BAYESIAN ANALYSIS OF STOCHASTIC TRENDS AND SEASONALITY

Freie wissenschaftliche Arbeit
zur Erlangung des akademischen Grades
"Dr. rer. pol."

vorgelegt bei

Professor Dr. Susanne Rässler
Lehrstuhl für Statistik und Ökonometrie
Wirtschafts- und Sozialwissenschaftliche Fakultät
Otto-Friedrich-Universität Bamberg

von

Diplom Sozialwirt Alexander Vosseler

12. September 2013

Promotionsvermerk:

Erstgutachterin: *Prof. Dr. Susanne Rässler*

Zweitgutachter: *Prof. Dr. Ingo Klein*

Datum der Disputation: *18.02.2014*

Contents

Part B: Analysis of seasonal time series

List of Tables

List of Figures

List of notations

I. General symbols

$=$	equals
$\propto$	proportional to
$\equiv$	equals by definition
$\Rightarrow$	implies
$\Leftrightarrow$	is equivalent to
$\approx$	approximately equal to
$\sim$	is distributed as
$\in$	element of
$\forall$	for all
$\exists$	exists
$\subset$	subset of
$\bigcup$	union
$\bigcap$	intersection
$\wedge$	conjunction ('and')
$\vee$	disjunction ('or')
mod	modulo operator
$\lfloor x \rfloor$	floor(x) - function
$\sum$	summation sign
$\prod$	product sign
$\mapsto$	maps to
$\rightarrow$	converges to, approaches
$\xrightarrow{d}$	weakly converges, converges in distribution
Δ_s	s-th differencing operator
lim	limes
min	minimum
max	maximum
sin	sine function
cos	cosine function
exp	exponential function
ln	natural logarithm
log	logarithm to base 10

sign	sign function
Γ	gamma function
det	determinant
$\mathbf{X}^{-1}$	inverse matrix
$\mathbf{X}'$	transposed matrix
ι	unit vector
$\mathbf{I}_d$	identity matrix of dimension d
$dim(\mathbf{X})$	dimension of matrix $\mathbf{X}$
$\overline{x}$	arithmetic mean of x
$Med(x)$	median of x
$\widehat{\theta}$	estimator of θ
$\mathbf{1}_{(.)}$	indicator function
$D_{s,t}$	seasonal dummy variable
L	lag operator
E	expectation
Var	variance
Cov	covariance
se	standard error
$f(X)$	marginal probability density or mass function of X
$f(X\|Y)$	conditional probability density or mass function of X given $Y = y$
$f(X,Y)$	joint probability density or mass function of X and Y
$F(X)$	cumulative probability density or mass function of X
$\|J\|$	Jacobian determinant
$I(\theta)$	Fisher information with respect to θ
C	normalizing constant
$\|z\|$	absolute value or modulus of a complex number
$\mathbb{R}$	real numbers
$\mathbb{R}^m$	m-dimensional Euclidean space
$\mathbb{C}$	complex numbers
$\mathbb{N}$	positive integers
$\mathbb{Z}$	integers
$\mathscr{D}$	decision (or action) space
$\mathscr{M}$	model space

II. Probability distributions and stochastic processes

$Be(\alpha,\beta)$	Beta distribution
$B(n,p)$	Binomial distribution
$\chi^2_{(k)}$	Chi-squared distribution
$G(\alpha,\beta)$	Gamma distribution
$G_2(\nu,s)$	Gamma-2 distribution
$IG_2(\nu,s)$	Inverse Gamma-2 distribution
$L(\mu,b)$	Laplace (or 'Double exponential') distribution
$M_k(n;p_1,...,p_k)$	Multinomial distribution
$F_{(\nu_1,\nu_2)}$	Fisher's F-distribution
$N(\mu,\sigma^2)$	Univariate Normal distribution
$N_d(\mu,\sigma^2)$	Multivariate Normal distribution
$\mathscr{T}_d(\theta,\ \Sigma,\nu)$	Multivariate Student-t distribution
$U(a,b)$	Continuous uniform distribution
$\{y_t\}_{t=-\infty}^{\infty}$	Stochastic process
AR(p)	Autoregressive process of order p
MA(q)	Moving average process of order q
ARMA(p,q)	Autoregressive moving average process of order (p,q)
SARMA$(P,Q)\times(p,q)$	Seasonal ARMA process of order (P,Q) and (p,q)
SETAR(p)	Self-exiting threshold autoregressive model of order p
PAR(p)	Periodic autoregressive process of order p
PMA(q)	Periodic moving average process of order q
PARMA(p,q)	Periodic autoregressive moving average process of order (p,q)
$I(d)$	Integrated of order d (process)
SI(1)	Seasonally integrated of order one (process)
PI(1)	Periodically integrated of order one (process)
ARI(1)	Integrated AR(1) process
PARI	PAR process for an integrated series
PIAR	Periodically integrated AR process
VAR(P)	Vector autoregressive process of order P
W(r)	(Standard) Wiener process
$WN(0,\sigma^2)$	White noise process

III. Abbreviations

ADF	Augmented Dickey-Fuller
AIC	Akaike's information criterion
BF	Bayes factor or Boswijk-Franses
BIC	Bayesian information criterion
BL	Broemeling-Land
BMA	Bayesian model averaging
BPAR	Bayesian periodic autoregressive
DGP	Data generating process
HEGY	Hylleberg-Engle-Granger-Yoo
HPD	Highest posterior density
IAB	Institute for Employment Research
iid	Independently identically distributed
IN	Independently normally distributed
LR	Likelihood ratio
MAP	Maximum a posteriori
MAPE	Mean absolute percentage error
MAD	Mean absolute deviation
MC	Monte Carlo
MCMC	Markov chain Monte Carlo
MH	Metropolis-Hastings
ML	Maximum Likelihood
NAIRU	Non-accelerating inflation rate of unemployment
(N)SC	(No) structural change
OECD	Organisation for economic cooperation and development
OLS	Ordinary least squares
pdf	Probability density function
PEL	Posterior expected loss
PMEANS	Periodic means
(P)MSE	(Predictive) mean squared error
pmf	Probability mass function
RSS	Residual sum of squares
S(P)ACF	Sample (partial) autocorrelation function
WSR	Wilcoxon Signed Rank

Introduction

1.1. Summary and structure of the thesis

This doctoral thesis consists of two main parts, which are devoted to the analysis of non-seasonal and seasonal time series data, respectively. These two parts are then further subdivided according to three research articles, where in the first of these articles, presented in chapter 2, a fully Bayesian approach to model selection in testing regressions for a non-seasonal unit root with multiple structural breaks is proposed. In the second part of the thesis the focus is on testing and forecasting of seasonal time series data. Since the reader might not be so familiar with some of the unit root concepts used in the second research article presented in chapter 4, the most important concepts related to nonstationarity in seasonal time series models are introduced as a preliminary in chapter 3. In the second paper, Bayesian testing approaches for different kinds of unit roots within the class of periodic autoregressive (PAR) models with a possible mean break are proposed. Further, since all these approaches assume seasonality of quite general form also two Bayesian pretests for periodic variation in the mean of a process are considered, respectively. The third article, presented in chapter 5, is devoted to the prediction of quarterly and monthly time series. Here a model averaging approach for Bayesian PAR models of unknown order, number of breaks and break dates is proposed in order to improve forecasting accuracy. Moreover the joint posterior predictive distribution for multistep ahead forecasts is analyzed and a sampling approach to obtain the marginal predictive distributions is presented. In order to compare the predictive ability of the presented forecasting model with those of other models, a Bayesian sign test is introduced.

In each of the three presented articles an extensive Monte Carlo study is conducted to analyze the statistical methods for different data generating processes. Further in the empirical sections of the first two papers (see chapters 2 and 4) the proposed unit root testing

procedures are used to answer the question if there is empirical evidence for unemployment persistence or hysteresis in 17 OECD countries after a labor market shock. In the empirical application of chapter 5 the suggested prediction approach is applied to forecast the unadjusted monthly unemployment rates of the 16 German federal states and of East- and West-Germany. Finally, in chapter 6 the major contributions and results of this thesis are summarized and a short discussion with potential future research is given.

Before proceeding a brief motivation with regard to the focus of this thesis, namely testing the unit root hypothesis and in particular the chosen Bayesian frame of reference, will be given. The aim here is not to provide a general discussion of Bayesian and classical (or frequentist) statistics per se, but to motivate the use of Bayesian methods for the analysis and prediction of time series data. A review of the various arguments for and against the use of the Bayesian paradigm in statistical inference can be found in many excellent textbooks as for example Berger (1980) or Robert (2007), and the references therein.

1.2. Motivation

In the economic and econometric literature the unit root hypothesis has gained much interest since the seminal paper of Nelson and Plosser (1982). Among the economic theories for which (non)stationarity of the considered dynamic system has important implications are the permanent income theory (cf. Hall (1978)), the business cycle theory (cf. King et al. (1988)) or the insider-outsider theory and the theory of unemployment hysteresis in labor market research (cf. Blanchard and Summers (1986), Blanchard and Summers (1987)). The latter provides the theoretical background for the empirical analyses presented in chapters 2 and 4.

In the nineties there was a heated controversy in the literature on classical and Bayesian unit root testing and in particular on the appropriate prior distribution, starting with Sims (1988). The main contributions to this debate are summarized in a special issue of the *Journal of Applied Econometrics* (1991, volume 6, number 4). A summary of the many arguments put forward by the involved authors can be found in Maddala and Kim (1998), chapter 8, Bauwens et al. (1999), chapter 6, also Uhlig (1994). One of the most striking points of the Bayesian advocates is, that, in contrast to classical theory, Bayesian inference

in dynamic models is largely unaffected by the presence of a unit root, see Sims (1988), Sims and Uhlig (1991). As a simple comparison of classical and Bayesian unit root inference consider the following AR(1) model:[1]

$$y_t = \phi y_{t-1} + \varepsilon_t \ , \ \ y_0 = 0, \ \varepsilon_t \overset{i.i.d.}{\sim} N(0, \sigma^2) \tag{1.1}$$

with ordinary least squares estimator $\widehat{\phi} = \sum_{t=1}^{T} y_t y_{t-1} / \sum_{t=1}^{T} y_{t-1}^2$.

It can be shown that the asymptotic distribution of $\widehat{\phi}$ has a discontinuity at $\phi = 1$ and thus is not the usual Gaussian distribution as in the stationary case (see Hamilton (1994), p.475 ff., for details). Kadane et al. (1996) summarize the asymptotic behavior of $\widehat{\phi}$, conditional on the considered subset of the parameter space of ϕ, as

$$(\widehat{\phi} - \phi) / \sqrt{\sum_{t=1}^{T} y_{t-1}^2} \xrightarrow{d} \begin{cases} N(0,1) & \text{, for } \phi < 1 \\ \frac{0.5W(1)^2 - 1}{(\int_0^1 W(r)^2 dr)^{1/2}} & \text{, for } \phi = 1 \\ \mathscr{T}_1(0,1,\ \nu = 1) & \text{, for } \phi > 1 \end{cases} \tag{1.2}$$

where $\xrightarrow{d}$ denotes convergence in distribution as T tends to infinity, $\mathscr{T}_1(.)$ is the univariate standard Student-t density with $\nu = 1$ degree of freedom, i.e. a Cauchy distribution, and $W(r)$ denotes a standard Wiener process with $r \in [0,1]$, cf. Banerjee et al. (1993).

