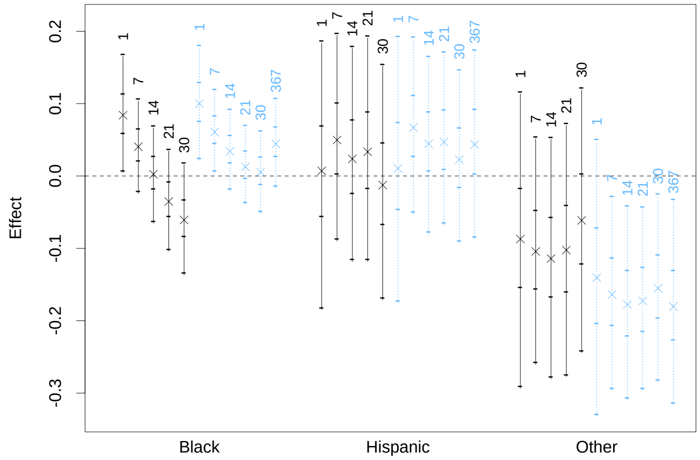

Figure 6.9: 95% posterior intervals for *Racial* effects $\zeta_z(t)$ relative to *White*.

uncertain, though. Only at $t = 30$ in the analysis excluding censored data the effect is really significant. At $t = 1$ the effect $\chi_2(1)$ is rather uncertain, though the posteriors do give more weight to negative values, which would indicate a small decrease in the probability of an immediate return for women relative to men. There may be two potentially offsetting effects for *Gender*. In general, women can be more compliant with follow-up care effects; however, the dominance of male patients in the VA system can encourage women to shy away from VA care or return to private sector care.

Figure 6.9 presents *Racial* effects $\zeta_z(t)$. The graph illustrates that there are no important differences in the effects across levels. For example, the effects in the *Hispanic* category are around zero, there is no evidence for a difference with respect to *White*; and the effects in the *Other* category are lower but highly uncertain. The effects in the *Black* category indicate a monotonic decreasing pattern in the posterior medians that is at least suggestive of a small but persistent decrease in return probabilities, relative to *Whites*, at later return-times. The median estimates alone are consistent with slightly higher return probabilities for *Blacks* relative to *Whites* at the early return-times up to 7 days, though the magnitude of the difference is practically insignificant. In general *non-Whites* are more likely to use VA out-

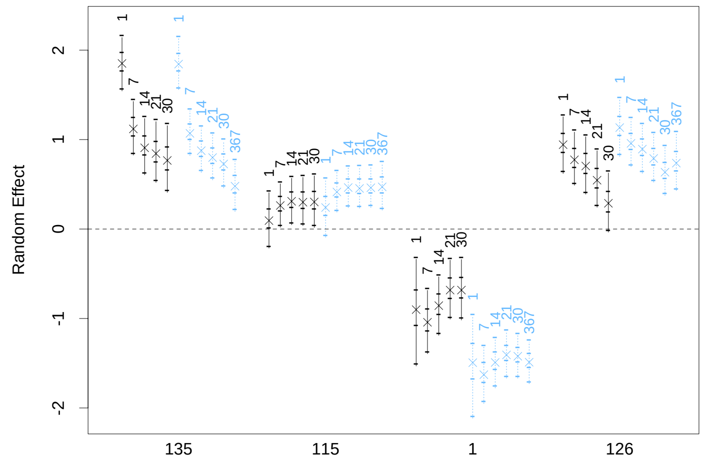

Figure 6.10: 95% posterior intervals for hospital specific random effects $\epsilon_j(t)$ for four selected hospitals indicated by station numbers.

patient services (Burgess and DeFiore, 1994), however, conditional on admission, *Whites* might follow-up more effectively after discharge.

Random Effects

Figure 6.10 displays inferences on hospital random effects for four arbitrarily selected hospitals, those with station numbers 135, 115, 1 and 126. The most important general finding is that there are general differences between hospitals and the differences are consistent across analyses. These specific hospitals would be ranked (from low quality to higher quality) as 1, 115, 126, 135. Of these, hospital # 1 is clearly below average, with effects $\epsilon_1(t)$ clearly negative. There is very clear evidence that the effects represent non-proportional odds structure in the return-time regression model for at least one of the hospitals; consider hospital # 135, and note that the clear differences in inferences for this hospital's effect at $t = 1$ compared to the other values of return-times. The parameter $\epsilon_{135}(1)$ is obviously higher than the rest, indicating that the probability of return in exactly one day for this hospital is increased by more, relative to the average, than is the probability of return at later days. For the remaining cut-offs at this hospital, the intervals overlap so the differences are unclear;

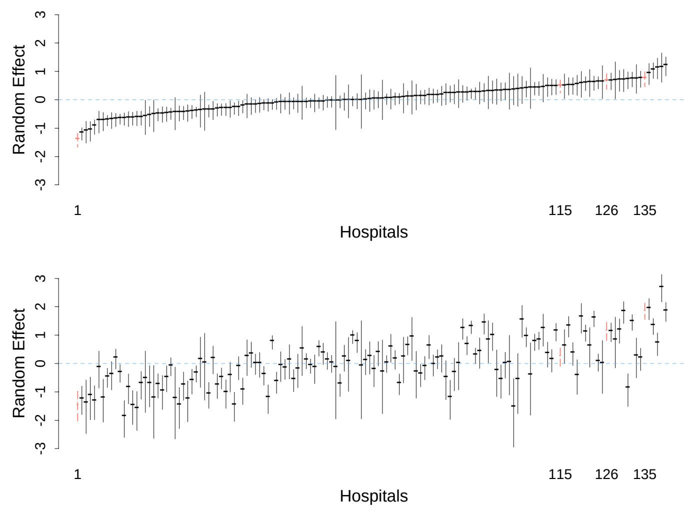

Figure 6.11: 95% posterior intervals for Hospital random effects $\epsilon_j(t)$ in analysis with cut-off $t = 30$ (top) and $t = 1$ (bottom). Hospitals are ordered by the posterior medians of the effects under the analysis with $t = 30$.

however, the decreasing pattern of the posterior medians as t increases is suggestive. This would imply that the increased quality, in terms of increased probability of a return, exhibits a diminishing effect at later times. The same may be said for hospital # 126. For hospitals # 115 and # 1, the intervals are suggestive of a constant effect across cut-offs, i.e., $\epsilon_j(t) \approx \epsilon_j$ for $j = 115, 126$. The results are similar between the two data sets (with and without censored cases) for all but hospital # 1. Here the effects are apparently lower in the analysis of the full data set, consistent with the raw frequencies that indicate over 60% censored cases at this facility.

The hospital specific random effects $\epsilon_j(t)$ $t = 1, 30$ (all data set) are graphed using line plots in Figure 6.11. The lines represent approximate 95% posterior intervals and the segments indicate the median of each effect. The graphs serve to reinforce the conclusion that some or many of the hospital effects are indeed different at different return-time cut-offs t, consistent with a general non-proportional odds structure. The hospitals have a common ordering across the two figures, chosen simply as the order of the posterior medians on the effects in the analysis with $t = 30$. In the analyses

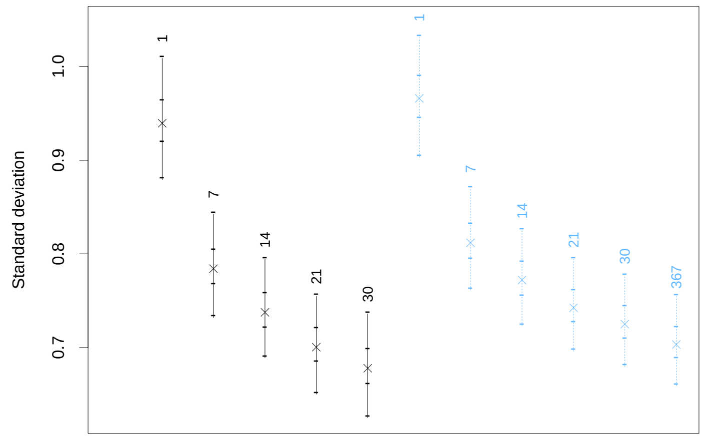

Figure 6.12: 95% posterior intervals for the standard deviation $w(t)$ of the population distribution of hospital specific random effects ϵ_j.

with different cut-offs, the appearance is no longer monotonic, indicating that the relative quality levels of hospitals is indeed quite variable with cut-off. Indications of uncertainty are also clear in these graphs where hospitals with smaller samples sizes present wider intervals.

The standard deviations $w(t)$ determine the dispersion in the random effects distribution on the population of hospitals. Figure 6.12 provides posterior intervals and medians for the $w(t)$ in each of the analyses. The major point to note here is the apparent decreasing pattern as the value of the cut-off t increases. This is particularly significant in moving from $t = 1$ onward. Thus, variability in the distribution of hospital-specific effects is significantly more marked at $t = 1$ than at later times, and tends to decrease with increasing t. This is consistent with features evident in Figure 6.11 where the variance of the effects for $t = 1$ is clearly higher than with $t = 30$. And it is evident also in Table 6.1 where the differences in raw frequencies across the 4 selected hospitals are less marked at higher cut-offs.

6.5 Summary

This chapter presents a class of hierarchical logistic model to profile 140 VA facilities within the fiscal year 1997. The key motivating concern is to evaluate differences in return-time distributions that are specific to each hospital in the context of a range of possible individual-level explanatory variables. Summary conclusions arising from this analysis are as follows.

- Hospital-specific random effects are the means to compare hospital performance within the VA system. Higher positive effect is translated as higher return probability and therefore higher quality.

- The posterior distribution of the effects with the two data sets (with and without censored cases) lead to similar conclusions in terms of profiling facilities. It is believed that censored cases does not give significant information about the effectiveness of health policy interventions, therefore the VA Management is not interested in including specific modeling of the censored cases.

- Exploratory and confirmatory analyses of FY97 data alone help to understand key aspects of the structure of this massive data set. In relation to the covariates, there are two explanatory variables, *Marital Status* and *Race*, where the overall effects for all levels are small in terms of their impact on return probabilities. These results suggest that *Marital Status* and *Race* should be dropped for future analyses. On the other hand, within the two covariates *Age* and *Diagnosis*, there are some levels with significant effects and some with no major differences between them, therefore suggesting groupings within these variables. Posterior distributions of the covariate effects and the exploratory analysis using the BIC lead to similar conclusions.

- A time to return threshold of 30 days has been commonly considered in many studies in health-care profiling; actually, it is used as an industrial standard. In this work, the analysis with a 30 day cut-off leads to somewhat less “noisy” results than other potential cut-off times since it appears to limit the effects of unobserved heterogeneity at the individual-level. This can be seen in the hospital level comparisons and the random effects variances above.

The next chapter will extend the logistic model presented here to include a time series structure to analyze ten years of data, and hence to develop the quality of care comparisons across years both within and between hospitals.

Chapter 7

Non-Gaussian Hierarchical Time Series Structure for Models in Institutional Profiling

This chapter contains extension of the models for profiling VA facilities as presented in Chapter 6, to include dependencies between hospital-specific effects from year-to-year. The motivation is based on policy interests to monitor impact of changes on internal policy by assessing changes over time in measures of hospital-level performance. This is of key interest for the VA system due to the close connection with the development of management and economic incentives designed to encourage and promote care provision at sustained and acceptable levels.

The idea is to extend the simple model presented in the previous chapter by including a simple time series structure that relate the hospital-specific effects between years, while maintaining the same natural random effects/hierarchical model within years. The time series components are introduced to adequately capture aspects of heterogeneity across the VA hospital system and to understand patterns of return-time variability over time. This is a new class of time series models developed for longitudinal data that relates individual and hospital level effects.

7.1 Multi-year Model

Ten years of return-time data, 1988-1997 inclusive, are analyzed and discussed in this chapter. Prior studies have similar objectives and developed models at a very highly aggregated level, using as data only the total numbers of patients with return-times at a specific cut-off, i.e, considering no covariates other than *hospital*.

The current work is therefore a generalization of previous models, aiming to assess similar issues but now using the hugely larger data set based on data at the patient level. The class of models adopted is again a set of logistic regressions with outcomes classified by return-times below or exceeding one of a set of specified cut-offs. The data analysis reported below is based on the return-time cut-off $t = 30$, using all data.

7.1.1 Data Specification

A total of 463,015 individual psychiatric discharges were recorded across the 136 profiled hospitals in the VA system from 1988-1997. Note that there are 136 hospitals compared to the 140 used in the FY1997 analysis above. The reduction is because there are only 136 hospitals with annual records of patients for each year of study in the psychiatric area. The annual numbers of patients among the 136 hospitals are:

1988	1989	1990	1991	1992	1993	1994	1995	1996	1997
52,837	51,451	48,514	46,163	47,054	45,175	47,325	46,438	43,062	34,996

Based on the results of the single-year analysis, some modifications were made on the selection and specification of categorical covariates defined in Section 6.2.

- ***Hospital/station***: a total of 136 facilities with patients recorded in the substance abuse psychiatric care area.

