MARKOV TRENDS IN MACROECONOMIC TIME SERIES

Cover design: Mirjam Bode

The book is no. 171 of the Tinbergen Institute Research Series. This series is established through cooperation between Thesis Publishers and the Tinbergen Institute. A list of books which already appeared in the series can be found in the back.

Markov Trends in Macroeconomic Time Series

(Markov Trends in Macroeconomische Tijdreeksen)

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam op gezag van de
Rector Magnificus
Prof. dr P.W.C. Akkermans M.A.
en volgens besluit van het College voor Promoties

De openbare verdediging zal plaatsvinden op
donderdag 27 november 1997 om 11.00 uur

door

Richard Paap
geboren te Rotterdam

Promotiecommissie

Promotor:	Prof. dr H.K. van Dijk
Overige leden:	dr P.H.B.F. Franses
	Prof. dr T. Kloek
	Prof. dr G. Ridder

Preface

Four years ago I started my PhD project as a classical econometrician unaware of the Bayesian thought. After some months I got acquainted with the Bayesian way of thinking and started to write papers about Bayesian analysis of macroeconomic time series. At the same time, I continued my classical research on the modelling of seasonal time series and on the presence of convergence in per capita income among countries. This enabled me to compare the Bayesian and classical approach. In my thesis I have chosen for a Bayesian approach to analyse trends in macroeconomic time series, although this does not imply that I have become a true Bayesian now.

Many persons have contributed to this thesis. First of all, I would like to thank my promotor Herman van Dijk for his suggestions and comments. He allowed me to explore my own research ideas and his enthusiasm was a good stimulus for my research. I also thank the members of the small committee, Philip Hans Franses, Teun Kloek and Geert Ridder for their comments.

Parts of this thesis are based on joint work with several colleagues. In this respect I thank my co-authors Philip Hans Franses, Henk Hoek, Frank Kleibergen and Herman van Dijk for the pleasant cooperation and helpful discussions. Also discussions with Jan Groen, André Lucas, Gerbert Romijn and Dick van Dijk on econometric issues in general have contributed to this thesis.

This thesis was written, while I was employed by the Netherlands Organisation for Scientific Research (N.W.O.). I thank the N.W.O. for sponsoring this PhD project. During the last four years I was also fortunate to visit several conferences and universities. This enabled me to present my work and to exchange ideas with other researchers. The travel funds of the N.W.O., the Tinbergen Institute, Shell, the European Community and the Econometric Society are gratefully acknowledged.

I thank my roommates and other colleagues at the Tinbergen Institute for providing a pleasant atmosphere for doing research. In addition, I am grateful to the secretarial staff of the Tinbergen Institute for their support. I especially enjoyed the teatalks with Carien and Cor, which were a nice break during the day.

Finally, I thank my family and friends. Although most of them do not understand what I am doing, they all showed interest in my work during the last four years. In particular, I thank my parents for their moral support during the difficult moments.

Contents

1 Introduction **1**
1.1 Motivation 1
1.2 Short Overview of Markov Switching Literature 3
1.3 Outline 4

I Univariate Analysis **7**

2 Univariate Stochastic Trends **9**
2.1 Introduction 9
2.2 Trend Specifications 10
2.3 Model Specifications 13
2.4 Illustration of Trend Modelling 15
2.5 Outline of Part I 20

3 Univariate Markov Trend Model **21**
3.1 Introduction 21
3.2 The Markov Trend Model 22
3.3 Bayesian Analysis 25
3.3.1 Prior Specification 26
3.3.2 Unit Root Analysis 27
3.3.3 Markov Trend Analysis 28
3.4 Simulating Posterior Distributions 29
3.4.1 Full Conditional Posterior Distributions 30
3.5 Forecasting 32
3.6 Application 33
3.6.1 Simulated Series 33
3.6.2 German Industrial Production 37
3.7 Concluding Remarks 43

4 Seasonal Markov Trend Model **45**
4.1 Introduction 45
4.2 Modelling Seasonality 46

4.3 The Seasonal Markov Trend Model . . . 49
4.4 Bayesian Analysis . . . 53
4.4.1 Prior Specification . . . 53
4.4.2 Posterior Distributions . . . 55
4.4.3 Seasonal Unit Root Analysis . . . 56
4.5 Illustration of Seasonal Unit Root Analysis . . . 57
4.6 German Unemployment . . . 61
4.6.1 Seasonal Adjusted Series . . . 63
4.6.2 Seasonal Unadjusted Series . . . 64
4.7 Concluding Remarks . . . 69

II Multivariate Analysis 71

5 Multivariate Stochastic Trends 73
5.1 Multivariate Trend Specifications . . . 73
5.2 Common Trends . . . 76
5.3 Cointegration . . . 78
5.4 Outline of Part II . . . 80

6 Cointegration Analysis 81
6.1 Introduction . . . 81
6.2 The Cointegration Model . . . 82
6.3 Prior Framework . . . 85
6.4 Posterior Distributions . . . 89
6.4.1 Simulating Posterior Distributions . . . 92
6.5 Bayesian Lagrange Multiplier Statistics . . . 93
6.5.1 Bayesian LM Statistic in the Linear Regression Model . . . 94
6.5.2 Bayesian LM Statistic in the Cointegration Model . . . 97
6.6 Posterior Odds Analysis . . . 98
6.6.1 Computation of Bayes Factors . . . 99
6.7 Invariant Specification . . . 101
6.7.1 Prior Specification . . . 102
6.7.2 Posterior Distributions . . . 103
6.7.3 Bayes Factors . . . 105
6.8 Illustrative Examples . . . 107
6.8.1 Simulated Series . . . 107
6.8.2 Small Monetary Model for the UK . . . 108
6.8.3 Danish Money Demand . . . 110
6.9 Concluding Remarks . . . 114
6.A Derivation of Prior Distributions . . . 116
6.B Derivation of Posterior Distributions . . . 118
6.C Derivations for the Invariant Specification . . . 120

7 Multivariate Markov Trend Model **125**
7.1 Introduction . . . 125
7.2 The Model . . . 127
7.3 Prior Specification . . . 131
7.4 Simulating Posterior Distributions . . . 132
7.4.1 Full Conditional Posterior Distributions . . . 133
7.5 Bayesian Analysis of Common Trends . . . 135
7.5.1 Cointegration Analysis . . . 135
7.5.2 Common Markov Trends Analysis . . . 136
7.6 Application . . . 138
7.6.1 Simulated Series . . . 138
7.6.2 US Income and Consumption . . . 140
7.7 Concluding Remarks . . . 147

8 Summary and Conclusions **149**

Nederlandse Samenvatting (Summary in Dutch) **153**

Bibliography **159**

Chapter 1

Introduction

1.1 Motivation

Many macroeconomic time series, like industrial production, gross national product and consumption are characterised by long periods of positive growth, expansions, and short periods of decline, recessions. The overall positive growth in these series may be modelled by a linear deterministic trend with a positive slope. If however the short periods of decline have a large impact on future values of the series, a trend specification with constant growth rate may not be flexible enough to capture the trend in the time series. Therefore, the deterministic trend is often extended.

One interesting possibility is to include a stochastic trend. A popular stochastic trend specification is the random walk, which consists of a cumulative sum of uncorrelated stochastic shocks generated by a continuous distribution. This random walk specification assumes that positive and negative shocks, occurring in every time period, have a permanent effect on future values of the series. However, if only shocks during recession periods have a permanent effect, then the random walk may be too flexible to model the trend.

An alternative strategy to model the shocks in the recession periods, which are likely to have a permanent effect, is to use appropriate dummy variables. Due to the deterministic character of this solution, this implies that in forecasting exercises the model does not take into account the possibility of a period of negative growth in the future and provides unjustly less forecast uncertainty.

Hamilton (1989) exploits the fact that recessions are recurrent events. Based on the ideas of modelling the business cycle by Neftçi (1984), he formulates a trend specification, where shocks during recessions have a different effect on future values of the series than shocks during expansions. This so-called Markov trend is a segmented trend with two slopes to model the different growth rates in the recession and expansion periods. The direction of the slope in every period depends on the value of an unobserved first-order two-state Markov process. Instead of a continuous shock in every period like in the random walk specification, the stochastic Markov trend is determined by discrete shocks, which do not have to occur in every period. The Markov trend can be represented in a way,

where the growth in expansion periods is modelled by a linear deterministic trend and the growth during recessions by discrete shocks. The Markov structure in the trend implies that the current slope of the trend depends on the value of the slope in the previous period to model persistency in the time series. This is not the case for the random walk, which consists of a cumulative sum of uncorrelated shocks. Therefore, the Markov trend is more restricted than the random walk.

The Markov trend has been applied to analyse the business cycle in macroeconomic time series. The transition probabilities of the Markov process determine the expected duration of recessions and expansions and inference on regime changes can be used to determine turning points. Hamilton (1989) models real gross national product of the United States using a Markov trend, where the deviations from this trend are assumed to be an autoregressive process. He assumes that the trend in this series can be described by a Markov trend plus a random walk, resulting from a unit root in the autoregressive polynomial. Hence, the future values of the series depend on discrete shocks which have occurred during recession periods and continuous shocks which have happened in every period in the past. Lam (1990) however assumes that the series is stationary around a Markov trend so that only shocks during recessions matter. His results show that the estimated duration of recessions and the dating of turning points depend on the specification of the trend, see also Gordon (1997). This is not surprising, since the Markov trend specification models the trend and the business cycle simultaneously and adding an extra stochastic trend influences inference on the Markov trend. It is therefore for business cycle analysis necessary to have an accurate description of the trend in the series. Furthermore, the trend specification also plays an important role in forecast exercises and impulse-response analysis. Overspecification of stochastic trends leads to superfluous forecast uncertainty, while underspecification assumes unfairly less forecast uncertainty.

In this thesis we focus on the analysis of trends in macroeconomic time series using Markov trend models. For the univariate analysis, we consider model specification, the impact of shocks on future values of the series for several trend specifications and forecast uncertainty. Additionally, we link the slopes of the Markov trend with the stages of the business cycle. The univariate Markov trend model is extended with a seasonal component to analyse the business cycle and the seasonal variation in time series simultaneously. To analyse multivariate macroeconomic time series, we consider a straightforward multivariate extension of the univariate Markov trend model of Hamilton (1989) and Lam (1990). The model consists of a multivariate version of a univariate Markov trend specification and the deviations from this multivariate Markov trend are assumed to be a vector autoregressive process. We focus on model specification and on the role of the trend and in particular on common stochastic trend specifications. Again, the impact of shocks under different trend specifications is discussed. The slopes of the multivariate Markov trend can be linked with the stages of business cycle and common Markov trends may imply the presence of common business cycles in multivariate time series.

To analyse univariate and multivariate Markov trend models, we use Bayesian methods. Recent developments in simulation techniques, like Gibbs sampling, has made the

Bayesian analysis of Markov trend models possible as shown by *e.g.* Albert and Chib (1993) and McCulloch and Tsay (1994b). Inference on for instance the business cycle is not only based on single optimal parameter values like in classical analyses but also takes into account parameter uncertainty. The same arguments holds for out-of-sample inference, like the analysis of forecast distributions. The Bayesian inference in this thesis is based on diffuse prior specifications and can be seen as an extensive way to analyse the information in the likelihood function. It enables us to examine the amount of information in macroeconomic time series as posterior variances are completely dominated by variation in the time series and not heavily influenced by prior specification. Informative priors can however easily be included in the analysis.

To determine the appropriate trend specification in univariate Markov trend models we can build on existing Bayesian methods. However, the analysis of stochastic trends in multivariate Markov trend models is not straightforward and requires new techniques. Therefore, we provide in this thesis a new general Bayesian framework to analyse the number of (common) stochastic trends in vector autoregressive models. This framework is adapted to analyse the presence of common Markov trends in the multivariate Markov trend models to analyse common business cycles.

In the next section, we provide a short overview of the literature about Markov switching including model extensions. Although we usually do not consider all of these extensions, several of them can easily be incorporated in the models considered in this thesis. The last section of this chapter provides a more detailed outline of the thesis.

1.2 Short Overview of Markov Switching Literature

The introduction of the Markov trend model by Hamilton (1989) has resulted in a rapidly expanding literature on the implementation of Markov processes in econometric modelling. Just as for business cycle analysis, the Markov processes are used to model switches between regimes. In this section we provide a short overview of some major contributions of Markov switching in the modelling of economic time series.

One of the main applications of Markov switching models is in business cycle analysis. Hamilton (1989) proposes a time series model, where the growth rate of the series depends on the stage of the business cycle. The business cycle is modelled by an unobserved first-order Markov process like in Neftçi (1984) and Falk (1986). The Hamilton Markov trend model has been used to analyse business cycles in macroeconomic time series. Hamilton (1989) and Lam (1990) use this model to analyse the business cycle in gross national product [GNP] of the United States, Goodwin (1993) analyses the business cycle in real GNP of seven other countries, Layton (1994) considers Australian gross domestic product and Diebold and Rudebusch (1996) compare several composite indexes of coincident indicators.

Several extensions to the original Hamilton model have been proposed. Durland and McCurdy (1994) consider duration-dependent transition probabilities to change from recession to expansion regimes and *vice versa*. In their model the value of transition proba-

bility depends on the number of periods that the process has been in the regime. Filardo (1994) and Diebold, Lee and Weinbach (1994) model the transition probabilities using logistic functions, which include explanatory variables. The Hamilton model assumes a constant dynamic pattern over the business cycle. To allow for different dynamics during recession and expansion periods, Markov switching in the autoregressive parameters have been proposed, see *e.g.* McCulloch and Tsay (1994b) and Hansen (1992). Additionally, for seasonally unadjusted series Ghysels (1994) and Ghysels, McCulloch and Tsay (1994) propose transition probabilities, which are different in every season. Finally, McCulloch and Tsay (1994a) consider Markov switching between trend stationary and difference stationary regimes.

A few studies have extended the univariate Markov trend model to a multivariate model. Phillips (1991a) considers a simple two-dimensional version of the Hamilton to analyse transmissions of business cycles between countries using industrial production series. Kim and Yoo (1995) propose a multivariate Markov switching factor model to construct a composite index of coincident indicators, while Dwyer and Potter (1996) consider multivariate versions of the Markov trend.

Finally, Markov switching dynamics are also found in financial time series. For instance, Hamilton (1988), Ceccheti, Lam and Mark (1990) and Garcia and Perron (1996) detect Markov switching in the conditional mean of interest rates. Engel and Hamilton (1990), Kaminsky (1993) and Engel (1994) model depreciation and appreciation periods in exchange rates with a simple Markov switching model. Tyssedal and Tjøstheim (1988) use Markov switching in the autoregressive structure of dividend rates, while Ceccheti, Lam and Mark (1990) and Bonomi and Garcia (1994) only consider Markov switching in the mean and variance. An extension to regime changes in the conditional variance dynamics of stock returns can be found in Hamilton and Susmel (1994), see also Cai (1994) and Dueker (1997). Hamilton and Lin (1996) investigate the relation between stock market volatility and the business cycle.

In this thesis we focus on the original Markov trend specification of Hamilton (1989) to model macroeconomic time series. We only consider Markov switching structure in the mean of the series and not in the variance. The possibilities and limitations of several of the above mentioned model extensions are discussed. The next section provides a more detailed outline of the thesis.

1.3 Outline

In this section we provide a general outline of the thesis. A detailed outline with references to related literature can be found in the introduction of each chapter. The thesis is partitioned in two parts. In the first part we consider the modelling of the trend component in univariate macroeconomic time series. The second part deals with a multivariate analysis of trends in time series.

The outline of the first part is as follows. In Chapter 2 we provide a short introduction into the modelling of the trend component in univariate time series. We focus on three

trend specifications: the linear deterministic trend, the Markov trend and the even more flexible random walk plus drift. The latter two specifications are stochastic trends and result in a better within-sample description of the trend component in time series, but imply more forecast uncertainty. Additionally, we briefly discuss the modelling of the deviations from the trend component using autoregressive models. We show that a unit root in the autoregressive structure also implies the presence of a stochastic trend in the model.

The univariate Markov trend model is considered in Chapter 3. The ideas in this chapter are based on Hoek and Paap (1994). Following Hamilton (1989) the model is autoregressive in deviation from a Markov trend. We discuss the presence of stochastic trends and especially the impact of exogenous shocks on the level and the first differences of the series under various specifications of the model. The chapter provides a Bayesian strategy to analyse the presence of stochastic trends in time series using the Markov trend model. This includes prior specification, a simulation scheme to obtain posterior results, posterior odds ratios for the presence of stochastic trends and a predictive Bayesian analysis. Finally, we investigate using the Markov trend model the trend in quarterly observed German industrial production and compare forecast distributions generated by Markov and non-Markov trend models.

In Chapter 4 the univariate Markov trend model is extended with a seasonal component to analyse quarterly observed seasonally unadjusted series. This chapter is based on ideas in Franses and Paap (1996) and uses techniques from Franses, Hoek and Paap (1997). We first discuss the consequences of seasonal adjustment on business cycle analysis using Markov switching models and conclude that it is preferable to model the business cycle and the seasonal pattern in a time series simultaneously. In the model the deviations from the Markov trend component now consist of a seasonal component and an autoregressive model. To allow for changes in the seasonal pattern over time, the autoregressive structure may include seasonal unit roots, which lead to the presence of seasonal stochastic trends. The seasonal component consists of seasonal dummies and incorporates via seasonal mean shifts the possibility that changes in the seasonal pattern coincide with changes in the business cycle. The Bayesian techniques in Chapter 3 are extended to analyse the presence of seasonal stochastic trends and seasonal mean shifts. Finally, we consider quarterly observed seasonally adjusted and non-adjusted German unemployment. The estimated business cycle resulting from the model for the adjusted data is compared with the business cycle resulting from the model for the seasonally unadjusted series. We notice differences in the dating of turning points and in the expected duration of contraction and expansion periods.

The second part of the thesis is organised as follows. Just as in the univariate part we start with a short introduction in the modelling of trends in multivariate time series in Chapter 5. The univariate deterministic and stochastic trend specifications of Chapter 2 are generalised to multivariate trend specifications. We discuss parameter restrictions which lead to the presence of common trends. We speak of common trends if the number of trends needed to model the multivariate time series is smaller than the number

of trends needed to model the individual time series separately. In addition, we discuss the modelling of the deviations from the trend component using vector autoregressive moving average models. Just as in the univariate case, unit roots in the autoregressive structure imply the presence of stochastic trends. We focus on the restrictions for common stochastic trends in vector autoregressive models, which is usually referred to as cointegration.

Contrary to univariate Bayesian analysis of unit roots in autoregressive models, there is no complete Bayesian method to analyse cointegration in vector autoregressive models. Therefore, we provide in Chapter 6 a complete Bayesian framework to analyse cointegration in vector autoregressive models. This chapter is an abbreviated version of Kleibergen and Paap (1996). The Bayesian analysis is based on a new decomposition of the parameter matrix which models the error correction. Although the Bayesian framework is developed for the standard linear vector autoregressive model, it turns out to be easy to implement in more complicated models like the multivariate Markov trend model in Chapter 7. The framework includes prior specification, posterior simulation, and Bayesian Lagrange multiplier statistics and posterior odds ratios to analyse the number of cointegration relations. To illustrate the Bayesian cointegration analysis we compare posterior results with classical outcomes of cointegration analysis for the Johansen and Juselius (1990) Danish series and the United Kingdom series of Hendry and Doornik (1994).

In Chapter 7 we consider a multivariate Markov trend model. This model is a multivariate generalisation of the univariate Markov trend model of Chapter 3. It consists of a vector autoregressive model in deviation from a multivariate Markov trend specification. First, we discuss several simplifications of the most general model specification including common Markov trends, cointegration and reduced rank Markov trend cointegration. We focus on the impact of exogenous shocks on the level and the first differences of the series under several specifications. The Bayesian framework to analyse the models includes prior specification, a simulation scheme to obtain posterior results and posterior odds ratios for common Markov trends and cointegration. The analysis of common stochastic trends is based on the techniques presented in Chapter 6. Using the multivariate Markov trend model we conclude that quarterly observed per capita income and consumption of the United States contains one common Markov trend and one cointegration relation. Neglecting the Markov structure leads to favouring the hypothesis of no common stochastic trends.

Finally, we end this thesis with a summary of the main results and conclusions in Chapter 8. Additionally, we provide some directions for further research.

Part I

Univariate Analysis

Chapter 2

Univariate Stochastic Trends

2.1 Introduction

Many macroeconomic variables, like real gross national product [GNP], real consumption, total investment, industrial production have a tendency to grow over time. It is important for policy makers to have accurate forecasts of these variables. For forecasting trending time series, it is crucial to have a well specified trend component. A misspecification of the trend component can lead to a forecast path which lies totally above or below future realisations of the series, see *e.g.* Stock and Watson (1988). The trend of the logarithm of macroeconomic time series is often modelled with a linear deterministic trend. This trend specification corresponds to the assumption of constant exponential growth in the level of the series and implies no forecast uncertainty since it is fully deterministic. In practice, a deterministic trend may not be flexible enough to model the trend in a time series. Therefore, the linear deterministic trend is often extended with a stochastic trend, like a random walk. The random variable in this trend specification allows for a different direction of the trend in every period and produces a better within-sample description of the trend. However, including random variables in the trend specification leads to forecast uncertainty. Hence, from a forecasting point of view it is necessary to avoid overfitting of the trend component by making it too flexible with random variables. The same arguments hold for impulse-response analysis. A misspecification of the trend component may lead to incorrect conclusions about the impact of shocks on future values of the time series. In this part of the thesis, we investigate techniques to analyse trends in univariate time series. As an introduction we provide in this chapter a short overview of the various trend specifications, which have been used to model the trend in macroeconomic time series. Additionally, we discuss the way in which the trend component can be included in a time series model.

The outline of this chapter is as follows. In Section 2.2 we discuss various possibilities for the specification of the trend component in macroeconomic time series. In Section 2.3 we discuss how these trend components may enter a time series model. A simple illustration of the modelling of the trend in GNP of the United States is given in Section 2.4.

Finally, in Section 2.5 we give a short outline of the chapters in the first part of the thesis.

2.2 Trend Specifications

In this section, we consider the most applied trend specifications, for modelling macroeconomic time series. The simplest trend specification n_t, $t = 1, 2, 3, \ldots$ is

$$n_t = n_{t-1} + \gamma_0, \tag{2.1}$$

which corresponds to a linear deterministic trend

$$n_t = n_1 + \gamma_0(t-1) \tag{2.2}$$

with slope γ_0 and initial value n_1. This specification assumes a constant growth rate. For known values of γ_0 and n_1 there is no forecast uncertainty so that the expectation and the variance of n_t at $t = 1$ is $n_1 + \gamma_0(t-1)$ and zero, respectively.[1] If this linear deterministic trend is too restrictive, one can replace the linear specification in (2.2) with a polynomial in t of degree > 1. This however implies that $\lim_{t\to\infty} |n_t - n_{t-1}| \to \infty$, which may be unrealistic for the time series under consideration. Of course, any function of time, $g(t)$, can be used to model the trend, see for instance Granger (1989, p. 28) for useful alternatives. Note that the trend of the logarithm of an exponential growing series can be modelled by a deterministic trend, since $\ln(\exp(\gamma_0 t)) = \gamma_0 t$.

Another possibility for a more flexible trend is to introduce a random variable in the trend specification

$$n_t = n_{t-1} + \gamma_0 + \gamma_1 u_t, \tag{2.3}$$

where $u_t \sim \text{NID}(0,1)$. If $\gamma_0 = 0$ this trend is called a random walk, while for $\gamma_0 \neq 0$ we refer to this trend as a random walk plus drift ($=\gamma_0$). The direction of the trend (2.3) at time t is γ_0 plus an unanticipated shock u_t and is therefore more flexible than the linear deterministic trend (2.2). This follows directly from the backward solution of (2.3)

$$n_t = n_1 + \gamma_0(t-1) + \gamma_1 \sum_{i=2}^{t} u_i. \tag{2.4}$$

The random walk plus drift consists of a deterministic trend plus accumulated shocks $\sum_{i=2}^{t} u_i$, which will be referred to as a stochastic trend in this thesis. The expectation of n_t at $t = 1$ is $n_1 + \gamma_0(t-1)$ and its variance is $\gamma_1^2(t-1)$, which implies that the forecast uncertainty about this trend increases linear over time. Note that instead of a standard normal distribution for u_t, one can also assume other continuous distributions, like t-distributions. Sometimes (2.3) is extended with a stochastic drift term

$$\begin{aligned} n_t &= n_{t-1} + \gamma_{0,t} + \gamma_1 u_t, \\ \gamma_{0,t} &= \gamma_{0,t-1} + \eta_t, \end{aligned} \tag{2.5}$$

[1] Strictly speaking, the expectation and variance are not defined since the trend component does not contain a random variable.

where $\eta_t \sim \text{NID}(0, \sigma_\eta^2)$ and $\text{E}[\eta_t u_t] = 0$, see *e.g.* Harvey (1989). The backward solution of this trend, $n_t = n_1 + \gamma_{0,1}(t-1) + \sum_{i=2}^{t}(\gamma_1 u_i + \sum_{j=2}^{i} \eta_j)$, shows that the expectation of n_t at $t = 1$ is still linear in t, $n_t + \gamma_{0,1}(t-1)$. However contrary to the random walk (2.3), the variance of n_t at $t = 1$, $\gamma_1^2(t-1) + \frac{1}{2}\sigma_\eta^2(t^2 - t)$, increases quadratic over time, since it includes the uncertainty in the drift term $\gamma_{0,t}$. In general, this implies that the random walk plus drift (2.3) has smaller forecasts intervals than the random walk plus stochastic drift term (2.5).

An alternative to the random walk plus deterministic drift was proposed by Hamilton (1989), which is now known in the literature as a Markov trend

$$n_t = n_{t-1} + \gamma_0 + \gamma_1 s_t, \qquad s_t = 0, 1, \tag{2.6}$$

where s_t is an unobserved first-order Markov process with transition probabilities

$$\begin{array}{ll} \Pr[s_t = 0 | s_{t-1} = 0] = p, & \Pr[s_t = 1 | s_{t-1} = 0] = 1 - p, \\ \Pr[s_t = 1 | s_{t-1} = 1] = q, & \Pr[s_t = 0 | s_{t-1} = 1] = 1 - q, \end{array} \tag{2.7}$$

with $0 < p < 1$ and $0 < q < 1$. The direction of this trend at time t is γ_0 if $s_t = 0$ and $\gamma_0 + \gamma_1$ if $s_t = 1$. Instead of a cumulative sum of a continuous random variable u_t, the stochastic trend now is a cumulative sum of a discrete random variable s_t

$$n_t = n_1 + \gamma_0(t-1) + \gamma_1 \sum_{i=2}^{t} s_i. \tag{2.8}$$

We will refer to $\sum_{i=2}^{t} s_i$ as a stochastic Markov trend. Note that this trend does not consist of a cumulative sum of independent shocks, since $\text{E}[s_t s_{t-1}] \neq 0$. To derive the expectation of n_t at $t = 1$, we consider the following first-order autoregressive representation of s_t

$$s_t = (1 - p) + \varphi s_{t-1} + v_t, \tag{2.9}$$

where $\varphi = (-1 + p + q)$ and conditional on $s_{t-1} = 0$

$$\begin{array}{ll} v_t = -(1-p) & \text{with probability } p, \\ v_t = p & \text{with probability } (1-p), \end{array} \tag{2.10}$$

and conditional on $s_{t-1} = 1$

$$\begin{array}{ll} v_t = (1-q) & \text{with probability } q, \\ v_t = -q & \text{with probability } (1-q), \end{array} \tag{2.11}$$

so that $\text{E}[v_t | s_{t-1} = 0] = \text{E}[v_t | s_{t-1} = 1] = 0$, see Hamilton (1989) and Hamilton (1994, p. 683–684). If $\varphi = 0$ (*i.e.* $p + q = 1$), s_t is not correlated with the past and the Markov structure disappears. In this case s_t has an uncorrelated Bernoulli distribution with

$\Pr[s_t = 1] = 1 - p$ and $\Pr[s_t = 0] = p$. Using (2.9) it is easy to show that the expectation of s_t at $t = 1$ given the initial value s_1 equals

$$\begin{aligned} \mathrm{E}[s_t|s_1] &= \sum_{i=0}^{t-2}(1-p)\varphi^i + \varphi^{t-1}s_1 \\ &= \frac{(1-p)(1-\varphi^{t-1})}{(1-\varphi)} + \varphi^{t-1}s_1 \\ &= \pi + \varphi^{t-1}(s_1 - \pi) = \Pr[s_t = 1|s_1], \end{aligned} \tag{2.12}$$

where $\pi = (1-p)/(2-p-q)$. The unconditional probability that $s_t = 1$, $\Pr[s_t = 1]$, equals $\lim_{t\to\infty} \mathrm{E}[s_t|s_1]$, which is π if $\varphi < 1$. Also the expectation $\mathrm{E}[s_t|\Pr[s_1 = 1] = \pi]$ equals π. Hence, the expectation of n_t at $t = 1$ equals

$$\begin{aligned} \mathrm{E}[n_t|n_1, s_1] &= \mathrm{E}\left[n_0 + \gamma_0(t-1) + \gamma_1 \sum_{i=2}^{t} s_i \middle| n_1, s_1\right] \\ &= n_0 + \gamma_0(t-1) + \gamma_1 \sum_{i=2}^{t}(\pi + \varphi^{i-1}(s_1 - \pi)) \\ &= n_1 + (\gamma_0 + \pi\gamma_1)(t-1) + \gamma_1(s_1 - \pi)\sum_{i=1}^{t-1}\varphi^i \\ &= n_1 + (\gamma_0 + \pi\gamma_1)(t-1) + \varphi\gamma_1(s_1 - \pi)\frac{1-\varphi^{t-1}}{1-\varphi}, \end{aligned} \tag{2.13}$$

see also Hamilton (1989). In practice the value of s_1 is seldom known. If we have no useful information about s_1 apart from $\Pr[s_1 = 1] = \pi$, the expectation of n_t at $t = 1$ simplifies to $n_1 + (\gamma_0 + \pi\gamma_1)(t-1)$. In the same way it can be shown that the variance of n_t at $t = 1$ given that $\Pr[s_1 = 1] = \pi$ equals

$$\begin{aligned} &\mathrm{E}[(n_t - \mathrm{E}[n_t])^2|\Pr[s_1 = 1] = \pi] \\ &= \gamma_1^2 \mathrm{E}\left[\left(\sum_{i=2}^{t} s_i - \mathrm{E}\left[\sum_{i=2}^{t} s_i\right]\right)^2 \middle| \Pr[s_1 = 1] = \pi\right] \\ &= \gamma_1^2 \mathrm{E}\left[\left(\sum_{i=2}^{t} s_i\right)^2 \middle| \Pr[s_1 = 1] = \pi\right] - \gamma_1^2 \mathrm{E}\left[\sum_{i=2}^{t} s_i \middle| \Pr[s_1 = 1] = \pi\right]^2 \\ &= \gamma_1^2 \mathrm{E}\left[\sum_{i=2}^{t} s_i^2 + 2\sum_{i=2}^{t}\sum_{j=i+1}^{t} s_j s_i \middle| \Pr[s_1 = 1] = \pi\right] - \gamma_1^2 \mathrm{E}\left[\sum_{i=2}^{t} s_i \middle| \Pr[s_1 = 1] = \pi\right]^2 \\ &= \gamma_1^2 \pi(1-\pi)\left((t-1) + 2\varphi\frac{((1-\varphi)(t-1) - (1-\varphi^{t-1})}{(1-\varphi)^2}\right), \end{aligned} \tag{2.14}$$

where we use that $\mathrm{E}[s_j s_i|\Pr[s_1 = 1] = \pi] = \Pr[s_j = 1|s_i = 1]\Pr[s_i = 1|\Pr[s_1 = 1] = \pi]$. Note that the variance of the Markov trend for large t increases linear over time.

In summary, we have discussed in this section the properties of four trend specifications. We started with the linear deterministic trend, which incorporates no forecast uncertainty. The random walk plus drift and the random walk plus stochastic drift term trend specification are more flexible but lead to forecast uncertainty. Finally, we have considered the properties of the Markov trend. Since this trend allows for two possible slope values, it is less flexible than a random walk but more flexible than a linear deterministic trend. It can be seen as a segmented trend, where the slope changes according to a first-order Markov process. Although it is possible to think of more trend specifications, we limit ourselves in this section to the most applied trends. In Section 2.4 we discuss as an illustration the fit and the forecast uncertainty of the linear deterministic trend, the random walk and the Markov trend in modelling gross national product of the United States. In the next section, we discuss the role of trend components in time series models.

2.3 Model Specifications

To include a trend in a time series models there are several possibilities. One can model the time series in deviation from a trend component n_t or one can simply add the trend component to the time series model in a linear way. The former corresponds to the assumption that a univariate time series $\{y_t\}_{t=1}^T$ can be decomposed as

$$y_t = n_t + z_t, \tag{2.15}$$

where n_t is the trend component and z_t represents the deviations from the trend. Sometimes (2.15) is extended with a seasonal component and/or a cyclical component. The deviations from the trend are usually assumed to be an autoregressive moving average [ARMA(k,l)] process

$$z_t - \sum_{i=1}^{k} \phi_i z_{t-i} = \epsilon_t - \sum_{i=1}^{l} \psi_i \epsilon_{t-i}, \tag{2.16}$$

where $\epsilon_t \sim \text{NID}(0, \sigma^2)$.[2] If we use the lag operator L, defined as $L^i z_t = z_{t-i}$, $i = 0, 1, 2, \ldots$, we can write (2.16) as

$$\begin{aligned}(1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_k L^k) z_t &= (1 - \psi_1 L - \psi_2 L^2 - \cdots - \psi_l L^l)\epsilon_t, \\ \phi(L) z_t &= \psi(L)\epsilon_t\end{aligned} \tag{2.17}$$

where $\phi(L)$ and $\psi(L)$ are polynomials in the lag operator L of order k and l respectively.

The process z_t is stationary if the roots of the polynomial $\phi(x)$ are outside the unit circle and also invertible if the roots of the polynomial $\psi(x)$ are outside the unit circle, see among others Granger and Newbold (1987). If this is the case y_t is a stationary and invertible ARMA process around the trend n_t. If however one of roots of $\phi(x)$ is one (unit

[2] It is of course possible to assume other continuous distributions.

root), the process z_t contains a stochastic trend. We refer to z_t in this case as a unit root process or an integrated process of order one, $z_t \sim \mathrm{I}(1)$. We can write the polynomial $\phi(L)$ as

$$\phi(L) = (1 - L)\phi^*(L), \tag{2.18}$$

where $\phi^*(L) = (1 - \phi_1^* L - \cdots - \phi_{k-1}^* L^{k-1})$ is a polynomial in the lag operator of order $(k-1)$. This implies an ARMA$(k-1, l)$ model for the first difference of z_t

$$\begin{aligned} \phi^*(L)(1-L)z_t &= \psi(L)\epsilon_t, \\ (1 - \phi_1^* L - \cdots - \phi_{k-1}^* L^{k-1})(1-L)z_t &= (1 - \psi_1 L - \cdots - \psi_l L^l)\epsilon_t, \\ \Delta z_t - \sum_{i=1}^{k-1} \phi_i^* \Delta z_{t-i} &= \epsilon_t - \sum_{i=1}^{l} \psi_i \epsilon_{t-i}, \end{aligned} \tag{2.19}$$

where $\Delta z_t = (1-L)z_t = z_t - z_{t-1}$. To show that z_t contains a stochastic trend we write Δz_t as a MA(∞)

$$\begin{aligned} \phi^*(L)\Delta z_t &= \psi(L)\epsilon_t, \\ \Delta z_t &= (\phi^*(L))^{-1}\psi(L)\epsilon_t, \\ \Delta z_t &= c(L)\epsilon_t, \end{aligned} \tag{2.20}$$

where $c(L) = (\phi^*(L))^{-1}\psi(L)$ is a lag polynomial with $c_1 = 1$ and $\sum_{i=0}^{\infty} i|c_i| < \infty$. The lag polynomial $c(L)$ can be decomposed as

$$c(L) = c(1) + (1 - L)c^*(L), \tag{2.21}$$

where $c^*(L)$ is a lag polynomial see Johansen (1995, p. 47), so that

$$z_t = z_1 + c(1)\sum_{i=2}^{t} \epsilon_i + c^*(L)\epsilon_t. \tag{2.22}$$

The process z_t is the sum of a stochastic trend $\sum_{i=2}^{t} \epsilon_i$ and a MA(∞) process $c^*(L)\epsilon_t$. This decomposition was first made by Beveridge and Nelson (1981) and is known as the *Beveridge-Nelson* decomposition.

From (2.22) if follows that under the decomposition (2.15) n_t does not have to be the sole trend component. If the ARMA process z_t contains a unit root, $y_t (= n_t + z_t)$ consists of the sum of two trends, for instance a stochastic Markov trend $\sum_{i=2}^{t} s_i$ in n_t plus a stochastic trend $\sum_{i=2}^{t} \epsilon_i$ through z_t, or a linear deterministic trend t in n_t plus a stochastic trend $\sum_{i=2}^{t} \epsilon_i$ through z_t.

Finally, it is also possible to have two unit roots in the polynomial $\Phi(x)$, *i.e.* z_t is integrated of the order two, $z_t \sim \mathrm{I}(2)$. In this case the stochastic trend consists of a double sum of unanticipated shocks, like for the trend specification (2.5). In practice, there are not many examples of economic series, which seem to have two unit roots. Therefore, we consider the analysis of these processes beyond the scope of this thesis.

An alternative model specification is

$$y_t - \sum_{i=1}^{k} \phi_i y_{t-i} = n_t + \epsilon_t - \sum_{i=1}^{l} \psi_i \epsilon_{t-i}. \tag{2.23}$$

Instead of modelling the time series in deviation from its trend, the trend component n_t is added in a linear way. The interpretation of n_t is difficult since it does not represent the trend component in y_t. Note that if n_t is a linear deterministic trend, it is possible to find a one-to-one relation between the parameters in the decomposition in (2.15) and in (2.23) unless $\phi(L)$ contains a unit root. In that case n_1 is not identified, if we use decomposition (2.15) but still identified in (2.23).

If $\phi(L)$ contains a unit root, the first difference of y_t still contains the trend component n_t

$$\Delta y_t - \sum_{i=1}^{k-1} \phi_i^* \Delta y_{t-1} = n_t + \epsilon_t - \sum_{i=1}^{l} \psi_i \epsilon_{t-i} \tag{2.24}$$

and it is straightforward to show that

$$y_t = y_1 + c(1) \sum_{i=2}^{t} (n_i + \epsilon_i) + c^*(L)\epsilon_t \tag{2.25}$$

so that y_t consists of a stochastic trend $\sum_{i=2}^{t} \epsilon_i$ plus the accumulated sum of the trend component n_i. Loosely speaking, n_t now acts as a drift term. If n_t contains a linear deterministic trend, like in the linear deterministic trend specification (2.2) but also in the random walk plus drift (2.4) and the Markov trend (2.8), a unit root in $\phi(x)$ implies the presence of a quadratic trend in y_t.

In this thesis we use specification (2.15). In this specification the component n_t can be interpreted as a trend in y_t, which is not the case in (2.23). Furthermore, the role and hence the interpretation of the trend component n_t does not change if a unit root enters the autoregressive part of the model.

2.4 Illustration of Trend Modelling

To illustrate the modelling of the trend, we consider in this section seasonally adjusted quarterly observed real GNP of the United States [US], 1951.I–1984.IV. Figure 2.1 displays a plot of the level of the series. We notice that the series increases over time with short periods of decrease, for instance in the middle of the 1970s and in the beginning of the 1980s. Since the growth in the series seems exponential, one usually models the logarithm of this series. To model the trend component of the logarithm of the series, several specifications have been proposed. In this section, we focus on four trend specifications, which play an important role in this thesis. The aim of this section is to show the

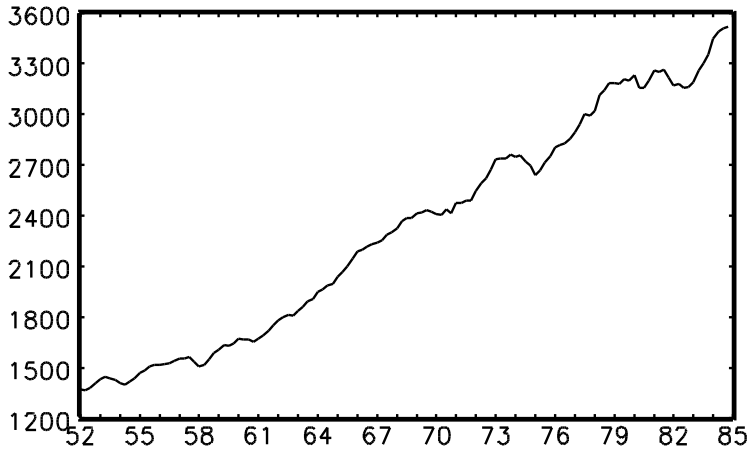

Figure 2.1. Seasonally adjusted US real GNP, 1951.I–1984.IV.

properties and implications of the various trend specifications and not to find the best trend specification for this series.

To analyse the series we consider $y_t = 100 \times \ln(\text{GNP}_t)$. A simple model for this transformed series is the trend stationary [TS] model. One assumes that the series is a stationary process around a linear deterministic trend (2.2)[3]

$$\begin{aligned} y_t &= n_t + z_t, \\ n_t &= n_{t-1} + 0.75, \\ z_t &= 1.31 z_{t-1} - 0.36 z_{t-1} + \epsilon_t, \qquad \sigma = 0.99. \end{aligned}$$

The slope of the linear deterministic trend is 0.75, which implies an average quarterly growth rate of 0.75%. The deviations from the trend follow an AR(2) process. The first row of Figure 2.2 shows the transformed GNP series plus the fitted linear deterministic trend and the series minus this linear trend. In the deviations from the trend one can see the several recessions in the US economy. The roots of the z_t process are 0.92 and 0.39 and hence a unit root in the z_t process may be plausible. Imposing a unit root in the autoregressive part of the model leads to the following difference stationary [DS] model for y_t[3]

$$\begin{aligned} y_t &= n_t + z_t, \\ n_t &= n_{t-1} + 0.72, \\ \Delta z_t &= 0.34 \Delta z_{t-1} + \epsilon_t, \qquad \sigma = 1.00. \end{aligned}$$

The first difference of y_t is a stationary process with mean 0.72. This implies that the estimated average quarterly growth percentage (=0.72) is somewhat smaller than for

[3]Parameter values are obtained using ordinary least squares.

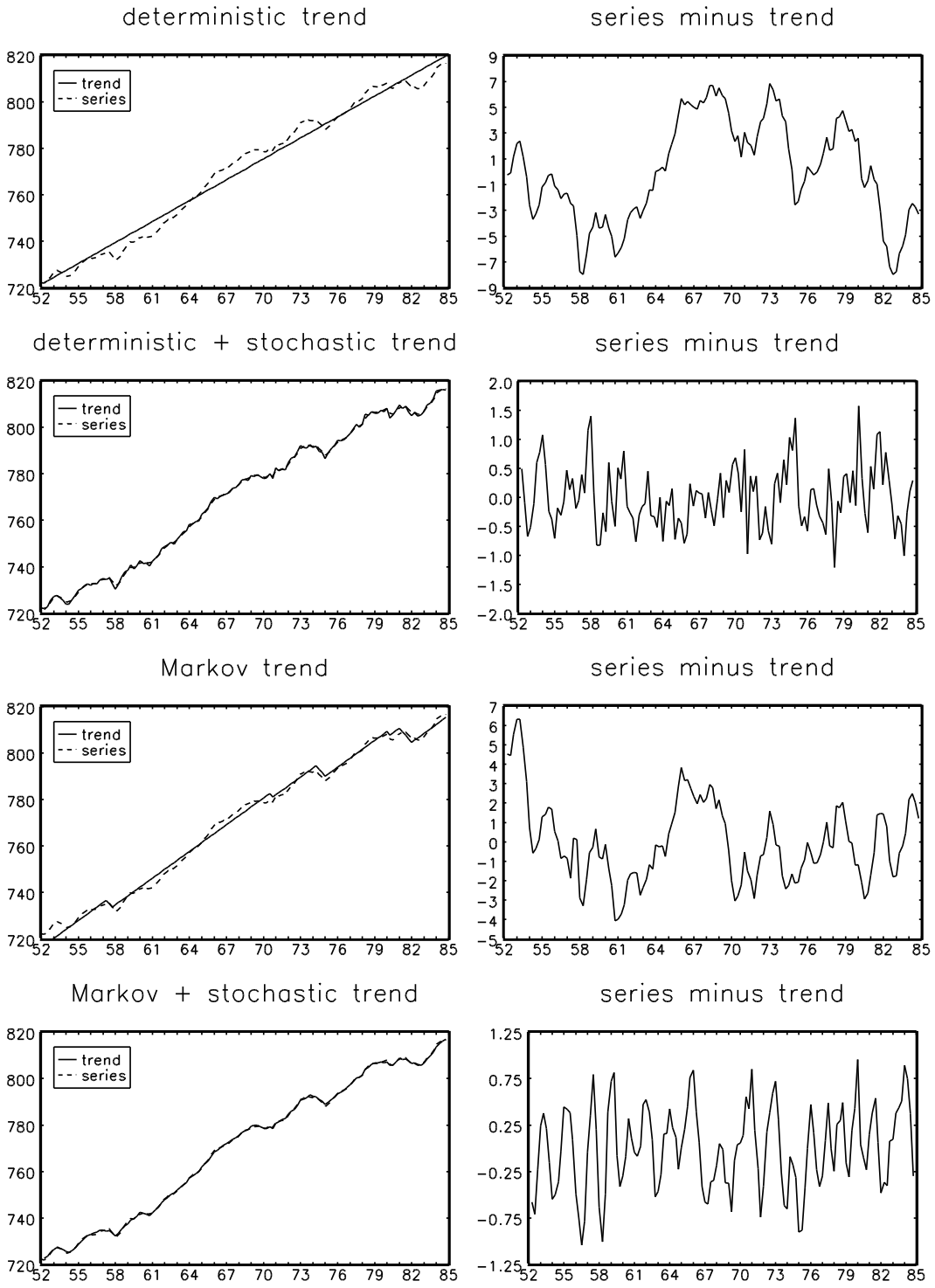

Figure 2.2. US real gross national product and fitted trends plus deviations from these trends, 1952.I–1984.IV.

the trend stationary model. The y_t series contains a deterministic linear trend n_t and a stochastic trend resulting from the unit root in z_t, see (2.22). The second row of Figure 2.2 shows the series and the sum of the two trends. The second column shows the deviations from the total trend. Contrary to the deterministic trend specification, we cannot detect recessions any more and hence the stochastic trend seems to capture the recessions.

The stochastic trend allows for a continuum of slope values for the trend in every period. A more restrictive stochastic trend, which allows for only two different slopes is the Markov trend (2.6). Usually, the two different slopes are related to the stages of the business cycle, *i.e* modelling an expansion and a recession regime. Lam (1990) assumes that the deviations from a Markov trend in US real GNP follow a stationary AR process. We refer to this model specification as a Markov trend stationary [MTS] model[4]

$$
\begin{aligned}
y_t &= n_t + z_t, \\
n_t &= n_{t-1} + 0.96 - 2.45 s_t, && p = 0.96,\ q = 0.51, \\
z_t &= 1.24 z_{t-1} - 0.38 z_{t-2} + \epsilon_t, && \sigma = 0.77.
\end{aligned}
$$

The Markov trend allows for a different direction of the trend during recession ($s_t = 1$) and expansion periods ($s_t = 0$). During an expansion period the quarterly growth rate of GNP is 0.96%, while during a recession -1.49% ($=0.96-2.45$). The probability of staying in an expansion period is 0.96 and the probability of staying in a recession is 0.51. Hence, the unconditional probability of being in a recession is $(1-p)/(2-p-q) = 0.08$. Using this probability and (2.13) the unconditional expectation of the slope of the Markov trend is $0.96 - 2.45 \times 0.08 = 0.76$, which is close to the value of the slope of the trend stationary model. The third row of Figure 2.2 shows y_t and the Markov trend.[5] The trend captures some of the recessions but not all. This is even more clear from the deviations from the Markov trend, which are shown in the second column.

Hamilton (1989) imposes a unit root in the deviations from the Markov trend[6]

$$
\begin{aligned}
y_t &= n_t + z_t \\
n_t &= n_{t-1} + 1.16 - 1.52 s_t, && p = 0.90,\ q = 0.76 \\
\Delta z_t &= 0.01 \Delta z_{t-1} - 0.06 \Delta z_{t-2} - 0.25 \Delta z_{t-3} - 0.21 \Delta z_{t-4} + \epsilon_t, && \sigma = 0.77.
\end{aligned}
$$

We refer to this model a Markov difference stationary [MDS] model, since the first difference of y_t minus the Markov drift term $(1.16 - 1.52 s_t)$ is stationary. The average quarterly growth rate during an expansion period is 1.16% and during a recession -0.36% $(= 1.16 - 1.52)$. The unconditional probability of being in a recession is 0.29 and hence the unconditional expectation of the slope of the Markov trend equals $1.16 - 1.52 \times 0.29 = 0.72$, which is almost the same as the drift term in the difference stationary model. The final

[4]Parameter values are obtained from Table 1 in Lam (1990).

[5]Since s_t is not observed, we have to construct the Markov trend using the ex-post probabilities of being in a recession $\Pr[s_t = 1 | y_1, \ldots, y_T]$, which result from the estimation procedure, see Table 3 in Lam (1990). We have set $s_t = 1$ if $\Pr[s_t = 1 | y_1, \ldots, y_T] > 0.5$ and zero elsewhere.

[6]Parameter values are obtained from Table I in Hamilton (1989).

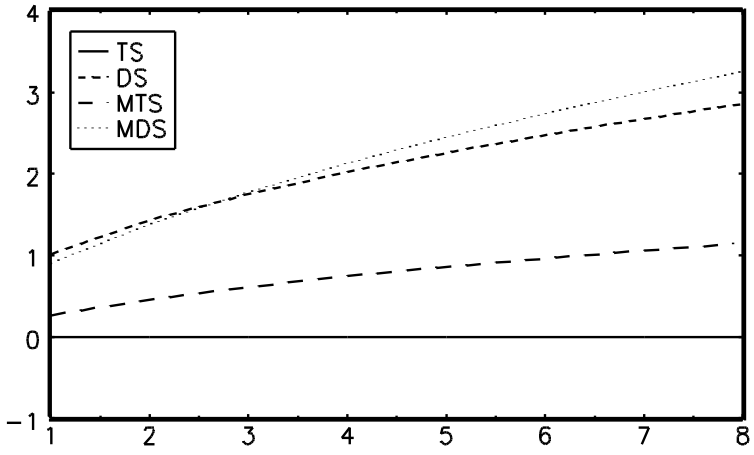

Figure 2.3. Standard deviations of the one- through eight-step ahead forecast distribution of the trend in US GNP.

row of Figure 2.2 shows a plot of the Markov trend[7] plus the stochastic trend resulting from the unit root in z_t. The deviations from the Markov and stochastic trend do not display any recessions. Note that this model produces the smallest deviations from the trend, followed by the difference stationary model and the Markov trend stationary model. This does not have to imply that the Markov difference stationary model is the best model, since it may suffer from overfitting the trend component.

Finally, we consider the forecast uncertainty of the various trend specifications. We treat the parameter values as fixed and hence the forecast uncertainty results only from the random variables in the trend component. Figure 2.3 shows the standard deviations of the one- through eight-step head forecast distributions of the trend for the four models. Since the trend in the trend stationary model is deterministic the standard deviation is zero. This does not imply that we can obtain perfect forecasts of the trend in US GNP with this model, since the trend may be misspecified. Since $\mathrm{E}[s_t s_{t-i}] \neq 0$, the forecast uncertainty resulting from the Markov trend depends on the value of the state in the last period of the sample s_T. We assume that the probability that $s_T = 1$ equals the unconditional probability that s_T equals one, *i.e.* $\Pr[s_T = 1] = \pi$. The variance of the forecast distribution of the Markov trend simplifies to (2.14), where $(t-1)$ has to replaced by the number of periods ahead h, one wants to forecast. The forecast variance of the stochastic trend, which results from the unit root in z_t, is $\sigma^2 h$.

We see from Figure 2.3 that the forecast standard deviation of the trend of the Markov trend [MT] stationary model is the smallest. For one- and two-step ahead the forecast standard deviation of the trend of the Markov difference stationary [MDS] model is smaller than for the difference stationary [DS] model but for more periods ahead it is the oppo-

[7]The Markov trend is constructed using the ex-post probabilities, see Figure 1 in Hamilton (1989).

site way. In general, it seems that imposing more flexible trends leads to more forecast uncertainty. It is therefore necessary to specify a trend which is flexible enough to provide a good within-sample description of the trend without implying too much forecast uncertainty.

2.5 Outline of Part I

In this part of the thesis, we propose Bayesian methods to analyse univariate Markov trend models. The Bayesian analysis focuses on the analysis of the presence of stochastic trends in univariate time series. In Chapter 3 we discuss the properties of the four models, discussed in the previous section, in more detail and we apply the proposed Bayesian trend analysis to German industrial production. To evaluate the out-of-sample performance we compare the forecast distributions of the four models.

Markov trend models are also used to analyse business cycles in time series. The unobserved state variables represent the stage of the business cycle. In Chapter 4 we pay more attention to modelling the business cycle with Markov trend models. We extend the Markov trend model with a seasonal component to analyse seasonally unadjusted quarterly observed time series. We propose Bayesian methods to analyse the seasonal component, including the analysis of the presence of seasonal stochastic trends. Additionally, we examine the consequences of seasonal adjustment on business cycle analysis in German unemployment.

Chapter 3

Univariate Markov Trend Model

3.1 Introduction

Until the middle of the 1970s, it was widely believed that the trend component of macroeconomic series, like the logarithm of output, could be described accurately by a deterministic linear trend, see for instance Lucas (1973). As we already have seen in Chapter 2 this corresponds to the assumption of a constant growth rate in the series. However, Nelson and Plosser (1982), using the unit root tests of Fuller (1976) and Dickey and Fuller (1979), were unable to reject the hypothesis of a unit root in the autoregressive representation of the deviations from a linear trend for thirteen macroeconomic time series of the United States [US]. This result suggests the presence of a stochastic trend so that all shocks in the past have a permanent effect on the future level of the series.

Perron (1989) argues that only a few shocks, like the 1929 Great Depression and the 1973 oil crisis, are likely to have had permanent effects. Modelling these events by appropriate dummy variables, Perron rejects the null hypothesis of a stochastic trend against the alternative hypothesis of a shifting and segmented linear trend for many of the Nelson and Plosser series. Subsequent econometric inference has shown that Perron's conclusions are sensitive to his assumptions about the breakpoints. Treating the breakpoints as unknown parameters results in non-rejection of the stochastic trend hypothesis, see among others Zivot and Andrews (1992), Chu and White (1992) and Hoek (1997).

Hamilton (1989) recognizes that the growth of series like real gross national product [GNP] depends on the stage of the business cycle. He uses the idea of modelling the business cycle with a Markov process by Neftçi (1984) in a time series model, where the value of the growth rate of a series depends on the stage of the business cycle, see also Goldfeld and Quandt (1973) and Lindgren (1978). The model consists of a Markov trend driven by an unobserved two-state first-order Markov process. By considering growth rates, the unit root is implicitly imposed. Therefore, Hamilton assumes that the series contains a stochastic trend and all shocks have a permanent effect on the level of the series. Lam (1990) avoids imposing the unit root in the autoregressive part of the model by extending Hamilton's model to analyse the level of the series instead of the first differences.

Lam's model only contains a Markov trend, where the slope depends on the stage of the business cycle. Now only shocks during recessions have a permanent effect on the future level of the series. The trend component in the model proposed by Lam (1990) is therefore more restricted than the overall trend in the model proposed by Hamilton (1989) or in an autoregressive model in first differences, but less restricted than the deterministic trend.

In Chapter 2 we have seen that introducing more stochastic trends in a time series model may lead to a better within-sample fit, but leads to larger forecast intervals and hence more forecast uncertainty. It is therefore necessary for forecasting exercises to have an adequate impression of the nature and the number of stochastic trends in a time series. This chapter focuses on econometric inference in so-called Markov trend models via a likelihood approach. We consider Bayesian techniques to provide a posterior comparison of the above mentioned trend specifications for modelling the trend in time series. The within-sample comparison is combined with a predictive Bayesian analysis for out-of-sample evaluation of the models.

The outline of this chapter is as follows. In Section 3.2 the Markov trend model is presented and the implications of a unit root in the autoregressive component are discussed. Section 3.3 provides a Bayesian framework to analyse the Markov trend models. Furthermore, a posterior odds ratio test for a unit root and for the presence of a Markov trend is proposed. To obtain posterior results, a simulation procedure is discussed in Section 3.4. This simulation procedure is extended in Section 3.5 to compute predictive densities. In Section 3.6 the Bayesian analysis of stochastic trends in Markov trend models is illustrated with two simulated series and quarterly observed German industrial production series. For the latter series we also analyse the business cycle. Forecasts from the Markov trend models are compared with forecasts from difference and trend stationary models. Finally, Section 3.7 concludes.

3.2 The Markov Trend Model

Suppose that a time series $\{y_t\}_{t=1}^T$ can be decomposed into two parts,

$$y_t = n_t + z_t, \tag{3.1}$$

where n_t is a Markov trend and z_t represents the deviations from this trend. The Markov trend is defined as

$$n_t = n_{t-1} + \gamma_0 + \gamma_1 s_t, \qquad s_t = 0, 1, \tag{3.2}$$

where the unobserved state variable s_t is a two-state first-order Markov process[1] with transition probabilities,

$$\begin{aligned} \Pr[s_t = 0|s_{t-1} = 0] = p, \quad \Pr[s_t = 1|s_{t-1} = 0] = 1 - p, \\ \Pr[s_t = 1|s_{t-1} = 1] = q, \quad \Pr[s_t = 0|s_{t-1} = 1] = 1 - q. \end{aligned} \tag{3.3}$$

[1] If $s_t \sim \text{NID}(0, 1)$ the model becomes a structural time series model, see *e.g.* Harvey (1989).

The state variable s_t represents the stage of the business cycle, see Hamilton (1989). We assume in this thesis that $s_t = 0$ corresponds to an expansion regime, while $s_t = 1$ denotes a recession. The Markov trend (3.2) is a stochastic segmented trend, whose slope is γ_0 during an expansion and $\gamma_0 + \gamma_1$ during a recession. Since s_t can only take the value 0 or 1 the Markov trend is more restricted than a Gaussian random walk process, see also Section 2.2.

The deviations from the Markov trend $\{z_t\}_{t=1}^T$ are modelled by an autoregressive process of order k [AR(k)],

$$\begin{aligned}(1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_k L^k) z_t &= \epsilon_t, \\ \phi(L) z_t &= \epsilon_t,\end{aligned} \tag{3.4}$$

where $\phi(L)$ is a polynomial in the lag operator L, defined by $L^i z_t = z_{t-i}$, $i = 0, 1, 2, \ldots$, and $\epsilon_t \sim \text{NID}(0, \sigma^2)$. Replacing z_t by $(y_t - n_t)$ and substituting the backward solution of (3.2)

$$n_t = \gamma_0(t-1) + \gamma_1 \sum_{i=2}^{t} s_i + n_1 \tag{3.5}$$

equation (3.4) can be rewritten as

$$\phi(L)(y_t - \gamma_0(t-1) - \gamma_1 \sum_{i=2}^{t} s_i - n_1) = \epsilon_t. \tag{3.6}$$

Stochastic Trends

From the backward solution of the Markov trend (3.5) we can see that past stochastic shocks, denoted by s_t, have a permanent effect on the future level of the series y_t. The influence of the these shocks on the first differences of y_t depends on the presence of a unit root in the polynomial $\phi(x)$. To analyse the implications of a unit root in the AR polynomial, we decompose $\phi(L)$ as follows

$$\phi(L) = (1 - \rho L) + \bar{\phi}(L)(1 - L), \tag{3.7}$$

where $\rho = \sum_{j=1}^{k} \phi_j$ and $\bar{\phi}(L) = (1 - \bar{\phi}_1 L - \cdots - \bar{\phi}_{k-1} L^{k-1})$, a lag polynomial of order $(k-1)$ with $\bar{\phi}_i = -\sum_{j=i+1}^{k} \phi_j$, $i = 1, \ldots, k-1$. Note that for $\rho = 1$ the polynomial $\phi(x)$ contains a unit root. Using (3.7) we can write (3.6) as

$$(y_t - n_t) = \rho(y_{t-1} - n_{t-1}) + \sum_{i=1}^{k-1} \bar{\phi}_i \Delta(y_{t-i} - n_{t-i}) + \epsilon_t, \tag{3.8}$$

where Δ is a difference operator defined as $\Delta y_t = (1 - L) y_t = y_t - y_{t-1}$.

Hamilton (1989) assumes that the polynomial $\phi(L)$ contains a unit root, in other words $\rho = 1$. In that case the model simplifies to

$$(\Delta y_t - \gamma_0 - \gamma_1 s_t) = \sum_{i=1}^{k-1} \bar{\phi}_i (\Delta y_{t-i} - \gamma_0 - \gamma_1 s_{t-i}) + \epsilon_t. \tag{3.9}$$

This implies that the series is described by a random walk with a stochastic drift term $\gamma_0 + \gamma_1 s_t$. If the roots of the polynomial $\bar{\phi}(x)$ are outside the unit circle, past shocks s_t only have a transitory effect on the first difference of y_t. Stochastic shocks, denoted by ϵ_t, have a permanent effect on the deviation from the Markov trend and therefore on the level of the series y_t but not on the first differences. As we have seen in Section 2.3 y_t contains a stochastic Markov trend $\sum_{i=2}^{t} s_i$ plus the stochastic trend $\sum_{i=2}^{t} \epsilon_i$. Note that the initial value of the Markov trend n_1 does not enter (3.9).

Lam (1990) assumes that the roots of the polynomial $\phi(x)$ are outside the unit circle so that the series is a stationary AR process around the Markov trend (3.2). From (3.8) it can be seen that past s_t shocks have a permanent effect on the first difference of y_t. The past ϵ_t shocks have only a transitory effect on y_t. We will call such a series Markov trend stationary.

The Markov process s_t introduces a non-linear trend in a time series model. If $s_t = 0$ the slope of the Markov trend (3.2) is γ_0, while if $s_t = 1$ the slope is $\gamma_0 + \gamma_1$. Under the restriction $\gamma_1 = 0$ the Markov trend (3.2) degenerates to a deterministic trend with slope γ_0 and intercept n_1. Model (3.8) with $\rho < 1$[2] becomes a trend stationary AR(k) model,

$$(y_t - \gamma_0(t-1) - n_1) = \rho(y_{t-1} - \gamma_0(t-2) - n_1) + \sum_{i=1}^{k-1} \bar{\phi}_i \Delta(y_{t-i} - \gamma_0) + \epsilon_t, \tag{3.10}$$

while model (3.9) results in a difference stationary model with drift γ_0

$$(\Delta y_t - \gamma_0) = \sum_{i=1}^{k-1} \bar{\phi}_i (\Delta y_{t-i} - \gamma_0) + \epsilon_t. \tag{3.11}$$

The Likelihood Function

To analyse the Markov trend model we have to specify the likelihood function. First, we consider the density function of y_t given the past observations $y^{t-1} = \{y_1, \ldots, y_{t-1}\}$ and given the past and current states $s^t = \{s_1, \ldots, s_t\}$ for model (3.8). This conditional density function reads

$$f(y_t | y^{t-1}, s^t, \gamma_0, \gamma_1, n_1, \sigma, \rho, \bar{\phi}) = \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{1}{2\sigma^2}\epsilon_t^2), \tag{3.12}$$

where ϵ_t is given in (3.8) and $\bar{\phi} = \{\bar{\phi}_1, \ldots, \bar{\phi}_{k-1}\}$. Note that we consider the initial value of the Markov trend n_1 as a parameter. Conditional on the states and on the initial observations $y^k = \{y_1, \ldots, y_k\}$ the likelihood function can be written as the product of the conditional densities in (3.12)

$$\mathcal{L}(y^T | y^k, s^T, \theta) = p^{\mathcal{N}_{00}} (1-p)^{\mathcal{N}_{01}} q^{\mathcal{N}_{11}} (1-q)^{\mathcal{N}_{10}} \prod_{t=k+1}^{T} f(y_t | y^{t-1}, s^t, \gamma_0, \gamma_1, n_1, \sigma, \rho, \bar{\phi}), \tag{3.13}$$

[2]The exact restriction is that the roots of the polynomial $\phi(x)$ are outside the unit circle.

where $\theta = \{\gamma_0, \gamma_1, n_1, \sigma, \rho, \bar{\phi}, p, q\}$, $\mathcal{N}_{ij}$ denotes the number of transitions from state i to state j. Note that we use that $\{y^{t-1}, s^{t-2}\}$ does not Granger cause s_t, *i.e.* the conditional distribution $p(s_t|s^{t-1}, y^{t-1})$ equals the conditional distribution $p(s_t|s_{t-1})$.

The unconditional (on the states) likelihood function is given by

$$\mathcal{L}(y^T|y^k, \theta) = \sum_{s_1=0}^{1} \sum_{s_2=0}^{1} \cdots \sum_{s_T=0}^{1} \mathcal{L}(y^T|y^k, s^T, \theta). \tag{3.14}$$

The evaluation of this likelihood requires the summation and evaluation of 2^T unconditional likelihood functions. To avoid this huge summation Hamilton (1989) derives a filter to compute the unconditional likelihood function in case $\rho = 1$. Lam (1990) adjusts this computationally intensive filter for the $\rho < 1$ case, see also Kim (1994). Fortunately, in our Bayesian analysis we can circumvent the evaluation of the unconditional likelihood function, see Section 3.4.

Finally, the unconditional likelihood function of the Markov trend model with $\rho = 1$ (3.9) denoted by $\mathcal{L}_0(y^T|y^k, \theta_0)$ is given by

$$\mathcal{L}_0(y^T|y^k, \theta_0) = \mathcal{L}(y^T|y^k, \theta)|_{\rho=1} \tag{3.15}$$

where $\theta_0 = \theta \backslash \{\rho, n_1\}$.

In the next section we propose a Bayesian framework to analyse the stochastic trends in Markov trend models. We specify prior distributions and discuss posterior odds ratios for the presence of a unit root in the polynomial $\phi(x)$ versus stationarity and for the absence of a Markov trend, *i.e.* $\gamma_1 = 0$.

3.3 Bayesian Analysis

Classical inference concerning the stochastic trends in Markov trend models turns out to be far from standard. The distribution of a likelihood ratio [LR] test for the presence of a unit root in the autoregressive component ($\rho = 1$) is unknown and technical difficulties are such that no formal strategy is available for a classical test procedure. Instead, Lam (1990) performs a small Monte Carlo experiment (100 trials) to obtain a critical value for an LR test for $\rho = 1$ in his model for US real GNP. At a 5% significance level the presence of a unit root is rejected. The same simulation experiment shows that, like in the standard AR model [see, *e.g.* Dickey and Fuller (1979)], the estimator of the dominant root of the autoregressive component is downward biased. Furthermore, a standard LR test for $\gamma_1 = 0$ is not possible since under $\gamma_1 = 0$ the transition probabilities p and q are not identified. Hansen (1992, 1996) provides a very computationally intensive procedure to perform a classical test for $\gamma_1 = 0$, see also Garcia (1995).