From (1.2) it can be seen that the asymptotic sampling distribution $f(\widehat{\phi}|\phi)$ is symmetric around zero in the stationary and the explosive case, but is asymmetric, more precisely skewed to the left, in the presence of a unit root. Moreover, adding a constant or a trend to the model complicates the resulting asymptotic distribution in a non-trivial way, see Hamilton (1994) for details. Since conditional on initial values an AR model can be treated as an ordinary linear regression model all the analytical results of the conjugate normal regression case can be utilized in a Bayesian framework. For example, assuming a Normal-Inverse-Gamma-2 prior on the unknown quantities ϕ and σ^2 in (1.1) the marginal posterior distribution $f(\phi|y_T, ..., y_1, y_0)$ equals a univariate Student-t density or a univariate normal density when $\sigma^2 = 1$, i.e. known, see Raiffa and Schlaifer (2000), chapter 13, also Zellner (1971), chapter 7.[2] It is important to note that the latter is true irrespective

[1] In order to keep the notation simple I will not discriminate between a random variable and its realization in the following.

[2] Another useful analytical result can be utilized by assuming prior independence of both parameters and

of the sample size T, nonstationarity of the data generating process or the inclusion of deterministic components. Another appealing feature of choosing a Bayesian framework is that one does not only get a point or an interval estimate of the quantity of interest, e.g. a break date or the number of breaks, but a whole distribution, viz. the posterior distribution which, for a given data set, is a sound way to capture the uncertainty with regard to any further inference.

Sims and Uhlig (1991) compute the joint posterior distribution of ϕ and $\widehat{\phi}$ under a flat prior using Monte Carlo methods and examine the behavior of the conditional densities $f(\widehat{\phi}|\phi = 1)$ and $f(\phi|\widehat{\phi} = 1)$ to compare the Bayesian and the classical approach to unit root testing. They conclude that classical methods based on the resulting asymmetric small sample distribution $f(\widehat{\phi}|\phi = 1)$ of the OLS estimator can be misleading by assigning too much density to large ϕ values, see ibid. for details. Although hypothesis testing can also be considered as a point estimation problem (cf. Robert (2007), chapter 5), unit root testing is a striking example where it is not possible to recover classical results by using a flat prior, see Bauwens et al. (1999). Moreover the results of classical and Bayesian unit root tests can differ considerably. Since the unit root hypothesis is a point hypothesis it is highly controversial to test for this in a Bayesian framework. In the continuous case, this implies comparing a parameter interval receiving positive probability mass under the alternative $H_1 : \phi < 1$ with a singleton of zero probability mass under the null hypothesis $H_0 : \phi = 1$, see equation (1.1). An extreme illustration of the conflicting outcomes of frequentist and Bayesian testing of a point null, namely rejecting the null almost surely with an arbitrarily low p-value, while on the other hand obtaining a posterior probability of H_0 close to one when the sample size increases, is called the 'Jeffreys-Lindley's paradox', see Berger (1980), p.156, also Shafer (1982) for details. Besides the appropriate choice of the prior, which becomes less influential when more information is available, the chosen parametrization of the testing regression can play a pivotal role for the performance of the test. In frequentist unit root testing the distribution theory becomes much more involved when structural breaks in the model parameters are allowed, cf. Maddala and Kim (1998) for an overview. Then for testing purposes two additional kinds of unknown entities have to be selected, viz. the number of breaks and the corresponding break dates. In a Bayesian context estimation of both entities can be accomplished simultaneously by

then using a Student-t prior on ϕ and an Inverse-Gamma-2 prior on the variance of the error term, see Dreze (1977), Richard and Tompa (1980).

using Monte Carlo techniques, as will be shown in chapter 2, where this task becomes much more challenging when using classical methods, cf. Bai and Perron (1998), Bai and Perron (2003). Another useful approach presented in chapter 4, which avoids the selection of a single testing regression is to use model averaging techniques. It is exactly for these reasons why a comparison of classical and Bayesian unit root testing procedures can bring new insights into the analysis of nonstationarity when different Bayesian approaches to this model uncertainty problem are considered.

Finally, also with regard to the prediction of future values, y_{T+k}, $k \geq 1$, Bayesian methods can provide an interesting alternative to existing frequentist time series methods. In chapter 5 a model averaging prediction approach for seasonal time series models with possible breaks is presented. This requires the computation of a mixture of density functions, where in a Bayesian framework the mixture weights are the posterior probabilities of the different candidate models M_i. In general, model averaging demands the predefinition of a set of candidate models, $\mathscr{M} = \{M_1, ..., M_I\}$, where each element M_i represents a certain (non)nested model specification, which can be indicated by the introduction of a model index $\gamma \in \{1, ..., I\}$, see Raftery et al. (1997).[3] Although there also exist frequentist model averaging approaches that obtain the required model weights as the solutions of an optimization problem (cf. Hansen (2007)), the majority of the existing literature assumes a Bayesian frame of reference. One reason for this is that here it is natural to consider the model indicator γ as an additional random parameter with assigned prior probability in order to express model uncertainty. Then model averaging implies 'integrating out' this nuisance parameter by averaging over the support of γ, using the posterior model probabilities as weights. It is interesting to note that the frequentist estimators of the model weights can also be obtained by applying a Laplace approximation (see Tierney and Kadane (1986), Tierney et al. (1989)) to the joint posterior density as in Raftery (1995), p.130 ff. In this case the posterior probability mass function of model M_i can be expressed as a function of the Bayesian information criterion (BIC) (see Schwarz (1978)):

$$f(M_i|\,\mathbf{y}) \approx \exp(-1/2 \cdot BIC_i) / \sum_{j=1}^{I} \exp(-1/2 \cdot BIC_j)$$

which is also one of the sampling estimates for the weights used by Hansen (2007) within

[3] In this respect a model selection problem can also be perceived as a point estimation problem, namely that of the model indicator γ, see Robert (2007), p.342.

a classical context.

After this motivation on the topic and the Bayesian framework has been given, next an approach to test for a nonseasonal unit root in the case of multiple structural breaks is presented.

Analysis of nonseasonal time series

Bayesian model selection for unit root testing with multiple breaks

2.1. Introduction

There has been a growing literature to unit root testing in economic time series over the last three decades starting with the seminal papers of Dickey and Fuller (1979) and Nelson and Plosser (1982). As stressed by many authors the misspecification of the considered test regressions can lead to substantially biased inferences in the class of autoregressive integrated moving average (ARIMA) models, see Banerjee et al. (1993), Stock (1994), Maddala and Kim (1998) for discussions. The specification regards on the one hand the structure of the stochastic component, i.e. the autoregressive and/or moving average lag orders (see Hall (1994), Ng and Perron (2001)) and on the other hand the specification of the deterministic components like the inclusion of time trends, the number of possible structural breaks and also the timing of these breaks, see Perron (1989), Christiano (1992), Vogelsang and Perron (1992), inter alia. In the Bayesian unit root literature a heated controversy was devoted to the adequate prior use, model specification issues and the proper modeling of initial conditions, see e.g. Sims (1988), Phillips (1991*b*), Uhlig (1994), inter alia, and Bauwens et al. (1999) for an overview. In contrast to the classical literature there are only a few approaches to account for structural breaks when testing for stochastic trends, see e.g. Zivot and Phillips (1994), Koop and Steel (1994), DeJong (1996), Marriott and Newbold (2000). Most of these works treat the model order, in particular the lag order and/or the number of breaks as fixed quantities. In general, the process of model selection induces uncertainty with respect to any subsequent analysis and thus should be captured in order to improve statistical inference. Although there exist many classical approaches to model selection in dynamic models with structural breaks,

which are mainly based on information criteria, a different approach is chosen here using a Bayesian framework.

In the following a stochastic model selection approach is presented, which can be used to determine the optimal specification for unit root testing in the case of multiple structural breaks. In a nutshell, the proposed sampling scheme can be regarded as an extension of the approach presented in Wang and Zivot (2000) by estimating the number of structural breaks, the associated break dates as well as the number of autoregressive lags simultaneously with all other unknown model parameters. This is accomplished by the introduction of two discrete valued state variables that indicate certain model combinations in the space of candidate models. Since the joint distribution of all unknown parameters is of varying dimension, i.e. depends on the specific model complexity, usual sampling techniques to generate random draws from this distribution, like Gibbs sampling, can not be applied without further modifications. For this purpose a flexible Markov chain Monte Carlo (MCMC) approach is introduced, which enables to jump between parameter spaces of differing dimensionality. The performance of this method is demonstrated on the basis of several Monte Carlo (MC) experiments which indicate great reliability in finding the true values of the data generating process (DGP). Using Bayesian methods for model selection in structural break models has the advantage that most of these methods are technically simpler than their classical counterparts, allow for finite-sample inferences that are optimal given the framework, and also allow for nonnested model comparisons (see Wang and Zivot (2000)). Furthermore with regard to unit root testing a Bayesian model framework is appealing, because unlike in classical approaches inference stays the same here for trending and nontrending data (see Sims and Uhlig (1991)). So far many approaches to model selection in time series models have been proposed in the Bayesian literature. Among the works that mainly focus on lag order determination are Huerta and West (1999), Vermaak et al. (2004), Ehlers and Brooks (2004), Philippe (2006), inter alia. Many of the existing works that deal with the detection of change points treat the selection of the number of breaks as a successive problem, which is solved by using information criteria or Bayes factors, but do not treat the number of change points together with the number of lags as additional model parameters explicitly in their sampling schemes, see for example Chib (1998), Wang and Zivot (2000), Koop and Potter (2004).

Therefore the present work aims to provide contributions in the following directions: a

stochastic model selection approach for multiple structural breaks models is proposed, where the autoregressive lag order, the break dates and also the number of breaks can be estimated simultaneously with all other model parameters. As a result the joint posterior distribution of these model indicators is obtained, which can be used for further inference. Moreover the model selection approach presented below focuses on the application in unit root testing problems and the use of Augmented-Dickey-Fuller-type regressions (see Said and Dickey (1984)) and thus provides an alternative to classical model selection strategies used in this context. Unlike in standard autoregressive models alternative Bayesian model selection approaches to lag order determination, as for example the stochastic search variable selection method of George et al. (1993) can be cumbersome to apply here, because of the special structure of such test regressions. Besides model determination a second focus lies on testing for a (zero frequency) unit root when there are multiple structural breaks in the deterministic components of the process. This is done by computing the posterior probability of a unit root under several prior distributions, which is then used to construct a Bayesian test. The proposed Bayesian unit root test is then compared with several classical unit root tests by means of simulated power functions. Monte Carlo experiments indicate a clear superiority of the Bayes test in terms of power especially in moderate and small samples, i.e. when the asymptotic distribution theory underlying the classical tests is not valid anymore.

In an empirical application, the unemployment rates of 17 OECD countries for the years 1960 to 2010 are analyzed to answer the question if there is persistent behavior after a labor market shock. The majority of empirical works to test for persistence effects in European unemployment rates has been done by using classical methods. Most of these works apply univariate tests without structural breaks and can not reject the unit root null hypothesis, see Mitchell (1993), Roed (1996) also Hassler and Wolters (2009). For the US the results are mostly reverse and therefore no high degree of persistence has been found in the unemployment rates, see Nelson and Plosser (1982), Blanchard et al. (1992) also Roed (1996). It is by now well recognized that not allowing for structural breaks in the test regression can bias the results toward a unit root. Therefore a second group of studies uses methods that allow for (multiple) structural breaks. Not surprisingly the results of these studies show a clearer tendency against a unit root, see Arestis and Biefang-Frissancho Mariscal (1999), Papell et al. (2000), Papell and Prodan (2004) and also Pascalau (2007). So far only few authors used Bayesian methods to analyze the

trend characteristics of unemployment rates, exceptions are Summers (2004), Mikhail et al. (2006) and Berger and Everaert (2008). Since the time span of the data covers the first financial crisis of the year 2008 it is also possible to capture the impact of this event. Thereby, besides the main statistical focus, the present work also aims to provide some new empirical evidence to the question if certain OECD countries are more likely then others to recover to their natural rate of unemployment after an exogenous shock or if such an event has a permanent impact on a country's long run unemployment rate. As a benchmark the Bayesian estimates of the break numbers and the break points are compared with the estimates from an application of the classical methods of Bai and Perron (2003). To gain further insights into the countries' convergence properties and also in order to control for uncertainty induced through the model selection step the estimated model posterior distribution is used to compute the model averaged half life of a shock for each OECD country.