- ***Age factor***: refined to classify cases into just two groups: *Age groups* 1 (age$\leq$ 64 years) and 2 (age$\geq$ 65 years).

- ***DRG factor***: four levels, unchanged.

- ***Priority code status***: three groups, unchanged.

- ***Gender***: 1 (*Male*) and 2 (*Female*), unchanged.

- ***Diagnosis***: refined to classify cases into one of just 3 groups (1) a *Dependence* group that combines the original groups 1,2,3 and 5; (2) a *Psychoses* group that combines the original groups 4,6,9,10 and 11; and (3) an *Abuse* group that combines the original groups 7 and 8.

- ***Marital status*** and ***Racial*** covariates are not now included.

Given this categorical covariate structure, there are a 19,584 cells in the cross-classification (*Hospital*×*Age*×*DRG*×*Priority*×*Diagnosis*×*Gender*). Of these, there are a total of 7,848 cells that are non-empty for at least one year.

7.1.2 Model Structure and Implementation

The basic logistic regression model of equations (6.1) and (6.2) is modified to incorporate multiple consecutive years. Independently across individuals $i = 1, \ldots, N$, and over all years $r = 1, \ldots, 10$, the data are assumed to arise from the set of $10N$ Bernoulli models

$$z_{i,r} | p_{i,r} \sim \text{Bernoulli}(p_{i,r})$$

with logistic regression on the explanatory covariates and random effects:[1]

$$\text{logit}(p_r) = \beta_{0,r} + \epsilon_{j,r} + \begin{cases} \delta_{d,r}, & \text{for } \textit{Age group } d = 2, \\ \gamma_{g,r}, & \text{for } \textit{DRG level } g = 2,3,4, \\ \eta_{e,r}, & \text{for } \textit{Priority level } e = 2,3, \\ \xi_{x,r}, & \text{for } \textit{Diagnosis group } x = 2,3, \\ \chi_{c,r}, & \text{for } \textit{Gender group } c = 2 \ (\textit{women}). \end{cases} \tag{7.1}$$

[1]Note that there is no explicit indication of cut-off in the model. This is because the analysis will focus on a chosen return-time cut-off $t = 30$, dropped from the notation. However, it should be borne in mind that all parameters are cut-off specific.

Observe that the model does not relate the population parameters $\beta_{0,r}$ over the years, hence baseline quality levels are unconstrained in how they vary between years. This neutral standpoint is adopted in order to "let the data speak" about any patterns of variation, or lack thereof, in the baselines. The same attitude is adopted for the effects parameters for all covariates with the exception of the hospital effects. Hence, the covariate effects are treated as nuisance parameters without anticipating systematic structure over time.

The model extensions developed relate to the hospital-specific effects $\epsilon_{j,r}$ across years, in a way similar to prior work with aggregated data (West and Aguilar,1997; Aguilar and West,1998 and West *et al.*,1998). The specific structure adopted involves simple time series panel models to incorporate the view that the $\epsilon_{j,r}$ are expected to remain relatively stable within each hospital from year-to-year, while allowing for unexplained sources of variability at the hospital level that may induce random changes. Unless policies and protocols in the care area 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 will reflect random variations due to hospital specific practices and the characteristics of the patient sample present at the hospital (West and Aguilar, 1997). Therefore, each $\epsilon_{j,r}$ term is modeled as a simple AR(1) time series over the years. This structure captures much of the random-effects variability across hospitals within each year as well as the systematic dependencies within hospitals from year-to-year.

The dependence structure allows for both between-year correlations and hospital specific within year variability. For hospital j, the standard AR(1) model is considered, namely

$$\epsilon_{j,r} = \phi_j \epsilon_{j,r-1} + \omega_{j,r},$$

for years $r = 2, \ldots, 10$, where ϕ_j is the autoregressive parameter which will generally be close to one and is constrained to lie in part of stationary region $0 < \phi < 1$, and the $\omega_{j,r}$ are independent *innovations* distributed as

$$\omega_{j,r} \sim N(\omega_{j,r}|0, u_j^2)$$

for some innovations variance u_j^2. The AR(1) model is such that, for all r and including the first year $r = 1$, the implied *marginal* distribution of hospital effects within the year is simply

$$\epsilon_{j,r} \sim N(\epsilon_{j,r}|0, w_j^2)$$

with marginal variance $w_j^2 = u_j^2/(1 - \phi_j^2)$.

Whereas prior studies had assumed common parameters across hospitals, here the parameters (ϕ_j, u_j) are hospital specific, allowing for variations in both systematic dependency and overall levels of variation in the $\epsilon_{j,r}$ across hospitals.

A hierarchical model structure is adopted for the hospital-specific parameters (ϕ_j, u_j). Assuming that (ϕ_j, u_j) are exchangeable parameters drawn from a hospital-population prior delivers a class of Bayesian hierarchical models. Furthermore, modeling the (ϕ_j, u_j) as a random sample from a common prior implies that the resulting random effects $\epsilon_{j,r}$ follow a common marginal distribution within each year. This hierarchical framework provides flexibility to assess different degrees of dependency for random effects (through ϕ_j) and levels of contribution of the systematic component of variation (through u_j) across hospitals, while maintaining a "neutral" view, initially, as to possible differences between hospitals and over years.

The specific prior distributions adopted here have the following main characteristics. First, ϕ_j and u_j are exchangeable across hospitals. Second, the dependence parameters are drawn from a prior beta distribution,

$$\phi_j \sim \mathrm{Beta}(\phi_j|a\mu, a(1 - \mu))$$

where the underlying average dependence level is represented by the hyperparameter μ, and variations among the ϕ_j are determined by the precision hyperparameter a. These two hyperparameters are to be estimated, along with the ϕ_j themselves. Third, the population of dispersion parameters u_j^2 is assumed an inverse-Gamma distribution,

$$u_j^{-2} \sim \text{Gamma}(u_j^{-2}|c, c\rho)$$

where the hyperparameter ρ represents an underlying average dispersion level, and variations among the u_j are determined by the precision hyperparameter c. In the same way as for ϕ_j, these two hyperparameters are to be estimated, along with the u_j themselves.

Posterior analysis uses Markov Chain Monte Carlo (MCMC) methods to simulate iteratively from the full joint posterior distribution of all model effects and hyperparameters. The sampler was implemented using WinBUGS software (Gilks *et al.*, 1996; Spiegelhalter *et al.*, 1999).

7.1.3 Analysis Results

Summaries of the posterior distributions of the parameters for model (7.1) are displayed in Figures 7.1 to 7.8 inclusive. The graphs present, for selected model parameters, 95% posterior intervals with posterior medians and quartiles marked as in the previous chapter.

Baseline Duration Model Parameters

Figure 7.1 displays posterior intervals and estimates for the baseline return-time probabilities $p_{0,r} = 1/(1 + \exp(-\beta_{0,r}))$ in each of the ten years $r = 1, \ldots, 10$. The graph reveals that there is an increasing trend in probabilities of return over time. Moreover, it shows a distinct change from FY96 to FY97, there being no overlap of the

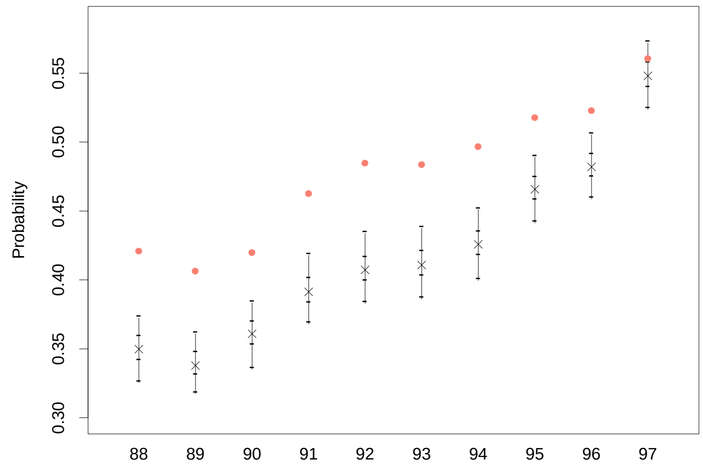

Figure 7.1: 95% posterior intervals for baseline return-time probabilities $p_{0,r}$ over years $r = 1, \ldots, 10$. The points correspond to the overall proportions returning within 30 days from the raw data.

FY97 interval with the rest. As it turns out, FY97 corresponds to the time period where management efforts in the VA system were put in place aimed at reducing psychiatric inpatient days and discharges, along with an emphasis on continuity of care. Figure 7.1 also indicates an increase in the probability of return within 30 days at an "average" hospital from around 35% in 1990 to nearly 55% in 1997. From the raw data sets, the crude aggregate proportions of returners in each year, in terms of percent returning within 30 days, are:

Year:	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997
Percent:	42.0	40.5	41.9	46.1	48.3	48.2	49.5	51.6	52.2	55.9

These observed values are also displayed in the graph as points next to the corresponding interval estimates. Their pattern over the years mainly agree with that of the inferred baseline parameters. However, these data summaries generally exceed the baseline parameters. This is simply a result of the fact that the majority of the observations lie in covariate groups with generally positive effects. For example, results discussed later in the chapter reveal that for FY92, the only categories with negative effects are 65+ in *Age* group and *Others* in *Priority* group with 6% and 3%

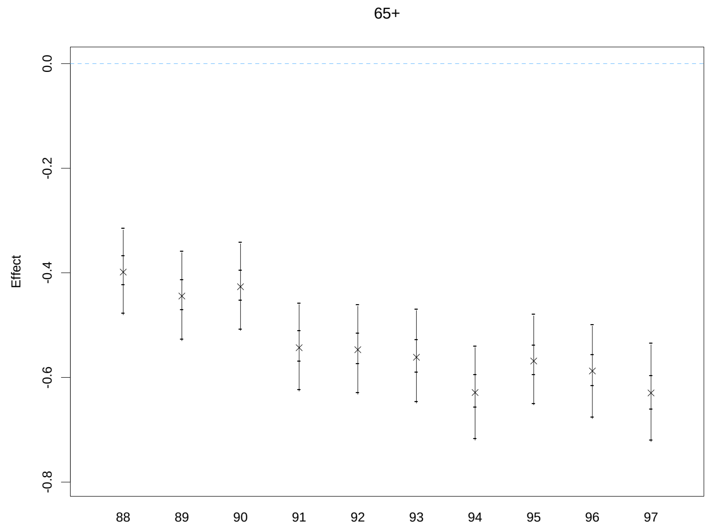

Figure 7.2: 95% posterior intervals for *Age group* effects $\delta_{2,r}$ over years $r = 1, \ldots, 10$ relative to *Age group* 1 (≤ 65 years).

of the FY92 data lying in these levels respectively. Thus, the proportion of returners in less than 30 days for FY92, 48.3%, includes the baseline plus predominantly positive effects. The pattern of change over the years reflects hospital system-wide effects on 30 day return-time probabilities. The effects of system-wide policy changes and common management practices are clear in these improvements in 30-day returns.

Analysis of Covariates

Figures 7.2 to 7.6 present posterior intervals of the fixed effects; recall that the effects are referenced to zero in the base level of each covariate. Figure 7.2 displays posterior intervals of the fixed effects $\delta_{2,r}$ for the *Age* covariate at level $d = 2$ (65+). The elderly group effect is clearly negative in each year, and the graph shows a mild though significant decreasing pattern over the ten year period. Elderly veterans may have greater transportation problems or may find it easier to have out-patient psychiatric follow-up care covered in the private sector, since they are also eligible for Medicare. Therefore, this category has a large number of non-returners. However, with changes in the VA system beginning to mirror private sector psychiatric treatment patterns,

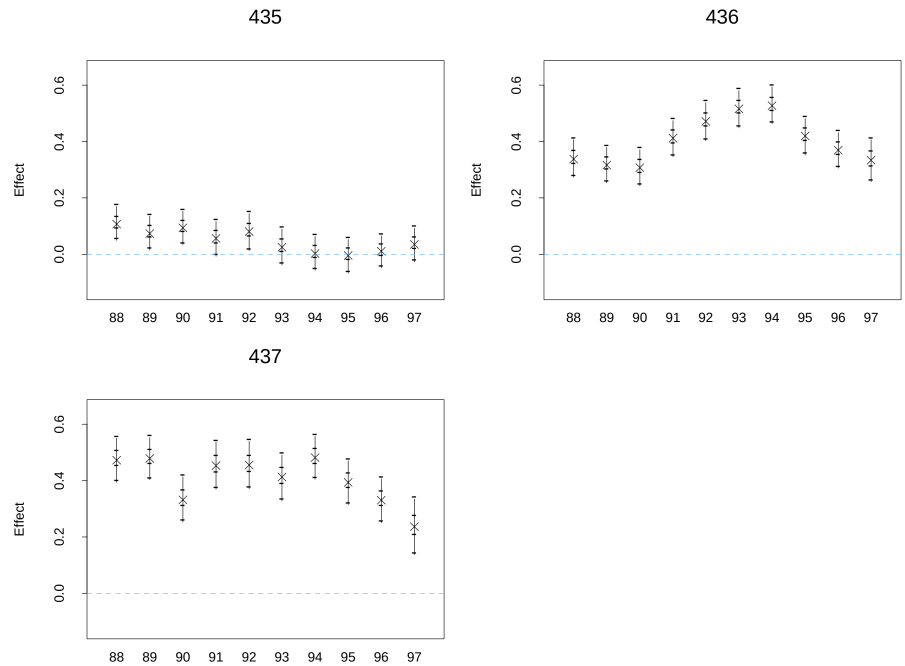

Figure 7.3: 95% posterior intervals for *DRG* group effects $\gamma_{g,r}$ over years $r = 1, \ldots, 10$ relative to *DRG* 434.

this effect should be lessening and not increasing. Thus, the decreasing pattern may be reflecting the possibility that patients from this *Age group* are admitted on expensive drug treatments that they receive at VA prices,[2] and later seek follow-up out-patient visits in the private sector which would be reimbursable through Medicare. This explanation appears likely as the increasing prices of many psychiatric pharmaceuticals can stretch elderly budgets, whether by direct payment or insured through Medigap or Medicare+Choice plans purchased by the elderly.