In this section we provide a Bayesian framework to analyse the Markov trend model. The Bayesian analyses of the non-Markov models proceed in a similar way. The discussed theory is limited to the specific problems under consideration. For a more general introduction in Bayesian analysis, we refer to Zellner (1971) or Box and Tiao (1973). In Section

3.3.1 we specify priors for the model parameters. In Section 3.3.2 and Section 3.3.3 we derive posterior odds ratios for the presence of a unit root and for the absence of the Markov trend.

3.3.1 Prior Specification

The Markov trend model (3.8) is non-linear in certain parameters. This phenomenon often leads to local non-identification for certain parameters in the model. It is easy to see that for $\rho = 1$, the parameter n_1 is not identified. Specifying a diffuse prior on n_1 implies that the conditional posterior of n_1 given ρ is constant and non-zero when $\rho = 1$. The integral over this conditional posterior at $\rho = 1$ is therefore infinity, favouring the unit root. Hoek and Paap (1994) correct for this problem by imposing a conditional normal prior on n_1 with mean y_1 and variance $\sigma^2/(1-\rho^2)$, which is based on the ideas of Schotman and van Dijk (1991a, 1991b). This prior reflects the uncertainty about n_1 if ρ approaches one by letting the variance of n_1 going to infinity if $\rho \to 1$. This solution provides a proper marginal posterior for ρ. Here, we specify a simplified version of this prior for n_1

$$n_1 \mid y_1, \sigma \sim \mathrm{N}(y_1, \sigma^2), \tag{3.16}$$

where y_1 is the first observation. The advantage of this simple prior specification is that the computation of the marginal posterior of ρ becomes easier, see Section 3.4. Furthermore, Hoek (1997, Chapter 2) shows that this simplification leads to almost the same answers in standard unit root analysis. Although this prior is more flexible since it allows for values of ρ larger than one, we specify a uniform prior distribution for ρ on the interval $[\rho_{lb}, 1]$ to exclude explosive AR structures

$$p(\rho) = \frac{1}{(1-\rho_{lb})}\mathbb{I}_{[\rho_{lb},1]}, \tag{3.17}$$

where $\mathbb{I}$ is an indicator function, which is one on $[\rho_{lb}, 1]$ and zero elsewhere. The value of ρ_{lb} will be discussed in the next subsection.

As already mentioned in the previous section the transition parameters p and q are not identified if $\gamma_1 = 0$. In a Bayesian framework this does not cause any problems as long as we specify proper priors on the non-identified parameters. The prior distributions for p and q are independent and uniform on the open interval $(0, 1)$

$$\begin{aligned} p(p) &= \mathbb{I}_{(0,1)} \\ p(q) &= \mathbb{I}_{(0,1)}. \end{aligned} \tag{3.18}$$

Under flat priors for p and q special attention must be payed to the priors for γ_0 and γ_1. It is easy to show that the likelihood has the same value if we switch the role of the states and change the values of γ_0, γ_1, p and q into $\gamma_0 + \gamma_1$, $-\gamma_1$, q and p respectively. This complicates proper posterior analysis if we specify uninformative priors on γ_0 and γ_1.

A simple solution to this problem is to restrict the intervals on which the priors have probability mass

$$p(\gamma_0) = \frac{1}{(\gamma_{0,ub} - \gamma_{0,lb})} \mathbb{I}_{[\gamma_{0,lb}, \gamma_{0,ub}]} \tag{3.19}$$

$$p(\gamma_1) = \frac{1}{(\gamma_{1,ub} - \gamma_{1,lb})} \mathbb{I}_{[\gamma_{1,lb}, \gamma_{1,ub}]}, \tag{3.20}$$

where $\gamma_{i,lb}$ and $\gamma_{i,ub}$, $i = 0, 1$ denote the upper- and lowerbound of the intervals. Another option is to specify informative priors for γ_0 and γ_1, for instance normal priors, see Diebolt and Robert (1994) for a discussion on prior specification in mixture models.

Finally, the priors for σ and $\bar{\phi}$ are given by

$$p(\sigma) \propto \sigma^{-1} \tag{3.21}$$

$$p(\bar{\phi}) \propto \mathbb{I}_{[stat]}, \tag{3.22}$$

where $\mathbb{I}_{[stat]}$ is an indicator function, which is one if the autoregressive parameters ϕ imply that the roots of the autoregressive polynomial are outside the unit circle and zero elsewhere.

The joint prior $p(\theta)$ is given by the product of the marginal priors (3.16)–(3.22). The joint prior for the model parameters under the restriction $\rho = 1$, $p_0(\theta_0)$, is just the product of the marginal priors (3.18)–(3.22).

3.3.2 Unit Root Analysis

Several methods for a Bayesian analysis of unit roots have been proposed. The differences result from prior specification, the initial value problem and model representation, see for an overview Hoek (1997, Section 2.2) and Schotman (1994). Here we follow a standard posterior odds ratio analysis to analyse the presence of a unit root. Since our null hypothesis, $H_0 : \rho = 1$, is a sharp hypothesis, we have to assign a discrete prior probability to the event $\rho = 1$. The alternative hypothesis is $H : \rho < 1$. The relative prior beliefs about the two hypotheses is given by the prior odds ratio [PROR]

$$\text{PROR}(\rho) = \frac{\Pr[H_0]}{\Pr[H]}, \tag{3.23}$$

where $\Pr[H_0]$ and $\Pr[H]$ denote the prior probabilities for both hypotheses. The posterior odds ratio is the product of the prior odds ratio and the ratio of the averaged likelihoods with the priors under the competing hypotheses serving as weighting functions see, *e.g.*, Leamer (1978, Chapter 4) and Zellner (1971, Chapter X). Formally, the posterior odds ratio [POR] is defined as

$$\begin{aligned} \text{POR}(\rho) &= \text{PROR}(\rho) \times \frac{\Pr[H_0|y^T]}{\Pr[H|y^T]} \\ &= \frac{\Pr[H_0]}{\Pr[H]} \times \frac{\int p_0(\theta_0) \mathcal{L}_0(y^T|y^k, \theta_0)\, d\theta_0}{\int p(\theta) \mathcal{L}(y^T|y^k, \theta)\, d\theta}, \end{aligned} \tag{3.24}$$

where $\mathcal{L}(y^T|y^k,\theta)$ and $\mathcal{L}_0(y^T|y^k,\theta_0)$ denote the unconditional likelihood functions given in (3.14) and (3.15), respectively. If the PROR(ρ)=1, *i.e.* both hypotheses are *a priori* equally likely, the POR(ρ) equals the Bayes factor [BF], which corresponds to the last term in (3.24).

Computation of the Bayes factor requires the evaluation of two integrals. Fortunately, Dickey (1971) shows that the Bayes factor also equals the ratio of the height of the marginal posterior and the height of the marginal prior, both evaluated in the point of interest ($\rho = 1$)

$$\mathrm{BF}(\rho) = \frac{p(\rho|y^T)|_{\rho=1}}{p(\rho)|_{\rho=1}}, \tag{3.25}$$

where $p(\rho|y^T)$ denotes the marginal posterior for ρ. This ratio is known as the Savage-Dickey density ratio. Two regularity conditions are required for the use of the Savage-Dickey density ratio, see also Verdinelli and Wasserman (1995). The total prior under the alternative hypothesis must equal the conditional prior under the null hypothesis in the point of interest, *i.e.*,

$$p(\theta_0|\rho)|_{\rho=1} = p_0(\theta_0). \tag{3.26}$$

Furthermore, the marginal posterior of the parameter of interest and the total prior of all parameters must be bounded in the parameter point of interest. In our case the conditions are

$$0 < p(\rho|y^T)|_{\rho=1} < \infty \tag{3.27}$$

$$0 < p(\theta)|_{\rho=1} < \infty. \tag{3.28}$$

The prior specification in the previous subsection ensures that condition (3.26) is satisfied. Since we have specified a proper prior on n_1 it is easy to see that condition (3.27) is also fulfilled. Although condition (3.28) is formally not satisfied in our prior specification, we can construct theoretically equal uniform priors for the parameters $\bar{\phi}$ and $\ln(\sigma)$ in both competing models such that condition (3.28) is satisfied with the same posterior results as for the improper priors.

It is easily seen from (3.25) that the role of the prior on ρ is important. Under our uniform prior for ρ the value of ρ_{lb} is important. If we choose ρ_{lb} too small the POR becomes very large favouring the unit root hypothesis. We follow the solution by Schotman and van Dijk (1991a), who choose ρ_{lb} such that $[\rho_{lb}, 1]$ corresponds to the 99% highest posterior density [HPD] region.

3.3.3 Markov Trend Analysis

Testing for the absence of a Markov trend can be done in the same way. Koop and Potter (1995) use the Savage-Dickey density ratio to test for $\gamma_1 = 0$ versus $\gamma_1 \neq 0$. The Bayes

factor for this hypothesis simplifies to

$$\mathrm{BF}(\gamma_1) = \frac{p(\gamma_1|y^T)|_{\gamma_1=0}}{p(\gamma_1)|_{\gamma_1=0}} \tag{3.29}$$

where $p(\gamma_1|y^T)$ is the marginal posterior of γ_1 and $p(\gamma_1)$ the prior for γ_1, defined in (3.19), see Koop and Potter (1995) for more details. The priors for the models under $\gamma_1 = 0$ have to satisfy a condition like (3.26). For the other conditions the same arguments as for the test for $\rho = 1$ hold. If one wants to specify a flat prior on γ_1, a HPD region for γ_1 is required to avoid favouring the hypothesis $\gamma_1 = 0$.

To compute the Bayes factors we need the marginal posterior density of ρ and γ_1. In the next section we use Markov chain Monte Carlo [MCMC] methods to obtain posterior results.

3.4 Simulating Posterior Distributions

The posterior distribution is proportional with the product of the marginal priors in (3.16)–(3.22) and the unconditional likelihood in (3.14). This posterior distribution is too complicated to derive analytical posterior results. As Albert and Chib (1993), McCulloch and Tsay (1994b) and Chib (1996) demonstrate, the Gibbs sampling algorithm of Geman and Geman (1984) is very useful tool for the computation of posterior results for models with unobserved states. The state variables $\{s_t\}_{t=1}^T$ can be treated as unknown parameters and simulated alongside the model parameters. This technique is known as data augmentation, see Tanner and Wong (1987).

To describe the Gibbs sampler, let ψ be a random vector which can be divided in d blocks $(\psi_1, \ldots, \psi_j, \ldots, \psi_d)$, Also, let $\psi_j \mid \psi_{-j}$ denote the distribution of ψ_j conditional on the other random variables $\psi_{-j} = \psi \backslash \psi_j$. The simulation algorithm to sample from the joint distribution of ψ works as follows:

Step 1: Specify starting values $\psi^0 = (\psi_1^0, \ldots, \psi_j^0, \ldots, \psi_d^0)$ and set $i = 0$.

Step 2: Simulate

$$\begin{aligned}
&\psi_1^{i+1} \text{ from } \psi_1 \mid \psi_2^i, \psi_3^i, \ldots, \psi_d^i, \\
&\psi_2^{i+1} \text{ from } \psi_2 \mid \psi_1^{i+1}, \psi_3^i, \ldots, \psi_d^i, \\
&\psi_3^{i+1} \text{ from } \psi_3 \mid \psi_1^{i+1}, \psi_2^{i+1}, \psi_4^i, \ldots, \psi_d^i, \\
&\quad\vdots \qquad\qquad\qquad\qquad \vdots \\
&\psi_d^{i+1} \text{ from } \psi_d \mid \psi_1^{i+1}, \psi_2^{i+1}, \ldots, \psi_{d-1}^{i+1}.
\end{aligned}$$

Step 3: Set $i = i + 1$ and go to step 2.

The described iterative scheme generates a Markov chain. After the chain has converged, say at H iterations, the simulated values $\{\psi^i, i \geq H\}$ can be used as a sample from the joint distribution of ψ to compute means, variances and marginal densities. For instance, the marginal density of ψ_j is obtained by

$$p(\psi_j) = \frac{1}{N} \sum_{i=H+1}^{N+H} p(\psi_j | \psi_1^i, \ldots, \psi_{j-1}^i, \psi_{j+1}^i, \ldots, \psi_d^i), \tag{3.30}$$

where N denotes the number of useful draws. For an overview and some details on convergence, see among others Smith and Roberts (1993) and Tierney (1994).

In the remainder of this section, we show how the Gibbs sampler can be implemented to obtain posterior results from the Markov trend model. To apply the Gibbs sampler we need the full conditional posterior densities of the model parameters, which play the role of $\psi_j|\psi_{-j}$ in step 2. Additionally, the full conditional posterior distributions of the state variables are needed. We focus on the unrestricted Markov trend model (3.8) with $\rho < 1$. The full conditional posterior densities of the models (3.9)–(3.11) can be derived in a similar way. The Gibbs sampling scheme described above results in draws θ^i from the posterior distribution, which can be used to compute posterior means, variances and marginal densities.

3.4.1 Full Conditional Posterior Distributions

Full Conditional Posterior of the States

To sample the states, we need the full conditional posterior density of s_t, denoted by $p(s_t|s^{-t}, \theta, y^T)$, $t = 1, \ldots, T$, where $s^{-t} = s^T \backslash \{s_t\}$. Since s_t follows a first-order Markov process, it is easily seen that

$$p(s_t|s^{-t}) \propto p(s_t|s_{t-1})\, p(s_{t+1}|s_t), \tag{3.31}$$

due to the Markov property. Following Albert and Chib (1993), we can write

$$\begin{aligned} p(s_t|s^{-t}, \theta, y^T) &= \frac{p(s_t|s^{-t}, \theta, y^t)\, f(y_{t+1}, \ldots, y_T|y^t, s^{-t}, s_t, \theta)}{f(y_{t+1}, \ldots, y_T|y^t, s^{-t}, \theta)} \\ &\propto p(s_t|s^{-t}, \theta, y^t)\, f(y_{t+1}, \ldots, y_T|y^t, s^{-t}, s_t, \theta). \end{aligned} \tag{3.32}$$

Using the rules of conditional probability, the first term of (3.32) can be simplified as

$$\begin{aligned} p(s_t|s^{-t}, \theta, y^t) &\propto p(s_t|s^{-t}, \theta, y^{t-1})\, f(y_t, s_{t+1}, \ldots, s_T|y^{t-1}, s^t, \theta) \\ &\propto p(s_t|s_{t-1}, \theta)\, f(y_t|y^{t-1}, s^t, \theta) \\ &\qquad p(s_{t+1}|s^t, \theta, y^t)\, p(s_{t+2}, \ldots, s_T|s^{t+1}, \theta, y^t) \\ &\propto p(s_t|s_{t-1}, \theta)\, f(y_t|y^{t-1}, s^t, \theta)\, p(s_{t+1}|s_t, \theta), \end{aligned} \tag{3.33}$$

where we use the fact that $\{s_{t+2}, \ldots, s_T\}$ is independent of s_t given s_{t+1}. The second term of (3.32) is proportional to

$$f(y_{t+1}, \ldots, y_T | y^t, s^t, \theta) \propto \prod_{i=t+1}^{T} f(y_i | y^{i-1}, s^i, \theta). \tag{3.34}$$

Next, using (3.33) and (3.34) the full conditional distribution of s_t for $t = k+1, \ldots, T$ is given by

$$p(s_t | s^{-t}, \theta, y^T) \propto p(s_t | s_{t-1}, \theta)\, p(s_{t+1} | s_t, \theta) \prod_{i=t}^{T} f(y_i | y^{i-1}, s^i, \theta), \tag{3.35}$$

where $f(y_t | y^{t-1}, s^t, \theta)$ is defined in (3.12) and the constant of proportionality can be obtained by summing over the two possible values of s_t. At time $t = T$ the term $p(s_{T+1} | s_T, \theta)$ drops out. The first k states can be sampled from the full conditional distribution

$$p(s_t | s^{-t}, \theta, y^T) \propto p(s_t | s_{t-1}, \theta)\, p(s_{t+1} | s_t, \theta) \prod_{i=k+1}^{T} f(y_i | y^{i-1}, s^i, \theta), \tag{3.36}$$

for $t = 1, \ldots, k$, where at time $t = 1$ the term $p(s_t | s_{t-1}, \theta)$ is replaced by the unconditional density $p(s_1 | \theta)$, which is a binomial density with probability $\pi = (1-p)/(2-p-q)$.

As Albert and Chib (1993) show, sampling of the state variables if a unit root is present in $\phi(x)$ is easier. Under the restriction $\rho = 1$ only the first $(k-1)$ future conditional densities of y_t depend on s_t instead of all future conditional densities. However, sampling is possible in the same way: take the most recent value of s^T and sample the states backward in time, one after another, starting with s_T. After each step, the t-th element of s^T is replaced by its most recent draw.

Full Conditional Posterior of p and q

It is easy to see from the conditional likelihood function (3.13) that the full conditional posterior densities of the transition parameters are given by

$$p(p | s^T, \theta \backslash \{p\}, y^T) \propto p^{\mathcal{N}_{00}} (1-p)^{\mathcal{N}_{01}} \tag{3.37}$$

and

$$p(q | s^T, \theta \backslash \{q\}, y^T) \propto q^{\mathcal{N}_{11}} (1-q)^{\mathcal{N}_{10}}, \tag{3.38}$$

where $\mathcal{N}_{ij}$ again denotes the number of transitions from state i to state j. This implies that the transition probabilities can be sampled from beta distributions.

Full Conditional Posterior of γ_0, γ_1 and n_1

To derive the full conditional posterior distribution of γ_0, γ_1 and n_1 we rewrite (3.8) in a linear regression equation representation

$$\phi(L)y_t = \phi(L)\begin{pmatrix} t-1 & \sum_{i=2}^{t} s_t & 1 \end{pmatrix}\begin{pmatrix} \gamma_0 \\ \gamma_1 \\ n_1 \end{pmatrix} + \epsilon_t, \tag{3.39}$$

for $t = k+1, \ldots, T$. Using the properties of the likelihood of a regression model it is easy to show that the full conditional density of $(\gamma_0 \ \gamma_1 \ n_1)'$ is normal, see *e.g.* Zellner (1971, p. 65–67). Note that we also have to add the normal prior for n_1 (3.16) to the regression equation (3.39) via $y_1 = (0 \ 0 \ 1)(\gamma_0 \ \gamma_1 \ n_1)' + \epsilon_1$. Normal priors for γ_0 and/or γ_1 can be included in the same way.

Full Conditional Posterior of ρ and $\bar{\phi}$

If we condition on γ_0, γ_1 and n_1 and the states $\{s_t\}_{t=1}^{T}$, model (3.8) can be seen as a regression model in the parameters ρ and $\bar{\phi}$. Using the same argument as above the full conditional posterior distribution of $(\rho, \bar{\phi})$ is normal. The restriction $\rho_{lb} \leq \rho \leq 1$ can be incorporated by rejecting draws for which $\rho > 1$ or by drawing from a truncated normal distribution.

Full Conditional Posterior of σ

To derive the full conditional posterior distribution of σ we consider the conditional likelihood function (3.13). This function is proportional to an inverted gamma-2 density, so the sampling of σ can be based on

$$\frac{\sum_{t=k+1}^{T} \epsilon_t^2 + (y_1 - n_1)^2}{\sigma^2} \ \Big| \ \theta \backslash \{\sigma\}, y^T \sim \chi^2(T-k+1), \tag{3.40}$$

see Zellner (1971, p. 61–62). The term $(y_1 - n_1)^2$ results from the normal prior specification on n_1 (3.16).

3.5 Forecasting

Time series models are not only used to analyse trends, business cycles, seasonal patterns within-sample, but also to generate out-of-sample forecasts. In a Bayesian analysis this implies the analysis of predictive densities. The one-step ahead predictive density $f(y_{T+1}|y^T)$ for the Markov trend model (3.8) results from

$$f(y_{T+1}|y^T) = \iint f(y_{T+1}|y^T, s_{T+1}, s^T, \theta)\, ds_{T+1}\, d\theta, \tag{3.41}$$

where $f(y_{T+1}|y^T, s_{T+1}, s^T, \theta)$ is given in (3.12). Note that contrary to standard classical approaches this predictive density incorporates state uncertainty and parameter uncertainty. The computation of the one-step ahead predictive density $f(y_{T+1}|y^T)$ can easily been done by extending the sampling scheme from the previous section, see Albert and Chib (1993). To obtain posterior results we simulate in each iteration of the Gibbs sampling procedure from the full conditional distribution of s^{T+1} and y^{T+1}. In other words, for each draw (s^T, θ), which results from the Gibbs sampler we generate

- s_{T+1} from $p(s_{T+1}|s^T, \theta)$ and
- y_{T+1} from $f(y_{T+1}|y^T, s_{T+1}, s^T, \theta)$ given in (3.12).

These extra steps can be implemented in the Gibbs sampling scheme without any difficulties. The same is true for the computation of h-step ahead predictive densities. The h-step ahead predictive density $f(y_{T+h}|y^T)$ is given by

$$f(y_{T+h}|y^T) = \iiint f(y_{T+h}|y^{T+h-1}, s^{T+h}, \theta)\, d(y_{T+h-1}, \ldots, y_{T+1})\, d(s_{T+h}, \ldots, s_{T+1})\, d\theta, \tag{3.42}$$

where $f(y_{T+h}|y^{T+h-1}, s^{T+h}, \theta)$ is given in (3.12). For the computation of the h-step ahead predictive density we use the simulation results for the computation of the one- through $(h-1)$-step ahead predictive densities as the following simulation scheme shows:

- Generate s_{T+h} from $p(s_{T+h}|s_{T+h-1}, \theta)$, where s_{T+h-1} results from the simulation of the $(h-1)$-step ahead predictive density.
- Generate y_{T+h} from $f(y_{T+h}|y^{T+h-1}, s^{T+h}, \theta)$ given in (3.12), where $\{s_{T+h-1}, \ldots, s_{T+1}\}$ and $\{y_{T+h-1}, \ldots, y_{T+1}\}$ are draws resulting from the simulation of the 1- through $(h-1)$-step ahead predictive densities.

The simulation output can be used to compute predictive means, variances and the predictive densities like in (3.30).

3.6 Application

To illustrate the proposed Bayesian analysis of a unit root in a Markov trend model we first analyse in Section 3.6.1 two simulated series. In Section 3.6.2 we consider German industrial production.

3.6.1 Simulated Series

We consider the following data generating process [DGP],

$$\begin{aligned} y_t &= n_t + z_t, \\ n_t &= n_{t-1} + 2 - 4\, s_t, \qquad & n_1 &= 0, \\ z_t &= \rho z_{t-1} + \epsilon_t, \qquad & \epsilon_t &\sim \text{NID}(0,1), \end{aligned} \tag{3.43}$$

Table 3.1. Posterior means with posterior standard deviations between parentheses of the model parameters in (3.8) with Bayes factors for $\rho = 1$, for the two DGPs.

series[1]	γ_0	γ_1	n_1	p	q	ρ	σ	BF(ρ)[2]
DGP I	1.98 (0.04)	−4.05 (0.24)	0.57 (0.79)	0.90 (0.03)	0.60 (0.11)	0.79 (0.09)	0.88 (0.07)	0.61
DGP II	1.94 (0.06)	−3.81 (0.23)	0.34 (0.91)	0.89 (0.04)	0.58 (0.11)	0.94 (0.06)	0.85 (0.07)	3.00

[1] The DGP is given in (3.43) with $\rho = 0.8$ (DGP I) and $\rho = 1$ (DGP II).
[2] BF(ρ) denotes the Bayes factor. A Bayes factor exceeding one implies that $\rho = 1$ is *a posteriori* more likely than $\rho < 1$.

where $\{s_t\}_{t=1}^{100}$ is generated by a first-order Markov process with transition probabilities $p = 0.9$ and $q = 0.6$. We analyse two DGPs using the same set of disturbances $\{\epsilon_t\}_{t=1}^{100}$ and state variables $\{s_t\}_{t=1}^{100}$ but with different values for the autoregressive coefficient, $\rho = 0.8$ and $\rho = 1$ respectively. The simulated series are denoted by DGP I and DGP II and are plotted in the top left cells of Figure 3.1 and 3.2.

To analyse the two series we consider the Markov trend model (3.8) with a first-order AR component, which corresponds to the AR order of the DGPs. The priors for the model parameters are given in (3.16)–(3.21). For γ_0 and γ_1 we take flat priors on the intervals $[1, \infty)$ and $(-\infty, 0]$ respectively. Table 3.1 displays the posterior means and posterior standard deviations between parentheses of the model parameters for both DGPs. Posterior results are obtained using the Gibbs sampling algorithm explained in Section 3.4. The posterior means of the parameters match the parameter values of the DGPs. The posterior results are almost identical for the two DGPs except of course for ρ. The posterior mean of ρ for DGP I is 0.79, which is clearly below unity. For DGP II we have a posterior mean for ρ of 0.94. Note that due to truncation to the right the mode of the marginal posterior density is closer to one. The final column of Table 3.1 shows the Bayes factors for $\rho = 1$, which are computed using the Savage-Dickey density ratio (3.25). The value ρ_{lb} has been chosen such that $[\rho_{lb}, 1]$ corresponds to the 99% HPD region for ρ. Under equal prior odds, these Bayes factors are equal to posterior odds ratios for $\rho = 1$. For DGP I the Bayes factor for $\rho = 1$ equals 0.61, which indicates that the hypothesis of $\rho < 1$ is *a posteriori* more likely than the hypothesis $\rho = 1$. However, the unit root hypothesis is strongly favoured for the second DGP: the Bayes factor 3.00 clearly exceeds unity. Note that the posterior standard deviation of n_1 for the Markov trend stationary DGP I is smaller than for DGP II, which contains a unit root.

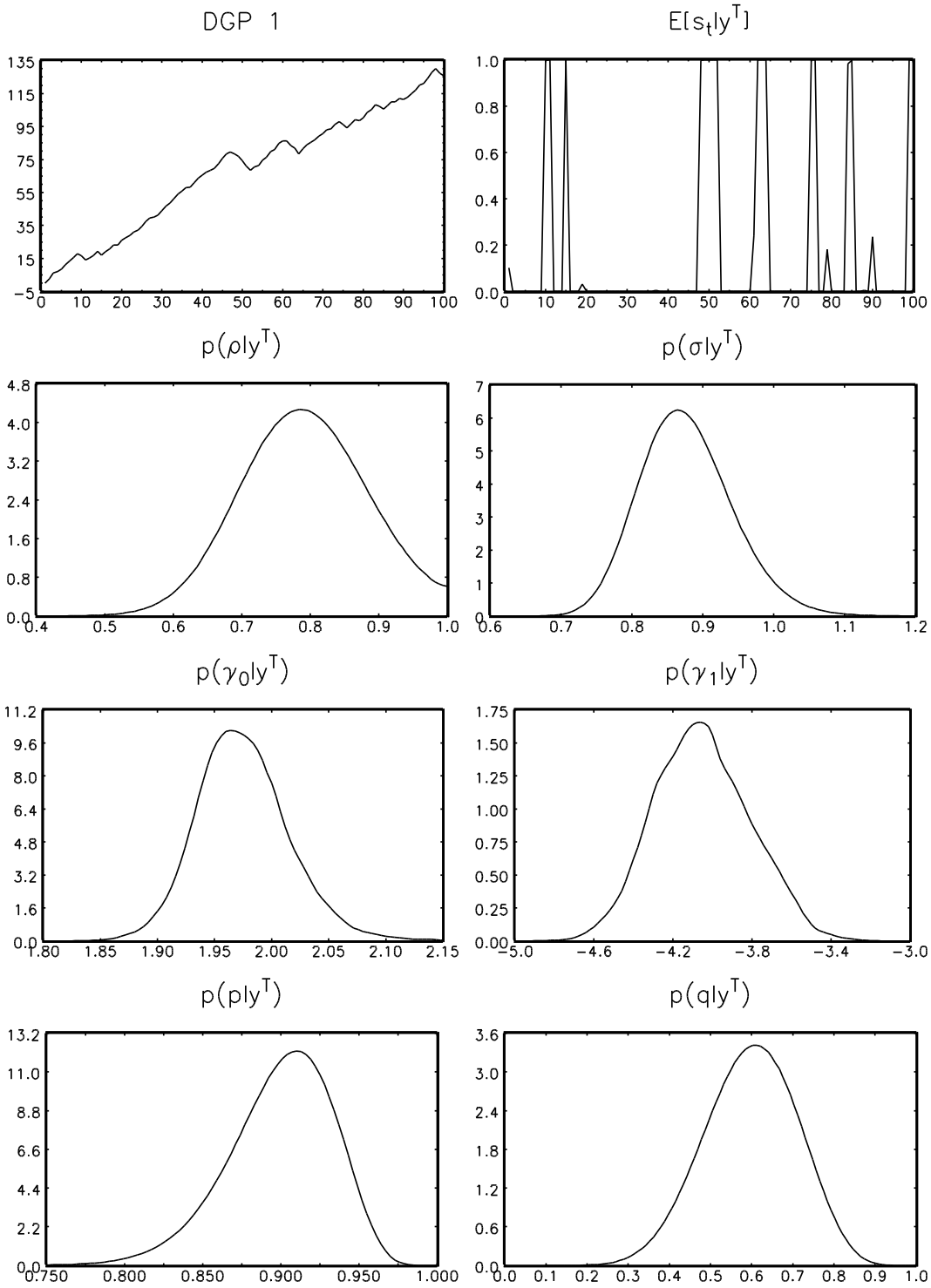

Figure 3.1. Simulated series and marginal posterior densities for DGP I.

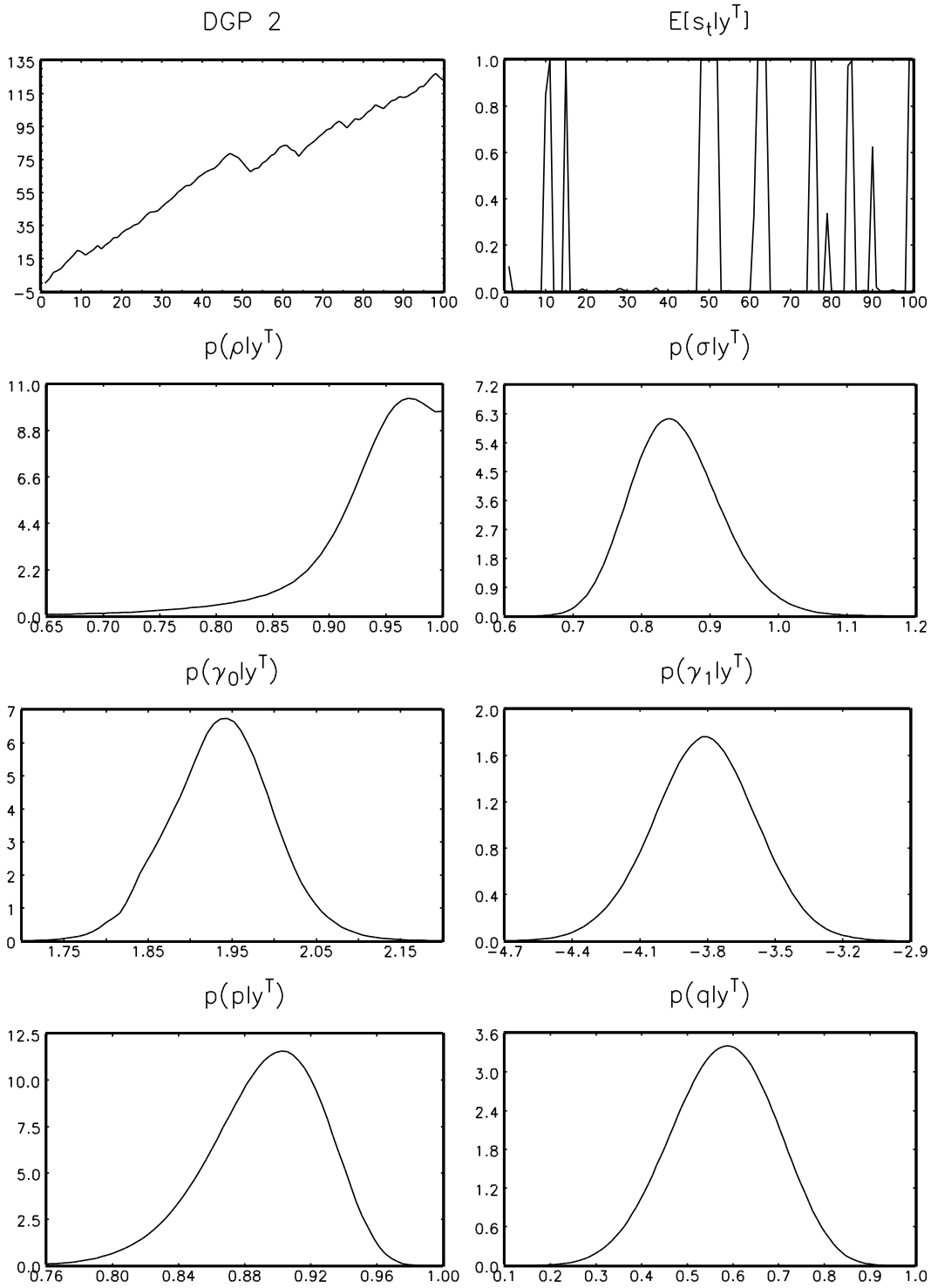

Figure 3.2. Simulated series and marginal posterior densities for DGP II.

Figures 3.1 and 3.2 display the marginal posterior densities of the model parameters. The shapes of the marginal posterior densities of ρ support the results of the Bayes factors. For DGP I the density mass is situated far away from one, while for DGP II the mode is very close to one. The top right cells of the figures denote the posterior expectations of the states $\mathrm{E}[s_t|y^T]$, $t = 1, \ldots, 100$. The peaks in these graph correspond reasonably well with the periods of negative growth. The posterior distributions of γ_0 and γ_1 and the transition probabilities reflect the fact that the DGPs consist of a relatively small number of observations with $s_t = 1$ and many observations with $s_t = 0$. The posterior variances of q are much larger than the posterior variances of p. The same is true for the posterior variances of γ_1 and γ_0. Finally, the marginal posteriors of γ_1 show that Bayes factors for $\gamma_1 = 0$ are zero for both DGPs, since there is no probability mass in $\gamma_1 = 0$.

The simulation experiment gives some indication of the practical usefulness of Bayesian unit root analysis in Markov trend models. Since only two simulated series have been considered, it must be emphasized that no general conclusion can be drawn about the performance of the approach. In the next subsection we analyse seasonally adjusted quarterly observed German industrial production series.

3.6.2 German Industrial Production

In this subsection we consider quarterly observed seasonally adjusted industrial production of Germany for the period 1957.I–1993.IV. The data source is the International Financial Statistics. The top left cell of Figure 3.3 shows a plot of the series. Since the industrial production series is an index and does not exhibit exponential growth, we do not apply a logarithmic transformation. There are three major periods of negative growth in the industrial production 1966.III–1967.II, 1973.III–1975.III and 1992.II–1993.II. Also in the period 1980.II–1982.IV there is an overall tendency of decrease in German industrial production. The periods of positive growth are much longer than the periods of negative growth, which points to an asymmetric cycle in the series. Furthermore, the average increase during a positive growth regime is smaller than the average decrease during a negative growth regime. A Markov trend model seems to be a suitable model to analyse this series.

First, we consider the Markov trend model without the unit root in the autoregressive component. The priors for the model parameters are given by (3.16)–(3.22). For γ_0 and γ_1 we take flat priors on the intervals $[0.2, \infty)$ and $(-\infty, 0]$ respectively. This ensures that $s_t = 0$ corresponds to an expansion regime and $s_t = 1$ corresponds to a contraction regime. The lag order k of the model is determined using Bayes factors tests. We start with an AR model of order six and decrease the order with one until the Bayes factor for the zero restriction on the highest order $\bar{\phi}_j$ coefficient is smaller than one.[3]

The first column of Table 3.2 shows the posterior results for the Markov trend stationary model. The posterior mean of the slope of the Markov trend is 0.70 during the

[3]The Bayes factors are computed using a Savage-Dickey density ratio like in (3.25) and based on 99% HPD regions.

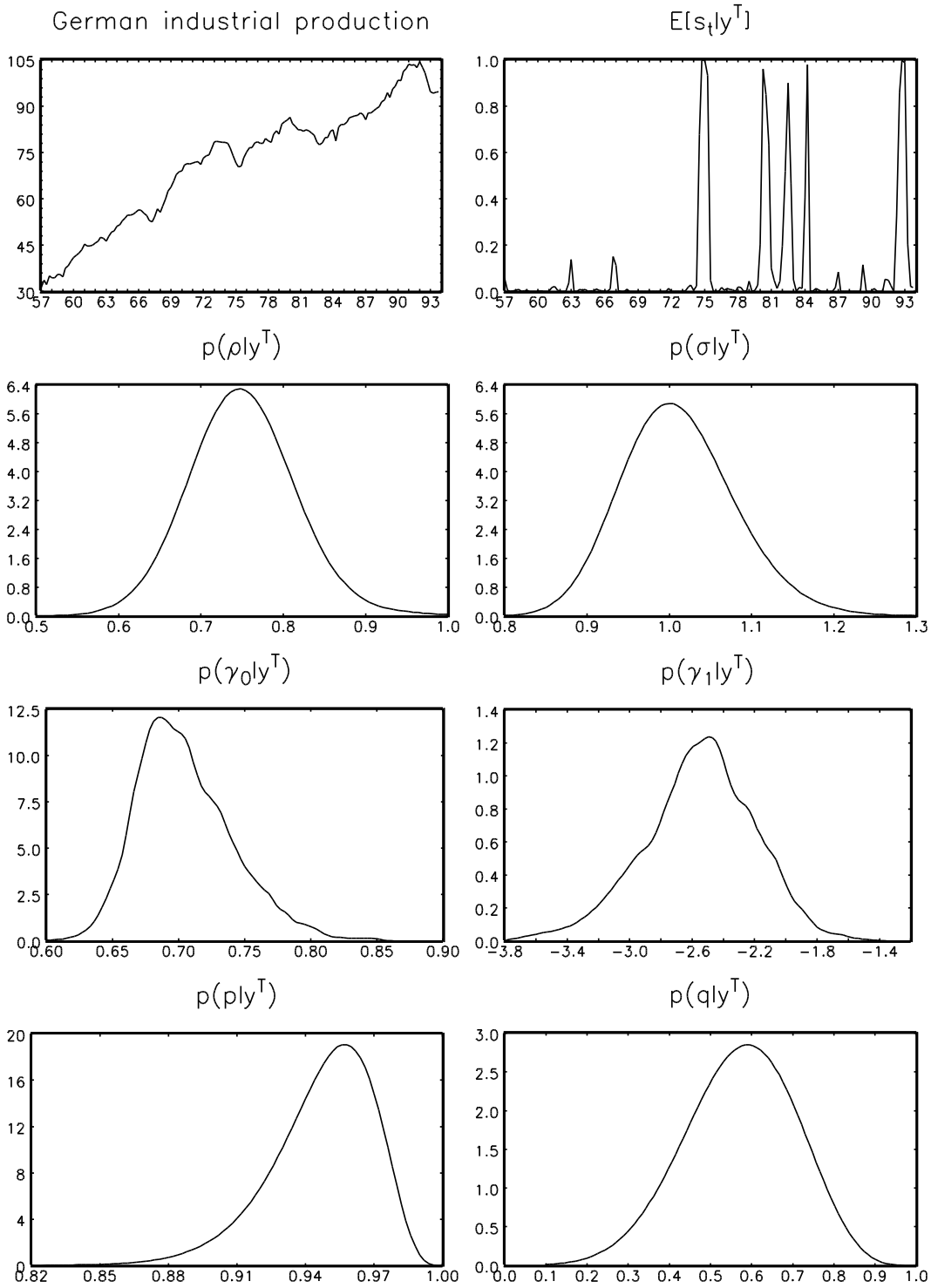

Figure 3.3. German industrial production and marginal posterior densities.

Table 3.2. Posterior means with posterior standard deviations between parentheses and Bayes factors for $\rho = 1$ for German industrial production, 1957.I–1993.IV.

para-	MTS[1]		MDS		TS		DS	
meters	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.
γ_0	0.70	(0.04)	0.72	(0.17)	0.47	(0.08)	0.44	(0.18)
γ_1	−2.56	(0.36)	−2.31	(0.56)	0	–	0	–
n_1	31.41	(0.79)	–	–	31.46	(1.35)	–	–
p	0.95	(0.02)	0.92	(0.06)	–	–	–	–
q	0.57	(0.14)	0.53	(0.16)	–	–	–	–
ρ	0.75	(0.07)	1	–	0.98	(0.02)	1	–
$\bar{\phi}_1$	0.02	(0.09)	−0.01	(0.11)	0.14	(0.08)	0.14	(0.09)
$\bar{\phi}_2$	0.19	(0.09)	0.16	(0.10)	0.16	(0.08)	0.16	(0.09)
$\bar{\phi}_3$	0.35	(0.10)	0.24	(0.10)	0.14	(0.08)	0.16	(0.08)
$\bar{\phi}_4$	0.29	(0.09)	0.10	(0.10)	0	–	0.03	(0.09)
$\bar{\phi}_5$	0	–	−0.23	(0.10)	0	–	−0.18	(0.09)
σ	1.01	(0.07)	1.08	(0.09)	1.32	(0.08)	1.33	(0.08)
BF(ρ)[2]	0.02		–		1.74		–	

[1] MTS=Markov trend stationary model (3.8), MDS=Markov difference stationary model (3.9), TS=trend stationary model (3.10) and DS=difference stationary model (3.11).

[2] BF(ρ) denotes the Bayes factor. A Bayes factor exceeding one implies that $\rho = 1$ is *a posteriori* more likely than $\rho < 1$.

expansion regime and -1.86 $(-2.56 + 0.70)$ during a recession. The posterior mean of the probability of staying in the expansion regime is 0.95, which is larger than the posterior mean of the probability of staying in a recession 0.57. The posterior probability that p is larger than q is 0.99, which indicates an asymmetric business cycle. The transition probabilities determine the expected duration of recessions and expansions. The expected duration of staying in an expansion regime ($s_t = 0$) conditional on being in an expansion regime is

$$\sum_{i=1}^{\infty} i p^{i-1}(1-p) = \frac{1}{1-p}, \tag{3.44}$$

see Hamilton (1989). The posterior mean of this expected duration is 23.31 quarters.

Likewise, the expected duration of staying in a recession ($s_t = 1$) equals

$$\sum_{i=1}^{\infty} iq^{i-1}(1-q) = \frac{1}{1-q}. \tag{3.45}$$

The posterior mean of the expected duration of a recession is 2.64 quarters. There is clear evidence for an asymmetric business cycle. Note that there is again more posterior uncertainty for γ_1 and q than for γ_0 and p.

The posterior mean of ρ is 0.75, which is far away from one. The Bayes factor for $\rho = 1$ is 0.02, which indicates that *a posteriori* the unit root is not plausible. Hence, a Markov trend stationary model is more appropriate than a Markov difference stationary model. Figure 3.3 displays the marginal posterior densities for the most important parameters of the model. The marginal posterior of ρ supports the outcome of the Bayes factor for $\rho = 1$. There is very little probability mass in $\rho = 1$. The marginal posterior densities of γ_0 and γ_1 are shown in the third row in Figure 3.3. Since the height of the marginal posterior density of γ_1 in $\gamma_1 = 0$ is almost zero, the Bayes factor for $\gamma_1 = 0$ is also very close to zero. Hence, for the industrial production series a Markov trend stationary model is *a posteriori* more likely than a linear trend stationary model.

The top right cell of Figure 3.3 denotes the posterior expectations of the states $\mathrm{E}[s_t|y^T]$. Using the 0.5 rule as in Hamilton (1989) we can distinguish the two stages of the business cycle. We define a recession by two consecutive data points for which $\mathrm{E}[s_t|y^T] > 0.5$. A peak is defined by the last expansion observation before a recession. A trough is defined by the last observation in a recession. Using this rule we detect four recessions in German industrial production. The following quarters are labelled as peaks: 1974.II, 1980.I, 1982.I and 1992.II. The troughs are found in 1975.II, 1980.IV, 1982.III and 1993.I. Note that the states do not completely pick up the period of decrease in industrial production in the 1960s. This may be explained by the fact that the growth rate in the period just after this recession was larger than the growth rate after the other recessions and the level of industrial production returned to the same growth path as before the recession in the 1960s, see the top left cell of Figure 3.3. The Markov trend model considers this recession as a large temporary deviation from the Markov trend, since it did not have a permanent effect on the level of the series like the other periods of negative growth. Similar findings are reported in Sichel (1994) who suggest that there exist an extra phase in the business cycle of the United States. In this so-called recovery phase the series returns to its original growth path so that recessions do not have a permanent effect on the level of US GNP, see also Beaudry and Koop (1993).

Some alternative models for the German industrial production series are the Markov trend model with a unit root in the autoregressive component (3.9), the trend stationary model (3.10) and the difference stationary model (3.11). Note that these models are nested in the Markov trend stationary model but that Bayes factors indicate that neither of these models are preferred. Table 3.2 displays the posterior results for these three models. The lag order for these models is determined in the same way as for Markov trend stationary model. Also, the priors for the model parameters are the same as in the

Markov trend stationary model except that we take a diffuse prior for γ_0 on $(-\infty, \infty)$ for the models without Markov trend. The second column shows that the posteriors means of the parameters of the Markov difference stationary model are roughly the same as for the Markov trend stationary model except for the autoregressive parameters. The posterior standard deviations of the parameters are however larger. Note that the lag order of the model in first differences is larger than of the model in levels, probably due to overdifferentiation.

The final two columns of Table 3.2 display the posterior means for the trend stationary and difference stationary model. The posterior mean of the slope of the trend and the drift term are roughly 0.45. The posterior mean of the variance of the error process σ^2 is larger than for the Markov trend models. This leads to more forecast uncertainty due to the noise component. The last row of the table displays the Bayes factor for $\rho = 1$ in the trend stationary model. Contrary to the Markov trend model, the Bayes factor now favours the unit root hypothesis, indicating that all past shocks have a permanent effect on the future level of the series.

As Markov trend models are designed to model the recession in a time series one may expect that these models produce superior forecasts than standard AR models for recession periods. To evaluate the out-of-sample performance of our Markov trend models, we remove the last eight observations (1992.I–1993.IV) from the sample for a forecast exercise. Note that the posterior expectation of the states indicate a recession during this forecast evaluation period. We reanalyse the models again for the sample 1957.I–1991.IV and consider one- through eight-step ahead forecast distributions, see Section 3.5.

Table 3.3 displays the posterior means with posterior standard deviations between parentheses of the one- through eight-step ahead forecasts distributions together with the true values of the series. The results are obtained by extending the Gibbs sampling procedure with the extra steps mentioned in Section 3.5. Considering posterior means, we see that in 1992.I the trend stationary model, in 1992.II the difference stationary model and in 1992.III the Markov difference stationary model produces the best forecasts. After 1992.III the Markov trend stationary model is superior. Contrary to the non-linear AR models, the Markov models take into account the possibility of a recession, which occurred after 1992.II. The standard deviations of the forecast distribution increase if we forecast more periods ahead, although this does not have to be the case for non-linear models. Note that imposing the unit root leads to a larger standard deviation of the forecast distribution for the Markov difference stationary model than for the Markov trend stationary model. This is also the case for the non-Markov models. The smaller posterior variances of the disturbances of the Markov trend models than of the standard AR models lead to smaller standard deviations in the forecast distributions. This is even more clear from Figure 3.4, which shows the one- through eight-step ahead predictive error densities for the four models. These densities are obtained by a horizontal shift of the predictive densities of minus the true value. The predictive densities of the Markov trend models are unimodal but skewed. It is very clear from these figures that the Markov trend model produces superior forecast distributions after 1992.II. However in the first three periods

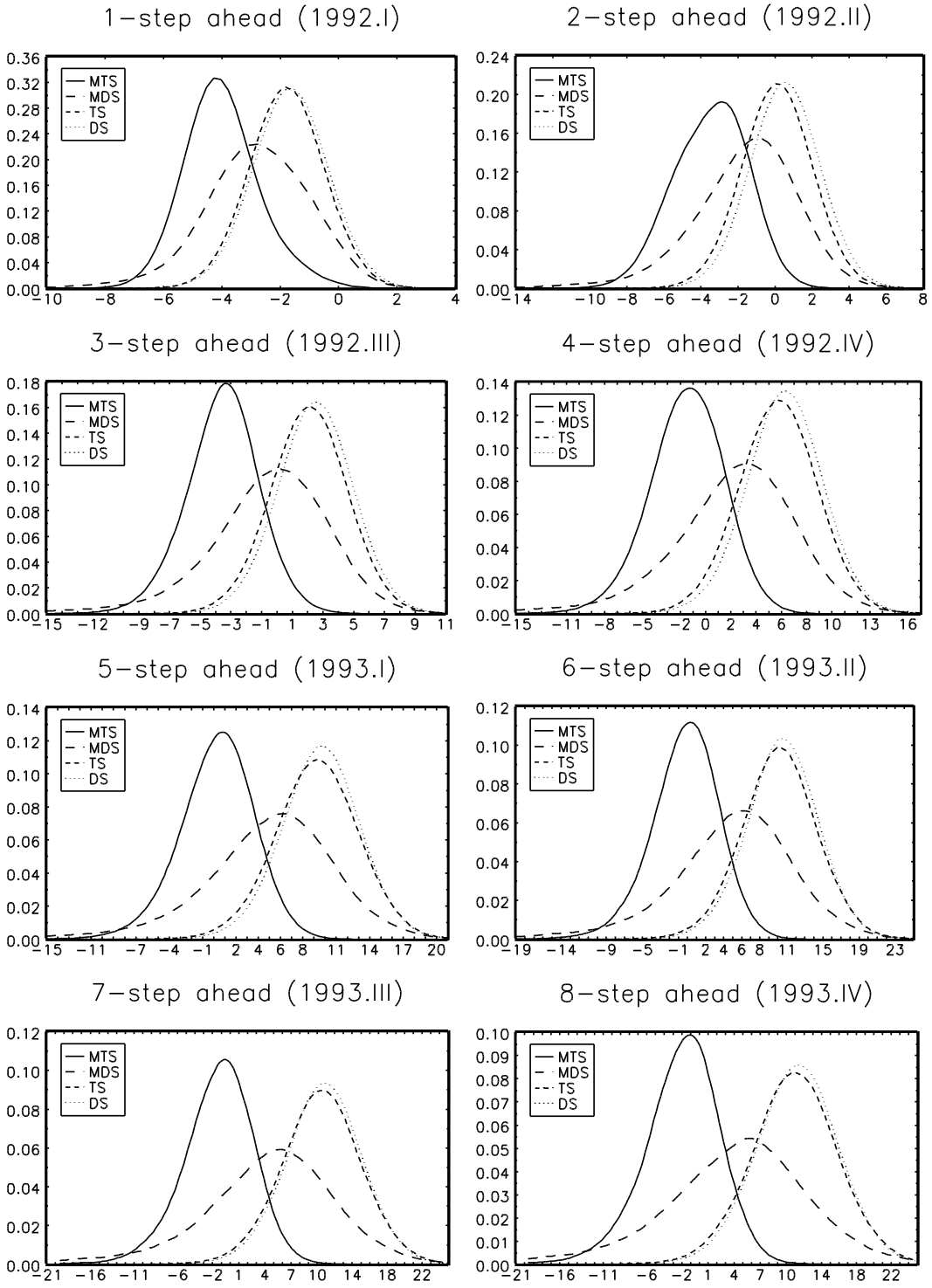

Figure 3.4. One- through eight- step ahead predictive error densities for German industrial production, 1992.I–1993.IV.

Table 3.3. Posterior means with posterior standard deviations between parentheses of the one- through eight-step ahead forecast distributions for the period 1992.I-1993.IV.[1]

period	true value	MTS[2] mean	MTS[2] s.d.	MDS mean	MDS s.d.	TS mean	TS s.d.	DS mean	DS s.d.
1992.I	104.40	100.40	(1.31)	101.59	(1.84)	102.74	(1.29)	102.60	(1.29)
1992.II	102.50	99.02	(2.01)	100.85	(2.93)	103.03	(1.89)	102.59	(1.92)
1992.III	100.80	97.26	(2.40)	100.19	(4.06)	103.27	(2.44)	102.86	(2.49)
1992.IV	97.50	95.99	(2.97)	99.64	(5.16)	103.72	(3.00)	103.19	(3.15)
1993.I	94.60	94.88	(3.38)	98.40	(6.25)	104.19	(3.50)	103.84	(3.74)
1993.II	94.20	93.94	(3.79)	98.97	(7.28)	104.66	(3.97)	104.35	(4.19)
1993.III	94.50	92.15	(4.10)	98.64	(8.31)	105.22	(4.38)	104.92	(4.62)
1993.IV	94.60	92.58	(4.42)	98.27	(9.32)	105.76	(4.75)	105.50	(5.00)

[1]Posterior results for the forecasts are conditional on the sample 1957.I–1991.IV.
[2]MTS=Markov trend stationary model (3.8), MDS=Markov difference stationary model (3.9), TS=trend stationary model (3.10) and DS=difference stationary model (3.11).

the Markov trend stationary model performs very bad compared to the other models.

In summary the Markov trend model seems a useful model to describe the trend and the business cycle in German industrial production. However, the model does not detect the recession in the 1960s. This recession is characterized by the fact that the growth in the period after the recession was so large that that the same growth path as before the recession was reached. This recession did not have a permanent effect on the level of German industrial production, like for the other recessions. One could extend the Markov trend with an extra regime to model the recession in the 1960s, see for instance Boldin (1996). However inference about the parameters modelling this extra regime will be totally based on a single event.

3.7 Concluding Remarks

In this chapter we have considered Markov trend models and discussed the role of the stochastic trends in these models. A Bayesian method to test for the presence of a unit root in the autoregressive part of the model is proposed and a test for the presence of Markov switching is discussed. The theory is applied to two simulated series and to quarterly observed seasonally adjusted German industrial production. For the latter series, a stationary AR model around a Markov trend is *a posteriori* preferred to a Markov difference stationary model and to non-linear AR models. The model detects the recessions

in German industrial production, which have had a permanent impact on the level of the series. The recession in the 1960s is however not detected. Contrary to the other recession periods, this recession was followed by a period of very high growth to reach the same growth path as before the recession. Finally, the Markov trend stationary provides superior multi-step ahead forecast distributions for the period 1992.III–1993.IV.

We conclude with some remarks concerning model extensions. The first extension concerns time varying transition probabilities. Durland and McCurdy (1994) propose duration-dependent transition probabilities, *i.e.* the value of transition probability depends on the number of periods that the process has been in the regime. Filardo (1994) and Diebold, Lee and Weinbach (1994) model the transition probabilities using logistic functions, which include explanatory variables. McCulloch and Tsay (1994b) and Paap (1995) among others allow for regime dependent autoregressive parameters, like in threshold autoregressive models [see *e.g.* Potter (1995)] or smoothed threshold autoregressive models [see *e.g.* Teräsvirta and Anderson (1992)]. The dynamic properties of these switching autoregressive models are however not easy to derive, see Holst, Lindgren, Holst and Thuvesholmen (1994) and Warne (1996). Obviously, also extensions to moving average models are possible, see Billio, Monfort and Robert (1996). Boldin (1996) considers a third-order transition process to model the regime switches and extends the Markov trend with an extra slope. Since the number of observations in a recession are usually very small, it remains however to be seen whether the data contain enough information to allow for a useful analysis of these flexible models. Finally, the models are extended with seasonal components to analyse seasonal time series, see for instance Ghysels, McCulloch and Tsay (1994). In the next chapter we extend the Markov trend model with a seasonal component to analyse seasonal unadjusted series.

Chapter 4

Seasonal Markov Trend Model

4.1 Introduction

To date business cycle turning points in the past is one of the applications of Markov trend models. This dating is usually based on seasonally adjusted time series, see among others Hamilton (1989), Goodwin (1993) and Filardo (1994) since one expects that the seasonal movements possibly blur inference on business cycles. Recent studies have questioned the accuracy of business cycle analysis using seasonal adjusted series, see among others Ghysels (1994, 1997) and Franses and Paap (1996).

A widely applied seasonal adjustment method is Census X-11, which transforms a seasonal time series via a sequence of moving average filters to an approximately non-seasonal time series. To construct seasonally adjusted series for business cycle analysis it is necessary that the seasonal adjustment filter fully removes the seasonal pattern from the series without affecting the non-seasonal pattern. The application of the Census X-11 seasonal adjustment filter has however a number of implications on the dynamic structure of time series. Ghysels, Granger and Siklos (1996) show that seasonal adjustment can introduce non-linear features in a time series with a linear dynamic pattern. Ghysels and Perron (1996) show that for time series with a one-time structural break, the probability of detecting this break decreases if one analyses the series after seasonal adjustment. Finally, seasonal adjustment biases inference from Markov switching models. Since Census X-11 consists of a moving average filtering process, a current recession observation is replaced by a weighted average of previous and forthcoming observations, which may include expansion observations. This leads to a decrease in the transition probability of changing regimes and an increase in the probability of staying in a regime, see Franses and Paap (1996) for a simulation study. In turn, this causes a bias in the estimated duration of recession and expansion periods. Furthermore, their results suggest that seasonal adjustment may influence the correct dating of business cycle turning points.

Apart from the above mentioned negative side effects of seasonal adjustment, the Census X-11 filter is also not capable of fully removing the seasonal pattern from a time series. In fact, Ghysels (1994) using seasonally adjusted data still detects seasonal patterns

in turning points, in the sense that the turning points are not equally distributed over the year. This may result from the fact that seasonal fluctuations and the business cycle are not independent [see *e.g.* Miron (1994)] and hence, it may be impossible to fully remove the seasonal pattern without affecting the business cycle. Canova and Ghysels (1994) discover, using the official NBER peaks and troughs, a significant change in the seasonal means during recessions, which indicates that seasonal fluctuations can contain valuable information about the business cycle.

The above arguments suggest that it is in most cases better for business cycle analysis to consider seasonally unadjusted series. The seasonal fluctuations have to be modelled simultaneously with the business cycle. In this chapter we introduce a seasonal variant of the Markov trend model of the previous chapter. The model allows for different means in every season. To allow for changes in the seasonal pattern we examine the presence of seasonal stochastic trends caused by seasonal unit roots.[1] Since it may be the case that changes in the seasonal means coincide with changes in stage of the business cycle [see Canova and Ghysels (1994)], we also extend the model with different seasonal means in a recession and an expansion period.

The outline of this chapter is as follows. In Section 4.2 we briefly discuss the modelling of seasonal patterns in quarterly observed time series. In Section 4.3 we propose a seasonal Markov trend model. This model extends the Markov trend model from Chapter 3 to include different means for every season and a changing seasonal pattern during recessions. Furthermore, we allow for the presence of seasonal unit roots. The Bayesian framework to analyse this seasonal variant of the Markov trend model is discussed in Section 4.4. We propose prior distributions and discuss posterior odd ratios for the presence of seasonal unit roots. To illustrate the Bayesian analysis of seasonal unit roots, we consider in Section 4.5 several simulated series. In Section 4.6 we examine the business cycle in quarterly observed German unemployment and its relation with the seasonal fluctuations. To analyse the consequences of seasonal adjustment on the dating of turning points and the estimation of the expected duration of recession and expansion periods, we consider seasonal adjusted and non-adjusted data. Finally, Section 4.7 concludes.

4.2 Modelling Seasonality

The modelling of the seasonal pattern in quarterly observed macroeconomic time series has been an important issue in the past decade. In this section we discuss some basic topics of modelling seasonal fluctuations, which concern the issues raised in this chapter. For a recent and detailed overview of modelling seasonality we refer to Franses (1996).

The top left cell of Figure 4.1 shows a plot of quarterly observed seasonally unadjusted German unemployment for the period 1962.I–1991.IV. The series clearly exhibits intra-year differences, which are referred to as seasonal fluctuations. The easiest way to model the seasonal pattern in the series is to use deterministic seasonal dummies. This

[1] See Hylleberg *et al.* (1990) for a discussion about seasonal stochastic trends.

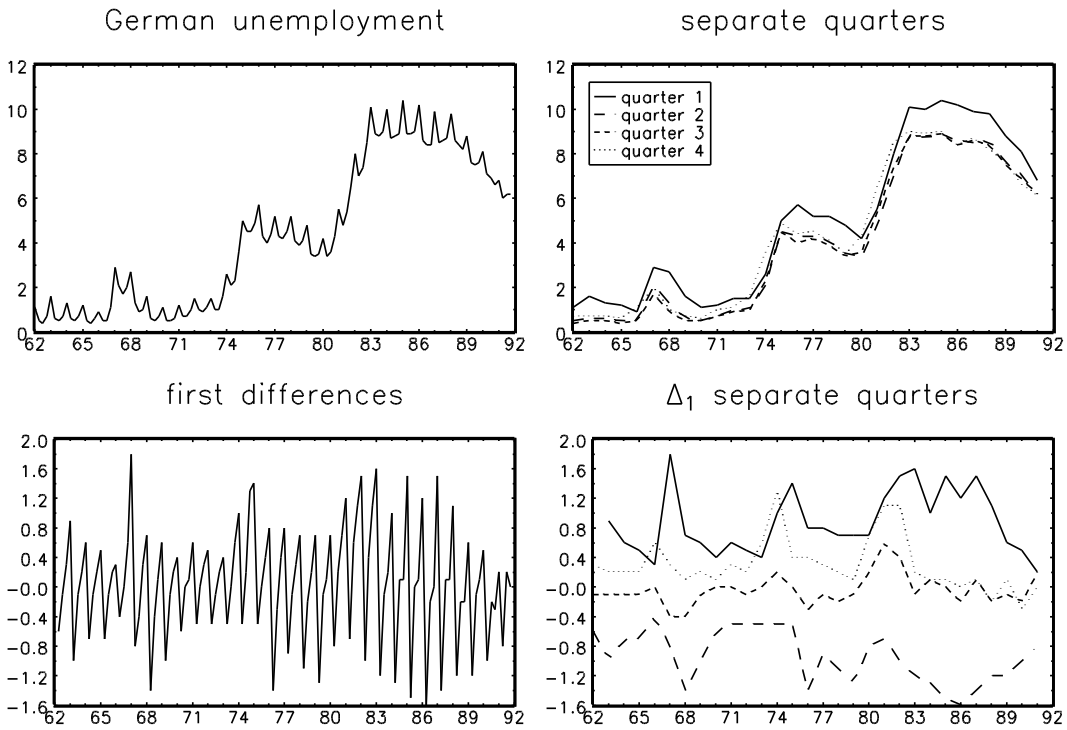

Figure 4.1. Seasonally unadjusted German unemployment, 1961.I–1991.IV.

corresponds to the assumption of a constant seasonal pattern over time in the time series. To display the seasonal pattern more explicitly, we consider a plot of the quarters of unemployment as four separate series, see the top right cell of Figure 4.1. We see that the unemployment rate in the first quarter (solid line) is almost always larger than in the other quarters. In the years 1966, 1973-1974 and 1980–1982 however the unemployment rate in the fourth quarter is larger than in the first quarter. The second row of Figure 4.1 shows a plot of the first differences of German unemployment and a plot of the first differences split up in a series for each separate quarter. The first plot shows that the amplitude of the seasonal pattern does not seem to be constant over time. The second plot shows that the distance between the first differences in each separate quarter changes over time. Hence, at first sight the seasonal pattern does not seem to be constant over time and deterministic seasonal dummies are probably not sufficient to model the seasonal fluctuations.

Apart from the seasonal fluctuations, we notice in the German unemployment series periods of overall increase and periods of decrease or constant unemployment, see Figure 4.1. To model the business cycle in this series, one may apply a Markov trend model with a seasonal component. To use this Markov trend model for business cycle analysis it is necessary to have an accurate modelling of the seasonal pattern in the time series to have precise dating of turning points. In general, changes in the seasonal pattern may be interpreted by this Markov trend model as a change in the growth rate and hence as a

recession. On the other hand, if one models changes in the seasonal pattern too flexible, a change from an expansion to a recession period or *vice versa*, may be interpreted as changes in the seasonal means of the growth rate of the series. Although it is not likely that the model considers the complete recession as a continuous change in the seasonal pattern, it can however lead to a mistake in the dating of a turning point. The top right cell of Figure 4.1 shows that that in our example the change in the order of magnitude of quarter one and quarter four almost always coincide with the periods of increase in unemployment. From the plot of the first differences it seems that the changes in the seasonal pattern coincide with the changes in the stage of the business cycle, see Canova and Ghysels (1994) for similar phenomena in macroeconomic time series of the United States. In Section 4.6.1 we analyse the seasonal pattern and the business cycle in the German unemployment series in more detail.

To deal with changing seasonal patterns, quarterly observed series are often transformed by taking fourth differences. The fourth difference filter $\Delta_4 = (1 - L^4)$ can be decomposed as

$$\begin{aligned}(1 - L^4) &= (1 - L)(1 + L)(1 - iL)(1 + iL) \\ &= (1 - L)(1 + L)(1 + L^2) \\ &= (1 - L)(1 + L + L^2 + L^3).\end{aligned} \tag{4.1}$$

It is easy to see that a time series which needs fourth differences to obtain stationarity has four roots on the unit circle. Such a series is said to be seasonally integrated. The non-seasonal root at the zero frequency (1) corresponds to a non-seasonal stochastic trend. The seasonal unit root at the frequency $1/2$ (-1) corresponds to two cycles per year and the seasonal unit roots at the frequencies $1/4$ and $3/4$ (i and $-i$) correspond to one cycle per year. A seasonal unit root corresponds to the presence of a seasonal stochastic trend, see Hylleberg *et al.* (1990) and Engle *et al.* (1993). For instance, the process $(1+L)y_t = \epsilon_t$ consists of the seasonal stochastic trend $\sum_{j=0}^{t-1}(-1)^j\epsilon_{t-j}$. Therefore, a seasonal unit root implies a continuously changing seasonal pattern. Shocks, denoted by ϵ_t, have a permanent effect on the seasonal pattern. Notice from (4.1) that the $(1-L^4)$ filter can be decomposed in a part with a non-seasonal unit root and a part with three seasonal unit roots.

A typical outcome for test for the presence of seasonal unit roots is that quarterly observed macroeconomic time series tend to have one or more seasonal stochastic trends. For example, Osborn (1990) detects seasonal unit roots in six out of thirty UK macroeconomic variables. Otto and Wirjanto (1990) obtain similar results for Canadian macroeconomic time series and Hylleberg, Jørgensen and Sørensen (1993) find that several country-specific gross domestic product series have one or more seasonal unit roots. As we already have seen, the presence of a seasonal unit root in a time series implies a continuously changing seasonal pattern. However, it may be the case that changes in the seasonal patterns only occur during changes in the stage of the business cycle, see also Canova and Ghysels (1994). In that case a seasonal stochastic trend is too flexible to model the changing seasonal pattern. For instance, each stage of the business cycle can have different seasonal means, so that a change in regime leads to a seasonal mean shift and hence the changing

seasonal pattern can be described by recurrent seasonal mean shifts.