This chapter is structured as follows: in section 2.2 the statistical model is presented and in section 2.3 the MCMC sampling algorithm is introduced and its performance is analyzed in section 2.4. Then in section 2.5 the Bayesian testing approach for a unit root with multiple structural breaks is presented. Further, in section 2.6, the sensitivity of the test results with regard to the assumed prior distributions is analyzed and the Bayesian unit root test is compared with some classical unit root tests. In section 2.7 the presented methods are applied to annual OECD unemployment rates and section 2.8 concludes.

2.2. Bayesian unit root testing with multiple breaks

2.2.1. Model and definitions

In the following it is assumed that the series of interest y_t can be described by an autoregressive process of order p together with some deterministic components, i.e.

$$A_p(L) \cdot y_t = \mu + \beta \cdot t + u_t \text{ with } u_t \overset{i.i.d.}{\sim} N(0, \sigma^2) \text{ , } t = 1, ..., T \tag{2.1}$$

where the intercept μ and slope β are the coefficients of the deterministic trend function and $A_p(L)$ is defined in terms of the lag operator $x_{t-j} \equiv L^j \cdot x_t$, $j \in \mathbb{Z}$, i.e.

$$A_p(L) \equiv 1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p \tag{2.2}$$

The main focus here is to investigate if the autoregressive lag polynomial $A_p(L)$ can be factorized according to $A_{p-1}(L) \cdot (1-L)$, where in this case y_t is said to exhibit a (nonseasonal) unit root, i.e. $A_p(1) = 0$. To test for a root at the zero spectral frequency it is convenient to rewrite the above polynomial as

$$A_p(L) = (1 - \theta L) - A^{\star}_{p-1}(L) \cdot (1-L) \tag{2.3}$$

with $\theta = \sum_{s=1}^{p} \phi_s$ the long-run impact coefficient. The polynomial $A^{\star}_{p-1}(L) = \psi_1 \cdot L + ... + \psi_{p-1} \cdot L^{p-1}$ is assumed to have all roots outside the complex unit circle with the coefficients $\psi_j = -\sum_{s=j+1}^{p} \phi_s$, $j = 1...p-1$, measuring transient dynamics (see Hamilton (1994), p.517). Substituting (2.3) into (2.1) then leads to

$$y_t = \mu + \beta \cdot t + \theta \cdot y_{t-1} + \sum_{j=1}^{p-1} \psi_j \cdot \Delta y_{t-j} + u_t \;\; \text{with} \;\; u_t \overset{i.i.d.}{\sim} N(0, \sigma^2) \tag{2.4}$$

This is an Augmented Dickey-Fuller (ADF) regression for testing the unit root hypothesis $H_0 : \theta = 1$, which implies that (at least) one of the p characteristic roots $z_j \in \mathbb{C}$ of $A_p(z)$ has modulus equal to unity (see Said and Dickey (1984), also Nelson and Plosser (1982)). On the other hand the (trend)stationary alternative $H_1 : |\theta| < 1$ implies that all roots are strictly outside the unit circle. Hence the relevant parameter region for this testing problem is $\theta \in [0;1]$. In order to allow for multiple structural breaks in the DGP the above regression is now extended to (cf. Perron and Vogelsang (1992)):

$$y_t = \sum_{i=1}^{m+1} \mathbf{1}_{\{k_{i-1} \leq t < k_i\}} (\alpha_i + \beta_i \cdot t) + \theta \cdot y_{t-1} + \sum_{j=1}^{p-1} \psi_j \cdot \Delta y_{t-j} + u_t \;\; , \;\; u_t \overset{i.i.d.}{\sim} N(0, \sigma^2) \tag{2.5}$$

where k_i denotes the i-th break date with $1 < k_1 < ... < k_m \leq T$ and $\mathbf{1}_{\{A\}}$ denotes an indicator variable that equals 1 if the statement A is true and 0 otherwise. Note that the effect of an intervention is modeled here as a step function at date $t = k_i$, which is typically used to represent an instantaneous impact on the level of a series y_t. Setting $k_0 = 1$ and

$k_{m+1} = T+1$ for the lower and upper margins, respectively, the T observations can be separated into $m+1$ regimes, see **?**. In contrast to the latter authors I treat the number of structural breaks $m = 0,1,...,m_{max}$ and also the autoregressive lag order $p = 1,...,p_{max}$ as unknown model indicators stacked together in a vector $\gamma \equiv (p,m)'$ which has to be estimated.

The above ADF-type multiple structural breaks model can be written more compactly in matrix form, where it is convenient to separate the vector of model parameters $\lambda_\gamma \equiv (\mathbf{B}'_\gamma, \sigma^2, \mathbf{k}'_\gamma)'$ of dimension $h(\gamma) = 3 \cdot (m+1) + p$ in the 'drift-and-trend' case and of $h(\gamma) = 2 \cdot (m+1) + p$ in the 'drift-only' case from the vector of model indicators denoted by γ. The model in (2.5) can then be expressed in the usual linear regression form[1]

$$\mathbf{y} = \mathbf{X}_\gamma \mathbf{B}_\gamma + \mathbf{u} \tag{2.6}$$

with $\mathbf{y} = (y_{p+1},...,y_T)'$, the first p observations $\mathbf{y}_0 = (y_1,...,y_p)'$ used as initial values, $\mathbf{u} = (u_{p+1},...,u_T)'$ the vector of innovations and the matrix $\mathbf{X}_\gamma = [\mathbf{x}_{p+1},...,\mathbf{x}_T]'$ with row vectors

$$\mathbf{x}_t \equiv [\mathbf{1}_{\{k_0 \le t < k_1\}}, \ldots, \mathbf{1}_{\{k_m \le t < k_{m+1}\}}, \mathbf{1}_{\{k_0 \le t < k_1\}} \cdot t, \ldots, \mathbf{1}_{\{k_m \le t < k_{m+1}\}} \cdot t, \\ y_{t-1}, \Delta y_{t-1}, \ldots, \Delta y_{t-p+1}]$$

of dimension $d(\gamma) = m+1+p$ in the 'drift-only' case or dimension $2 \cdot (m+1) + p$ in the 'drift-and-trend' case.[2] The vector of all unknown quantities is thus $(\mathbf{B}'_\gamma, \sigma^2, \mathbf{k}'_\gamma, \gamma')'$ with $\mathbf{B}_\gamma = (\alpha_1, \ldots, \alpha_{m+1}, \beta_1, \ldots, \beta_{m+1}, \theta, \psi_1, \ldots, \psi_{p-1})'$ the vector of regimewise regression coefficients and $\mathbf{k}_\gamma = (k_1, \ldots, k_m)'$ the vector of break dates.[3]

[1] Technically speaking the $\mathbf{X}_\gamma$ matrix is of course also dependent on the realizations in $\mathbf{k}_\gamma$. In this sense the parameters in $\mathbf{k}_\gamma$, i.e. the break dates, could also be considered as further model indicators.

[2] For notational brevity I will write $d = d(p,m)$ and $h = h(p,m)$ and also for example $d^\star = d(p^\star, m)$ in the sequel.

[3] Since this is a conditional likelihood approach, rather than an exact (cf. Bauwens et al. (1999), p.135), the first p observations of the series get lost due to lagging and so the first break can not be detected until $t = p+2$. This is the reason why $\mathbf{k}_\gamma$ is also dependent on the lag order.

2.2.2. Likelihood and prior specifications

Utilizing the first p observations as initial values $\mathbf{y_0}$ yields the conditional data density, which when viewed as a function of the parameters is the approximate likelihood function

$$f(\mathbf{y}|\lambda_\gamma,\gamma;\mathbf{y_0}) \propto \sigma^{-(T-p)} \cdot \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{X}_\gamma\mathbf{B}_\gamma)' \cdot (\mathbf{y}-\mathbf{X}_\gamma\mathbf{B}_\gamma)\right\} \tag{2.7}$$

Due to the introduction of the model indicators and the break dates this is a mixture of discrete and continuous distributions. But given values for $\mathbf{k}_\gamma$, p and m, this is the kernel of the Normal-Inverse-Gamma-2 distribution.[4] For the unknown quantities I use the following prior distributions:

$$f(\mathbf{B}_\gamma|\sigma^2) = N_d(\mathbf{B}_0,\ \sigma^2 \cdot \mathbf{M}^{-1}) \tag{2.8a}$$

$$f(\sigma^2) = IG_2(a,b)\ , \quad a,b>0 \tag{2.8b}$$

$$f(p,m\,;T) \propto T^{-\frac{d\gamma}{2}}\ , \quad p \in \mathbb{N},\ m \in \mathbb{N}_0 \tag{2.8c}$$

$$f(k_i|k_{i-1}) \propto 1 \quad \text{, if } k_i \in \left]k_{i-1}\ ;\ k_{i+1}\right[\text{ and 0 otherwise , } i = 1...m+1 \tag{2.8d}$$

$$f(\mathbf{k}_\gamma) = \prod_{i=1}^{m+1} f(k_i|k_{i-1}) \text{ with } k_0 = p+1 \text{ and } k_{m+1} = T+1 \tag{2.8e}$$

The priors in (2.8a) and (2.8b) are the conjugate prior distributions for a Normal linear regression model, namely a multivariate Normal distribution with mean vector $\mathbf{B}_0 \in \mathbb{R}^d$ and $\mathbf{M}$ a positive definite symmetric $d \times d$ matrix[5] and the Inverse-Gamma-2 (IG_2) density with scale and shape parameters b and a, respectively. The assumptions (2.8d) and (2.8e) express lack of prior knowledge with respect to the break dates. As it can be observed from (2.8c) the two random variables m and p are assumed as stochastically independent of each other, for example in the 'drift-and-trend'-case the prior can be factorized as $f(\gamma) = f(p) \cdot f(m) = T^{-\frac{p}{2}} \cdot T^{-(m+1)}$. This is a strictly decreasing function in p and m, which implies that higher lag orders and/or number of breaks are considered to be

[4] Strictly speaking one could also condition on the upper bounds p_{max} and m_{max}, respectively, in the following. However in order not to overload the notation this conditioning is omitted.

[5] Here I choose $\mathbf{M} = \mathbf{I}_d/100$ and $\mathbf{B}_0 = \mathbf{0}$, respectively, to express a lack of knowledge with respect to the prior location and scaling of $\mathbf{B}_0$.

less likely, given a sample of size T. Furthermore it is a data-dependent[6] and informative prior in the sense that more complex, i.e. higher parameterized models are assigned a lower prior weight compared to less complex models and thus serves as a penalty factor in the acceptance ratios below.[7] The above priors together with the likelihood function in (2.7) lead to the following Bayesian hierarchy:

$$f(\lambda_\gamma, \gamma | \mathbf{y}, \mathbf{y_0}) \propto f(\mathbf{y}|\lambda_\gamma, \gamma; \mathbf{y_0}) \cdot f(\mathbf{B}_\gamma|\sigma^2, \mathbf{k}_\gamma, \gamma) \cdot f(\sigma^2|\mathbf{k}_\gamma,\ \gamma) \cdot f(\mathbf{k}_\gamma|\gamma) \cdot f(\gamma) \quad (2.9)$$

which is equal the joint posterior density of the parameters up to a normalizing constant. Before presenting an MCMC algorithm to generate random draws from the joint posterior distribution (2.9) for ADF-type models of varying dimension, I will first sketch the general idea of the underlying stochastic model selection scheme. In the following the conditioning on $\mathbf{y_0}$ will be dropped for notational convenience.