Figure 7.3 provides intervals and estimates of the fixed effects $\gamma_{d,r}$ for *DRG*. As can be seen from the graph, *DRG 436* and *437* have positive effects and meaningful year-to-year variations in the *DRG* effects within each category. There is also an apparently persistent deterioration in return probabilities in *DRG 436* and *437* over 1995-1997. Actually, according to the VA, this may be more a result of relative

[2]Either free to the veteran or with a nominal co-pay depending on their eligibility class.

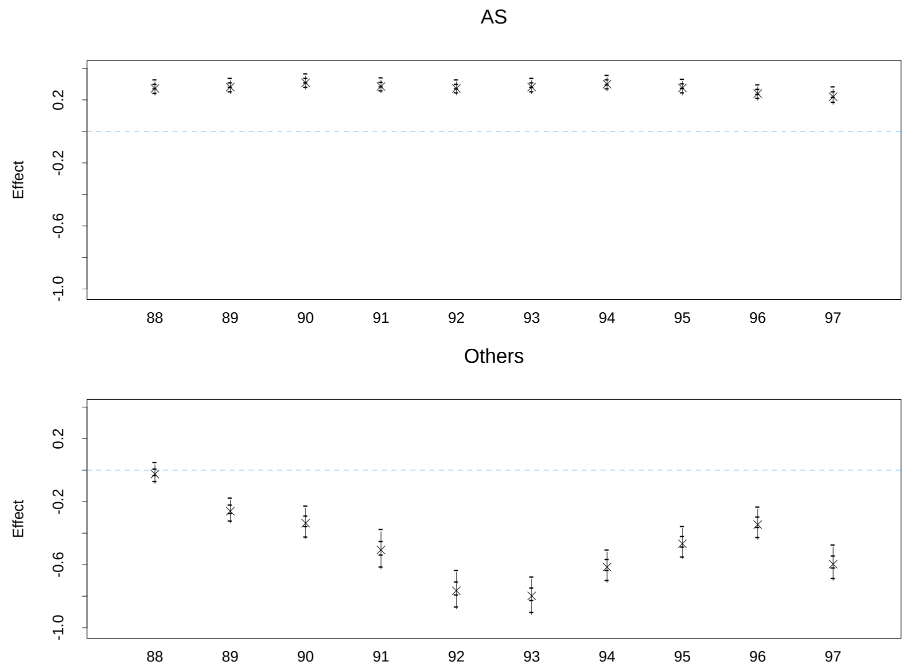

Figure 7.4: 95% posterior intervals for Priority group effects $\eta_{e,r}$ over years $r = 1, \ldots, 10$ relative to AN (poor veterans).

improvements in *DRG 434*, the reference group for this covariate, as well as *DRG 435* rather than a deterioration in *DRG 436* and *437*.

Plots in Figure 7.4 show posterior intervals of the fixed effects $\eta_{e,r}$ for the *Priority* variable. Priority eligibility for using VA health-care services is obtained either from having service connected injuries (AS) or from having income and wealth below a means test threshold (AN). The plots suggest that the effects in *Priority* group AS are accurately estimated, very stable and positive over the years, indicating a significant and sustained relative level of 30-day return probability. By contrast, the effect in *Priority* group *Others* is quite variable and generally negative; there is an indication of improvement in later years following deterioration during 1989-92/3, but a sharp drop-off in 1997. This corresponds with national budget policies to balance the budget that mandated a constant nominal budget at $17 billion from 1996 through 2002 and started to put pressure on utilization in the non-priority category.

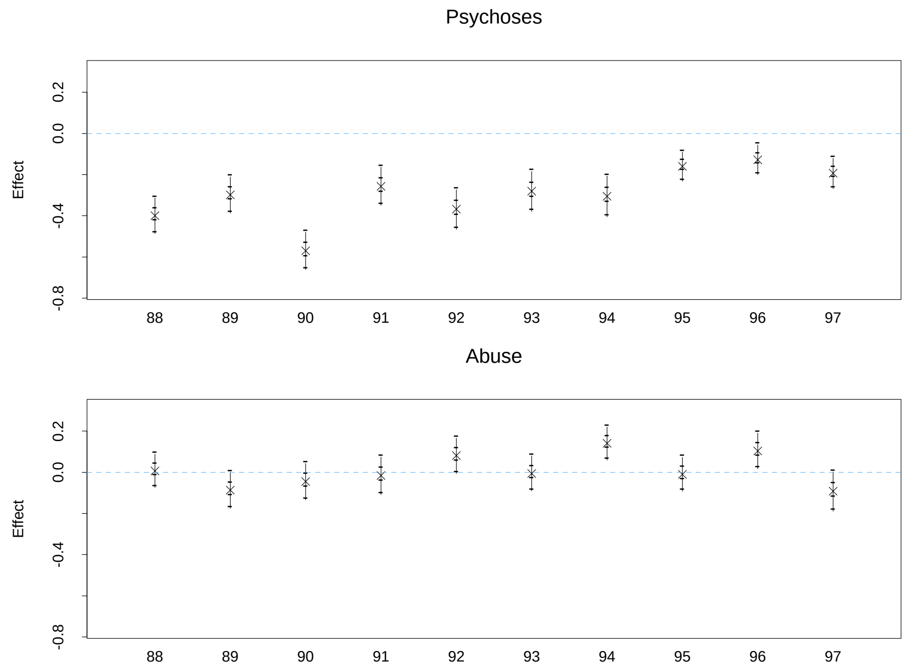

Figure 7.5: 95% posterior intervals for *Diagnosis* group effects $\xi_{x,r}$ over years $r = 1, \ldots, 10$ relative to *Dependence* category.

The categorization for *Diagnosis* groups, described in Section 6.2, are very broad in order to obtain sample sizes large enough to recognize effects. Figure 7.5 displays posterior intervals of the fixed effects $\xi_{j,r}$ for *Diagnosis* categories: *Psychoses* and *Abuse* relative to *Dependence*. The effect in the *Psychoses* group is generally negative and consistent with lower return probabilities than in either the *Dependence* or *Abuse* groups. The *Psychosis* patients are most likely to have psychological impairments that would make follow-up care more difficult.

Figure 7.6 shows posterior intervals of the fixed effects $\chi_{2,r}$ for the *Female* category relative to the *Male* effect. Effects are always positive over the full ten year span of the data. However, the levels are somewhat lower for FY92-93 and FY97; this requires further consideration and interpretation from VA personnel.

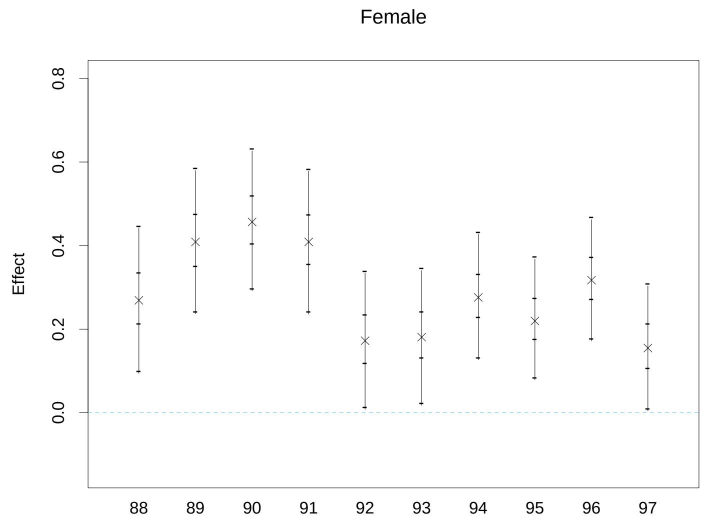

Figure 7.6: 95% posterior intervals for *Gender* effects $\chi_{2,r}$ over years $r = 1, \ldots, 10$, *Female* category relative to *Male*.

Random Effects

Figure 7.7 presents posterior intervals of the random effects $\epsilon_{j,r}$ for five selected hospitals; these are the four hospitals selected in the FY97 analysis above, plus a fifth, hospital number #139.

These hospital effects display diverse patterns over the years. Hospitals # 135 and # 115 have apparently positive effects across all ten years, consistent with higher return probabilities than the norm. Hospital # 1 is well below the norm across the board with no evidence of improvement in recent years. Hospital # 126 has tended to vary mildly about the system-wide norm, but has seen a marked increase in return probability in 1997 implied by the large and positive effect in that year. The additional hospital, number 139, has experienced return probabilities much lower than the norm during the first nine years, but has seen a very marked increase in 1997, with a magnitude of improvement that exceeds that of hospital # 126. This is an atypical behavior for a hospital, since it is not expected that effects will vary so wildly year-to-year, and further investigation of circumstances at this hospital in

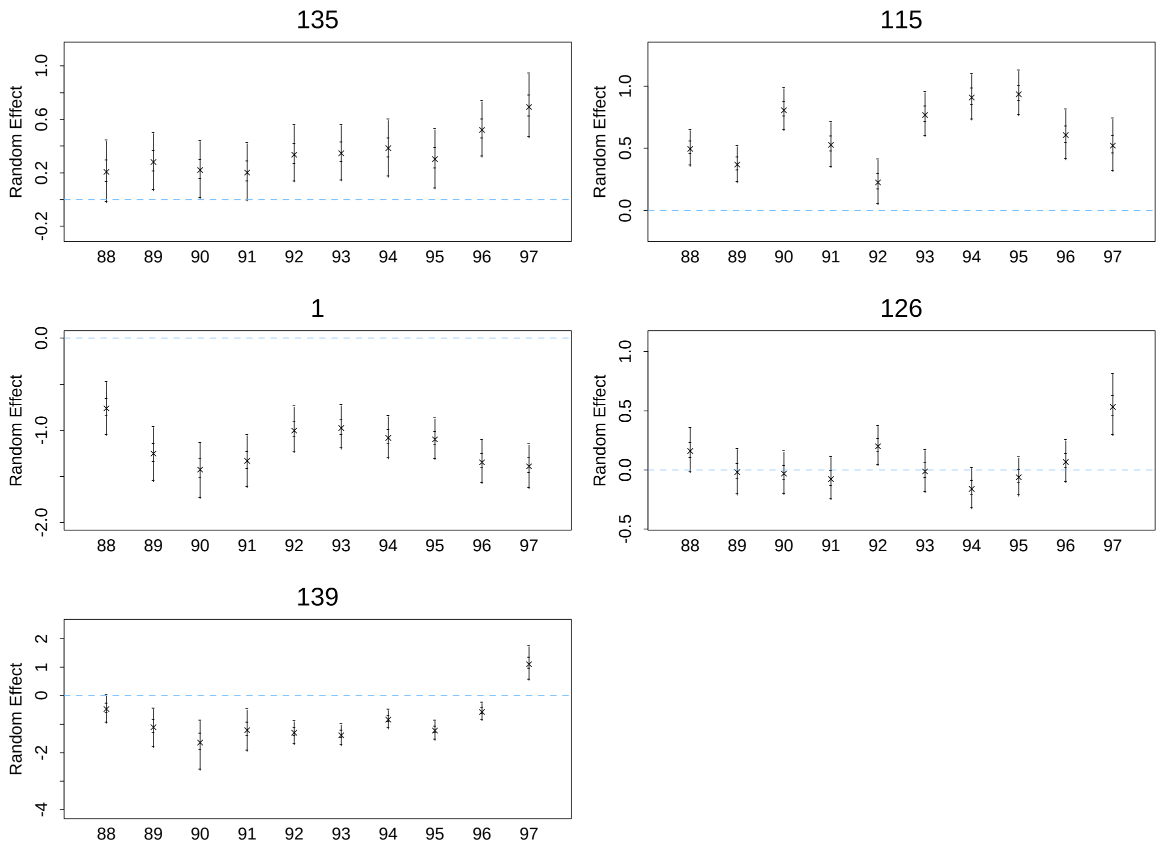

Figure 7.7: 95% posterior intervals for hospital-specific random effects $\epsilon_{j,r}$ over years $r = 1, \ldots, 10$ in multi-year analysis with cut-off $t = 30$ days. Intervals are given for the four hospitals selected in the earlier, single year analysis, as labeled, plus one additional hospital (#139).