The influence of seasonal mean shifts on seasonal unit root inference is analysed in several recent papers. Simulation results in Paap, Franses and Hoek (1997) show that neglecting seasonal means shift, when they are present, yields evidence of seasonal unit roots. In practice, Franses, Hoek and Paap (1997) show in a Bayesian analysis that the evidence for seasonal unit roots in three consumption series tends to disappear when one allows for a possible seasonal mean shift. Likewise, Franses and Vogelsang (1998) find, using classical methods, that the evidence for the bi-annual unit root -1 in US industrial production disappears when allowing for a seasonal mean shift. Finally, Paap, Franses and Hoek (1997) show that neglecting seasonal mean shifts and modelling these shifts using seasonal unit roots may lead to inferior forecasts.

In the next section, we propose the Markov trend model to analyse the business cycle in seasonally unadjusted time series. The model incorporates the possibility of changing seasonal patterns due to seasonal unit roots. To account for possible different seasonal means during the stages of the business cycle, we allow for different seasonal means during recessions and expansions.

4.3 The Seasonal Markov Trend Model

To analyse quarterly observed seasonally unadjusted series $\{y_t\}_{t=1}^T$, we extend the decomposition in (3.1) with a seasonal component

$$y_t = n_t + d_t + z_t, \tag{4.2}$$

where n_t is a trend component, d_t is a seasonal component and z_t represents the deviation from the trend and seasonal component. The trend component is again a Markov trend

$$n_t = n_{t-1} + \gamma_0 + \gamma_1 s_t, \qquad s_t = 0, 1, \tag{4.3}$$

with $n_1 = 0$.[2] The unobserved state variable s_t follows a first-order Markov process with transition probabilities

$$\begin{array}{ll} \Pr[s_t = 0 | s_{t-1} = 0] = p, & \Pr[s_t = 1 | s_{t-1} = 0] = 1 - p, \\ \Pr[s_t = 1 | s_{t-1} = 1] = q, & \Pr[s_t = 0 | s_{t-1} = 1] = 1 - q. \end{array} \tag{4.4}$$

In the simplest case the seasonal component consists of four seasonal dummies

$$d_t = \sum_{s=1}^{4} \delta_{0,s} D_{s,t}, \tag{4.5}$$

where $D_{s,t}$ represents deterministic seasonal dummies, *i.e.* $D_{s,t} = 1$ if t lies in the s-th quarter and zero elsewhere, and $\delta_{0,s}$, $s = 1, \ldots, 4$ model the seasonal means. The

[2] This restriction enables us to define four seasonal dummies in d_t, see (4.5).

deviations from the trend and seasonal component are modelled by an autoregressive process of order k [AR(k)]

$$\phi(L)z_t = \epsilon_t, \tag{4.6}$$

where $\phi(L) = (1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_k L^k)$ is a polynomial in the lag operator L, defined by $L^i y_t = y_{t-i}$, $i = 0, 1, \ldots$.

Seasonal Unit Roots

In Section 3.2, we have shown in (3.7) that it is possible to rewrite the polynomial $\phi(L)$ such that a restriction on a single parameter (*i.e.* $\rho = 1$) implies the presence of a unit root in the polynomial. Hylleberg *et al.* (1990) [HEGY] show that it is also possible to rewrite the polynomial such that zero restrictions on parameters imply the presence of the non-seasonal root 1, and the seasonal unit roots -1, i and $-i$

$$\begin{aligned}\phi(L) = -\pi_1 L(1 + L + L^2 + L^3) - \pi_2 L(-1 + L - L^2 + L^3) \\ - (\pi_3 L^2 + \pi_4 L)(-1 + L^2) + \bar{\bar{\phi}}(L)(1 - L^4),\end{aligned} \tag{4.7}$$

where $\bar{\bar{\phi}}(L) = (1 - \bar{\bar{\phi}}_1 L - \cdots - \bar{\bar{\phi}}_{k-4} L^{k-4})$, a lag polynomial of order $(k-4)$ and $\bar{\bar{\phi}}_i$ $i = 1, \ldots, k-4$ and π_j $j = 1, \ldots, 4$ are functions of the ϕ_i parameters. Applying (4.7) to (4.6) results in the so-called HEGY test equation

$$\Delta_4 z_t = \pi_1 z_{1,t-1} + \pi_2 z_{2,t-1} + \pi_3 z_{3,t-2} + \pi_4 z_{3,t-1} + \sum_{i=1}^{k-4} \bar{\bar{\phi}}_i \Delta_4 z_{t-i} + \epsilon_t, \tag{4.8}$$

where

$$\begin{aligned} z_{1,t} &= (1 + L + L^2 + L^3) z_t = z_t + z_{t-1} + z_{t-2} + z_{t-3} \\ z_{2,t} &= (-1 + L - L^2 + L^3) z_t = -z_t + z_{t-1} - z_{t-2} + z_{t-3} \\ z_{3,t} &= (-1 + L^2) z_t = -z_t + z_{t-2}. \end{aligned} \tag{4.9}$$

If $\pi_1 = 0$ the series contains a unit root at the zero frequency. A unit root at the frequency $1/2$ (-1) corresponds to $\pi_2 = 0$. If $\pi_3 = \pi_4 = 0$ the series contains the complex roots i and $-i$. For details we refer to Hylleberg *et al.* (1990) and Engle *et al.* (1993).

Identification Problem

We have seen in Section 3.2 that in case of a unit root the initial value of the Markov trend n_1 is not identified, since it drops out of the model. Likewise, in case of seasonal unit roots, linear combinations of the seasonal dummy parameters $\delta_{0,s}$ are not identified. To make this identification problem explicit, consider the following one-to-one transformation of the seasonal mean parameters $\delta_0 = (\delta_{0,1}\ \delta_{0,2}\ \delta_{0,3}\ \delta_{0,3})'$ into the parameters ω_s, $s = 1, \ldots, 4$,

$$\begin{pmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \\ \omega_4 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{pmatrix} \begin{pmatrix} \delta_{0,1} \\ \delta_{0,2} \\ \delta_{0,3} \\ \delta_{0,4} \end{pmatrix} \tag{4.10}$$

or in matrix notation $\omega = F\delta_0$, where $\omega = (\omega_1\ \omega_2\ \omega_3\ \omega_4)'$. Using this transformation and replacing z_t by $(y_t - n_t - d_t)$ in (4.8) results in

$$\begin{aligned}\Delta_4\tilde{y}_t = \pi_1(\tilde{y}_{1,t-1} - \omega_1) + \pi_2(\tilde{y}_{2,t-1} - \omega_2(-1)^t) + \pi_3(\tilde{y}_{3,t-2} - \omega_3\kappa_{t-1} - \omega_4\kappa_{t-2}) \\ + \pi_4(\tilde{y}_{3,t-1} - \omega_3\kappa_t - \omega_4\kappa_{t-1}) + \sum_{i=1}^{k-4}\bar{\bar{\phi}}_i\Delta_4\tilde{y}_{t-i} + \epsilon_t, \quad (4.11)\end{aligned}$$

where $\tilde{y}_t = y_t - n_t$, $\kappa_t = \frac{1}{2}(i^t + (-i)^t)$ and where we assume that $t = 1$ corresponds to a first quarter observation. It is easy to see that in (4.11) the parameter ω_1 is not identified if $\pi_1 = 0$. Analogously, if $\pi_2 = 0$ the parameter ω_2 is not identified. Roots at the frequencies 1/4 and 3/4 ($\pi_3 = \pi_4 = 0$) imply that ω_3 and ω_4 are not identified, see also Franses, Hoek and Paap (1997). Note that zero restrictions on the π parameters do not lead to non-identification of γ_0 and γ_1. This follows directly from the annual growth of the trend component n_t at time t

$$\begin{aligned}\Delta_4 n_t &= n_t - n_{t-4} \\ &= (\gamma_0(t-1) + \gamma_1\sum_{i=2}^{t} s_i) - (\gamma_0(t-5) + \gamma_1\sum_{i=2}^{t-4} s_i) \\ &= 4\gamma_0 + \gamma_1\sum_{i=t-3}^{t} s_i,\end{aligned} \quad (4.12)$$

where we use the backward solution of n_t given in (3.5). Note that the interpretation of γ_0 and γ_1 does not change under zero restrictions on the π parameters. Under the restriction $\gamma_1 = 0$ the model simplifies to an AR model with a deterministic trend. The annual growth in y_t equals $4\gamma_0$. This model specification is examined in Franses, Hoek and Paap (1997).

Markov Switching in Seasonal Means

Franses, Hoek and Paap (1997) correct in their model for a possible seasonal mean shift in a series by extending the seasonal component (4.5) with a different seasonal mean after a certain point in time. Likewise, to allow for changes in the seasonal means during recessions we replace (4.5) by

$$d_t = \sum_{s=1}^{4}(\delta_{0,s} + \delta_{1,s}s_t)D_{s,t}, \quad (4.13)$$

so that during a recession the seasonal means change from $\delta_{0,s}$ to $(\delta_{0,s} + \delta_{1,s})$, $s = 1, \ldots, 4$. Since

$$\begin{aligned}\Delta_4 d_t &= \sum_{s=1}^{4}(\delta_{0,s} + \delta_{1,s}s_t)D_{s,t} - (\delta_{0,s} + \delta_{1,s}s_{t-4})D_{s,t-4} \\ &= \sum_{s=1}^{4}(\delta_{1,s}s_t - \delta_{1,s}s_{t-4})D_{s,t}\end{aligned} \quad (4.14)$$

the $\delta_{1,s}$ parameters $s = 1, \ldots, 4$ are identified if there is at least one observation in every quarter, which corresponds to a recession, $s_t = 0$. To see the consequences of these recurrent seasonal mean shifts on the growth of the series y_t we consider

$$\begin{aligned}\Delta y_t &= \Delta n_t + \Delta d_t + \Delta z_t \\ &= \gamma_0 + \gamma_1 s_t + \sum_{s=1}^{4}((\delta_{0,s} - \delta_{1,s-1}) + (\delta_{1,s} s_t - \delta_{1,s-1} s_{t-1}))D_{s,t} + \Delta z_t,\end{aligned} \tag{4.15}$$

where $\delta_{i,0} = \delta_{i,4}$, $i = 0, 1$. If $s_t = s_{t-1} = 0$ the quarterly growth in y_t in season s is γ_0 plus the seasonal variation $(\delta_{0,s} - \delta_{0,s-1})$. Note that the sum of the seasonal variations over a year is 0, *i.e.* $\sum_{s=1}^{4}(\delta_{0,s} - \delta_{0,s-1}) = 0$. If $s_t = s_{t-1} = 1$, the growth rate is $(\gamma_0 + \gamma_1)$ plus the seasonal variation $(\delta_{0,s} + \delta_{1,s}) - (\delta_{0,s-1} - \delta_{1,s-1})$, so that under no regime changes, the sum of the seasonal variation over a year is again zero. However, if $s_t \neq s_{t-1}$ the quarterly growth is corrected by the factor $(\delta_{1,s} s_t - \delta_{1,s-1} s_{t-1})$ for the difference in seasonal variation in the two regimes.

The Likelihood Function

The derivation of the likelihood function of the seasonal Markov trend model proceeds in the same way as for the non-seasonal Markov trend model in Section 3.2. Under the assumption $\epsilon_t \sim \text{NID}(0, \sigma^2)$, the conditional density of y_t given the past observations $y^{t-1} = \{y_1, \ldots, y_{t-1}\}$ and given the past and current states $s^t = \{s_1, \ldots, s_t\}$ reads

$$f(y_t | y^{t-1}, s^t, \gamma_0, \gamma_1, \delta_0, \delta_1, \sigma, \pi, \bar{\bar{\phi}}) = \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{1}{2\sigma^2}\epsilon_t^2), \tag{4.16}$$

where $\delta_1 = \{\delta_{1,1}, \delta_{1,2}, \delta_{1,3}, \delta_{1,4}\}$, $\bar{\bar{\phi}} = \{\bar{\bar{\phi}}_1, \ldots, \bar{\bar{\phi}}_{k-4}\}$, $\pi = \{\pi_1, \pi_2, \pi_3, \pi_4\}$ and ϵ_t is defined in (4.11). Hence, the likelihood function for model (4.11) conditional on the states $s^T = \{s_1, \ldots, s_T\}$ and conditional on the initial k observations $y^k = \{y_1, y_2, \ldots, y_k\}$ is given by

$$\begin{aligned}\mathcal{L}(y^T | y^k, s^T, \theta) = p^{\mathcal{N}_{00}} (1-p)^{\mathcal{N}_{01}} q^{\mathcal{N}_{11}} (1-q)^{\mathcal{N}_{10}} \\ \prod_{t=k+1}^{T} f(y_t | y^{t-1}, s^t, \gamma_0, \gamma_1, \delta_0, \delta_1, \sigma, \pi, \bar{\bar{\phi}}),\end{aligned} \tag{4.17}$$

where $\theta = \{\gamma_0, \gamma_1, \omega, \delta_1, \sigma, \pi, \bar{\bar{\phi}}, p, q\}$[3], $\mathcal{N}_{ij}$ denotes the number of transitions from state i to state j. Again, we have used that $\{y^{t-1}, s^{t-2}\}$ does not Granger cause s_t, *i.e.* the conditional distribution $p(s_t | s^{t-1}, y^{t-1})$ equals the conditional distribution $p(s_t | s_{t-1})$. The unconditional (on the states) likelihood function $\mathcal{L}(y^T | y^k, \theta)$ reads

$$\mathcal{L}(y^T | y^k, \theta) = \sum_{s_1=0}^{1} \sum_{s_2=0}^{1} \cdots \sum_{s_T=0}^{1} \mathcal{L}(y^T | y^k, s^T, \theta). \tag{4.18}$$

[3]Note that we have specified the likelihood as a function of ω instead of δ_0. This turns out to be more convenient for the seasonal unit root analysis, see Section 4.4.

This unconditional likelihood function can be evaluated using the algorithm of Lam (1990). As we already have seen in Chapter 3 we do not have to compute this unconditional likelihood for our Bayesian analysis.

The unconditional likelihood functions under zero restrictions on the π parameters are defined in the same way. The unconditional likelihood function under $\pi_1 = 0$ equals

$$\mathcal{L}_1(y^T|y^k, \theta_1) = \mathcal{L}(y^T|y^k, \theta)|_{\pi_1=0}, \tag{4.19}$$

where $\theta_1 = \theta \backslash \{\pi_1, \omega_1\}$.[4] Likewise, the unconditional likelihood functions given the states under $\pi_2 = 0$ and $\pi_3 = \pi_4 = 0$ are

$$\begin{aligned} \mathcal{L}_2(y^T|y^k, \theta_2) &= \mathcal{L}(y^T|y^k, \theta)|_{\pi_2=0} \\ \mathcal{L}_{34}(y^T|y^k, \theta_{34}) &= \mathcal{L}(y^T|y^k, \theta)|_{\pi_3=\pi_4=0}, \end{aligned} \tag{4.20}$$

where $\theta_2 = \theta \backslash \{\pi_2, \omega_2\}$ and $\theta_{34} = \theta \backslash \{\pi_3, \pi_4, \omega_3, \omega_4\}$.

In the next section we consider prior specification, posterior simulation and the analysis of the presence of seasonal unit roots for the seasonal Markov trend model.

4.4 Bayesian Analysis

The Bayesian analysis of the seasonal Markov trend model is similar to the analysis of the non-seasonal model in Chapter 3. Therefore, we focus in this section on the Bayesian treatment of the seasonal part of the model: the prior specification for the parameters, which model the seasonal pattern, and posterior odds ratios to examine the presence of seasonal unit roots.

4.4.1 Prior Specification

In the previous section we have seen that the presence of (seasonal) unit roots implies that certain linear combination of the seasonal dummy parameters δ_0 are not identified. This corresponds to the non-identifiedness of the n_1 parameter when $\rho = 1$ in the non-seasonal Markov trend model. Hence, using the same arguments as in the previous chapter specifying diffuse priors on δ_0 results in improper posterior distributions for the π parameters favouring the presence of (seasonal) unit roots. In Section 3.3.1 we have used the initial observation y_1 to define a prior for n_1. Likewise, we use the first four observations $\{y_1, y_2, y_3, y_4\}$ to define a prior on the seasonal dummy parameters. Consider the following initial model for the first four observations

$$\begin{pmatrix} \tilde{y}_1 - \delta_{1,1} s_1 \\ \tilde{y}_2 - \delta_{1,2} s_2 \\ \tilde{y}_3 - \delta_{1,3} s_3 \\ \tilde{y}_4 - \delta_{1,4} s_4 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \delta_{0,1} \\ \delta_{0,2} \\ \delta_{0,3} \\ \delta_{0,4} \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \end{pmatrix} \tag{4.21}$$

[4] Note that the subscript correspond to which π parameters is equal to zero.

or in matrix notation

$$\tilde{\tilde{y}} = \mathbf{I}_4 \, \delta_0 + e, \tag{4.22}$$

where $e = (\epsilon_1 \ \epsilon_2 \ \epsilon_3 \ \epsilon_4)'$, $\tilde{\tilde{y}}$ is a (4×1) vector containing $y_i - n_i - \delta_{1,i} s_i$ for $i = 1, \ldots, 4$ and $\mathbf{I}_4$ a (4×4) identity matrix. Instead of assuming that e is normally distributed with zero mean and the covariance matrix of an AR(4) model, like in Franses, Hoek and Paap (1997), we assume that $e \sim \mathrm{N}(0, \sigma^2 \mathbf{I}_4)$. This simplifies the computation of marginal posteriors for the π parameters, see also Hoek (1997) and Chapter 3 for a similar argument concerning the ρ parameter. The presence of (seasonal) unit roots implies that certain linear combinations of the elements of δ_0 are not identified. To compare models with different number of (seasonal) unit roots it is more convenient to have a single parameter which is not identified under a certain hypothesis, like $\omega_1 = 0$. Therefore, we transform the seasonal dummy parameters δ_0 to ω using (4.10). Hence, we consider the parameterisation in (4.11). Applying the transformation (4.10) (or $\delta_0 = F^{-1}\omega$) to (4.22) results in

$$\tilde{\tilde{y}} = F^{-1} \, \omega + e. \tag{4.23}$$

This implies the following conditional prior for ω

$$\omega \mid \delta_1, s^4, y^4 \ \sim \ \mathrm{N}(F\tilde{\tilde{y}}, FF'). \tag{4.24}$$

Note that this prior is conditional on the first four observations y^4, the first four states s^4, and δ_1.

The priors for the π_i parameters, $i = 1, \ldots, 4$, are uniform. Since we want to test the presence of unit roots against roots outside the unit circle we define uniform and independent priors for π_1 and π_2 on the the intervals $[\pi_{1,lb}, 0]$ and $[\pi_{2,lb}, 0]$

$$\begin{aligned} p(\pi_1) &= \frac{1}{-\pi_{1,lb}} \mathbb{I}_{[\pi_{1,lb},0]} \\ p(\pi_2) &= \frac{1}{-\pi_{2,lb}} \mathbb{I}_{[\pi_{2,lb},0]}, \end{aligned} \tag{4.25}$$

where $\mathbb{I}$ is again an indicator function, which is one on $[\pi_{i,lb}, 0]$ and zero elsewhere. The values $\pi_{1,lb}$ and $\pi_{2,lb}$ are chosen such that they define highest posterior density [HPD] regions for π_1 and π_2, respectively, see Section 4.4.3 for a discussion. Since the presents of the roots i and $-i$ corresponds to the restriction $\pi_3 = \pi_4 = 0$, we define a joint prior for (π_3, π_4) on the HPD region for the join posterior of π_3 and π_4

$$p(\pi_3, \pi_4) = \begin{cases} \frac{1}{\mathrm{area}(\mathcal{P})} & \text{if } (\pi_3, \pi_4) \in \mathcal{P} \\ 0 & \text{elsewhere} \end{cases} \tag{4.26}$$

where $\mathcal{P}$ denotes the HPD region for (π_3, π_4) and $\mathrm{area}(\mathcal{P})$ corresponds to the area of this region. Of course $\mathcal{P}$ has to be inside the stationary region.

As we already have seen, the $\delta_{1,s}$ parameters, $s = 1, \ldots, 4$, are identified if there is at least one s_t in quarters s which corresponds to a recession observation $s_t = 1$. Since there is no guarantee that this is the case we impose informative priors to avoid identification problems. We take normal prior distributions for $\delta_{1,s}$ parameters

$$\delta_{1,s} \sim \mathrm{N}(0, \sigma^2_{\delta_{1,s}}) \qquad s = 1, \ldots, 4, \tag{4.27}$$

so that we do not *a priori* suggest a seasonal mean shift and we can control the amount of information in the prior via the prior variance $\sigma^2_{\delta_{1,s}}$.

The priors on the remaining parameters follow directly from Section 3.3.1. Briefly, the priors for p and q are again uniform and independent on the interval $(0, 1)$

$$\begin{aligned} p(p) &= \mathbb{I}_{(0,1)} \\ p(q) &= \mathbb{I}_{(0,1)}. \end{aligned} \tag{4.28}$$

For the γ_0 and γ_1 we take uniform and independent priors on bounded intervals to identify the two regimes

$$\begin{aligned} p(\gamma_0) &= \frac{1}{(\gamma_{0,ub} - \gamma_{0,lb})} \mathbb{I}_{[\gamma_{0,lb}, \gamma_{0,ub}]} \\ p(\gamma_1) &= \frac{1}{(\gamma_{1,ub} - \gamma_{1,lb})} \mathbb{I}_{[\gamma_{1,lb}, \gamma_{1,ub}]}. \end{aligned} \tag{4.29}$$

and the priors for σ and $\bar{\bar{\phi}}$ are given by

$$\begin{aligned} p(\sigma) &\propto \sigma^{-1} \\ p(\bar{\bar{\phi}}) &\propto \mathbb{I}_{[stat]}, \end{aligned} \tag{4.30}$$

where $\mathbb{I}_{[stat]}$ is an indicator function, which is one if the autoregressive parameters ϕ imply that the roots of the autoregressive polynomial are outside the unit circle and zero elsewhere.

The joint prior of the parameters $\theta = \{\gamma_0, \gamma_1, \omega, \delta_1, \sigma, \pi, \bar{\bar{\phi}}, p, q\}$, $p(\theta)$, is given by the product of (4.24)–(4.30). The joint priors for the parameters under zero restrictions on π, *i.e.* $p_1(\theta_1)$, $p_2(\theta_2)$ and $p_{34}(\theta_{34})$, are defined by simply neglecting the marginal prior for the parameters which are not in the parameter set.

4.4.2 Posterior Distributions

The joint posterior distribution of the parameters is proportional to the product of the priors (4.24)–(4.30) and the unconditional likelihood function (4.18). To obtain marginal results, we use again the Gibbs sampling techniques discussed in Section 3.4. Since the seasonal Markov trend model is a generalisation of the Markov trend model of Chapter 3, simulation from the posterior can be done in a similar way. Again, the state variables $\{s_t\}_{t=1}^T$ are sampled alongside the model parameters from similar full conditional distributions as in (3.35). The full conditional distributions of p and q are beta distributions, see

(3.37) and (3.38). The variance σ given the states and the other parameters is inverted gamma-2 distributed. The full conditional distributions of the remaining parameters π, ω, δ_1 and $\bar{\bar{\phi}}$ are (truncated) normal.

4.4.3 Seasonal Unit Root Analysis

To analyse the presence of (seasonal) unit roots in a quarterly observed time series we consider the following four hypotheses

$$\begin{aligned} H &: \pi_{1,lb} < \pi_1 < 0 \ \wedge \ \pi_{2,lb} < \pi_2 < 0 \ \wedge \ (\pi_3, \pi_4) \in \mathcal{P}, \\ H_1 &: \quad \pi_1 = 0 \quad \wedge \ \pi_{2,lb} < \pi_2 < 0 \ \wedge \ (\pi_3, \pi_4) \in \mathcal{P}, \\ H_2 &: \pi_{1,lb} < \pi_1 < 0 \ \wedge \quad \pi_2 = 0 \quad \wedge \ (\pi_3, \pi_4) \in \mathcal{P}, \\ H_{34} &: \pi_{1,lb} < \pi_1 < 0 \ \wedge \ \pi_{2,lb} < \pi_2 < 0 \ \wedge \ \pi_3 = \pi_4 = 0 \end{aligned} \tag{4.31}$$

and hence H_1 corresponds to the hypothesis of the presence of the root 1, H_2 to the presence of the root -1 and H_{34} to the presence of i and $-i$. Of course it is also possible to consider other joint tests, for instance for the presence of the non-seasonal and the seasonal unit roots 1 and -1, but for our purpose it is sufficient to consider the hypotheses in (4.31).

To analyse the presence of (seasonal) unit roots we compare the three hypotheses H_i, $i = 1, 2, 34$, against the hypothesis H using posterior odds ratios. To illustrate the computation of the posterior odds ratios, we focus on the test for the presence of the complex root i and $-i$, *i.e.* $\pi_3 = \pi_4 = 0$. *A priori* we assign prior probabilities to the hypotheses H_{34} and H. These prior probabilities imply the prior odds ratio

$$\text{PROR}(\pi_3, \pi_4) = \frac{\Pr[H_{34}]}{\Pr[H]}. \tag{4.32}$$

This prior odds ratio times the Bayes factor [BF] provides the posterior odds ratio [POR]

$$\begin{aligned} \text{POR}(\pi_3, \pi_4) &= \text{PROR}(\pi_3, \pi_4) \times \text{BF}(\pi_3, \pi_4) \\ &= \frac{\Pr[H_{34}]}{\Pr[H]} \times \frac{\int p_{34}(\theta_{34}) \mathcal{L}_{34}(y^T | y^k, \theta_{34}) \, d\theta_{34}}{\int p(\theta) \mathcal{L}(y^T | y^k, \theta) \, d\theta}. \end{aligned} \tag{4.33}$$

To compute the Bayes factor we again use the Savage-Dickey density ratio of Dickey (1971), see Section 3.3.2 for a discussion. The Bayes factor for $\pi_3 = \pi_4 = 0$ equals the ratio of the marginal posterior of (π_3, π_4) evaluated in $\pi_3 = \pi_4 = 0$ and the prior for (π_3, π_4) evaluated in $\pi_3 = \pi_4 = 0$

$$\text{BF}(\pi_3, \pi_4) = \frac{p(\pi_3, \pi_4 | y^T)|_{\pi_3 = \pi_4 = 0}}{p(\pi_3, \pi_4)|_{\pi_3 = \pi_4 = 0}} \tag{4.34}$$

where $p(\pi_3, \pi_4 | y^T)$ denotes the marginal posterior of (π_3, π_4) and the prior $p(\pi_3, \pi_4)$ is defined in (4.26). We choose the region $\mathcal{P}$ such that it corresponds to the 99% HPD

region for (π_3, π_4) to avoid favouring the H_{34} hypothesis, see Section 3.3.2 for a similar argument concerning the prior for the ρ parameter.

The Bayes factors for $\pi_1 = 0$ and $\pi_2 = 0$ can be constructed in the same way. The Savage-Dickey density ratios for both hypotheses equal

$$\begin{aligned} \mathrm{BF}(\pi_1) &= \frac{p(\pi_1|y^T)|_{\pi_1=0}}{p(\pi_1)|_{\pi_1=0}} \\ \mathrm{BF}(\pi_2) &= \frac{p(\pi_2|y^T)|_{\pi_2=0}}{p(\pi_2)|_{\pi_2=0}}, \end{aligned} \tag{4.35}$$

where $p(\pi_1|y^T)$ and $p(\pi_2|y^T)$ are the marginal posteriors of π_1 and π_2 and the priors $p(\pi_1)$ and $p(\pi_2)$ are defined in (4.25). The lowerbounds $\pi_{1,lb}$ and $\pi_{2,lb}$ in these priors define the 99% HPD region for π_1 and π_2 respectively.

An alternative Bayesian approach for seasonal unit root analysis can be found in Koop and Pitarakis (1992). They, however, consider an linear model specification like in (2.23) instead of the specification from the previous section and hence can follow the Bayesian unit root analysis of DeJong and Whiteman (1991) and Zellner and Siow (1980). Since they do not consider a time series in deviation from a trend and seasonal component, the interpretation of intercept, seasonal dummies and trend parameters changes under the various hypotheses. Furthermore, in this linear trend stationary AR specification the trend parameter does not represent the growth rate of a series unless the AR order is zero, see also Section 2.3.

In the next section we illustrate the Bayesian seasonal unit root analysis, presented in this section, on simulated series. In Section 4.6 we test for the presence seasonal unit roots in seasonally unadjusted German unemployment.

4.5 Illustration of Seasonal Unit Root Analysis

To illustrate the Bayesian seasonal unit root analysis we consider four simulated series. The data generating process [DGP] is given by

$$\begin{aligned} y_t &= n_t + d_t + z_t, \\ n_t &= n_{t-1} + 1, \qquad\qquad n_1 = 0, \\ d_t &= -D_{1,t} + D_{2,t} - D_{3,t} + D_{4,t} \end{aligned} \tag{4.36}$$

with

$$\begin{aligned} \mathrm{I} &: z_t = \epsilon_t, \\ \mathrm{II} &: z_t = z_{t-1} + \epsilon_t, \\ \mathrm{III} &: z_t = -z_{t-1} + \epsilon_t, \\ \mathrm{IV} &: z_t = z_{t-4} + \epsilon_t, \end{aligned} \tag{4.37}$$

Table 4.1. Posterior means with posterior standard deviations between parentheses of the parameters of a trend stationary model with seasonal dummies and Bayes factors for the presence of (seasonal) unit roots for the four DGPs.

para-	DGP I[1]		DGP II		DGP III		DGP IV	
meters	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.
γ_0	0.99	(0.00)	0.98	(0.03)	1.00	(0.00)	1.00	(0.01)
w_1	0.60	(0.59)	−0.73	(1.69)	−0.29	(0.63)	−2.91	(1.81)
w_2	3.67	(0.25)	4.16	(0.14)	4.69	(2.11)	3.46	(1.70)
w_3	0.17	(0.24)	0.24	(0.19)	0.03	(0.21)	1.56	(1.46)
w_4	−0.36	(0.24)	−0.07	(0.19)	−0.04	(0.21)	−0.89	(1.36)
π_1	−0.29	(0.06)	−0.01	(0.01)	−0.39	(0.09)	−0.04	(0.03)
π_2	−0.34	(0.06)	−0.63	(0.10)	−0.01	(0.00)	−0.03	(0.02)
π_3	−0.49	(0.07)	−0.43	(0.09)	−0.57	(0.09)	−0.02	(0.01)
π_4	−0.03	(0.07)	−0.46	(0.09)	0.47	(0.10)	−0.01	(0.02)
σ	0.80	(0.06)	1.08	(0.06)	1.06	(0.08)	0.95	(0.07)
$\text{BF}(\pi_1)$[2]	0.00		2.07		0.00		1.62	
$\text{BF}(\pi_2)$	0.00		0.00		1.47		1.31	
$\text{BF}(\pi_3, \pi_4)$	0.00		0.00		0.00		5.36	

[1] The DGPs are given in (4.36) and (4.37).

[2] $\text{BF}(\pi_i)$ denotes the Bayes factor. A Bayes factor exceeding one implies that $\pi_i = 0$ is *a posteriori* more likely than $\pi_i < 0$. $\text{BF}(\pi_3, \pi_4)$ denotes the Bayes factor for $\pi_3 = \pi_4 = 0$.

where $\epsilon_t \sim \text{NID}(0, 1)$. DGP I does not contain any unit roots, DGP II contains the non-seasonal unit root 1, DGP III contains the seasonal unit root -1 and the final DGP contains the roots 1, -1, i and $-i$. The simulated series denoted by DGP I–IV are analysed using the model presented in Section 4.3. Since we have not introduced a Markov trend in the DGPs, we remove the Markov structure from the model, *i.e.* $\gamma_1 = 0$ and $\delta_1 = \mathbf{0}$. In the previous section, we have seen that we need at least four lags to test for seasonal unit roots. Since the maximum lag order of the AR process in our DGPs is four, we take $k = 4$. The joint prior is given by the product of (4.24)–(4.26) and a flat prior for γ_0 on $(-\infty, \infty)$. We assume equal prior probabilities for the hypothesis under consideration so that the posterior odds ratios equal the Bayes factors. Table 4.1 shows posterior results for the four DGPs.

For the first DGP we see that the posterior means of π_1, π_2 and π_3 are more than two posterior standard deviations away from zero. However, since the posterior densities are

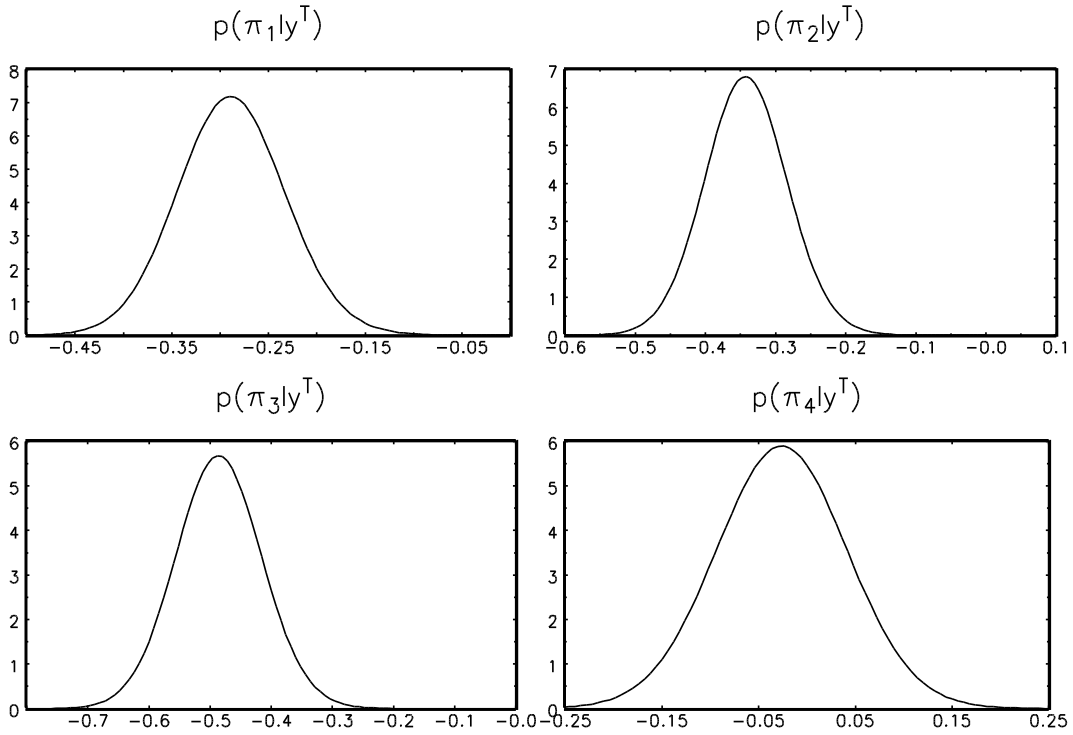

Figure 4.2. Marginal posterior densities of the π parameters for DGP I.

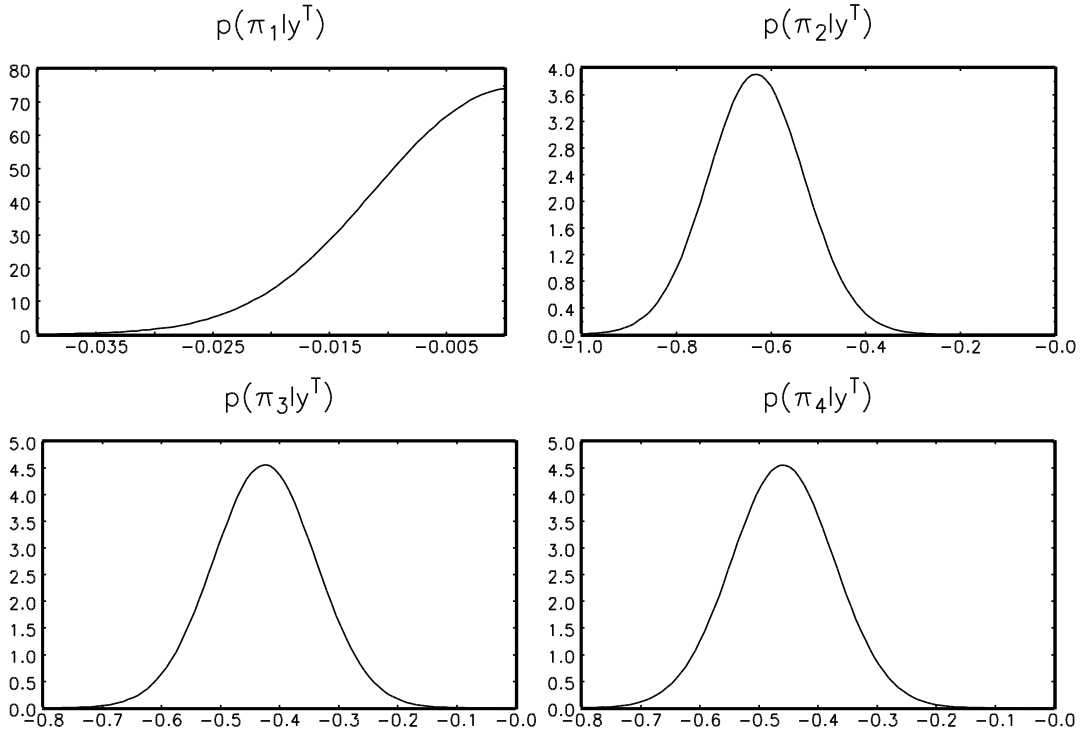

Figure 4.3. Marginal posterior densities of the π parameters for DGP II.

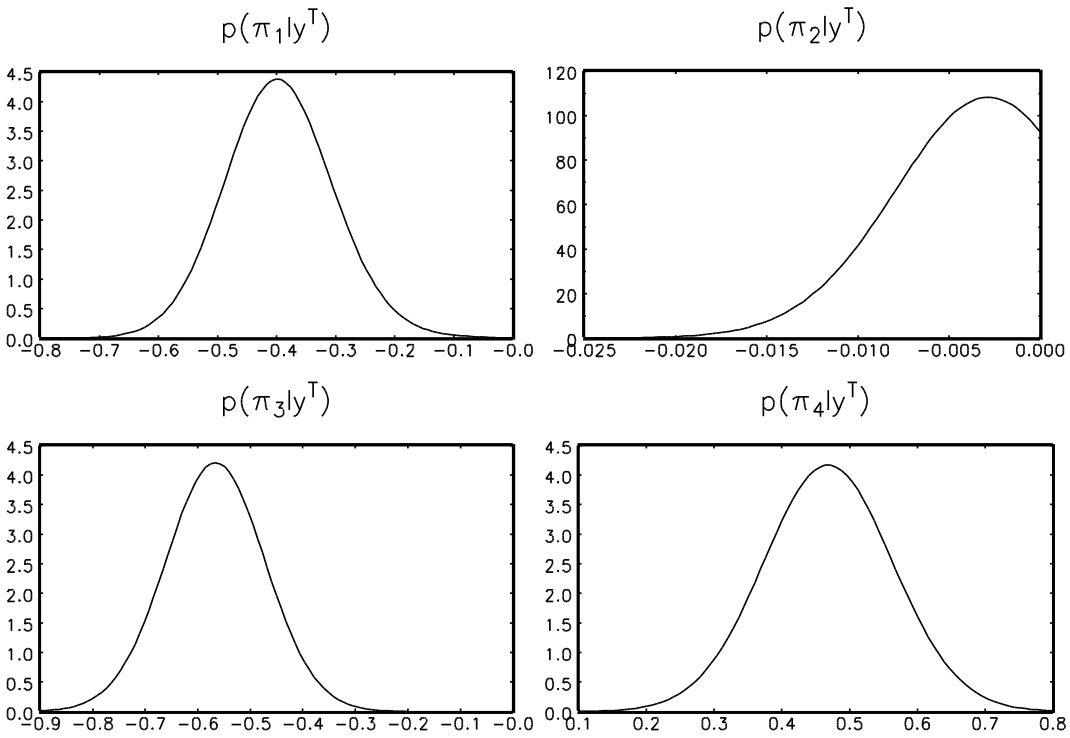

Figure 4.4. Marginal posterior densities of the π parameters for DGP III.

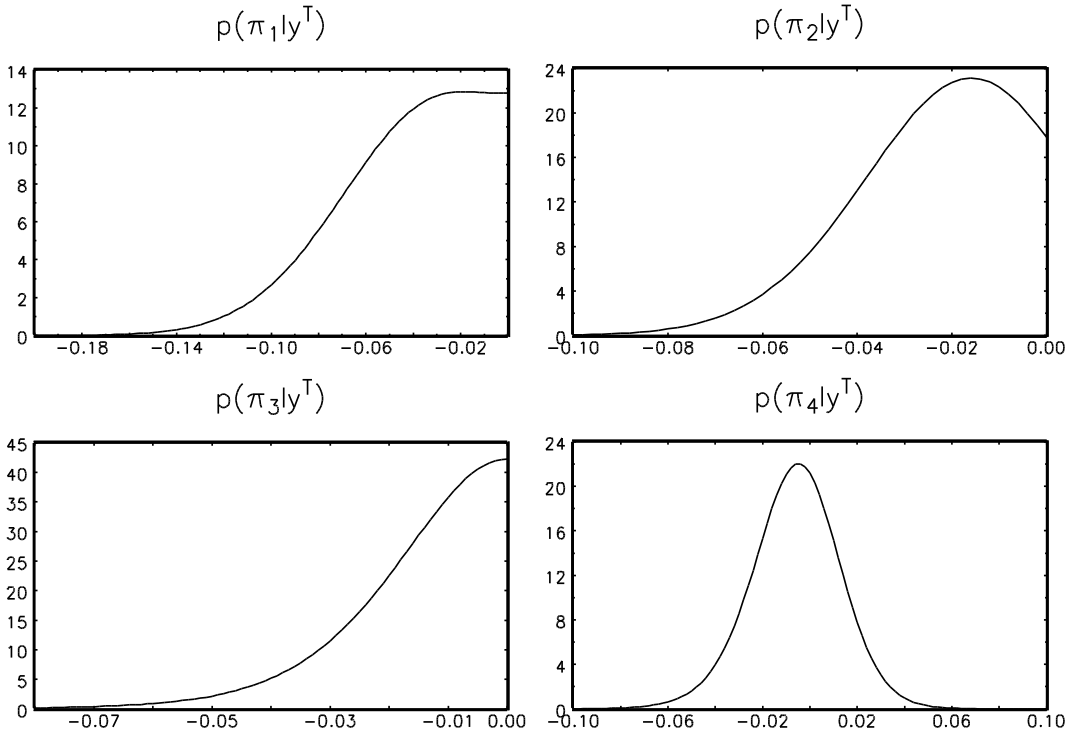

Figure 4.5. Marginal posterior densities of the π parameters for DGP IV.

truncated, care must be exercised in interpreting the standard deviations. In Figure 4.2 we depict the marginal posterior densities of the π parameters. The modes of the distributions are far away from zero, except for π_4. However, this does not imply the presence of a seasonal unit root, since the condition for the presence of the roots i and $-i$ is $\pi_3 = \pi_4 = 0$. The Bayes factors for this joint test is clearly smaller than one, see Table 4.1. The same is true for the Bayes factors for the hypothesis of the presence of the roots 1 and -1.[5] The overwhelming evidence for the absence of any unit roots is not surprising, since the z_t process in the DGP has only roots equal to zero. The posterior means of σ and γ_0 are close to their true value. The posterior means of ω_1, $(\omega_2 - 4)$, ω_3 and ω_4 lie within two posterior standard deviations from zero as expected.

The second column of Table 4.1 shows the posterior results for the second DGP. Now the posterior mean of π_1 parameter is near zero and the Bayes factor for the presence of the root 1 is 2.07. The posterior means of π_2, π_3 and π_4 are more than two posterior standard deviations away from zero, see also Figure 4.3. The Bayes factors for the presence of the seasonal unit roots are clearly below one. Again the posterior means of the remaining parameters are near their true values. Note that the posterior standard deviation of ω_1, is relatively large.

For the third DGP the Bayes factor for the presence of the seasonal unit root -1 exceeds one, as expected. Figure 4.4 shows that the marginal posterior densities of the π_1, π_3 and π_4 are situated far away from zero, which leads to very small Bayes factors for the presence of the roots 1, i and $-i$. This is not surprising since the DGP contains only the root -1 and no other non-zero roots. The posterior means of σ, γ_0, and the ω parameters are near their true values. Note that now the posterior standard deviation of the ω_2 parameter is relatively large compared to the posterior standard deviations of the other ω parameters.

The marginal posterior densities of the π parameters for the fourth DGP are shown in Figure 4.5. The modes of the posterior densities are near zero. Table 4.1 show that the three Bayes factors exceed one and hence indicate that a fourth difference filter is necessary to obtain stationarity. Now, the posterior standard deviations of all ω parameters are relatively large. The posterior means of σ and γ_0 do not differ more than two times the posterior standard deviation from their true value.

In summary, the posterior results of the four DGPs show the applicability of Bayesian seasonal unit root analysis. However, since we have only considered four DGPs, no general conclusions can be drawn about the performance of this approach. In the next subsection we apply the seasonal unit root analysis on German unemployment.

4.6 German Unemployment

A macroeconomic variable which obviously displays seasonal patterns is unemployment. In this section we analyse the business cycle in quarterly observed German unemployment,

[5]The Bayes factors reported in this section are based on flat priors on the 99% HPD region for the parameters of interest, namely π_1, π_2 and (π_3, π_4), as explained in the previous section.

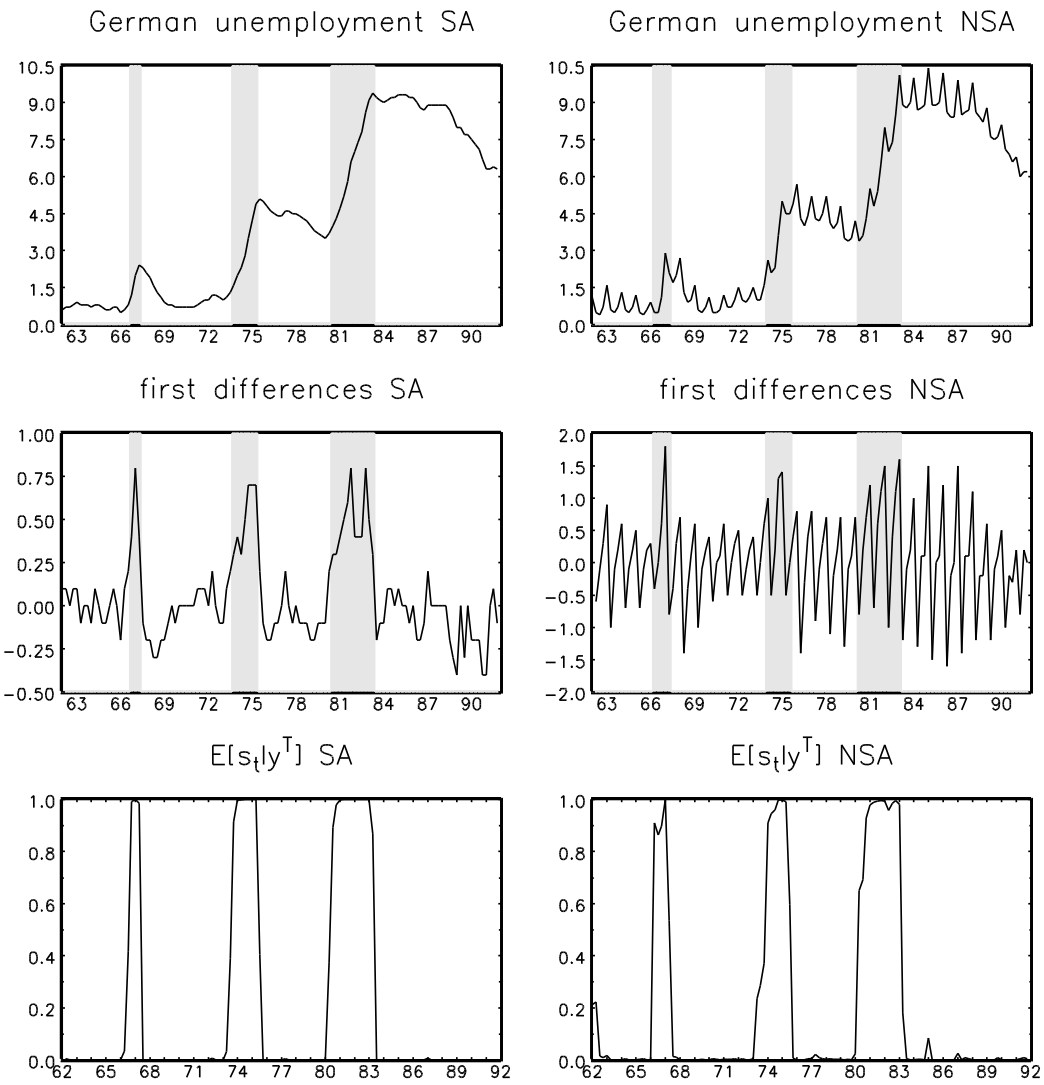

Figure 4.6. Seasonally adjusted [SA] and seasonally unadjusted [NSA] German unemployment. The shaded areas correspond to recessions.

1962.I–1991.IV. Figure 4.6 shows plots of the level and the first differences of the officially seasonally adjusted [SA] and seasonally unadjusted [NSA] German unemployment series. It is clear from the plots that there is more variation in the unadjusted series than the adjusted series. The periods with an increase in unemployment are shorter than the periods with a decrease, which points towards asymmetry in the series. Furthermore, the increase per quarter in the recessions is much larger than the decrease per quarter in the expansion periods.

In this section we analyse the business cycle in German unemployment. To investigate the consequences of seasonal adjustment on the business cycle, we first consider in Section 4.6.1 the seasonal adjusted series. Section 4.6.2 deals with the analysis of the seasonal unadjusted series to examine the relation between seasonal fluctuations and the business cycle.

4.6.1 Seasonal Adjusted Series

To analyse the business cycle in seasonally adjusted German unemployment, we use the Markov trend model from Chapter 3. Although it is not likely from Figure 4.6 that the series is stationary around a Markov trend, we start with a Markov trend model without unit roots in the autoregressive component (3.8) and test for the presence of a unit root. A model with one lag turns out to be the best model to analyse the business cycle in the adjusted series. The priors for the model parameters n_1, p, q and σ are given by (3.16)–(3.18) and (3.21). For γ_0 and γ_1 we take flat priors on the intervals $(-\infty, 0.2]$ and $[0, -\infty)$ respectively. This ensures us that $s_t = 0$ corresponds to an expansion regime and $s_t = 1$ corresponds to an recession regime. The first row of Table 4.2 shows the posterior results of this model. The posterior mean of the ρ parameter (0.98) is very close to one and the presence of a unit root seems likely. Indeed, the Bayes factor for $\rho = 1$, which equals 2.76, favours the unit root hypothesis. This Bayes factor is again constructed such that $[\rho_{lb}, 1]$ corresponds to the 99% HPD region for ρ, see Section 3.3.2 for details.

The second row of Table 4.2 shows the posterior results for a Markov trend model with a unit root in the autoregressive component (3.9). The marginal priors for the model parameters are the same as the marginal priors for the Markov trend model without the unit root imposed. Note that the posterior means and posterior standard deviations of the parameters are almost exactly the same as for the Markov trend stationary model. During an expansion regime the unemployment rate decreases on average with 0.05% per quarter, while during a recession there is an average increase of $-0.05 + 0.53 = 0.48\%$ per quarter. The posterior mean of the probability of staying in this recession is 0.84. The posterior mean of the probability of staying in an expansion regime equals 0.96. The solid lines in Figure 4.7 on page 66 show the marginal posterior densities of these transition probabilities and the γ_0 and γ_1 parameters. The posterior uncertainty about γ_0 and p is smaller than the posterior uncertainty about γ_1 and q. As we already have seen in Section 3.6.1, this may follow from the fact that the number of expansion observations is larger than the number of recession observations. Using (3.44) and (3.45) we can compute the expected duration of a recession and an expansion, which turn out to be 8.0 and 31.4

Table 4.2. Posterior means with posterior standard deviations between parentheses and Bayes factors for $\rho = 1$ for seasonally adjusted German unemployment 1962.I–1991.IV.

model[1]	γ_0	γ_1	n_1	p	q	ρ	σ	BF(ρ)[2]
MTS	−0.05 (0.01)	0.53 (0.04)	0.62 (0.15)	0.96 (0.02)	0.84 (0.07)	0.98 (0.02)	0.15 (0.01)	2.76
MDS	−0.06 (0.04)	0.53 (0.04)	– –	0.96 (0.02)	0.84 (0.07)	1 –	0.15 (0.01)	–

[1] MTS=Markov trend stationary model (3.8) and MDS=Markov difference stationary model (3.9).

[2] BF(ρ) denotes the Bayes factor. A Bayes factor exceeding one implies that $\rho = 1$ is *a posteriori* more likely than $\rho < 1$.

quarters respectively. In the next subsection, we analyse the business cycle in seasonally unadjusted German unemployment

4.6.2 Seasonal Unadjusted Series

The Bayesian analysis of the seasonal unadjusted series is based on the seasonal Markov trend model proposed in Section 4.3. However, first we examine the nature of the seasonal pattern in the series using a model with a deterministic trend instead of a Markov trend. This model follows from the seasonal Markov trend model of Section 4.3 with $\gamma_1 = 0$ and $\delta_{1,s} = 0$, $s = 1, \ldots, 4$. The prior for the model parameters is given by the product of (4.24)–(4.26), (4.30) and a flat prior for γ_0 on $(-\infty, \infty)$. Posterior odds test for zero $\bar{\phi}_i$ parameters, like described in Section 3.6.1 indicate that a model with $k = 9$ lags is necessary to analyse the series. The first column of Table 4.3 shows posterior results for this model. The posterior mean of π_1 is almost zero, but its posterior standard deviation is relatively small. The Bayes factor for $\pi_1 = 0$ equals 1.02, which suggests that trend stationarity and a unit root are *a posteriori* roughly equally likely. This Bayes factor is based on a flat prior for π_1 on the region $[\pi_{1,lb}, 0]$, where the prior parameter $\pi_{1,lb}$ is chosen such that $[\pi_{1,lb}, 0]$ corresponds to the 99% HPD region for π_1, see Section 4.4.3. Likewise, the Bayes factor for $\pi_2 = 0$ is constructed. This Bayes factor equals 1.45 and hence indicates the presence of the seasonal unit root -1. The Bayes factor for $\pi_3 = \pi_4 = 0$ is smaller than one and hence the presence of the seasonal unit roots i and $-i$ is *a posteriori* not likely.

The second column of Table 4.3 shows the posterior results for a seasonal Markov

Table 4.3. Posterior means with posterior standard deviations between parentheses and Bayes factors for $\pi_i = 0$ for seasonally unadjusted German unemployment.

para-	TS+SD[1]		MTS+SD		MDS+SD		MDS+SM[3]	
meters	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.
γ_0	0.06	(0.02)	−0.05	(0.02)	−0.03	(0.05)	0.04	(0.11)
γ_1	0	–	0.54	(0.06)	0.57	(0.08)	0.62	(0.11)
w_1	2.28	(0.60)	2.66	(0.85)	–	–	–	–
w_2	−0.64	(0.35)	−0.64	(0.27)	−0.75	(0.23)	−0.76	(0.23)
w_3	0.16	(0.15)	0.23	(0.12)	0.21	(0.11)	0.21	(0.10)
w_4	−1.01	(0.16)	−1.00	(0.13)	−1.01	(0.11)	−1.05	(0.09)
$\delta_{1,1}$	0	–	0	–	0	–	0.04	(0.05)
$\delta_{1,2}$	0	–	0	–	0	–	0.00	(0.05)
$\delta_{1,3}$	0	–	0	–	0	–	0.02	(0.05)
$\delta_{1,4}$	0	–	0	–	0	–	0.01	(0.05)
p	–	–	0.95	(0.02)	0.95	(0.03)	0.95	(0.03)
q	–	–	0.81	(0.09)	0.75	(0.13)	0.60	(0.22)
π_1	−0.01	(0.00)	−0.02	(0.02)	–	–	–	–
π_2	−0.10	(0.07)	−0.12	(0.07)	−0.13	(0.07)	−0.14	(0.08)
π_3	−0.16	(0.10)	−0.14	(0.10)	−0.13	(0.08)	−0.13	(0.08)
π_4	−0.25	(0.12)	−0.24	(0.10)	−0.30	(0.11)	−0.37	(0.12)
$\bar{\bar{\phi}}_1$	0.98	(0.15)	0.58	(0.11)	0.55	(0.10)	0.58	(0.02)
$\bar{\bar{\phi}}_2$	−0.35	(0.16)	0	–	0	0	0	0
$\bar{\bar{\phi}}_3$	0.16	(0.17)	0	–	0	0	0	0
$\bar{\bar{\phi}}_4$	0.29	(0.16)	0	–	0	0	0	0
$\bar{\bar{\phi}}_5$	0.15	(0.08)	0	–	0	0	0	0
σ	0.29	(0.02)	0.23	(0.02)	0.24	(0.02)	0.24	(0.02)
BF(π_1)[2]	1.02		1.45		–		–	
BF(π_2)	1.45		0.63		0.44		0.46	
BF(π_3, π_4)	0.27		0.06		0.07		0.02	

[1] TS+SD=trend stationary model + seasonal dummies (4.5), MTS+SD=Markov trend stationary model + seasonal dummies (4.5), MDS+SD=Markov difference stationary model + seasonal dummies (4.5) and MDS+SM=Markov difference stationary model + seasonal dummies (4.13) where the $\delta_{0,s}$ parameters are transformed into w_s parameters using (4.10), $s = 1, \ldots, 4$.

[2] BF(π_i) denotes the Bayes factor. A Bayes factor exceeding one implies that $\pi_i = 0$ is *a posteriori* more likely than $\pi_i < 0$. BF(π_3, π_4) denotes the Bayes factor for $\pi_3 = \pi_4 = 0$.

[3] Results are based on normal priors for $\delta_{1,s}$ with zero mean and variance 0.05^2, $s = 1, \ldots, 4$.

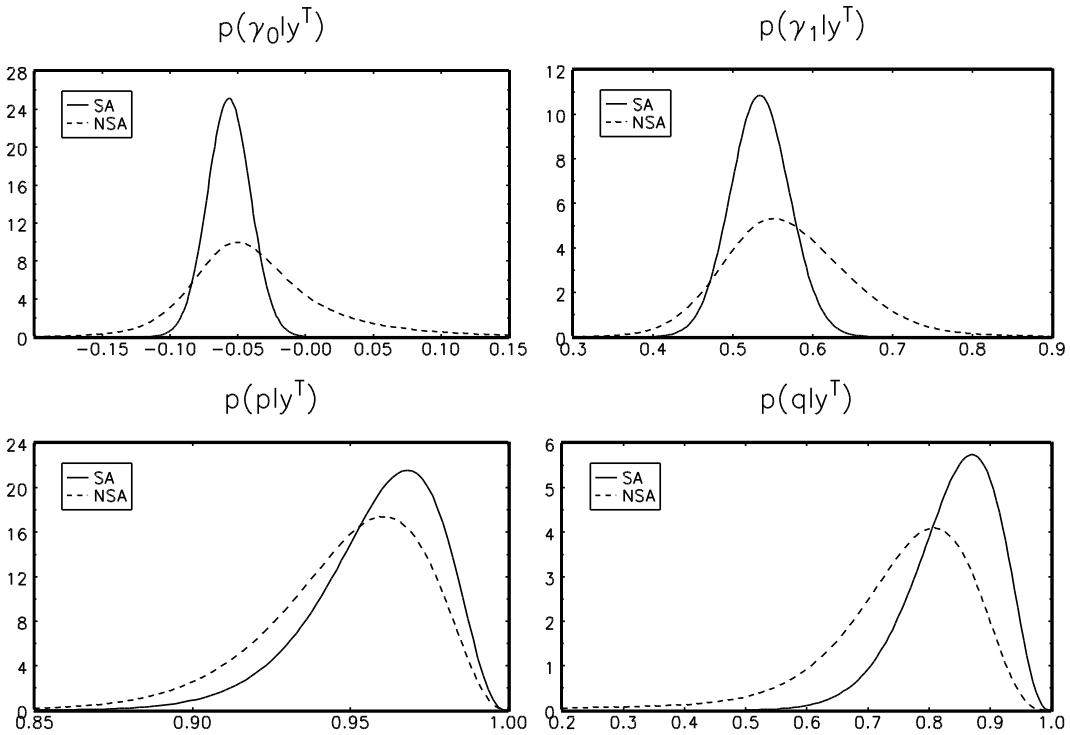

Figure 4.7. Some marginal posterior densities for the seasonal adjusted and unadjusted German unemployment series.

trend model with $\delta_1 = \mathbf{0}$. These results are based on a prior specification (4.24)–(4.30) with a uniform prior for γ_0 on $(-\infty, 0.2]$ and a uniform prior for γ_1 on $[0, \infty)$. Posterior odds tests for $\bar{\bar{\phi}}_i = 0$ suggest that a model of order 5 is enough to analyse the series.[6] Figure 4.8 shows the marginal posterior densities for the π parameters for the seasonal Markov trend model (dashed lines) and for the deterministic trend stationary model discussed above (solid lines). The introduction of a Markov trend does almost have no effect on the modes of the marginal posterior densities of π_3 and π_4. It results only in a minor increase in the variance. The Bayes factor for $\pi_3 = \pi_4 = 0$ decreases from 0.27 to 0.06, see Table 4.3. The mode of the marginal posterior of π_2 however shifts to the left. This leads to a decrease in the Bayes factor for $\pi_2 = 0$ from 1.45 to 0.63 and the presence of the seasonal unit root -1 is not favoured any more. A possible explanation for this phenomenon is that model without the Markov trend interprets a change in the regime as a sudden increase in the seasonal fluctuation so that the hypothesis of a seasonal unit root is more likely, see also Section 4.2 for a discussion. The largest effect of the introduction of a Markov trend can be found in the marginal posterior of π_1. The mode of this posterior shifts to the left but the posterior variance increases substantially, see also Table 4.3. The Bayes factor for $\pi_1 = 0$ (=1.45) clearly favours the non-seasonal unit root hypothesis.

[6]The Bayes factors for the presence of (non)-seasonal unit roots favour the same hypotheses, if we take $k = 9$.

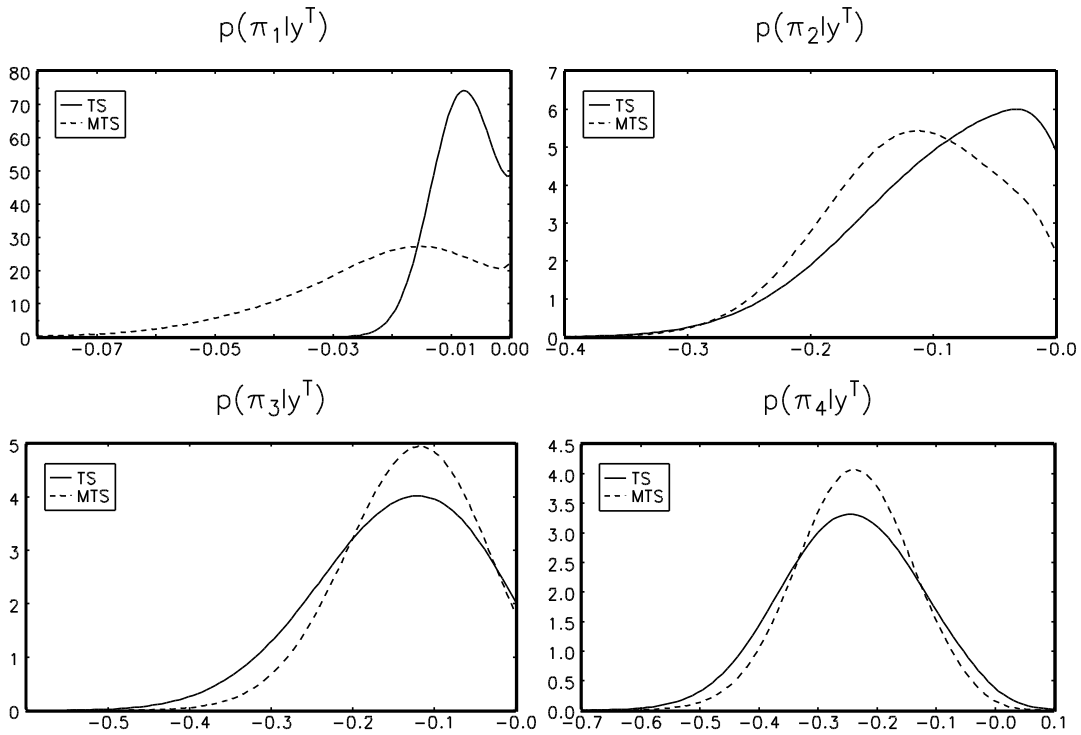

Figure 4.8. Marginal posterior densities of the π parameters under a deterministic trend stationary [TS] specification and under a Markov trend stationary [MTS] specification.

Note that all Bayes factors are again based on uniform priors on the 99% HPD regions for the parameters of interest.

The Bayes factors suggest that a model in first differences without seasonal unit roots is appropriate to analyse the unadjusted series. The third column of Table 4.3 shows the posterior results for a seasonal Markov trend model under the restriction $\pi_1 = 0$ and $\delta_1 = \mathbf{0}$. Note that this model can be rewritten in a model in first differences with drift term $\gamma_0 + \gamma_1 s_t$ and three seasonal dummies with zero mean. Bayes factors for $\pi_2 = 0$ and $\pi_3 = \pi_4$ favour the absence of seasonal units. The parameters in the seasonal Markov trend model have the same interpretation as in the non-seasonal Markov trend model and hence we can compare the results directly. Because of the seasonal fluctuations, we may expect that the posterior standard deviations of the various parameters are larger. The marginal posterior densities of γ_0, γ_1 p and q are displayed in Figure 4.7 together with the marginal posteriors of the same parameters of the Markov difference stationary model for the adjusted series analysed in the previous subsection. The posterior variances of the parameters of the Markov model for the seasonal unadjusted series are clearly larger than for the adjusted series. The posterior means of γ_0 and γ_1 indicate that during an expansion the unemployment rate decreases apart from seasonal fluctuations with 0.03% per quarter and during a recession increases apart from seasonal fluctuations with 0.54% per quarter. The latter percentage is about 0.06% higher than we have found for the seasonal adjusted

Table 4.4. Peaks and troughs dates for German unemployment.[1]

	German unemployment	
	SA	NSA
peak	1966.III	1966.I
trough	1967.II	1967.II
peak	1973.III	1973.IV
trough	1975.II	1975.III
peak	1980.II	1980.I
trough	1983.II	1983.I

[1] Turning points are based on the posterior expectations of the states $\mathrm{E}[s_t|y^T]$.

series. The posterior mean of the probability of staying in a recession for the unadjusted series equals 0.75, which is smaller than the same probability for the seasonally adjusted series, which equals 0.84. The posterior mean of staying in an expansions is 0.95, which is roughly the same as for the adjusted series analysed in the previous subsection. This leads to a drop in the expected duration of a recession from 8.0 to 5.1 quarters and a drop in the expected duration of an expansion from 31.4 to 25.7 quarters for the seasonally unadjusted series.