2.3. Stochastic model selection via MCMC

The main task in model determination is to find a single parametrization which describes the data best with respect to goodness of fit criteria. In Bayesian statistics this fit is measured in probabilistic terms by means of the posterior probability of a certain model M_i. In accordance with the model selection literature the parameter and the model space are distinguished in the following, where the former can be viewed as embedded in the latter, that is each model M_i is represented as a point in the space of candidate models. In the sequel a sampling scheme to conduct jumps between parameter spaces of varying dimensionality is proposed, that is to make moves within the model space. For the model in (2.5) let $\mathscr{M} = \{M_1, M_2, ...\}$ be a countable set of candidate models each of which is associated with a likelihood function $f(\mathbf{y}|\lambda_\gamma, \gamma)$, with unknown parameter vector $\lambda_\gamma \in \Lambda_\gamma = \mathbb{R}^d \times \mathbb{R}^+ \times \mathbb{N}^{m_{max}}_{[p+2;\ T-1]}$ and a vector of model indicators $\gamma \in \Gamma = \mathbb{N}_{[1;\ p_{max}]} \times \mathbb{N}_{[0;\ m_{max}]}$. For example, consider M_1 and M_2 to be two models indicated by $\gamma_1 = (p_1, m_1)'$ and $\gamma_2 = (p_2, m_2)'$ of dimensions h_1 and h_2 within the set of candidate models Γ. Then for each of these two models three types of model moves can be distinguished when

[6] This can be considered as an empirical Bayes approach, see Casella (1985).

[7] Note that this prior trades off the reduction in the residual variance against the inclusion of additional (regimewise) regressors in the spirit of information criteria, cf. Schwarz (1978).

moving from one state γ_1 to another state γ_2. The possible transitions are called 'Birth' (i.e. upward), 'Death' (i.e. downward) and 'Life' moves, where the first two are between-model moves and the latter are within-model moves. For example in the case of a 'p-Birth' move the Markov chain would jump from γ_1 to γ_2 with $p_2 > p_1$, and for the 'Death' move vice versa. The same applies of course to jumps in the m-dimension. As the name suggests 'Life' moves leave the dimension with respect to p and m unchanged, i.e. $p_2 = p_1$ and $m_2 = m_1$.

The present approach is most similar to that of Troughton and Godsill (1997a), who propose a sampling scheme for a lag order selection in autoregressive models, see also Godsill (2001), Ehlers and Brooks (2002). In accordance to the cited works I generate new values λ_2 as a full vector directly in the h_2-dimensional parameter space leaving the current value of the error variance σ^2 unchanged.[8] Conducting 'between' model moves, i.e. 'Birth' and 'Death' moves, with respect to the number of breaks $m^\star$ and/or the number of lags $p^\star$ is accomplished by Metropolis-Hastings (MH) steps (see Hastings (1970), Chib and Greenberg (1995)) with acceptance probabilities of a candidate move $\gamma^\star = (p, m^\star)'$ or $\gamma^\star = (p^\star, m)'$ depending on the context:

$$\alpha(\gamma,\ \gamma^\star) = \min\left\{1, \frac{f(\mathbf{B}_{\gamma^\star}, \gamma^\star, \mathbf{k}_{\gamma^\star}|\sigma^2, \mathbf{y})}{f(\mathbf{B}_\gamma, \gamma, \mathbf{k}_\gamma|\sigma^2, \mathbf{y})} \cdot \frac{\pi(\gamma^\star,\ \gamma)}{\pi(\gamma,\ \gamma^\star)} \cdot \frac{q(\mathbf{u}_\gamma|\mathbf{u}_{\gamma^\star})}{q(\mathbf{u}_{\gamma^\star}|\mathbf{u}_\gamma)}\right\} \tag{2.10}$$

Here $f(.)$ denotes the target density, $\pi(.)$ the proposal density for a model move according to $\gamma^\star$ and $q(.)$ a proposal density to draw a vector of model parameters $\mathbf{u}_{\gamma^\star}$ under the candidate model. Conducting 'Life' moves is done by leaving the components in γ unchanged and updating the remaining model parameters via Gibbs sampling steps (see Casella and George (1992)). This hybrid strategy to sample from the joint posterior distribution in (2.9) then proceeds as outlined in algorithm (1).

[8] This is the 'full parameter vector proposal' approach of (Troughton and Godsill, 1997a, p.5). The same approach is used by (Ehlers and Brooks, 2002, p.23) for their 'class B moves'.

Algorithm 1 : Hybrid sampler

Step 1: *Set the iteration counter on $j = 1$ and initialize $\lambda_\gamma^{(0)}, p^{(0)}, m^{(0)}$ randomly or deterministically.*

Step 2: *Propose a candidate for the lag order $p^\star$ from a proposal density π:*

- *if $p^\star > p_{max}$: set $p^\star = p_{max}$,*
- *if $p^\star < p_{min}$: set $p^\star = p_{min}$,*
- *otherwise accept the candidate move $p^\star$ with probability:*

$$\alpha_1(p,\ p^\star) = \min\left\{1, \frac{f(\mathbf{y}|\ p^\star, m, \mathbf{k}_{(p^\star,m)})}{f(\mathbf{y}|\ p, m, \mathbf{k}_{(p,m)})} \cdot \frac{\pi(p^\star, p)}{\pi(p, p^\star)} \cdot \frac{f(p^\star)}{f(p)}\right\} \tag{2.11}$$

- *Set $p^{(j)} = p^{\star(j)}$ if accepted, otherwise $p^{(j)} = p^{(j-1)}$*

Step 3: *Propose a new number of structural breaks $m^\star$ similar to step 2:*

- *if $m^\star > m_{max}$: set $m^\star = m_{max}$,*
- *if $m^\star < m_{min}$: set $m^\star = m_{min}$,*
- *otherwise accept the candidate move $m^\star$ with probability:*

$$\alpha_2(m,\ m^\star) = \min\left\{1, \frac{f(\mathbf{y}|\ m^\star, p, \mathbf{k}_{(p,m^\star)})}{f(\mathbf{y}|\ m, p, \mathbf{k}_{(p,m)})} \cdot \frac{\pi(m^\star, m)}{\pi(m, m^\star)} \cdot \frac{f(m^\star)}{f(m)}\right\} \tag{2.12}$$

- *Set $m^{(j)} = m^{\star(j)}$ if accepted, otherwise $m^{(j)} = m^{(j-1)}$*

Step 4: *Draw the i-th break date k_i from the full conditional multinomial posterior distribution $f(k_i^{(j)}|\ k_{i-1}^{(j-1)},\ k_{i+1}^{(j-1)},\ p^{(j)},\ m^{(j)},\ \mathbf{B}_\gamma^{(j-1)},\ \sigma^{2(j-1)},\ \mathbf{y})$ on the sample space $]k_{i-1}^{(j)},\ k_{i+1}^{(j)}[\ ,\ i = 1, ..., m$*

Step 5: *Draw a random vector $\mathbf{B}_\gamma^{(j)}$ from the full conditional multivariate normal posterior distribution $f(\mathbf{B}_\gamma^{(j)}|\ \mathbf{k}^{(j)},\ p^{(j)},\ m^{(j)},\ \sigma^{2(j-1)},\ \mathbf{y})$.*

Step 6: *Draw $\sigma^{2(j)}$ from the full conditional inverse gamma posterior distribution $f(\sigma^{2(j)}|\ \mathbf{B}_\gamma^{(j)},\ \mathbf{k}^{(j)},\ p^{(j)},\ m^{(j)},\ \mathbf{y})$.*

Step 7: *Set $j = j + 1$, return to step 2.*

As a proposal density $\pi(.)$ for the respective model moves I use a discretized Laplacian density (see Johnson and Kotz (1970)) centered over the current model as proposed by Godsill (2001), Ehlers and Brooks (2004) among others. Since this is a Random Walk proposal most moves are conducted in the neighborhood of the current model and thus most jumps will be small. However occasionally large jumps are also conducted and this ensures nice mixing and convergence properties of the corresponding Markov chains of m and p.[9] For example, in the case of a jump from state p to $p^\star$ I choose a Laplacian density of the form

$$\pi(p,p^\star) = \frac{1}{2\tau}\exp\left\{-\frac{|p^\star - p|}{\tau}\right\} \quad , \quad \tau \geq 0\ ,\ p^\star \in [1,\ p_{max}] \tag{2.13}$$

with mean p and variance $2\tau^2$. The same strategy is used in order to model jumps in the 'm'- dimension. To achieve the expressions of the MH acceptance ratios in (2.11) and (2.12) I use the full conditional posterior distribution of the vector of regression coefficients (see step 5) as a proposal density:[10]

$$\mathbf{u}_{\gamma^\star} \sim q(\mathbf{u}_{\gamma^\star}|\ \mathbf{k}_{\gamma^\star}, \gamma^\star, \mathbf{u}_\gamma,\ \sigma^2;\ \mathbf{y}) = N_{d^\star}(\mu_{\gamma^\star},\ \mathbf{\Sigma}_{\gamma^\star}) \tag{2.14a}$$

$$\text{with}\ \ \mu_{\gamma^\star} = \sigma^{-2}\cdot\mathbf{\Sigma}_{\gamma^\star}\cdot\mathbf{X}'_{\gamma^\star}\cdot\mathbf{y}\ , \tag{2.14b}$$

$$\mathbf{\Sigma}_{\gamma^\star} = \sigma^2\cdot(\mathbf{X}'_{\gamma^\star}\mathbf{X}_{\gamma^\star} + \mathbf{M})^{-1} \tag{2.14c}$$

and $d^\star$ the dimension of the proposed candidate vector. For example Ehlers and Brooks (2002) show by applying their 'second order method' that this constitutes an efficient proposal, see also Troughton and Godsill (1997*b*), Dellaportas et al. (2002) among others. Instead of drawing new values $\mathbf{u}_{\gamma^\star}$ from the proposal density in (2.14a) and substituting it together with the likelihood and the priors into equation (2.10), which could lead to numerical problems in the computation, I follow Troughton and Godsill (1997*b*) here by applying the 'Candidate's identity' of Besag (1989) to the present context, namely[11]

$$f(\mathbf{y}|\gamma,\mathbf{k}_\gamma,\sigma^2) = \frac{f(\mathbf{y}|\gamma,\mathbf{k}_\gamma,\mathbf{B}_\gamma,\sigma^2)\cdot f(\mathbf{B}_\gamma|\gamma,\mathbf{k}_\gamma,\sigma^2)}{f(\mathbf{B}_\gamma|\gamma,\mathbf{k}_\gamma,\sigma^2,\mathbf{y})} \tag{2.15}$$

[9] Note that for $\tau \to \infty$ the result is the uniform proposal for $p^\star \in [1,\ p_{max}]$.
[10] This approach is sometimes called 'Independent Reversible Jump MCMC sampler', see Lopes (2006).
[11] This is the 'basic marginal likelihood identity' used in Chib (1995).