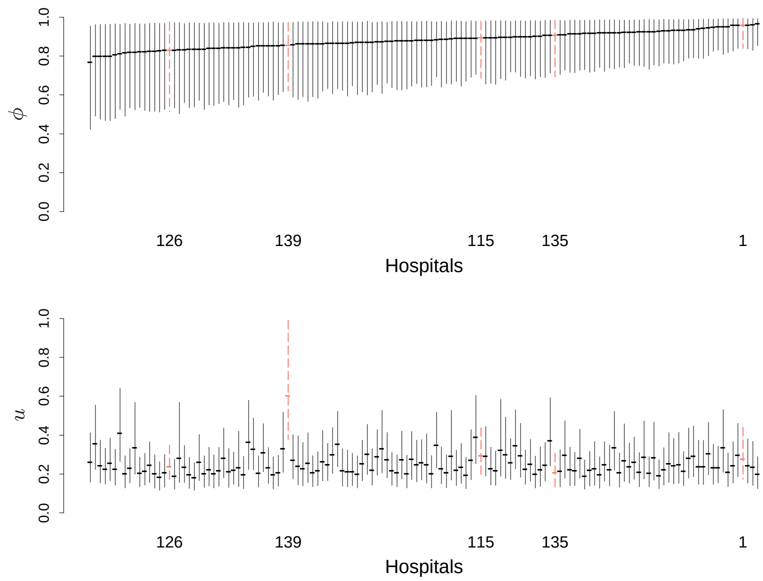

Figure 7.8: 95% posterior intervals for the hospital-specific dependence parameters ϕ_j (top) and intervals for the hospital-specific innovation variance u_j (bottom) for all 136 hospitals. Hospitals ordered by posterior median of ϕ_j. The five hospitals selected earlier are highlighted.

1996-1997 may prove informative.

Persistence parameters ϕ_j and associated hyperparameters

Parameter μ represents a system-wide average value for year-to-year correlations between hospital effects within hospitals. The approximate posterior median and end points of a 95% interval for μ are 0.82-0.85-0.88, indicating a high correlation structure generally across hospitals. The precision hyperparameter a measures the dispersion of the actual, hospital-specific ϕ_j values about μ. A high value indicates the ϕ_j are very tightly distributed around μ, lower values indicate more variability. The approximate posterior median and end-points of a 95% interval for a are 6.67-10.62-16.68; this range of fairly high values indicates that there is actually rather little variability in the ϕ_j parameters across hospitals. This is confirmed in the top panel of Figure 7.8 where posterior estimates and intervals for the ϕ_j are displayed for all 136 hospitals. Hospitals are ordered according to increasing values of the posterior medians of the ϕ_j. The plot shows generally large values for the persistence parameters, implying that a hospital that is been generally "good" ("bad") in one year will have a high probability of remaining "good" ("bad") the next year. On the other hand, the concordance illustrated in the graph suggests that the model could be reduced to a common value of $\phi_j = \phi$ for all j. A reanalysis of the data was done under this constraint and, as expected, the common parameter ϕ is close to μ and the inferences on all other model parameters and effects are basically unchanged.

Variability parameters u_j and associated hyperparameters

Parameter ρ represents a system wide average value for innovation variance; the approximate posterior median and end points of a 95% interval for ρ are 0.04-0.05-0.06. Parameter c measures the dispersion of the actual, hospital-specific innovation variances u_j^2 values about ρ. A high value of c would indicate that the u_j^2 take similar values, whereas lower values of c indicate more variation in the u_j^2 across hospitals. The approximate posterior median and end-points of a 95% interval for c are 2.06-3.07-4.73. This range of rather low values indicates that there is a fair degree of heterogeneity in the actual set of 136 u_j^2 quantities across hospitals. This is confirmed in the bottom panel of Figure 7.8, where posterior estimates and intervals for the standard deviations u_j are displayed. The plot shows some variation in both estimates and uncertainties across the hospitals. The five hospitals noted earlier are highlighted, hospital # 139 clearly stands out as a case of relatively high variability, consistent with the earlier discussion about the significantly varying $\epsilon_{j,r}$ parameters in this hospital.

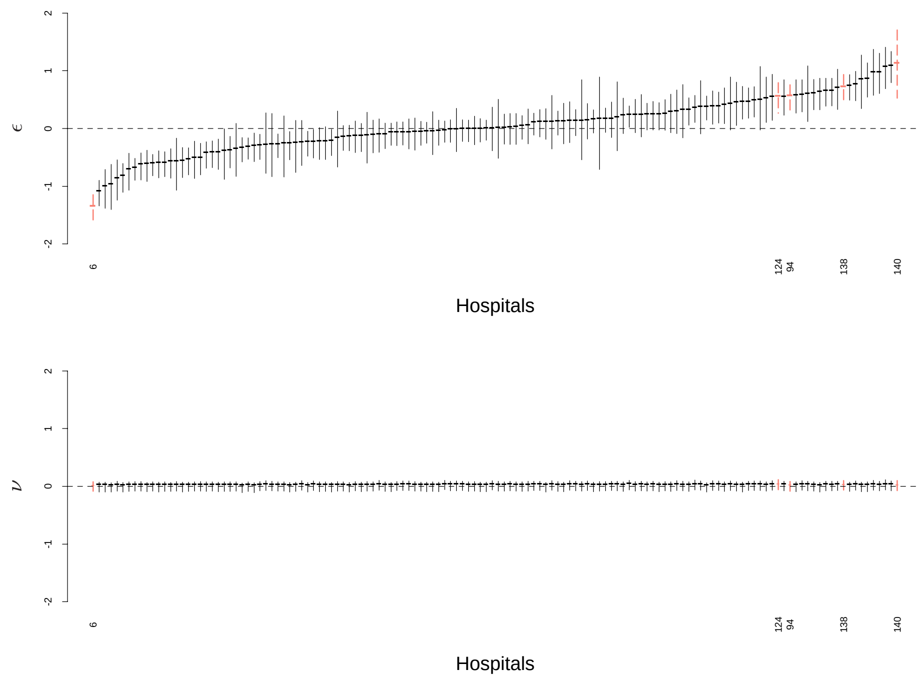

Figure 7.9: 95% posterior intervals for the hospital-specific random effects ϵ_j (top) and hospital-specific residual terms ν_j (bottom) of all 136 hospitals for FY97. The hospitals are ordered by posterior medians of the random effects.

7.2 Extensions

In this section, extensions of model (7.1) are proposed and discussed. First, the addition of residual terms to model extra-binomial variation not explained by either the covariates or the correlated hospital-specific random effects is considered. Second, the analysis of model (7.1) under different cut-off values $t = 1, 7, 14, 21, 30$ and 367 is discussed.

7.2.1 Model with Additional Residual Terms

The AR(1) structure from the random effects $\epsilon_{j,r}$ of model (7.1) may not fully account for the levels of overall extra-binomial variation apparent in the data, as discussed in West and Aguilar (1997), Aguilar and West (1998) and West *et al.* (1998) for aggregated data. To deal with this and to generalize previous work, residual or

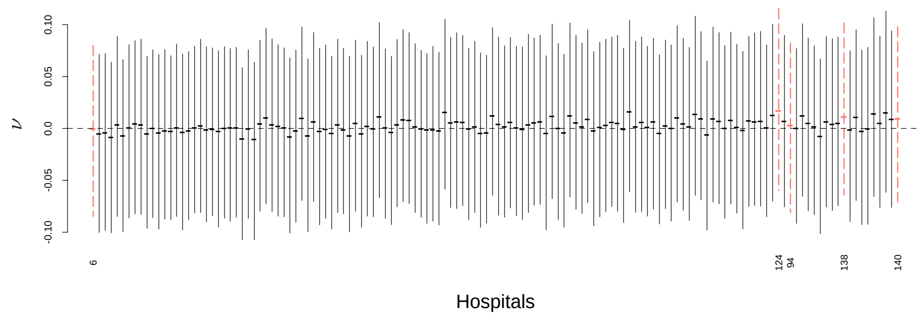

Figure 7.10: 95% posterior intervals for the hospital-specific residual terms ν_j of all 136 hospitals for FY97. The hospitals are ordered by posterior medians of the random effects. The five hospitals selected earlier are highlighted.

"idiosyncratic" random effects are included in the model to account for possible residual variation not yet explained by the explanatory variables and hospital-specific random effects. Namely, extending model (7.1),

$$\operatorname{logit}(p_r) = \beta_{0,r} + \epsilon_{j,r} + \nu_{j,r} + \begin{cases} \delta_{d,r}, & \text{for } \textit{Age group } d = 2, \\ \gamma_{g,r}, & \text{for } \textit{DRG level } g = 2,3,4, \\ \eta_{e,r}, & \text{for } \textit{Priority level } e = 2,3, \\ \xi_{x,r}, & \text{for } \textit{Diagnosis group } x = 2,3, \\ \chi_{c,r}, & \text{for } \textit{Gender group } c = 2 \ (\textit{women}) \end{cases} \tag{7.2}$$

where the new, residual terms are

$$\nu_{j,r} \sim N(\nu_{j,r}|0, v^2)$$

for $j = 1, ..., 136$ and $r = 1, ..., 10$.

The addition of the residual terms to the AR(1) process above would modify the original correlation structure in $\epsilon_{j,r}$ over time. Actually, the combined hospital-specific random effects $\epsilon_{j,r} + \nu_{j,r}$ represent a time series with autoregressive and moving-average structure: ARMA(1,1).

Figures 7.9 and 7.10 display 95% posterior intervals for hospital-specific random effects $\epsilon_{j,r}$ and hospital-specific residual terms $\nu_{j,r}$ for FY97 under model (7.2). It is clear that the residual terms are essentially negligible compared to the levels of variation in the random effects for FY97. This is supported by Figure 7.11 which shows

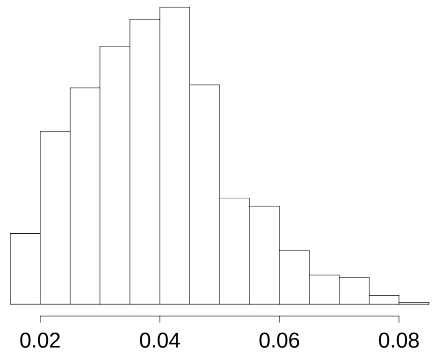

Figure 7.11: Histogram of a sample from the posterior distribution of υ, the standard deviation of the hospital-specific residual/noise terms $\nu_{j,r}$.

the histogram of a sample from the posterior distribution of the small standard deviation υ of the "unpredictable" components of variation in the random effects model. Similar behavior is present across the other years. As it turns out, the extended model (7.2) produces similar posterior distributions for the rest of the parameters compared to earlier models. Hence, it can be concluded that the random effects and the explanatory variables adequately explained most of the extra-binomial variation, and therefore no additional terms are necessary.

7.2.2 Model Estimation for More Cut-off Values for Return-Times

The analysis of Section 7.1.3 focused on a standard return-time cut-off $t = 30$. In this section, similar analyses are presented for model (7.1) with different cut-off values and using all the data. This is interesting because it will help to explore potential differences among cut-off values and offer valuable information to the VA management in terms of the effectiveness of a certain policy over different return-times. Figures 7.12 to 7.16 inclusive include similar displays to those presented in previous sections: 95% posterior intervals for the hyperparameters and random and fixed effects. In addition, summaries of the implied differences relative to the baseline on a probability scale are displayed in this section. The differences in return probabilities relative to

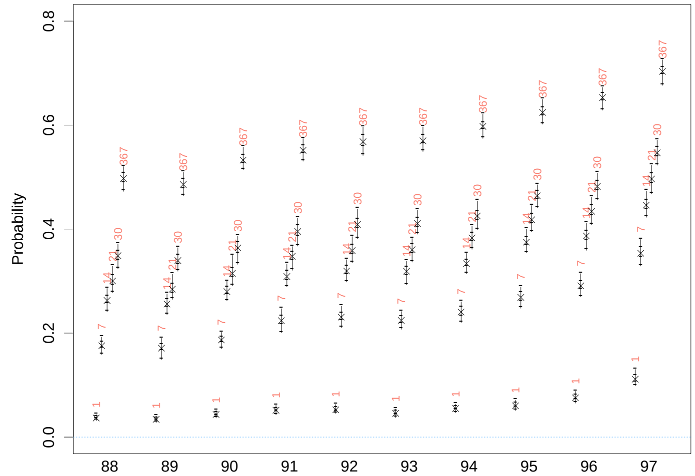

Figure 7.12: 95% posterior intervals for baseline return-time probabilities $p_{0,r}(t)$ over years $r = 1, \ldots, 10$ and cut-offs $t = 1, 7, 14, 21, 30, 367$.

baseline are defined as $p_{j,r}(t) - p_{0,r}(t)$ for each covariate level j, year r and cut-off t. As previously stated, $p_{0,r}(t) = 1/(1 + \exp(-\beta_{0,r}(t))$ is the baseline return probability; moving to level j of any one of the covariates shifts this to $p_{j,r}(t) = 1/(1+\exp(-\beta_{0,r}(t)-\tau_{j,r}(t)))$ where $\tau_{j,r}(t)$ represents the corresponding effect for that level of the chosen covariate (e.g., $\tau_{j,r}(t) = \delta_{j,r}(t)$ for *Age group* j, $\tau_{j,r}(t) = \xi_{j,r}(t)$ for *Diagnosis* group j, and so forth).