The last row of of Figure 4.6 shows the posterior expectation of the states $\mathrm{E}[s_t|y^T]$, $t = 1, \ldots, T$ for the Markov difference stationary model (3.9) for the seasonally adjusted series and the Markov difference stationary model with seasonal dummies for the unadjusted series. Using the 0.5 rule as in Hamilton (1989) we can determine the turning points. Again, we define a recession by two consecutive data points for which $\mathrm{E}[s_t|y^T] > 0.5$. A peak is defined by the last expansion observation, while a trough is defined by the last observation in a recession. Table 4.4 shows the peaks and troughs based on the posterior results of the Markov models for the seasonal adjusted and seasonal unadjusted series. For the seasonally unadjusted data, we find that the recession in the eighties starts and ends one quarter earlier than with the adjusted series. In the seventies, however the recession starts and ends one quarter later with the unadjusted series. We detect the largest difference in the sixties, where based on seasonal adjusted series the recession starts in 1966.IV and based on the unadjusted series two quarters earlier. The shaded areas in Figure 4.6 denote the recession periods.

Although the Bayes factors do not indicate the presence of seasonal unit roots, it is

still possible that the seasonal means change during recessions. Note that the graphs in Section 4.2 seem to indicate some change in the seasonal pattern during recessions. To analyse this possible change, we consider the seasonal Markov trend with changing means as in (4.13). The prior specification for the parameters is the same as for the other Markov seasonal trend models analysed in this section. Unfortunately, the likelihood does not contain enough information to impose a flat prior on the $\delta_{1,s}$ parameters, $s = 1, \ldots, 4$. Therefore, we opt for the normal prior specification in (4.27) and compute Bayes factors for $\delta_{1,1} = \delta_{1,2} = \delta_{1,3} = \delta_{1,4} = 0$ for different values of $\sigma_{\delta_{1,s}}$. These Bayes factors are 0.94, 0.62, 0.64 for $\sigma_{\delta_{1,s}} = 0.01$, 0.05 and 0.1, $s = 1, \ldots, 4$, respectively and hence a seasonal mean shift is *a posteriori* not preferred. For lower values of $\sigma_{\delta_{1,s}}$ the Bayes factor are around one and we do not learn from the data. It seems that after correcting for possible changes in the slope of the trend during recessions the changes in the seasonal pattern are not so pronounced any more. The final column of Table 4.3 shows the posterior results for $\sigma_{\delta_{1,s}} = 0.05$, $s = 1, \ldots, 4$. In comparison with the third column we see a decrease in the posterior mean of q and an increase in the posterior mean of γ_0 and γ_1. The posterior means of $\delta_{1,s}$ do not differ more than two standard deviations from zero.

To summarize the empirical findings in this section, we can simply state that transition probabilities and hence the expected duration of a recession and the business cycle turning points obtained from a Markov trend model for German unemployment, differ across seasonal adjusted and unadjusted series. Additionally, there is little evidence in German unemployment for a seasonal mean shift during recessions. After correcting for a change in the growth rate during recessions, there is no posterior indication for changing seasonal patterns due to seasonal unit roots.

4.7 Concluding Remarks

In this chapter we have proposed a seasonal Markov trend model to analyse seasonally unadjusted time series. This model extends the Markov trend model from the previous chapter with seasonal dummies, possible seasonal unit roots and allows for different seasonal means during the stages of the business cycle. The model is used to analyse the business cycle in seasonally unadjusted German unemployment. This analysis indicates that there exists a substantial difference in posterior results, obtained from a Markov trend model, about the transition probabilities and the dating of turning points for seasonal unadjusted and seasonal adjusted series. Additionally, posterior results indicate that there does not seem to be a major change in the seasonal means during recessions.

Just as for the non-seasonal Markov trend model of the previous chapter, we can extend the seasonal Markov trend model with the extensions suggested in Section 3.7. Further extensions of the model concern the modelling of the seasonality in the series. We may for instance use periodic varying transition probabilities as in Ghysels, McCulloch and Tsay (1994) or consider seasonally varying autoregressive parameters as *e.g.* in Osborn and Smith (1989) and Franses and Paap (1994). Finally, one may argue that the dynamics in a time series changes with the business cycle, see for instance, Teräsvirta (1995) and Tiao

and Tsay (1994). However, these extension imply extra parameters and non-linearities in our already ample specified model.

Part II

Multivariate Analysis

Chapter 5

Multivariate Stochastic Trends

In the first part of this thesis, we have used Markov trend models to analyse business cycles and the presence of stochastic trends in univariate time series. In this part we extend these models for multivariate analysis. In this chapter, we give a short introduction into the modelling of trends in multivariate time series. We consider trend specifications, which are multivariate generalisations of the univariate trend specification of Chapter 2. Special attention will be given to situations where time series have one or more common trends. This occurs if the number of trends needed to model the time series in a multivariate model is less than the total number of trends needed to model the separate series in univariate models. The deviations from the trend are usually modelled by a vector autoregressive model. Just as in the univariate case, a unit root in the autoregressive structure implies the presence of a stochastic trend.

The outline of this chapter is as follows. In Section 5.1, we consider some multivariate trend specifications. In Section 5.2 the conditions for common trends in the multivariate trend specifications are discussed. Section 5.3 deals with stochastic trends caused by unit roots in vector autoregressive models. We discuss the conditions for common stochastic trends or cointegration. Finally, we give in Section 5.4 a brief outline of the contents of the second part of this thesis.

5.1 Multivariate Trend Specifications

To model the trend in multivariate time series, we can take for each series a separate univariate trend component. However, the trends of the univariate time series may be correlated and hence this valuable information may be used in a multivariate trend specification. In this section we discuss some multivariate trend specifications, which are generalisations of the univariate trend specifications in Section 2.2. We define the $(n \times 1)$ vector N_t, $t = 1, \ldots, T$ as an n-dimensional trend component.

It is easy to generalise the linear deterministic trend (2.2) to an n-dimensional trend specification

$$N_t = N_{t-1} + \Gamma_0, \tag{5.1}$$

where Γ_0 is now an $(n \times 1)$ parameter vector. This trend implies deterministic linear trends for each element in N_t with n different slopes Γ_0 and n different initial values N_1,

$$N_t = N_1 + \Gamma_0(t-1). \tag{5.2}$$

Again, there is no forecast uncertainty for known values of Γ_0 and N_1. Since this trend does not contain random variables, there is no correlation between the elements of N_t. The multivariate version of the random walk plus drift (2.3) is given by

$$N_t = N_{t-1} + \Gamma_0 + \Gamma_1 U_t, \tag{5.3}$$

where Γ_0 is an $(n \times 1)$ parameter vector, Γ_1 an $(n \times n)$ parameter matrix and the $(n \times 1)$ vector $U_t \sim \text{NID}(\mathbf{0}, \mathbf{I}_n)$. The direction of this trend is given by the n drift terms, which are the elements of Γ_0 plus a linear combination (rows of Γ_1) of n random shocks U_t. The backward solution of this multivariate random walk

$$N_t = N_1 + \Gamma_0(t-1) + \Gamma_1 \sum_{i=2}^{t} U_i. \tag{5.4}$$

consists of a multivariate linear deterministic trend $\Gamma_0(t-1)$ and an n-dimensional stochastic trend $\sum_{i=2}^{t} U_i$. Since the covariance between the elements of the n-dimensional stochastic trend is zero, the trend consists of n independent stochastic trends. The matrix $\Gamma_1\Gamma_1'$ denotes the covariance between these n independent stochastic trends. The expectation of N_t at $t = 1$ is $N_1 + \Gamma_0(t-1)$ and the variance is $\Gamma_1\Gamma_1'(t-1)$. Hence, the forecast uncertainty increases linear over time.

A multivariate generalisation of the Markov trend can be represented as follows[1]

$$N_t = N_{t-1} + \Gamma_0 + \Gamma_1 S_t, \tag{5.5}$$

where Γ_0 is an $(n \times 1)$ parameter vector, Γ_1 an $(n \times n)$ parameter matrix and the n-dimensional vector $S_t = (s_{1,t}, \ldots, s_{n,t})'$ a random variable with $s_{j,t} = 0, 1$, $j = 1, \ldots, n$, which evolves according to a first-order Markov process. The backward solution of the Markov trend

$$N_t = N_1 + \Gamma_0(t-1) + \Gamma_1 \sum_{i=2}^{t} S_i \tag{5.6}$$

consist of the linear deterministic trend $\Gamma_0(t-1)$ and an n-dimensional stochastic Markov trend $\sum_{i=2}^{t} S_i$.

Since the state variable S_t can take 2^n different values, the Markov trend allows many slope directions for each univariate series for large values of n. To model the transitions between these 2^n states there are several possibilities. The easiest possibility is to assume

[1]Other multivariate generalisations of the Markov trend can be found in Phillips (1991a), Kim and Yoo (1995) and Dwyer and Potter (1996).

that each element $s_{j,t}$, $j = 1, \ldots, n$ evolves according to an independent first-order Markov process with transition probabilities

$$\begin{array}{ll} \Pr[s_{j,t} = 0 | s_{j,t-1} = 0] = p_j, & \Pr[s_{j,t} = 1 | s_{j,t-1} = 0] = 1 - p_j, \\ \Pr[s_{j,t} = 1 | s_{j,t-1} = 1] = q_j, & \Pr[s_{j,t} = 0 | s_{j,t-1} = 1] = 1 - q_j. \end{array} \tag{5.7}$$

The trend N_t now consists of n independent stochastic Markov trends and the matrix Γ_1 models the correlation between these Markov trends. Since S_t consist of independent Markov processes, we can use the expectation (2.13) and the variance (2.14) of the univariate Markov trend specification to compute the expectation and variance of N_t at $t = 1$. For instance, the unconditional probability that $s_{j,t} = 1$ equals $\Pr[s_{j,t} = 1] = \pi_j = (1 - p_j)/(2 - p_j - q_j)$, $j = 1, \ldots, n$, see Section 5.1, so that the unconditional expectation of N_t at $t = 1$ equals $(\Gamma_0 + \Gamma_1 P)(t - 1)$ where $P = (\pi_1 \ \pi_2 \ \ldots \ \pi_n)'$. The variance of the Markov trend can be derived in the same way using the result in (2.14).

The most general possibility to model the transitions is to define $(2^n - 1)2^n$ transition probabilities between the 2^n possible realisations of S_t. These transition probabilities can be put in an $(2^n \times 2^n)$ transition matrix, see below for an example. The $(2^n \times 1)$ eigenvector of this transition matrix belonging to the eigenvalue one contains the unconditional probabilities of the 2^n possible realisations of S_t, see Hamilton (1994, p. 681–682) for details. Using the 2^n unconditional probabilities we can compute the unconditional expectation of the Markov trend as before. The derivation of the variance of the Markov trend is however more complicated but can easily be obtained using simulation. Note that the specification with independent Markov processes with transition probabilities (5.7) is just a restricted version of the latter possibility.

To illustrate the multivariate Markov trend, assume that $n = 2$ and $S_t = (s_{1t} \ s_{2t})'$. Now define a new state variable s_t^* which equals

$$s_t^* = \begin{cases} 1 & \text{if } S_t = (0 \ 0)' \\ 2 & \text{if } S_t = (1 \ 0)' \\ 3 & \text{if } S_t = (0 \ 1)' \\ 4 & \text{if } S_t = (1 \ 1)'. \end{cases} \tag{5.8}$$

The direction of the Markov trend at time t is Γ_0 if $s_t^* = 1$, Γ_0 plus the first row of Γ_1 if $s_t^* = 2$, Γ_0 plus the second row of Γ_1 if $s_t^* = 3$ and Γ_0 plus the sum of the rows of Γ_1 if $s_t^* = 4$.

To model the transitions between the four realisations of S_t we define a (4×4) transition matrix. The (i, j)-th element of this matrix denotes the transition probability $\Pr[s_t^* = i | s_{t-1}^* = j]$, $i, j = 1, \ldots, 4$. The most general transition matrix for s_t^* contains $4 \times 3 = 12$ free parameters. If we however assume that $S_t = (s_{1,t} \ s_{2,t})'$ consists of two independent Markov processes with transition probabilities (5.7) the (4×4) transition matrix equals

$$\begin{pmatrix} p_1 p_2 & (1 - q_1) p_2 & p_1 (1 - q_2) & (1 - q_1)(1 - q_2) \\ (1 - p_1) p_2 & q_1 p_2 & (1 - p_1)(1 - q_2) & q_1 (1 - q_2) \\ p_1 (1 - p_2) & (1 - q_1)(1 - p_2) & p_1 q_2 & (1 - q_1) q_1 \\ (1 - p_1)(1 - p_2) & q_1 (1 - p_2) & (1 - p_1) q_2 & q_1 q_2 \end{pmatrix}. \tag{5.9}$$

This transition matrix only consists of four parameters, p_1, p_2, q_1 and q_2.

In this section we have specified n-dimensional trend components to model the trend in multivariate time series. In practice, it is sometimes not realistic to assume different trends for each separate time series. In the next section, we show the conditions for common trends in time series. These conditions imply restrictions on the parameter matrices in the multivariate trend specifications of this section.

5.2 Common Trends

Just as in the univariate case, there are two possibilities to include the trend component in a time series model. One can model the time series in deviation from a trend component N_t or one can simply add the trend component to the time series model in a linear way. In the former approach we assume that an n-dimensional time series $\{Y_t\}_{t=1}^T$ can be decomposed as

$$Y_t = N_t + Z_t, \tag{5.10}$$

where N_t represents the trend component and Z_t the deviations from this trend. As we already have seen in Section 2.2 we can interpret N_t in this specification as a trend in Y_t. In this section we focus on the role of the trend component and assume that Z_t is a stationary process around the multivariate trend component N_t. In the next section we consider the specification of the Z_t component and discuss the second possibility to include the trend component.

The trend components N_t, which we have discussed in the previous section, define for each of the n univariate series in Y_t a single trend. These n trends can be correlated. It may however be the case that less than n trends are sufficient to describe the trend in the n-dimensional time series Y_t. In other words, the univariate series in Y_t have a common trend. We speak of a common trend if a linear combination of two or more series which contain the trend does not contain the trend. Hence, if there are m linearly independent combinations of Y_t, which do not contain the trend, they are $(n-m)$ common trends in Y_t.

It is easy to see that if we specify a linear deterministic trend for N_t (5.1) we assume that there are $(n-1)$ common trends. Define the $(n \times (n-1))$ matrix $\Gamma_{0\perp}$, which is orthogonal to Γ_0, *i.e.* $\Gamma_{0\perp}'\Gamma_0 = \mathbf{0}$. Since $\Gamma_{0\perp}'N_t = \Gamma_{0\perp}'(N_1 + \Gamma_1)$, the linear combinations $\Gamma_{0\perp}'Y_t$ do not contain a linear deterministic trend so that we have one common linear deterministic trend.

To have common stochastic trends, we need linear combinations of N_t (5.4), which cancel out the stochastic trend $\sum_{i=2}^t U_i$. From (5.4) it is clear that if $\text{rank}(\Gamma_1) = n$ it is not possible to find a non-zero matrix, which by premultiplying cancels out $\sum_{i=2}^t U_i$. It is therefore only possible to have common stochastic trends if the rank of Γ_1 is less than n. If the $0 < \text{rank}(\Gamma_1) = m < n$ we can write Γ_1 as the product of two $(n \times m)$ full rank matrices γ and δ

$$\Gamma_1 = \gamma\delta'. \tag{5.11}$$

The $(n-m)$ linear combinations, which remove the stochastic trend from Y_t are given by the rows of the $(n \times (n-m))$ matrix $\gamma_\perp$, which is defined such that $\gamma_\perp'\gamma = \mathbf{0}$. Hence, if the rank of Γ_1 is m there are m common stochastic trends. These m trends are given by $\delta' \sum_{i=2}^{t} U_i$. The matrix γ models the magnitude of the impact of the m common stochastic trends on the elements in N_t and therefore on each of the univariate time series in Y_t. Note that the $(n-m)$ linear combinations $\gamma_\perp'$ which remove the stochastic trend from Y_t do not automatically remove the deterministic trend $\Gamma_0(t-1)$ from N_t in (5.4), since $\gamma_\perp'\Gamma_0$ does not have to be a zero matrix.

Under rank reduction of Γ_1 there are only $(nm + (n-m)m)$ free parameters and hence the δ and γ have to be restricted to become estimable. We propose the following restriction

$$\delta = \begin{pmatrix} \mathbf{I}_m \\ \delta_2 \end{pmatrix}, \tag{5.12}$$

where δ_2 is an $((n-m) \times m)$ matrix. This phenomenon of common stochastic trends is a simple case of cointegration, which will be discussed in the next section in more detail.

Likewise, we can define common stochastic Markov trends in specification (5.5). To have common stochastic Markov trends Γ_1 has to have reduced rank so that it can be written as $\gamma\delta'$ like in (5.11). The $(n-m)$ linear combinations, which remove the trend from Y_t are given by the rows of $\gamma_\perp$ and the m stochastic common Markov trends are $\delta' \sum_{i=2}^{t} S_i$. Note again that the linear combinations which define the common stochastic Markov do not have to remove the deterministic linear trend $\Gamma_0(t-1)$ from (5.6) unless $\gamma_\perp'\Gamma_0 = \mathbf{0}$.

To understand the meaning of a common Markov trend, consider the example of the two-dimensional Markov trend with transition probabilities (5.7) discussed in the previous section. We already have discussed the direction of the Markov trend if Γ_1 has full rank. If the rank of the (2×2) matrix Γ_1 is one, we can write Γ_1 as the product (5.11) of the two (2×1) vectors γ and δ. Hence, the common Markov trend is given by $\delta' \sum_{i=2}^{t} S_i$. Since under rank reduction of Γ_1 there are only three free parameters, we have to to restrict one parameter in δ or γ to make them estimable. We impose restriction (5.12), *i.e.* $\delta' = (1 \ -\delta_2)$. This implies that the common Markov trend at time t apart from the slope in Γ_0 can have four directions: 0, 1, $-\delta_2$ and $(1-\delta_2)$. If $\delta_2 = 1$ or $\delta_2 = -1$ it has three directions (-1, 1 and 0) and for $\delta_2 = 0$ it has only two directions, 1 and 0. The stochastic trend is common since under under rank reduction of Γ_1 the stochastic Markov trend for the first series in Y_t is proportional to the stochastic trend for the second series in Y_t. The elements of γ determines the magnitude of the impact of the common Markov trend on each of the two univariate series.

In the next section, we consider the deviations from the trend $Z_t = Y_t - N_t$. We model these deviations with a vector autoregressive moving average model. Just as in the univariate case, unit roots in the autoregressive part of Z_t correspond to the presence of stochastic trends in Y_t.

5.3 Cointegration

In the previous section, we have discussed the possibility of common stochastic trends in the multivariate trend specification N_t. In this section we consider the deviations Z_t from this trend specification. These deviations are often assumed to be a vector autoregressive moving average [VARMA(k,l)] model

$$Z_t - \sum_{i=1}^{k} \Phi_i Z_{t-i} = \varepsilon_t - \sum_{j=1}^{l} \Psi_i \varepsilon_{t-j}, \tag{5.13}$$

where Φ_i, $i = 1, \ldots, k$ and Ψ_j, $j = 1, \ldots, l$ are $(n \times n)$ parameter matrices and the $(n \times 1)$ vector $\varepsilon_t \sim \text{NID}(0, \Sigma)$ with a positive definite symmetric $(n \times n)$ covariance matrix Σ. Using the lag operator L we can write

$$\begin{aligned} (\mathbf{I}_n - \Phi_1 L - \cdots - \Phi_k L^k) Z_t &= (\mathbf{I}_n - \Psi_1 L - \cdots - \Psi_k L^l)\varepsilon_t, \\ \Phi(L) Z_t &= \Psi(L)\varepsilon_t. \end{aligned} \tag{5.14}$$

The process Z_t is stationary if the roots of $|\Phi(x)|$ are outside the unit circle and invertible if the roots of $|\Psi(x)|$ are outside the unit circle, see Lütkepohl (1993) for an introduction into VARMA models. Just as in the univariate case, a unit root in $\Phi(x)$ corresponds to the presence of a stochastic trend in Z_t. To make this more explicit we neglect the MA component for notational convenience and rewrite the model (5.13) in error correction form

$$\Delta Z_t = \Pi Z_{t-1} + \sum_{i=1}^{k-1} \bar{\Phi}_i \Delta Z_{t-i} + \varepsilon_t, \tag{5.15}$$

where the $(n \times n)$ matrix $\Pi = \sum_{j=1}^{k} \Phi_j - \mathbf{I}_n$ and the $(n \times n)$ matrices $\bar{\Phi}_i = -\sum_{j=i+1}^{k} \Phi_j$, $i = 1, \ldots, k-1$, see *e.g.* Johansen (1991). Since $\Phi(1) = -\Pi$ unit roots enter the model if Π has reduced rank. If $\Pi = \mathbf{0}$, Z_t contains n unit roots and we can write (5.15) as[2]

$$\begin{aligned} \bar{\Phi}(L)\Delta Z_t &= \varepsilon_t, \\ \Delta Z_t &= (\bar{\Phi}(L))^{-1}\varepsilon_t, \end{aligned} \tag{5.16}$$

where $\bar{\Phi}(L) = (\mathbf{I}_n - \bar{\Phi}_1 L \cdots - \bar{\Phi}_{k-1} L^{k-1})$. Defining $C(L) = (\bar{\Phi}(L))^{-1}$ and using (2.21) we obtain the multivariate version of the *Beveridge-Nelson* decomposition

$$Z_t = Z_1 + C(1) \sum_{i=2}^{t} \varepsilon_i + C^*(L)\varepsilon_t, \tag{5.17}$$

where $C(1) = (\bar{\Phi}(1))^{-1}$ and $C^*(L)\varepsilon_t$ is a stationary vector moving average process. Since $C(1)$ is of full rank Z_t contains n stochastic trends $\sum_{i=2}^{t} \varepsilon_i$.

[2]Remember that we do not consider I(2) type trends in this thesis.

If the rank of Π is r with $0 < r < n$, Z_t contains $(n-r)$ unit roots. The matrix Π can be written as a product of two full rank $(n \times r)$ matrices α and β

$$\Pi = \alpha\beta'. \tag{5.18}$$

The multivariate version of the *Beveridge-Nelson* decomposition has the same form as (5.17) but now $C(1) = \beta_\perp(\alpha'_\perp\bar{\Phi}(1)\beta_\perp)^{-1}\alpha'_\perp$ and $\beta_\perp$ and $\alpha_\perp$ are defined such that $\beta'\beta_\perp = \mathbf{0}$ and $\alpha'\alpha_\perp = \mathbf{0}$, see Johansen (1991, p. 49). Note that if $\Pi = \mathbf{0}$, which corresponds to $\beta = \mathbf{0}$ and $\alpha = \mathbf{0}$, $\beta_\perp(\alpha'_\perp\bar{\Phi}(1)\beta_\perp)^{-1}\alpha'_\perp = (\bar{\Phi}(1))^{-1}$. Since $\beta'Z_t = \beta'Z_1 + \beta'C^*(L)\varepsilon_t$ does not contain a stochastic trend, there are $(n-r)$ common stochastic trends. Engle and Granger (1987) refer to this phenomenon as cointegration and the β matrix is called the cointegration vector. The rank r of the matrix Π denotes the number of common stochastic trends and is called the cointegration rank.[3] The cointegrating vector β reflects the stationary long run relations between the elements of Y_t. Since the number of free parameters in Π under rank reduction $(= nr + (n-r)r)$ is smaller than the number of parameters in α and β $(=2nr)$ we have to restrict some parameters. Common used restrictions are $\beta'\beta = \mathbf{I}_r$ or

$$\beta' = (\mathbf{I}_r \ -\beta'_2), \tag{5.19}$$

where β_2 is an $((n-r) \times r)$ matrix. The vector α contains the so-called adjustment parameters, which take care of the adjustment of deviations from long term equilibrium to the equilibrium. From $C(1) = \beta_\perp(\alpha'_\perp\bar{\Phi}(1)\beta_\perp)^{-1}\alpha'_\perp$ it can be seen that the $(n-r)$ common stochastic trends are represented by $\alpha'_\perp\sum_{i=2}^{t}\varepsilon_i$. An extensive treatment of cointegration can be found in Johansen (1995) and Banerjee *et al.* (1993).

As we already have seen in Section 2.3 the trend component N_t can also be added in a linear way to a time series model instead of modelling the time series in deviation from the trend component as in (5.10). The former approach usually leads to a linear model specification. Therefore from an estimation point of view it is more convenient to extend the error correction model (5.15) with deterministic elements in a linear way

$$\Delta Y_t = \mu + \tau(t-1) + \Pi Y_{t-1} + \sum_{i=1}^{k-1}\bar{\Phi}_i\Delta Y_{t-i} + \varepsilon_t, \tag{5.20}$$

where μ and τ are $(n \times 1)$ parameter vectors. If the roots $|\Phi(x)|$ are outside the unit circle it is possible to rewrite Y_t in deviations from a linear deterministic trend like in the univariate case. This is not possible if $(n-r)$ roots of $|\Phi(x)|$ are equal to one. The *Beveridge-Nelson* decomposition in this case reads

$$Y_t = Y_1 + C(1)\sum_{i=2}^{t}(\mu + \tau(i-1) + \varepsilon_i) + C^*(L)(\mu + \tau(t-1) + \varepsilon_t), \tag{5.21}$$

[3]Note that rank reduction in Π implies the presence of more stochastic trends, while rank reduction in Γ_1 in Section 5.1 implies less stochastic trends.

with $C(1) = \beta_\perp(\alpha'_\perp\bar{\Phi}(1)\beta_\perp)^{-1}\alpha'_\perp$. The process Y_t contains $(n-r)$ stochastic trends plus a quadratic deterministic trend, which results from $\sum_{i=2}^{t}\tau(i-1) = \frac{1}{2}\tau(t^2-t)$. In general, this quadratic trend disappears if $\tau = \mathbf{0}$. Note however that under cointegration the restriction $\alpha'_\perp\tau = \mathbf{0}$ also corresponds to the absence of a quadratic trend in Y_t. Likewise, the deterministic linear trend in Y_t disappears if $\tau = \mathbf{0}$ and $\alpha'_\perp\mu = \mathbf{0}$. If however $C(1)$ has full rank the latter condition changes to $\mu = \mathbf{0}$.

In the next chapter we consider a Bayesian analysis of cointegration. For analytical simplicity and to meet with classical studies, we use the error correction model specification (5.20), where the deterministic trend is added in a linear way. As we have seen in (5.21) it is still possible in this specification to relate the deterministic elements to the trend in the series. This is however not the case if we replace $\mu + \tau(t-1)$ by a multivariate random walk plus drift (5.4) or the Markov trend (5.6). Therefore, we consider in Chapter 7 specification (5.10) to analyse multivariate Markov trend models.

5.4 Outline of Part II

Contrary to the Bayesian analysis of the presence of unit roots in univariate time series, there does not exist a Bayesian framework for the analysis of unit roots and cointegration in multivariate time series. Therefore, we propose in Chapter 6 a complete framework for Bayesian cointegration analysis in VAR models. This framework includes, prior specification, posterior odds ratio analysis for determining the number of unit roots or cointegration relations and simulation techniques to obtain the posterior distributions of the cointegration vectors and adjustment parameters. This standard framework will be used in Chapter 7 to analyse stochastic trends in a multivariate version of the Markov trend model of Chapter 3. This model is used to analyse the presence of common Markov trends and cointegration in per capita consumption and income of the United States.

Chapter 6

Cointegration Analysis

6.1 Introduction

In the previous chapters we have analysed the presence of stochastic trends in univariate time series. In this chapter we consider a multivariate analysis of stochastic trends. The introduction of the concept of common stochastic trends or cointegration by Engle and Granger (1987) has introduced a rapidly expanding literature on multivariate analysis of stochastic trends. This has lead to a largely unified theory of classical statistical analysis of cointegration, see among others Johansen (1991) and Phillips (1991b). However, there does not exist a complete framework for Bayesian analysis of cointegration, like in the classical literature. The main contributions to Bayesian analysis of cointegration are: Koop (1991) analyses implied moving averages/impulse responses resulting from the Wold decomposition of a time series, DeJong (1992) considers the posterior distributions of the roots of vector autoregressive models, Kleibergen and van Dijk (1994) analyse the consequences of local non-identification and prior specification on the posteriors of the parameters, Dorfman (1995) tests for the number of cointegrating vectors by analysing the difference between the number of unit roots in the different univariate models and the number of unit roots in the multivariate model, and Geweke (1996) proposes posterior simulators using the Gibbs sampler.

These studies typically consider a specific problem in the sequence of steps involved in Bayesian cointegration analysis. They do not provide a full cointegration model selection strategy which allows one to start with an unrestricted multivariate time series model and to end with the posteriors of the parameters in a cointegration model. In this chapter we provide a full modelling strategy for Bayesian cointegration analysis. The sections, which discuss the different steps in the model selection strategy are organised as follows.

In Section 6.2 we give a short introduction of cointegration in vector autoregressive models and provide the notation we will use in this chapter. A vector autoregressive model is rewritten in an error correction form in which a zero restriction on a parameter matrix reflects cointegration. In Section 6.3 a prior framework for the Bayesian analysis of this error correction model is proposed. This prior framework does not depend on

the functional form of the prior. Therefore, we also derive the functional forms under a diffuse (Jeffreys') and a natural conjugate prior specification. In Section 6.4 the posterior distributions are derived. The posteriors of the cointegration model do not belong to a known class of distribution. In Section 6.4.1 a Metropolis-Hastings simulation algorithm is proposed to simulate from the unknown posterior distributions.

To analyse the number of cointegration relations, we consider in Section 6.5 a Bayesian version of a Lagrange multiplier [LM] statistic. This Bayesian LM statistic can be seen as an alternative to a highest posterior density [HPD] region type test. The statistic is based on a transformation of a multidimensional HPD region into a one-dimensional HPD region. This transformation resembles the functional form of a classical Lagrange multiplier statistic but has a different interpretation. Since the derivation and interpretation of the Bayesian version of an LM statistic to test for cointegration is not straightforward, we use a linear regression model to show the involved steps. In Section 6.6 we propose posterior odds/Bayes factors to compare models with different number of cointegration relations and therefore different number of unit roots. We consider Bayes factors under a natural conjugate prior specification and propose a Bayes factor in case of diffuse priors.

Unfortunately, the outcomes of the Bayes factors and the LM statistics depend on the order of the variables in the error correction model. Therefore, we propose in Section 6.7 an alternative error correction model representation which leads to posterior results that are invariant with respect to the order of the variables. Using the same prior framework as in Section 6.3 we derive priors, posteriors and Bayes factors. Since it is under the invariant specification not possible to derive any analytical marginal prior and posterior results, we propose new simulation procedures to obtain these results.

Finally, Section 6.8 shows some illustrative examples of the derived procedures using four simulated series, the United Kingdom [UK] data analysed in Hendry and Doornik (1994) and the Danish data analysed in Johansen and Juselius (1990). Section 6.9 concludes.

6.2 The Cointegration Model

Consider a vector autoregressive model of order k [VAR(k)] for an n-dimensional vector of time series $\{Y_t\}_{t=1}^T$

$$Y_t = \mu + \tau(t-1) + \sum_{i=1}^{k} \Phi_i Y_{t-i} + \varepsilon_t, \tag{6.1}$$

where ε_t is an independent n-dimensional vector normal process with zero mean and $(n \times n)$ positive definite symmetric covariance matrix Σ. The $(n \times 1)$ vectors μ and τ contain the constant and trend coefficients and Φ_i, $i = 1, \ldots, k$ are $(n \times n)$ matrices with autoregressive coefficients. The initial values $Y_1, \ldots, Y_k$ are fixed. The VAR model in (6.1) can be rewritten in the error correction form

$$\Delta Y_t = \mu + \tau(t-1) + \Pi\, Y_{t-1} + \sum_{i=1}^{k-1} \bar{\Phi}_i \Delta Y_{t-i} + \varepsilon_t, \tag{6.2}$$

where the $(n \times n)$ matrix $\Pi = \sum_{j=1}^{k} \Phi_j - \mathbf{I}_n$ and the $(n \times n)$ matrices $\bar{\Phi}_i = -\sum_{j=i+1}^{k} \Phi_j$, $i = 1, \ldots, k-1$, see *e.g.* Johansen (1991).

The characteristic polynomial of model (6.1) is equal to $|\Phi(z)| = |\mathbf{I}_n z^k - \sum_{i=1}^{k} \Phi_i z^{k-i}|$. Since by definition $\Phi(1) = -\Pi$, unit roots enter the model when $\Phi(1)$ has a lower rank value. If Π is a zero matrix, the characteristic polynomial has n unit roots, which corresponds to n stochastic trends. Common stochastic trends appear if $(n-r)$ roots of the polynomial $|\Phi(z)|$ are equal to one, $0 < r < n$, see Section 5.3. In that case the rank of Π equals r and we say that series generated by model (6.1) are cointegrated. Hence, cointegration implies that we can write the matrix Π as a product of two full rank $(n \times r)$ matrices α and β

$$\Pi = \alpha\beta'. \tag{6.3}$$

The matrix β contains the cointegrating vectors, which reflect the stationary long term relations (or equilibria) between the univariate series in Y_t. The α matrix contains the adjustment parameters, which determine the speed of adjustment to the equilibria $\beta' Y_t$.

Since the number of parameters in $\alpha\beta'$, $2nr$ is larger than the number of free parameters in Π, under reduced rank $(= nr + (n-r)r)$ the α and/or β parameters have to be restricted to become estimable. Here we choose for the following restriction on the cointegration vectors β

$$\beta' = (\mathbf{I}_r \ \ -\beta_2'), \tag{6.4}$$

where β_2 is an $((n-r) \times r)$ matrix. Note that due to this normalization the β matrix has always full rank.

To save on notation we write the error correction model (6.2) in matrix notation,

$$\Delta Y = Y_{-1}\Pi' + X\Phi + \varepsilon, \tag{6.5}$$

where $\Delta Y = (\Delta Y_{k+1} \ldots \Delta Y_T)'$, $Y_{-1} = (Y_k \ldots Y_{T-1})'$, $\varepsilon = (\varepsilon_{k+1} \ldots \varepsilon_T)'$, $X = (X_{k+1}' \ldots X_T')'$, $X_t = (\Delta Y_{t-1}' \ldots \Delta Y_{t-k+1}' \ \ 1 \ \ (t-1))$, and $\Phi = (\bar{\Phi}_1 \ldots \bar{\Phi}_{k-1} \ \ \mu \ \ \tau)'$. To save even further on notation, we focus in the remainder of this chapter on a simple VAR(1) model without deterministic elements. This is not a serious restriction since under a diffuse prior specification on Φ, integrating out the Φ parameters from the likelihood function leads to analysing a VAR(1) model for the transformed data $M_X \Delta Y$ and $M_X Y_{-1}$, where $M_X = \mathbf{I}_{T-k} - X(X'X)^{-1}X'$. We refer to this VAR(1) model in error correction form as a linear error correction [lec] model

$$\Delta Y = Y_{-1}\boldsymbol{\Pi} + \varepsilon, \tag{6.6}$$

where we define $\boldsymbol{\Pi} = \Pi'$ for notational convenience. Under the restriction $\Pi = \alpha\beta'$ this model simplifies to an error correction cointegration [ecc] model

$$\begin{aligned} \Delta Y &= Y_{-1}\beta\boldsymbol{\alpha} + \varepsilon \\ &= Y_{1,-1}\boldsymbol{\alpha} - Y_{2,-1}\beta_2\boldsymbol{\alpha} + \varepsilon, \end{aligned} \tag{6.7}$$

where $Y_{1,-1}$ consists of the first r columns of Y_{-1}, $Y_{2,-1}$ consists of the last $(n-r)$ columns of Y_{-1} and $\boldsymbol{\alpha} = \alpha'$. To clarify the parameter restriction, which leads to rank reduction in Π, we reparameterise the linear error correction model (6.6) into an unrestricted error correction [uec] model

$$\Delta Y = Y_{1,-1}\boldsymbol{\alpha} - Y_{2,-1}\beta_2\boldsymbol{\alpha} + Y_{2,-1}(\mathbf{0} \;\; \lambda) + \varepsilon, \tag{6.8}$$

where λ is an $((n-r) \times (n-r))$ matrix. The error correction cointegration model (6.7) corresponds to $\lambda = \mathbf{0}$, see Kleibergen (1996b) and Kleibergen and van Dijk (1994) for details. We have rewritten the linear error correction model (6.6) in such a way that rank reduction corresponds to a parameter restriction. To see the relation between this unrestricted error correction model and the linear error correction model (6.6) we decompose $\boldsymbol{\Pi}$ in submatrices

$$\boldsymbol{\Pi} = \begin{pmatrix} \boldsymbol{\Pi}_{11} & \boldsymbol{\Pi}_{12} \\ \boldsymbol{\Pi}_{21} & \boldsymbol{\Pi}_{22} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\alpha}_1 & \boldsymbol{\alpha}_2 \\ -\beta_2\boldsymbol{\alpha}_1 & -\beta_2\boldsymbol{\alpha}_2 + \lambda \end{pmatrix}, \tag{6.9}$$

where $\boldsymbol{\Pi}_{11}$ is an $(r \times r)$ matrix, $\boldsymbol{\Pi}_{12}$ is an $(r \times (n-r))$ matrix, $\boldsymbol{\Pi}_{21}$ is an $((n-r) \times r)$ matrix, $\boldsymbol{\Pi}_{22}$ is an $((n-r) \times (n-r))$ matrix and where $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1 \;\; \boldsymbol{\alpha}_2)$ with $\boldsymbol{\alpha}_1$ an $(r \times r)$ matrix and $\boldsymbol{\alpha}_2$ an $(r \times (n-r))$ matrix.

Note that the decomposition of $\boldsymbol{\Pi}$ in (6.9) is not unique. A different order of the elements of Y_t results in a different definition of λ. This implies that the posterior of λ also depends on the order of the univariate series in Y_t. Tests, which are based on this posterior, like the posterior odds ratios in Section 6.6, are therefore sensitive to the order of the series in Y_t. The Bayesian version of the LM statistic, which is discussed in Section 6.5.2, is less sensitive for the order since it is calculated using the posterior of the parameters of the error correction cointegration model (6.7) in which $\lambda = \mathbf{0}$, see also Lucas (1996, chapter 8) for a discussion about this phenomenon. In Section 6.7 we discuss a decomposition of $\boldsymbol{\Pi}$, which leads to posterior results which are invariant to the order of the variables in Y_t. The disadvantage of this decomposition is that contrary to the simple decomposition in (6.9) it does not allow for an analytical decomposition of the prior and posterior in known conditional/marginal densities. Posterior results are obtained using complicated simulation methods. Therefore, we first discuss as an introduction in the next sections Bayesian cointegration analysis using the simple decomposition in (6.9). In Section 6.7 we consider the invariant decomposition and discuss the computation of the order invariant posterior results.

The Likelihood Function

The likelihood function of the unrestricted error correction model (6.8) conditional on the initial observations Y_1 is given by

$$\begin{aligned} \mathcal{L}_{uec}(Y|\Sigma, \boldsymbol{\alpha}, \lambda, \beta_2) &= (\sqrt{2\pi})^{-(T-1)n} |\Sigma \otimes \mathbf{I}_n|^{-\frac{1}{2}} \exp(-\frac{1}{2}(\mathrm{vec}(\varepsilon)'(\Sigma^{-1} \otimes \mathbf{I}_n^{-1})\mathrm{vec}(\varepsilon))) \\ &= (\sqrt{2\pi})^{-(T-1)n} |\Sigma|^{-\frac{1}{2}(T-1)} |\mathbf{I}_{T-1}|^{-\frac{1}{2}n} \exp(-\frac{1}{2}\mathrm{tr}(\Sigma^{-1}\varepsilon'\varepsilon)), \end{aligned} \tag{6.10}$$

where ε is given in (6.8). In addition, the likelihood function of the cointegration model (6.7) equals the likelihood function of the unrestricted model (6.10) evaluated in $\lambda = \mathbf{0}$

$$\mathcal{L}_{ecc}(Y|\Sigma, \boldsymbol{\alpha}, \beta_2) = \mathcal{L}_{uec}(Y|\Sigma, \boldsymbol{\alpha}, \lambda, \beta_2)|_{\lambda=\mathbf{0}}. \tag{6.11}$$

In the next section we propose a prior framework to analyse the unrestricted error correction models (6.8) and the cointegration models (6.7).

6.3 Prior Framework

Traditional Bayesian analysis of the cointegration model starts directly with specifying priors on the parameters Σ, $\boldsymbol{\alpha}$ and β_2 in the cointegration model (6.7). The cointegration model (6.7) is non-linear in the parameters $\boldsymbol{\alpha}$ and β_2. It is easy to see that the parameter β_2 is not identified when $\boldsymbol{\alpha} = \mathbf{0}$ (or when $\boldsymbol{\alpha}$ is of reduced rank), see Phillips (1989) for more discussion on local non-identification. Consequently, if a diffuse prior is used, such that the joint posterior of the parameters is proportional to the likelihood, the conditional posterior of β_2 given $\boldsymbol{\alpha}$ is constant and non-zero when $\boldsymbol{\alpha} = \mathbf{0}$. The integral over this conditional posterior at $\boldsymbol{\alpha} = \mathbf{0}$, which is part of the marginal posterior of $\boldsymbol{\alpha}$, is therefore proportional to the volume of the parameter region of β_2 ($\mathbb{R}^{(n-r)r}$), which is infinity. This leads to a *a posteriori* favour for locally non-identified parameter values when diffuse priors are used for the parameters $(\boldsymbol{\alpha}, \beta_2)$, see Kleibergen and van Dijk (1994) for a more elaborate discussion of this phenomenon. Diffuse priors for models which are non-linear in the parameters like the cointegration model (6.7) do not lead to posteriors with similar properties as posteriors of linear models under a diffuse prior specification. Hence, from a posterior perspective, diffuse priors in non-linear models like the cointegration model (6.7) are not the natural extension of diffuse priors in linear models.

Since the error correction cointegration model (6.7) is nested in the unrestricted error correction model (6.8) it seems natural that the joint posterior distribution of the unrestricted error correction model (6.8) evaluated in $\lambda = \mathbf{0}$ is proportional to the joint posterior of the cointegration model (6.7). Note that in case of linear restrictions, this requirement is automatically fulfilled, as we can condition on the parameters modelling the restriction. For non-linear restrictions, like cointegration, however, we cannot condition on the restrictions and we have to use the joint posterior explicitly, see also Kleibergen (1996a). Since the likelihood of the cointegration model (6.7) equals the likelihood of the unrestricted error correction model (6.8) evaluated in $\lambda = \mathbf{0}$ the prior has to obey the same rule. Note that the unrestricted error correction model (6.8) is also non-linear and suffers from the same identification problem as the cointegration model. However, the unrestricted error correction model is observationally equivalent with the *linear* error correction model (6.6). Therefore, specifying diffuse priors for the Σ and $\boldsymbol{\Pi}$ parameters in the latter model implies priors for the Σ, $\boldsymbol{\alpha}$, λ and β_2 parameters in the unrestricted error correction model (6.8), which leads to well-behaved posterior distributions. The joint prior of the error correction cointegration model (6.7) equals the joint prior of the unrestricted error correction model (6.8) evaluated in $\lambda = \mathbf{0}$. This prior framework can be

seen as a natural extension of defining diffuse prior in non-linear error correction models and does not lead to improper posteriors, see also Kleibergen (1996a). Notice that every cointegration model, $r = 1, \ldots, n$, is nested in the linear error correction model (6.6), which ensures also that there exists a natural transition between the joint posteriors of cointegration models with different number of cointegration relations.

The ideas behind this prior framework can also be applied to natural conjugate priors, like the informative Minnesota priors of Doan, Litterman and Sims (1984) and Litterman (1986). In general one starts with specifying a prior on the parameters $\mathbf{\Pi}$ and Σ of the linear error correction model (6.6)

$$\begin{array}{c} p_{lec}(\Sigma), \\ p_{lec}(\mathbf{\Pi}|\Sigma). \end{array} \tag{6.12}$$

Note that we have not specified a functional form for the priors yet. The prior on $\mathbf{\Pi}$ implies a prior on the submatrices of $\mathbf{\Pi}$ defined in (6.9)

$$\begin{array}{c} p_{lec}(\mathbf{\Pi}_{11}, \mathbf{\Pi}_{12}|\Sigma), \\ p_{lec}(\mathbf{\Pi}_{21}, \mathbf{\Pi}_{22}|\mathbf{\Pi}_{11}, \mathbf{\Pi}_{12}, \Sigma). \end{array} \tag{6.13}$$

The decomposition in (6.9) shows the relation between the submatrices of $\mathbf{\Pi}$ and the parameters $\boldsymbol{\alpha}$, λ and β_2. The prior distribution of $\boldsymbol{\alpha}$, λ and β_2 can now be constructed using the Jacobian of the transformation of $\mathbf{\Pi}$ to $(\boldsymbol{\alpha}, \lambda, \beta_2)$. The priors for $\boldsymbol{\alpha}$, λ and β_2 implied by the prior for $\mathbf{\Pi}_{11}$, $\mathbf{\Pi}_{21}$, $\mathbf{\Pi}_{12}$ and $\mathbf{\Pi}_{22}$ are constructed such that they obey the sequence, in which the parameter matrices should be analysed conditional on one another, dictated by the model: λ has to be analysed given Σ and $\boldsymbol{\alpha}$, and the cointegrating vectors β_2 have to be analysed given Σ, $\boldsymbol{\alpha}$, and λ. Only this sequence allows for an analytical decomposition of the joint prior/posterior into conditional posteriors/priors as will be shown in Section 6.4. The marginal/conditional priors read

$$\begin{array}{c} p_{uec}(\Sigma), \\ p_{uec}(\boldsymbol{\alpha}|\Sigma), \\ p_{uec}(\lambda|\Sigma, \boldsymbol{\alpha}), \\ p_{uec}(\beta_2|\Sigma, \boldsymbol{\alpha}, \lambda). \end{array} \tag{6.14}$$

Note that of course $p_{uec}(\Sigma) = p_{lec}(\Sigma)$ and $p_{uec}(\boldsymbol{\alpha}|\Sigma) = p_{lec}(\mathbf{\Pi}_{11}, \mathbf{\Pi}_{12}|\Sigma)_{(\mathbf{\Pi}_{11}\ \mathbf{\Pi}_{12})=\boldsymbol{\alpha}}$. The joint prior of the unrestricted error correction model (6.8) $p_{uec}(\Sigma, \boldsymbol{\alpha}, \lambda, \beta_2)$ is given by the product of the conditional/marginal priors in (6.14).

Finally, the joint prior for the error correction cointegration model (6.7) is proportional to the joint prior for the unrestricted error correction model (6.8) evaluated in $\lambda = \mathbf{0}$

$$p_{ecc}(\Sigma, \boldsymbol{\alpha}, \beta_2) = \frac{1}{c_r}\, p_{uec}(\Sigma, \boldsymbol{\alpha}, \lambda, \beta_2)|_{\lambda=\mathbf{0}}, \tag{6.15}$$

where the constant c_r is a correction factor, which corrects the integrating constant in p_{uec}. This correction factor is defined by

$$c_r = \iiint p_{uec}(\Sigma, \boldsymbol{\alpha}, \lambda, \beta_2)|_{\lambda=\mathbf{0}}\, d\Sigma\, d\boldsymbol{\alpha}\, d\beta_2. \tag{6.16}$$

To demonstrate the implications of this prior framework, we derive the implicit conditional priors the $\boldsymbol{\alpha}$, β_2, and λ parameters if we specify a diffuse (Jeffreys') and a natural conjugate prior for the $\boldsymbol{\Pi}$ parameter.

Diffuse Prior Specification

A Jeffreys' prior for Π and Σ is proportional to the square root of the determinant of the information matrix of the linear error correction model (6.6). Since the information matrix is block diagonal [see Lütkepohl (1993, section 3.4)] the Jeffreys' prior for Π results from the information matrix of Π given Σ

$$-\mathrm{E}\left[\frac{\partial^2 \ln \mathcal{L}(Y|\Sigma,\boldsymbol{\Pi})}{\partial(\mathrm{vec}(\boldsymbol{\Pi}))\partial(\mathrm{vec}(\boldsymbol{\Pi}))'}\,\middle|\,\Sigma\right] = (\Sigma^{-1}\otimes(Y_{-1}'Y_{-1})). \tag{6.17}$$

Note that we only take the expectation of the disturbances and we, as Bayesians, treat Y_1 as fixed and given so that the expectation equals its realisation. This implies that the diffuse prior for the linear error correction model (6.6) reads

$$\begin{aligned} p_{lec}(\Sigma) &\propto |\Sigma|^{-\frac{1}{2}(n+1)}, \\ p_{lec}(\boldsymbol{\Pi}|\Sigma) &\propto |\Sigma|^{-\frac{1}{2}n}|Y_{-1}'Y_{-1}|^{\frac{1}{2}n}. \end{aligned} \tag{6.18}$$

The conditional prior for $\boldsymbol{\Pi}$ implies the following conditional priors for the submatrices of $\boldsymbol{\Pi}$

$$\begin{aligned} p_{lec}(\boldsymbol{\Pi}_{11},\boldsymbol{\Pi}_{12}|\Sigma) &\propto |\Sigma|^{-\frac{1}{2}r}|Y_{1,-1}'M_{Y_{2,-1}}Y_{1,-1}|^{\frac{1}{2}n}, \\ p_{lec}(\boldsymbol{\Pi}_{21},\boldsymbol{\Pi}_{22}|\boldsymbol{\Pi}_{11},\boldsymbol{\Pi}_{12},\Sigma) &\propto |\Sigma|^{-\frac{1}{2}(n-r)}|Y_{2,-1}'Y_{2,-1}|^{\frac{1}{2}n}, \end{aligned} \tag{6.19}$$

where $M_{Y_{2,-1}} = \mathbf{I}_{T-1} - Y_{2,-1}(Y_{2,-1}'Y_{2,-1})^{-1}Y_{2,-1}'$. Appendix 6.A shows the derivation of the priors for unrestricted error correction model (6.8) based on the prior (6.19) including the Jacobian for the transformation of $(\boldsymbol{\Pi}_{11},\boldsymbol{\Pi}_{12},\boldsymbol{\Pi}_{21},\boldsymbol{\Pi}_{22})$ to $(\boldsymbol{\alpha},\ \lambda,\ \beta_2)$. These priors read

$$\begin{aligned} p_{uec}(\Sigma) &\propto |\Sigma|^{-\frac{1}{2}(n+1)}, \\ p_{uec}(\boldsymbol{\alpha}|\Sigma) &\propto |\Sigma|^{-\frac{1}{2}r}|Y_{1,-1}'M_{Y_{2,-1}}Y_{1,-1}|^{\frac{1}{2}n}, \\ p_{uec}(\lambda|\Sigma,\boldsymbol{\alpha}) &\propto |\boldsymbol{\alpha}_\perp\Sigma\boldsymbol{\alpha}_\perp'|^{-\frac{1}{2}(n-r)}|Y_{2,-1}'Y_{2,-1}|^{\frac{1}{2}(n-r)}, \\ p_{uec}(\beta_2|\Sigma,\boldsymbol{\alpha},\lambda) &\propto |\boldsymbol{\alpha}\Sigma^{-1}\boldsymbol{\alpha}'|^{\frac{1}{2}(n-r)}|Y_{2,-1}'Y_{2,-1}|^{\frac{1}{2}r}, \end{aligned} \tag{6.20}$$

where $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1\ \boldsymbol{\alpha}_2)$ with $\boldsymbol{\alpha}_1$ an $(r\times r)$ matrix and $\boldsymbol{\alpha}_2$ an $(r\times(n-r))$ matrix and $\boldsymbol{\alpha}_\perp = (-\boldsymbol{\alpha}_2'(\boldsymbol{\alpha}_1^{-1})'\ \ \mathbf{I}_{n-r})$. The joint prior $p_{uec}(\Sigma,\boldsymbol{\alpha},\lambda,\beta_2)$ is given by the product of the marginal/conditional priors in (6.20).

Now, the joint prior for the error correction cointegration model (6.7) is given by

$$p_{ecc}(\Sigma,\boldsymbol{\alpha},\beta_2) \propto p_{uec}(\Sigma,\boldsymbol{\alpha},\lambda,\beta_2)|_{\lambda=\mathbf{0}}. \tag{6.21}$$

or

$$
\begin{aligned}
p_{ecc}(\Sigma, \boldsymbol{\alpha}) &\propto |\Sigma|^{-\frac{1}{2}(n+r+1)} |Y'_{1,-1} M_{Y_{2,-1}} Y_{1,-1}|^{\frac{1}{2}n} |Y'_{2,-1} Y_{2,-1}|^{\frac{1}{2}(n-r)} |\boldsymbol{\alpha}_\perp \Sigma \boldsymbol{\alpha}'_\perp|^{-\frac{1}{2}(n-r)}, \\
p_{ecc}(\beta_2 | \Sigma, \boldsymbol{\alpha}) &\propto |\boldsymbol{\alpha} \Sigma^{-1} \boldsymbol{\alpha}'|^{\frac{1}{2}(n-r)} |Y'_{2,-1} Y_{2,-1}|^{\frac{1}{2}r}.
\end{aligned} \tag{6.22}
$$

Note that in the case of diffuse priors we cannot define c_r according to (6.16). For the derivation of the posteriors the value of c_r is not important since it is only an integrating constant. However, for model comparison via posterior odds this integrating constant is important. In Section 6.6.1 we discuss the value of c_r in posterior odds analysis in case of a diffuse prior specification.

Natural Conjugate Prior Specification

In case of natural conjugate priors we specify an inverted Wishart prior for Σ and a matrix normal prior for $\boldsymbol{\Pi}$ given Σ

$$
\begin{aligned}
p_{lec}(\Sigma) &\propto |S|^{\frac{1}{2}h} |\Sigma|^{-\frac{1}{2}(h+n+1)} \exp(-\frac{1}{2} \mathrm{tr}(\Sigma^{-1} S)), \\
p_{lec}(\boldsymbol{\Pi} | \Sigma) &\propto |\Sigma|^{-\frac{1}{2}n} |A|^{\frac{1}{2}n} \exp(-\frac{1}{2} \mathrm{tr}(\Sigma^{-1} (\boldsymbol{\Pi} - P)' A (\boldsymbol{\Pi} - P))),
\end{aligned} \tag{6.23}
$$

where h and the positive definite symmetric [PDS] $(n \times n)$ matrix S are prior parameters for the inverted Wishart and the PDS $(n \times n)$ matrix A and the $(n \times n)$ matrix P prior parameters for the matrix normal prior. The matrices A and P can be decomposed as

$$
A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \quad \text{and} \quad P = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix}, \tag{6.24}
$$

where A_{11}, P_{11} are $(r \times r)$ matrices, A_{12}, P_{12} $(r \times (n-r))$ matrices, A_{21}, P_{21} $((n-r) \times r)$ matrices and A_{22}, P_{22} $((n-r) \times (n-r))$ matrices. Appendix 6.A shows the derivation of the priors for the unrestricted error correction model (6.8) implied by the priors for the parameters of the linear error correction model (6.23). The marginal/conditional priors read

$$
\begin{aligned}
p_{uec}(\Sigma) &\propto |S|^{\frac{1}{2}h} |\Sigma|^{-\frac{1}{2}(h+n+1)} \exp(-\frac{1}{2} \mathrm{tr}(\Sigma^{-1} S)), \\
p_{uec}(\boldsymbol{\alpha} | \Sigma) &\propto |\Sigma|^{-\frac{1}{2}r} |A_{11.2}|^{\frac{1}{2}n} \exp(-\frac{1}{2} \mathrm{tr}(\Sigma^{-1} (\boldsymbol{\alpha} - (P_{11}\ P_{12}))' A_{11.2} (\boldsymbol{\alpha} - (P_{11}\ P_{12})))), \\
p_{uec}(\lambda | \Sigma, \boldsymbol{\alpha}) &\propto |\boldsymbol{\alpha}_\perp \Sigma \boldsymbol{\alpha}'_\perp|^{-\frac{1}{2}(n-r)} |A_{22}|^{\frac{1}{2}(n-r)} \\
&\qquad \exp(-\frac{1}{2} \mathrm{tr}((\boldsymbol{\alpha}_\perp \Sigma \boldsymbol{\alpha}'_\perp)^{-1} (\lambda - l)' A_{22} (\lambda - l))), \\
p_{uec}(\beta_2 | \Sigma, \boldsymbol{\alpha}, \lambda) &\propto |\boldsymbol{\alpha} \Sigma^{-1} \boldsymbol{\alpha}'|^{\frac{1}{2}(n-r)} |A_{22}|^{\frac{1}{2}r} \exp(-\frac{1}{2} \mathrm{tr}(A_{22} (\beta_2 - b_2) \boldsymbol{\alpha} \Sigma^{-1} \boldsymbol{\alpha}' (\beta_2 - b_2)')),
\end{aligned} \tag{6.25}
$$

where $A_{11.2} = A_{11} - A_{12} A_{22}^{-1} A_{21}$, $(Q_{21}\ Q_{22}) = (P_{21}\ P_{22}) - A_{22}^{-1} A_{21} ((\boldsymbol{\Pi}_{11}\ \boldsymbol{\Pi}_{12}) - (P_{11}\ P_{12}))$, $l = (Q_{21}\ Q_{22}) \boldsymbol{\alpha}'_\perp$ and $b_2 = -(Q_{21}\ (Q_{22} - \lambda)) \Sigma^{-1} \boldsymbol{\alpha}' (\boldsymbol{\alpha} \Sigma^{-1} \boldsymbol{\alpha}')^{-1}$. The joint prior for the

parameters in the unrestricted error correction model (6.8), $p_{uec}(\Sigma, \boldsymbol{\alpha}, \lambda, \beta_2)$, is given by the product of the marginal/conditional priors in (6.25). Note that it is also possible to specify the marginal/conditional priors (6.25) for a cointegration model with rank r directly without starting with the prior specification for the linear error correction model (6.23). These priors automatically imply the priors for the cointegration models with rank $\neq r$. Furthermore, to reflect the prior belief about a cointegration rank, one can assign prior probabilities to every cointegration rank in a posterior odds analysis, see Section 6.6.

Now, the joint prior for the error correction cointegration model (6.7) is given by

$$p_{ecc}(\Sigma, \boldsymbol{\alpha}, \beta_2) = \frac{1}{c_r} p_{uec}(\Sigma, \boldsymbol{\alpha}, \lambda, \beta_2)|_{\lambda=\mathbf{0}}. \tag{6.26}$$

or

$$\begin{aligned} p_{ecc}(\Sigma, \boldsymbol{\alpha}) &\propto p_{uec}(\Sigma) p_{uec}(\boldsymbol{\alpha}|\Sigma) p_{uec}(\lambda|\Sigma, \boldsymbol{\alpha})|_{\lambda=\mathbf{0}}, \\ p_{ecc}(\beta_2|\Sigma, \boldsymbol{\alpha}) &\propto p_{uec}(\beta_2|\Sigma, \boldsymbol{\alpha}, \lambda)|_{\lambda=\mathbf{0}} \end{aligned} \tag{6.27}$$

where p_{uec}'s are defined in (6.25). The value of c_r can be computed according to (6.16). In Section 6.6 we provide a simulation procedure to obtain c_r.

In the next section we derive posterior distributions for the parameters of the unrestricted error correction and the cointegration models both for the diffuse and natural conjugate prior specification.

6.4 Posterior Distributions

As already mentioned, only for the specific sequence of the parameters, in which we stated the conditional priors of the parameters in the previous section, it is possible to derive analytical expressions for the conditional posteriors of the parameters from the unrestricted error correction model (6.8). If we follow this specific conditioning sequence, λ has to be analysed given $(\Sigma, \boldsymbol{\alpha})$ and β_2 has to be analysed given $(\Sigma, \boldsymbol{\alpha}, \lambda)$. First, we consider the posterior distributions of the restricted and unrestricted error correction models in case of diffuse priors.

Posterior under Diffuse Prior Specification

The posterior of the unrestricted error correction model (6.8) is proportional to the likelihood (6.10) times the product of the marginal/conditional priors in (6.20). Due to the structure of the model it is only possible to derive an analytical expression for marginal posteriors of Σ and $\boldsymbol{\alpha}$. Fortunately, there exist a decomposition of the posterior in known

densities, see Appendix 6.B. The kernels of these marginal/conditional posteriors read

$$\begin{aligned}
p_{uec}(\Sigma|Y) &\propto |\Sigma|^{-\frac{1}{2}(T+n)} \exp(-\frac{1}{2}\mathrm{tr}(\Sigma^{-1}\Delta Y' M_{Y_{-1}} \Delta Y)), \\
p_{uec}(\boldsymbol{\alpha}|\Sigma, Y) &\propto |\Sigma|^{-\frac{1}{2}r} |Y'_{1,-1} M_{Y_{2,-1}} Y_{1,-1}|^{\frac{1}{2}n} \\
&\quad \exp(-\frac{1}{2}\mathrm{tr}(\Sigma^{-1}(\boldsymbol{\alpha}-\hat{\boldsymbol{\alpha}})' Y'_{1,-1} M_{Y_{2,-1}} Y_{1,-1}(\boldsymbol{\alpha}-\hat{\boldsymbol{\alpha}}))), \\
p_{uec}(\lambda|\Sigma, \boldsymbol{\alpha}, Y) &\propto |\boldsymbol{\alpha}_\perp \Sigma \boldsymbol{\alpha}'_\perp|^{-\frac{1}{2}(n-r)} |Y'_{2,-1} Y_{2,-1}|^{\frac{1}{2}(n-r)} \\
&\quad \exp(-\frac{1}{2}\mathrm{tr}((\boldsymbol{\alpha}_\perp \Sigma \boldsymbol{\alpha}'_\perp)^{-1}(\lambda-\hat{\lambda})' Y'_{2,-1} Y_{2,-1}(\lambda-\hat{\lambda}))), \\
p_{uec}(\beta_2|\Sigma, \boldsymbol{\alpha}, \lambda, Y) &\propto |\boldsymbol{\alpha}\Sigma^{-1}\boldsymbol{\alpha}'|^{\frac{1}{2}(n-r)} |Y'_{2,-1} Y_{2,-1}|^{\frac{1}{2}r} \\
&\quad \exp(-\frac{1}{2}\mathrm{tr}(Y'_{2,-1} Y_{2,-1}((\beta_2-\hat{\beta}_2)\boldsymbol{\alpha}\Sigma^{-1}\boldsymbol{\alpha}'(\beta_2-\hat{\beta}_2)'),
\end{aligned} \tag{6.28}$$

where

$$\begin{aligned}
\hat{\boldsymbol{\alpha}} &= (Y'_{1,-1} M_{Y_{2,-1}} Y_{1,-1})^{-1} Y'_{1,-1} M_{Y_{2,-1}} \Delta Y, \\
\hat{\lambda} &= (Y'_{2,-1} Y_{2,-1})^{-1} Y'_{2,-1} \Delta Y \boldsymbol{\alpha}'_\perp, \\
\hat{\beta}_2 &= -(Y'_{2,-1} Y_{2,-1})^{-1} Y'_{2,-1} (\Delta Y - Y_{1,-1}\boldsymbol{\alpha} - Y_{2,-1}(\mathbf{0} \;\; \lambda)) \Sigma^{-1}\boldsymbol{\alpha}'(\boldsymbol{\alpha}\Sigma^{-1}\boldsymbol{\alpha}')^{-1},
\end{aligned} \tag{6.29}$$

and $M_{Y_{-1}} = \mathbf{I}_{T-1} - Y_{-1}(Y'_{-1}Y_{-1})^{-1}Y'_{-1}$. All marginal/conditional posteriors belong to a known class of probability densities functions: inverted Wishart for the Σ parameter and matrix normal for the remaining parameters, see Zellner (1971, appendix B) for a definition of these densities. Marginal results for the λ and β_2 parameters can be obtained via simulation in a straightforward way.

Since the prior and the likelihood of the error correction cointegration model (6.7) equal the prior and the likelihood of the unrestricted error correction model (6.8) in $\lambda = \mathbf{0}$, the posterior of the cointegration model is proportional to the posterior of the unrestricted error correction model in $\lambda = \mathbf{0}$

$$\begin{aligned}
p_{ecc}(\Sigma, \boldsymbol{\alpha}, \beta_2|Y) &\propto p_{uec}(\Sigma, \boldsymbol{\alpha}, \lambda, \beta_2|Y)|_{\lambda=\mathbf{0}} \\
&\propto p_{uec}(\Sigma, \boldsymbol{\alpha}|Y)\, p_{uec}(\lambda|\Sigma, \boldsymbol{\alpha}, Y)|_{\lambda=\mathbf{0}}\, p_{uec}(\beta_2|\Sigma, \boldsymbol{\alpha}, \lambda, Y)|_{\lambda=\mathbf{0}}.
\end{aligned} \tag{6.30}$$

The posterior of the cointegration model (6.30) can not be decomposed in known densities. The conditional posterior of β_2 restricted in $\lambda = \mathbf{0}$ is still matrix normal, which means that we can simulate β_2 if we know how to simulate from the marginal posterior of Σ and $\boldsymbol{\alpha}$. In Section 6.4.1 we construct a simulation procedure to simulate from the marginal posterior distribution of Σ and $\boldsymbol{\alpha}$. This simulation procedure is based on the ratio of the marginal posterior of Σ and $\boldsymbol{\alpha}$ in the error correction cointegration model (6.30) and the marginal posterior of the same parameters in the unrestricted error correction model (6.28)

$$\begin{aligned}
\frac{p_{ecc}(\Sigma, \boldsymbol{\alpha}|Y)}{p_{uec}(\Sigma, \boldsymbol{\alpha}|Y)} &\propto |\boldsymbol{\alpha}_\perp \Sigma \boldsymbol{\alpha}'_\perp|^{-\frac{1}{2}(n-r)} |Y'_{2,-1} Y_{2,-1}|^{\frac{1}{2}(n-r)} \\
&\quad \exp(-\frac{1}{2}\mathrm{tr}(\boldsymbol{\alpha}_\perp \Sigma \boldsymbol{\alpha}'_\perp)^{-1} \hat{\lambda}' Y'_{2,-1} Y_{2,-1} \hat{\lambda})).
\end{aligned} \tag{6.31}$$

Note that this ratio is proportional to the conditional posterior of λ given $\boldsymbol{\alpha}$ and Σ defined in (6.28) evaluated in the parameter point $\lambda = \mathbf{0}$, $p_{uec}(\lambda|\Sigma, \boldsymbol{\alpha}, Y)|_{\lambda=\mathbf{0}}$. Not surprisingly, this ratio plays an important role in the computation of posterior odds ratios to compare the error correction cointegration model with the unrestricted error correction model, see Section 6.6.