so that the density in (2.14a) can equivalently be written as

$$q(\mathbf{u}_{\gamma^\star}|\mathbf{u}_\gamma) = \frac{f(\mathbf{y}|\gamma^\star,\mathbf{k}_{\gamma^\star},\mathbf{B}_{\gamma^\star},\sigma^2)\cdot f(\mathbf{B}_{\gamma^\star}|\gamma^\star,\mathbf{k}_{\gamma^\star},\sigma^2)}{f(\mathbf{y}|\gamma^\star,\mathbf{k}_{\gamma^\star},\sigma^2)} \tag{2.16}$$

Now substituting expression (2.16) into (2.10) and integrating out σ^2 in both the numerator and the denominator analytically, the acceptance ratio simplifies to:

$$\alpha(\gamma,\gamma^\star) = \min\left\{1, \frac{f(\mathbf{y}|\gamma^\star,\mathbf{k}_{\gamma^\star})}{f(\mathbf{y}|\gamma,\mathbf{k}_\gamma)}\cdot\frac{f(\gamma^\star,\mathbf{k}_{\gamma^\star})}{f(\gamma,\mathbf{k}_\gamma)}\cdot\frac{\pi(\gamma^\star,\gamma)}{\pi(\gamma,\gamma^\star)}\right\} \tag{2.17}$$

The above acceptance probability thus reduces to the posterior odds ratio in favor of a model with $\gamma^\star$ and $\mathbf{k}_{\gamma^\star}$. Given the prior assumptions (2.8c) - (2.8e) the expression in (2.17) is proportional to the ratio of the marginalized (model-specific) likelihoods times the ratio of the jump probabilities.[12] The model likelihoods in (2.17) have the form of multivariate Student-t densities, $\mathscr{T}_{T-p}(\mu = \mathbf{X}_\gamma\mathbf{B}_0,\ \mathbf{P} = (\mathbf{I}_{T-p} + \mathbf{X}_\gamma\mathbf{M}^{-1}\mathbf{X}'_\gamma)^{-1}/b,\ \nu = a)$, with a degrees of freedom (see appendix G for details), which are given by

$$f(\mathbf{y}|\gamma,\mathbf{k}_\gamma) = C\cdot\left[1 + (\mathbf{y}-\mathbf{X}_\gamma\mathbf{B}_0)'\cdot\frac{\left(\mathbf{I}_{T-p}+\mathbf{X}_\gamma\mathbf{M}^{-1}\mathbf{X}'_\gamma\right)^{-1}}{b}\cdot(\mathbf{y}-\mathbf{X}_\gamma\mathbf{B}_0)\right]^{-\frac{T-p+a}{2}} \tag{2.18}$$

with normalizing constant

$$C \equiv \pi^{-\frac{T-p}{2}}\cdot\left|\mathbf{I}_{T-p}+\mathbf{X}_\gamma\mathbf{M}^{-1}\mathbf{X}'_\gamma\right|^{-\frac{1}{2}}\cdot b^{\frac{a}{2}}\cdot\left[\Gamma\left(\frac{T-p+a}{2}\right)/\Gamma\left(\frac{a}{2}\right)\right] \tag{2.19}$$

Furthermore in step 5 of the sampling scheme above the random vectors $\mathbf{B}_\gamma$ are drawn from a multivariate normal distribution with mean vector $\mu_\gamma = \sigma^{-2}\cdot\mathbf{\Sigma}_\gamma\cdot\mathbf{X}'_\gamma\cdot\mathbf{y}$ and covariance matrix $\mathbf{\Sigma}_\gamma = \sigma^2\cdot(\mathbf{M}+\mathbf{X}'_\gamma\mathbf{X}_\gamma)^{-1}$. The $IG(a^\star,b^\star)$ posterior distribution of step 6 has shape parameter $a^\star \equiv a + \frac{T}{2}$ and scale parameter $b^\star \equiv b + \frac{1}{2}(\mathbf{y}-\mathbf{X}_\gamma\mathbf{B}_\gamma)'\cdot(\mathbf{y}-\mathbf{X}_\gamma\mathbf{B}_\gamma)$. Before turning to the empirical analysis of OECD unemployment rates, it is useful to evaluate the performance of the outlined sampling algorithm for model selection and also the power of the Bayesian unit root test.

[12] A similar result is stated by Dellaportas et al. (2002) in context of their 'Metropolised Carlin and Chib' approach when using the posterior distribution for each model M_i as a *pseudo-prior*, see Dellaportas et al. (2002), p.30 for details.

2.4. Monte Carlo evidence: MCMC based model selection

With the stochastic model selection procedure presented in the last section, the (conditional) posterior distributions of the parameters p, m and $\mathbf{k}_\gamma$ can be approximated. These can then be used for further analysis. For example, Koop and Potter (1999) point out, that in a Bayesian approach no single model has to capture the true DGP, instead we can weight features of interest (as the long run coefficient θ in the present context) from different models by their respective posterior model probabilities. Working with such model averaged quantities is particularly attractive in disciplines like economics where theoretical considerations do not always suggest which model specification is best. In contrary to classical methods where the induced uncertainty of the model selection step can not be fully captured through the statistical measures, the underlying Bayesian approach allows a probabilistic representation of this uncertainty through the shape (e.g. multimodality, platykurtosis) of the corresponding posterior distributions. To get an impression of the sampler's performance, especially with regard to model selection, trajectories of moderate lengths ($T = 200$) are simulated for three ARMA(p,q)-processes without breaks and also three processes with breaks. Then the above sampler is run for 10000 iterations omitting the first 1000 random draws due to burn-in. The scale parameter of the Laplacian jump proposal (2.13) and the maximum number of lags are chosen to be $\tau = 5$ and $p_{max} = 15$, respectively. The DGPs considered here are as follows:

- No-break design 1: $y_t = 0.8y_{t-1} - 0.35y_{t-2} + \varepsilon_t$
- No-break design 2: $y_t = 0.2 + 0.1 \cdot t + 0.5y_{t-1} + 0.2y_{t-2} - 0.3y_{t-3} + 0.5y_{t-4} - 0.3y_{t-5} - 0.21y_{t-6} + \varepsilon_t$
- No-break design 3: $y_t = 0.7y_{t-1} + 0.12y_{t-2} + 0.22y_{t-3} - 0.15y_{t-4} - 0.5y_{t-5} + 0.4y_{t-6} - 0.35y_{t-7} + 0.23y_{t-8} + 0.21y_{t-9} - 0.4y_{t-10} + 0.2y_{t-11} - 0.2y_{t-12} + \varepsilon_t$

each with $\varepsilon_t \overset{i.i.d.}{\sim} N(0, \sigma = 0.25)$.

Since in the above three examples interest primary is on the identification of the true lag order, the number of structural breaks is fixed at $m = 0$ with the exception of simulation

design 3, where a maximum number of $m_{max} = 5$ breaks is allowed. Figures 2.1-2.3 show simulated trajectories together with the resulting posterior distributions of the lag order $p|(m = 0)$, the long run coefficient $\theta|(\widehat{p_{MAP}},\ m = 0)$, and in case of design 3 also of the posterior of the number of breaks m, where $\widehat{p_{MAP}}$ denotes the maximum a posteriori (MAP) estimate of p.[13] As is evident from the three examples, the sampler generates

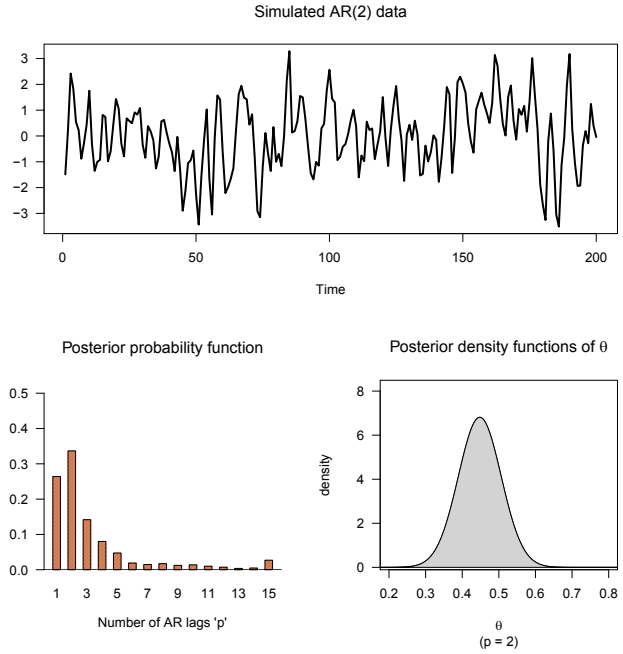


Figure 2.1.: No-break design 1: trajectory of an AR(2) process together with the posteriors of the lag order and of the sum of AR coefficients θ.

unimodal distributions with modes equal to the true parameter values so that the MAP estimator yields consistent results. Next I simulate three series with breaks in the level and/or trend of the process. The data are generated according to

- Break design 1: $y_t = a_t + b_t \cdot t + 0.55y_{t-1} + 0.35\varepsilon_t \ \ , \ \varepsilon_t \sim N(0,1)$

with regime specific trend parameters $a_t = \alpha_1 = 1$, $b_t = \beta_1 = 0.01$ for $1 \leq t \leq 33$, $a_t = \alpha_2 = -0.2$, $b_t = \beta_2 = 0.01$ for $33 < t \leq 82$, $a_t = \alpha_3 = 0.5$, $b_t = \beta_3 = 0.015$ for $82 < t \leq 121$, $a_t = \alpha_4 = 1.6$, $b_t = \beta_4 = 0.02$ for $121 < t \leq 151$, and $a_t = \alpha_5 = 0.4$, $b_t = \beta_5 = 0.02$ for $151 < t \leq 200$, so that the vector of break dates is $\mathbf{k} = (33, 82, 121, 151)'$.

[13] This point estimator is also called the *generalized* maximum likelihood estimator, see DeGroot (1970), p.236.

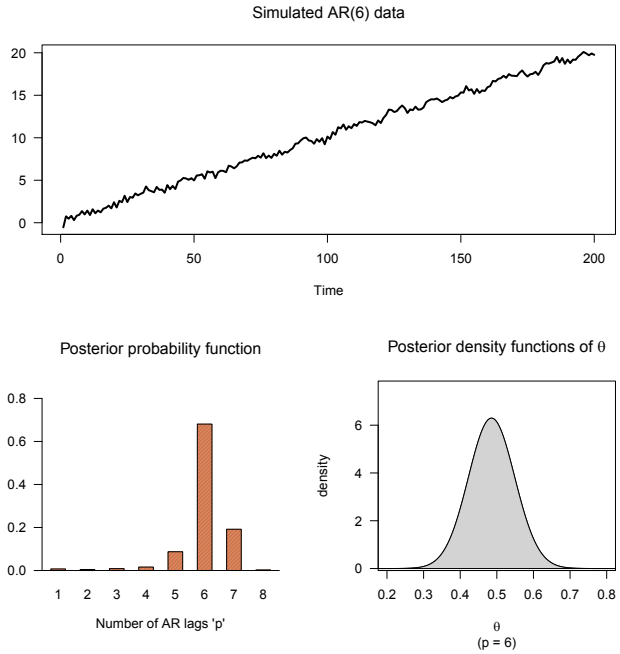


Figure 2.2.: No-break design 2: trajectory of an AR(6) process with drift and trend together with the posteriors of the lag order and of the sum of AR coefficients θ.