Baseline Duration Model Parameters

Figure 7.12 shows posterior intervals and estimates for the baseline return-time probabilities $p_{0,r}(t) = 1/(1 + \exp(-\beta_{0,r}(t)))$ for $r = 1, \ldots, 10$ and return-time cut-offs $t = 1, 7, 14, 21, 30$ and 367. The graph reveals a general increasing trend over time and also an increasing pattern of $p_{0,r}(t)$ as a function of t, as expected from previous analyses above.

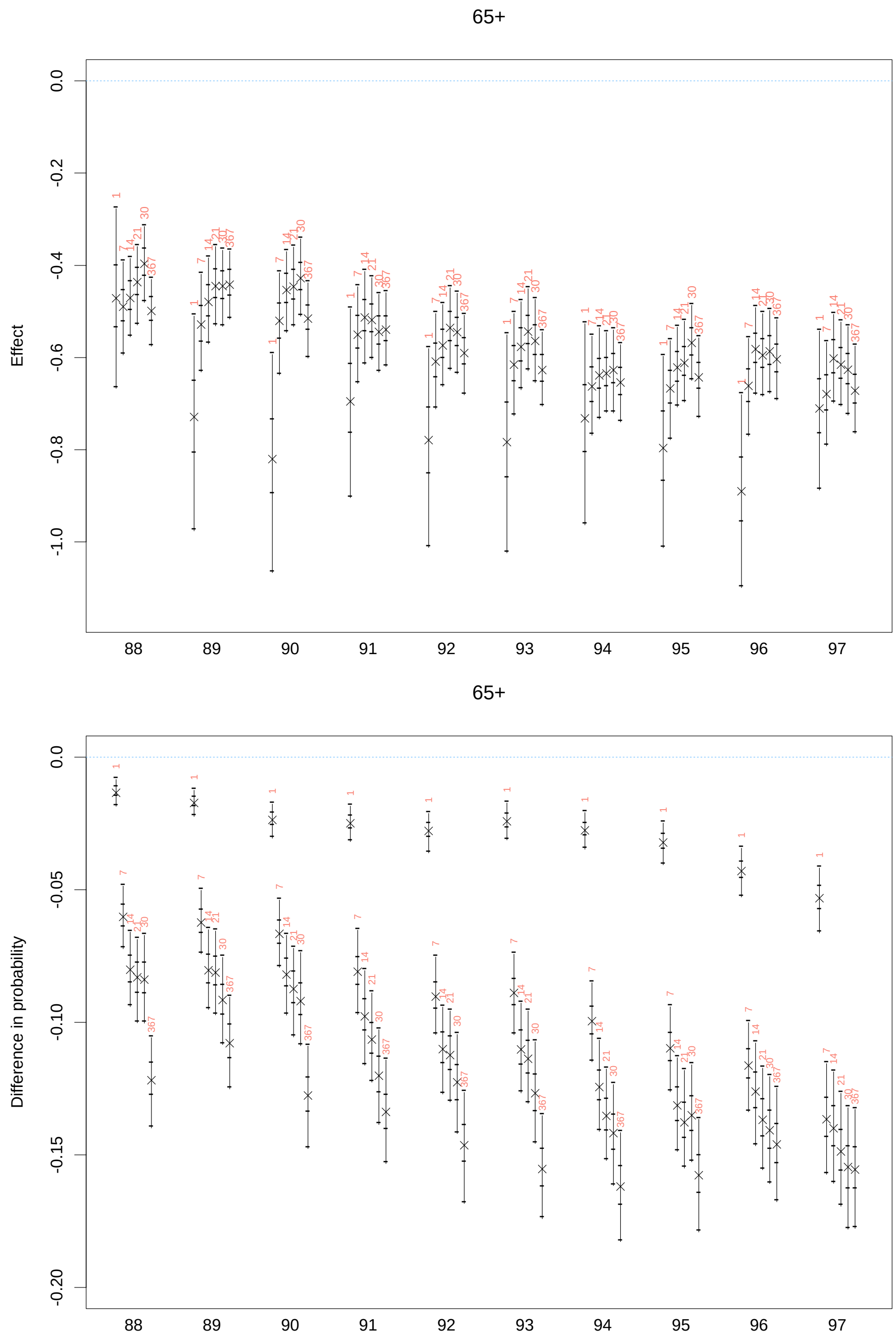

Figure 7.13: 95% posterior intervals for *Age group* effects $\delta_{2,r}(t)$ (upper frame) and the differences in implied return-time probabilities relative to the baseline (lower frame).

Analysis of Covariates

Figure 7.13 shows intervals and estimates of the fixed effects $\delta_{2,r}(t)$ for the *Age* covariate. The lower frame displays intervals for the implied differences in return-time probabilities relative to baseline (i.e., for $p_{2,r}(t) - p_{0,r}(t)$ where $p_{2,r}(t) = 1/(1 + \exp(-\beta_{0,r}(t) - \delta_{2,r}(t)))$ for each t and r). The top graph suggests evidence of non-proportional odds behavior for: FY89, FY90 and FY96, where clearly the intervals with return-time cut-off $t = 1$ have a different level than the rest. The bottom graph shows that the difference in probability of return in exactly one day for the elderly is closer to the baseline probability than any other cut-off.

DRG and *Priority* effect estimates (not displayed) follow expected patterns given the previous analyses above. On the other hand *Diagnosis* presents little evidence of non-proportional odds behavior and for different cut-offs, the effects of this covariates follow the same patterns as with cut-off $t = 30$.

Figure 7.14 shows intervals for the fixed effects $\chi_{2,r}(t)$ of the *Gender* covariate for different cut-off values. The lower frame displays similar intervals for the implied differences in return-time probabilities relative to baseline (i.e., for $p_{2,r}(t) - p_{0,r}(t)$ where $p_{2,r}(t) = 1/(1 + \exp(-\beta_{0,r}(t) - \chi_{2,r}(t)))$ for each t and r).

Although there was the suspicion of a generally increased effect with longer return-times for FY97, the upper frame Figure 7.14 shows that for FY89 the posterior medians observe a decreasing trend not present for FY92-93, where there is no apparent trend.

Random Effects for one Hospital

Figure 7.15 displays 95% intervals of the random effects for hospital #115 (randomly selected). These plots illustrate posterior estimates of hospital-specific random effects over time for different cut-offs. Intervals in the upper frame are those of the random

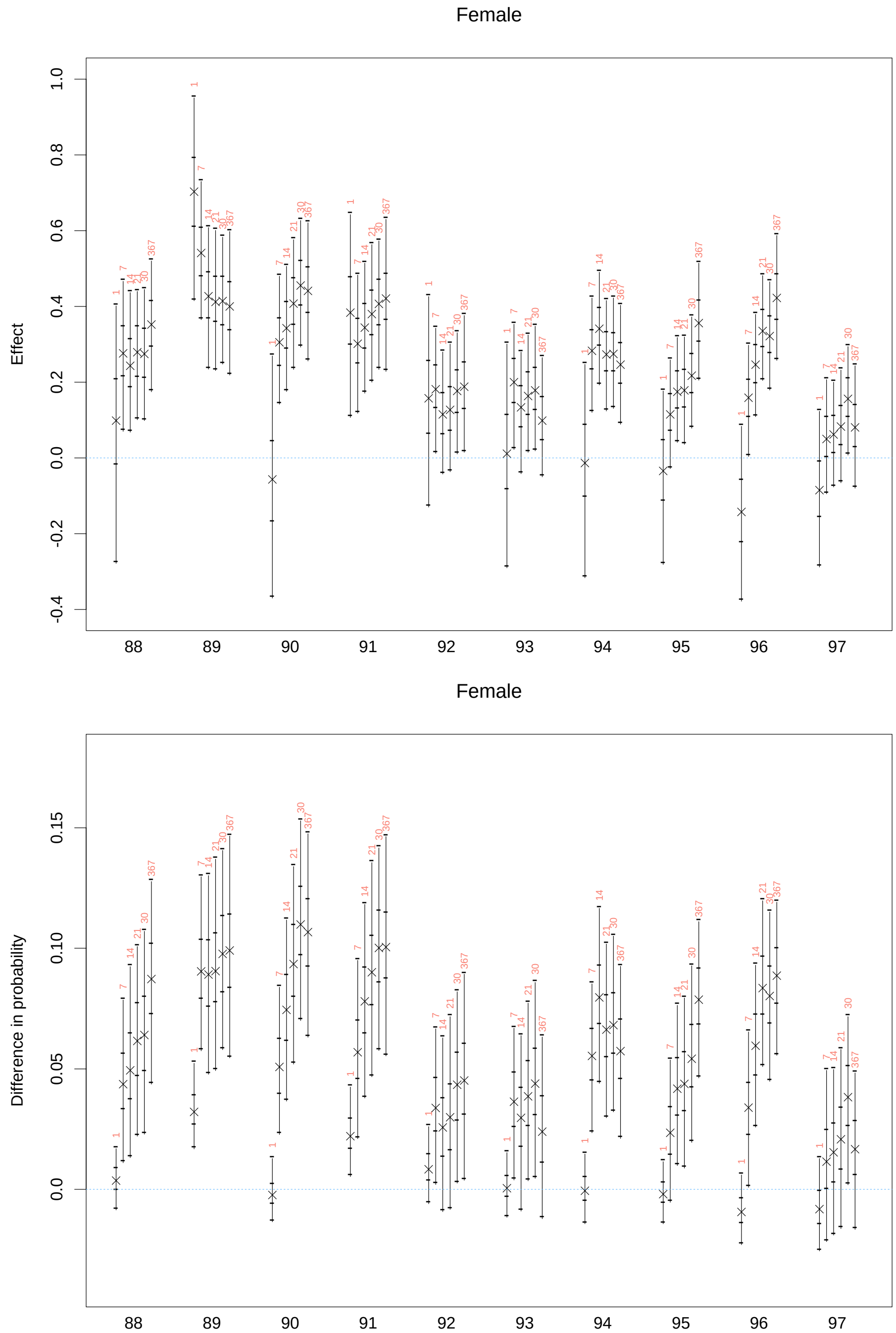

Figure 7.14: 95% posterior intervals for *Gender* effects $\chi_{2,r}(t)$ (upper frame) and the differences in implied return-time probabilities relative to the baseline (lower frame).

effects $\epsilon_{j,r}(t)$ for $j = 115$. The lower frame displays similar intervals for the implied differences in return time probabilities relative to baseline (i.e., for $p_{j,r}(t) - p_{0,r}(t)$ where $p_{j,r}(t) = 1/(1 + \exp(-\beta_{0,r}(t) - \epsilon_{j,r}(t)))$ for $j = 115$ and for each r and t). Although there is very clear evidence that the effects represent non-proportional odds structure for this hospital, the top graph in Figure 7.15 shows that the pattern among cut-offs is just slightly different every year except for FY93-95. On one hand, for FY96-97 the intervals overlap and differences are unclear, on the other hand, for FY94 the parameter $\epsilon_{115,r}(1)$ is obviously lower than the rest indicating that the probability of return in exactly one day for this hospital is approximately the same as the average hospital. Moreover, the bottom graph in Figure 7.15 shows that for FY93-95, the probabilities $p_{115,r}(1)$ are the lowest over the ten years and $p_{115,r}(t)$ are among the highest for the rest of the cut-offs t. That is, there are three years where hospital # 115 exceeds the baseline probability by approximately 20% for all cut-offs except for $t = 1$ where, on the contrary, the effects on the probability scale are the same as the baseline probability.