Posterior under Natural Conjugate Prior Specification

In a similar way it is possible to derive marginal/conditional posteriors of the unrestricted error correction model (6.8) in case of natural conjugate priors, see Appendix 6.B. The kernels of these posteriors are

$$\begin{aligned}
p_{uec}(\Sigma|Y) &\propto |S + P'AP + \Delta Y'\Delta Y - \tilde{\boldsymbol{\Pi}}'(A + Y'_{-1}Y_{-1})\tilde{\boldsymbol{\Pi}}|^{\frac{1}{2}(T+h)}|\Sigma|^{-\frac{1}{2}(T+h+n)} \\
&\quad \exp(-\frac{1}{2}\mathrm{tr}(\Sigma^{-1}(S + P'AP + \Delta Y'\Delta Y - \tilde{\boldsymbol{\Pi}}'(A + Y'_{-1}Y_{-1})\tilde{\boldsymbol{\Pi}}))), \\
p_{uec}(\boldsymbol{\alpha}|\Sigma, Y) &\propto |\Sigma|^{-\frac{1}{2}r}|(A + Y'_{-1}Y_{-1})_{11.2}|^{\frac{1}{2}n} \\
&\quad \exp(-\frac{1}{2}\mathrm{tr}(\Sigma^{-1}(\boldsymbol{\alpha} - (\tilde{\boldsymbol{\Pi}}_{11}\ \tilde{\boldsymbol{\Pi}}_{12}))'(A + Y'_{-1}Y_{-1})_{11.2}(\boldsymbol{\alpha} - (\tilde{\boldsymbol{\Pi}}_{11}\ \tilde{\boldsymbol{\Pi}}_{12}))), \\
p_{uec}(\lambda|\Sigma, \boldsymbol{\alpha}, Y) &\propto |\boldsymbol{\alpha}_\perp\Sigma\boldsymbol{\alpha}'_\perp|^{-\frac{1}{2}(n-r)}|(A + Y'_{-1}Y_{-1})_{22}|^{\frac{1}{2}(n-r)} \\
&\quad \exp(-\frac{1}{2}\mathrm{tr}((\boldsymbol{\alpha}_\perp\Sigma\boldsymbol{\alpha}'_\perp)^{-1}(\lambda - \tilde{\lambda})'(A + Y'_{-1}Y_{-1})_{22}(\lambda - \tilde{\lambda}))), \\
p_{uec}(\beta_2|\Sigma, \boldsymbol{\alpha}, \lambda, Y) &\propto |\boldsymbol{\alpha}\Sigma^{-1}\boldsymbol{\alpha}'|^{\frac{1}{2}(n-r)}|(A + Y'_{-1}Y_{-1})_{22}|^{\frac{1}{2}r} \\
&\quad \exp(-\frac{1}{2}\mathrm{tr}((A + Y'_{-1}Y_{-1})_{22}(\beta_2 - \tilde{\beta}_2)\boldsymbol{\alpha}\Sigma^{-1}\boldsymbol{\alpha}'(\beta_2 - \tilde{\beta}_2)')),
\end{aligned} \tag{6.32}$$

with

$$\begin{aligned}
\tilde{\boldsymbol{\Pi}} &= \begin{pmatrix} \tilde{\boldsymbol{\Pi}}_{11} & \tilde{\boldsymbol{\Pi}}_{12} \\ \tilde{\boldsymbol{\Pi}}_{21} & \tilde{\boldsymbol{\Pi}}_{22} \end{pmatrix} = (A + Y'_{-1}Y_{-1})^{-1}(AP + Y'_{-1}Y_{-1}\hat{\boldsymbol{\Pi}}), \\
\tilde{\lambda} &= (R_{21}\ R_{22})\boldsymbol{\alpha}'_\perp, \\
\tilde{\beta}_2 &= -(R_{21}\ (R_{22} - \lambda))\Sigma^{-1}\boldsymbol{\alpha}'(\boldsymbol{\alpha}\Sigma^{-1}\boldsymbol{\alpha}')^{-1},
\end{aligned} \tag{6.33}$$

where $(R_{21}\ R_{22}) = ((\tilde{\boldsymbol{\Pi}}_{21}\ \tilde{\boldsymbol{\Pi}}_{22}) - (\boldsymbol{\alpha} - (A + Y'_{-1}Y_{-1})^{-1}_{22}(A + Y'_{-1}Y_{-1})_{21}(\tilde{\boldsymbol{\Pi}}_{11}\ \tilde{\boldsymbol{\Pi}}_{12})))$ and $(A + Y'_{-1}Y_{-1}) = \begin{pmatrix} (A + Y'_{-1}Y_{-1})_{11} & (A + Y'_{-1}Y_{-1})_{12} \\ (A + Y'_{-1}Y_{-1})_{21} & (A + Y'_{-1}Y_{-1})_{22} \end{pmatrix}$. Since all the marginal/conditional densities are of a known class of distributions, the marginal posteriors of λ and β_2 can be obtained via straightforward simulation.

Just as in the diffuse case, the posterior of the cointegration model (6.7) is proportional to the posterior of the unrestricted error correction model (6.8) evaluated in $\lambda = \mathbf{0}$, see (6.30). Again the ratio of the marginal posterior of Σ and $\boldsymbol{\alpha}$ resulting from the error correction cointegration model (6.7) and the marginal posterior of $\boldsymbol{\alpha}$ and Σ in the unrestricted error correction model plays an important role in the simulation from the posterior distribution and the computation of posterior odds ratios. In the next subsection we show a simulation scheme to sample from the posterior of the cointegration model.

6.4.1 Simulating Posterior Distributions

To evaluate the posterior distributions of the error correction cointegration models (6.7) with diffuse or natural conjugate priors, we use Markov chain Monte Carlo techniques. Since not all of the full conditional posterior distributions are of a known type, standard Gibbs sampling is not possible. Therefore, we apply the Metropolis-Hastings sampler of Metropolis *et al.* (1953) and Hastings (1970).

To describe the Metropolis-Hastings [M-H] sampling algorithm, let ψ be a random variable with density function $f(\psi)$. Let $g(\zeta|\psi)$ be a *candidate-generating* density function in ζ. The simulation algorithm to sample from the density $f(\psi)$ works as follows:

Step 1: Specify starting values ψ^0 and set $i = 0$.

Step 2: Simulate ζ from $g(\zeta|\psi^i)$.

$$\text{Define } a(\zeta, \psi^i) = \begin{cases} \min\left(\frac{f(\zeta)g(\psi^i|\zeta)}{f(\psi^i)g(\zeta|\psi^i)}, 1\right) & f(\psi^i)g(\zeta|\psi^i) > 0 \\ 1 & f(\psi^i)g(\zeta|\psi^i) = 0. \end{cases}$$

Choose $\psi^{i+1} = \zeta$ with probability $a(\zeta, \psi^i)$
and $\psi^{i+1} = \psi^i$ with probability $(1 - a(\zeta, \psi^i))$.

Step 3: Set $i = i + 1$ and go to step 2.

The described iterative scheme generates a Markov chain. After the chain has converged, say at H iterations, the simulated values $\{\psi^i, i \geq H\}$ can be used as a sample from the distribution of ψ to compute means, variances, *etc.* Different choices for the *candidate-generating* function result in different specific forms of the algorithm. For example, if $g(\psi^i|\zeta) = g(\zeta|\psi^i)$ the acceptance probability simplifies to $a(\zeta, \psi^i) = \min(f(\zeta)/f(\psi^i), 1)$. This describes the original Metropolis algorithm. If $g(\zeta, \psi^i) = g(\zeta)$, we get $a(\psi^i, \zeta) = \min(w(\zeta)/w(\psi^i), 1)$, where $w(\zeta) = f(\zeta)/g(\zeta)$, which can be interpreted as importance weights, see also Section 6.6.1. For details we refer to Smith and Roberts (1993) and Tierney (1994).

The simulation framework to sample from the posterior of the cointegration model (6.7) is the same for diffuse and natural conjugate priors. The sampling scheme is based on the fact that the posterior of the error correction cointegration model (6.7) is proportional to the posterior of the unrestricted error correction model (6.8) evaluated in $\lambda = \mathbf{0}$

$$\begin{aligned} p_{ecc}(\Sigma, \boldsymbol{\alpha}, \beta_2|Y) &\propto p_{uec}(\Sigma|Y)\, p_{uec}(\boldsymbol{\alpha}|\Sigma, Y)\, p_{uec}(\lambda|\Sigma, \boldsymbol{\alpha}, Y)|_{\lambda=\mathbf{0}}\, p_{uec}(\beta_2|\Sigma, \boldsymbol{\alpha}, \lambda, Y)|_{\lambda=\mathbf{0}} \\ &\propto p_{uec}(\Sigma|Y)\, p_{uec}(\boldsymbol{\alpha}|\Sigma, Y)\, w(\Sigma, \boldsymbol{\alpha}|Y)\, p_{uec}(\beta_2|\Sigma, \boldsymbol{\alpha}, \lambda, Y)|_{\lambda=\mathbf{0}}, \end{aligned} \tag{6.34}$$

where $w(\Sigma, \boldsymbol{\alpha}|Y) = p_{uec}(\lambda|\Sigma, \boldsymbol{\alpha}, Y)|_{\lambda=\mathbf{0}}$. If we ignore the function $w(\Sigma, \boldsymbol{\alpha}|Y)$ in (6.34), simulation from the posterior distribution is easy, since the remainder consists of a product of standard densities. Since $w(\Sigma, \boldsymbol{\alpha}|Y)$ is a bounded function, we can use an acceptance-rejection simulation algorithm. This may, however, lead to large rejection frequencies if

the cointegration rank is not correctly specified. Chib and Greenberg (1995) show that in this case a M-H algorithm can speed up the simulation process. Since β_2 does not enter the weight function w, the M-H step only enters the simulation scheme for the sampling of the Σ and the $\boldsymbol{\alpha}$ parameters. If we take as *candidate-generating* density $p_{uec}(\Sigma, \boldsymbol{\alpha}|Y)$ or $p_{uec}(\Sigma|Y)\, p_{uec}(\boldsymbol{\alpha}|\Sigma, Y)$, the acceptance-rejection probability simplifies to a ratio of weight functions w, as discussed above. Given the draws of Σ and $\boldsymbol{\alpha}$, we sample β_2 conditional on $(\Sigma, \boldsymbol{\alpha})$ from a matrix normal distribution.

The three steps to sample from the posterior distribution of the cointegration model can be summarized as follows:

Step 1: Draw Σ^{i+1} from $p_{uec}(\Sigma|Y)$.

Draw $\boldsymbol{\alpha}^{i+1}$ from $p_{uec}(\boldsymbol{\alpha}|\Sigma^{i+1}, Y)$.

Step 2: Accept $(\Sigma^{i+1}, \boldsymbol{\alpha}^{i+1})$ with probability $\min\left(\frac{w(\Sigma^{i+1}, \boldsymbol{\alpha}^{i+1}|Y)}{w(\Sigma^{i}, \boldsymbol{\alpha}^{i}|Y)}, 1\right)$,

otherwise $(\Sigma^{i+1}, \boldsymbol{\alpha}^{i+1}) = (\Sigma^{i}, \boldsymbol{\alpha}^{i})$.

Step 3: Draw β_2^{i+1} from $p_{uec}(\beta_2|\Sigma^{i+1}, \boldsymbol{\alpha}^{i+1}, \lambda, Y)|_{\lambda=\mathbf{0}}$.

This simulation scheme has advantages if one wants to analyse the model under every cointegration rank r. Since the sampling distribution of Σ does not depend on the rank r, one only needs one draw Σ for every cointegration rank. Furthermore, using the properties of the matrix normal distribution, the sampling of $\boldsymbol{\alpha}$ parameters can be accelerated. Instead of sampling an $\boldsymbol{\alpha}$ matrix for every rank r, one can sample the $\boldsymbol{\alpha}$ matrices at once using one draw $\boldsymbol{\Pi}^{i+1}$ from $p(\boldsymbol{\Pi}|\Sigma, Y)$, which is a matrix normal density. The candidate draws of $\boldsymbol{\alpha}^{i+1}$ under the cointegration rank r, $r = 1, \ldots, n$, are obtained by taking the first r rows of the draw $\boldsymbol{\Pi}^{i+1}$.

Finally, the presented sampling scheme is not unique. It is possible to use a different decomposition than the one proposed in (6.34). Furthermore, the decomposition (6.34) is also suitable for importance sampling, see Kloek and van Dijk (1978) and Geweke (1989). The weight functions w evaluated in the draws represent in that case the importance weights. However, the M-H sampling method has the advantage that it is easier to apply in more complicated cointegration models, like for instance cointegration models with a break in the constant and/or in the cointegration relation or threshold cointegration models. These more complicated models are often analysed in a Gibbs framework. The sampling of the block $(\Sigma, \boldsymbol{\alpha}, \beta_2)$ given the remaining parameters in the model can then be done using the simulation steps presented in this subsection, see Section 7.4 for an example.

6.5 Bayesian Lagrange Multiplier Statistics

In the previous section, we assumed for the derivation of the posterior simulators, that the number of cointegrating vectors r is known *a priori*. This is in practice seldom the

case and procedures, which analyse whether the chosen number of cointegrating vectors is plausible, are needed. In classical statistical analysis diagnostic test statistics like Lagrange Multiplier or score statistics are intended for this purpose. In this section, we will construct the Bayesian analog of these classical LM statistics to test whether the assumed number of cointegrating vectors is plausible. Since the computation and interpretation of the Bayesian LM statistic to test for cointegration is not straightforward, we start in Section 6.5.1 with a Bayesian LM test for a zero regression coefficient in a simple linear regression model. The Bayesian LM cointegration statistics are extensions of the LM statistic in a linear regression model. To save on notation we only consider the Bayesian LM statistics under a diffuse prior specification. The results can easily be extended for the natural conjugate prior case.

6.5.1 Bayesian LM Statistic in the Linear Regression Model

Consider a linear regression model with two explanatory variables,

$$\begin{aligned} y &= z_1\gamma_1 + z_2\gamma_2 + \eta \\ &= Z\gamma + \eta, \end{aligned} \tag{6.35}$$

where y, η, z_1, z_2 are $(T \times 1)$ matrices, $Z = (z_1\ z_2)$, $\gamma = (\gamma_1\ \gamma_2)'$ and $\eta \sim N(0, \sigma^2\mathbf{I}_T)$. If we are interested whether the parameter γ_1 is zero, we can test this hypothesis using a highest posterior density region, see Box and Tiao (1973). An alternative method to test the hypothesis $\gamma_1 = 0$, is to use a Bayesian analog of an LM statistic. Since in the linear regression model the marginal posterior distributions are of a known type, it is possible to derive the LM statistic for $\gamma_1 = 0$ analytically. In the cointegration model, however, the marginal distributions are of an unknown form and we use a M-H sampler to simulate from the posterior distribution. Therefore, we also show in the linear regression model how to calculate the LM statistic for the restriction $\gamma_1 = 0$ using M-H output. The latter approach mimics the computation of the more complicated LM statistic for cointegration, which will be discussed in Section 6.5.2.

Analytical Approach

Assuming diffuse priors for the different parameters,

$$p(\gamma_1, \gamma_2, \sigma^2) \propto \sigma^{-3}, \tag{6.36}$$

some conditional and marginal posteriors of the parameters of the linear regression model (6.35) read,

$$p(\gamma_2|\gamma_1, \sigma^2, y, Z) \propto \sigma^{-1} \exp(-\frac{1}{2\sigma^2}(\gamma_2 - \hat{\gamma}_2)' z_2 z_2 (\gamma_2 - \hat{\gamma}_2)), \tag{6.37}$$

$$p(\sigma^2|\gamma_1, y, Z) \propto \sigma^{-(T+3)} \exp(-\frac{1}{2\sigma^2}(y - z_1\gamma_1)' M_{z_2}(y - z_1\gamma_1)), \tag{6.38}$$

$$p(\gamma_1|\sigma^2, y, Z) \propto \sigma^{-1} \exp(-\frac{1}{2\sigma^2}(\gamma_1 - \hat{\gamma}_1)' z_1' M_{z_2} z_1 (\gamma_1 - \hat{\gamma}_1)), \tag{6.39}$$

where $\hat{\gamma}_1 = (z_1' M_{z_2} z_1)^{-1} z_1' M_{z_2} y$, $\hat{\gamma}_2 = (z_2' z_2)^{-1} z_2' (y - z_1\gamma_1)$ and $M_{z_2} = \mathbf{I}_T - z_2(z_2' z_2)^{-1} z_2'$.

To derive the distribution of the Bayesian version of the LM statistic for the hypothesis $\gamma_1 = 0$ under the alternative hypothesis we use the conditional posterior of γ_1 given σ^2 (6.37):

$$\begin{aligned} (\gamma_1 - \hat{\gamma}_1) &\sim N(0, \sigma^2 (z_1' M_{z_2} z_1)^{-1}) \quad \Leftrightarrow \\ \sigma^{-1}(z_1' M_{z_2} z_1)^{\frac{1}{2}} (\gamma_1 - \hat{\gamma}_1) &\sim N(0,1) \quad \Leftrightarrow \\ \sigma^{-1}(z_1' M_{z_2} z_1)^{-\frac{1}{2}} z_1' M_{z_2} (y - z_1\gamma_1) &\sim N(0,1), \end{aligned} \tag{6.40}$$

and hence given σ

$$\sigma^{-2}(y - z_1\gamma_1)' M_{z_2} z_1 (z_1' M_{z_2} z_1)^{-1} z_1' M_{z_2} (y - z_1\gamma_1) \sim \chi^2(1). \tag{6.41}$$

This results holds regardless of the value of σ^2 so that this property is not lost when we go to the marginal result for γ_1 by integrating out σ^2,

$$\mathrm{E}_{\sigma^2}[\sigma^{-2}(y - z_1\gamma_1)' M_{z_2} z_1 (z_1' M_{z_2} z_1)^{-1} z_1' M_{z_2} (y - z_1\gamma_1)] \sim \chi^2(1). \tag{6.42}$$

If we substitute $\gamma_1 = 0$ in this expression and take the expectation with respect to the conditional posterior of σ^2 in (6.38) with $\gamma_1 = 0$, we obtain the value of the LM statistic under the hypothesis $\gamma_1 = 0$

$$\mathrm{LM}(\gamma_1 = 0) = \mathrm{E}_{\sigma^2}[\sigma^{-2}(y - z_1\gamma_1)' M_{z_2} z_1 (z_1' M_{z_2} z_1)^{-1} z_1' M_{z_2} (y - z_1\gamma_1)|_{\gamma_1 = 0}]. \tag{6.43}$$

We reject the hypothesis $\gamma_1 = 0$ when the resulting LM statistic (6.43) lies outside the one-sided HPD region of a $\chi^2(1)$ distribution. This can be seen as an alternative of testing whether $\gamma_1 = 0$ using a HPD region for the marginal distribution of γ_1, which is t-distributed.

Likewise, it can be shown that the expression

$$\sigma^{-2}(y - Z\gamma)' Z (Z'Z)^{-1} Z' (y - Z\gamma). \tag{6.44}$$

is $\chi^2(2)$ distributed, when γ_1 and γ_2 are non-zero. So, the LM test for $\gamma_1 = \gamma_2 = 0$ is obtained by evaluating (6.44) in $\gamma_1 = \gamma_2 = 0$ and taking the expectation over σ^2 given $\gamma_1 = \gamma_2 = 0$.

It is also possible to compute the LM statistic (6.43) by adjusting the expression (6.44). This expression can be decomposed as

$$\begin{aligned} \sigma^{-2}(y - Z\gamma)' Z (Z'Z)^{-1} Z' (y - Z\gamma) = \\ \sigma^{-2}(y - Z\gamma)' M_{z_2} z_1 (z_1' M_{z_2} z_1)^{-1} z_1' M_{z_2} (y - Z\gamma) + \\ \sigma^{-2}(y - Z\gamma)' z_2 (z_2' z_2)^{-1} z_2' (y - Z\gamma), \end{aligned} \tag{6.45}$$

see Davidson and MacKinnon (1993, section 3.5). Under $\gamma_1 = 0$ the second part has a $\chi^2(1)$ distribution and the first part equals (6.41), which implies that (6.43) is equal to

$$\begin{aligned} \mathrm{LM}(\gamma_1 = 0) &= \mathrm{E}_{\sigma^2}[\sigma^{-2}(y - z_1\gamma_1)' M_{z_2} z_1 (z_1' M_{z_2} z_1)^{-1} z_1' M_{z_2} (y - z_1\gamma_1)|_{\gamma_1=0}] \\ &= \mathrm{E}_{\gamma_2}\mathrm{E}_{\sigma^2}[\sigma^{-2}(y - Z\gamma)' Z(Z'Z)^{-1} Z'(y - Z\gamma)|_{\gamma_1=0}] - \mathrm{E}[\chi^2(1)] \\ &= \mathrm{E}_{\gamma_2}\mathrm{E}_{\sigma^2}[\sigma^{-2}(y - Z\gamma)' Z(Z'Z)^{-1} Z'(y - Z\gamma)|_{\gamma_1=0}] - 1. \end{aligned} \tag{6.46}$$

This results extends also to other kind of hypotheses on γ_1 and γ_2, and can be used in any kind of linear model. For certain non-linear hypotheses on the parameters of a linear model, like the reduced rank restriction for cointegration models, Bayesian LM statistics can only be constructed by using a generalization of the result in (6.46).

Metropolis-Hastings Sampling Approach

Consider the case that we do not construct the Bayesian LM statistic using the marginal and conditional posteriors assuming that $\gamma_1 = 0$, but use the marginal posterior of σ^2, and the posterior of γ_2 given σ^2 from the unrestricted model in a M-H sampling approach. So, the marginal/conditional densities from which σ^2 and γ_2 are sampled read,

$$\begin{aligned} p(\sigma^2|y, X) &\propto \sigma^{-(T+2)} \exp(-\frac{1}{2\sigma^2} y' M_{(z_1\, z_2)} y), \\ p(\gamma_2|\sigma^2, y, X) &\propto \sigma^{-1} \exp(-\frac{1}{2\sigma^2}(\gamma_2 - \tilde{\gamma}_2)' z_2' M_{z_1} z_2 (\gamma_2 - \tilde{\gamma}_2)), \end{aligned} \tag{6.47}$$

where $\tilde{\gamma}_2 = (z_2' M_{z_1} z_2)^{-1} z_2' M_{z_1} y$. To correct for not sampling from the true posterior, we need again the ratio of the true posterior and the sampling density, which is needed to construct the acceptance probability a in the Metropolis-Hastings sampler, see Section 6.4.1. This ratio is given by

$$\sigma^{-1} \exp(-\frac{1}{2\sigma^2} \tilde{\gamma}_1' z_1' z_1 \tilde{\gamma}_1), \tag{6.48}$$

where $\tilde{\gamma}_1 = (z_1' z_1)^{-1} z_1' (y - z_2\gamma_2)$, *i.e.* the mean of the conditional posterior of γ_1 given γ_2.

The LM statistic to test $\gamma_1 = 0$ can be computed by evaluating (6.43) using the M-H output. We can also use the M-H output to calculate the LM statistic using (6.46). Although (6.44) can be decomposed in the expression in the exponent of the kernel of the sampling density of γ_2 (6.47) and in the expression of the exponent of (6.48),

$$\begin{aligned} &\sigma^{-2}(y - Z\gamma)' Z(Z'Z)^{-1} Z'(y - Z\gamma)_{\gamma_1=0} \\ &\quad = \sigma^{-2}(y - z_2\gamma_2)'(M_{z_1} z_2 (z_2' M_{z_1} z_2)^{-1} z_2' M_{z_1} + z_1 (z_1' z_1)^{-1} z_1')(y - z_2\gamma_2) \\ &\quad = \sigma^{-2}((\gamma_2 - \tilde{\gamma}_2)' z_2' M_{z_1} z_2 (\gamma_2 - \tilde{\gamma}_2) + \tilde{\gamma}_1' z_1' z_1 \tilde{\gamma}_1) \end{aligned} \tag{6.49}$$

the Bayesian LM statistic does **not** correspond to the expectation of $\sigma^{-2}\tilde{\gamma}_1' z_1' z_1 \tilde{\gamma}_1$ based on the M-H output. Since σ^2 and γ_2 are sampled given $\gamma_1 = 0$ using the M-H sampler, $\mathrm{E}_{\sigma^2}[\sigma^{-2}(\gamma_2 - \tilde{\gamma}_2)' z_2' M_{z_1} z_2 (\gamma_2 - \tilde{\gamma}_2)]$ does not have a $\chi^2(1)$ distribution.

For the cointegration hypotheses, discussed in the next subsection, the specific dependence of the parameters on one another does only allow for the kind of decompositions as in (6.45). Closed form expressions of the Bayesian LM cointegration statistic, like (6.43), do therefore not exist. However, these Bayesian LM statistics can still be calculated using a multivariate generalization of the result in (6.46) together with a M-H sampling procedure, like discussed above.

6.5.2 Bayesian LM Statistic in the Cointegration Model

The parameter restriction for rank reduction in the unrestricted error correction model (6.8) is $\lambda = \mathbf{0}$. Since the marginal posterior of the parameter reflecting cointegration, λ, in the unrestricted error correction model, cannot be constructed analytically, the Bayesian LM statistic to test for cointegration does not have a closed form analytical expression as in the linear regression model. Therefore, we need the M-H sampling approach in combination with a multivariate extension of the result in (6.46) to calculate the Bayesian LM statistic. The Bayesian LM statistic is based on the expression

$$\begin{aligned}\mathrm{E}_{\Sigma}[\mathrm{tr}(\Sigma^{-1}(\mathbf{\Pi}-\hat{\mathbf{\Pi}})'Y_{-1}'Y_{-1}(\mathbf{\Pi}-\hat{\mathbf{\Pi}})] = \\ \mathrm{E}_{\Sigma}[\mathrm{tr}(\Sigma^{-1}(Y-Y_{-1}\mathbf{\Pi})'Y_{-1}(Y_{-1}'Y_{-1})^{-1}Y_{-1}'(Y-Y_{-1}\mathbf{\Pi}))], \quad (6.50)\end{aligned}$$

where $\hat{\mathbf{\Pi}} = (Y_{-1}'Y_{-1})^{-1}Y_{-1}'\Delta Y$. It is straightforward to show that this expression has a χ^2 distribution with n^2 degrees of freedom. Hence, the Bayesian LM statistic for $\mathbf{\Pi} = \mathbf{\Pi}_0$ is given by $\mathrm{E}_{\Sigma}[\mathrm{tr}(\Sigma^{-1}(Y-Y_{-1}\mathbf{\Pi})'Y_{-1}(Y_{-1}'Y_{-1})^{-1}Y_{-1}'(Y-Y_{-1}\mathbf{\Pi}))|_{\mathbf{\Pi}=\mathbf{\Pi}_0}]$, which has to be compared with a $\chi^2(n^2)$ distribution.

To compute the Bayesian LM statistic for $\lambda = \mathbf{0}$ we consider (6.50) in $\lambda = \mathbf{0}$ and take the expectation with respect to $\boldsymbol{\alpha}$ and β_2 given $\lambda = \mathbf{0}$

$$\begin{aligned}&\mathrm{E}_{\Sigma,\boldsymbol{\alpha},\beta_2}[\mathrm{tr}(\Sigma^{-1}(Y-Y_{-1}\beta\boldsymbol{\alpha})'Y_{-1}(Y_{-1}'Y_{-1})^{-1}Y_{-1}'(Y-Y_{-1}\beta\boldsymbol{\alpha}))] \\ &\quad = \mathrm{E}_{\boldsymbol{\alpha},\beta_2,\Sigma}[\mathrm{tr}(\Sigma^{-1}(\boldsymbol{\alpha}-\hat{\boldsymbol{\alpha}})'Y_{1,-1}M_{Y_{2,-1}}Y_{1,-1}(\boldsymbol{\alpha}-\hat{\boldsymbol{\alpha}})) + \\ &\qquad \mathrm{tr}((\boldsymbol{\alpha}_{\perp}\Sigma\boldsymbol{\alpha}_{\perp}')^{-1}\hat{\lambda}'Y_{2,-1}'Y_{2,-1}\hat{\lambda}) + \mathrm{tr}(Y_{2,-1}'Y_{2,-1}(\beta_2-\hat{\beta}_2)\boldsymbol{\alpha}\Sigma^{-1}\boldsymbol{\alpha}'(\beta_2-\hat{\beta}_2)')], \quad (6.51)\end{aligned}$$

where $\hat{\boldsymbol{\alpha}}$, $\hat{\lambda}$ and $\hat{\beta}_2$ are given in (6.29) with $\lambda = \mathbf{0}$. Under the restriction of cointegration, $\lambda = \mathbf{0}$, draws from Σ, $\boldsymbol{\alpha}$ and β_2 can be obtained using a M-H sampler, as outlined in Section 6.4.1. Since the same reasoning holds for equation (6.51) as for equation (6.49), the Bayesian LM statistic for testing for cointegration, $\lambda = \mathbf{0}$, does **not** correspond to $\mathrm{E}_{\boldsymbol{\alpha},\beta_2,\Sigma}[\mathrm{tr}((\boldsymbol{\alpha}_{\perp}\Sigma\boldsymbol{\alpha}_{\perp}')^{-1}\hat{\lambda}'Y_{2,-1}'Y_{2,-1}\hat{\lambda})]$. Therefore, we have to apply a generalization of (6.46) to calculate the Bayesian version of the LM statistic to test for rank reduction versus full rank using the M-H output

$$\begin{aligned}&\mathrm{LM}(r|n) = \mathrm{LM}(\lambda = \mathbf{0}) \\ &= \mathrm{E}_{\boldsymbol{\alpha},\beta_2,\Sigma}[\mathrm{tr}(\Sigma^{-1}(Y-Y_{-1}\beta\boldsymbol{\alpha})'Y_{-1}(Y_{-1}'Y_{-1})^{-1}Y_{-1}'(Y-Y_{-1}\beta\boldsymbol{\alpha}))] - \mathrm{E}[\chi^2(r(2n-r)] \\ &= \mathrm{E}_{\boldsymbol{\alpha},\beta_2,\Sigma}[\mathrm{tr}(\Sigma^{-1}(Y-Y_{-1}\beta\boldsymbol{\alpha})'Y_{-1}(Y_{-1}'Y_{-1})^{-1}Y_{-1}'(Y-Y_{-1}\beta\boldsymbol{\alpha}))] - r(2n-r). \quad (6.52)\end{aligned}$$

The correction factor $r(2n-r)$ equals the number of parameters in $\boldsymbol{\alpha}$ and β_2. The resulting Bayesian LM cointegration statistic has to be compared with a χ^2 distribution with $(n-r)^2$ degrees of freedom. If it is not plausible that the calculated statistic has been generated by this distribution, the hypothesis that $\lambda = \mathbf{0}$ is not considered plausible, which implies that a cointegration model with r cointegration relations is not likely. Typical extensions of the cointegration hypothesis $\lambda = \mathbf{0}$ towards hypotheses including parameters of deterministic components, for example to test whether deterministic components lie in the cointegration space, can be dealt with in a straightforward way.

6.6 Posterior Odds Analysis

The in the previous sections developed procedures for calculating the posteriors of the parameters of the cointegration model for different number of cointegrating vectors r, allow us to compare models with different cointegration ranks using posterior odds analysis. Since the number of cointegration vectors r can only take $n+1$ different discrete values, we can consider prior and posterior probabilities of the cointegration rank r and the implied number of unit roots $(n-r)$, $r = 0, \ldots, n$.

First we assign prior probabilities to every cointegration rank r,

$$\Pr[\text{rank} = r] \qquad r = 0, \ldots, n. \tag{6.53}$$

These prior probabilities imply prior odds ratios [PROR] to compare *a priori* the cointegration models with different number of cointegration relations. Since every cointegration model (6.7) is nested in the full rank model (6.8) it is convenient to consider

$$\text{PROR}(r|n) = \frac{\Pr[\text{rank} = r]}{\Pr[\text{rank} = n]}, \qquad r = 0, \ldots, n. \tag{6.54}$$

The Bayes factor [BF] which compares the cointegration model (6.7) with the unrestricted error correction model (6.8) is given by

$$\begin{aligned} \text{BF}(r|n) &= \frac{\Pr[Y|\text{rank} = r]}{\Pr[Y|\text{rank} = n]}, \qquad r = 0, \ldots, n \\ &= \frac{\iiint p_{ecc}(\Sigma, \boldsymbol{\alpha}, \beta_2)\, \mathcal{L}_{ecc}(Y|\Sigma, \boldsymbol{\alpha}, \beta_2)\, d\Sigma\, d\boldsymbol{\alpha}\, d\beta_2}{\iiiint p_{uec}(\Sigma, \boldsymbol{\alpha}, \lambda, \beta_2)\, \mathcal{L}_{uec}(Y|\Sigma, \boldsymbol{\alpha}, \lambda, \beta_2)\, d\Sigma\, d\boldsymbol{\alpha}\, d\lambda\, d\beta_2}, \end{aligned} \tag{6.55}$$

see *e.g.* Poirier (1995) for a formal definition of a Bayes factor. Now we can define the posterior odds ratios [POR] to compare *a posteriori* a cointegration model with r cointegrating vectors with a model with n cointegrating vectors

$$\text{POR}(r|n) = \text{PROR}(r|n) \times \text{BF}(r|n), \qquad r = 0, \ldots, n. \tag{6.56}$$

These posterior odds ratios imply posterior probabilities for every cointegration rank. The posterior probability for a cointegration model with rank r equals

$$\Pr[\text{rank} = r|Y] = \frac{\text{POR}(r|n)}{\sum_{i=0}^{n} \text{POR}(i|n)}, \qquad r = 0, \ldots, n. \tag{6.57}$$

The posterior probabilities can be used to choose the cointegration rank, or as weights in further analyses, like forecasting exercises.

Bayes factors are only well defined in case of proper priors. Especially for the λ parameter a proper prior is required, see Section 3.3.2 for a discussion. In the next subsection we show how the Bayes factors can be computed in case of natural conjugate priors. Additionally, we provide a Bayes factor under a diffuse prior specification, which can be seen as a limiting case of a natural conjugate prior specification on λ.

6.6.1 Computation of Bayes Factors

Since the prior and the likelihood of the error correction cointegration model (6.7) equal the prior and the likelihood of the unrestricted error correction model (6.8) evaluated in $\lambda = \mathbf{0}$, times the constant $1/c_r$, defined in (6.16), the Bayes factor (6.55) is equal to

$$\begin{aligned}
\mathrm{BF}(r|n) &= \frac{\iiint \frac{1}{c_r} p_{uec}(\Sigma, \boldsymbol{\alpha}, \lambda, \beta_2)|_{\lambda=\mathbf{0}}\, \mathcal{L}_{uec}(Y|\Sigma, \boldsymbol{\alpha}, \lambda, \beta_2)|_{\lambda=\mathbf{0}}\, d\Sigma\, d\boldsymbol{\alpha}\, d\beta_2}{\iiiint p_{uec}(\Sigma, \boldsymbol{\alpha}, \lambda, \beta_2)\, \mathcal{L}_{uec}(Y|\Sigma, \boldsymbol{\alpha}, \lambda, \beta_2)\, d\Sigma\, d\boldsymbol{\alpha}\, d\lambda\, d\beta_2} \\
&= \frac{\iiint \frac{1}{c_r} p_{uec}(\Sigma, \boldsymbol{\alpha}|Y)\, p_{uec}(\lambda|\Sigma, \boldsymbol{\alpha}, Y)|_{\lambda=\mathbf{0}}\, p_{uec}(\beta_2|\Sigma, \boldsymbol{\alpha}, \lambda, Y)|_{\lambda=\mathbf{0}}\, d\Sigma\, d\boldsymbol{\alpha}\, d\beta_2}{\iiiint p_{uec}(\Sigma, \boldsymbol{\alpha}|Y)\, p_{uec}(\lambda|\Sigma, \boldsymbol{\alpha}, Y)\, p_{uec}(\beta_2|\Sigma, \boldsymbol{\alpha}, \lambda, Y)\, d\Sigma\, d\boldsymbol{\alpha}\, d\lambda\, d\beta_2} \\
&= \frac{\frac{1}{c_r} \iint p_{uec}(\Sigma, \boldsymbol{\alpha}|Y)\, p_{uec}(\lambda|\Sigma, \boldsymbol{\alpha}, Y)|_{\lambda=\mathbf{0}}\, d\Sigma\, d\boldsymbol{\alpha}}{\iint p_{uec}(\Sigma, \boldsymbol{\alpha}|Y)\; d\Sigma\, d\boldsymbol{\alpha}},
\end{aligned} \tag{6.58}$$

where the last step is obtained by integrating over β_2 and λ. We can calculate the ratio of integrals of conditional posteriors efficiently by simulating Σ and $\boldsymbol{\alpha}$ from $p_{uec}(\Sigma, \boldsymbol{\alpha}|Y)$ which is the product of an inverted Wishart for Σ and a matrix normal for $\boldsymbol{\alpha}$ given Σ. For the sampled parameters we calculate the ratio of the integrands

$$w(\Sigma^i, \boldsymbol{\alpha}^i|Y) = \frac{p_{uec}(\Sigma^i, \boldsymbol{\alpha}^i|Y)\, p_{uec}(\lambda|\Sigma^i, \boldsymbol{\alpha}^i, Y)|_{\lambda=\mathbf{0}}}{p_{uec}(\Sigma^i, \boldsymbol{\alpha}^i|Y)} = p_{uec}(\lambda|\boldsymbol{\alpha}^i, \Sigma^i, Y)|_{\lambda=\mathbf{0}}, \tag{6.59}$$

where i corresponds to the i-th draw of $(\Sigma, \boldsymbol{\alpha})$. The average of the simulated weights $w(\Sigma^i, \boldsymbol{\alpha}^i|Y)$ then converges to the ratio of the integrals (6.58)

$$\sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N} w(\Sigma^i, \boldsymbol{\alpha}^i|Y) - \frac{\iint p_{uec}(\Sigma^i, \boldsymbol{\alpha}^i|Y)\, p_{uec}(\lambda|\Sigma^i, \boldsymbol{\alpha}^i, Y)|_{\lambda=\mathbf{0}}\, d\Sigma\, d\boldsymbol{\alpha}}{\iint p_{uec}(\Sigma^i, \boldsymbol{\alpha}^i|Y)\, d\Sigma\, d\boldsymbol{\alpha}}\right) \Rightarrow N(0, v), \tag{6.60}$$

where N is the number of draws, $v = \mathrm{var}(w(\Sigma, \boldsymbol{\alpha}|Y))$, $\Rightarrow$ stands for weak convergence, and $\left(\frac{1}{N}\sum_{i=1}^{N} w(\Sigma^i, \boldsymbol{\alpha}^i|Y)^2 - (\frac{1}{N}\sum_{i=1}^{N} w(\Sigma^i, \boldsymbol{\alpha}^i|Y))^2\right) \Rightarrow v$, see Geweke (1989).

To compute the Bayes factor we need the value of c_r. As we have already seen, this correction factor (6.16) is only properly defined in case of proper priors. Under a natural conjugate prior specification the factor c_r can easily be computed using the above

mentioned simulation method. Since we can write (6.16) as

$$\begin{aligned} c_r &= \frac{\iiint p_{uec}(\Sigma, \boldsymbol{\alpha}, \lambda, \beta_2)|_{\lambda=\mathbf{0}}\, d\Sigma\, d\boldsymbol{\alpha}\, d\beta_2}{\iiiint p_{uec}(\Sigma, \boldsymbol{\alpha}, \lambda, \beta_2)\, d\Sigma\, d\boldsymbol{\alpha}\, d\lambda\, d\beta_2} \\ &= \frac{\iiint p_{uec}(\Sigma, \boldsymbol{\alpha})\, p_{uec}(\lambda|\Sigma, \boldsymbol{\alpha})|_{\lambda=\mathbf{0}}\, p_{uec}(\beta_2|\Sigma, \boldsymbol{\alpha}, \lambda)|_{\lambda=\mathbf{0}}\, d\Sigma\, d\boldsymbol{\alpha}\, d\beta_2}{\iiiint p_{uec}(\Sigma, \boldsymbol{\alpha})\, p_{uec}(\lambda|\Sigma, \boldsymbol{\alpha})\, p_{uec}(\beta_2|\Sigma, \boldsymbol{\alpha}, \lambda)\, d\Sigma\, d\boldsymbol{\alpha}\, d\lambda\, d\beta_2} \\ &= \frac{\iint p_{uec}(\Sigma)\, p_{uec}(\boldsymbol{\alpha}|\Sigma)\, p_{uec}(\lambda|\Sigma, \boldsymbol{\alpha})|_{\lambda=\mathbf{0}}\, d\Sigma\, d\boldsymbol{\alpha}}{\iint p_{uec}(\Sigma)\, p_{uec}(\boldsymbol{\alpha}|\Sigma)\, d\Sigma\, d\boldsymbol{\alpha}} \end{aligned} \tag{6.61}$$

we can sample Σ and $\boldsymbol{\alpha}$ from the prior $p_{uec}(\Sigma, \boldsymbol{\alpha})$, which is again the product of an inverted Wishart $p_{uec}(\Sigma)$ and a matrix normal $p_{uec}(\boldsymbol{\alpha}|\Sigma)$, and calculate the mean of the ratio of the integrands

$$\frac{p_{uec}(\Sigma^i, \boldsymbol{\alpha}^i)\, p_{uec}(\lambda|\Sigma^i, \boldsymbol{\alpha}^i)|_{\lambda=\mathbf{0}}}{p_{uec}(\Sigma^i, \boldsymbol{\alpha}^i)} = p_{uec}(\lambda|\Sigma^i, \boldsymbol{\alpha}^i)|_{\lambda=\mathbf{0}}, \tag{6.62}$$

which converges to c_r. Note that the simulation steps to compute the Bayes factor basically consists of the computation of the ratio of the marginal posterior distribution of λ evaluated in $\lambda = \mathbf{0}$, $p_{uec}(\lambda|Y)|_{\lambda=\mathbf{0}}$, and the marginal prior of λ in $\lambda = \mathbf{0}$, $p_{uec}(\lambda)|_{\lambda=\mathbf{0}}$. This implies that we in fact calculate the Savage-Dickey density ratio of Dickey (1971), see also Section 3.3.2.

In case of diffuse priors the value of c_r is infinity. We can again apply the empirical Bayes rule of Schotman and van Dijk (1991a) and choose a flat prior for λ on the 99% HPD region, like in Section 3.3.2. This is however not a straightforward solution since we have specified a conditional prior for λ, see (6.20). In this chapter we follow another approach. In order to define an interpretable Bayes factor, we start with natural conjugate priors and degenerate these proper priors such that they become diffuse priors. As we already have seen, the correction factor c_r depends on the ratio of the prior of the error correction cointegration model (6.7) and the prior of the unrestricted error correction model (6.8), which is the conditional prior of λ evaluated in $\lambda = \mathbf{0}$. Hence, to determine the correction factor in case of diffuse priors we degenerate the proper conditional prior for λ (6.25) in $\lambda = \mathbf{0}$. The conditional prior for λ reads

$$\begin{aligned} p_{uec}(\lambda|\Sigma, \boldsymbol{\alpha}) = (2\pi)^{-\frac{1}{2}(n-r)^2} |\boldsymbol{\alpha}_\perp \Sigma \boldsymbol{\alpha}_\perp'|^{-\frac{1}{2}(n-r)} |A_{22}|^{\frac{1}{2}(n-r)} \\ \exp(-\frac{1}{2}\mathrm{tr}((\boldsymbol{\alpha}_\perp \Sigma \boldsymbol{\alpha}_\perp')^{-1}(\lambda - l)' A_{22}(\lambda - l))), \end{aligned} \tag{6.63}$$

where l is defined below (6.25). To degenerate this proper prior we consider the density function of $\dot{\lambda} = A_{22}^{\frac{1}{2}}\lambda$,

$$p_{uec}(\dot{\lambda}|\Sigma, \boldsymbol{\alpha}) = (2\pi)^{-\frac{1}{2}(n-r)^2} |\boldsymbol{\alpha}_\perp \Sigma \boldsymbol{\alpha}_\perp'|^{-\frac{1}{2}(n-r)} \exp(-\frac{1}{2}\mathrm{tr}((\boldsymbol{\alpha}_\perp \Sigma \boldsymbol{\alpha}_\perp')^{-1}(\dot{\lambda} - \dot{l})'(\dot{\lambda} - \dot{l}))), \tag{6.64}$$

where $\dot{l} = A_{22}^{\frac{1}{2}} l$. Substituting $\lambda = \mathbf{0}$ (which implies $\dot{\lambda} = \mathbf{0}$) and degenerating by letting $\|\dot{l}\|$ going to zero, results in

$$\lim_{\|\dot{l}\| \to 0} p_{uec}(\dot{\lambda}|\Sigma, \boldsymbol{\alpha})|_{\dot{\lambda}=\mathbf{0}} = (2\pi)^{-\frac{1}{2}(n-r)^2} |\boldsymbol{\alpha}_\perp \Sigma \boldsymbol{\alpha}_\perp'|^{-\frac{1}{2}(n-r)}. \tag{6.65}$$

The second part of this limit ends up in the marginal prior of the error correction cointegration model for Σ and $\boldsymbol{\alpha}$, see (6.22), while the first part equals the correction factor c_r.

The Bayes factors (6.55) in case of a diffuse prior specification can now be computed by averaging the simulated $w(\Sigma^i, \boldsymbol{\alpha}^i|Y)$ as for the natural conjugate priors and by replacing c_r by

$$(2\pi)^{-\frac{1}{2}(n-r)^2}, \qquad r = 0, \ldots, n. \tag{6.66}$$

The resulting Bayes factor is closely related to the Posterior Information Criterion [PIC] of Phillips and Ploberger (1994, 1996), see Kleibergen and Paap (1996) for more details.

6.7 Invariant Specification

As we already have discussed, cointegration implies that the long run multiplier $\boldsymbol{\Pi}$ of the linear error correction model (6.6) has reduced rank. In the previous sections, we have analysed rank reduction as a restriction in the unrestricted error correction model (6.8). When the parameter showing the deviation from cointegration λ equals $\mathbf{0}$, this model simplifies to the error correction cointegration model (6.7). A drawback of this specification is however that different orderings of the variables in Y_t can lead to different results for the Bayes factors and the marginal posteriors under cointegration. This results as λ is correlated with the other parameters, even at $\lambda = \mathbf{0}$. In order to obtain a specification whose posterior of the parameters of the cointegration model is invariant with respect to parameter transformations, we need to model the restriction such that when it holds, the parameter resembling the restriction is (locally) uncorrelated with the other parameters. We therefore specify the parameter reflecting deviations from cointegration such that it only captures deviations which lie in spaces orthogonal to the cointegrating vectors β and their multiplicators $\boldsymbol{\alpha}$

$$\boldsymbol{\Pi} = \beta\boldsymbol{\alpha} + \beta_\perp\lambda^*\boldsymbol{\alpha}_\perp = (\beta\ \beta_\perp)\begin{pmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \lambda^* \end{pmatrix}\begin{pmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\alpha}_\perp \end{pmatrix}, \tag{6.67}$$

where $\beta = (\mathbf{I}_r\ -\beta_2')'$ and $\boldsymbol{\alpha}_\perp$ and $\beta_\perp$ are specified such that $\boldsymbol{\alpha}_\perp\boldsymbol{\alpha}' = \mathbf{0}$ with $\boldsymbol{\alpha}_\perp\boldsymbol{\alpha}_\perp' = \mathbf{I}_{n-r}$ and $\beta'\beta_\perp = \mathbf{0}$ with $\beta_\perp'\beta_\perp = \mathbf{I}_{n-r}$. Note that now $\boldsymbol{\alpha}_\perp$ does not equal $(-\boldsymbol{\alpha}_2'(\boldsymbol{\alpha}_1^{-1})'\ \mathbf{I}_{n-r})$ any more. When $\lambda^* = \mathbf{0}$, the long run multiplier $\boldsymbol{\Pi}$ in (6.67) displays rank reduction and cointegration occurs. This decomposition leads to an information matrix in the parameters $\boldsymbol{\alpha}$, λ^* and β_2, which is in $\lambda^* = \mathbf{0}$ block diagonal indicating the (local) uncorrelatedness between λ^* and $(\boldsymbol{\alpha}, \beta_2)$. The decomposition of $\boldsymbol{\Pi}$ in (6.67) is identical to a singular value decomposition of $\boldsymbol{\Pi}$,

$$\boldsymbol{\Pi} = U\,S\,V', \tag{6.68}$$

where U and V are $(n\times n)$ orthonormal matrices, S is an $(n\times n)$ diagonal matrix containing the positive singular values of $\boldsymbol{\Pi}$ (in decreasing order), see *e.g.* Magnus and Neudecker

(1988). If we write

$$U = \begin{pmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{pmatrix}, \; S = \begin{pmatrix} S_{11} & \mathbf{0} \\ \mathbf{0} & S_{22} \end{pmatrix} \text{ and } V = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix} \tag{6.69}$$

with U_{11}, S_{11}, V_{11} $(r \times r)$, U_{22}, S_{22}, V_{22} $((n-r) \times (n-r))$, U_{21}, V_{12} $((n-r) \times r)$ and U_{12}, V_{21} $(r \times (n-r))$ matrices, we obtain the following expressions for $\boldsymbol{\alpha}$, λ^* and β_2

$$\begin{aligned} \boldsymbol{\alpha} &= U_{11}\, S_{11}\, (V_{11}'\; V_{21}') \\ \lambda^* &= (U_{22}U_{22}')^{-\frac{1}{2}} U_{22}\, S_{22}\, V_{22}' (V_{22}V_{22}')^{-\frac{1}{2}} \\ \beta_2 &= -U_{21}\, U_{11}^{-1}, \end{aligned} \tag{6.70}$$

where we use that for a positive definite real symmetric matrix M, $M^{\frac{1}{2}} = C\Lambda^{\frac{1}{2}}C'$ where Λ is a diagonal matrix containing the eigenvalues of M and C contains the orthonormal eigenvectors of M and $M^{-\frac{1}{2}} = C\Lambda^{-\frac{1}{2}}C'$, see *e.g.* Johansen (1995, p. 222). The singular value decomposition (6.69) also shows how λ^* is identified, namely through the $(n-r)$ smallest singular values of $\boldsymbol{\Pi}$, which end up in S_{22}.

Note that under specification (6.67) for $\lambda^* = \mathbf{0}$ the error correction cointegration model is still the same as in (6.7), while for $\lambda^* \neq \mathbf{0}$ we have a different parameterisation of the unrestricted error correction than in (6.8). The same is true for the likelihood functions, but **not** for the prior and the posterior distributions. Since we have a new decomposition of $\boldsymbol{\Pi}$ the prior for the parameters of the unrestricted error correction model is different than in Section 6.3. We will use an asterisk to denote the priors, posteriors, likelihood functions and Bayes factors resulting from the new specification. Note that the prior for the cointegration model still equals the prior for the unrestricted error correction model evaluated in $\lambda^* = \mathbf{0}$. In the following subsections, we show how priors, posteriors and Bayes factors are constructed using the decomposition of $\boldsymbol{\Pi}$ in (6.67).

6.7.1 Prior Specification

The prior densities for the new decomposition of $\boldsymbol{\Pi}$ (6.67) can be constructed using the prior framework in Section 6.3. We specify a prior on Σ and $\boldsymbol{\Pi}$, which implies a prior for $(\Sigma, \boldsymbol{\alpha}, \lambda^*, \beta_2)$

$$p_{uec}^*(\Sigma, \boldsymbol{\alpha}, \lambda^*, \beta_2) = p_{lec}^*(\Sigma, \boldsymbol{\Pi}(\boldsymbol{\alpha}, \lambda^*, \beta_2))|J(\boldsymbol{\alpha}, \lambda^*, \beta_2)|, \tag{6.71}$$

where $\boldsymbol{\Pi}(\boldsymbol{\alpha}, \lambda^*, \beta_2)$ means that we evaluate $\boldsymbol{\Pi}$ in $(\boldsymbol{\alpha}, \lambda^*, \beta_2)$ and $|J(\boldsymbol{\alpha}, \lambda^*, \beta_2)|$ is the Jacobian of the transformation from $\boldsymbol{\Pi}$ to $(\boldsymbol{\alpha}, \lambda^*, \beta_2)$. In Appendix 6.C we show the derivation and the analytical expression of this Jacobian transformation. This Jacobian transformation can be combined with a diffuse or natural conjugate prior for $\boldsymbol{\Pi}$. The prior density (6.71) cannot be decomposed into a product of conditional and marginal densities belonging to a known class of probability density functions, like for the previous specification of the unrestricted error correction model (6.8), see Section 6.3. In case of a natural

conjugate prior for $\mathbf{\Pi}$, the properties of the priors for $\boldsymbol{\alpha}$, λ^* and β_2 therefore have to be obtained through simulation. The simulation algorithm constructed in the next subsection to obtain random draws from the posterior, can also be used to simulate from a natural conjugate prior.

For the cointegration model, the prior equals the prior (6.71) evaluated in $\lambda^* = \mathbf{0}$,

$$\begin{aligned} p^*_{ecc}(\Sigma, \boldsymbol{\alpha}, \beta_2) &= \frac{1}{c_r} p^*_{uec}(\Sigma, \boldsymbol{\alpha}, \lambda^*, \beta_2)|_{\lambda^*=\mathbf{0}} \\ &= \frac{1}{c_r} p^*_{uec}(\Sigma, \mathbf{\Pi}(\boldsymbol{\alpha}, \lambda^*, \beta_2))|_{\lambda^*=\mathbf{0}} |J(\boldsymbol{\alpha}, \lambda^*, \beta_2)|_{\lambda^*=\mathbf{0}}|, \end{aligned} \tag{6.72}$$

where c_r is the integrating constant like in (6.16).

In case of a diffuse prior for $\mathbf{\Pi}$ given Σ, $p_{lec}(\mathbf{\Pi}|\Sigma) \propto |\Sigma^{-1} \otimes (Y'_{-1}Y_{-1})|^{\frac{1}{2}}$, the quadratic form of $|J(\boldsymbol{\alpha}, \beta_2, \lambda^*)|_{\lambda^*=\mathbf{0}}|$ with $(\Sigma^{-1} \otimes (Y'_{-1}Y_{-1}))$, *i.e.* the information matrix, is block diagonal. This quadratic form (information matrix) is not block diagonal when $\lambda^* \neq \mathbf{0}$. The block diagonality implies that when cointegration occurs, the posterior of $\boldsymbol{\alpha}$ and β_2 is invariant with respect to the specification of $\boldsymbol{\alpha}$ and β_2, *i.e.* the posteriors of different specifications of $\boldsymbol{\alpha}$ and β_2 can be constructed from one another. When the information matrix of the unrestricted error correction model (6.8) is not block diagonal, the posteriors of $\boldsymbol{\alpha}$ and β_2 do depend on the chosen order of the elements of Y_t.

6.7.2 Posterior Distributions

The posterior of the unrestricted error correction model for the invariant specification is proportional to the prior (6.71) times the likelihood $\mathcal{L}^*_{uec}(\Sigma, \boldsymbol{\alpha}, \lambda^*, \beta_2)$, which is defined like in (6.10),

$$p^*_{uec}(\Sigma, \boldsymbol{\alpha}, \lambda^*, \beta_2|Y) \propto p^*_{uec}(\Sigma, \boldsymbol{\alpha}, \lambda^*, \beta_2)\, \mathcal{L}^*_{uec}(Y|\Sigma, \boldsymbol{\alpha}, \lambda^*, \beta_2). \tag{6.73}$$

This posterior density cannot be decomposed into a product of conditional and marginal densities belonging to a known class of probability density functions, like for the specification of the unrestricted cointegration model (6.8) in Section 6.4. We can however still simulate from this posterior since it is possible to sample from the posterior of the linear error correction model (6.6) which is the product of an inverted Wishart and a matrix normal distribution

Step 1: Draw Σ^i from $p_{lec}(\Sigma|Y)$

Draw $\mathbf{\Pi}^i$ from $p_{lec}(\mathbf{\Pi}|\Sigma^i, Y)$.

Step 2: Perform a singular value decomposition of $\mathbf{\Pi}^i = U^i\, S^i\, V^{i\prime}$.

Step 3: Compute $\boldsymbol{\alpha}^i$, λ^{*i} and β_2^i using (6.70).

The simulated values $\boldsymbol{\alpha}^i$, λ^{*i} and β_2^i can be used to compute marginal results. Likewise, we can use this simulation scheme to obtain marginal prior results for $\boldsymbol{\alpha}$, λ^* and β_2 if we specify a natural conjugate prior for Σ and $\boldsymbol{\Pi}$.

The posterior of the cointegration model equals the posterior of the error correction model evaluated in $\lambda^* = \mathbf{0}$

$$\begin{aligned} p^*_{ecc}(\Sigma, \boldsymbol{\alpha}, \beta_2|Y) &\propto p^*_{ecc}(\Sigma, \boldsymbol{\alpha}, \beta_2)\, \mathcal{L}^*_{ecc}(Y|\Sigma, \boldsymbol{\alpha}, \beta_2) \\ &\propto p^*_{uec}(\Sigma, \boldsymbol{\alpha}, \lambda^*, \beta_2|Y)|_{\lambda^*=\mathbf{0}}. \end{aligned} \tag{6.74}$$

Also under $\lambda^* = \mathbf{0}$ the conditional posterior of β_2 is of an unknown type. For instance under a diffuse prior specification for $(\Sigma, \boldsymbol{\Pi})$, $p_{lec}(\Sigma, \boldsymbol{\Pi}) \propto |\Sigma|^{-\frac{1}{2}(n+1)}|\Sigma^{-1} \otimes (Y'_{-1}Y_{-1})|^{\frac{1}{2}}$, the posterior of the cointegration model reads

$$\begin{aligned} p^*_{ecc}(\Sigma, \boldsymbol{\alpha}, \beta_2|Y) \propto |\Sigma|^{-\frac{1}{2}(T+n+1)} \left| \begin{pmatrix} (\boldsymbol{\alpha}\Sigma^{-1}\boldsymbol{\alpha}') \otimes (Y'_{2,-1}Y_{2,-1}) & (\boldsymbol{\alpha}\Sigma^{-1}) \otimes (Y'_{2,-1}Y_{-1}\beta) \\ (\Sigma^{-1}\boldsymbol{\alpha}') \otimes (\beta' Y'_{-1}Y_{2,-1}) & \Sigma^{-1} \otimes (\beta' Y'_{-1}Y_{-1}\beta) \end{pmatrix} \right|^{\frac{1}{2}} \\ |\boldsymbol{\alpha}_\perp \Sigma^{-1} \boldsymbol{\alpha}'_\perp|^{\frac{1}{2}(n-r)} |\beta'_\perp Y'_{-1}Y_{-1}\beta_\perp|^{\frac{1}{2}(n-r)} \exp(-\frac{1}{2}\mathrm{tr}(\Sigma^{-1}(\Delta Y - Y_{-1}\beta\boldsymbol{\alpha})'(\Delta Y - Y_{-1}\beta\boldsymbol{\alpha}))), \end{aligned} \tag{6.75}$$

where we use the expression for the Jacobian transformation $|J(\boldsymbol{\alpha}, \lambda^*, \beta_2)|$ evaluated in $\lambda^* = \mathbf{0}$ given in Appendix 6.C. This means that we cannot sample from the marginal posterior of Σ and $\boldsymbol{\alpha}$ using a Metropolis-Hastings algorithm and sample β_2 from the conditional posterior of β_2 given $(\Sigma, \boldsymbol{\alpha})$ like in Section 6.4.1. Hence, we need to sample Σ, $\boldsymbol{\alpha}$ and β_2 at once. If we opt for a M-H algorithm, we can take the posterior of the unrestricted error correction model (6.73) as *candidate-generating* density function, since we have already shown how to sample from this distribution. However, in this case we also sample λ^* which does not show up in the posterior of the cointegration model (6.74). To circumvent this problem we extend the posterior of the cointegration model (6.74) with a proper conditional density $g(\lambda^*|\Sigma, \boldsymbol{\alpha}, \beta_2, Y)$[1]

$$g(\lambda^*|\Sigma, \boldsymbol{\alpha}, \beta_2, Y)\, p_{ecc}(\Sigma, \boldsymbol{\alpha}, \beta_2|Y) \propto g(\lambda^*|\Sigma, \boldsymbol{\alpha}, \beta_2, Y)\, p^*_{uec}(\Sigma, \boldsymbol{\alpha}, \lambda^*, \beta_2|Y)|_{\lambda^*=\mathbf{0}}, \tag{6.76}$$

and sample from this distribution using the M-H approach. Since g is a proper density, the draws Σ, $\boldsymbol{\alpha}$ and β_2 can be seen as draws from the posterior (6.74). The acceptance-rejection step now depends on the ratio of the extended posterior of the cointegration model (6.76) and the posterior of the unrestricted error correction model (6.73)

$$w^*(\Sigma, \boldsymbol{\alpha}, \lambda^*, \beta_2|Y) = \frac{g(\lambda^*|\Sigma, \boldsymbol{\alpha}, \beta_2, Y)\, p^*_{uec}(\Sigma, \boldsymbol{\alpha}, \lambda^*, \beta_2|Y)|_{\lambda^*=\mathbf{0}}}{p^*_{uec}(\Sigma, \boldsymbol{\alpha}, \lambda^*, \beta_2|Y)}. \tag{6.77}$$

Implementing this in a M-H sampler results in

Step 1: Draw $(\Sigma^{i+1}, \boldsymbol{\alpha}^{i+1}, \lambda^{*i+1}, \beta_2^{i+1})$ from (6.73).

[1]This solution is based on the ideas in Chen (1994).

Step 2: Accept $(\Sigma^{i+1}, \boldsymbol{\alpha}^{i+1}, \lambda^{*i+1}, \beta_2^{i+1})$ with probability $\min\left(\frac{w^*(\Sigma^{i+1},\boldsymbol{\alpha}^{i+1},\lambda^{*i+1},\beta_2^{i+1}|Y)}{w^*(\Sigma^{i},\boldsymbol{\alpha}^{i},\lambda^{*i},\beta_2^{i}|Y)}, 1\right)$

otherwise $(\Sigma^{i+1}, \boldsymbol{\alpha}^{i+1}, \lambda^{*i+1}, \beta_2^{i+1}) = (\Sigma^{i}, \boldsymbol{\alpha}^{i}, \lambda^{*i}, \beta_2^{i})$.

Since the *candidate-generating* density function has to approximate the density from which one wants to sample, it is necessary to take for $g(\lambda^*|\Sigma, \boldsymbol{\alpha}, \beta_2, Y)$ a density function which is close to the conditional posterior of λ^*. Therefore, the choice of g depends on the functional form of the prior for $(\Sigma, \mathbf{\Pi})$. The decomposition of the trace of the likelihood function given in Appendix 6.C shows that under a diffuse prior specification a convenient choice for g is

$$g(\lambda^*|\Sigma, \boldsymbol{\alpha}, \beta_2, Y) = (2\pi)^{-\frac{1}{2}(n-r)^2} |\boldsymbol{\alpha}_\perp \Sigma^{-1} \boldsymbol{\alpha}_\perp'|^{\frac{1}{2}(n-r)} |\beta_\perp' Y_{-1}' Y_{-1} \beta_\perp|^{\frac{1}{2}(n-r)} \exp(-\frac{1}{2}\mathrm{tr}(\beta_\perp' Y_{-1}' Y_{-1} \beta_\perp (\lambda^* - \hat{\lambda}^*) \boldsymbol{\alpha}_\perp \Sigma^{-1} \boldsymbol{\alpha}_\perp' (\lambda^* - \hat{\lambda}^*)')), \quad (6.78)$$

with $\hat{\lambda}^* = (\beta_\perp' Y_{-1}' Y_{-1} \beta_\perp)^{-1} \beta_\perp' Y_{-1}' \Delta Y \Sigma^{-1} \boldsymbol{\alpha}_\perp' (\boldsymbol{\alpha}_\perp \Sigma^{-1} \boldsymbol{\alpha}_\perp')^{-1}$. This results in the following expression for the weight function w^*

$$w^*(\Sigma, \boldsymbol{\alpha}, \lambda^*, \beta_2|Y) = (2\pi)^{-\frac{1}{2}(n-r)^2} \frac{|J(\boldsymbol{\alpha}, \lambda^*, \beta_2)|_{\lambda^*=\mathbf{0}}|}{|J(\boldsymbol{\alpha}, \lambda^*, \beta_2)|} |\boldsymbol{\alpha}_\perp \Sigma^{-1} \boldsymbol{\alpha}_\perp'|^{\frac{1}{2}(n-r)} |\beta_\perp' Y_{-1}' Y_{-1} \beta_\perp|^{\frac{1}{2}(n-r)} \exp(-\frac{1}{2}\mathrm{tr}((\beta_\perp' Y_{-1}' Y_{-1} \beta_\perp) \hat{\lambda}^* (\boldsymbol{\alpha}_\perp \Sigma^{-1} \boldsymbol{\alpha}_\perp') \hat{\lambda}^{*\prime})). \quad (6.79)$$

The functional form of the Jacobian is given in Appendix 6.C. It is straightforward to show how the density g and the weight function w^* change when we, instead of a diffuse prior, use a natural conjugate prior for $(\Sigma, \mathbf{\Pi})$: $(Y_{-1}' Y_{-1})$ changes to $(Y_{-1}' Y_{-1} + A)$ and $(Y_{-1}' \Delta Y)$ changes to $(AP + Y_{-1}' \Delta Y)$, see also Section 6.4

The Metropolis-Hastings sampler presented in this section may lead to high rejection frequencies and therefore slow convergence. An alternative approach is importance sampling, see Kloek and van Dijk (1978) and Geweke (1989). The weight function w^* in (6.79), evaluated in the draws, represents in that case important weights, see also Chen (1994) and Verdinelli and Wasserman (1995, p. 615). In the next subsection we show how we can use importance sampling techniques to compute posterior odds to compare cointegration models with different number of cointegration relations for the specification (6.67).