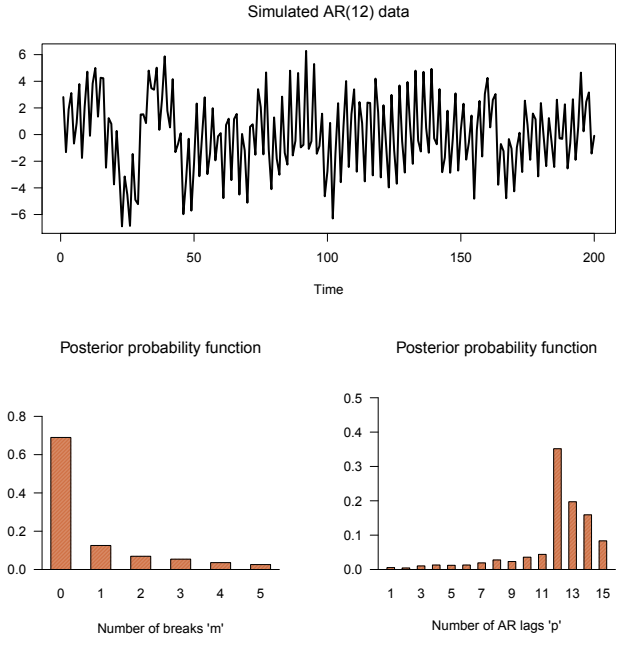


Figure 2.3.: No-break design 3: trajectory of an AR(12) process together with the posterior distributions of the lag order and the number of breaks.

- Break design 2: $y_t = a_t + y_{t-1} + 0.55\varepsilon_t \ , \ \varepsilon_t \sim N(0,1)$

with regime specific trend parameters $a_t = \alpha_1 = 0.15$ for $1 \leq t \leq 33$, $a_t = \alpha_2 = 0.75$ for $33 < t \leq 151$, and $a_t = \alpha_3 = 0.1$ for $151 < t \leq 200$, so that the vector of break dates is $\mathbf{k} = (33, 151)'$.

- Break design 3: $y_t = a_t + b_t \cdot t + 0.95y_{t-1} - 0.35y_{t-2} + 0.3\varepsilon_{t-1} + \varepsilon_t \ , \ \varepsilon_t \sim N(0, 0.1)$

with regime specific trend parameters $a_t = \alpha_1 = 0.12$, $b_t = \beta_1 = -0.01$ for $1 \leq t \leq 121$, and $a_t = \alpha_2 = 0.3$, $b_t = \beta_2 = -0.009$ for $121 < t \leq 200$, with the single break date $k_1 = 121$.

Figures 2.4-2.6 show the simulated paths together with the posterior distributions of p, m and $\mathbf{k}_\gamma$. From figure 2.4 we observe that the lag order and the number of breaks are

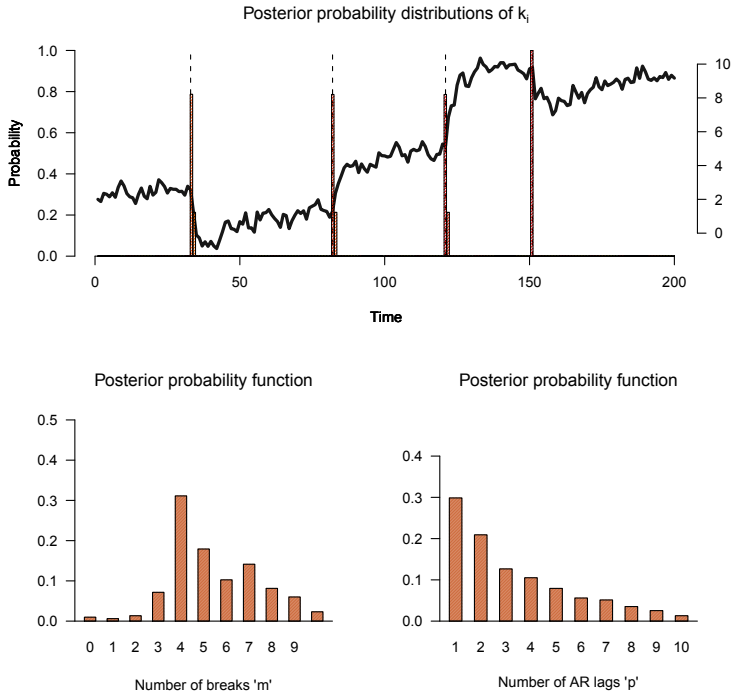


Figure 2.4.: Break design 1: trajectory of a stationary AR(1) with four breaks in level and trend (dashed lines for true break dates) together with the posterior distributions of the lag order, the number of breaks and the break dates k_i given the MAP estimate of m.

chosen correctly with $p = 1$ and $m = 4$. The corresponding four posterior distributions of the break dates are all concentrated around a single mass point which coincides with

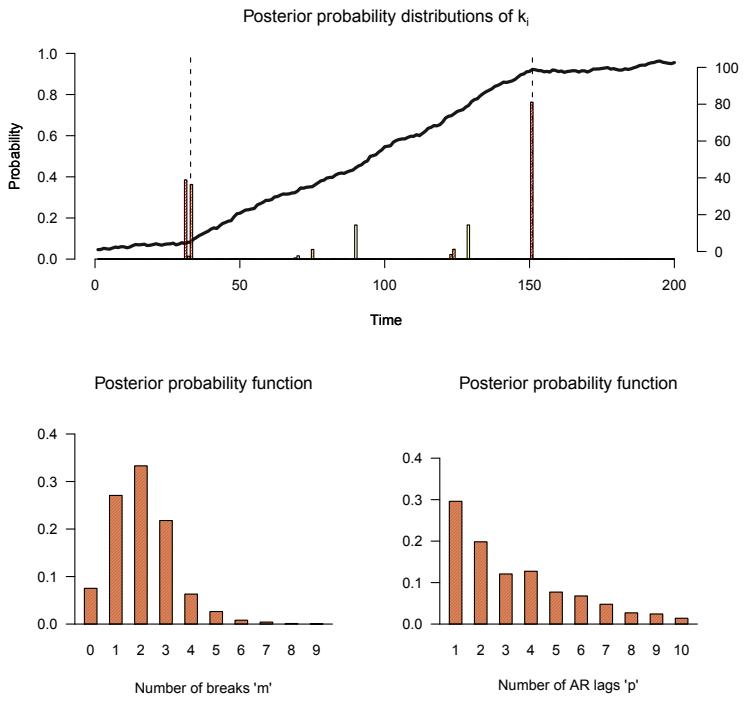


Figure 2.5.: Break design 2: trajectory of a Random Walk with two drift breaks (dashed lines for true break dates) together with the posterior distributions of the lag order, the number of breaks and the break dates k_i given the MAP estimate of m.

the respective true break date.[14] In design 2 the DGP is specified to give an impression of how the model selection procedure performs in the context of nonstationary processes with structural breaks. Therefore a Random Walk with two drift breaks is simulated for $T = 200$. Even in this case the two break points can be identified relatively precisely, although the posterior distribution of k_1 shows more variation compared to that of k_2. The distributions of m and p have the expected modes at the parameter values of the DGP, i.e. $m = 2$ and $p = 1$. The third series allows the investigation of more general ARMA processes. Figure 2.6 depicts the selection results when the data are generated by an ARMA(2,1)-process with one level break. The selection of the AR order and the number of structural breaks are not appreciably influenced by the addition of extra noise due to an MA(1)-component in this case. In summary, the results of the MC experiments indicate favorable performance of the proposed MCMC approach in finding the true parameter values of the DGP. Before turning now to the real data analysis the chosen Bayesian unit

[14]The lack of variation in the posterior distributions of the k_i's are due to the choice of the scale parameter in the Inverse-Gamma prior for σ^2 with $b = 0.001$ (see equation (2.8b)).

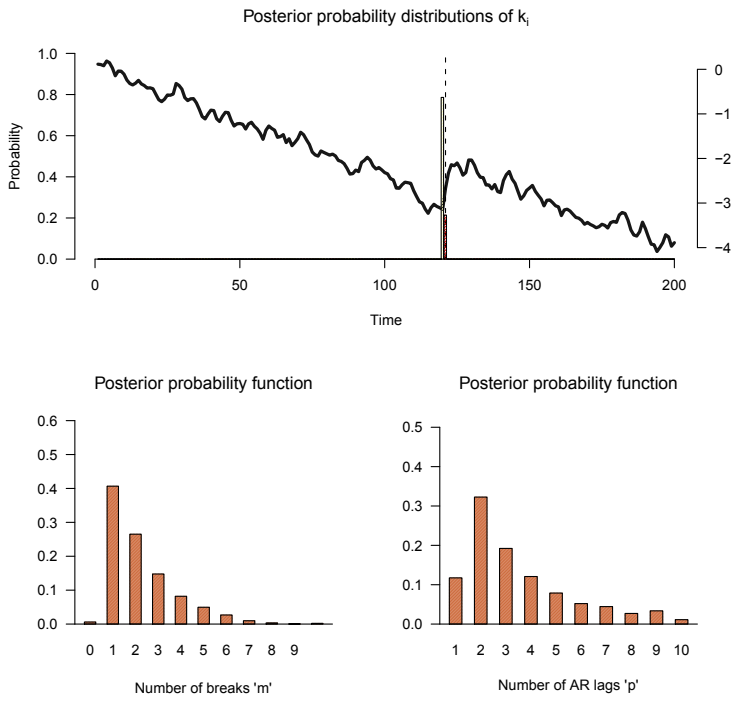


Figure 2.6.: Break design 3: trajectory of a stationary ARMA(2,1) with one level break (dashed line for true break date) together with the posterior distributions of the lag order, the number of breaks and the break dates k_i given the MAP estimate of m.

root testing approach is introduced.

2.5. Testing the unit root null hypothesis

The Bayesian key device for unit root testing, using the structural breaks model in (2.5), is the likelihood function $f(\mathbf{y}|\theta,\gamma,\mathbf{k}_\gamma)$ with $\theta \in \Theta = [0;1]$. Let the parameter set Θ be partitioned into $\Theta_0 = \{1\}$ and $\Theta_1 = \Theta \backslash \{1\}$. For testing the sharp null hypothesis of a unit root $H_0: \ \theta \in \Theta_0$ against the alternative of a covariance stationary process $H_1: \ \theta \in \Theta_1$ it is natural to compare the corresponding posterior mass of these two disjoint sets and to reject the null if $P(\Theta_1|\mathbf{y}) > 0.5$, see Robert (2007), p.225 for details.[15] The above pair of hypotheses is also the starting point for many other authors to test for a unit root, see Maddala and Kim (1998) and Bauwens et al. (1999) for overviews concerning other

[15]Note that in classical terminology $1 - P(\Theta_0|\mathbf{y})$ plays the role of a test statistic, as a function of the sample.

Bayesian approaches to unit root testing. In order to give the unit root null more 'weight', a mixed prior density as in Kadane et al. (1996) is used that assigns a positive prior probability $\pi_0 \equiv f(\theta|H_0) > 0$ to the singleton Θ_0 and uses a continuous density $f(\theta|H_1)$ with weight $1-\pi_0$ elsewhere, cf. Berger and Delampady (1987), Berger and Sellke (1987) for details. Note that by assuming a continuous prior on $\Theta = \Theta_0 \bigcup \Theta_1$ the simple unit root hypothesis would always be rejected, since Θ_0 has zero Lebesgue measure.[16] Given values for γ and $\mathbf{k}_\gamma$ the posterior probability of H_0 can be expressed as a function of the conditional Bayes factor (BF) in favor of the null hypothesis.[17] The posterior probability of $\theta_0 = 1$ is then obtained analogously to the unconditional case (see Berger and Delampady (1987), also DeGroot (1970), p.238) as

$$P(H_0|\gamma,\mathbf{k}_\gamma;\mathbf{y}) = \frac{f(\mathbf{y}|\theta_0,\gamma,\mathbf{k}_\gamma)\cdot\pi_0}{\pi_0\cdot f(\mathbf{y}|\theta_0,\gamma,\mathbf{k}_\gamma)+(1-\pi_0)\cdot f(\mathbf{y}|H_1,\gamma,\mathbf{k}_\gamma)} \tag{2.20a}$$

$$= \left[1+\frac{1-\pi_0}{\pi_0}\cdot\frac{1}{BF}\right]^{-1} \tag{2.20b}$$

$$\text{with } BF = \frac{f(\mathbf{y}|\theta_0,\gamma,\mathbf{k}_\gamma)}{f(\mathbf{y}|H_1,\gamma,\mathbf{k}_\gamma)} \tag{2.20c}$$

$$\text{and } f(\mathbf{y}|H_1,\gamma,\mathbf{k}_\gamma) = \int_{\Theta_1} f(\mathbf{y}|\theta,\gamma,\mathbf{k}_\gamma)\cdot f(\theta|H_1)\cdot d\theta \tag{2.20d}$$

where the posterior probability of the null, given in (2.20), is abbreviated by P_0 henceforth.