Hyperparameters

Figure 7.16 displays approximately 95% posterior intervals of the posterior distribution of the hyperparameters associated with the autocorrelation parameter $\phi_j(t)$ (top frames) and the hospital-specific innovation variances $u_j(t)$ (lower frames). In general, intervals for the hyperparameters $\mu(t)$, $a(t)$ and $c(t)$ overlap over all return-time cut-offs, leading to similar conclusions to those discussed in Section 7.1.3. However, intervals for $\rho(t)$, the system-wide average value for innovation variance, show a clear decreasing trend with cut-off. This implies that the innovations have higher variance under return-time probability cut-off $t = 1$ and in consequence the random effects will have higher variance than with the rest of the cut-offs as observed earlier in

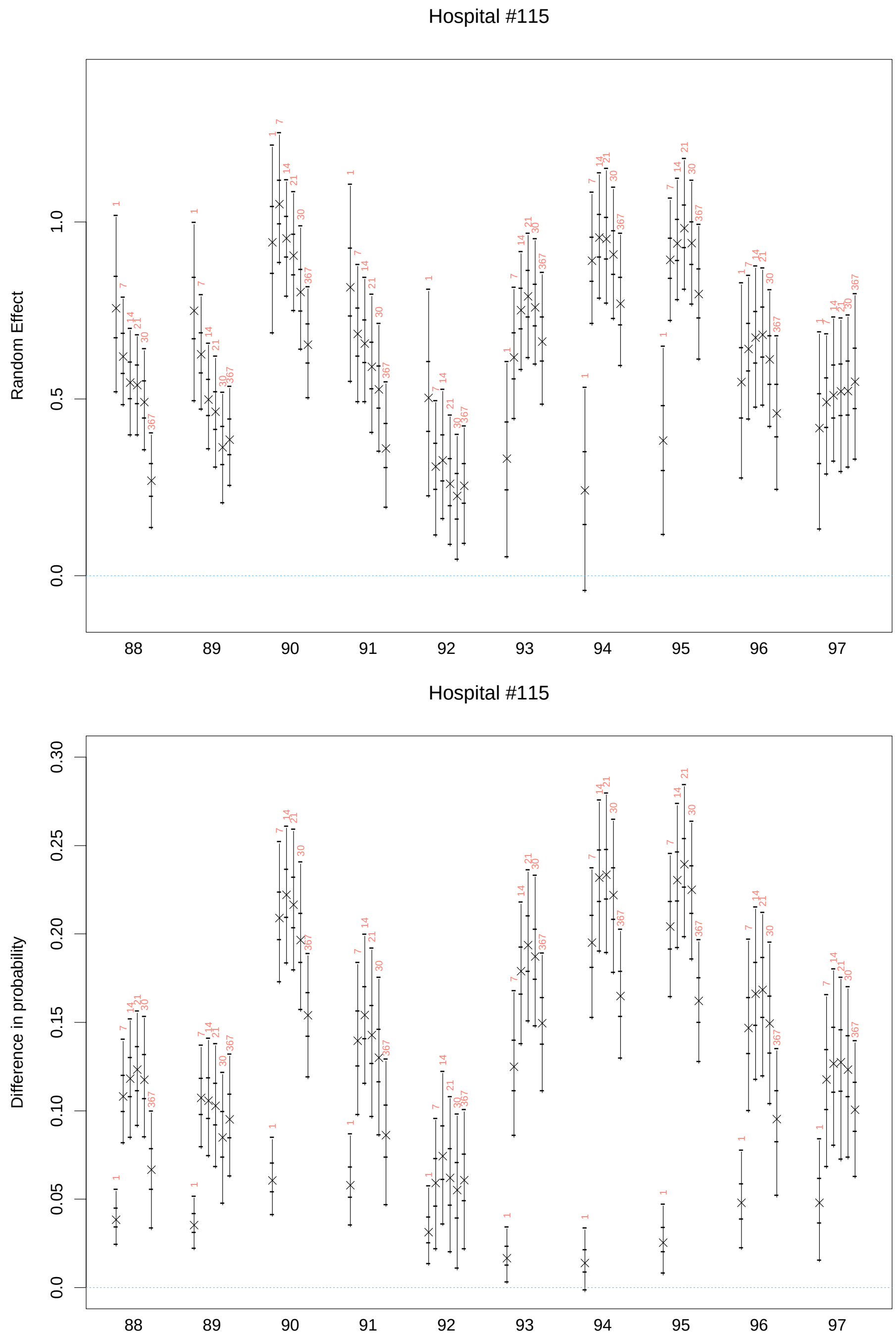

Figure 7.15: 95% posterior intervals for hospital-specific random effects $\epsilon_{j,r}(t)$ for hospital j =115. Graphs display the effects $\epsilon_{115,r}(t)$ (upper frame) and the differences in implied return-time probabilities relative to the baseline (lower frame).

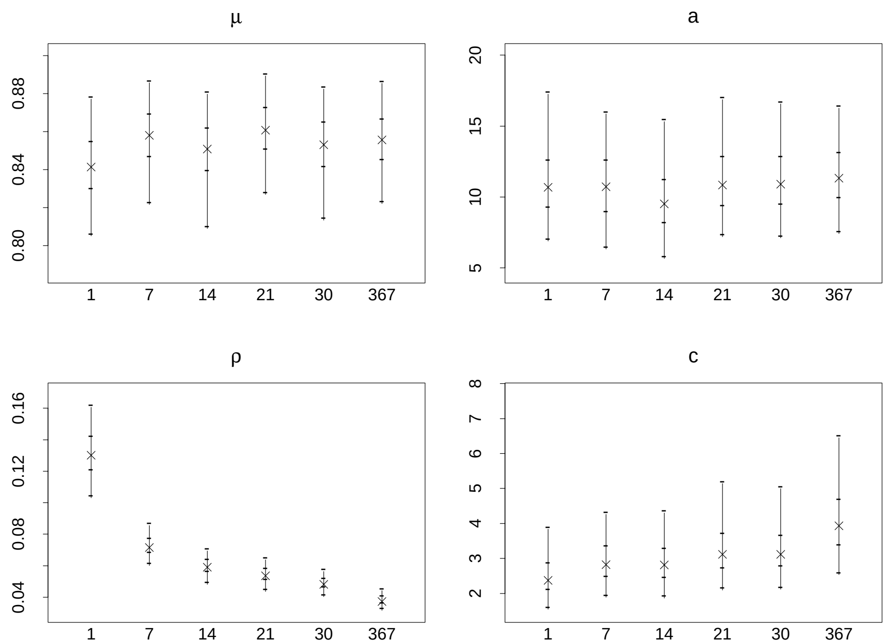

Figure 7.16: 95% posterior intervals for model hyperparameters for different return-time cut-offs $t = 1, 7, 14, 21, 30, 367$ days.

Figure 6.11.

7.3 Summary

In the past two chapters, complex non-Gaussian hierarchical models are developed to profile VA facilities with the final goals of interpreting patterns of variability across hospitals and years, and usefully summarizing the complex data for possible use by VA policy makers concerned with improving quality of care and efficient and appropriate budgetary decisions. The models presented, first for a single year of data and then extended to include a time series structure, help to quantify differences in performance in *out-patient substance abuse* programs at hospital level across the system. Important issues arise from the discussion above and can be summarized as follows:

- The diversity of inferred behavior of the hospital-specific random effects over the years is of use in providing insights to managers and policy analysts in profiling providers.

- There are evident changes in the baseline parameter over the years that require consideration and interpretation by VA personnel.

- The time series structure imposed in the model at the hospital level corroborated the consistency and persistence of hospital effects over time. However, in this case, the hospital autoregressive parameters can be reduced to a common value $\phi_j = \phi$ yielding to a more parsimonious representation.

- The autoregressive structure of the random effects and the covariates explain most of the extra-binomial variation observed in the data, suggesting that no additional structure remains unmodeled. This conclusion differs radically from previous work for aggregated data in West and Aguilar (1997), Aguilar and West (1998) and West *et al.* (1998). They discussed the decomposition of the hospital-specific random effects into two sources of variability, one coming from the autoregressive structure and other from a very significant unpredictable component. Evidently, this new analysis confirms that the chosen categorical covariates, and model form, very adequately describe patterns of variability in this large and complex data set at the highly disaggregated, individual patient level.

- Different cut-off values were used to analyze the data including the industrial standard threshold of 30 days. Analyses show different patterns of variability and structure among thresholds over time. This material require special consideration from the VA Management Science Group to assess different health-care policies beyond the industrial standards.

Chapter 8

Summary and Future Directions

This dissertation develops within a Bayesian framework, the structure and analysis of realistic and complex mathematical models arising in the social and economic sciences. Important results and future directions are outlined in this chapter.

8.1 Multivariate DLMs for Futures Prices

In the first part of the dissertation a new class of dynamic linear models was developed to explore, understand, estimate and eventually predict latent processes that are directly connected to crude oil futures pricing. These models are based on theoretical foundations of continuous time models introduced recently in the financial literature (Schwartz and Smith, 2000).

8.1.1 General Results

- A class of Bayesian multivariate dynamic linear models for oil future prices is developed based on a theoretical financial model that assumes two latent factor processes: a notional equilibrium price level and a process representing short-term deviations from equilibrium levels. Novel and customized MCMC sampling algorithms are developed to sample from the joint posterior distribution of this new class of complicated models.

- The observational variance matrix plays a central role in the estimation of the two latent processes inherent in commodity futures pricing. Different structures for the observational variance were proposed and analyzed starting with traditional assumptions of uncorrelated full rank covariance matrices and potential spatial structures. In all cases, results indicated singular or nearly singular observation variance matrices which in turn affected the convergence rates of the MCMC algorithm.

- A factor model decomposition of the observational variance was implemented to split the variability into two main components: a common source for all series and idiosyncratic terms. As it turns out, two common factors were identified and related to short term and longer term maturities, respectively. Moreover, these two common factors explain more than 80% of the total variability across the board. In addition, in some cases, informative priors have to be used to avoid having zero idiosyncratic variances for some maturities.

- General definitions of normal and Wishart distributions are presented, which include a singular variance for the normal density and a singular scale matrix for the Wishart.

- Filtering algorithms and updating mechanisms are extended and outlined for the general dynamic linear model when the observational variance matrix is singular or nearly singular. The key modification is to consider generalized inverse algorithms where needed.

- The Gibbs/Metropolis-Hastings algorithm was implemented for the so-called modified discrete time model assuming a fixed value of four for the rank of the observational variance matrix. In general, convergence rates were reasonable and estimation yield to nice interpretations of model parameters consistent with financial theory and earlier studies in this area.

- Volatile periods commonly observed in commodity pricing are difficult to capture without "heavier than normal tails" assumptions. Outlying residuals are observed especially in the Gulf War period in the series of oil futures prices.

- Results under singular distribution assumptions, and with a factor decomposition of the observational variance, yield similar conclusions for most model parameters.

- Out-of-sample forecast distributions were computed and compared to actual observed values yielding promising improvements in short term forecasting which could eventually represent a useful contribution in financial applications.

8.1.2 Future Directions

One of the main extensions of the oil futures prices model is the addition of potential stochastic volatility structure suggested from factor model decomposition and singular matrix assumptions.

First, multivariate stochastic volatility models could be implemented on the common factors variance along the lines of Aguilar (1998) and Aguilar and West (2000). The time varying covariance matrix $\mathbf{V}_t$ will then be defined by $\mathbf{V}_t = \mathbf{X}\mathbf{H}_t\mathbf{X} + \mathbf{\Psi}_t$, where $\mathbf{H}_t = \text{diag}(h_{t1}, \ldots, h_{tk})$, a diagonal matrix of instantaneous factor variances, $\mathbf{f}_t \sim \mathbf{N}(\mathbf{f}_t|\mathbf{0}, \mathbf{H}_t)$ are conditionally independent, and $\mathbf{\Psi}_t$ is a diagonal matrix of possibly time-varying idiosyncratic variances. The idea is to adopt a stationary vector autoregression model of order one for the log volatilities of the factors, centered around a mean $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_k)$. That is, for $t = 1, 2, \ldots$, $\boldsymbol{\lambda}_t = \boldsymbol{\alpha} + \mathbf{\Phi}(\boldsymbol{\lambda}_{t-1} - \boldsymbol{\alpha}) + \boldsymbol{\eta}_t$

where $\lambda_{ti} = \log(h_{ti})$ for each $i = 1, \ldots, k$, $\boldsymbol{\lambda}_t = (\lambda_{t1}, \ldots, \lambda_{tk})'$, $\boldsymbol{\Phi} = \mathrm{diag}(\phi_1, \ldots, \phi_k)$ a matrix of individual AR parameters and $\boldsymbol{\eta}_t \sim \mathbf{N}(\boldsymbol{\eta}_t|\mathbf{0}, \mathbf{U})$ are independent innovations for some innovations variance matrix $\mathbf{U}$. In the same fashion, a multivariate stochastic volatility model could be implemented on the observational variance under the singular matrix assumption, for example by allowing some of the idiosyncratic variances in the factor model to be exactly zero.

Second, unrelated univariate stochastic volatility models could be adopted for the idiosyncratic variances as in Aguilar and West (2000) and Pitt and Shephard (1999). The idea is to assume an autoregression model of order one on the log of the idiosyncratic variances. That is, $\zeta_{tj} = \delta_j + \phi_j(\zeta_{t-1,j} - \delta_j) + \eta_{tj}$ for $j = 1, \ldots, r$, where $\zeta_{tj} = \log(\psi_{tj})$ are mutually independent series; $\eta_{tj} \sim N(\eta_{tj}|0, s_j)$ and $0 < \phi_j < 1$ for each j. It is important to emphasize that the inclusion of stochastic volatility structure on the mentioned variances would potentially improve the forecast performance of the model, as has been the experience with multivariate stochastic volatility models in portfolio applications.