6.7.3 Bayes Factors

We can perform posterior odds analysis for the new decomposition of $\mathbf{\Pi}$ (6.67) the same way as in Section 6.6. However, since we cannot decompose the posterior in known densities the computation of the Bayes factors is more complicated. Again the Bayes factors are defined as the ratio of the marginal likelihoods of the cointegration model and the unrestricted error correction model like in (6.55). Since it is still true that the prior and the likelihood of the error correction cointegration model equals the prior and the

likelihood of the unrestricted error correction model times $1/c_r$, the Bayes factor simplifies to

$$\begin{aligned}
\mathrm{BF}^*(r|n) &= \frac{\iiint p^*_{ecc}(\Sigma, \boldsymbol{\alpha}, \beta_2)\, \mathcal{L}^*_{ecc}(Y|\Sigma, \boldsymbol{\alpha}, \beta_2)\, d\Sigma\, d\boldsymbol{\alpha}\, d\beta_2}{\iiiint p^*_{uec}(\Sigma, \boldsymbol{\alpha}, \lambda^*, \beta_2)\, \mathcal{L}^*_{uec}(Y|\Sigma, \boldsymbol{\alpha}, \lambda^*, \beta_2)\, d\Sigma\, d\boldsymbol{\alpha}\, d\lambda^*\, d\beta_2} \\
&= \frac{\iiint \frac{1}{c_r} p^*_{uec}(\Sigma, \boldsymbol{\alpha}, \lambda^* \beta_2)|_{\lambda^*=\mathbf{0}}\, \mathcal{L}^*_{uec}(Y|\Sigma, \boldsymbol{\alpha}, \lambda^*, \beta_2)|_{\lambda^*=\mathbf{0}}\, d\Sigma\, d\boldsymbol{\alpha}\, d\beta_2}{\iiiint p^*_{uec}(\Sigma, \boldsymbol{\alpha}, \lambda^*, \beta_2)\, \mathcal{L}^*_{uec}(Y|\Sigma, \boldsymbol{\alpha}, \lambda^*, \beta_2)\, d\Sigma\, d\boldsymbol{\alpha}\, d\lambda^*\, d\beta_2} \\
&= \frac{\iiint \frac{1}{c_r} p^*_{uec}(\Sigma, \boldsymbol{\alpha}, \lambda^*, \beta_2|Y)|_{\lambda^*=\mathbf{0}}\, d\Sigma\, d\boldsymbol{\alpha}\, d\beta_2}{\iiiint p^*_{uec}(\Sigma, \boldsymbol{\alpha}, \lambda^*, \beta_2|Y)\, d\Sigma\, d\boldsymbol{\alpha}\, d\lambda^*\, d\beta_2}. \qquad (6.80)
\end{aligned}$$

For the computation of this Bayes factor we encounter differences with Section 6.6.1. Since we cannot decompose the posterior in known densities, we cannot integrate out β_2 analytically in the numerator and denominator. The same is true for the λ^* in the denominator. Since λ^* enters the integral in the denominator but not the in the numerator we cannot use the simulation procedure, which is proposed in Section 6.6.1. However, Chen (1994) shows that we can adjust the simulation procedure by extending the numerator with the integral $\int g(\lambda^*|\Sigma, \boldsymbol{\alpha}, \beta_2) d\lambda^*$, where g is a proper density function

$$\begin{aligned}
\mathrm{BF}^*(r|n) &= \frac{\iiint \frac{1}{c_r} p^*_{uec}(\Sigma, \boldsymbol{\alpha}, \lambda^*, \beta_2|Y)|_{\lambda^*=\mathbf{0}} \left(\int g(\lambda^*|\Sigma, \boldsymbol{\alpha}, \beta_2)\, d\lambda^*\right) d\Sigma\, d\boldsymbol{\alpha}\, d\beta_2}{\iiiint p^*_{uec}(\Sigma, \boldsymbol{\alpha}, \lambda^*, \beta_2|Y)\, d\Sigma\, d\boldsymbol{\alpha}\, d\lambda^*\, d\beta_2} \\
&= \frac{\frac{1}{c_r} \iiiint p^*_{uec}(\Sigma, \boldsymbol{\alpha}, \lambda^*, \beta_2|Y)|_{\lambda^*=\mathbf{0}}\, g(\lambda^*|\Sigma, \boldsymbol{\alpha}, \beta_2)\, d\Sigma\, d\boldsymbol{\alpha}\, d\lambda^*\, d\beta_2}{\iiiint p^*_{uec}(\Sigma, \boldsymbol{\alpha}, \lambda^*, \beta_2|Y)\, d\Sigma\, d\boldsymbol{\alpha}\, d\lambda^*\, d\beta_2}. \qquad (6.81)
\end{aligned}$$

An appropriate candidate for $g(\lambda^*|\Sigma, \boldsymbol{\alpha}, \beta_2)$ is a density function which is close to the conditional posterior of λ^*, see Chen (1994) for details. Therefore, the conditional density function (6.78) is again a good candidate. We can now calculate the ratio of the two integrals by simulating Σ, $\boldsymbol{\alpha}$, λ^* and β_2 from the posterior of the unrestricted error correction model (see previous subsection). For the sampled parameters, we calculate the ratio of the two integrands, which equals the importance weights (6.79). The sum of these simulated importance weights converges to the ratio of the two integrals in (6.81), see Geweke (1989).

To compute the Bayes factors we need the value of c_r. For a diffuse prior specification we take again for c_r the factor (6.66). To obtain c_r for a natural conjugate prior specification we can use the above mentioned simulation technique of Chen (1994)

$$\begin{aligned}
c_r &= \iiint p^*_{uec}(\Sigma, \boldsymbol{\alpha}, \lambda^*, \beta_2)|_{\lambda^*=\mathbf{0}}\, d\Sigma\, d\boldsymbol{\alpha}\, d\beta_2 \\
&= \frac{\iiint p^*_{uec}(\Sigma, \boldsymbol{\alpha}, \lambda^*, \beta_2)|_{\lambda^*=\mathbf{0}}\, d\Sigma\, d\boldsymbol{\alpha}\, d\beta_2}{\iiiint p^*_{uec}(\Sigma, \boldsymbol{\alpha}, \lambda^*, \beta_2)\, d\Sigma\, d\boldsymbol{\alpha}\, d\lambda^* d\beta_2} \\
&= \frac{\iiint p^*_{uec}(\Sigma, \boldsymbol{\alpha}, \lambda^*, \beta_2)|_{\lambda^*=\mathbf{0}} \left(\int h(\lambda^*|\Sigma, \boldsymbol{\alpha}, \beta_2)\, d\lambda^*\right) d\Sigma\, d\boldsymbol{\alpha}\, d\beta_2}{\iiiint p^*_{uec}(\Sigma, \boldsymbol{\alpha}, \lambda^*, \beta_2)\, d\Sigma\, d\boldsymbol{\alpha}\, d\lambda^* d\beta_2} \\
&= \frac{\iiiint p^*_{uec}(\Sigma, \boldsymbol{\alpha}, \lambda^*, \beta_2)|_{\lambda^*=\mathbf{0}}\, h(\lambda^*|\Sigma, \boldsymbol{\alpha}, \beta_2)\, d\Sigma\, d\boldsymbol{\alpha}\, d\lambda^*\, d\beta_2}{\iiiint p^*_{uec}(\Sigma, \boldsymbol{\alpha}, \lambda^*, \beta_2)\, d\Sigma\, d\boldsymbol{\alpha}\, d\lambda^* d\beta_2}, \qquad (6.82)
\end{aligned}$$

where $h(\lambda^*|\Sigma, \boldsymbol{\alpha}, \beta_2)$ is a proper conditional density function. Simulate from the prior $p^*_{uec}(\Sigma, \boldsymbol{\alpha}, \lambda^*, \beta_2)$ (6.71) and compute the ratio of the integrands of the numerator and denominator. The sum of these ratios converges to c_r. An appropriate density function h for the prior specification (6.23) is a density function which is close to the conditional prior of λ^*

$$h(\lambda^*|\Sigma, \boldsymbol{\alpha}, \beta_2) = (2\pi)^{-\frac{1}{2}(n-r)^2} |\boldsymbol{\alpha}_\perp \Sigma^{-1} \boldsymbol{\alpha}'_\perp|^{\frac{1}{2}(n-r)} |\beta'_\perp A \beta_\perp|^{\frac{1}{2}(n-r)} \exp(-\frac{1}{2}\mathrm{tr}(\beta'_\perp A \beta_\perp (\lambda^* - l^*) \boldsymbol{\alpha}_\perp \Sigma^{-1} \boldsymbol{\alpha}'_\perp (\lambda^* - l^*)')), \quad (6.83)$$

with $l^* = (\beta'_\perp A \beta_\perp)^{-1} \beta'_\perp A P \Sigma^{-1} \boldsymbol{\alpha}'_\perp (\boldsymbol{\alpha}_\perp \Sigma^{-1} \boldsymbol{\alpha}'_\perp)^{-1}$.

Finally, in Section 6.6.1 we have seen that the Bayes factor can be computed using the Savage-Dickey density ratio of Dickey (1971). In fact, the above proposed simulation method to compute the Bayes factor basically computes the height of the marginal posterior of λ^* in $\mathbf{0}$. Hence, an alternative strategy to compute the Bayes factor is to use the simulated values of λ^* and a kernel estimator to compute the height of this marginal posterior, see Silverman (1986). The same reasoning holds for the computation of c_r.

In the next section we use the invariant specification to compute Bayes factors for simulated and real time series.

6.8 Illustrative Examples

To illustrate the applicability of the, in the previous sections, constructed methods and procedures for Bayesian cointegration analyses, we consider four simulated time series, the UK data analysed in Hendry and Doornik (1994) and the Danish data analysed in Johansen and Juselius (1990).

6.8.1 Simulated Series

We consider the following four data generating processes [DGPs],

$$\begin{aligned}
\text{I} \;:\; \Delta Y_t &= \begin{pmatrix} 0.1 \\ 0.1 \\ 0.1 \end{pmatrix} + \varepsilon_t \\
\text{II} \;:\; \Delta Y_t &= \begin{pmatrix} 0.1 \\ 0.1 \\ 0.1 \end{pmatrix} + \begin{pmatrix} -0.2 \\ 0.2 \\ 0.2 \end{pmatrix} \begin{pmatrix} 1 & 0 & -1 \end{pmatrix} Y_{t-1} + \varepsilon_t \\
\text{III} \;:\; \Delta Y_t &= \begin{pmatrix} 0.1 \\ 0.1 \\ 0.1 \end{pmatrix} + \begin{pmatrix} -0.2 & -0.2 \\ 0.2 & -0.2 \\ 0.2 & 0.2 \end{pmatrix} \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{pmatrix} Y_{t-1} + \varepsilon_t \\
\text{IV} \;:\; \Delta Y_t &= \begin{pmatrix} 0.1 \\ 0.1 \\ 0.1 \end{pmatrix} + \begin{pmatrix} -0.2 & -0.2 & -0.2 \\ 0.2 & -0.2 & -0.2 \\ 0.2 & 0.2 & -0.2 \end{pmatrix} \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix} Y_{t-1} + \varepsilon_t,
\end{aligned} \quad (6.84)$$

where $\varepsilon_t \sim \text{NID}(\mathbf{0}, \mathbf{I}_3)$ and the sample size T is 100 observations. The four DGPs contain 0, 1, 2 and 3 cointegration relations, respectively. DGP I contains three unit roots, DGP II contains 2 unit roots and a root 0.6, DGP III contains the roots 1, 0.6 and 0.6, and DGP IV contains the roots 0.8, 0.6 and 0.6.

To analyse the simulated series, we consider a VAR(1) model with a constant term, which corresponds to the specification in the DGP. The first step in the Bayesian analysis is to specify a prior on the vector autoregressive parameters Π and on the covariance matrix Σ. We use the diffuse prior specification (6.18).

First, we discuss Bayes factors for rank reduction. We give each cointegration rank the same prior probability $\Pr[\text{rank} = r] = \frac{1}{4}$, $r = 0, \ldots, 3$, see (6.53). Given the priors and prior probabilities, we can compare models with reduced rank (cointegration models) with the full rank unrestricted error correction model. The Bayes factors are based on posterior distributions resulting from the invariant decomposition of Π given in (6.67). The second column of Table 6.1 displays the Bayes factors (6.81) for the four DGPs. The Bayes factors are calculated using the correction factor (6.66). A Bayes factor exceeding one (or $\ln(\text{BF}^*)$ exceeding zero) indicates that rank r is preferred above the full rank situation. For instance, for DGP I every rank reduction is preferred, while for DGP IV the full rank situation is always preferred. The Bayes factors can be translated into posterior probabilities for the cointegration ranks, see (6.57). These are displayed in the second column of Table 6.1. They assign in all cases more than 85% probability to the correct cointegration rank.

The fourth column of Table 6.1 contains the Bayesian LM statistics. These statistics indicate whether rank reduction in Π (cointegration) is plausible with respect to the full rank situation, *i.e.* three cointegration relations, see Section 6.5.2. The LM statistics are calculated using the posterior distribution resulting from decomposition (6.9). These LM statistics have to be compared with a χ^2 distribution with $(3-r)^2$ degrees of freedom which number is shown in the fifth column. The sixth column of Table 6.1 shows the p-values of the calculated statistics. For instance, for DGP IV none of the models with reduced rank is plausible, while for DGP III only a model with two cointegration relations is plausible. In general, the Bayesian LM results lead to the right cointegration rank. In the last two columns we report the results of the classical Johansen trace tests denoted by $\text{LR}(r|3)$. Notice that the values of these statistics are often roughly of the same magnitude as the computed Bayesian LM statistics. The p-values based on the classical asymptotic distribution show that the Johansen trace statistics also point out the right decision about the cointegration rank.

6.8.2 Small Monetary Model for the UK

Hendry and Doornik (1994) construct a small linear dynamic monetary model for the United Kingdom. The model consist of the variables nominal M_1, denoted by m_t, total final expenditure y_t, the total final expenditure deflator p_t, and the differential between the three-month local authority interest rate and the M_1 retail sight-deposit interest rate denoted by r_t. The latter represents the opportunity cost of holding M_1. All variables

Table 6.1. Bayes factors, posterior probabilities, Bayesian LM statistics and classical likelihood ratio tests for the four DGPs.

r	$\ln(\mathrm{BF}^*(r\|3))$[1]	$\Pr[r\|Y]$[2]	$\mathrm{LM}(r\|3)$	dof	p-value	$\mathrm{LR}(r\|3)$[3]	p-value[4]
			DGP I				
0	25.06	1.00	10.71	9	0.30	10.97	0.96
1	12.59	0.00	5.72	4	0.22	5.10	0.80
2	3.61	0.00	0.72	1	0.39	1.38	0.24
3	0.00	0.00					
			DGP II				
0	6.82	0.01	38.20	9	0.00	44.98	0.00
1	12.04	0.99	6.85	4	0.14	6.39	0.64
2	3.78	0.00	1.75	1	0.18	1.09	0.29
3	0.00	0.00					
			DGP III				
0	−23.87	0.00	74.76	9	0.00	94.86	0.00
1	−4.18	0.00	28.59	4	0.00	33.05	0.00
2	2.10	0.89	2.84	1	0.09	2.95	0.09
3	0.00	0.11					
			DGP IV				
0	−23.96	0.00	75.52	9	0.00	90.93	0.00
1	−8.86	0.00	36.15	4	0.00	39.68	0.00
2	−2.88	0.05	11.69	1	0.00	12.32	0.00
3	0.00	0.95					

[1] A Bayes factor $\ln(\mathrm{BF}^*(r|3)) > 0$ denotes that a cointegration model with r cointegration relations is more likely than a model with n cointegration relations.

[2] Posterior probability of the cointegration rank (6.57) is based on equal prior probabilities (6.53) for every rank r.

[3] Johansen (1991) trace test statistic.

[4] The p-values are based on the asymptotic classical distribution.

are in logs except for the interest rate r_t.

In this section we analyse the same UK data as in Hendry and Doornik (1994). We have the same quarterly observed series of m_t, y_t, p_t and r_t for the period 1963.I–1989.II. The data are seasonally adjusted. The first step in the modelling strategy is to specify an unrestricted VAR model. Hendry and Doornik (1994) propose a VAR(2) model for the four-dimensional vector of time series $Y_t = (m_t - p_t, y_t, \Delta p_t, r_t)'$

$$\Delta Y_t = \mu + \tau t + \Pi Y_{t-1} + \bar{\Phi}_1 \Delta Y_{t-1} + \xi_1 \mathrm{DOIL}_t + \xi_2 \mathrm{DOUT}_t + \varepsilon_t \tag{6.85}$$

where $\varepsilon_t \sim \mathrm{NID}(\mathbf{0}, \Sigma)$, ξ_1 and ξ_2 are (4×1) parameter vectors and DOIL_t and DOUT_t are dummy variables to capture outlying observations caused by the Heath-Barber boom and the first effects of the Thatcher government, and the two oil crises respectively, see Hendry and Doornik (1994) for details.[2] The trend t and the dummy variable DOUT_t are restricted in the cointegration space, *i.e.* $\alpha_\perp' \tau = \mathbf{0}$ and $\alpha_\perp' \xi_2 = \mathbf{0}$, which means that the vector $(t \ \mathrm{DOUT}_t)'$ is added to the Y_{t-1} vector and that Π becomes a (6×4) matrix.

The first part of Table 6.2 displays the results of a Bayesian cointegration analysis for the model (6.85). In the first row the results for a model without the dummy variables DOIL_t and DOUT_t are reported. The results are based on a diffuse (Jeffreys') prior for the parameters in (6.85). The Bayes factor are computed using the posterior distributions resulting from the invariant decomposition in (6.67), while the Bayesian LM statistics are based on the posterior distributions resulting from decomposition (6.9). We assume equal prior probabilities $\Pr[\mathrm{rank} = r] = \frac{1}{5}$, $r = 0, \ldots, 4$. The second and third column of the table show the Bayes factors and the implied posterior probabilities over the cointegration rank. The Bayes factors favour every rank reduction over a full rank model. The posterior probabilities assign about 100% probability to rank one and 0% to every other rank. The results of the classical Johansen trace tests are reported in the last two columns. These tests indicate no cointegration relation at a 5% level of significance. On basis of the Bayesian LM statistics we even opt for two cointegration relations.

The results change if we include the dummy variables DOUT_t and DOIL_t like in Hendry and Doornik (1994), see second row of Table 6.2. The posterior probabilities now also indicate two cointegration relations between the series in Y_t. The same is true for the Johansen trace statistics. However, a model with three cointegration relation is not unlikely according to the Bayesian LM statistic. Note that the degrees of freedom are different from $(n - r)^2$ due to the restricted trend. The difference in results between the Bayesian LM statistics and the Johansen trace statistics follow from the fact the LM statistic is exactly χ^2 distributed and not asymptotically a functional of Brownian motions since we consider the data as given.

6.8.3 Danish Money Demand

Johansen and Juselius (1990) analyse the demand function for money for the Danish economy using a VAR model. Their model consist of M_2 denoted by m_t, real income y_t,

[2] DOUT_t is zero except for unity in 1972.IV, 1973.I and 1979.II and DOIL_t is zero except in 1973.III, 1973.IV and 1979.III.

Table 6.2. Bayes factors, posterior probabilities, Bayesian LM statistics and classical likelihood ratio tests for the UK and Danish data.

r	$\ln(\mathrm{BF}^*(r\|4))$[1]	$\Pr[r\|Y]$[2]	$\mathrm{LM}(r\|4)$	dof	p-value	$\mathrm{LR}(r\|4)$[3]	p-value[4]
			UK data				
			no dummies and restricted trend				
0	24.58	0.00	94.84	20	0.00	119.38	0.00
1	37.30	1.00	39.75	12	0.00	40.89	0.08
2	28.02	0.00	15.12	6	0.02	12.12	0.80
3	14.96	0.00	6.22	2	0.05	4.48	0.68
4	0.00	0.00					
			dummies and restricted trend				
0	25.71	0.05	119.07	24	0.00	152.85	0.00
1	18.50	0.00	69.88	15	0.00	71.42	0.00
2	28.56	0.95	19.28	8	0.01	19.65	0.24
3	20.84	0.00	6.32	3	0.10	6.43	0.40
4	0.00	0.00					
			Danish data				
			unrestricted constant				
0	20.65	0.08	37.86	16	0.00	45.67	0.08
1	23.06	0.92	17.54	9	0.04	17.07	0.63
2	15.92	0.00	7.75	4	0.10	6.71	0.61
3	5.18	0.00	1.18	1	0.27	0.38	0.54
4	0.00	0.00					
			restricted constant				
0	44.36	1.00	40.63	20	0.00	49.14	0.11
1	26.10	0.00	19.03	12	0.08	19.06	0.79
2	21.43	0.00	10.09	6	0.12	8.69	0.77
3	10.69	0.00	3.21	2	0.20	2.35	0.70
4	0.00	0.00					

[1] A Bayes factor $\ln(\mathrm{BF}^*(r|4)) > 0$ denotes that a cointegration model with r cointegration relations is more likely than a model with n cointegration relations.

[2] Posterior probability of the cointegration rank (6.57) is based on equal prior probabilities (6.53) for every rank r.

[3] Johansen (1991) trace test statistic.

[4] The p-values are based on the asymptotic classical distribution.

price level p_t and the costs of holding money. The costs of holding money is approximated by a difference between the bank deposit rate i_t^d for interest bearing deposits and the bond rate i_t^b. All variables are in logs. Since the inflation rate Δp_t does not alter the cointegration analysis significantly, this variable is not considered in the Johansen and Juselius study.

In this subsection we analyse the same Danish data as in Johansen and Juselius (1990). We have quarterly observed series of m_t, i_t^d, i_t^b and y_t for the period 1974.1–1987.3. The cointegration analysis is performed in the following VAR(2) model,

$$\begin{pmatrix} \Delta m_t \\ \Delta y_t \\ \Delta i_t^b \\ \Delta i_t^d \end{pmatrix} = \mu + \sum_{s=1}^{3} \bar{\delta}_s \bar{D}_{s,t} + \Pi \begin{pmatrix} m_{t-1} \\ y_{t-1} \\ i_{t-1}^b \\ i_{t-1}^d \end{pmatrix} + \bar{\Phi}_1 \begin{pmatrix} \Delta m_{t-1} \\ \Delta y_{t-1} \\ \Delta i_{t-1}^b \\ \Delta i_{t-1}^d \end{pmatrix} + \varepsilon_t, \tag{6.86}$$

where $\bar{D}_{s,t}$ represents seasonal dummies with zero mean and $\bar{\delta}_s$ is a four-dimensional parameter vector, $s = 1, \ldots, 3$. Although it is not likely that real income does not have a linear trend, Johansen and Juselius restrict the constant in the cointegrating space, *i.e.* $\alpha_\perp' \mu = \mathbf{0}$, see the end of Section 5.3. Hence, the Π matrix is extended with an extra row and the Y_{-1} matrix with an extra column.

The second part of Table 6.2 displays the results of a Bayesian cointegration analysis for the Danish data. The results are based on a diffuse (Jeffreys') prior for the parameters and equal prior probabilities (6.53) $\Pr[\text{rank} = r] = \frac{1}{5}$, $r = 0, \ldots, n$. Again, the results are based on posterior distributions resulting from the decomposition (6.9) except for the Bayes factor analyses, which uses the posterior of the invariant decomposition (6.67). First, we consider a model where the constant is not restricted in the cointegrating space ($\alpha_\perp' \mu \neq \mathbf{0}$). The second and third column show the Bayes factors and the implied posterior probabilities over the cointegration rank. The Bayes factors favour every rank reduction over a full rank model and lead to 92% posterior probability for a model with one cointegration relation. The fourth column of Table 6.2 display the outcomes of the Bayesian LM statistics. Only the LM(3|4) and LM(2|4) statistics lie inside the 95% one-sided HPD interval, which implies that two cointegration relations between m_t, i_t^d, i_t^b and y_t are plausible. The Johansen trace test statistics indicate no cointegration relation at a 5% level of significance as in the classical approach the asymptotic distribution is not χ^2 but a functional of Brownian motions.

In case we restrict the constant in the cointegrating space, we see that the Bayes factors again favour every rank reduction over a full rank model. The posterior probabilities assign 100% probability to the model with zero cointegration relations and no probability to the other models. The Bayes factor that compares within the model with one cointegration relation whether the constant has to be restricted in the cointegrating space, equals the ratio of the Bayes factors of the analysis with the restricted and the unrestricted constant. The natural logarithm of this ratio (26.10−23.06=3.04) is positive, which indicates that the restriction of the constant in the cointegration space ($\alpha_\perp' \mu = \mathbf{0}$) is more likely than an unrestricted constant ($\alpha_\perp' \mu \neq \mathbf{0}$). Since only the LM(0|4) statistic is outside the 95%

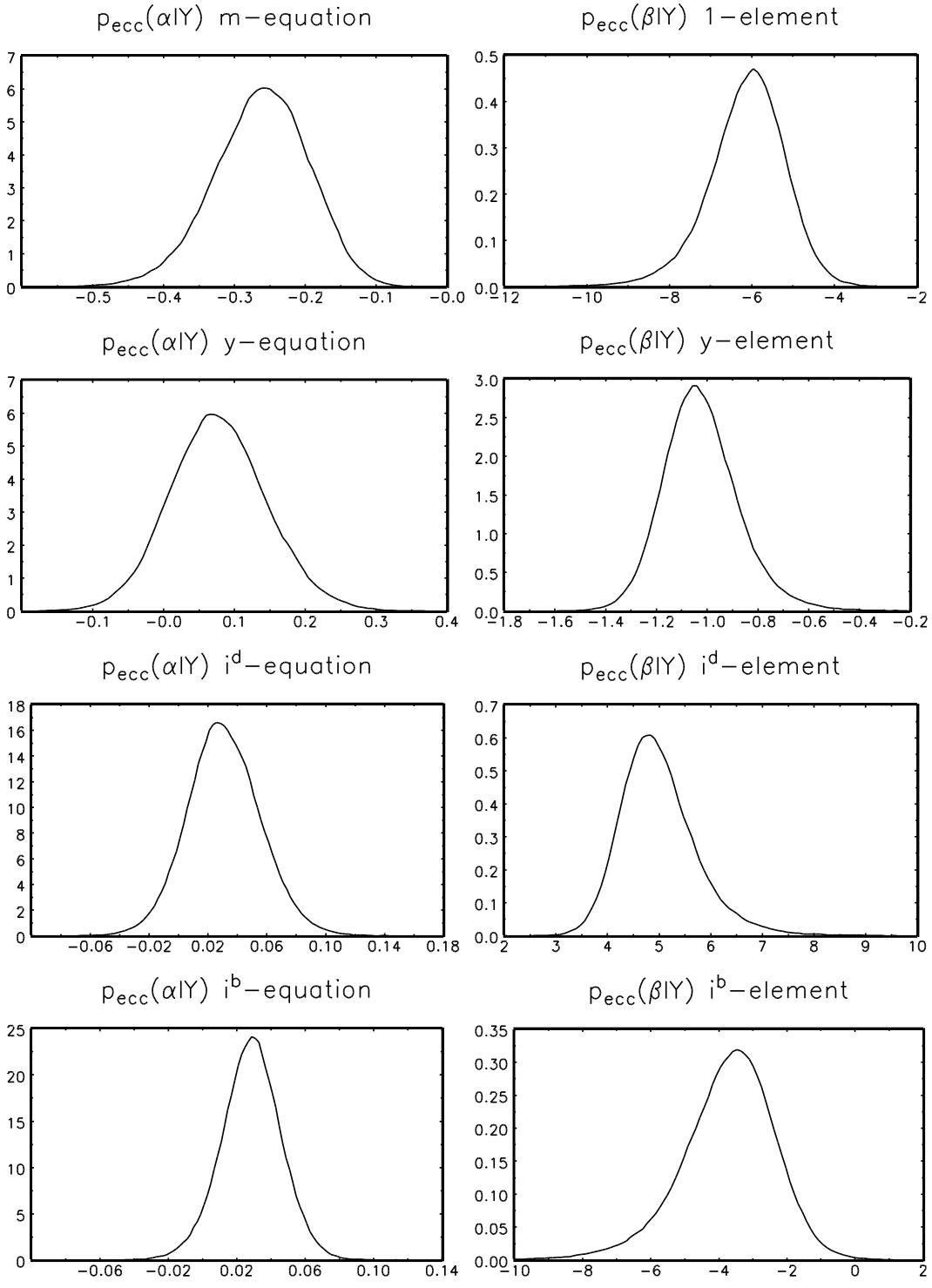

Figure 6.1. Marginal posterior densities of the adjustment parameters α and the cointegrating vector β for the UK series.

one-sided HPD interval, the LM tests indicate one cointegration relation between m_t, i_t^d, i_t^b and y_t. Note that the degrees of freedom are different from $(n-r)^2$ due to the restricted constant. In the classical analysis, we have to test at 11% level of significance to find one cointegration relation according to the Johansen trace statistics.

Using the M-H simulation algorithm in Section 6.4.1 we compute posterior results for the cointegration model with one cointegration relation and the constant restricted in the cointegrating space ($\alpha'_\perp \mu = \mathbf{0}$) under a diffuse prior specification (6.22). This is the model proposed by Johansen and Juselius (1990). The posterior means of the α and β parameters are

$$\begin{array}{cccccc} & m_t & y_t & i_t^b & i_t^d & \\ \alpha' = (& -0.26 & 0.08 & 0.03 & 0.03 &) \\ & (0.07) & (0.07) & (0.02) & (0.02) & \end{array} \text{ and } \begin{array}{cccccc} & m_t & y_t & i_t^b & i_t^d & 1 \\ \beta' = (& 1 & -1.02 & 4.99 & -3.76 & -6.23 \ ), \\ & (-) & (0.15) & (0.76) & (1.42) & (0.95) \end{array}$$

where the posterior standard deviations are between parentheses. Note that the posterior means correspond reasonably well to the maximum likelihood estimates in Table 2 of Johansen and Juselius (1990). Figure 6.1 shows the marginal posterior densities of the adjustment parameters α and the cointegrating vector β. The first column shows the marginal posteriors of the α parameters. The marginal posterior of the adjustment parameter for the money equation is situated far away from zero. This is not the case for the other adjustment parameters, where zero lies within the 95% one-sided HPD regions of the marginal posteriors. The second column of Figure 6.1 shows the marginal posteriors of the β parameters. These marginal posterior distributions are more skewed and have fatter tails. The posterior masses of the marginal posteriors of the cointegration parameters are situated far away from zero except for the i^b-element.

In summary, although the examples in this section are simple, they show that Bayesian techniques provide useful tools to analyse cointegration. Bayes factors and Bayesian LM statistics indicate whether rank reduction is plausible. In the empirical examples, the Bayesian LM statistics often result in more cointegration relations than the Bayes factors and the classical LR tests. This is due to the fact that the Bayesian LM statistic have to be compared with standard χ^2 distributions instead of functionals of Brownian motions. The Bayes factors can be used to calculate posterior probabilities for each cointegration rank, to show the best model. If there is no clear preference for one of the cointegration ranks, it is also possible to use the posterior probabilities as weights in a forecasting exercise.

6.9 Concluding Remarks

The chapter discusses a Bayesian modelling framework for the analysis of cointegration models. This framework is based on a specification of an unrestricted error correction model which contains a parameter reflecting cointegration, *i.e.* it is equal to zero when

cointegration occurs. Posteriors for parameters in the cointegration model are then proportional to conditional posteriors of the parameters in the unrestricted error correction model given that the parameter reflecting cointegration is equal to zero. This is identical to the classical analysis where the likelihood of the cointegration model is proportional to the conditional likelihood of the unrestricted error correction model given that the parameter reflecting cointegration is equal to zero. The difference is though that we can only construct these conditional posteriors given the parameters reflecting cointegration numerically as analytical expressions do not exist. A Metropolis-Hastings sampler is used to calculate the posteriors of the cointegration model. The prior framework, which is proposed, allows for a full Bayesian treatment of all aspects of a cointegration model under various prior specifications, like diffuse and natural conjugate. The prior is specified on the autoregressive parameter matrices of the vector autoregressive model. This prior implies the prior for the unrestricted error correction model and the prior for the cointegration models. Therefore, one specification of the prior for the parameters of the VAR suffices as it implies the functional specification of the priors for the cointegration models. For posterior odds analysis it is possible to give prior probabilities to models with different number of cointegration relations. These prior probabilities can be combined with Bayes factors to determine posterior probabilities. The proposed Bayes factors under a diffuse prior specification are related to the posterior information criterium of Phillips and Ploberger (1994, 1996). We also compare different cointegration models using the Bayesian analog of a Lagrange Multiplier statistic.

The Bayesian cointegration analysis framework in this chapter can be extended for cointegration analysis in more complicated models. We can for instance allow for structural breaks or consider non-linear cointegration models, like Markov switching cointegration and threshold cointegration. Also we can change the assumption of normal distributed errors and extend the analysis to t-distributed errors or vector moving average errors. In the next chapter we use this Bayesian cointegration framework to analyse multivariate Markov trend models.

6.A Derivation of Prior Distributions

Diffuse Prior Specification

The priors for the parameters α, λ and β_2 follow from the prior (6.19) and the Jacobian of the transformations of $(\mathbf{\Pi}_{11}, \mathbf{\Pi}_{12}, \mathbf{\Pi}_{21}, \mathbf{\Pi}_{22})$ to $(\alpha,\ \lambda, \beta_2)$. As $\alpha = (\mathbf{\Pi}_{11}, \mathbf{\Pi}_{12})$ the prior for α given Σ reads

$$p_{uec}(\boldsymbol{\alpha}|\Sigma) = p_{lec}(\mathbf{\Pi}_{11}, \mathbf{\Pi}_{12}|\Sigma)|_{(\mathbf{\Pi}_{11},\mathbf{\Pi}_{12})=\boldsymbol{\alpha}} \propto |\Sigma|^{-\frac{1}{2}r} |Y'_{1,-1} M_{Y_{2,-1}} Y_{1,-1}|^{\frac{1}{2}n}.$$

Since $(\mathbf{\Pi}_{21}, \mathbf{\Pi}_{22})$ can only be transformed to (λ, β_2) when $\boldsymbol{\alpha} = (\mathbf{\Pi}_{11}, \mathbf{\Pi}_{12})$ is known, we consider a prior for λ and β_2 given Σ and $\boldsymbol{\alpha}$. This prior is proportional to the product of the prior $p_{lec}(\mathbf{\Pi}_{21}, \mathbf{\Pi}_{22}|\mathbf{\Pi}_{11}, \mathbf{\Pi}_{12}, \Sigma)$ given in (6.19) and the Jacobian. Since $(\mathbf{\Pi}_{21}, \mathbf{\Pi}_{22}) = \lambda(\mathbf{0}\ \ \mathbf{I}_{n-r}) - \beta_2\boldsymbol{\alpha}$ the Jacobian $|J(\boldsymbol{\alpha}, \lambda, \beta_2)| = |(J_1\ J_2)|$ results from

$$\begin{aligned} J_1 = \frac{\partial(\mathrm{vec}(\mathbf{\Pi}_{21}, \mathbf{\Pi}_{22}))}{\partial(\mathrm{vec}(\beta_2))'}\bigg|\boldsymbol{\alpha} &= -(\boldsymbol{\alpha}' \otimes \mathbf{I}_{n-r}), \\ J_2 = \frac{\partial(\mathrm{vec}(\mathbf{\Pi}_{21}\ \mathbf{\Pi}_{22}))}{\partial(\mathrm{vec}(\lambda))'}\bigg|\boldsymbol{\alpha} &= \Big(\begin{pmatrix} \mathbf{0} \\ \mathbf{I}_{n-r} \end{pmatrix} \otimes \mathbf{I}_{n-r}\Big). \end{aligned}$$

Therefore, the priors for λ and β_2 read

$$\begin{aligned} &p_{uec}(\lambda, \beta_2|\Sigma, \boldsymbol{\alpha}) \\ &\quad\propto p_{lec}(\mathbf{\Pi}_{21}, \mathbf{\Pi}_{22}|\mathbf{\Pi}_{11}, \mathbf{\Pi}_{12}, \Sigma)\, |J(\boldsymbol{\alpha}, \lambda, \beta_2)| \\ &\quad\propto |J(\boldsymbol{\alpha}, \lambda, \beta_2)'\, p_{lec}(\mathbf{\Pi}_{21}, \mathbf{\Pi}_{22}|\mathbf{\Pi}_{11}, \mathbf{\Pi}_{12}, \Sigma)\, J(\boldsymbol{\alpha}, \lambda, \beta_2)|^{\frac{1}{2}} \end{aligned}$$

and since $|\Sigma|^{-\frac{1}{2}(n-r)} |Y'_{2,-1}Y_{2,-1}|^{\frac{1}{2}n} = |\Sigma^{-\frac{1}{2}} \otimes Y'_{2,-1}Y_{2,-1}|^{\frac{1}{2}}$

$$\begin{aligned} &\propto |J(\boldsymbol{\alpha}, \lambda, \beta_2)'\, (\Sigma^{-1} \otimes Y'_{2,-1}Y_{2,-1})\, J(\boldsymbol{\alpha}, \lambda, \beta_2)|^{\frac{1}{2}} \\ &\propto \left| \begin{pmatrix} (\boldsymbol{\alpha}\Sigma^{-1}\boldsymbol{\alpha}') \otimes (Y'_{2,-1}Y_{2,-1}) & \Big(\begin{pmatrix} \mathbf{0} \\ \mathbf{I}_{n-r} \end{pmatrix}'\Sigma^{-1}\boldsymbol{\alpha}'\Big) \otimes (Y'_{2,-1}Y_{2,-1}) \\ \Big(\boldsymbol{\alpha}\Sigma^{-1}\begin{pmatrix} \mathbf{0} \\ \mathbf{I}_{n-r} \end{pmatrix}\Big) \otimes (Y'_{2,-1}Y_{2,-1}) & \Big(\begin{pmatrix} \mathbf{0} \\ \mathbf{I}_{n-r} \end{pmatrix}'\Sigma^{-1}\begin{pmatrix} \mathbf{0} \\ \mathbf{I}_{n-r} \end{pmatrix}\Big) \otimes (Y'_{2,-1}Y_{2,-1}) \end{pmatrix} \right|^{\frac{1}{2}} \\ &\propto |(\boldsymbol{\alpha}\Sigma^{-1}\boldsymbol{\alpha}') \otimes (Y'_{2,-1}Y_{2,-1})|^{\frac{1}{2}} \\ &\qquad \Big|\Big(\begin{pmatrix} \mathbf{0} \\ \mathbf{I}_{n-r} \end{pmatrix}'(\Sigma^{-1} - \Sigma^{-1}\boldsymbol{\alpha}'(\boldsymbol{\alpha}\Sigma^{-1}\boldsymbol{\alpha}')^{-1}\boldsymbol{\alpha}\Sigma^{-1})\begin{pmatrix} \mathbf{0} \\ \mathbf{I}_{n-r} \end{pmatrix}\Big) \otimes (Y'_{2,-1}Y_{2,-1})\Big|^{\frac{1}{2}} \\ &\propto |\boldsymbol{\alpha}\Sigma^{-1}\boldsymbol{\alpha}'|^{\frac{1}{2}(n-r)} \Big|\begin{pmatrix} \mathbf{0} \\ \mathbf{I}_{n-r} \end{pmatrix}'\boldsymbol{\alpha}'_\perp(\boldsymbol{\alpha}_\perp\Sigma\boldsymbol{\alpha}'_\perp)^{-1}\boldsymbol{\alpha}_\perp\begin{pmatrix} \mathbf{0} \\ \mathbf{I}_{n-r} \end{pmatrix}\Big|^{\frac{1}{2}(n-r)} |Y'_{2,-1}Y_{2,-1}|^{\frac{1}{2}n} \\ &\propto |\boldsymbol{\alpha}\Sigma^{-1}\boldsymbol{\alpha}'|^{\frac{1}{2}(n-r)} |\boldsymbol{\alpha}_\perp\Sigma\boldsymbol{\alpha}'_\perp|^{-\frac{1}{2}(n-r)} |Y'_{2,-1}Y_{2,-1}|^{\frac{1}{2}n} \end{aligned}$$

so that

$$\begin{aligned} p_{uec}(\lambda|\Sigma, \boldsymbol{\alpha}) &\propto |\boldsymbol{\alpha}_\perp\Sigma^{-1}\boldsymbol{\alpha}'_\perp|^{-\frac{1}{2}(n-r)} |Y'_{2,-1}Y_{2,-1}|^{\frac{1}{2}(n-r)}, \\ p_{uec}(\beta_2|\Sigma, \boldsymbol{\alpha}, \lambda) &\propto |\boldsymbol{\alpha}\Sigma^{-1}\boldsymbol{\alpha}'|^{\frac{1}{2}(n-r)} |Y'_{2,-1}Y_{2,-1}|^{\frac{1}{2}r}, \end{aligned}$$

where $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1\ \boldsymbol{\alpha}_2)$, $\boldsymbol{\alpha}_\perp = (-\boldsymbol{\alpha}'_2(\boldsymbol{\alpha}_1^{-1})'\ \mathbf{I}_{n-r})$ and $\boldsymbol{\alpha}_1$ is an $(r \times r)$ and $\boldsymbol{\alpha}_2$ an $(r \times (n-r))$ matrix and we use that $(\Sigma^{-1} - \Sigma^{-1}\boldsymbol{\alpha}'(\boldsymbol{\alpha}\Sigma^{-1}\boldsymbol{\alpha}')^{-1}\boldsymbol{\alpha}\Sigma^{-1})$ equals $\boldsymbol{\alpha}'_\perp(\boldsymbol{\alpha}_\perp\Sigma\boldsymbol{\alpha}'_\perp)^{-1}\boldsymbol{\alpha}_\perp$.

Natural Conjugate Prior Specification

The natural conjugate prior for $\mathbf{\Pi}$ given Σ (6.23) implies the following conditional priors for the submatrices of $\mathbf{\Pi}$ defined in (6.9)

$$\begin{aligned} p_{lec}(\mathbf{\Pi}_{11}, \mathbf{\Pi}_{12}|\Sigma) &\propto |\Sigma|^{-\frac{1}{2}r}|A_{11.2}|^{\frac{1}{2}n}\exp(-\frac{1}{2}\mathrm{tr}(\Sigma^{-1}((\mathbf{\Pi}_{11}\ \mathbf{\Pi}_{12}) - (P_{11}\ P_{12}))' \\ &\quad A_{11.2}((\mathbf{\Pi}_{11}\ \mathbf{\Pi}_{12}) - (P_{11}\ P_{12})))), \\ p_{lec}(\mathbf{\Pi}_{21}, \mathbf{\Pi}_{22}|\mathbf{\Pi}_{11}, \mathbf{\Pi}_{12}, \Sigma) &\propto |\Sigma|^{-\frac{1}{2}(n-r)}|A_{22}|^{\frac{1}{2}n}\exp(-\frac{1}{2}\mathrm{tr}(\Sigma^{-1}((\mathbf{\Pi}_{21}\ \mathbf{\Pi}_{22}) - (Q_{21}\ Q_{22})' \\ &\quad A_{22}((\mathbf{\Pi}_{21}\ \mathbf{\Pi}_{22}) - (Q_{21}\ Q_{22})))), \end{aligned}$$

where $A_{11.2} = A_{11} - A_{12}A_{22}^{-1}A_{21}$, $(Q_{21}\ Q_{22}) = (P_{21}\ P_{22}) - A_{22}^{-1}A_{21}((\mathbf{\Pi}_{11}\ \mathbf{\Pi}_{12}) - (P_{11}\ P_{12}))$. Since $\boldsymbol{\alpha} = (\mathbf{\Pi}_{11}\ \mathbf{\Pi}_{12})$ the conditional prior for $\boldsymbol{\alpha}$ equals the conditional prior of $(\mathbf{\Pi}_{11}, \mathbf{\Pi}_{12})$ evaluated in $\boldsymbol{\alpha}$: $p_{uec}(\boldsymbol{\alpha}|\Sigma) = p_{lec}(\mathbf{\Pi}_{11}, \mathbf{\Pi}_{12}|\Sigma)|_{(\mathbf{\Pi}_{11}\ \mathbf{\Pi}_{12})=\boldsymbol{\alpha}}$. To derive the conditional priors of λ and β_2 we need the Jacobian $|J(\boldsymbol{\alpha}, \lambda, \beta_2)|$ of the transformation of $(\mathbf{\Pi}_{21}, \mathbf{\Pi}_{22})$ to (λ, β_2) derived above. The following two decompositions

$$\begin{aligned} |\Sigma|^{-\frac{1}{2}(n-r)}|A_{22}|^{\frac{1}{2}n}|J| &= |J(\boldsymbol{\alpha}, \lambda, \beta_2)'\,(\Sigma^{-1} \otimes A_{22})J(\boldsymbol{\alpha}, \lambda, \beta_2)|^{\frac{1}{2}} \\ &= |\boldsymbol{\alpha}\Sigma^{-1}\boldsymbol{\alpha}'|^{\frac{1}{2}(n-r)}|\boldsymbol{\alpha}_\perp\Sigma\boldsymbol{\alpha}_\perp'|^{-\frac{1}{2}(n-r)}|A_{22}|^{\frac{1}{2}n} \end{aligned}$$

and

$$\begin{aligned} &\exp(-\frac{1}{2}\mathrm{tr}(\Sigma^{-1}((\mathbf{\Pi}_{21}\ \mathbf{\Pi}_{22}) - (Q_{21}\ Q_{22}))'A_{22}((\mathbf{\Pi}_{21}\ \mathbf{\Pi}_{22}) - (Q_{21}\ Q_{22})))) = \\ &\quad \exp(-\frac{1}{2}(\mathrm{tr}((\boldsymbol{\alpha}_\perp\Sigma\boldsymbol{\alpha}_\perp')^{-1}(\lambda - l)'A_{22}(\lambda - l)) + \mathrm{tr}(A_{22}(\beta_2 - b_2)\boldsymbol{\alpha}\Sigma^{-1}\boldsymbol{\alpha}'(\beta_2 - b_2)'))), \end{aligned}$$

where $l = (Q_{21}\ Q_{22})\boldsymbol{\alpha}_\perp'$ and $b_2 = -(Q_{21}\ (Q_{22} - \lambda))\Sigma^{-1}\boldsymbol{\alpha}'(\boldsymbol{\alpha}\Sigma^{-1}\boldsymbol{\alpha}')^{-1}$ lead to the conditional priors $p_{uec}(\lambda|\Sigma, \boldsymbol{\alpha})$ and $p_{uec}(\beta_2|\Sigma, \boldsymbol{\alpha}, \lambda)$ in (6.25).

6.B Derivation of Posterior Distributions

Posterior under Diffuse Prior Specification

The joint posterior $p_{uecm}(\Sigma, \boldsymbol{\alpha}, \lambda, \beta_2|Y)$ is proportional to the product of the prior (6.20) and likelihood (6.10). Due to the structure of the unrestricted error correction model (6.8) it only possible to decompose the joint posterior in marginal and conditional posteriors, $p_{uecm}(\Sigma|Y)\, p_{uecm}(\boldsymbol{\alpha}|\Sigma, Y)\, p_{uecm}(\lambda|\Sigma, \boldsymbol{\alpha}, Y)\, p_{uecm}(\beta_2|\Sigma, \boldsymbol{\alpha}, \lambda, Y)$, which is the product of an inverted Wishart and three matrix normal densities. The marginal/conditional posteriors follow from the following decomposition of the elements in the trace operator from the likelihood (6.8)

$$
\begin{aligned}
\operatorname{tr}(\Sigma^{-1}\varepsilon'\varepsilon) &= \operatorname{tr}(\Sigma^{-1}(\Delta Y - Y_{1,-1}\boldsymbol{\alpha} + Y_{2,-1}\beta_2\boldsymbol{\alpha} - Y_{2,-1}(\mathbf{0}\ \lambda))' \\
&\qquad (\Delta Y - Y_{1,-1}\boldsymbol{\alpha} + Y_{2,-1}\beta_2\boldsymbol{\alpha} - Y_{2,-1}(\mathbf{0}\ \lambda))) \\
&= \operatorname{tr}(\Sigma^{-1}(\Delta Y' M_{Y_{-1}} \Delta Y + (\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}})' Y_{1,-1}' M_{Y_{2,-1}} Y_{1,-1}(\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}}) \\
&\qquad + (\beta_2\boldsymbol{\alpha} - (\mathbf{0}\ \lambda) - \hat{\boldsymbol{\Pi}}_2)' Y_{2,-1}' Y_{2,-1}(\beta_2\boldsymbol{\alpha} - (\mathbf{0}\ \lambda) - \hat{\boldsymbol{\Pi}}_2)) \\
&= \operatorname{tr}(\Sigma^{-1}(\Delta Y' M_{Y_{-1}} \Delta Y + (\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}})' Y_{1,-1}' M_{Y_{2,-1}} Y_{1,-1}(\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}})) \\
&\qquad + \operatorname{tr}(Y_{2,-1}' Y_{2,-1}(\beta_2\boldsymbol{\alpha} - (\mathbf{0}\ \lambda) - \hat{\boldsymbol{\Pi}}_2)\Sigma^{-1}(\beta_2\boldsymbol{\alpha} - (\mathbf{0}\ \lambda) - \hat{\boldsymbol{\Pi}}_2)') \\
&= \operatorname{tr}(\Sigma^{-1}(\Delta Y' M_{Y_{-1}} \Delta Y + (\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}})' Y_{1,-1}' M_{Y_{2,-1}} Y_{1,-1}(\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}})) \\
&\qquad + \operatorname{tr}(Y_{2,-1}' Y_{2,-1}((\beta_2 - \hat{\beta}_2)\boldsymbol{\alpha}\Sigma^{-1}\boldsymbol{\alpha}'(\beta_2 - \hat{\beta}_2)' + \\
&\qquad\quad + ((\mathbf{0}\ \lambda) - \hat{\boldsymbol{\Pi}}_2)(\Sigma^{-1} - \Sigma^{-1}\boldsymbol{\alpha}(\boldsymbol{\alpha}'\Sigma^{-1}\boldsymbol{\alpha})^{-1}\boldsymbol{\alpha}'\Sigma^{-1})((\mathbf{0}\ \lambda) - \hat{\boldsymbol{\Pi}}_2)') \\
&= \operatorname{tr}(\Sigma^{-1}(\Delta Y' M_{Y_{-1}} \Delta Y + (\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}})' Y_{1,-1}' M_{Y_{2,-1}} Y_{1,-1}(\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}})) \\
&\qquad + \operatorname{tr}(Y_{2,-1}' Y_{2,-1}(\beta_2 - \hat{\beta}_2)\boldsymbol{\alpha}\Sigma^{-1}\boldsymbol{\alpha}'(\beta_2 - \hat{\beta}_2)') \\
&\qquad\quad + \operatorname{tr}(Y_{2,-1}' Y_{2,-1}((\mathbf{0}\ \lambda) - \hat{\boldsymbol{\Pi}}_2)\boldsymbol{\alpha}_\perp'(\boldsymbol{\alpha}_\perp\Sigma\boldsymbol{\alpha}_\perp')^{-1}\boldsymbol{\alpha}_\perp((\mathbf{0}\ \lambda) - \hat{\boldsymbol{\Pi}}_2)') \\
&= \operatorname{tr}(\Sigma^{-1}(\Delta Y' M_{Y_{-1}} \Delta Y + (\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}})' Y_{1,-1}' M_{Y_{2,-1}} Y_{1,-1}(\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}})) \\
&\qquad + \operatorname{tr}(Y_{2,-1}' Y_{2,-1}(\beta_2 - \hat{\beta}_2)\boldsymbol{\alpha}\Sigma^{-1}\boldsymbol{\alpha}'(\beta_2 - \hat{\beta}_2)') \\
&\qquad\quad + \operatorname{tr}((\boldsymbol{\alpha}_\perp\Sigma\boldsymbol{\alpha}_\perp')^{-1}\boldsymbol{\alpha}_\perp((\mathbf{0}\ \lambda) - \hat{\boldsymbol{\Pi}}_2)' Y_{2,-1}' Y_{2,-1}((\mathbf{0}\ \lambda) - \hat{\boldsymbol{\Pi}}_2)\boldsymbol{\alpha}_\perp') \\
&= \operatorname{tr}(\Sigma^{-1}(\Delta Y' M_{Y_{-1}} \Delta Y + (\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}})' Y_{1,-1}' M_{Y_{2,-1}} Y_{1,-1}(\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}})) \\
&\qquad + \operatorname{tr}(Y_{2,-1}' Y_{2,-1}(\beta_2 - \hat{\beta}_2)\boldsymbol{\alpha}\Sigma^{-1}\boldsymbol{\alpha}'(\beta_2 - \hat{\beta}_2)') \\
&\qquad\quad + \operatorname{tr}((\boldsymbol{\alpha}_\perp\Sigma\boldsymbol{\alpha}_\perp')^{-1}(\lambda - \hat{\lambda})' Y_{2,-1}' Y_{2,-1}(\lambda - \hat{\lambda})),
\end{aligned}
$$

where we have used that $\boldsymbol{\alpha}_\perp'(\boldsymbol{\alpha}_\perp\Sigma\boldsymbol{\alpha}_\perp')^{-1}\boldsymbol{\alpha}_\perp = (\Sigma^{-1} - \Sigma^{-1}\boldsymbol{\alpha}'(\boldsymbol{\alpha}\Sigma^{-1}\boldsymbol{\alpha}')^{-1}\boldsymbol{\alpha}\Sigma^{-1})$ and

$$
\begin{aligned}
\hat{\boldsymbol{\Pi}}_2 &= (Y_{2,-1}' Y_{2,-1})^{-1} Y_{2,-1}'(\Delta Y - Y_{1,-1}\boldsymbol{\alpha}), \\
\hat{\boldsymbol{\alpha}} &= (Y_{1,-1}' M_{Y_{2,-1}} Y_{1,-1})^{-1} Y_{1,-1}' M_{Y_{2,-1}} \Delta Y, \\
\hat{\lambda} &= (Y_{2,-1}' Y_{2,-1})^{-1} Y_{2,-1}' \Delta Y, \boldsymbol{\alpha}_\perp' \\
\hat{\beta}_2 &= -(Y_{2,-1}' Y_{2,-1})^{-1} Y_{2,-1}'(\Delta Y - Y_{1,-1}\boldsymbol{\alpha} - Y_{2,-1}(\mathbf{0}\ \lambda))\Sigma^{-1}\boldsymbol{\alpha}'(\boldsymbol{\alpha}\Sigma^{-1}\boldsymbol{\alpha}')^{-1}.
\end{aligned}
$$

Combining this decomposition with the prior specification (6.20) provides the conditional/marginal posterior distributions in (6.28).

Posterior under Natural Conjugate Prior Specification

The conditional posteriors of $\boldsymbol{\alpha}$, λ and β_2 (6.32) under a natural conjugate prior specification for Σ and $\mathbf{\Pi}$ (6.23) follow directly from the joint posterior of Σ and $\mathbf{\Pi}$, which reads

$$\begin{aligned} p_{lec}(\Sigma, \mathbf{\Pi}|Y) &\propto |S|^{\frac{1}{2}h}|A|^{\frac{1}{2}n}|\Sigma|^{-\frac{1}{2}(T+h+n)} \exp(-\frac{1}{2}\mathrm{tr}(\Sigma^{-1}(S + \\ &\quad (\mathbf{\Pi}-P)'A(\mathbf{\Pi}-P) + (\Delta Y - Y_{-1}\mathbf{\Pi})'(\Delta Y - Y_{-1}\mathbf{\Pi})))) \\ &\propto |S|^{\frac{1}{2}h}|A|^{\frac{1}{2}n}|\Sigma|^{-\frac{1}{2}(T+h+n+1)} \exp(-\frac{1}{2}\mathrm{tr}(\Sigma^{-1}(S + \Delta Y'\Delta Y + P'AP \\ &\quad -\tilde{\mathbf{\Pi}}'(A + Y_{-1}'Y_{-1})\tilde{\mathbf{\Pi}} + (\mathbf{\Pi}-\tilde{\mathbf{\Pi}})'(A + Y_{-1}'Y_{-1})(\mathbf{\Pi}-\tilde{\mathbf{\Pi}})))), \end{aligned}$$

where

$$\begin{aligned} \hat{\mathbf{\Pi}} &= (Y_{-1}'Y_{-1})^{-1}Y_{-1}'\Delta Y, \\ (A + Y_{-1}'Y_{-1}) &= \begin{pmatrix} (A + Y_{-1}'Y_{-1})_{11} & (A + Y_{-1}'Y_{-1})_{12} \\ (A + Y_{-1}'Y_{-1})_{21} & (A + Y_{-1}'Y_{-1})_{22} \end{pmatrix}, \\ \tilde{\mathbf{\Pi}} &= \begin{pmatrix} \tilde{\mathbf{\Pi}}_{11} & \tilde{\mathbf{\Pi}}_{12} \\ \tilde{\mathbf{\Pi}}_{21} & \tilde{\mathbf{\Pi}}_{22} \end{pmatrix} = (A + Y_{-1}'Y_{-1})^{-1}(AP + Y_{-1}'Y_{-1}\hat{\mathbf{\Pi}}). \end{aligned}$$

6.C Derivations for the Invariant Specification

Jacobian for the Invariant Specification

For the derivation of the Jacobian transformation, it is convenient to split up the transformation from $\mathbf{\Pi}$ to $\boldsymbol{\alpha}$, λ^* and β_2 in two steps, firstly from $\mathbf{\Pi}$ to $(\boldsymbol{\alpha}_1, \vartheta_2, \lambda^*, \beta_2)$, where $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1 \ \boldsymbol{\alpha}_2)$ and $\vartheta_2 = -\boldsymbol{\alpha}_1^{-1}\boldsymbol{\alpha}_2$, and secondly from $(\boldsymbol{\alpha}_1, \vartheta_2, \lambda^*, \beta_2)$ to $(\boldsymbol{\alpha}, \lambda^*, \beta_2)$. In the following we construct the Jacobians for the two transformations. We can denote $\mathbf{\Pi}$ as a function of $(\boldsymbol{\alpha}_1, \vartheta_2, \lambda^*, \beta_2)$

$$\mathbf{\Pi} = \begin{pmatrix} \beta & \beta_\perp \end{pmatrix} \begin{pmatrix} \boldsymbol{\alpha}_1 & \mathbf{0} \\ \mathbf{0} & \lambda^* \end{pmatrix} \begin{pmatrix} \vartheta \\ \vartheta_\perp \end{pmatrix}$$

where $\vartheta = (\mathbf{I}_r \ -\vartheta_2)$ with $\boldsymbol{\alpha} = \boldsymbol{\alpha}_1\vartheta$, $\beta = (\mathbf{I}_r \ -\beta_2')'$, $\vartheta_\perp = (\mathbf{I}_{n-r} + \vartheta_2'\vartheta_2)^{-\frac{1}{2}}(\vartheta_2' \ \mathbf{I}_{n-r})$ and $\beta_\perp = (\beta_2 \ \mathbf{I}_{n-r})'(\mathbf{I}_{n-r} + \beta_2\beta_2')^{-\frac{1}{2}}$ so that $\vartheta_\perp\vartheta_\perp' = \mathbf{I}_{n-r}$ and $\beta_\perp'\beta_\perp = \mathbf{I}_{n-r}$.[3] The derivatives of $\mathbf{\Pi}$ with respect to $\boldsymbol{\alpha}_1$, ϑ_2, λ^* and β_2 read

$$\begin{aligned}
J_1 &= \frac{\partial \operatorname{vec}(\mathbf{\Pi})}{\partial(\operatorname{vec}(\boldsymbol{\alpha}_1))'} = (\vartheta' \otimes \beta) \\
J_2 &= \frac{\partial \operatorname{vec}(\mathbf{\Pi})}{\partial(\operatorname{vec}(\vartheta_2))'} = -\Big(\begin{pmatrix} \mathbf{0} \\ \mathbf{I}_{n-r} \end{pmatrix} \otimes \beta\boldsymbol{\alpha}_1\Big) + (\mathbf{I}_n \otimes \beta_\perp\lambda^*)\frac{\partial \operatorname{vec}(\vartheta_\perp)}{\partial(\operatorname{vec}(\vartheta_2))'} \\
J_3 &= \frac{\partial \operatorname{vec}(\mathbf{\Pi})}{\partial(\operatorname{vec}(\lambda^*))'} = (\vartheta_\perp' \otimes \beta_\perp) \\
J_4 &= \frac{\partial \operatorname{vec}(\mathbf{\Pi})}{\partial(\operatorname{vec}(\beta_2))'} = -\Big(\vartheta'\boldsymbol{\alpha}_1' \otimes \begin{pmatrix} \mathbf{0} \\ \mathbf{I}_{n-r} \end{pmatrix}\Big) + (\vartheta_\perp'\lambda^{*\prime} \otimes \mathbf{I}_n)\frac{\partial \operatorname{vec}(\beta_\perp)}{\partial(\operatorname{vec}(\beta_2))'},
\end{aligned}$$

with

$$\begin{aligned}
\frac{\partial \operatorname{vec}(\vartheta_\perp)}{\partial(\operatorname{vec}(\vartheta_2))'} &= (\mathbf{I}_n \otimes \theta_n^{-\frac{1}{2}})\frac{\partial \operatorname{vec}(\vartheta_2' \ \mathbf{I}_{n-r})}{\partial(\operatorname{vec}(\vartheta_2))'} \\
&\quad + ((\vartheta_2' \ \mathbf{I}_{n-r})' \otimes \mathbf{I}_{n-r})\frac{\partial \operatorname{vec}(\vartheta_n^{-\frac{1}{2}})}{\partial\,(\operatorname{vec}(\vartheta_n^{\frac{1}{2}}))'}\frac{\partial \operatorname{vec}(\vartheta_n^{\frac{1}{2}})}{\partial(\operatorname{vec}(\vartheta_n))'}\frac{\partial \operatorname{vec}(\vartheta_n)}{\partial(\operatorname{vec}(\vartheta_2))'} \\
\frac{\partial \operatorname{vec}(\beta_\perp)}{\partial(\operatorname{vec}(\beta_2))'} &= (\beta_n^{-\frac{1}{2}\prime} \otimes \mathbf{I}_n)\frac{\partial \operatorname{vec}((\beta_2 \ \mathbf{I}_{n-r})')}{\partial(\operatorname{vec}(\beta_2))'} \\
&\quad + (\mathbf{I}_{n-r} \otimes (\beta_2 \ \mathbf{I}_{n-r})')\frac{\partial \operatorname{vec}(\beta_n^{-\frac{1}{2}})}{\partial\,(\operatorname{vec}(\beta_n^{\frac{1}{2}}))'}\frac{\partial \operatorname{vec}(\beta_n^{\frac{1}{2}})}{\partial(\operatorname{vec}(\beta_n))'}\frac{\partial \operatorname{vec}(\beta_n)}{\partial(\operatorname{vec}(\beta_2))'}
\end{aligned}$$

[3] If M is a positive definite real symmetric matrix, then $M^{\frac{1}{2}} = C\Lambda^{\frac{1}{2}}C'$ where Λ is a diagonal matrix containing the eigenvalues of M and C contains the orthonormal eigenvectors of M and $M^{-\frac{1}{2}} = C\Lambda^{-\frac{1}{2}}C'$, see *e.g.* Johansen (1995, p. 222).

where we define for notational convenience $\vartheta_n = (\mathbf{I}_{n-r} + \vartheta_2'\vartheta_2)$ and $\beta_n = (\mathbf{I}_{n-r} + \beta_2\beta_2')$ so that $\vartheta_\perp = \vartheta_n^{-\frac{1}{2}}(\vartheta_2' \ \mathbf{I}_{n-r})$, $\beta_\perp = (\beta_2 \ \mathbf{I}_{n-r})'\beta_n^{-\frac{1}{2}}$ with $\vartheta_n^{\frac{1}{2}}\vartheta_n^{\frac{1}{2}} = \vartheta_n$, $\beta_n^{\frac{1}{2}}\beta_n^{\frac{1}{2}} = \beta_n$,

$$\begin{aligned}
\frac{\partial \operatorname{vec}(\vartheta_2' \ \mathbf{I}_{n-r})}{\partial(\operatorname{vec}(\vartheta_2))'} &= \left(\begin{pmatrix}\mathbf{I}_r \\ \mathbf{0}\end{pmatrix} \otimes \mathbf{I}_{n-r}\right)\mathbb{K}_{n-r,r} \\
\frac{\partial \operatorname{vec}(\vartheta_n^{-\frac{1}{2}})}{\partial(\operatorname{vec}(\vartheta_n^{\frac{1}{2}}))'} &= -(\vartheta_n^{-\frac{1}{2}\prime} \otimes \vartheta_n^{-\frac{1}{2}}) \\
\frac{\partial \operatorname{vec}(\vartheta_n^{\frac{1}{2}})}{\partial(\operatorname{vec}(\vartheta_n))'} &= ((\vartheta_n^{\frac{1}{2}\prime} \otimes \mathbf{I}_{n-r}) + (\mathbf{I}_{n-r} \otimes \vartheta_n^{\frac{1}{2}}))^{-1} \\
\frac{\partial \operatorname{vec}(\vartheta_n)}{\partial(\operatorname{vec}(\vartheta_2))'} &= (\mathbf{I}_{n-r} \otimes \vartheta_2') + (\vartheta_2' \otimes \mathbf{I}_{n-r})\mathbb{K}_{n-r,r}
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial \operatorname{vec}((\beta_2 \ \mathbf{I}_{n-r})')}{\partial(\operatorname{vec}(\beta_2))'} &= \left(\mathbf{I}_{n-r} \otimes \begin{pmatrix}\mathbf{I}_r \\ \mathbf{0}\end{pmatrix}\right)\mathbb{K}_{n-r,r} \\
\frac{\partial \operatorname{vec}(\beta_n^{-\frac{1}{2}})}{\partial(\operatorname{vec}(\beta_n^{\frac{1}{2}}))'} &= -(\beta_n^{-\frac{1}{2}\prime} \otimes \beta_n^{-\frac{1}{2}}) \\
\frac{\partial \operatorname{vec}(\beta_n^{\frac{1}{2}})}{\partial(\operatorname{vec}(\beta_n))'} &= ((\beta_n^{\frac{1}{2}\prime} \otimes \mathbf{I}_{n-r}) + (\mathbf{I}_{n-r} \otimes \beta_n^{\frac{1}{2}}))^{-1} \\
\frac{\partial \operatorname{vec}(\beta_n)}{\partial(\operatorname{vec}(\beta_2))'} &= (\beta_2 \otimes \mathbf{I}_{n-r}) + (\mathbf{I}_{n-r} \otimes \beta_2)\mathbb{K}_{n-r,r}
\end{aligned}$$

where $\mathbb{K}_{i,j}$ are so-called commutation matrices. For any $(i \times j)$ matrix W, $\operatorname{vec}(W) = \mathbb{K}_{i,j}\operatorname{vec}(W')$, $\operatorname{vec}(W') = \mathbb{K}_{j,i}\operatorname{vec}(W)$, and $\mathbb{K}_{i,j}' = \mathbb{K}_{j,i}$, see Lütkepohl (1993, p. 466). The Jacobian of the transformation from $\mathbf{\Pi}$ to $(\boldsymbol{\alpha}_1, \vartheta_2, \lambda^*, \beta_2)$ becomes

$$\left|\frac{\partial \operatorname{vec}(\mathbf{\Pi})}{\partial(\operatorname{vec}(\boldsymbol{\alpha}_1)' \ \operatorname{vec}(\vartheta_2)' \ \operatorname{vec}(\lambda^*)' \ \operatorname{vec}(\beta_2)')}\right| = |(J_1 \ J_2 \ J_3 \ J_4)|.$$

Since $\vartheta_2 = -\boldsymbol{\alpha}_1^{-1}\boldsymbol{\alpha}_2$ the derivatives of $(\boldsymbol{\alpha}_1, \vartheta_2, \lambda^*, \beta_2)$ with respect to $\boldsymbol{\alpha}_1$, $\boldsymbol{\alpha}_2$, λ^* and β_2 are respectively

$$\begin{aligned}
G_1 &= \frac{\partial(\operatorname{vec}(\boldsymbol{\alpha}_1)' \ \operatorname{vec}(\vartheta_2)' \ \operatorname{vec}(\lambda^*)' \ \operatorname{vec}(\beta_2)')'}{\partial(\operatorname{vec}(\boldsymbol{\alpha}_1))'} = \begin{pmatrix} \mathbf{I}_r \otimes \mathbf{I}_r \\ (\boldsymbol{\alpha}_1^{-1}\boldsymbol{\alpha}_2)' \otimes \boldsymbol{\alpha}_1^{-1} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix} \\
G_2 &= \frac{\partial(\operatorname{vec}(\boldsymbol{\alpha}_1)' \ \operatorname{vec}(\vartheta_2)' \ \operatorname{vec}(\lambda^*)' \ \operatorname{vec}(\beta_2)')'}{\partial(\operatorname{vec}(\boldsymbol{\alpha}_2))'} = \begin{pmatrix} \mathbf{0} \\ -\mathbf{I}_{n-r} \otimes \boldsymbol{\alpha}_1^{-1} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}
\end{aligned}$$

$$G_3 = \frac{\partial(\mathrm{vec}(\boldsymbol{\alpha}_1)'\ \mathrm{vec}(\vartheta_2)'\ \mathrm{vec}(\lambda^*)'\ \mathrm{vec}(\beta_2)')'}{\partial(\mathrm{vec}(\lambda^*))'} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{I}_{n-r} \otimes \mathbf{I}_{n-r} \\ \mathbf{0} \end{pmatrix}$$

$$G_4 = \frac{\partial(\mathrm{vec}(\boldsymbol{\alpha}_1)'\ \mathrm{vec}(\vartheta_2)'\ \mathrm{vec}(\lambda^*)'\ \mathrm{vec}(\beta_2)')'}{\partial(\mathrm{vec}(\beta_2))'} = \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{I}_r \otimes \mathbf{I}_{n-r} \end{pmatrix}$$

The Jacobians of the two transformations determine the Jacobian of the total transformation from $\boldsymbol{\Pi}$ to $(\boldsymbol{\alpha}, \lambda^*, \beta_2)$

$$\begin{aligned} &|J(\boldsymbol{\alpha}, \lambda^*, \beta_2)| \\ &\quad = \left| \frac{\partial\, \mathrm{vec}(\boldsymbol{\Pi})}{\partial(\mathrm{vec}(\boldsymbol{\alpha})'\ \mathrm{vec}(\lambda^*)'\ \mathrm{vec}(\beta_2)')} \right| \\ &\quad = \left| \frac{\partial\, \mathrm{vec}(\boldsymbol{\Pi})}{\partial(\mathrm{vec}(\boldsymbol{\alpha}_1)'\ \mathrm{vec}(\vartheta_2)'\ \mathrm{vec}(\lambda^*)'\ \mathrm{vec}(\beta_2)')} \right| \left| \frac{\partial(\mathrm{vec}(\boldsymbol{\alpha}_1)'\ \mathrm{vec}(\vartheta_2)'\ \mathrm{vec}(\lambda^*)'\ \mathrm{vec}(\beta_2)')'}{\partial(\mathrm{vec}(\boldsymbol{\alpha})'\ \mathrm{vec}(\lambda^*)'\ \mathrm{vec}(\beta_2)')'} \right| \\ &\quad = |(J_1\ J_2\ J_3\ J_4)|\ |(G_1\ G_2\ G_3\ G_4)|. \end{aligned}$$

Straightforward algebra shows that the Jacobian evaluated in $\lambda^* = \mathbf{0}$ equals

$$|J(\boldsymbol{\alpha}, \lambda^*, \beta_2)|_{\lambda^*=\mathbf{0}}| = |(\mathbf{I}_n \otimes \beta)\ \ (\boldsymbol{\alpha}' \otimes \begin{pmatrix} \mathbf{0} \\ \mathbf{I}_{n-r} \end{pmatrix})\ \ (\boldsymbol{\alpha}_\perp' \otimes \beta_\perp)|.$$

Decomposition of the Trace of the Likelihood Function

For the invariant decomposition (6.67) the trace in the likelihood function (6.10) can be decomposed as follows

$$\begin{aligned} \mathrm{tr}(\Sigma^{-1}\varepsilon'\varepsilon) &= \mathrm{tr}(\Sigma^{-1}(\Delta Y - \Delta Y_{-1}\boldsymbol{\Pi})'(\Delta Y - \Delta Y_{-1}\boldsymbol{\Pi})) \\ &= \mathrm{tr}(\Sigma^{-1}(\Delta Y M_{Y_{-1}} \Delta Y + (\boldsymbol{\Pi} - \hat{\boldsymbol{\Pi}})' Y_{-1}' Y_{-1} (\boldsymbol{\Pi} - \hat{\boldsymbol{\Pi}}))) \\ &= \mathrm{tr}(\Sigma^{-1}(\Delta Y M_{Y_{-1}} \Delta Y + (\beta\boldsymbol{\alpha} + \beta_\perp \lambda^* \boldsymbol{\alpha}_\perp - \hat{\boldsymbol{\Pi}})' Y_{-1}' Y_{-1} (\beta\boldsymbol{\alpha} + \beta_\perp \lambda^* \boldsymbol{\alpha}_\perp - \hat{\boldsymbol{\Pi}}))) \\ &= \mathrm{tr}(\Sigma^{-1}(\Delta Y M_{Y_{-1}} \Delta Y + (\beta\boldsymbol{\alpha} - \hat{\boldsymbol{\Pi}})' Y_{-1}' Y_{-1} (\beta\boldsymbol{\alpha} - \hat{\boldsymbol{\Pi}}) \\ &\qquad + (\beta_\perp \lambda^* \boldsymbol{\alpha}_\perp - \hat{\boldsymbol{\Pi}})' Y_{-1}' Y_{-1} (\beta_\perp \lambda^* \boldsymbol{\alpha}_\perp - \hat{\boldsymbol{\Pi}}) - \hat{\boldsymbol{\Pi}}' Y_{-1}' Y_{-1} \hat{\boldsymbol{\Pi}})) \\ &= \mathrm{tr}(\Sigma^{-1}(\Delta Y M_{Y_{-1}} \Delta Y) + \Sigma^{-1}(\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}})' \beta' Y_{-1}' Y_{-1} \beta (\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}})) \\ &\qquad - \Sigma^{-1} \hat{\boldsymbol{\alpha}}' (\beta' Y_{-1}' Y_{-1} \beta) \hat{\boldsymbol{\alpha}} + (\boldsymbol{\alpha}_\perp \Sigma^{-1} \boldsymbol{\alpha}_\perp')(\lambda^* - \hat{\lambda}^*)' \beta_\perp' Y_{-1}' Y_{-1} \beta_\perp (\lambda^* - \hat{\lambda}^*) \\ &\qquad\quad - (\boldsymbol{\alpha}_\perp \Sigma^{-1} \boldsymbol{\alpha}_\perp') \hat{\lambda}^{*\prime} \beta_\perp' Y_{-1}' Y_{-1} \beta_\perp \hat{\lambda}^* + \Sigma^{-1} \hat{\boldsymbol{\Pi}}' Y_{-1}' Y_{-1} \hat{\boldsymbol{\Pi}}) \\ &= \mathrm{tr}(\Sigma^{-1}(\Delta Y M_{Y_{-1}} \Delta Y) + \Sigma^{-1}(\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}})' \beta' Y_{-1}' Y_{-1} \beta (\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}}) \\ &\qquad + (\boldsymbol{\alpha}_\perp \Sigma^{-1} \boldsymbol{\alpha}_\perp')(\lambda^* - \hat{\lambda}^*)' \beta_\perp' Y_{-1}' Y_{-1} \beta_\perp (\lambda^* - \hat{\lambda}^*)) \end{aligned}$$

where we use that

$$\mathrm{tr}(\Sigma^{-1} \hat{\boldsymbol{\Pi}}' Y_{-1}' Y_{-1} \hat{\boldsymbol{\Pi}}) = \mathrm{tr}(\Sigma^{-1} \hat{\boldsymbol{\alpha}}' (\beta' Y_{-1}' Y_{-1} \beta) \hat{\boldsymbol{\alpha}}) + \mathrm{tr}((\boldsymbol{\alpha}_\perp \Sigma^{-1} \boldsymbol{\alpha}_\perp')(\hat{\lambda}^{*\prime} \beta_\perp' Y_{-1}' Y_{-1} \beta_\perp \hat{\lambda}^*))$$

and

$$\begin{aligned}
\hat{\mathbf{\Pi}} &= (Y'_{-1}Y_{-1})^{-1}Y'_{-1}\Delta Y \\
\hat{\boldsymbol{\alpha}} &= (\beta' Y'_{-1}Y_{-1}\beta)^{-1}\beta' Y'_{-1}\Delta Y \\
\hat{\lambda}^* &= (\beta'_\perp Y'_{-1}Y_{-1}\beta_\perp)^{-1}\beta'_\perp Y'_{-1}\Delta Y \Sigma^{-1}\boldsymbol{\alpha}'_\perp(\boldsymbol{\alpha}_\perp \Sigma^{-1}\boldsymbol{\alpha}'_\perp)^{-1}.
\end{aligned}$$

Chapter 7

Multivariate Markov Trend Model

7.1 Introduction

In Section 5.3 we have seen that cointegration implies that there exist a linear combination of univariate time series with stochastic trends, which can be described by a stationary process. Although shocks have a permanent effect on the level of the separate univariate series, they only have a temporary effect on the linear cointegration relation between the univariate series. A shock only leads to a temporary deviation from the linear relation between the series. The cointegration relations reflect the long run equilibria between series. Therefore, cointegration analysis is often used to detect long run relations (equilibria) between economic variables, see for instance King *et al.* (1991) and Hendry and Ericsson (1991).