To utilize the results from the above sections in order to derive the likelihood function of $\theta|(\gamma,\mathbf{k}_\gamma)$ simply define $\mathbf{B}_{-\theta} = \mathbf{B}_\gamma \backslash \{\theta\}$ to be the vector of regression coefficients without the long run coefficient θ and $\mathbf{y}(\theta) \equiv \mathbf{y} - \theta\mathbf{y}_{-1}$ a linear function in θ with $\mathbf{y} = (y_{p+1},...,y_T)'$ and $\mathbf{y}_{-1} = (y_p,...,y_{T-1})'$. Then the conditional likelihood of θ given γ and $\mathbf{k}_\gamma$ is obtained by integrating out all other parameters $\mathbf{B}_{-\theta}$ and σ^2 from (2.7) and therefore the derivation of $f(\mathbf{y}(\theta)|\gamma,\mathbf{k}_\gamma)$ is the same as for $f(\mathbf{y}|\gamma,\mathbf{k}_\gamma)$. Looking at the

[16] As point hypotheses can be perceived as approximations to interval hypotheses, the probability π_0 can be thought of as the mass that would have been assigned to the (more realistic) interval hypothesis $H_0 : \theta \in [\theta_0 - c,\ \theta_0 + c]$, $c > 0$, see Berger (1980), p.150, also Robert (2007), p.230.

[17] Strictly speaking, this is not a real Bayes factor, because the involved likelihood expressions are not the marginal likelihoods as in the definition of a Bayes factor (see Kass and Raftery (1995)), since they are still conditional on γ and $\mathbf{k}_\gamma$, i.e. a particular candidate model.

integrand of (2.20d) the first of these two densities is then given by

$$f(\mathbf{y}|\theta,\gamma,\mathbf{k}_\gamma) \propto |\mathbf{S}|^{-\frac{1}{2}} \cdot \left\{ b + (\mathbf{y} - \theta\mathbf{y}_{-1})' \cdot \mathbf{S}^{-1} \cdot (\mathbf{y} - \theta\mathbf{y}_{-1}) \right\}^{-(T-p+a)/2} \tag{2.21a}$$

$$\text{with } \mathbf{S} \equiv \left(\mathbf{I}_{T-p} + \mathbf{X}_\gamma \mathbf{M}^{-1} \mathbf{X}_\gamma' \right) \tag{2.21b}$$

As a prior density $f(\theta)$, with $\theta \in \Theta$, for the computation of (2.20d) I use three different distributions, namely a conjugate normal prior, as for example advocated by Uhlig (1994), a simple flat prior as a benchmark and the approximate Jeffreys prior as proposed by Phillips (1991*b*), see also Zivot and Phillips (1994), which has the form

$$f_J(\theta;T) = \begin{cases} \left(\frac{1}{1-\theta^2} \left[T - \frac{1-\theta^{2T}}{1-\theta^2} \right] \right)^{1/2} & \text{, for } \theta \neq 1 \\ \left(\frac{T(T-1)}{2} \right)^{1/2} & \text{, for } \theta = 1, \end{cases} \tag{2.22}$$

Figure 2.7 depicts the shapes of these priors, which allocate relatively similar weight over the stationary region up to, say $\theta < 0.8$. To check the sensitivity of the inferential results

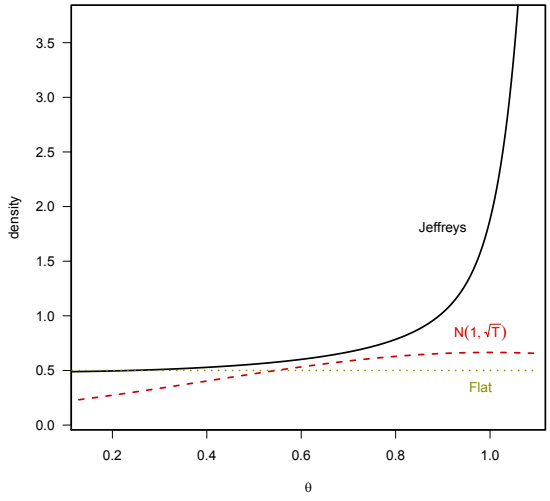


Figure 2.7.: Prior distributions for the long run coefficient θ for $T = 200$.

concerning the above prior specifications I compute the corresponding risk functions under squared error loss using the no-break model in (2.4), i.e. with $\gamma = (p,0)'$ and $\mathbf{k}_\gamma = \mathbf{0}'$:

$$R(\delta(\mathbf{y}),\theta_0) = \int_{\mathbb{R}^{T-p}} [\delta(\mathbf{y}) - \theta_0]^2 \cdot f(\mathbf{y}|\theta_0,\gamma,\mathbf{k}_\gamma) \cdot d\mathbf{y} \tag{2.23a}$$

$$\simeq \frac{1}{N}\sum_{i=1}^{N} [\delta(\mathbf{y}_i) - \theta_0]^2 \tag{2.23b}$$

$$\text{with } \delta(\mathbf{y}_i) \equiv E(\theta|\gamma,\mathbf{k}_\gamma;\ \mathbf{y}_i), \tag{2.23c}$$

where the vector $\mathbf{y}_i$, $i = 1...N$ denotes a draw from $f(\mathbf{y}|\theta_0,\gamma,\mathbf{k}_\gamma)$ of dimension $T-p$.

The risk functions are computed as in Bauwens et al. (1999), chapter 6, for the no-break model (2.4). Here I define a grid of values for $\theta \in [0.1;1.1]$ and then generate a trajectory $\{y_t\}_{t=1}^{T=200}$ from $y_t = 0.5 + \theta y_{t-1} + 0.5\varepsilon_{t-1} + \varepsilon_t,\ \varepsilon_t \sim IN(0,1)$. Next the conditional posterior expectation[18] of θ for the ADF-model in (2.4) without a time trend is computed, given a lag order of $p = \lfloor T^{\frac{1}{3}} \rfloor$, with $\lfloor x \rfloor$ the largest integer not greater than x. Repeating this for $i = 1...1500$ and then taking the sample average with respect to the N squared loss functions in the approximation of (2.23) finally yields the frequentistic risk of the Bayes estimator (2.23c) under the respective prior.[19] Figure 2.8 shows the risk functions under the three considered prior distributions. As most of the test decisions in this chapter are based upon the normal prior, a sensitivity analysis of the point estimation is conducted using different location/scale specifications. To compare the impact of the location parameter on the point estimation of θ, I choose a normal prior centered over a Random Walk and also a normal prior with mean $\mu_\theta = 0$, which is thus not informative with respect to the long run impact coefficient θ. The prior information gets more accentuated when represented by densities of moderate dispersion, like for example $\sigma_\theta^2 = 0.04$, and becomes less influential when considering flat densities using for example $\sigma_\theta^2 = 40$. Consequently the point estimation under the $N(1,0.2)$ prior performs best when approaching the null constraint $\theta = 1$. However, for short time series, e.g. $T < 50$,[20] this can be problematic since the prior weighting then dominates the likelihood and so the posterior will be shifted

[18] The one-dimensional integration for the posterior expectation in (2.23c) is computed numerically via Simpson's rule on 31 points over a grid of values $\theta \in [0.1;1.1]$ as in Bauwens et al. (1999), p.178. For the computations all priors are normalized to integrate to one, i.e. to be proper densities.

[19] Note that the posterior Bayes estimator (2.23c) minimizes the Bayes risk, or equivalently the posterior expected loss, under a squared loss function for a given prior.

[20] This situation is frequently encountered when working with annual data, for example most of the OECD series used in the empirical analysis (see section 2.7) have lengths between $T = 25$ and $T = 45$.

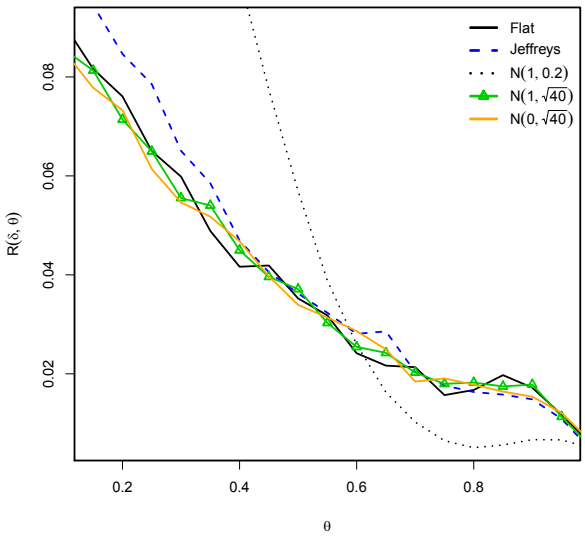


Figure 2.8.: Risk functions for δ under different prior distributions.

towards a Random Walk. For this reason I use a data-dependent $N(1, \sqrt{T})$ prior, which is rather flat and hence does not put so much weight on the unit root constraint.

2.6. Power comparison of unit root testing procedures

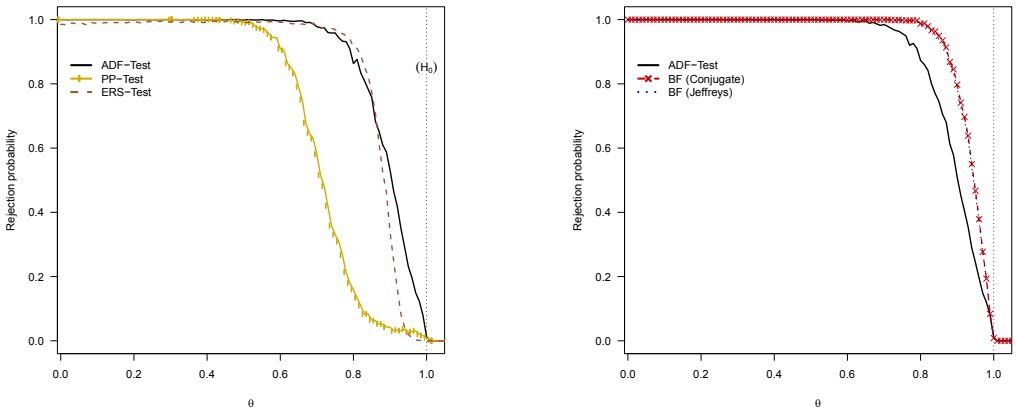
To compare the above Bayesian testing procedure under various prior distributions with some commonly used classical unit root tests, I compute the power functions $\beta_T(\theta)$ for the Augmented Dickey Fuller (ADF)-test (Dickey and Fuller (1979)), the Phillips and Perron (1988) (PP)-test and also the Elliot et al. (1996) (ERS)-test.[21] The rejection probabilities of the null hypothesis under a specific model are approximated by the average number of rejections, i.e.