Another issue under current development is to extend the model to consider a time-varying structure on the equilibrium growth rate μ_ξ, as suggested in Schwartz and Smith (2000) and also a time-varying structure on the persistence parameter ϕ as the results under the factor model suggested.

Finally, a formal assessment of rank uncertainty, using model selection methods for example, is necessary under the singular observational matrix assumption.

8.2 Hierarchical Models in Institutional Profiling

In the second part of the dissertation, complex non-Gaussian hierarchical models are developed to profile hospitals in the VA system. The main goals are to interpret patterns of variability across hospitals and years, and usefully summarize the complex and large data set for possible use by VA policy makers concerned with improving quality of care and efficient and appropriate budgetary decisions.

8.2.1 General Results

- Exploratory and confirmatory analyses of FY97 data alone help us to understand key aspects of the structure of the massive data set, especially in exploring the individual level covariates.

- A "time to return" threshold of 30 days has been commonly considered in many studies in health care profiling; actually, it is used as an industrial standard. In the analysis of a single year of data under 30 day cut-off, results are somewhat less "noisy" than under other potential cut-off times; it appears to limit the effects of unobserved heterogeneity at the individual level.

- The diversity of inferred behavior of the hospital-specific random effects over the years is of use in providing insights to managers and policy analysts in profiling providers.

- The autoregressive structure of the random effects and the covariates explain most of the extra-binomial variation observed in the ten years of data, suggesting that no additional structure remains unmodeled. The chosen categorical covariates and model form adequately describe patterns of variability in this large and complex data set at the highly disaggregated, individual patient level.

8.2.2 Future Directions

An important issue under current development is regarding model diagnostics, residual analyses and general model validation. It is of primary concern to identify and investigate any apparent residual variation in the data that could be still unmodeled.

Generally, Bayesian residual analyses[1] for a logistic regression model are based on direct examination of the posterior probability distribution of the difference between the observed proportion and the fitted proportion where the posterior distribution of the parameters determines the posterior distribution of the residuals. However, in this case, proportions are different per individual with certain attributes determined by the covariates which complicates the analysis. Another diagnostic tool to explore is the analysis of the posterior predictive distributions of observed values compared to the actual values which will provide out-of-sample or cross-validatory assessment of model adequacy. However, again, since there are many possible cells defined by the covariates, the analysis of the posterior distributions is also complex. These important issues are under further investigation and are obviously of real relevance in applications of large-scale, highly structured hierarchical models in many areas, as well as in institutional profiling.

[1] For more on Bayesian residual analysis and goodness-of-fit see, for example, Johnson and Albert (1999).

Appendix A

Some Results on Singular Variance Matrices

A.1 A Practical Issue of the Rank of a Matrix

Although the rank of a matrix is easy to define, it may be difficult to compute it in practice. For example, eigenvalues could be very small but not exactly zero, thus a tolerance error has to be settled in order to define when an eigenvalue is considered zero or not. For example, Strang (1988) suggests a threshold of 10^{-6}. For more on rank of a matrix and singular matrices see, for example, Barnett (1990), Strang (1988) and Graybill (1983).

A.2 Generalized Inverse

Generalized inverse matrix is a more general definition for an inverse matrix. Let $\boldsymbol{\Sigma}$ be an $r \times q$ matrix of arbitrary rank. A generalized inverse of $\boldsymbol{\Sigma}$ is a unique $q \times r$ matrix $\boldsymbol{\Sigma}^+$ which satisfies the following conditions:

$$\begin{aligned} \boldsymbol{\Sigma}\boldsymbol{\Sigma}^+\boldsymbol{\Sigma} &= \boldsymbol{\Sigma} \\ \boldsymbol{\Sigma}^+\boldsymbol{\Sigma}\boldsymbol{\Sigma}^+ &= \boldsymbol{\Sigma}^+ \\ (\boldsymbol{\Sigma}\boldsymbol{\Sigma}^+)' &= \boldsymbol{\Sigma}\boldsymbol{\Sigma}^+ \\ (\boldsymbol{\Sigma}^+\boldsymbol{\Sigma})' &= \boldsymbol{\Sigma}^+\boldsymbol{\Sigma} \end{aligned} \tag{A.1}$$

$\mathbf{\Sigma}^+$ is also called g-inverse, Moore-Penrose inverse or pseudo-inverse. Note that if $\mathbf{\Sigma}$ is square and non-singular, $\mathbf{\Sigma}^{-1}$ satisfies the conditions of a g-inverse. For more on generalized inverse see for example Rao and Mitra (1971), Graybill (1983) and Boullion and Odell (1971).

In particular, for symmetric matrices, the spectral decomposition can be used to compute their generalized inverses. $\mathbf{\Sigma} : r \times r$ can be factorized as $\mathbf{\Sigma} = \mathbf{HDH}'$ with the orthonormal eigenvectors in $\mathbf{H} : r \times r$ and the eigenvalues $l_1, \cdots, l_r$ in $\mathbf{D}$ =diag$(l_1, \cdots, l_r)$. If $\mathbf{\Sigma}$ is singular, there are some eigenvalues equal to zero; thus $\mathbf{\Sigma} = (\mathbf{H}_1, \mathbf{H}_2) \begin{pmatrix} \mathbf{D}_0 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \mathbf{H}_1' \\ \mathbf{H}_2' \end{pmatrix}$ where $\mathbf{H} = (\mathbf{H}_1, \mathbf{H}_2)$, $\mathbf{H}_1 : r \times \rho$ column orthonormal and $\mathbf{D}_0$ =diag$(l_1, \cdots, l_\rho)$ with $l_i > 0$ for all $i = 1, \cdots, \rho$. Therefore $\mathbf{\Sigma} = \mathbf{H}_1\mathbf{D}_0\mathbf{H}_1'$, rank$(\mathbf{\Sigma}) = \rho$, and the generalized inverse of $\mathbf{\Sigma}$ is $\mathbf{\Sigma}^+ = \mathbf{H}_1\mathbf{D}_0^{-1}\mathbf{H}_1'$ which clearly satisfies conditions (A.1). Given this decomposition, $\mathbf{\Sigma}$ and $\mathbf{\Sigma}^+$ can be rewritten as follows:

$$\begin{aligned} \mathbf{\Sigma} &= \mathbf{BB}' \\ \mathbf{\Sigma}^+ &= \mathbf{T}'\mathbf{T} \end{aligned} \tag{A.2}$$

where $\mathbf{B} = \mathbf{H}_1\mathbf{D}_0^{1/2}$ is $r \times \rho$ matrix and $\mathbf{T} = \mathbf{D}_0^{-1/2}\mathbf{H}_1'$ is $\rho \times r$. Two main points to note here, first, $\mathbf{TB} = I_\rho$, and second, that neither $\mathbf{B}$ nor $\mathbf{T}$ are unique since for example $\mathbf{\Sigma} = \bar{\mathbf{B}}\bar{\mathbf{B}}'$ where $\bar{\mathbf{B}} = \mathbf{BL}$ and $\mathbf{L}$ is any $\rho \times \rho$ orthogonal matrix.

A.3 Standardization of a Random Vector with Singular Variance

Let $\mathbf{y}$ be a random vector of dimension r with expected value zero and singular variance $\mathbf{\Sigma}$ of rank $\rho \leq r$. The standardization of $\mathbf{y}$ consists on the linear transformation $\mathbf{z} = \mathbf{Ty}$ where $\mathbf{\Sigma}^+ = \mathbf{TT}'$ implying that E$(\mathbf{z}) = \mathbf{0}$ and Var$(\mathbf{z}) = \mathbf{T\Sigma T}' = \mathbf{I}_\rho$.[1] It is important to note that the dimension of $\mathbf{z}$ is $\rho \leq r$.

[1]One way to define $\mathbf{T}$ is $\mathbf{T} = \mathbf{D}_0^{-1/2}\mathbf{H}_1'$ as is the previous section.

A.4 Conditional distribution of $\mathbf{y}_1$ given $\mathbf{y}_2$ when $\mathbf{y}_2$ is singular normally distributed

As in Muirhead (1982) (Theorem 1.2.11), let $\mathbf{y}$ be $\mathbf{N}_m(\mu, \mathbf{\Sigma})$ and partition $\mathbf{y}$, μ and $\mathbf{\Sigma}$ as: $\mathbf{y} = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix}$, $\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$, $\mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{pmatrix}$ where $\mathbf{y}_1$ and μ_1 are $k \times 1$ and $\mathbf{\Sigma}_{11}$ is $k \times k$. Let $\mathbf{\Sigma}_{22}^{+}$ be a generalized inverse of $\mathbf{\Sigma}_{22}$ and let $\mathbf{\Sigma}_{11\cdot 2} = \mathbf{\Sigma}_{11} - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{+}\mathbf{\Sigma}_{21}$. Then

(a) $\mathbf{y}_1 - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{+}\mathbf{y}_2$ is $N_k(\mu_1 - \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{+}\mu_2, \mathbf{\Sigma}_{11\cdot 2})$ and is independent of $\mathbf{y}_2$, and

(b) the conditional distribution of $\mathbf{y}_1$ given $\mathbf{y}_2$ is $\mathbf{N}_k(\mu_1 + \mathbf{\Sigma}_{12}\mathbf{\Sigma}_{22}^{+}(\mathbf{y}_2 - \mu_2), \mathbf{\Sigma}_{11\cdot 2})$.

A.5 Singular Wishart Distribution

If $\mathbf{W} \sim \text{Wishart}_r\,(n, \mathbf{\Sigma})$ $n \geq r$ and $\mathbf{\Sigma}$ is singular with rank r then $\mathbf{W}$ is distributed as a *Singular Wishart* and the density function for $\mathbf{W}$ is:

$$p(\mathbf{W}|\mathbf{\Sigma}) = \frac{1}{2^{\rho n/2}\Gamma_\rho\left(\frac{n}{2}\right)}|\mathbf{\Sigma}^{+}|^{n/2}|\mathbf{D}_0|^{(n-r-1)/2}\text{exp}\left\{-\frac{1}{2}\text{tr}\left(\mathbf{\Sigma}^{+}\mathbf{W}\right)\right\} \tag{A.3}$$

if $\left(\mathbf{I} - \mathbf{\Sigma}\mathbf{\Sigma}^{+}\right)(\mathbf{W}) = 0$ where $\mathbf{W} = \mathbf{H}_1\mathbf{D}_0\mathbf{H}_1'$, $\mathbf{D}_0 = \text{diag}(l_1, \ldots, l_\rho)$ $l_i > 0$ the ρ non-zero eigenvalues of $\mathbf{W}$, and $\mathbf{\Sigma}^{+}$ is the generalized inverse of $\mathbf{\Sigma}$ and $\text{E}(\mathbf{W}) = n\mathbf{\Sigma}$. Note that the density is concentrated over the hypersurface $\left(\mathbf{I} - \mathbf{\Sigma}\mathbf{\Sigma}^{+}\right)(\mathbf{W}) = 0$.

Derivation of the singular Wishart density for n integer.

Let $\mathbf{W} = \mathbf{Z}'\mathbf{Z}$, where $\mathbf{Z} = (\mathbf{Z}_1, \ldots, \mathbf{Z}_n)$ is $n \times \rho$ matrix, $\mathbf{z}_i \sim \mathbf{N}(\mathbf{0}, \mathbf{\Sigma})$ i.i.d with $\mathbf{\Sigma}$ $r \times r$ singular matrix. Since $\text{Var}(\mathbf{z}_i) = \mathbf{\Sigma}$ is singular, $\mathbf{z}_i$ has a singular normal distribution with (5.1) as density. Thus, the density for $\mathbf{Z}$ is given by:

$$(2\pi)^{-\rho n/2}\,|\mathbf{\Sigma}^{+}|^{n/2}\text{exp}\left\{-\frac{1}{2}\text{tr}\left(\mathbf{\Sigma}^{+}\mathbf{Z}'\mathbf{Z}\right)\right\}(d\mathbf{Z}) \tag{A.4}$$

if $(\mathbf{I}-\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{+})(\mathbf{Z}'\mathbf{Z})=0$ and zero otherwise. The Jacobian of the transformation from $\mathbf{Z}$ to $\mathbf{W}$ is:

$$(d\mathbf{Z}) = |\mathbf{D}_0|^{(n-\rho-1)/2}\frac{\pi^{\rho n/2}}{\Gamma_\rho\left(\frac{n}{2}\right)}(d\mathbf{W}) \tag{A.5}$$

which can be derived following parallel steps in Muirhead (1982), Section 3.2, for the non-singular Wishart Distribution. Replacing $\mathbf{Z}'\mathbf{Z}$ by $\mathbf{W}$ in (A.4) and substituting (A.5) in (A.4), the density (A.3) is obtained.