Standard analysis of cointegration is usually performed in linear vector autoregressive models as described in the previous chapter. However, as we already have seen in Section 3.6.2 neglecting the possibility of changes in the growth rates during recessions in univariate time series can favour the presence of unit root stochastic trends in this time series. Generalising this to a multivariate setting, it is not unlikely that neglecting changes in the growth rate of series may influence the analysis of long run equilibria in time series. Changes in growth rate of series may lead to temporary or even permanent changes in the mean of the cointegration relation. This happens for instance if changes in the growth rate do not occur simultaneously for every series, which belong to the cointegration relation. Additionally, even if growth rate changes in cointegrated time series occur at the same time, it is still possible that the mean of the cointegration relation changes. For instance, when during a recession the decrease in one variable is larger than the decrease in another variable. For instance, Figure 7.1 shows a candidate cointegration relation for the logarithm of quarterly observed per capita income and consumption of the United States.[1] As candidate we have taken the difference between the two series. The figure shows that the mean of the candidate cointegration relation is not constant over time but displays a more or less changing regime pattern. This may imply that the

[1] In Section 7.6.2 we consider the two series in more detail.

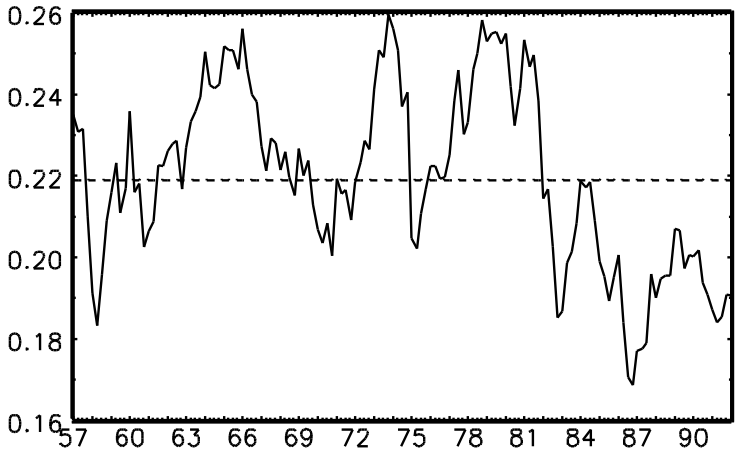

Figure 7.1. Difference between the logarithm of US per capita income and consumption, 1957.I–1992.IV.

cointegration relation is not a stationary process, but it is affected by occasionally shocks in the mean caused by regime changes, see also Krolzig (1996) for a discussion. One can interpret such a cointegration relation as a long run equilibrium between variables, which has a mean revision during or after recessions. A special case occurs if the growth rates of the series are affected by changes in regimes, but that the mean of the cointegration relation is not affected by regime changes. Dwyer and Potter (1996) refer to this situation as reduced rank Markov trend cointegration.

Neglecting permanent changes in the mean of cointegration relations, caused by regime changes, may lead to evidence in favour of no cointegration. To analyse cointegration in the presence of regime changes, we propose in this chapter a multivariate version of the Markov trend model in Chapter 3. The model allows for different Markov trends for each series, so that every series can have a different business cycle. Common business cycles in series result from common stochastic Markov trends. The deviation from the Markov trends are modelled by a vector autoregressive process, which may include stochastic trends. Hence, cointegration analysis can be performed in a regime changing environment. Markov trend cointegration occurs, if the linear cointegration relations also remove the Markov trend from the multivariate series.

The outline of this chapter is as follows. In Section 7.2 we propose the multivariate Markov trend model. We discuss the role of the stochastic trends in this model and the possibilities of common trends. Section 7.3 deals with prior specification. To obtain posterior results, we propose in Section 7.4 a Gibbs sampling algorithm to sample from the posterior distribution. The analysis of cointegration and common stochastic Markov trends is discussed in Section 7.5. In Section 7.6 we illustrate the analysis of multivariate Markov trends with some simulated series. Additionally, we analyse the role of stochastic

trends in quarterly observed real gross domestic product and consumption of the United States, on which Figure 7.1 is based. We conclude in Section 7.7.

7.2 The Model

Suppose that an n-dimensional vector of time series $\{Y_t\}_{t=1}^T$ can be decomposed as

$$Y_t = N_t + Z_t, \tag{7.1}$$

where N_t is a multivariate Markov trend and Z_t is a vector autoregressive [VAR] process of order k

$$Z_t = \sum_{i=1}^{k} \Phi_i Z_{t-i} + \varepsilon_t, \tag{7.2}$$

or using the lag polynomial $\Phi(L) = (\mathbf{I}_n - \Phi_1 L - \cdots - \Phi_k L^k)$

$$(\mathbf{I}_n - \Phi_1 L - \cdots - \Phi_k L^k) Z_t = \varepsilon_t, \tag{7.3}$$

where ε_t is an n-dimensional vector normally distributed process with zero mean and $(n \times n)$ positive definite symmetric covariance matrix Σ and Φ_i, $i = 1, \ldots, k$, are $(n \times n)$ parameter matrices. The n-dimensional Markov trend is defined as

$$N_t = N_{t-1} + \Gamma_0 + \Gamma_1 S_t, \tag{7.4}$$

where $S_t = (s_{1,t} \ldots s_{n,t})'$ an n-dimensional unobserved random variable with $s_{j,t} = 0, 1$, $j = 1, \ldots, n$, Γ_0 is an $(n \times 1)$ parameter vector and Γ_1 an $(n \times n)$ parameter matrix, see also Dwyer and Potter (1996) who use a different parameterisation. The random variable S_t can take 2^n different values and hence implies 2^n possible slope values for each of the n elements in N_t. To model the transitions between these 2^n states, we assume that each element $s_{j,t}$, $j = 1, \ldots, n$ is an unobserved first-order two-state Markov process with transition probabilities

$$\begin{array}{ll} \Pr[s_{j,t} = 0 | s_{j,t-1} = 0] = p_j, & \Pr[s_{j,t} = 1 | s_{j,t-1} = 0] = 1 - p_j, \\ \Pr[s_{j,t} = 1 | s_{j,t-1} = 1] = q_j, & \Pr[s_{j,t} = 0 | s_{j,t-1} = 1] = 1 - q_j. \end{array} \tag{7.5}$$

These transition probabilities imply a restricted $(2^n \times 2^n)$ transition matrix, which contains the transition probabilities between the 2^n possible outcomes of S_t, see also the end of Section 5.1.

Using the backward solution of (7.4)

$$N_t = \Gamma_0 (t-1) + \Gamma_1 \sum_{i=2}^{t} S_i + N_1 \tag{7.6}$$

and rewriting (7.2) in the error correction notation (6.6) results in

$$(\Delta Y_t - \Gamma_0 - \Gamma_1 S_t) = \Pi(Y_{t-1} - \Gamma_0\,(t-2) - \Gamma_1 \sum_{i=2}^{t-1} S_i - N_1) + \sum_{i=1}^{k-1} \bar{\Phi}_i(\Delta Y_{t-i} - \Gamma_0 - \Gamma_1 S_{t-i}) + \varepsilon_t, \quad (7.7)$$

where $\Pi = \sum_{j=1}^{k} \Phi_j - \mathbf{I}_n$ and $\bar{\Phi}_i = -\sum_{j=i+1}^{k} \Phi_j$, $i = 1, \ldots, k-1$. Note that the initial value of the Markov trend N_1 is unknown and plays the role of an intercept parameter vector.

As already mentioned in Section 5.1 the multivariate Markov trend implies the presence of n stochastic trends in Y_t. Additionally, unit roots in the vector autoregressive part of the Markov trend model also imply the presence of stochastic trends in Y_t. In practice it may therefore be wishful to limit the number of stochastic trends in the Markov trend model (7.7). In the remainder of this section we discuss the role of the stochastic trends and provide parameter restrictions, which limit the number of trends in the model. These restrictions can be imposed from a theoretical point of view or they can be tested using the Bayes factors, which are proposed in Section 7.5.

Common Markov trends

The transition probability structure in (7.5) corresponds to the presence of n independent stochastic Markov trends $\sum_{i=2}^{t} S_i$ in (7.6). The matrix Γ_1 models the correlation between the Markov trends. If the matrix Γ_1 is diagonal, each of the univariate series in Y_t contains a different univariate Markov trend as described in (3.2) with transition probabilities p_i and q_i. If we link the Markov trends with the business cycle, this implies n independent business cycles in the univariate time series in Y_t. In case of common business cycles less than n Markov trends suffice. An extreme case arises if the series have only one common business cycle. Now only one state variable $s_{j,t}$ suffices to model the business cycle. The matrix Γ_1 becomes an $(n \times 1)$ vector. Changes in the stage of the business cycle for each series in Y_t occur at the same time, see also Krolzig (1996). Although this may not always be the case in practice, there may exist some common business cycle structure among the series so that it is suffices to specify less than n stochastic Markov trends to model the series.

We have seen that if Γ_1 has full rank, Y_t consists of n correlated stochastic Markov trends $\Gamma_1 \sum_{i=2}^{t} S_i$. A reduction in the number of Markov trends, *i.e.* common Markov trends, occurs if $0 < \text{rank}(\Gamma_1) = m < n$, see Section 5.2. In that case we can write Γ_1 as a product of two $(n \times m)$ matrices γ and δ such that $\Gamma_1 = \gamma\delta'$. The m common Markov trends are given by $\delta' \sum_{i=2}^{t} S_i$ and $\gamma_\perp' N_t$ and therefore $\gamma_\perp' Y_t$ does not contain a stochastic Markov trend. Since $Y_t = N_t + Z_t$ shocks denoted by S_t have a permanent effect on Y_t. The impact of the S_t shocks on the first differences of Y_t depends on the number of unit roots (or the rank of Π) in the vector autoregressive component Z_t.

To discuss the impacts of the possible trends in this model specification, we first focus on the Markov trend.

Cointegration

In Section 5.3 we have seen that unit roots in the VAR process Z_t result in stochastic trends $\sum_{i=2}^{t} \varepsilon_i$ in Y_t. Several situations can occur, depending on the number of unit roots.

If the roots of the polynomial $|\Phi(x)|$ are outside the unit root circle, there are no stochastic trends in Z_t. Hence, Y_t is a stationary VAR process around a multivariate Markov trend N_t. The matrix Π in (7.7) has full rank and exogenous shocks represented by S_t have a permanent effect on future values of Y_t and ΔY_t, while shocks represented by ε_t only have a transitory effect.

As we already have seen in Section 5.3 stochastic trends enter Z_t and therefore Y_t if the Π matrix has reduced rank. If $\Pi = \mathbf{0}$ equation (7.7) simplifies to

$$(\Delta Y_t - \Gamma_0 - \Gamma_1 S_t) = \sum_{i=1}^{k-1} \bar{\Phi}_i (\Delta Y_{t-i} - \Gamma_0 - \Gamma_1 S_{t-i}) + \varepsilon_t, \tag{7.8}$$

and the first difference of Y_t is a stationary VAR process with a stochastically changing mean ($= \Gamma_0 + \Gamma_1 S_t$). This corresponds to the assumption of n stochastic trends, $\sum_{i=2}^{t} \varepsilon_i$, in Y_t. Now shocks S_t and ε_t have a permanent effect on future values of Y_t, however not on the first differences of Y_t. Note that the initial value of the Markov trend N_1 drops out of the model and the growth in Y_t does not depend on the initial value of the Markov trend.

A special case arises if $0 < \text{rank}(\Pi) = r < n$. In that case Π can be written as $\Pi = \alpha\beta'$, where α and β are full rank $(n \times r)$ matrices. Model (7.7) becomes

$$\begin{aligned}(\Delta Y_t - \Gamma_0 - \Gamma_1 S_t) = \alpha\beta'(Y_{t-1} - \Gamma_0(t-2) + \Gamma_1 \sum_{i=2}^{t-1} S_i - N_1) + \\ \sum_{i=1}^{k-1} \bar{\Phi}_i (\Delta Y_{t-i} - \Gamma_0 - \Gamma_1 S_{t-i}) + \varepsilon_t,\end{aligned} \tag{7.9}$$

with the following normalisation $\beta' = (\mathbf{I}_r \;\; \beta_2')$, where β_2 is an $((n-r) \times r)$ matrix, to identify β, see (5.19). We will refer to (7.9) as the multivariate Markov trend model in error correction cointegration [ecc] form. According to (5.17) there are $(n-r)$ common stochastic trends in Y_t given by $\alpha_\perp' \sum_{i=2}^{t} \varepsilon_i$ and r cointegration relations $\beta' Z_t$. Now the growth in Y_t depends on the initial value of the Markov trend unless $\beta' N_1 = \mathbf{0}$. The matrix $\beta'\Gamma_1$ represents the impact of the stochastic Markov trend $\sum_{i=2}^{t-1} S_i$ on the growth of Y_t at time t. In case of m common Markov trends this matrix becomes $\beta'\gamma\delta'$. Not that it is possible that β' contains one or more linear combinations which remove a stochastic Markov trend from N_t, *i.e.* $\beta \in \text{sp}(\gamma_\perp)$. If $r + m \leq n$ it is even possible that $\beta'\gamma = \mathbf{0}$ and

past shocks S_{t-i} do not have a permanent effect on ΔY_t since (7.9) simplifies to

$$(\Delta Y_t - \Gamma_0 - \Gamma_1 S_t) = \alpha\beta'(Y_{t-1} - \Gamma_0(t-2) - N_1) + \sum_{i=1}^{k-1} \bar{\Phi}_i(\Delta Y_{t-i} - \Gamma_0 - \Gamma_1 S_{t-i}) + \varepsilon_t. \quad (7.10)$$

Furthermore, the relation $\beta' Y_t = \beta' N_t + \beta' Z_t$ is (trend)stationary and does not contain a stochastic Markov trend. Dwyer and Potter (1996) refer to this phenomenon as reduced rank Markov trend cointegration.

The Likelihood Function

To perform a Bayesian analysis of the multivariate Markov trend model, we consider the likelihood conditional on the states S_t, see also Section 3.2. The conditional density of Y_t for the multivariate Markov trend stationary model (7.7) given the past and current states $S^t = \{S_1, \ldots, S_t\}$ and given the past observations $Y^{t-1} = \{Y_1, \ldots, Y_{t-1}\}$ is given by

$$f(Y_t|Y^{t-1}, S^t, \Gamma_0, \Gamma_1, N_1, \Sigma, \Pi, \bar{\Phi}) = \frac{1}{(\sqrt{2\pi})^n} |\Sigma|^{-\frac{1}{2}} \exp(-\frac{1}{2}\varepsilon_t' \Sigma^{-1} \varepsilon_t), \quad (7.11)$$

where ε_t is given in (7.7) and $\bar{\Phi} = \{\bar{\Phi}_1, \ldots, \bar{\Phi}_{k-1}\}$. The likelihood function for model (7.7) conditional on the states S^T and the first k initial observations Y^k equals

$$\mathcal{L}(Y^T|Y^k, S^T, \Theta) = \prod_{j=1}^{n} p_j^{\mathcal{N}_{j,00}} (1-p_j)^{\mathcal{N}_{j,01}} q_j^{\mathcal{N}_{j,11}} (1-q_j)^{\mathcal{N}_{j,10}} \prod_{t=k+1}^{T} f(Y_t|Y^{t-1}, S^t, \Gamma_0, \Gamma_1, N_1, \Sigma, \Pi, \bar{\Phi}), \quad (7.12)$$

where $\Theta = \{\Gamma_0, \Gamma_1, N_1, \Sigma, \Pi, \bar{\Phi}, p_j, q_j, j = 1, \ldots, n\}$, $\mathcal{N}_{j,il}$ denotes the number of transitions from state i to state l for j-th state variable. The unconditional likelihood function $\mathcal{L}(Y^T|Y^k, \Theta)$ can be obtained by summing over all possible realisations of S^T

$$\mathcal{L}(Y^T|Y^k, \Theta) = \sum_{S_1} \sum_{S_2} \cdots \sum_{S_T} \mathcal{L}(Y^T|Y^k, S^T, \Theta). \quad (7.13)$$

The unconditional likelihood function for the Markov trend model in error correction cointegration [ecc] form (7.9), $\mathcal{L}_{ecc}$, follows directly from (7.13)

$$\mathcal{L}_{ecc}(Y^T|Y^k, \Theta_{ecc}) = \mathcal{L}(Y^T|Y^k, \Theta)|_{\Pi=\alpha\beta'} \quad (7.14)$$

with $\Theta_{ecc} = \{\Gamma_0, \Gamma_1, N_1, \Sigma, \alpha, \beta_2, \bar{\Phi}, p_j, q_j, j = 1, \ldots, n\}$.

In the next section we discuss the prior distributions for the model parameters for the multivariate Markov trend stationary VAR model (7.7) and the Markov trend model in error correction cointegration form (7.9).

7.3 Prior Specification

The prior specification for the multivariate Markov trend model is a combination of the prior specification for the univariate model of Chapter 3 and the proposed prior framework for cointegration analysis as derived in the previous chapter.

As we have seen in the previous section, N_1 is not identified if $\Pi = \mathbf{0}$. To correct for this identification problem we propose the same solution as for the univariate Markov trend model, see Section 3.3.1. Hence, the prior distribution for N_1 conditional on the first observation Y_1 and Σ is normal with mean Y_1 and covariance Σ

$$N_1|Y_1, \Sigma \sim \mathrm{N}(Y_1, \Sigma). \tag{7.15}$$

The prior for Σ is again diffuse

$$p(\Sigma) \propto |\Sigma|^{-\frac{1}{2}}, \tag{7.16}$$

and the prior distributions for the transition probabilities p_j and q_j are independent and uniform on the unit interval $(0, 1)$

$$\begin{aligned} p(p_j) &= \mathbb{I}_{(0,1)} \qquad j = 1, \ldots, n, \\ p(q_j) &= \mathbb{I}_{(0,1)} \qquad j = 1, \ldots, n. \end{aligned} \tag{7.17}$$

Just as in the univariate Markov trend model, under a diffuse prior specification we need to restrict the parameter space on which Γ_0 and Γ_1 have probability mass to identify the regimes, see below (3.18). Note that in the multivariate case, also changing two columns in Γ_1, the corresponding two rows in S_t for $t = 1, \ldots, T$ and changing the corresponding transition probabilities, results in the same model and the same value of the likelihood function. To circumvent this problem we define the priors for Γ_0 and Γ_1 on subspaces $\mathcal{G}_0$ and $\mathcal{G}_1$, which uniquely identify the regimes,

$$p(\Gamma_0) \propto \begin{cases} 1 & \text{if } \Gamma_0 \in \mathcal{G}_0 \\ 0 & \text{elsewhere,} \end{cases} \tag{7.18}$$

where $\mathcal{G}_0 \in \mathbb{R}^n$ and

$$p(\Gamma_1) \propto \begin{cases} 1 & \text{if } \Gamma_1 \in \mathcal{G}_1 \\ 0 & \text{elsewhere,} \end{cases} \tag{7.19}$$

where $\mathcal{G}_1 \in \mathbb{R}^{n \times n}$. Another option to circumvent the identification problem is to specify appropriate matrix normal prior distributions for Γ_0 and Γ_1. The priors for the autoregressive parameters $\bar{\Phi}_i$ are diffuse

$$p(\bar{\Phi}_i) \propto 1, \qquad i = 1, \ldots, k-1. \tag{7.20}$$

The prior for the Π and the α and β_2 parameters is based on the prior framework, derived in the previous chapter. We start with a diffuse prior for Π given Σ

$$p(\Pi|\Sigma) \propto |\Sigma|^{-\frac{1}{2}n}. \tag{7.21}$$

To obtain the prior for the Markov trend error correction cointegration model (7.9) we decompose Π as follows

$$\Pi = \begin{pmatrix} \Pi_{11} & \Pi_{12} \\ \Pi_{21} & \Pi_{22} \end{pmatrix} = \begin{pmatrix} \alpha_1 & -\alpha_1\beta_2' \\ \alpha_2 & -\alpha_2\beta_2' + \lambda' \end{pmatrix}, \tag{7.22}$$

where $\alpha = (\alpha_1' \ \alpha_2')'$ with α_1 an $(r \times r)$ matrix and α_2 an $((n-r) \times r)$ matrix and λ is an $((n-r) \times (n-r))$ matrix, see also (6.9). The diffuse prior for Π given Σ (7.21) implies the following conditional priors for α, λ and β_2[2]

$$\begin{aligned} p_{uec}(\alpha|\Sigma) &\propto |\Sigma|^{-\frac{1}{2}r} \\ p_{uec}(\lambda|\Sigma,\alpha) &\propto |\alpha_\perp'\Sigma\alpha_\perp|^{-\frac{1}{2}(n-r)} \\ p_{uec}(\beta_2|\Sigma,\alpha,\lambda) &\propto |\alpha'\Sigma^{-1}\alpha|^{\frac{1}{2}(n-r)}, \end{aligned} \tag{7.23}$$

where $\alpha_\perp = (-\alpha_2\alpha_1^{-1} \ \mathbf{I}_{n-r})'$, see Section 6.3. Under reduced rank, *i.e.* $\lambda = \mathbf{0}$, the conditional priors for α and β_2 in the error correction cointegration [ecc] model (7.9) obey the rule that $p_{ecc}(\alpha,\beta_2|\Sigma) \propto p_{uec}(\alpha,\lambda,\beta_2|\Sigma)|_{\lambda=\mathbf{0}}$ so that

$$\begin{aligned} p_{ecc}(\alpha|\Sigma) &\propto |\Sigma|^{-\frac{1}{2}r}|\alpha_\perp'\Sigma\alpha_\perp|^{-\frac{1}{2}(n-r)} \\ p_{ecc}(\beta_2|\Sigma,\alpha,\lambda) &\propto |\alpha'\Sigma^{-1}\alpha|^{\frac{1}{2}(n-r)}, \end{aligned} \tag{7.24}$$

see also (6.22).

The joint prior $p(\Theta)$ for the Markov trend stationary VAR model (7.7) is given by the product of (7.15)–(7.21). For the error correction cointegration model (7.9) the joint prior $p_{ecc}(\Theta_{ecc})$ is given by the product of (7.15)–(7.20) times (7.24). Note that the joint prior of the Markov trend cointegration model (7.9) is proportional to the joint prior of the Markov trend stationary VAR model (7.7) restricted in $\lambda = \mathbf{0}$, the parameter restriction which implies cointegration.

7.4 Simulating Posterior Distributions

The posterior distribution of the multivariate Markov trend model (7.7) is proportional to the product of the prior $p(\Theta)$ and the unconditional likelihood (7.13). To obtain marginal posterior results, we use again the Gibbs sampling algorithm, see Section 3.4. The state variables $\{S_t\}_{t=1}^T$ are treated as unknown parameters and sampled alongside the model parameter Θ. The next subsection provides the full conditional posterior distributions, which are needed to sample the parameters. Special attention will be payed to the situation, where there is rank reduction in Π. In that case we need to build a Metropolis-Hastings step in the Gibbs sampler to generate the α and β_2 parameters.

[2]The subscript corresponds to the fact that the priors are defined in an unrestricted error correction [uec] model, see Section 6.2.

7.4.1 Full Conditional Posterior Distributions

Full Conditional Posterior of the States

The derivation of the full conditional posterior of the state variables follows directly from the derivation in the univariate case in Section 3.4. Since the state variables $s_{j,t}$, $j = 1, \ldots, n$ in S_t are independent, they can be sampled separately using the full conditional posterior distribution (3.35). The kernel of the full conditional distribution of $s_{j,t}$, $j = 1, \ldots, n$, $t = k+1, \ldots T$ now reads

$$p(s_{j,t}|S^T \backslash \{s_{j,t}\}, \Theta, Y^T) \propto p(s_{j,t}|s_{j,t-1}, \Theta)\, p(s_{j,t+1}|s_{j,t}, \Theta) \prod_{i=t}^{T} f(Y_i|Y^{i-1}, S^i, \Theta), \tag{7.25}$$

where $f(Y_i|Y^{i-1}, S^i, \Theta)$ is defined in (7.11). The initial k values of the states can be sampled using a full conditional posterior distribution like (3.36).

Full Conditional Posterior of p_j and q_j

The full conditional posterior distributions of p_j and q_j, $j = 1, \ldots, n$ are beta distributions since according to (7.12)

$$\begin{aligned} p(p_j|S^T, \Theta \backslash \{p_j\}, Y^T) &\propto p_j^{\mathcal{N}_{j,00}} (1-p_j)^{\mathcal{N}_{j,01}} \\ p(q_j|S^T, \Theta \backslash \{q_j\}, Y^T) &\propto q_j^{\mathcal{N}_{j,11}} (1-q_j)^{\mathcal{N}_{j,10}}, \end{aligned} \tag{7.26}$$

where $\mathcal{N}_{j,il}$ denotes the number of transitions from state i to state l for the j-th state variable.

Full Conditional Posterior of Σ

It is easy to see from the conditional likelihood (7.12) that the full conditional posterior of Σ is proportional to

$$p(\Sigma|S^T, \Theta \backslash \Sigma, Y^T) \propto |\Sigma|^{-\frac{1}{2}(T-k+n+2)} \exp(-\frac{1}{2}\mathrm{tr}(\Sigma^{-1}((Y_1 - N_1)(Y_1 - N_1)' + \sum_{t=k+1}^{T} \epsilon_t \epsilon_t'))) \tag{7.27}$$

and hence the covariance matrix Σ can be sampled from an inverted Wishart distribution, see Zellner (1971, p. 395).

Full Conditional Posterior of N_1, Γ_0 and Γ_1

To derive the full conditional posterior distribution of N_1, Γ_0 and Γ_1 we write (7.7) as

$$\begin{aligned} \Phi(L)Y_t &= \Phi(L)(\Gamma_0(t-1) + \Gamma_1 \sum_{i=2}^{t} S_i + N_1) + \varepsilon_t \\ &= -\sum_{j=1}^{k} \Phi_j (\Gamma_0 \;\; \Gamma_1 \;\; N_1) \begin{pmatrix} L^j(t-1) \\ L^j \sum_{i=2}^{t} S_i \\ 1 \end{pmatrix} + \varepsilon_t, \end{aligned} \tag{7.28}$$

where $\Phi_0 = -\mathbf{I}_n$. Without the Φ_j matrices, we have a multivariate regression model in the parameters N_1, Γ_0 and Γ_1 and the full conditional distribution would be matrix normal. To reverse the order of $\Phi(L)$ and the parameters $(\Gamma_0 \ \Gamma_0 \ N_1)$, we apply the vec operator to both sides of (7.28). Using the vec notation and the fact that $\text{vec}(ABC) = (C' \otimes A)\text{vec}(B)$, we can write (7.28) as a linear regression model and hence the full conditional distributions of $\text{vec}(N_1)$, $\text{vec}(\Gamma_0)$ and $\text{vec}(\Gamma_1)$ are normal.

Full Conditional Posterior of Π and $\bar{\Phi}$

To sample from the full conditional posterior of the autoregressive parameters we use that conditional on Γ_0, Γ_1, N_1 and the states $\{S_t\}_{t=1}^T$, equation (7.7) can be seen as a multivariate regression model in the parameters Π and $\bar{\Phi}$. From Zellner (1971, chapter VIII) if follows that the full conditional posterior distribution of the parameter matrices are matrix normal.

Sampling of α and β_2

To derive the full conditional posterior distributions for α and β_2 we rewrite (7.9) such that conditional on $\bar{\Phi}$, N_1, Γ_0, Γ_1 and the states $\{S_T\}_{t=1}^T$ it resembles the simple VAR(1) model (6.7) considered in the previous chapter. Using $Z_t = Y_t - N_t$ we can write

$$\begin{aligned} \Delta Z_t - \sum_{i=1}^{k-1} \bar{\Phi}_i \Delta Z_{t-i} &= \alpha\beta' Z_{t-1}, + \varepsilon_t \\ \Delta Z_t^* &= \alpha\beta' Z_{t-1}^* + \varepsilon_t, \end{aligned} \tag{7.29}$$

where $\Delta Z_t^* = \Delta Z_t - \sum_{i=1}^{k-1} \bar{\Phi}_i \Delta Z_{t-i}$ and $Z_{t-1}^* = Y_{t-1} - N_{t-1}$. The full conditional posterior distribution of (α, β_2) is given by the product of conditional posterior distribution of α given (Σ, Z^*), $p_{uec}(\alpha|\Sigma, Z^*)$, β_2 given (Σ, α, Z^*), $p_{uec}(\beta_2|\Sigma, \alpha, Z^*)$ and the conditional posterior of λ given (Σ, α, Z^*) evaluated in $\lambda = \mathbf{0}$, $p_{uec}(\lambda|\Sigma, \alpha, Z^*)|_{\lambda=\mathbf{0}}$. These distributions are given in (6.28), where $\Delta Z^* = (\Delta Z_k^* \dots \Delta Z_T^*)'$ takes the role of ΔY and $Z_{-1}^* = (Z_{k-1} \dots Z_{T-1})'$ takes the role of Y_{-1}. Note that this conditional posterior distribution is the full conditional distribution since Z^* is a function of the parameters $\bar{\Phi}$, N_1, Γ_0 and Γ_1.

As we already have discussed in Section 6.4 it is not possible to sample directly from this distribution and we have proposed a Metropolis-Hastings sampling algorithm in Section 6.4.1. Chib and Greenberg (1994, 1995) show that it is possible to build such a Metropolis-Hastings step into the Gibbs sampling procedure. Briefly, draw α from $p_{uec}(\alpha|\Sigma, Z^*)$ and β_2 from $p_{uec}(\beta_2|\Sigma, \alpha, Z^*)$, which are matrix normal distributions. The probability of accepting this draw is given by the ratio of the conditional posterior $p_{uec}(\lambda|\Sigma, \alpha, Z^*)|_{\lambda=\mathbf{0}}$ of the current drawing and the previous drawing, see Section 6.4.1 for details about the Metropolis-Hastings step.

In the next section we propose a method to analyse the presence of (common) stochastic trends in the multivariate Markov trend model.

7.5 Bayesian Analysis of Common Trends

The Bayesian analysis of the presence of common trends in the multivariate Markov trend model is based on the ideas in Chapter 6. In Section 7.5.1 we consider the analysis of stochastic trends caused by unit roots in the autoregressive part of the model. In Section 7.5.2 we propose a method to analyse common Markov trends, which proceeds in a similar way.

7.5.1 Cointegration Analysis

To investigate the presence of common stochastic trends (cointegration), we may analyse the posterior distribution of λ, which results from the decomposition of Π in (7.22). From (7.22) it follows that that $\lambda = \mathbf{0}$ implies that the matrix Π has reduced rank and hence the presence of common trends. However, in Section 6.2 we have already discussed that a posterior odds ratio for rank reduction in Π based on decomposition (7.22) depends on the order of the variables in Y_t. To analyse the number of common stochastic trends or the cointegration rank it is better to use the invariant decomposition of Π as proposed in Section 6.7

$$\Pi = (\alpha \ \alpha_\perp) \begin{pmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & (\lambda^*)' \end{pmatrix} \begin{pmatrix} \beta' \\ \beta_\perp' \end{pmatrix}, \tag{7.30}$$

where λ^* is an $((n-r) \times (n-r))$ matrix, $\beta_\perp'\beta_\perp = \mathbf{I}_{n-r}$, $\alpha_\perp'\alpha_\perp = \mathbf{I}_{n-r}$ and where $\alpha_\perp$ and $\beta_\perp$ are $(n \times (n-r))$ matrices defined such that $\alpha_\perp'\alpha = \mathbf{0}$ and $\beta_\perp'\beta = \mathbf{0}$ like in (6.67). Note that due to the normalisation restriction $\alpha_\perp$ does not equal $(-\alpha_2\alpha_1^{-1} \ \ \mathbf{I}_{n-r})'$ any more. The decomposition in (7.30) is identified trough a singular value decomposition on Π, see (6.68). The matrix λ^* is based on the r smallest singular values of Π and rank reduction in Π corresponds to $\lambda^* = \mathbf{0}$.

The diffuse prior for Π (7.21) implies via a Jacobian transformation the joint prior $p^*(\alpha, \lambda^*, \beta)$[3], see Section 6.7. Appendix 6.C provides the Jacobian for the transformation from Π to $(\alpha, \lambda^*, \beta)$. The joint posterior of the model parameters is the product of the joint prior times the likelihood function (7.13) as function of $(\alpha, \lambda^*, \beta)$. This likelihood function is denoted by $\mathcal{L}^*(Y^T|Y^k, \Theta^*)$, where we replace Π in Θ with $(\alpha, \lambda^*, \beta)$ to obtain Θ^*. Unfortunately it is not possible to derive any conditional posterior densities for α, λ^* and β. However it is easy to sample these parameters in the Gibbs framework from their full conditional distribution since we know how to sample Π. To sample $(\alpha, \lambda^*, \beta)$ from its full conditional posterior distribution, we perform a singular value decomposition on the sampled Π and compute α, λ^* and β according to (6.70).

The analysis of the presence of one or more stochastic trends, can be done using posterior odds ratios for $\lambda^* = \mathbf{0}$. We start with assigning prior probabilities to every possible rank of Π

$$\Pr[\text{rank} = r], \qquad r = 0, \ldots, n, \tag{7.31}$$

[3]The asterisk denotes again that the prior follows from the invariant decomposition, like in Chapter 6. The same notation will be used for the likelihood functions and the posterior distributions.

which implies prior probabilities to the number of stochastic trends $(n-r)$. These prior probabilities imply the following prior odds ratios [PROR]

$$\text{PROR}(r|n) = \frac{\Pr[\text{rank} = r]}{\Pr[\text{rank} = n]}, \qquad r = 0, \ldots, n. \tag{7.32}$$

The Bayes factor to compare rank r with rank n equals

$$\text{BF}^*(r|n) = \frac{\int \mathcal{L}^*_{ecc}(Y^T|Y^k, \Theta^*_{ecc})\, p^*_{ecc}(\Theta^*_{ecc})\, d\Theta^*_{ecc}}{\int \mathcal{L}^*(Y^T|Y^k, \Theta^*)\, p^*(\Theta^*)\, d\Theta^*}, \tag{7.33}$$

where $\mathcal{L}^*(Y^T|Y^k, \Theta)$ and $p^*(\Theta^*)$ denote the unconditional likelihood function and the joint prior of the Markov trend stationary model (7.7) as function of $(\alpha, \lambda^*, \beta)$ instead of Π. The prior and likelihood of the error correction cointegration model (7.9) are defined such that $\mathcal{L}^*_{ecc}(Y^T|Y^k, \Theta^*_{ecc}) = \mathcal{L}^*(Y^T|Y^k, \Theta^*)|_{\lambda^*=\mathbf{0}}$ and $p^*_{ecc}(\Theta^*_{ecc}) = p^*(\Theta^*)|_{\lambda^*=\mathbf{0}}$ with $\Theta^*_{ecc} = \Theta^* \backslash \{\lambda^*\}$. The posterior odds ratios to compare rank r with rank n equals prior odds ratio times the Bayes factor, $\text{POR}(r|n) = \text{PROR}(r|n) \times \text{BF}^*(r|n)$, and the posterior probabilities for every rank are simply

$$\Pr[\text{rank} = r|Y] = \frac{\text{POR}(r|n)}{\sum_{i=0}^{n} \text{POR}(i|n)}, \qquad r = 0, \ldots, n. \tag{7.34}$$

In Section 6.7.2 we have seen that the Bayes factor (7.33) can be computed using the Savage-Dickey density ratio of Dickey (1971), which states that the Bayes factor for $\lambda^* = \mathbf{0}$ equals the ratio of the marginal posterior density and the marginal prior density of λ^*, both evaluated in $\lambda^* = \mathbf{0}$

$$\text{BF}^*(r|n) = \frac{p^*(\lambda^*|Y^T)|_{\lambda^*=\mathbf{0}}}{p^*(\lambda^*)|_{\lambda^*=\mathbf{0}}}. \tag{7.35}$$

This means that we need the marginal posterior density of λ^* to compute this Savage-Dickey density ratio. Since the full conditional distribution of λ^* is of an unknown type, we have to use an approximation of the full conditional posterior density of λ^* in combination with importance weights to compute the marginal posterior of λ^*, see Chen (1994). A suitable approximation is given in (6.78) with ΔY and Y_{-1} replaced by ΔZ^* and Z^*_{-1}, which are defined at the end of Section 7.4. Another solution to compute the marginal posterior density of λ^* in zero is to use a non-parametric kernel estimator on the λ^* draws, see *e.g.* Silverman (1986). As we have specified an implicit diffuse prior for λ^*, the height of the marginal prior in $\lambda^* = \mathbf{0}$ is not defined. Therefore, we take for $p(\lambda^*)|_{\lambda^*=\mathbf{0}}$ the value $(2\pi)^{-\frac{1}{2}(n-r)^2}$ as in (6.66), which is related to the posterior information criterion [PIC] of Phillips and Ploberger (1994), see the end of Section 6.6.1 and Kleibergen and Paap (1996) for details.

7.5.2 Common Markov Trends Analysis

In Section 5.2 we have already seen that rank reduction in Γ_1 implies the presence of common stochastic Markov trends. Therefore, we can we can follow the same strategy as

for analysing rank reduction in Π to analyse the presence of common stochastic Markov trends. The decomposition for Γ_1 reads

$$\Gamma_1 = (\gamma \ \gamma_\perp) \begin{pmatrix} \mathbf{I}_m & \mathbf{0} \\ \mathbf{0} & \zeta^* \end{pmatrix} \begin{pmatrix} \delta' \\ \delta'_\perp \end{pmatrix}, \tag{7.36}$$

where ζ^* is an $((n-m) \times (n-m))$ matrix, $\gamma'_\perp \gamma_\perp = \mathbf{I}_{n-m}$, $\delta'_\perp \delta_\perp = \mathbf{I}_{n-m}$ and $\gamma_\perp$ and $\delta_\perp$ are $(n \times (n-m))$ matrices defined such that $\gamma'_\perp \gamma = \mathbf{0}$ and $\delta'_\perp \delta = \mathbf{0}$, *cf.* (7.30). Decomposition (7.36) corresponds to a singular value decomposition on Γ_1 like in (6.68). The ζ^* parameters are identified through the m smallest singular values of Γ_1 such that the presence of common stochastic Markov trends corresponds to $\zeta^* = \mathbf{0}$.

The analysis of rank reduction in Γ_1 proceeds in the same way as rank reduction in Π. The decomposition (7.36) relates the prior on Γ_1 (7.19) to the joint prior on $(\gamma, \zeta^*, \delta)$, see Section 6.7 and 7.5.1. The posterior is given by the product of the prior times the likelihood (7.13). Now, the Bayes factor for m versus n stochastic Markov trends can be computed using the Savage-Dickey density ratio of Dickey (1971) for $\zeta^* = \mathbf{0}$. This Bayes factor equals the ratio of the marginal posterior density of ζ^* and the marginal prior density of ζ^* both evaluated in $\zeta^* = \mathbf{0}$

$$\mathrm{BF}^*(m|n) = \frac{p^*(\zeta^*|Y^T)|_{\zeta^*=\mathbf{0}}}{p^*(\zeta^*)|_{\zeta^*=\mathbf{0}}}. \tag{7.37}$$

Prior odds ratios $\mathrm{PROR}(m|n)$, which lead to posterior odds ratios $\mathrm{POR}(m|n)$ can be constructed in the same way as in the previous section.

To calculate these Bayes factors we can perform a singular value decomposition on the sampled Γ_1 parameters, solve for ζ^* like in (6.70) and use a kernel estimator [see Silverman (1986)] to evaluate the marginal of ζ^* in zero. For the height of the marginal prior density of ζ^* we can perform a singular value decomposition on draws from the marginal prior together with a kernel estimator. In case of a diffuse prior for Γ_1 we can take the factor $(2\pi)^{-\frac{1}{2}(n-m)^2}$ for $p^*(\zeta^*)|_{\zeta^*=\mathbf{0}}$.

Since the number of possible values of the state variable S_t is 2^n, the multivariate Markov trend allows for large n many slope values of the trend for the univariate series in Y_t. In practice, it may be interested to link the state variables with stages of the business cycle so that an expansion and a contraction regime for each series is sufficient. Furthermore, one wants to test whether several series have a common business cycle so that only one $s_{j,t}$ is sufficient to describe the Markov trend and the business cycle. A direct method to test for this possibility it to restrict some parameters in Γ_1. For instance when $n = 2$ one may consider

$$\Gamma_1 = \begin{pmatrix} \Gamma_{1,11} & 0 \\ \Gamma_{1,21} & \Gamma_{1,22} \end{pmatrix} \tag{7.38}$$

so that $\Gamma_{1,22} = 0$ implies that $s_{1,t}$ is sufficient to describe the business cycle for the 2-dimensional time series Y_t. The Bayes factor for $\Gamma_{1,22} = 0$ can be computed using a

Savage-Dickey density ratio

$$\mathrm{BF}(\Gamma_{1,22}) = \frac{p(\Gamma_{1,22}|Y^T)|_{\Gamma_{1,22}=0}}{p(\Gamma_{1,22})|_{\Gamma_{1,22}=0}}, \tag{7.39}$$

where $p(\Gamma_{1,22}|Y^T)$ is the marginal posterior density and $p(\Gamma_{1,22})$ denotes the marginal prior density of $\Gamma_{1,22}$. Note that Bayes factors for a common Markov trend using decomposition (7.38) is likely to depend on the order of the variables in Y_t.

7.6 Application

In this section we analyse the presence of stochastic trends in three simulated series and in per capita income and consumption of the United States using multivariate Markov trend models. The main focus of this section is the determination of the number of Markov trends and stochastic trends caused by unit roots in the autoregressive part of the model.

7.6.1 Simulated Series

To illustrate the cointegration analysis in the presence of Markov trends we consider the following data generating process [DGP]

$$\begin{aligned} Y_t &= N_t + Z_t, \\ N_t &= N_{t-1} + \begin{pmatrix} 2 \\ 2 \end{pmatrix} - \begin{pmatrix} 4 \\ 4 \end{pmatrix} s_{1,t}, \qquad & N_1 &= \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \\ \Delta Z_t &= \Pi Z_{t-1} + \varepsilon_t, & \varepsilon_t &\sim \mathrm{NID}(\mathbf{0}, \mathbf{I}_2), \end{aligned} \tag{7.40}$$

with three different parameter values for Π

$$\begin{aligned} \mathrm{I} : \Pi &= \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \\ \mathrm{II} : \Pi &= \begin{pmatrix} -0.25 \\ 0.25 \end{pmatrix} (1 \ -1), \\ \mathrm{III} : \Pi &= \begin{pmatrix} -0.50 & -0.50 \\ 0.25 & -0.25 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}, \end{aligned} \tag{7.41}$$

and hence for DGP I Z_t contains two unit roots, for DGP II a unit root and a root 0.5 and for the last DGP two roots of 0.5. The state variables $\{s_{1,t}\}_{t=1}^T$ are generated according to a first-order Markov process with transition probabilities $p_1 = 0.9$ and $q_1 = 0.6$ and contains the same realised values for the three simulated series.[4] Note that we impose in (7.40) that Y_t has one common Markov trend. The number of observations is $T = 100$.

[4]The simulated states and therefore the Markov trends are exactly the same as for the simulated series in Section 3.6.1.

Table 7.1. Posterior means with posterior standard deviations between parentheses and Bayes factors for the cointegration rank for the three DGPs.

par.	DGP I	DGP II	DGP III
Γ_0	$\begin{pmatrix} 1.80 \\ (0.08) \\ 2.03 \\ (0.11) \end{pmatrix}$	$\begin{pmatrix} 2.06 \\ (0.06) \\ 2.06 \\ (0.06) \end{pmatrix}$	$\begin{pmatrix} 1.99 \\ (0.02) \\ 1.99 \\ (0.02) \end{pmatrix}$
Γ_1	$\begin{pmatrix} -3.45 \\ (0.24) \\ -4.12 \\ (0.31) \end{pmatrix}$	$\begin{pmatrix} -3.80 \\ (0.20) \\ -3.84 \\ (0.19) \end{pmatrix}$	$\begin{pmatrix} -3.91 \\ (0.12) \\ -3.96 \\ (0.12) \end{pmatrix}$
N_1	$\begin{pmatrix} 1.20 \\ (0.93) \\ 1.96 \\ (1.04) \end{pmatrix}$	$\begin{pmatrix} 0.68 \\ (0.80) \\ 1.16 \\ (0.74) \end{pmatrix}$	$\begin{pmatrix} -0.42 \\ (0.42) \\ 0.12 \\ (0.43) \end{pmatrix}$
p_1	0.90 (0.03)	0.90 (0.03)	0.91 (0.03)
q_1	0.61 (0.11)	0.64 (0.11)	0.63 (0.11)
Π	$\begin{pmatrix} -0.09 & -0.02 \\ (0.07) & (0.04) \\ -0.08 & -0.07 \\ (0.11) & (0.06) \end{pmatrix}$	$\begin{pmatrix} -0.35 & 0.27 \\ (0.09) & (0.09) \\ 0.35 & -0.47 \\ (0.08) & (0.09) \end{pmatrix}$	$\begin{pmatrix} -0.67 & 0.18 \\ (0.10) & (0.11) \\ 0.27 & -0.62 \\ (0.09) & (0.10) \end{pmatrix}$
Σ	$\begin{pmatrix} 0.88 & 0.01 \\ (0.14) & (0.10) \\ 0.01 & 1.09 \\ (0.10) & (0.16) \end{pmatrix}$	$\begin{pmatrix} 1.25 & -0.14 \\ (0.19) & (0.13) \\ -0.14 & 1.07 \\ (0.13) & (0.16) \end{pmatrix}$	$\begin{pmatrix} 1.15 & 0.09 \\ (0.17) & (0.10) \\ 0.09 & 0.85 \\ (0.10) & (0.17) \end{pmatrix}$

r	$\ln(\mathrm{BF}^*(r\|2))$[1]	$\Pr[r\|Y]$[2]	$\ln(\mathrm{BF}^*(r\|2))$	$\Pr[r\|Y]$	$\ln(\mathrm{BF}^*(r\|2))$	$\Pr[r\|Y]$
0	11.86	1.00	−8.41	0.00	−27.11	0.00
1	4.76	0.00	3.16	0.96	−2.11	0.10
2	0.00	0.00	0.00	0.04	0.00	0.90

[1] A Bayes factor $\ln(\mathrm{BF}^*(r|2)) > 0$ denotes that the presence of $(2-r)$ common trends or r cointegration relations is more likely than two cointegration relations.

[2] Posterior probability of the cointegration rank (7.34) is based on equal prior probabilities (7.31) for every rank r.

To analyse the three simulated series, we consider the multivariate Markov trend model (7.7), where we restrict Γ_1 to be a (2×1) parameter vector and hence we only allow for one unobserved state variable $s_{1,t}$ in the model as in the DGP (7.40). The lag order of the model k is equal to one, which is the same as in the DGPs. The priors for model parameters are given in (7.15)–(7.17) and (7.21). The priors for Γ_0 and Γ_0 are flat on the regions $\mathcal{G}_0 = \{\Gamma_0 \in \mathbb{R}^2 | \Gamma_0 > \mathbf{0}\}$ and $\mathcal{G}_1 = \{\Gamma_1 \in \mathbb{R}^2 | \Gamma_0 + \Gamma_1 \leq \mathbf{0}\}$ respectively, which identifies the two regimes.

Table 7.1 shows the posterior results for the three simulated series. The posterior means of parameters match the true values of the DGP quite well. All posterior means are within two posterior standard deviations from their true value. The fact that the simulated series contain more observations with $s_{1,t} = 0$ than with $s_{1,t} = 1$ shows up in the posterior standard deviations of the parameters. These standard deviations are larger for q_1 and Γ_1 than for p_1 and Γ_0. Note further that the posterior standard deviations of the elements of N_1 increase with the number of unit roots in the DGP. Similar findings are reported in Section 3.6.1.

To analyse the number of non-Markov stochastic trends in the series, or in other words the rank of Π, we consider the invariant decomposition of Π in (7.30). Bayes factors for the rank of Π are computed using the Savage-Dickey density ratio, as described in Section 7.5.1. Since we have specified an uninformative prior for Π, these Bayes factors are based on the factor $(2\pi)^{-\frac{1}{2}(n-r)^2}$ for the prior on λ^* in (7.35). We assign equal prior probabilities to every cointegration rank, *i.e.* $\Pr[\text{rank} = r] = \frac{1}{3}$, $r = 0, 1, 2$. The Bayes factors and the posterior probabilities for the cointegration rank are shown in the bottom half of the table. The results for DGP I show that every rank reduction in Π is very likely and the posterior probability for two unit roots (zero cointegration relations) is one. For DGP II, a model with rank one is more likely than a model with rank two, but a model with $\Pi = \mathbf{0}$ is not very likely. The Bayes factors imply 0.96 posterior probability on the true cointegration rank. For DGP III the posterior probability for the true cointegration rank is 0.90. The Bayes factors show that rank reduction in Π is unlikely.

The three simulated series show the possibility of Bayesian analysis of the cointegration rank in the presence of Markov trends. Since only three simulated series have been considered, it must be stressed that no general conclusion can be drawn about the performance of the approach. In the next subsection we analyse per capita consumption and income of the United States in more detail.

7.6.2 US Income and Consumption

In this section we consider seasonally adjusted real gross domestic product [GDP] and consumption per capita for the period 1957.I–1992.IV. The series are obtained from Citibase. Figure 7.2 shows a plot of the logarithm of the two series. Both series increase over the sample period with short periods of decline, for instance in the middle and the end of the 1970s. These periods of decline are more pronounced in the income series than in the consumption series but seem to occur roughly simultaneously. The average growth rate

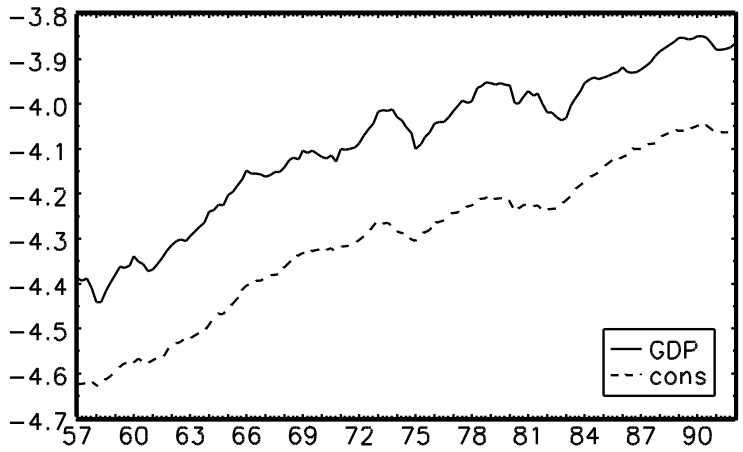

Figure 7.2. The logarithm of US per capita income and consumption, 1957.I–1992.IV.

of consumption series is 0.41% and for the income series 0.38%.

To analyse the series we consider the multivariate Markov trend stationary model (7.7) for $Y_t = (y_t \; c_t)'$, where y_t and c_t are 100 times the logarithm of the per capita income and the consumption series, respectively. The order of the VAR part of the model k is arbitrarily set to one to avoid overparametrisation, since the multivariate Markov trend specification already contains many parameters. The prior distributions of the model parameters are given by (7.15)–(7.21). To identify the regimes for the Markov trend we define $\mathcal{G}_0 = \{\Gamma_0 | \Gamma_0 > \mathbf{0}\}$ and $\mathcal{G}_1 = \{\Gamma_1 | \Gamma_0 + \Gamma_{1,1} < \Gamma_0 + \Gamma_{1,1} + \Gamma_{1,2} < \Gamma_0\}$ where $\Gamma_{1,i}$ denotes the i-th column of Γ_1. This implies the following four Markov trend slope values for Y_t in order of magnitude $\Gamma_0 + \Gamma_{1,1} < \Gamma_0 + \Gamma_{1,1} + \Gamma_{1,2} < \Gamma_0 < \Gamma_0 + \Gamma_{1,2}$. Note that this prior already imposes a structure on the Markov trend. It assumes the same order of magnitude in the slopes of the Markov trend for both series.

The posterior means with posterior standard deviations between parentheses of the posterior distributions of the model parameters read

$$
\begin{aligned}
Y_t &= N_t + Z_t, \\
N_t &= -\begin{pmatrix} 438.9 \\ (0.8) \\ 462.1 \\ (0.5) \end{pmatrix} + \begin{pmatrix} 0.26 \\ (0.18) \\ 0.31 \\ (0.06) \end{pmatrix} (t-1) + \begin{pmatrix} -2.02 & 0.89 \\ (0.59) & (0.35) \\ -1.31 & 0.69 \\ (0.35) & (0.16) \end{pmatrix} \begin{pmatrix} s_{1,t} \\ s_{2,t} \end{pmatrix}, \\
\Delta Z_t &= \begin{pmatrix} -0.29 & 0.54 \\ (0.13) & (0.44) \\ -0.06 & -0.03 \\ (0.08) & (0.14) \end{pmatrix} Z_{t-1} + \varepsilon_t, \text{ with } \Sigma = \begin{pmatrix} 0.69 & 0.18 \\ (0.18) & (0.10) \\ 0.18 & 0.25 \\ (0.10) & (0.06) \end{pmatrix}.
\end{aligned}
\tag{7.42}
$$

The posterior means of the transition probabilities of $s_{1,t}$ are $p_1 = 0.93$ and $q_1 = 0.45$ with

posterior standard deviations 0.03 and 0.20 respectively. The posterior means of p_2 and q_2 are 0.81 (0.07) and 0.65 (0.18) with posterior standard deviations between parentheses. Before analysing the business cycles in the series, we first focus on the trend specification. The results in (7.42) imply the posterior means of the four Markov trend slope values. These are -1.76, -0.87, 0.26 and 1.15 for the income series and -1.00, -0.31, 0.31 and 1.00 for the consumption series. Note that the negative slopes of the Markov trend in income are larger in absolute value than in consumption. The same is true for the largest positive slope.

To analyse the presence of the stochastic trends in the two series, we first consider the possibility of a common Markov trend. The prior on the parameter matrix Γ_1 implies via decomposition (7.36) a prior on ζ^*. As we already have seen in Section 7.5.2 a common Markov trend corresponds to $\zeta^* = 0$. Using the Savage-Dickey density ratio we can compute the Bayes factor for $\zeta^* = 0$ (7.37). The logarithm of this Bayes factor equals 1.73, which suggest the presence of a common Markov trend in consumption and income. Since we have used a diffuse prior for the Γ_1 parameter, the Bayes factor is constructed based on the $(\sqrt{2\pi})^{-1}$ factor for the prior of ζ^*.

To analyse further simplifications of the Markov trend specification in the model, we use the decomposition of Γ_1 in (7.38). Note that an analysis based on this decomposition may be sensitive to the ordering of the variables in Y_t. Since the Γ_0 parameter belonging to y_t (=0.26) in the previous model lies within two posterior standard deviations from zero, we restrict the number of possible states in the income series, *i.e.* we put a zero restriction in Γ_1 which concerns income variable y_t ($\Gamma_{1,12} = 0$). The prior for the multivariate Markov trend stationary model with Γ_1 equal to (7.38) is again given by (7.15)–(7.21) with $\mathcal{G}_0 = \{\Gamma_0 | \Gamma_0 < \mathbf{0}\}$ and $\mathcal{G}_1 = \{\Gamma_1 | \Gamma_0 + \Gamma_{1,1} < \mathbf{0}\}$. The posterior means with posterior standard deviations between parentheses of the model parameters of this model are given by

$$
\begin{aligned}
Y_t &= N_t + Z_t, \\
N_t &= -\begin{pmatrix} 439.0 \\ (0.8) \\ 462.2 \\ (0.6) \end{pmatrix} + \begin{pmatrix} 0.86 \\ (0.09) \\ 0.31 \\ (0.07) \end{pmatrix} (t-1) + \begin{pmatrix} -1.46 & 0 \\ (0.57) & \\ -0.56 & 0.38 \\ (0.16) & (0.13) \end{pmatrix} \begin{pmatrix} s_{1,t} \\ s_{2,t} \end{pmatrix} \\
\Delta Z_t &= \begin{pmatrix} -0.54 & 0.63 \\ (0.15) & (0.22) \\ -0.26 & 0.29 \\ (0.07) & (0.09) \end{pmatrix} Z_{t-1} + \varepsilon_t, \text{ with } \Sigma = \begin{pmatrix} 0.62 & 0.23 \\ (0.11) & (0.06) \\ 0.23 & 0.34 \\ (0.06) & (0.06) \end{pmatrix},
\end{aligned}
\tag{7.43}
$$

with $p_1 = 0.85$ (0.06), $q_1 = 0.76$ (0.07), $p_2 = 0.32$ (0.23) and $q_2 = 0.53$ (0.27). The posterior means of the slopes of the Markov trend for the income series equal -0.60 and 0.86. For the consumption series we have one negative slope -0.25 and three positive slopes 0.13, 0.31 and 0.69. The posterior mean of $\Gamma_{1,22}$ is 0.38 does not lie within two posterior standard deviations from zero indicating no common business cycle. However, the posterior expectations of the state variable $s_{2,t}$ fluctuate intensively within the region 0.25 and 0.90 without displaying a clear switching pattern, see Figure 7.3. Hence, it seems

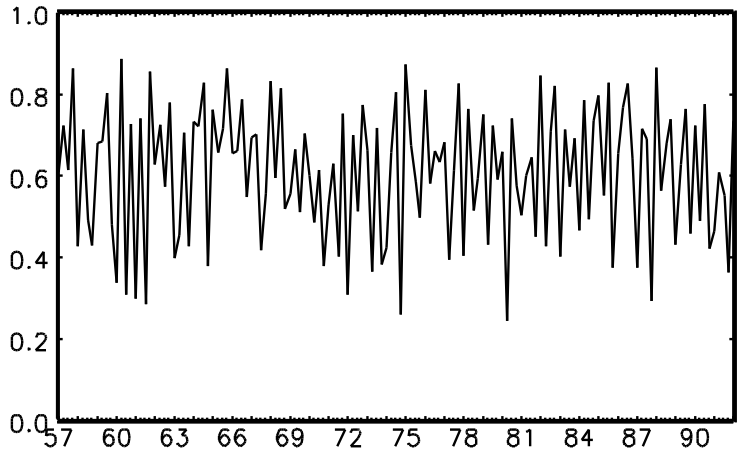

Figure 7.3. Posterior expectation of the state variable $s_{2,t}$ for model (7.43).

that this state variable mimics random walk behaviour. The posterior expectation of the first state variable is similar to Figure 7.4 and displays clear switches between two states. Unreported results show that a zero restriction in Γ_1 which concerns the c_t variable leads to the same conclusion. Since we already have seen in Section 5.3 that unit roots in Z_t introduce a cumulative sum of ε_t type stochastic trend in Y_t, we decide to remove the state variable $s_{2,t}$ from the multivariate Markov trend specification and analyse the number of unit roots in the autoregressive part of the model.