Bibliography

Aguilar, O. (1998) Latent structure in Bayesian multivariate time series models. Ph.D. Thesis. Institute of Statistics and Decision Sciences. Duke University.

Aguilar, O. and West, M. (1998) Analysis of hospital quality monitors using hierarchical time series models. In *Case Studies in Bayesian Statistics* (ed. C. Gatsonis et al), vol. 4, pp. 287–302. New York: Springer-Verlag.

Aguilar, O. and West, M. (2000) Bayesian dynamic factor models and portfolio allocation. *Journal of Business & Economic Statistics*; Forthcoming.

Barnett, S. (1990) *Matrices: Methods and Applications*. Oxford: Oxford University Press.

Berger, J.O. (1985) *Statistical Decision Theory and Bayesian Analysis*, Second edn. New York: Springer-Verlag.

Berger, J.O., J.K.Ghosh and Mukhopadhyay, N. (2000) Approximations to the Bayes factor in model selection problems and consistency issues. Discussion Paper # 00-10. Institute of Statistics and Decision Sciences, Duke University.

Bernardo, J.M. and Smith, A.F.M. (1994) *Bayesian Theory*. New York: John Wiley & Sons.

Best, N., Cowles, M.K. and Vines, K. (1995) Convergence diagnosis and output analysis software for gibbs sampling output (CODA). Software available at http://www.mrc-bsu.cam.ac.uk/bugs/.

Bhimasankaram, P. and Sengupta, D. (1991) Testing for the mean vector of a multivariate normal distribution with a possibly singular dispersion matrix and related results. *Statistics & Probability Letters*, **11,** 473–478.

Bodie, Z., Kane, A. and Marcus, A.J. (1999) *Investments*, Fourth edn. Burr Ridge, Illinois: Richard D. Irwin, INC.

Boullion, T.L. and Odell, P.L. (1971) *Generalized Inverse Matrices*. New York: Wiley Interscience. John Wiley & Sons.

Brooks, S.P. (1998) Markov chain Monte Carlo method and its application. *The Statistician*, **47,** 69–100.

Burgess, J.F., Christiansen, C.L., Michalak, S.E. and Morris, C.N. (2000) Medical profiling: Improving standards and risk adjustments using hierarchical models. *Journal of Health Economics*; Forthcoming.

Burgess, J.F. and DeFiore, D.A. (1994) The effect of distance to VA facilities on the choice and level of utilization of VA outpatient services. *Social Science and Medicine*, **39,** 95–103.

Carter, C.K. and Kohn, R. (1994) On Gibbs sampling for state space models. *Biometrika*, **81,** 541–553.

Chib, S. and Greenberg, E. (1995) Understanding the Metropolis-Hastings algorithm. *American Statistician*, **49,** 327–335.

Clubley, S. (1998) *Trading in Oil Futures and Options*. Cambridge: Woodhead Publishing Limited.

Cowles, M.K. and Carlin, B.P. (1996) Markov chain monte-carlo convergence diagnostics: a comparative study. *Journal of the American Statistical Association*, **91,** 883–904.

Cressie, N.A.C. (1993) *Statistics for Spatial Data*, Revised edn. New York: John Wiley & Sons, Inc.

Dadkhah, K.M. (1992) Futures market for crude oil. In *The Oil Market in the 1980s. A Decade of Decline* (eds S. Shojai and B.S. Katz), pp. 207–219. New York: Praeger.

Dawid, P. (1981) Some matrix-variate distribution theory: Notational considerations and a Bayesian application. *Biometrika*, **68,** 265–274.

Dixit, A.K. and Pindyck, R.S. (1994) *Investment Under Uncertainty*. Princeton, NJ: Princeton University Press.

Frühwirth-Schnatter, S. (1994) Data augmentation and dynamic linear models. *Journal of Time Series Analysis*, **15,** 183–202.

Gelfand, A.E. and Smith, A.F.M. (1990) Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, **85,** 398–409.

Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B. (1995) *Bayesian Data Analysis*. London: Chapman & Hall.

Geweke, J.F. and Zhou, G. (1996) Measuring the pricing error of the arbitrage pricing theory. *The Review of Financial Studies*, **9,** 557–587.

Gibson, R. and Schwartz, E.S. (1990) Stochastic convenience yield and the pricing of oil contingent claims. *The Journal of Finance*, **45,** 959–976.

Gilks, W., Richardson, S. and Spiegelhalter, D. (1996) *Markov Chain Monte Carlo in Practice*. London: Chapman and Hall.

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.

Graybill, F.A. (1983) *Matrices with Applications in Statistics*, Second edn. Belmont, California: Wadsworth International Group.

Hull, J.C. (1997) *Options, Futures, and Other Derivatives*, Third edn. New Jersey: Prentice Hall.

Hull, J.C. (1998) *Introduction to Futures and Options Markets*, Third edn. New Jersey: Prentice Hall.

Johnson, V.E. and Albert, J.H. (1999) *Statistics for Social Science and Public Policy*. New York: Springer-Verlag.

Kass, R.E. and Raftery, A.E. (1995) Bayes factors. *Journal of the American Statistical Association*, **90,** 773–795.

Kass, R.E. and Wasserman, L. (1995) A reference Bayesian test for nested hypothesis and its relationship to the Schwarz criterion. *Journal of the American Statistical Association*, **90,** 928–934.

Khatri, C. G. (1968) Some results for the singular normal multivariate regression models. *Sankhyā, Series A, Indian Journal of Statistics*, **30,** 267–280.

Lee, H.K.H. (1998) Model selection and model averaging for neural networks. Ph.D. Thesis. Carnegie Mellon University.

Lindsey, J.K. (1997) *Applying Generalized Linear Models*. New York: Springer-Verlag.

Lopes, H.F. and West, M. (1999) Model uncertainty in factor analysis. Discussion Paper # 98-38. Institute of Statistics and Decision Sciences, Duke University.

Luenberger, D.G. (1998) *Investment Science*. New York, NY: Oxford University Press.

Macfarlane, A. (1996) Discussion of League tables and their limitations: Statistical issues in comparisons of institutional performance, Goldstein H. and Spiegelhalter, D.J. *Journal of the Royal Statistical Society*, **159,** 414–415.

Mardia, K.V., Kent, J.T. and Bibby, J.M. (1979) *Multivariate Analysis*. New York: Academic Press.

Marsaglia, G. (1964) Conditional means and covariances of normal variables with singular covariance matrix. *Journal of the American Statistical Association*, **59,** 1203–1204.

McCullagh, P. (1980) Regression models for ordinal data. *Journal of the Royal Statistical Society, series* B, **42,** 109–142.

McCullagh, P. and Nelder, J.A. (1989) *Generalized Linear Models*, Second edn. Cambridge: Chapman and Hall.

McNeil, B. J. (1992) Current issues in profiling quality of care. *Inquiry*, **29,** 298–307.

Mengersen, K.L., Robert, C.P. and Guihenneuc-Jouyaux, C. (1999) MCMC convergence diagnostics: a "reviewww". In *Bayesian Statistics 6, Proceedings of the Sixth Valencia International Meeting* (eds J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith), pp. 415–440. London: Oxford University Press.

Mitra, S. K. and Rao, C. R. (1968) Some results in estimation and tests of linear hypotheses under the Gauss-Markoff model. *Sankhyā, Series A, Indian Journal of Statistics*, **30,** 281–290.

Muirhead, R.J. (1982) *Aspects of Multivariate Statistical Theory*. New York: John Wiley & Sons.

Normand, S.T., Glickman, M.E. and Gatsonis, C.A. (1997) Statistical methods for profiling providers of medical care: Issues and applications. *Journal of the American Statistical Association*, **92,** 803–814.

NYMEX (1998a) Glossary of terms. New York, NY. Brochure.

NYMEX (1998b) Light, sweet crude oil futures and options. New York, NY. Brochure.

NYMEX (1998c) New York Mercantile Exchange Energy Complex. New York, NY. Brochure.

Oktaba, W. and Kieloch, A. (1993) Wishart distributions in the multivariate Gauss-markoff model with singular covariance matrix. *Aplikace Matematiky*, **38,** 61–66.

Pitt, M.K. and Shephard, N. (1999) Time varying covariances: A factor stochastic volatility approach (with discussion). In *Bayesian Statistics 6, Proceedings of the Sixth Valencia International Meeting* (eds J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith), pp. 547–570. London: Oxford University Press.

Polasek, W. (1999) Factor analysis and outliers: A Bayesian approach. Discussion Paper. University of Basel.

Press, S.J. (1985) *Applied Multivariate Analysis: Using Bayesian and Frequentist Methods of Inference*. California: Krieger.

Radhakrishna, R. (1962) A note on a generalized inverse of a matrix with applications to problems in mathematical statistics. *Journal of the Royal Statistical Society, Series* B, **24,** 152–158.

Raftery, A.E. (1995) Bayesian model selection in social research (with discussion). In *Sociological Methodology 1995* (ed. P.V. Marsden), pp. 111–196. Cambridge: Blackwells.

Rao, C. R. (1978) Least squares theory for possibly singular models. *The Canadian Journal of Statistics*, **6,** 19–23.

Rao, C.R. and Mitra, S.K. (1971) *Generalized Inverse of Matrices and its Applications*. New York: John Wiley & Sons.

Rauscher, M. (1989) *OPEC and the Price of Petroleum*. Berlin: Springer Verlag.

Schwartz, E. and Smith, J.E. (2000) Short-term variations and long-term dynamics in commodity prices. *Management Science*; Forthcoming.

Schwartz, E.S. (1997) The stochastic behavior of commodity prices: Implications for valuation and hedging. *Journal of Finance*, **52,** 923–973.

Schwarz, G. (1978) Estimating the dimension of a model. *The Annals of Statistics*, **6,** 461–464.

Siotani, M., Hayakawa, T. and Fujikoshi, Y. (1985) *Modern Multivariate Statistical Analysis : A Graduate Course and Handbook*. Columbus, Ohio: American Sciences Press.

Smith, B.J. (2000) Bayesian output analysis program (BOA). Software available at http://www.public-health.uiowa.edu/boa/.

Smith, J.E. and Mccardle, K.F. (1999) Options in the real world: lessons learned in evaluating oil and gas investments. *Operation Research*, **47,** 1–15.

Spiegelhalter, D.J., Thomas, A. and Best, N.G. (1999) *WinBUGS Version 1.2 User Manual*. Cambridge: MRC Biostatistics Unit.

Stangl, D. and Huerta, G. (2000) Using Bayesian hierarchical models to assess the impact of managed-care strategies. *Lifetime Data Analysis*; Forthcoming.

Stoll, H.R. and Whaley, R.E. (1993) *Futures and Options: Theory and Applications*. Cincinnati, Ohio: South-Western Publishing Co.

Strang, G. (1988) *Linear Algebra and its Applications*, Third edn. Florida: Saunders College Publishing.

Styan, G. P. H. (1970) Notes on the distribution of quadratic forms in singular normal variables. *Biometrika*, **57,** 567–572.

Tierney, L. (1994) Markov chains for exploring posterior distributions (with discussion). *Annals of Statistics*, **22,** 1701–1762.

Uhlig, H. (1994) On singular Wishart and singular multivariate beta distributions. *The Annals of Statistics*, **22,** 395–405.

Verleger, P.K. (1993) *Adjusting to Volatile Energy Prices*. Washington, DC: Institute for International Economics.

Volinsky, C. (1996) S-plus functions for Bayesian model averaging and variable selection for generalized linear models. Software available at Statlib, http://www.stat.cmu.edu/S/bic.glm.

Wackernagel, H. (1998) *Multivariate Geostatistics*, Second edn. Berlin: Springer Verlag.

Wasserman, L. (1997) Bayesian model selection and model averaging. Technical Report # 666. Department of Statistics, Carnegie Mellon University.

West, M. (1984) Outlier models and prior distributions in Bayesian linear regression. *Journal of the Royal Statistical Society, Series* B, **46,** 431–439.

West, M. and Aguilar, O. (1997) Studies of quality monitor time series: The VA hospital system. Discussion Paper # 97-22a. Institute of Statistics and Decision Sciences, Duke University.

West, M., Aguilar, O. and Lourdes, V. (1998) VA hospital quality monitors: 1988-1997. Discussion Paper # 98-41. Institute of Statistics and Decision Sciences, Duke University.

West, M. and Harrison, P.J. (1997) *Bayesian Forecasting and Dynamic Models*, Second edn. New York: Springer Verlag.

Biography

Viridiana Lourdes was born in Mexico City, Mexico on April 4th, 1970. She obtained a Bachelor's degree in Actuarial Sciences and a Master's degree in Finance from the Autonomous Technological Institute of Mexico (ITAM) in September, 1994 and June, 1996 respectively. She arrived at Duke University in the Fall of 1996, she received a Master's degree in Statistics and Decision Sciences and became a Ph.D candidate in May, 1998.