To determine the cointegration rank, we consider the multivariate Markov trend model (7.7), where we restrict Γ_1 to be a (2×1) parameter vector and hence we allow only for one unobserved state variable $s_{1,t}$. The priors for the model parameters are the same as in the previous model. To identify the two regimes in the Markov trend we define $\mathcal{G}_0 = \{\Gamma_0 \in \mathbb{R}^2 | \Gamma_0 \geq \mathbf{0}\}$ and $\mathcal{G}_1 = \{\Gamma_1 \in \mathbb{R}^2 | \Gamma_0 + \Gamma_1 < \mathbf{0}\}$. This leads to the following posterior means with posterior standard deviations between parentheses of the model parameters

$$
\begin{aligned}
Y_t &= N_t + Z_t, \\
N_t &= - \begin{pmatrix} 439.1 \\ (0.8) \\ 462.1 \\ (0.6) \end{pmatrix} + \begin{pmatrix} 1.02 \\ (0.08) \\ 0.73 \\ (0.06) \end{pmatrix} (t-1) - \begin{pmatrix} 1.67 \\ (0.16) \\ 0.81 \\ (0.08) \end{pmatrix} s_{1,t}, \\
\Delta Z_t &= \begin{pmatrix} -0.52 & 0.60 \\ (0.11) & (0.15) \\ -0.24 & 0.27 \\ (0.06) & (0.08) \end{pmatrix} Z_{t-1} + \varepsilon_t, \text{ with } \Sigma = \begin{pmatrix} 0.62 & 0.23 \\ (0.10) & (0.05) \\ 0.23 & 0.38 \\ (0.05) & (0.05) \end{pmatrix},
\end{aligned}
\tag{7.44}
$$

with posterior means of the transition probabilities $p_1 = 0.87$ (0.04) and $q_1 = 0.76$ (0.08).

To analyse the number of non-Markov stochastic trends we consider the decomposition of Π as described in (7.30). The diffuse prior on Π implies a prior on λ^*. Bayes factors for rank reduction (7.33) are computed using the Savage-Dickey density ratio (7.35). Since we have specified an uninformative prior for Π, the Bayes factors are computed using the $(2\pi)^{-\frac{1}{2}(n-r)^2}$ factor. The logarithm of the Bayes factors for $\lambda^* = 0$ and $\Pi = \mathbf{0}$ are $\ln(\mathrm{BF}^*(1|2)) = 5.20$ and $\ln(\mathrm{BF}^*(0|2)) = -3.26$, respectively. Under equal prior probabilities $\Pr[\text{rank} = r] = \frac{1}{3}$, $r = 0, 1, 2$, this implies 99% posterior probability for the model with one cointegration relation, see (7.34).

The Bayes factors suggest a multivariate Markov trend model with one cointegration relation imposed (7.9). The priors for the model parameters are the same as for the previous model. The prior for α and β (7.24) follows from the diffuse prior on Π, see the end of Section 7.3. The following posterior means and standard deviation result from the Bayesian analysis[5]

$$
\begin{aligned}
Y_t &= N_t + Z_t, \\
N_t &= -\begin{pmatrix} 439.0 \\ (0.8) \\ 462.2 \\ (0.6) \end{pmatrix} + \begin{pmatrix} 0.74 \\ (0.05) \\ 0.48 \\ (0.05) \end{pmatrix}(t-1) - \begin{pmatrix} 1.57 \\ (0.17) \\ 0.68 \\ (0.13) \end{pmatrix} s_{1,t}, \\
\Delta Z_t &= -\begin{pmatrix} 0.52 \\ (0.13) \\ 0.25 \\ (0.06) \end{pmatrix} \begin{pmatrix} 1 & \underset{(0.14)}{-1.20} \end{pmatrix} Z_{t-1} + \varepsilon_t, \text{ with } \Sigma = \begin{pmatrix} 0.63 & 0.24 \\ (0.13) & (0.06) \\ 0.24 & 0.39 \\ (0.06) & (0.05) \end{pmatrix},
\end{aligned}
\tag{7.45}
$$

$p_1 = 0.87$ (0.04) and $q_1 = 0.76$ (0.08). Figure 7.4 shows the posterior expectations of the state variable $s_{1,t}$. If we identify the regimes based on the fact whether $\mathrm{E}[s_{1,t}|Y^T]$ is smaller or larger than 0.5, we detect the following periods which correspond to $s_{1,t} = 1$: 1957.I – 1958.II, 1960.II – 1960.IV, 1966.II – 1967.IV, 1968.IV – 1970.II, 1974.III – 1975.I, 1979.IV–1980.II, 1981.IV–1983.I, 1984.IV–1987.I and 1990.II–1991.I. These periods correspond reasonably well with the low growth periods in the income and consumption series. The posterior means of the growth rates of the income series are 0.74% during an expansion regime and -0.83% $(= 0.74 - 1.57)$ during a contraction regime. For the consumption series we find 0.48% and -0.20% $(= 0.48 - 0.68)$ respectively. Hence, during recessions the negative growth rate in consumption is smaller in absolute value than the growth rate in income. To correct for this difference in the growth rates, the growth rate in income has to be larger than the growth rate in consumption during expansions.

The posterior mean of cointegration relation parameter $\beta_2 = -1.20$ does not differ more than two standard errors from one. The adjustment parameters α are both negative, which indicates that there is no adjustment towards the equilibrium for the consumption equation. This phenomenon is not due to the non-linear Markov trend in the model, since unreported results show that this also arises in a simple VAR model with and even

[5]Note that we have to incorporate the Metropolis-Hastings step in the Gibbs framework to sample from the full conditional posterior of α, see the end of Section 7.4. In this step less than 10% of the candidate draws for α were rejected.

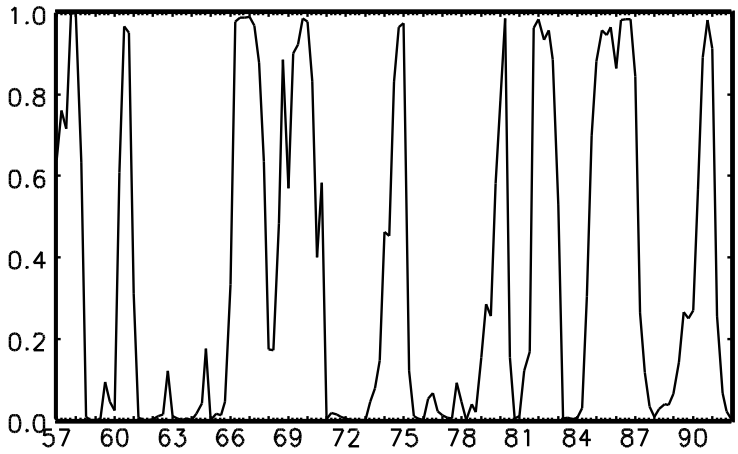

Figure 7.4. Posterior expectations of the state variable $s_{1,t}$ for model (7.45).

without a deterministic trend instead of a Markov trend, see also (7.46). Note that this does not imply that the series move away from the equilibrium, since the adjustment of income towards the equilibrium is larger than the non-adjustment in consumption, see also Johansen (1995, p. 39–42). Reduced rank Markov trend cointegration is not likely since based on the posterior means $\beta'\Gamma_1$ equals $1.57 - 1.20 \times 0.68 = 0.75$, which is relatively far away from zero.

Figure 7.5 shows the difference of the logarithm of US income and consumption. The shaded areas correspond to the periods where $\mathrm{E}[s_{1,t}|Y^T] > 0.5$. In these periods the posterior means of the slope of the Markov trend for the income series is smaller than for the consumption series, which results in a negative slope of about -0.63% $(= -0.83 + 0.20)$ in the cointegration relation. The positive slope of 0.26% in the cointegration in the other periods results from the larger positive slope of the Markov trend for the income series. The unconditional expectation of the slope of the Markov trend in the cointegration relation follows from the unconditional probability that $s_{1,t} = 1$, *i.e.* $\Pr[s_{1,t} = 1] = \pi_1 = (1 - p_1)/(2 - p_1 - q_1)$, see Section 2.2. Based on the posterior means of p_1 and q_1 of model (7.45) this unconditional expectation of the slope in the cointegration relation is about -0.03%, which is almost zero. This implies that the unconditional expectation of the slopes of the Markov trends in the per capita income and consumption series are roughly the same.

Finally, to analyse the role of the Markov trend, we consider a multivariate VAR(1) model without Markov trends, *i.e.* model (7.7) with $\Gamma_1 = \mathbf{0}$ and $k = 1$. The priors for the parameters of this model are given by (7.15)–(7.16) and (7.21). For Γ_0 we take a diffuse prior, $p(\Gamma_0) \propto 1$. The posterior means with standard deviations between parentheses of

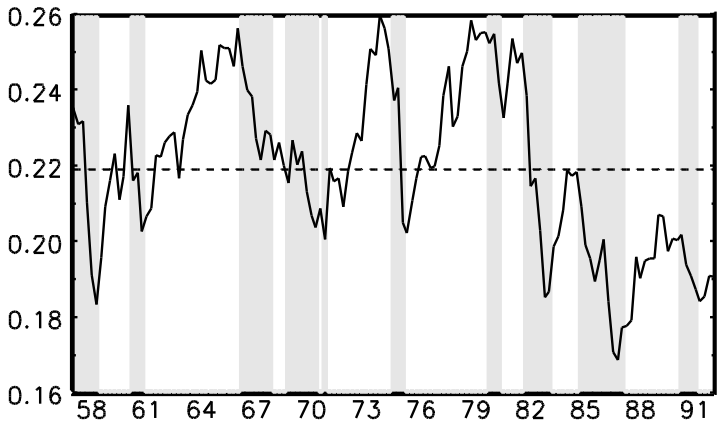

Figure 7.5. Difference between the logarithm of US per capita income and consumption, 1957.I–1992.IV. The shaded areas correspond to recessions.

this model are

$$
\begin{aligned}
Y_t &= N_t + Z_t, \\
N_t &= - \begin{pmatrix} 438.7 \\ (1.1) \\ 462.3 \\ (0.7) \end{pmatrix} + \begin{pmatrix} 0.41 \\ (0.09) \\ 0.44 \\ (0.08) \end{pmatrix} (t-1) \\
\Delta Z_t &= \begin{pmatrix} -0.17 & 0.20 \\ (0.05) & (0.06) \\ -0.06 & 0.07 \\ (0.03) & (0.04) \end{pmatrix} Z_{t-1} + \varepsilon_t, \text{ with } \Sigma = \begin{pmatrix} 1.32 & 0.52 \\ (0.16) & (0.08) \\ 0.52 & 0.49 \\ (0.08) & (0.06) \end{pmatrix}.
\end{aligned}
\tag{7.46}
$$

Removing the Markov trend from the model results in larger posterior means of the elements of the covariance matrix. The posterior means of the slopes of the deterministic trend are 0.41 for the income series and 0.44 for the consumption series. Note that the difference in the growth rate (=0.03%) is the same as the posterior mean of the unconditional expectation of the slope in the cointegration relation in the previous model. The flat prior on Π implies via decomposition (7.30) a prior on λ^*. To analyse rank reduction, we consider Bayes factors for $\lambda^* = \mathbf{0}$ and $\Pi = \mathbf{0}$. Since we have specified an uninformative prior for Π these Bayes factors are computed using the Savage-Dickey density ratio with the factor $(2\pi)^{-\frac{1}{2}(n-r)^2}$ in the denominator. The logarithm of the Bayes factors equal $\ln(\text{BF}^*(1|2)) = 4.88$ and $\ln(\text{BF}^*(0|2)) = 10.06$. Under equal prior probabilities $\Pr[\text{rank} = r] = \frac{1}{3}$, $r = 0, 1, 2$ the posterior odds ratios lead to assigning 99% posterior probability to the model with two stochastic trends. Hence, the evidence for cointegration disappears if we remove the Markov trend from the model. This may

be explained by the fact that the Markov trend in the model (7.45) implies that the cointegration relation between consumption and income does not have a constant mean, see Figure 7.5. Therefore, neglecting the Markov structure does not results in a stationary cointegration relationship between the two series.

In summary, the analysis of the presence stochastic trends in US per capita income and consumption using multivariate Markov trend models shows that specifying the Markov trend is the most difficult task. We have seen that if the model does not contain enough information for more than one Markov trend, the extra Markov trend can mimic random walk behaviour. It can also happen that overspecification of the Markov trend leads to modelling one or more outlying observations as a regime. Considering these outlying observations as a possible regime in the Markov trend may result in a better description of the multivariate time series, but is not likely to result in better forecasts since it leads to larger forecast uncertainty. If one however has strong beliefs about the presence of more than one Markov trend, it is better to impose informative priors to identify the regimes. Prior information can also enter the model by imposing *a priori* more structure on the multivariate Markov trend, for instance imposing that a regime change in one series happens one period earlier than in the other series or limiting the number of transitions between the states by restricting the transition probabilities, see also Dwyer and Potter (1996) for other suggestions.

7.7 Concluding Remarks

In this chapter we have presented a multivariate version of the Markov trend model of Chapter 3. The model incorporates a multivariate Markov trend specification, which allows for different stochastic Markov trends for the multivariate time series. The deviations from the multivariate Markov trend are modelled by a vector autoregressive model. Unit roots in this VAR specification also introduce stochastic trends. To analyse the number of stochastic trends, we have proposed a Bayesian framework based on the ideas in Chapter 6. The practical usefulness of this framework has been shown in stochastic trend analysis of three simulated series and in the analysis of seasonally adjusted US per capita income and consumption. The latter two series seem to have one common Markov trend plus one non-Markov common stochastic trend. The posterior results of the Markov trend show that during recessions the negative growth rate of consumption is smaller in absolute value than the growth rate of income. This is compensated by a larger positive growth rate in income than in consumption during the expansion periods. Removing the common Markov trend from the model suggest the presence of two stochastic trends and therefore no cointegration relation amongst the two series.

The model in this chapter can be extended to analyse seasonally unadjusted series, like in Chapter 4. Therefore, we need to extend the analysis of unit roots in vector autoregressive model to include the presence of possible seasonal unit roots and seasonal cointegration, see Engle *et al.* (1993) and Lee (1992). Other extensions include Markov switching

cointegration, where the parameters modelling the cointegration relations and/or adjustment parameters are different in the regimes like in threshold cointegration, see Balke and Fomby (1997).

Chapter 8

Summary and Conclusions

In this thesis we analysed trends in quarterly observed macroeconomic times series using Markov trend models. The analysis has been partitioned in two parts. In the first part we have focused on modelling univariate time series. The analysis of trends in multivariate time series has been considered in the second part.

The basic model in the first part of the thesis consists of a Markov trend where the deviations from the Markov trend are assumed to be an autoregressive process. The Markov trend is a stochastic trend where the direction of the slope depends on the value of an unobserved two-state first-order Markov process. Since the Markov trend allows for two possible slope values, the Markov trend is more flexible than a linear deterministic trend but more restricted than a random walk plus drift specification. Therefore, the Markov trend has less forecast uncertainty than the random walk.

The autoregressive deviations from the Markov trend may also results in the presence of a stochastic trend in the model. In most studies a unit root in the autoregressive polynomial is imposed, which implies that one assumes that the series can be described by a Markov trend plus a random walk. Sometimes this trend specification may be too variable to describe the trend in a time series and results in too much forecasts uncertainty. For instance, in Chapter 3 we show that quarterly observed seasonally adjusted German industrial production can be better described within-sample and out-of-sample by a Markov trend stationary model.

The two slopes of the Markov trend can be linked with the stage of the business cycle to model the growth rate in recession and expansion periods. The transition probabilities of the Markov process determine the expected duration of recessions and expansions and inference about regimes changes can be used to determine turning points. Since the Markov trend models the trend and the business cycle simultaneously, it is important for business cycle inference to have a proper specification of the trend of the time series in the Markov trend model. Furthermore, since the idea behind the Markov trend is that recessions are recurrent events, it is necessary for business cycle analysis that the characteristics of every recession are roughly the same. If this is not the case, the Markov trend may not capture all the recession periods. For instance, the empirical example of Chapter 3 shows that for German industrial production the Markov trend model does not

capture the recession in the 1960s. Contrary to the other recession periods, this recession was followed by a period of very high growth to reach the same growth path as before the recession. This problem may be solved by allowing for an extra regime in the Markov trend, but inference about the parameters modelling this extra regime will be totally based on the information in the recession period in the 1960s.

Business cycle analysis is usually based on seasonally adjusted time series. Seasonally adjusted series are often constructed from seasonally unadjusted series via a sequence of moving average operations. Therefore, some observations of a seasonal adjusted series may consist of a weighted average of recession and expansion observations. This may influence the inference on the business cycle, especially the dating of turning points. Additionally, many studies show that the business cycle and the seasonal fluctuations are not independent, which suggest that seasonal variation may contain valuable information about the business cycle. To model the business cycle and the seasonal variation simultaneously, we have extended the Markov trend model with a seasonal component. The seasonal component consists of seasonal dummies and incorporates the possibility that seasonal mean shifts may coincide with changes in the stage of the business cycle. Furthermore, we allow for changes in the seasonal pattern due to seasonal stochastic trends. The analysis of quarterly observed German unemployment with the (seasonal) Markov trend model shows that there is a difference in the estimated duration of recession and expansion periods and in the dating of turning points for the seasonally adjusted and unadjusted series. After correcting for different growth rates during recession and expansion periods, the seasonal pattern in the unemployment series seems to be constant.

In the second part of the thesis we have considered a multivariate version of the univariate Markov trend model for the analysis of multivariate time series. The model consist of a multivariate Markov trend and the deviations from this trend follow a vector autoregressive process. The multivariate Markov trend specification consists of correlated univariate Markov trends. Just as in the univariate model, imposing unit roots in the vector autoregressive polynomial corresponds to the assumption of random walk type stochastic trends in the multivariate time series.

Just as in the univariate case, the slopes of the Markov trend can be linked with the stages of the business cycle. It allows for different business cycles in every univariate series, for instance different turning points and durations of recession and expansion periods. However, there may be not enough information in the data to analyse a general multivariate Markov trend. In practice, it may happen that one of the Markov trends mimics the behaviour of a random walk stochastic trend, as the example in Chapter 7 shows, or that it models one or two more or less outlying observations as a regime. Although the latter possibility is interesting, this indicates an overspecification of the Markov trend. To limit the number of Markov trends, we have considered common Markov trends. Common Markov trends occur if a linear combination of two series, which contain a Markov trend, does not contain a Markov trend. Additionally, we have analysed the number of (common) stochastic trends resulting from unit roots in the vector autoregressive part of the model. The determination of the number of Markov trend can be based on inference,

but if the data do not contain enough information it may be wishful to impose *a priori* some structure, like common Markov trends. As a special case we have considered the situation of common business cycles, where the turning points and hence the expected duration of recession and expansion periods in time series are the same. The empirical illustration in Chapter 7 suggest that quarterly observed seasonally adjusted per capita income and consumption of the United States have one common Markov trend, which can be interpreted as a common business cycle and one common random walk trend. The posterior results of the Markov trend show that during recessions the negative growth rate of consumption is smaller in absolute value than the growth rate of income. This is compensated by a larger positive growth rate in income than in consumption during the expansion periods. If we remove the Markov trend from the model there is no posterior evidence for common stochastic trends in the series anymore.

For the inference in the Markov trend models we have chosen a Bayesian approach. The inference has been based on a relatively diffuse prior specification to let the data information in the likelihood dominate the inference. Informative priors can easily be included in the analysis. However, one has to be careful for including strong prior information, which is not in accordance with the information in the data. Since the number of observations in the recession periods is relatively small, one may expect that posterior results for parameters which model the recessions are very sensitive to prior information for these parameters. The determination of the appropriate trend specification in the time series has been based on posterior odds ratios. For the univariate Markov trend models, we have build on available Bayesian methods, while for the multivariate models, we have developed a new Bayesian framework to analyse the number of stochastic trends. The practical usefulness of these posterior odds ratios has been illustrated with simulated examples. A topic for further research is to analyse the sensitivity of the results with respect to several prior specifications.

Alternative specifications to model business cycle characteristics are threshold autoregressive [TAR] models, see *e.g.* Potter (1995) and smoothed threshold autoregressive [STAR] models, see *e.g.* Teräsvirta and Anderson (1992). Instead of transition probabilities like in the Markov trend models, a change in regime occurs if a lagged value of the times series exceeds a certain threshold level. These models allow for different dynamic structure during recession and expansion periods for the first or fourth differences of the series and permit even explosive dynamic behaviour in some regimes. A regime dependent constant is usually added in a linear way to the autoregressive model, so that there is no explicit modelling of the trend. A nice interpretation of different growth rates during recession and expansion periods as in the Markov trend models is therefore not possible. Due to the complex structure of these models, the characteristics of the estimated threshold model usually have to be analysed using Monte Carlo simulations.

Just as in the threshold models we may allow for different dynamic structures in recession and expansion regimes in the Markov trend models considered in this thesis. It remains however to be seen whether there is enough information in the relatively small number of recession observations for this extensions, since this may lead to a substantial

increase in the number of parameters modelling the recession. A more promising strategy to extend the basic Markov trend model is to consider time varying transition probabilities. Durland and McCurdy (1994) consider transition probabilities, which depend on the number of periods that the process has been in the regime. Filardo (1994) and Diebold, Lee and Weinbach (1994) consider logistic functions of explanatory variables to model the transition probabilities. Possible explanatory variables are functions of the lagged values of the time series or exogenous variables, for instance leading indicators. For a multivariate modelling of macroeconomic time series using Markov trend models, it seems better to develop *a priori* model structures than to simplify general models via a sequence of tests. Model structures, which model common business cycles as for instance in Kim and Yoo (1995) or where regime switches in one variable help to forecast regime switches in other variables seem interesting, see *e.g.* Phillips (1991a).

In this thesis we have modelled the deviations from the trend with autoregressive processes with uncorrelated normal distributed errors. Extensions to moving average models and to *e.g.* t-distributed errors are of course possible. Geweke (1993) provides Bayesian techniques to include t-distributed errors in a time series model. It will however make the Markov trend less important since under the assumption of t-distributed errors deviations from a deterministic linear trend are more likely, see Hoek, Lucas and van Dijk (1995) for a similar result. Kleibergen and Hoek (1995) and Chib and Greenberg (1994) show that Bayesian analysis of univariate time series models with moving average errors is feasible. Extending the vector autoregressive model with moving average errors is still a topic for further research. Furthermore, it seems interesting to examine how in a Bayesian analysis the initial observations in autoregressive moving average models have to be treated.

Nederlandse Samenvatting (Summary in Dutch)

Veel macro-economische tijdreeksen, zoals industriële productie en bruto nationaal product, worden gekarakteriseerd door lange perioden met positieve groei, de expansie perioden, en korte perioden met negatieve groei, de recessies. Wanneer de korte recessie perioden gezien worden als tijdelijke afwijkingen van de lange termijn positieve groei, dan wordt de lange termijn groei in deze tijdreeksen meestal beschreven met een lineaire deterministische trend. Vaak blijkt deze lineaire trendspecificatie met een constante groeivoet te restrictief om de trend in de tijdreeks te beschrijven, bijvoorbeeld wanneer de tijdelijke afwijkingen van de lange termijn groei een permanente invloed hebben op de toekomstige waarde van de tijdreeks. Daarom wordt de deterministische lineaire trend vaak uitgebreid.

Een populaire uitbreiding van de lineaire trend is het toevoegen van een kansvariable in de trendspecificatie, zodat de richting van de trend in elke periode niet deterministisch maar stochastisch is. De bekendste specificatie in dit verband is de stochastische wandeling. Deze bestaat uit een som van ongecorreleerde schokken, die voortkomen uit een continue kansverdeling. Schokken in elke periode hebben een permanente invloed op toekomstige waarden van de tijdreeks. Er wordt bij deze trendspecificatie echter geen onderscheid gemaakt tussen de invloed van schokken tijdens recessie en expansie perioden. Een alternatieve trendspecificatie, die schokken tijdens recessies en expansies wel asymmetrische behandelt, is de Markov trend. Dit is een stochastische trend, waarbij de richting van de trend bepaald wordt door een niet-waargenomen eerste orde Markov proces met twee toestanden. De overgangskansen van het Markov proces modelleren de overgangen tussen recessies en expansies. Dit impliceert dat de huidige richting van de Markov trend afhankelijk is van de richting van de trend in de vorige periode, hetgeen niet van toepassing is voor de stochastische wandeling. Aangezien het aantal mogelijke richtingen van de Markov trend ook nog beperkt is tot twee, is deze minder flexibel dan de stochastische wandeling, die een continu scala aan mogelijke richtingen van de trend toelaat.

Naast het modelleren van de trend in macro-economische tijdreeksen, is de Markov trend ook geschikt voor het analyseren van conjunctuurcycli. Aangezien de twee richtingen van de Markov trend corresponderen met de groeivoeten in recessie en expansie perioden, kan de analyse van regime veranderingen gebruikt worden voor het bepalen van omslagpunten. De overgangskansen van het Markov proces bepalen de verwachte duur

van recessie en expansie perioden. Voor het modelleren van de afwijkingen van de trend wordt meestal een autoregressief proces gekozen. Vaak modelleert men echter de eerste verschillen van de afwijkingen met een autoregressief proces, hetgeen komt overeen met veronderstelling van de aanwezigheid van een Markov trend plus een stochastische wandeling in de tijdreeks. Het analyseren van de lengte en de omslagpunten van de conjunctuurcyclus blijkt afhankelijk te zijn van de trendspecificatie. Dit is niet verwonderlijk aangezien de trend en de conjunctuurcyclus door de Markov trend simultaan gemodelleerd worden, en het toevoegen van een extra stochastische trend de schattingsresultaten van de Markov trend beïnvloedt. Een adequate trendspecificatie is daarom gewenst.

In dit proefschrift beschouwen we methoden voor het analyseren van trends in macro-economische tijdreeksen. We beperken ons tot de hiervoor besproken trendspecificaties. De analyse vindt plaats vanuit een Bayesiaans perspectief. In tegenstelling tot een klassieke benadering, wordt bij een Bayesiaanse analyse de onzekerheid in parameters meegenomen, bijvoorbeeld bij het bepalen van omslagpunten van de conjunctuurcyclus en bij het bepalen van voorspelonzekerheid. De Bayesiaanse analyse in dit proefschrift is gebaseerd op niet-informatieve prioren en kan gezien worden als uitgebreide analyse van de informatie in de aannemelijkheidsfunctie. De onzekerheid in de uitkomsten wordt gedomineerd door de variatie in de tijdreeksen zonder grote invloed van priorspecificatie. Informatieve prioren kunnen echter op eenvoudige wijze worden ingebracht in de analyse.

Het proefschrift is opgebouwd uit twee delen. In het eerste deel beschouwen we het modelleren van de trend in univariate tijdreeksen. Het tweede deel behandelt trends in multivariate tijdreeksen, waarbij we ons concentreren op het analyseren van gemeenschappelijke trends. De inleiding en motivatie voor het proefschrift worden gegeven in Hoofdstuk 1. Bovendien bevat dit hoofdstuk een kort literatuuroverzicht van het gebruik van Markov processen bij het modelleren van macro-economische en financiële tijdreeksen.

Hoofdstuk 2 geeft een korte inleiding in het modelleren van de trend in univariate macro-economische tijdreeksen. We beschrijven verschillende veel gebruikte trendspecificaties, waaronder de lineaire deterministische trend, de stochastische wandeling en de Markov trend. Voor de genoemde trendspecificaties wordt de nauwkeurigheid van het beschrijven van de trend binnen de steekproef vergeleken met voorspelonzekerheid. We laten zien dat een trendspecificatie, die in elke periode een groot aantal mogelijke richtingen van de trend toelaat, resulteert in een grotere voorspelonzekerheid dan een specificatie met slechts een beperkt aantal mogelijke richtingen. Daarom zal onderspecificatie van de trend in een tijdreeks ten onrechte leiden tot een kleine voorspelonzekerheid en een te flexibele beschrijving van de trend binnen de steekproef tot een te grote voorspelonzekerheid. Tot slot besteden we aandacht aan het modelleren van de afwijkingen van de trend door middel van autoregressieve modellen. We laten zien dat een eenheidswortel in de autoregressieve structuur de aanwezigheid van een stochastische trend, te weten een stochastische wandeling, in de tijdreeks impliceert.

In Hoofdstuk 3 beschouwen we het univariate Markov trend model. Dit model bestaat uit een univariate Markov trend, waarbij de afwijkingen van de trend worden gemodelleerd door een autoregressief model. We bespreken de aanwezigheid van stochastische trends en

de invloed van exogene schokken op het niveau en de eerste verschillen van de reeks onder verschillende trendspecificaties in het model. Voor het analyseren van de trendcomponent in het model stellen we een Bayesiaanse strategie voor. We beschouwen priorspecificatie, een simulatiemethode voor het verkrijgen van posterior resultaten en voorspelverdelingen, en posterior kansen voor de aanwezigheid van stochastische trends. Meestal wordt bij het modelleren van macro-economische tijdreeksen met behulp van een Markov trend model veronderstelt dat de reeks een Markov trend plus een stochastische wandeling bevat. Dit komt tot uitdrukking in het analyseren van de eerste verschillen van de reeks. Het analyseren van per kwartaal waargenomen seizoensgecorrigeerde industriële productie van Duitsland toont aan dat deze reeks beter beschreven kan worden door een stationair autoregressief model rond een Markov trend. Deze trendspecificatie resulteert in betere voorspellingen en minder voorspelonzekerheid dan de hiervoor genoemde veel gebruikte specificatie. Bovendien leert de analyse van Duitse tijdreeks ons dat het modelleren van recessies met een Markov trend vereist dat de karakteristieken van alle recessies ongeveer hetzelfde zijn. Voor Duitse industriële productie blijkt de Markov trend niet in staat om de recessie in de jaren zestig te traceren. In tegenstelling tot de andere recessies wordt deze recessies gevolgd door een periode van zeer snelle groei, waardoor de reeks weer terugkeert naar het oude groeipad van voor de recessie.

Het analyseren van de conjunctuurcyclus vindt meestal plaats met behulp van seizoensgecorrigeerde tijdreeksen, aangezien men verwacht dat het sterk dominerende seizoenspatroon het dateren van omslagpunten bemoeilijkt. Bij populaire seizoenscorrectiemethoden wordt de huidige waarde van de tijdreeks vervangen door een gewogen gemiddelde van toekomstige waarnemingen en waarnemingen uit het verleden. Dus ook waarnemingen die corresponderen met omslagpunten worden vervangen door een gewogen gemiddelde van waarnemingen die corresponderen met expansie en recessie perioden. Dit kan leiden tot een incorrecte datering van omslagpunten op basis van seizoensgecorrigeerde reeksen. In Hoofdstuk 4 onderzoeken we de invloed van seizoenscorrectie op het dateren van omslagpunten in de conjunctuurcyclus. Hiervoor wordt het univariate Markov trend model uit Hoofdstuk 3 uitgebreid met een seizoenscomponent bestaande uit seizoensdummies. Voor het modelleren van mogelijke veranderingen in het seizoenspatroon over de tijd, maken we gebruik van stochastische seizoenstrends, die ontstaan door de aanwezigheid van seizoenseenheidswortels in het autoregressieve deel van het Markov trend model. Aangezien veranderingen in het seizoenspatroon kunnen samenvallen met de omslagpunten in de conjunctuurcyclus bevat het model ook verschillende seizoensgemiddelden tijdens recessie en expansie perioden. Het uitgebreide Markov trend model biedt de mogelijkheid voor een simultane modellering van het seizoenspatroon en de conjunctuurcyclus. Om veranderingen in het seizoenspatroon en regime veranderingen goed uit elkaar te houden, is voor het analyseren van de conjunctuurcyclus naast een goede trendbeschrijving ook een adequate beschrijving van het seizoenspatroon nodig. De Bayesiaanse methoden uit Hoofdstuk 3 worden uitgebreid voor het analyseren van de aanwezigheid van stochastische seizoenstrends. De analyse van per kwartaal waargenomen Duitse werkloosheid toont verschillen in de datering van omslagpunten en in de verwachte lengte van

de conjunctuurcyclus voor de seizoensgecorrigeerde en de ongecorrigeerde tijdreeks. Voor de ongecorrigeerde reeks geldt bovendien dat na correctie voor verschillende groeivoeten tijdens recessie en expansie perioden er geen significante verandering in het seizoenspatroon te ontdekken valt.

In het tweede deel van het proefschrift beschouwen we het modelleren van trends in multivariate tijdreeksen. Hoofdstuk 5 geeft een korte inleiding in multivariate trendspecificaties. Deze trendspecificaties zijn multivariate generalisaties van de univariate trendspecificaties uit Hoofdstuk 2, waaronder een multivariate stochastische wandeling en een multivariate Markov trend. Speciale aandacht wordt besteed aan voorwaarden voor gemeenschappelijke trends tussen univariate tijdreeksen. We spreken van een gemeenschappelijke trend in twee of meer tijdreeksen als een lineaire combinatie van deze tijdreeksen, die ieder univariaat een bepaalde trend bevatten, deze trend niet meer bevat. Net zoals in de inleiding van het univariate gedeelte sluiten we af met het modelleren van de afwijkingen van de multivariate trendspecificatie met behulp van vector autoregressieve modellen. Eenheidswortels in de autoregressieve structuur impliceren de aanwezigheid van stochastische trends (multivariate stochastische wandelingen) in de multivariate tijdreeksen. Wanneer het aantal eenheidswortels in het multivariate model kleiner is dan het totaal aantal eenheidswortels in de univariate tijdreeksen, dan is er sprake van gemeenschappelijke stochastische trends. Dit verschijnsel staat bekend onder de term cointegratie.

In tegenstelling tot univariate analyse van stochastische trends veroorzaakt door eenheidswortels in de autoregressieve structuur, bestaat er nog geen makkelijk toepasbare Bayesiaanse benadering voor het analyseren van stochastische trends in vector autoregressieve modellen. Daarom stellen we in Hoofdstuk 6 een Bayesiaanse methode voor cointegratie-analyse in multivariate modellen voor. Deze methode is gebaseerd op een nieuwe decompositie van de parametermatrix die de foutencorrectie modelleert. We behandelen priorspecificatie, simulatie-algoritme voor het verkrijgen van posterior resultaten en posterior kansen voor het aantal (gemeenschappelijke) stochastische trends in tijdreeksen. Bovendien stellen we een Bayesiaanse versie van een Lagrange Multiplier toetsgrootheid voor. Ter illustratie analyseren we twee bekende vector autoregressieve modellen, die de Deense en Engelse geldvraag modelleren.

De in Hoofdstuk 6 ontwikkelde methode voor de analyse van stochastische trends blijkt redelijk eenvoudig uit te breiden naar complexere modellen. In Hoofdstuk 7 passen we de methode toe op een multivariate generalisatie van het Markov trend model uit Hoofdstuk 3. Dit model is vector autoregressief rond een multivariate Markov trendspecificatie. We behandelen de invloed van exogene schokken onder verschillende specificaties van het multivariate Markov model. De informatie in tijdreeksen kan te beperkt zijn om een algemeen multivariaat Markov trend te analyseren. Overspecificatie van de Markov trendcomponent kan leiden tot het modelleren van een of meer uitschieters in de tijdreeksen in plaats van een zinnig regime, of de Markov trend kan zelfs het gedrag van een stochastische wandeling imiteren. Het lijkt daarom wenselijk om *a priori* zinnige modelstructuren te specificiëren. Speciale aandacht wordt besteed aan gemeenschappelijke Markov trends,

cointegratie en Markov trend cointegratie. Voor het analyseren van deze gemeenschappelijke trends worden de Bayesiaanse technieken en methoden uit de Hoofdstukken 3 en 6 aangepast en uitgebreid. Net zoals in het univariate model kan de multivariate Markov trend worden gekoppeld aan de conjunctuurcyclus. Een gemeenschappelijke Markov trend kan worden geïnterpreteerd als een gemeenschappelijke conjunctuurcyclus in tijdreeksen. We concluderen op basis van posterior resultaten van het multivariate Markov trend model dat er tussen per capita inkomen en consumptie van de Verenigde Staten een gemeenschappelijke Markov trend en een cointegratierelatie bestaat. De posterior resultaten van de gemeenschappelijke Markov trend laten zien dat tijdens recessies de negatieve groeivoet van het inkomen in absolute waarde kleiner is dan de groeivoet van consumptie. Dit wordt gecompenseerd door een grotere positieve groeivoet in het inkomen dan in consumptie tijdens expansie perioden. Het belang van de Markov trend blijkt uit het feit dat er na verwijdering van de Markov trendspecificatie uit het model geen gemeenschappelijke trend in beide tijdreeksen wordt gevonden.

Hoofdstuk 8 geeft een overzicht van de belangrijkste conclusies die uit het proefschrift volgen. Het Markov trend model blijkt een bruikbaar model voor het modelleren van trends in macro-economische tijdreeksen. Bovendien kan het model ook gebruikt worden voor het analyseren van conjunctuurcycli. Voor het opsporen van recessies is het echter van belang dat de karakteristieken van recessies ongeveer gelijk zijn. Seizoenscorrectie beïnvloedt het dateren van omslagpunten en het analyseren van de verwachte lengte van recessie en expansie perioden met Markov trend modellen. Multivariate Markov trend modellen kunnen worden gebruikt voor het modelleren van verschillen in groeivoeten van variabelen tijdens recessie en expansie perioden en voor het analyseren van gemeenschappelijke conjunctuurcycli. Dankzij moderne simulatietechnieken blijkt een Bayesiaanse analyse van macroeconomische tijdreeksen met behulp van Markov trend modellen zeer geschikt. Tot slot bespreken we in het laatste hoofdstuk de gevoeligheid van de uitkomsten met betrekking tot priorspecificatie, alternatieve modelstructuren zoals drempelwaarde modellen en besteden aandacht aan mogelijke richtingen voor toekomstig onderzoek en modelextensies, waaronder het afhankelijk maken van de overgangskansen van het Markov proces van bijvoorbeeld exogene variabelen, vertraagde endogene variabelen of het aantal perioden dat men al in een recessie of expansie zit.

Bibliography

Albert, J.H. and S. Chib, 1993, Bayes Inference via Gibbs Sampling of Autoregressive Time Series subject to Markov Mean and Variance Shifts, *Journal of Business & Economic Statistics* **11**, 1–15.

Balke, N.S. and T.B. Fomby, 1997, Threshold Cointegration, *International Economic Review* **38**, 627–646.

Banerjee, A., J. Dolado, J.W. Gailbraith and D.F. Hendry, 1993, *Co-Integration, Error Correction, and the Econometric Analysis of Non-Stationary Data*, Oxford University Press, Oxford.

Beaudry, P. and G. Koop, 1993, Do Recessions Permanently Change Output?, *Journal of Monetary Economics* **31**, 149–163.

Beveridge, S. and C.R. Nelson, 1981, A New Approach to Decomposition of Economic Time Series into Permanent and Transitory Components with Particular Attention to Measurement of the Business Cycle, *Journal of Monetary Economics* **7**, 151–174.

Billio, M., A. Monfort and C.P. Robert, 1996, Bayesian Estimation of Switching ARMA Models, *Série des Documents de Travail du Crest 9650*, Centre de Recherche en Economie et Statistique.

Boldin, M.D., 1996, A Check on the Robustness of Hamilton's Markov Switching Approach to the Economic Analysis of Business Cycles, *Studies in Nonlinear Dynamics and Econometrics* **1**, 35–46.

Bonomi, M. and R. Garcia, 1994, Can a Well-Fitted Equilibrium Asset-Pricing Model Produce Mean Reversion?, *Journal of Applied Econometrics* **9**, 19–29.

Box, G.E.P. and G.C. Tiao, 1973, *Bayesian Inference in Statistical Analysis*, Addison-Wesley, Reading, MA.

Cai, J., 1994, A Markov Model of Switching-Regime ARCH, *Journal of Business & Economic Statistics* **12**, 309–316.

Canova, F. and E. Ghysels, 1994, Changes in Seasonal Patterns: Are They Cyclical?, *Journal of Economic Dynamics and Control* **18**, 1143–1171.

Ceccheti, S.G., P-S. Lam and N.C. Mark, 1990, Mean Reversion in Equilibrium Asset Prices, *American Economic Review* **80**, 398–418.

Chen, M-H., 1994, Importance-Weighted Marginal Bayesian Posterior Density Estimation, *Journal of the American Statistical Association* **89**, 818–824.

Chib, S., 1996, Calculating Posterior Distributions and Model Estimates in Markov Mixture Models, *Journal of Econometrics* **75**, 79–97.

Chib, S. and E. Greenberg, 1994, Bayes Inference in Regression Models with ARMA(p,q) Errors, *Journal of Econometrics* **64**, 183–206.

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

Chu, C-S. and H. White, 1992, A Direct Test for Changing Trend, *Journal of Business & Economic Statistics* **10**, 289–299.

Davidson, R. and J.G. MacKinnon, 1993, *Estimation and Inference in Econometrics*, Oxford University Press, Oxford.

DeJong, D.N., 1992, Co-Integration and Trend Stationarity in Macroeconomic Time Series: Evidence from the Likelihood Function, *Journal of Econometrics* **52**, 347–370.

DeJong, D.N. and C.H. Whiteman, 1991, Reconsidering "Trends and Random Walks in Macroeconomic Time Series", *Journal of Monetary Economics* **28**, 221–254.

Dickey, D.A. and W.A. Fuller, 1979, Distribution of Estimators for Autoregressive Time Series with a Unit Root, *Journal of the American Statistical Association* **74**, 427–431.

Dickey, J., 1971, The Weighted Likelihood Ratio, Linear Hypothesis on Normal Location Parameters, *The Annals of Mathematical Statistics* **42**, 240–223.

Diebold, F.X. and G.D. Rudebusch, 1996, Measuring Business Cycles: A Modern Perspective, *Review of Economics and Statistics* **78**, 67–77.

Diebold, F.X., J. Lee and G.C. Weinbach, 1994, Regime Switching and Endogenous Transition Probablities, in C.P. Hargreaves (editor), *Nonstationary Time Series Analysis and Cointegration*, Oxford University Press, Oxford, chapter 10, pp. 283–302.

Diebolt, J. and C.P. Robert, 1994, Estimation of Finite Mixture Distributions Through Bayesian Sampling, *Journal of the Royal Statistical Society B* **56**, 363–375.

Doan, T., R.B. Litterman and C. Sims, 1984, Forecasting and Conditional Projections using Realistic Prior Distributions, *Econometric Reviews* **3**, 1–144.

Dorfman, J.H., 1995, A Numerical Bayesian Test for Cointegration of AR processes, *Journal of Econometrics* **66**, 289–234.

Dueker, M.J., 1997, Markov Switching in GARCH Processes and Mean-Reverting Stock-Marker Volatility, *Journal of Business & Economic Statistics* **15**, 26–34.

Durland, J.M. and T.H. McCurdy, 1994, Duration-Dependent Transitions in a Markov Model, *Journal of Business & Economic Statistics* **12**, 279–288.

Dwyer, M.J. and S.M. Potter, 1996, Markov Trend Cointegration, unpublished working paper, University of California.

Engel, C., 1994, Can the Markov Switching Model Forecast Exchange Rates?, *Journal of International Economics* **36**, 151–165.

Engel, C. and J.D. Hamilton, 1990, Long Swings in the Dollar: Are They in the Data and Do Markets Know It?, *American Economic Review* **80**, 689–713.

Engle, R.F. and C.W.J. Granger, 1987, Co-Integration and Error Correction: Representation, Estimation, and Testing, *Econometrica* **55**, 251–276.

Engle, R.F., C.W.J. Granger, S. Hylleberg and H.S. Lee, 1993, Seasonal Cointegration: The Japanese Consumption Function, *Journal of Econometrics* **55**, 275–298.

Falk, B., 1986, Further Evidence on the Asymmetric Behavior of Economic Time Series over the Business Cycle, *Journal of Political Economy* **94**, 1096–1109.

Filardo, A.J., 1994, Business-Cycle Phases and Their Transitional Dynamics, *Journal of Business & Economic Statistics* **12**, 299–308.

Franses, P.H., 1996, *Periodicity and Stochastic Trends in Economic Time Series*, Oxford University Press, Oxford.

Franses, P.H., H. Hoek and R. Paap, 1997, Bayesian Analysis of Seasonal Unit Roots and Seasonal Mean Shifts, *Journal of Econometrics* **78**, 359–380.

Franses, P.H. and R. Paap, 1994, Model Selection in Periodic Autoregressions, *Oxford Bulletin of Economics and Statistics* **56**, 421–439.

Franses, P.H. and R. Paap, 1996, Does Seasonal Adjustment Change Inference from Markov Switching Models, *Econometric Institute Report 9615/A*, Erasmus University Rotterdam.

Franses, P.H. and T.J. Vogelsang, 1998, On Seasonal Cycles, Unit Roots, Mean Shifts, *Review of Economics and Statistics*, to appear.

Fuller, W.A., 1976, *Introduction to Statistical Time Series*, Wiley, New York.

Garcia, R., 1995, Asymptotic Null Distribution of the Likelihood Ratio Test in Markov Switching Models, *CIRANO Scientific Series 95s-7*.

Garcia, R. and P. Perron, 1996, An Analysis of the Real Interest Rate under Regime Shifts, *Review of Economics and Statistics* **78**, 111–125.

Geman, S. and D. Geman, 1984, Stochastic Relaxations, Gibbs Distributions, and the Bayesian Restoration of Images, *IEEE Transaction on Pattern Analysis and Machine Intelligence* **6**, 721–741.

Geweke, J., 1989, Bayesian Inference in Econometric Models Using Monte Carlo Integration, *Econometrica* **57**, 1317–1339.

Geweke, J., 1993, Bayesian Treatment of the Independent Student-t Linear Model, *Journal of Applied Econometrics* **8**, 519–540.

Geweke, J., 1996, Bayesian Reduced Rank Regression in Econometrics, *Journal of Econometrics* **75**, 121–146.

Ghysels, E., 1994, On the Periodic Structure of the Business Cycle, *Journal of Business & Economic Statistics* **12**, 289–298.

Ghysels, E., 1997, On Seasonality and Business Cycle Durations: A Nonparametric Investigation, *Journal of Econometrics* **79**, 269–290.

Ghysels, E., C.W.J. Granger and P.L. Siklos, 1996, Is Seasonal Adjustment a Linear or Nonlinear Data Filtering Process?, *Journal of Business & Economic Statistics* **14**, 374–386.

Ghysels, E. and P. Perron, 1996, The Effects of Linear Filters on Dynamic Time Series with Structural Change, *Journal of Econometrics* **70**, 69–98.

Ghysels, E., R.E. McCulloch and R.S. Tsay, 1994, Bayesian Inference for Periodic Regime-Switching Models, unpublished working paper, University of Montreal.

Goldfeld, S.M. and R.M. Quandt, 1973, A Markov Model for Switching Regressions, *Journal of Econometrics* **1**, 3–16.

Goodwin, T.H., 1993, Business-Cycle Analysis with a Markov-Switching Model, *Journal of Business & Economic Statistics* **11**, 331–339.

Gordon, S., 1997, Stochastic Trends, Deterministic Trends and Business Cycle Turning Points, *Journal of Applied Econometrics* **12**, 411–434.

Granger, C.W.J., 1989, *Forecasting in Business and Economics*, 2nd edition, Academic Press, San Diego.

Granger, C.W.J. and P. Newbold, 1987, *Forecasting Economic Time Series*, 2nd edition, Academic Press, San Diego.

Hamilton, J.D., 1988, Rational-Expectations Econometric Analysis of Changes in Regime: An investigation of the Term Structure of Interest Rates, *Journal of Economic Dynamics and Control* **12**, 385–423.

Hamilton, J.D., 1989, A New Approach to the Econometric Analysis of Nonstationary Time Series and Business Cycles, *Econometrica* **57**, 357–384.

Hamilton, J.D., 1990, Analysis of Time Series Subject to Changes in Regime, *Journal of Econometrics* **45**, 39–70.

Hamilton, J.D., 1994, *Time Series Analysis*, Princeton University Press, New Jersey.

Hamilton, J.D. and G. Lin, 1996, Stock Market Volatility and the Business Cycle, *Journal of Applied Econometrics* **11**, 573–593.

Hamilton, J.D. and R. Susmel, 1994, Autoregressive Conditional Heteroskedasticity and Changes in Regime, *Journal of Econometrics* **64**, 307–333.

Hansen, B.E., 1992, The Likelihood Ratio Test under Nonstandard Conditions: Testing the Markov Switching Model of GNP, *Journal of Applied Econometrics* **45**, S61–S82.

Hansen, B.E., 1996, Erratum: The Likelihood Ratio Test Under Nonstandard Conditions: Testing the Markov Switching Model of GNP, *Journal of Applied Econometrics* **11**, 195–198.

Harvey, A.C., 1989, *Forecasting Structural Time Series and the Kalman Filter*, Cambridge University Press, Cambridge.

Hastings, W.K., 1970, Monte Carlo Sampling using Markov Chains and their Applications, *Biometrika* **57**, 97–10.

Hendry, D.F. and J.A. Doornik, 1994, Modelling Linear Dynamic Econometrics Systems, *Scottish Journal of Political Economy* **41**, 1–33.

Hendry, D.F. and N.R. Ericsson, 1991, Modeling the Demand for Narrow Money in the United Kingdom and the United States, *European Economic Review* **35**, 833–886.

Hoek, H., 1997, *Variable Trends: A Bayesian Perspective*, PhD thesis, Tinbergen Institute, Thesis Publishers, Amsterdam.

Hoek, H., A. Lucas and H.K. van Dijk, 1995, Classical and Bayesian Aspects of Robust Unit Root Inference, *Journal of Econometrics* **69**, 27–59.

Hoek, H. and R. Paap, 1994, Bayesian Unit Root Inference in the Hamilton Model, *Tinbergen Institute discussion paper 94–142*, Erasmus University Rotterdam.

Holst, U., G. Lindgren, J. Holst and M. Thuvesholmen, 1994, Recursive Estimation in Switching Autoregressions with a Markov Regime, *Journal of Time Series Analysis* **15**, 489–506.

Hylleberg, S., C. Jørgensen and N.K. Sørensen, 1993, Seasonality in Macroeconomic Time Series, *Empirical Economics* **18**, 321–335.

Hylleberg, S., R.F. Engle, C.W.J. Granger and B. Yoo, 1990, Seasonal Integration and Cointegration, *Journal of Econometrics* **44**, 215–238.

Johansen, S., 1991, Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models, *Econometrica* **59**, 1551–1580.

Johansen, S., 1995, *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, Oxford University Press, Oxford.

Johansen, S. and K. Juselius, 1990, Maximum Likelihood Estimation and Inference on Cointegration - with Applications to the Demand for Money, *Oxford Bulletin of Economics and Statistics* **52**, 169–210.

Kaminsky, G., 1993, Is there a Peso Problem? Evidence from the Dollar/Pound Exchange Rate, *American Economic Review* **83**, 450–472.

Kim, C-J., 1994, Dynamic Linear Models with Markov-switching, *Journal of Econometrics* **60**, 1–22.

Kim, C-J. and J-S. Yoo, 1995, New Index of Coincident Indicators: A Multivariate Markov Switching Factor Model Approach, *Journal of Monetary Economics* **36**, 607–630.

King, R.G., C.I. Plosser, J.H. Stock and M.W. Watson, 1991, Stochastic Trends and Economic Fluctuations, *American Economic Review* **81**, 819–840.

Kleibergen, F., 1996a, Equality Restricted Random Variables: Densities and Sampling Algorithms, *Econometric Institute Report 9662/A*, Erasmus University Rotterdam.

Kleibergen, F., 1996b, Reduced Rank Regression using Generalized Method of Moments Estimators, *Center Discussion Papers 9620*, Tilburg University.

Kleibergen, F. and H. Hoek, 1995, Bayesian Analysis of ARMA models using Noninformative Priors, *Econometric Institute Report 9553/A*, Erasmus University Rotterdam.

Kleibergen, F. and H.K. van Dijk, 1994, On the Shape of the Likelihood/Posterior in Cointegration Models, *Econometric Theory* **10**, 514–551.

Kleibergen, F. and R. Paap, 1996, Priors, Posterior Odds and Lagrange Multiplier Statistics in Bayesian Analyses of Cointegration, *Econometric Institute Report 9668/A*, Erasmus University Rotterdam.

Kloek, T. and H.K. van Dijk, 1978, Bayesian Estimates of Equation System Parameters: An Application of Integration by Monte-Carlo, *Econometrica* **44**, 345–351.

Koop, G., 1991, Cointegration in Present Value Relationships, *Journal of Econometrics* **49**, 105–139.

Koop, G. and J-Y. Pitarakis, 1992, A Bayesian Analysis of Seasonality in Economic Time Series, unpublished working paper, University of Toronto.

Koop, G. and S.M. Potter, 1995, Bayes Factors and Nonlinearity: Evidence from Economic Time Series, unpublished working paper, University of Toronto.

Krolzig, H-M., 1996, Statistical Analysis of Cointegrated VAR Processes with Markovian Regime Shifts, unpublished working paper, Nuffield College, Oxford.

Lam, P-S., 1990, The Hamilton Model with a General Autoregressive Component, *Journal of Monetary Economics* **26**, 409–432.

Layton, A.P., 1994, Further on the Nature of the Australian Bysiness Cycle, *The Economic Record* **70**, 12–18.

Leamer, E.E., 1978, *Specification Searches: Ad Hoc Inference with Nonexperimental Data*, Wiley, New York.

Lee, H.S., 1992, Maximum Likelihood Inference on Cointegration and Seasonal Cointegration, *Journal of Econometrics* **54**, 351–365.

Lindgren, G., 1978, Markov Regime Models for Mixed Distributions and Switching Regressions, *Scandinavian Journal of Statistics* **5**, 81–91.

Litterman, R.B., 1986, Forecasting with Bayesian Vector Autoregressions: Five Years of Experience, *Journal of Business & Economic Statistics* **4**, 25–38.

Lucas, A., 1996, *Outlier Robust Unit Root Analysis*, PhD thesis, Tinbergen Institute, Thesis Publishers, Amsterdam.

Lucas, R.E., 1973, Some International Evidence on Output-Inflation Tradeoffs, *American Economic Review* **63**, 326–334.

Lütkepohl, H., 1993, *Introduction to Multiple Time Series Analysis*, 2nd edition, Springer-Verlag, Berlin.

Magnus, J.R. and H. Neudecker, 1988, *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Wiley & Sons, Chichester.

McCulloch, R.E. and R.S. Tsay, 1994a, Bayesian Inference of Trend- and Difference-Stationarity, *Econometric Theory* **10**, 596–608.

McCulloch, R.E. and R.S. Tsay, 1994b, Statistical Analysis of Economic Time Series via Markov Switching Models, *Journal of Time Series Analysis* **15**, 523–540.

Metropolis, N., A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller and E. Teller, 1953, Equations of State Calculations by Fast Computing Machines, *Journal of Chemical Physics* **21**, 1087–1092.

Miron, J.A., 1994, The Economics of Seasonal Cycles, in C.A. Sims (editor), *Advances in Econometrics, Sixth World Congress*, Vol. 1, Cambridge University Press, Cambridge, chapter 6, pp. 213–251.

Neftçi, S.N., 1984, Are Economic Time Series Asymmetric over the Business Cycle, *Journal of Political Economy* **92**, 307–328.

Nelson, C.R. and C.I. Plosser, 1982, Trends and Random Walks in Macroeconomic Time Series; Some Evidence and Implications, *Journal of Monetary Economics* **10**, 139–162.

Osborn, D.R., 1990, A Survey of Seasonality in UK Macroeconomic Variables, *International Journal of Forecasting* **6**, 327–336.

Osborn, D.R. and J.P. Smith, 1989, The Performance of Periodic Autoregressive Models in Forecasting Seasonal U.K. Consumption, *Journal of Business & Economic Statistics* **7**, 117–127.

Otto, G. and T. Wirjanto, 1990, Seasonal Unit Roots Tests on Canandian Macroeconomic Time Series, *Economics Letters* **34**, 117–120.

Paap, R., 1995, Industrial Production Forecasting Comparison using MSAR, STAR and AR Models, unpublished working paper, Erasmus University Rotterdam.

Paap, R., P.H. Franses and H. Hoek, 1997, Mean Shifts, Unit Roots and Forecasting Seasonal Time Series, *International Journal of Forecasting* **13**, 355–378.

Perron, P., 1989, The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis, *Econometrica* **57**, 1361–1401.

Phillips, K.L., 1991a, A Two-Country Model for Stochastic Output with Changes in Regime, *Journal of International Economics* **31**, 121–142.

Phillips, P.C.B., 1989, Partially Identified Econometric Models, *Econometric Theory* **5**, 181–240.

Phillips, P.C.B., 1991b, Optimal Inference in Cointegrated Systems, *Econometrica* **59**, 283–306.

Phillips, P.C.B. and W. Ploberger, 1994, Posterior Odds Testing for a Unit Root with Data-based Model Selection, *Econometric Theory* **10**, 774–808.

Phillips, P.C.B. and W. Ploberger, 1996, An Asymptotic Theory of Bayesian Inference for Time Series, *Econometrica* **64**, 381–412.

Poirier, D.J., 1995, *Intermediate Statistics and Econometrics: A Comparative Approach*, MIT Press, Cambridge.

Potter, S.M., 1995, A Nonlinear Approach to US GNP, *Journal of Applied Econometrics* **10**, 109–125.

Schotman, P.C., 1994, Priors for the AR(1) Model: Parameterization Issues and Time Series Considerations, *Econometric Theory* **10**, 579–595.

Schotman, P.C. and H.K. van Dijk, 1991a, A Bayesian Analysis of the Unit Root in Real Exchange Rates, *Journal of Econometrics* **49**, 195–238.

Schotman, P.C. and H.K. van Dijk, 1991b, On Bayesian Routes to Unit Roots, *Journal of Applied Econometrics* **6**, 387–401.

Sichel, D.E., 1994, Inventories and the Three Phases of the Business Cycle, *Journal of Business & Economic Statistics* **12**, 269–277.

Silverman, B.W., 1986, *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London.

Smith, A.F.M. and G.O. Roberts, 1993, Bayesian Computation via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods, *Journal of the Royal Statistical Society B* **55**, 3–23.

Stock, J.H. and M.W. Watson, 1988, Variable Trends in Economic Time Series, *Journal of Economic Perspectives* **2**, 147–174.

Tanner, M.A. and W.H. Wong, 1987, The Calculation of Posterior Distributions by Data Augmentation, *Journal of the American Statistical Association* **82**, 528–550.

Teräsvirta, T., 1995, Modelling Nonlinearity in U.S. Gross National Product 1889–1987, *Empirical Economics* **20**, 577–597.

Teräsvirta, T. and H.M. Anderson, 1992, Characterizing Nonlinearities in Business Cycles using Smooth Transition Autoregressive Models, *Journal of Applied Econometrics* **7**, S119–S136.

Tiao, G.C. and R.S. Tsay, 1994, Some Advances in Non-linear and Adaptive Modelling in Time-Series, *Journal of Forecasting* **13**, 109–131.

Tierney, L., 1994, Markov Chains for Exploring Posterior Distributions, *The Annals of Statistics* **22**, 1701–1762.

Tyssedal, J.S. and D. Tjøstheim, 1988, An Autoregressive Model with Suddenly Changing Parameters and an Application to Stock Market Prices, *Applied Statistics* **37**, 353–369.

Verdinelli, I. and L. Wasserman, 1995, Computing Bayes Factors Using a Generalization of the Savage-Dickey Density Ratio, *Journal of the American Statistical Association* **90**, 614–618.

Warne, A., 1996, Autocovariance Functions and Maximum Likelihood in a VAR Model under Markov Switching, unpublished working paper, Stockholm University.

Zellner, A., 1971, *An Introduction to Bayesian Inference in Econometrics*, Wiley, New York.

Zellner, A. and A. Siow, 1980, Posterior Odds for Selected Regression Hypotheses, in J. Bernardo, M.H. DeGroot, D.V. Lindley and A.F.M. Smith (editors), *Bayesian Statistics*, University Press, Valencia, pp. 585–603.

Zivot, E. and D.W.K. Andrews, 1992, Further Evidence on the Great Crash, the Oil-Price Shock, and the Unit-Root Hypothesis, *Journal of Business & Economic Statistics* **10**, 251–270.

The Tinbergen Institute is the Netherlands Research Institute and Graduate School for Economics and Business, which was founded in 1987 by the Faculties of Economics and Econometrics of the Erasmus University in Rotterdam, the University of Amsterdam and the Free University in Amsterdam. The Institute is named after the late Professor Jan Tinbergen, Dutch Nobel Prize laureate in economics in 1969. The Tinbergen Institute is located in Amsterdam and Rotterdam. If available trade editions of the books which are published in the Tinbergen Institute Research Series can be ordered through Thesis Publishers, P.O. Box 14791, 1001 LG Amsterdam, The Netherlands, phone: +3120 6255429; fax: +3120 6203395. The following books recently appeared in this series:

120. T.R.P.J. KROES, *Financial intermediation and monetary transmission.*
121. T.H.F. CHEUK, *Exotic options.*
122. K. TEPLA, *Accounting for the quality of work-life-putting price tags on sources of job (dis)satisfaction.*
123. R.H.M. EMMERINK, *Information and pricing in road transport: theoretical and applied models.*
124. T.J.S. OFFERMAN, *Beliefs and decision rules in public good games - theory and experiments.*
125. F. MERCURIO, *Claim pricing and hedging under market imperfections.*
126. A.W.M. ODÉ, *Migrant workers in the Dutch labour market today.*
127. R.E. WILDEMAN, *The art of grouping maintenance.*
128. M.T. ROCHA DE MAGALHAES MELO, *Stochastic lot-sizing in production planning - strategies for make-to-order and make-to-stock.*
129. G. RUSSO, *Firms' recruitment behaviour: an empirical investigation of the use of search channels, the rate of arrival of applicants, and the spatial radius of search.*
130. J.N. VAN OMMEREN, *Commuting and relocation of jobs and residences; a search perspective.*
131. F.N. GOUWELEEUW, *A general approach for computing loss probabilities in finite-buffer queues.*
132. M.R.H. MANDJES, *Rare event analysis of communication networks.*
133. E.J. VISSER, *Local sources of competitiveness: Spatial clustering and organisation dynamics in small-scale clothing in Lima, Peru.*
134. M.H.H. HUYSMAN, *Dynamics of organizational learning.*
135. J.K. GOEREE, *Applications of discrete choice models.*
136. N.J. VAN DER LIJN, *Consumer markets and welfare under central planning.*
137. E.A.A. DE LAAT, *Essays on patent policy: the multi-dimensionality of patents and asymmetric information*
138. C.M. VAN DER TAK, *International commodity-related environmental agreements: an analysis of international environmental agreements for domestic environmental externalities.*
139. M.P.E. MARTENS, *Interaction between financial markets.*
140. L.W.N. VAN DER TORRE, *Reasoning about obligations. Defeasibility in preference-based deontic logic.*
141. W. BOLT, *Theory of negotiation: a game-theoretic approach.*

142. M.A.J. SMITH, *On the availability of failure prone systems.*
143. R. SLOOF, *Game-theoretic models of the political influence of interest groups.*
144. J.N.M. WIT, *Dynamics and information aggregation in elections and markets.*
145. N.T. BUDINA, *Essays on consistency of fiscal and monetary policy in Eastern Europe.*
146. F. LIPPI, *Central Bank independence and credibility: essays on the delegation arrangements for monetary policy.*
147. H. HOEK, *Variable trends: a Bayesian perspective.*
148. E.J.S. PLUG, *Leyden welfare and beyond.*
149. C.S.J. JIE-A-JOEN, *Strategic trade policy, multinational firms, and vertically related industries.*
150. R.G.J. DEN HERTOG, *Empirical essays on growth, uncertainty and prediction.*
151. M. GELDERMAN, *Success of management support systems: a review of the literature and an empirical investigation.*
152. M.A. VAN DER ENDE, *On the values of jobs and specific training.*
153. M. KLAPWIJK, *Rural industry clusters in central Java, Indonesia; an empirical assessment of their role in rural industralization.*
154. R.J.G. VENNIKER, *Transition processes from constrained to Walrasian equilibrium.*
155. J.M. MORALEDA, *On the pricing of interest-rate options.*
156. J.F. STURM, *Primal-dual interior point approach to semidefinite programming.*
157. R. FRELING, *Models and techniques for integrating vehicle and crew scheduling.*
158. S. BORGHGRAEF, *Is more control better? The effects of organizational control on salespersons' job performance, job attitudes and sales organization effectiveness.*
159. C. VINKENBURG, *Managerial behavior and effectiveness. Determinants, measurement issues and gender differences.*
160. P.A. GAUTIER, *The flow approach to the labor market.*
161. A.A.R.J.M. VAN RIXTEL, *Informality, bureaucratic control and monetary policy: The case of Japan.*
162. Y. GAO, *Management support of information system project planning.*
163. J.G. PIEK, *Development and acquisition of technology in rural collective industries. The case of China's hinterland.*
164. F. LANGERAK, *The effect of market orientation on the performance of industrial firms: A study in the Dutch Metalelektro.*
165. A.K. SINHA, *Estimating failure probability when failure is rare: multi-dimensional case.*
166. J.P.M. SIJM, *Food security and policy interventions in sub-Saharan Africa. Lessons from the past two decades.*
167. W.A. LETTERIE, *Partisan politics, persuasion and information.*
168. R. VAN DIJK, *Corporate finance policy and equity investment: panel data analyses.*
169. M.A. CARREE, *Market dynamics, evolution and smallness.*
170. J.E. LIGTHART, *Environment, imperfect markets at public finance.*