$$P(\text{'Reject } H_0\text{'}|\theta_0, \gamma, \mathbf{k}_\gamma, \mathbf{y}) \approx \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}_{\{BF<1\}} \tag{2.24}$$

[21] These are readily available in the `R` package `urca`, see `www.r-project.org`.

where $\mathbf{1}_{\{.\}}$ is the indicator function and BF denotes the (model-specific) Bayes factor in favor of H_0 as given in (2.20c).[22]

For the computation I simulate ARMA(1,1) trajectories for different sample sizes according to $y_t = \theta y_{t-1} + 0.55\varepsilon_{t-1} + \varepsilon_t$, $\varepsilon_t \sim IN(0,1)$ over a grid $\theta \in [0;1.1]$ of 100 points and for each of these parameter values I calculate the posterior probability P_0 using the above priors. Further the above three classical unit root tests are applied for a nominal significance level of 5%. This is repeated $M = 5000$ times and at each run the rejections are counted to approximate the above probabilities in (2.24). As can be seen from figure 2.9(a) the PP test is clearly dominated by the two other tests. On the other hand, the ADF test is strictly dominated by the Bayes test (regardless of the prior) as figure 2.9(b) illustrates. The addressed dominance of the Bayes test gets more accentuated when the



(a) Classical (b) Classical vs. Bayesian

Figure 2.9.: Power functions of classical and Bayesian unit root tests without structural breaks.

sample size T increases.[23] This can be seen from the results of table 2.1, where the posterior probability of a unit root is shown for different sample sizes T and θ-values in the DGP. The results suggest that the Bayes test outlined above forms a consistent test in the sense that the power function $\beta_T(\theta) \to 1$ as $T \to \infty$, for $\theta \in \Theta_1$, which is equivalent to $P_0 \to 0$ as $T \to \infty$, for $\theta \in \Theta_1$.

[22] An equivalent decision rule in terms of the null posterior probability P_0 would be to reject the null hypothesis, if $1 - P_0 > 0.5$. Note that the null probability P_0 in (2.20) is an increasing function of the Bayes factor.

[23] The results are obtained using the conjugate normal distribution. For the other priors similar results are produced.

Table 2.1.: Posterior probability of a unit root as a function of T and θ

T	$\theta = 0.75$	$\theta = 0.8$	$\theta = 0.85$	$\theta = 0.9$	$\theta = 0.95$	$\theta = 1.0$
20	0.2575	0.3087	0.3443	0.4082	0.4881	0.6521
50	0.1686	0.2403	0.3109	0.4243	0.6034	0.8917
100	0.0722	0.1358	0.2102	0.3645	0.6038	0.9898
150	0.0240	0.0496	0.1382	0.3115	0.6074	0.9998
200	0.0065	0.0194	0.0709	0.2219	0.5472	0.9999
250	0.0016	0.0072	0.0272	0.1416	0.5278	1.0000

Overall the normal prior provides slightly better results than the ('objective') Jeffreys prior of Phillips (1991*b*), see figure 2.9(b), with respect to power and also with respect to frequentistic risk when taking the posterior mean as a point estimator. Looking at the power functions there seems no evidence against using the Jeffreys prior for testing, although experience shows that this prior is likely to produce bimodalities when approaching the nonstationarity region. To draw my conclusions the conjugate normal prior is used in the following empirical section and the Jeffreys prior is computed for reasons of comparison.

2.7. Empirical application using OECD data

Next the above two-stage procedure is utilized to test for possible unemployment persistence among 17 OECD countries. The data set consists of annual unemployment rates observed within the time interval from 1960 to 2010.[24] The high level of unemployment in countries of the European union compared to other countries of the OECD has been an object of investigation for many years. In the economic literature in principal there are two theoretical explanations for this phenomenon: the non-accelerating inflation rate of unemployment (NAIRU) and unemployment hysteresis. The former theory implies the unemployment rate to follow a trend stationary process, i.e. after an exogenous shock the rate will recover to its long run equilibrium. By contrast, hysteresis implies that temporary shocks have permanent effects on the level of unemployment and thus the

[24]The data was extracted from OECD online sources. The reported unemployment rates were computed as the ratio of the number of unemployed persons and the number of persons in the labour force, where the latter is defined as the sum of employed and unemployed persons.

underlying stochastic process has a unit root. This 'unit root hysteresis' definition is the most common in the literature on the stochastic properties of unemployment rates (see also Blanchard and Summers (1986)). Persistence can now be regarded as a special case of the NAIRU concept in the sense that the unemployment rate in fact follows a stationary process but also has a stochastic component that is nearly integrated of order one, for example with the sum of autoregressive coefficients being very close to one ('Quasi Random Walk'). Structuralist theories of unemployment as described in Phelps (1994) consider the natural rate of unemployment, as implied by the NAIRU, to be a function of different macroeconomic variables like the oil price, stock prices or the world real rate of interest (see Layard et al. (1991)). If the unemployment rate reaches its equilibrium path after a shock but this path is now higher (lower) than before, then the unemployment rates can be regarded as being generated by a trend stationary process with breaks. To test for possible unit root hysteresis in the OECD unemployment rates the corresponding posterior probabilities P_0, for $\pi_0 = 0.5$ are computed (see expression (2.20b)). In addition, the Bayesian tail probabilities $P(\theta \geq 0.975|\ \widehat{\gamma},\ \widehat{\mathbf{k}},\ \mathbf{y})$ to test for stochastic nonstationarity, as in Phillips (1991*b*), Summers (2004), are calculated.

Before proceeding with the analysis of the OECD unemployment rates one general note concerning the unit root testing of limited variables is in order. Because unemployment rates are bounded between zero and one they cannot be generated by a Random Walk process, because such a process would (at least theoretically) cross every boundary almost surely as the sample size increases. Since under the unit root hypothesis the data are postulated to be generated by a Random Walk, which is a special case of an $I(1)$ process,[25] namely linear with (symmetric and nontruncated) normal innovations, the above test regression would be inappropriate to describe the DGP (see also Koop and Potter (1999)). Cavaliere (2005*a*) points out that this neglect makes the interpretation of unit root tests controversial. He argues that given the researcher has rejected the $I(1)$ hypothesis, it is not clear if this rejection is due to the presence of $I(d)$ dynamics, $|d| < 1$, or due to the existing range constraints that have not been considered. In the present context of unemployment hysteresis this question is important, especially since the Random Walk hypothesis is usually rejected when applied to limited time series, e.g. unemployment rates or nominal interest rates, while it is not rejected in the majority of empirical applications to unlimited time series (see Cavaliere (2005*a*) for details). There are now two obvious ways to handle

[25] Here $I(d)$ means integrated of order d, see Banerjee et al. (1993).

this problem, namely either to change the test regression (2.5) in order to be able to handle truncated distributions or to make a change of variables that transforms the (double) truncated distribution of y_t to a nontruncated distribution $f(y_t^\star) \cdot \mathbf{1}_{\mathbb{R}}(y_t^\star)$. Here I choose the latter route. Following Wallis (1987), the unemployment rates y_t are transformed by means of the logistic transformation $f: \; y_t \mapsto y_t^\star = \ln(\frac{y_t}{1-y_t})$, i.e. $[0;1] \mapsto \mathbb{R}$. Then the hybrid sampler for the model in (2.5) without a time trend[26] is run to get estimates for the number of structural breaks, the number of lags and the corresponding break dates. As the sampler operates on both the parameter and the model space we get a sample $\{\gamma_l, \; l = 1,...,L\}$ for the model indicators, where γ denotes a certain (p,m)-combination in the $\Gamma = \mathbb{N}_{[1;\, p_{max}]} \times \mathbb{N}_{[0;\, m_{max}]}$ space. Thus the posterior probabilities $f(\gamma|\mathbf{y})$ can be estimated by (cf. Lopes (2006), p.3):

$$\widehat{f}(\gamma|\mathbf{y}) = \frac{1}{L} \sum_{l=1}^{L} \mathbf{1}_{\{\gamma=\gamma_l\}} \tag{2.25}$$

In order to account for any uncertainty induced through the model selection step the model probabilities in (2.25) are used as weights to compute a model average $\widehat{\theta}_{\overline{M}}$ over all considered submodels $\mathscr{M} = \{M_1,...,M_K\}$. Utilizing the Bayes estimates $\widehat{\theta}$ under the different priors and also the model averaged estimate $\widehat{\theta}_{\overline{M}}$ one can compute the half life $HL \equiv \ln(0.5) / \ln(\widehat{\theta})$ as a measure of persistence (or convergence) for each of the OECD series. In the present context, the half life means the expected number of years for an unemployment shock to decay by 50%. To have a classical benchmark I apply the original ADF test without a break and calculate the half life using the resulting point estimate of θ. Due to the relatively short series and also for reasons of comparison I allow for a maximum number of five level breaks and also for a maximum number of five lags. As can be seen from the results shown in table A.1 (see appendix A) for each of the 17 OECD countries at least two structural breaks are identified. In table A.2 the posterior probabilities of the number of lags together with the selected lags according to the AIC and BIC information criteria are reported.[27] The results indicate that for the OECD data one and two autoregressive lags have the highest posterior probabilities. With regard to the

[26] This specification is chosen mainly for reasons of comparison with the empirical results of other authors, in particular the results of Papell et al. (2000) and Summers (2004), where the latter author chooses a Bayesian frame of reference.

[27] For the models with break(s) these were calculated using the respective posterior estimates of break points from the MCMC output.

lag selection, BIC and an inspection of the sample partial autocorrelation function yield identical values in about 71% and 77% of all considered cases compared to the MCMC approach.[28] For the number of breaks there is an accordance with BIC in about 41% of all cases (see table A.1). Next the Bayesian results for the determination of the break numbers and also the break dates are compared with those obtained from an application of the methods described in Bai and Perron (2003).[29] A comparison of table A.3 with table A.4 reveals, that for about 41% of all countries there is an accordance of the selected number of breaks and that the corresponding change points are comparable in location for most of the countries. Summing up, the Bayesian model selection approach suggests slightly more parsimonious, i.e. less complex models than its classical competitors, although the results of the lag order selection are very similar.

Note that the data also cover events of possible structural changes like the two oil crises of 1973 and 1979 as well as more recent events like the (first) financial crisis of 2008. From A.3 it can be observed that the years 1973 and 1979 are identified as shocks to the national labor markets at least for some countries in the OECD. The impacts of these shocks appear with a lag of one or two years so that the posterior distributions of the break dates of Australia, Canada, Germany, Netherlands and the US have modes at the dates 1974, 1980 and 1981. The year(s) of the financial crisis 2008/9 are identified as break dates mainly for the European countries, viz. for Denmark, Greece, Ireland, Spain, Sweden and for the UK. Among the Non-European countries only Japan and the US show level shifts for the year 2008. One obvious feature in the results is that for some of the series there is much uncertainty concerning the identification of the break points. Looking for example at Germany (see figure D.7 in appendix D) the sampling results suggest two breaks, one with mode at 1980, which can be associated with the second oil crisis of 1979 and the other at 1990, i.e. the year of the German reunification. But there are also two almost equally likely break dates, namely the years 1992 and 1995, which suggests multimodality of the posterior distribution of the second change point. Furthermore there is a trade off between choosing a third break, which could then result in three unimodal posterior distributions and the use of just two breaks but with some uncertainty in determining the right timing. This feature can also be recognized when

[28] For AIC the matches are much lower (30%).

[29] These are readily available in the `R` package `strucchange`. To provide best possible comparability with the Bayesian results the minimum fraction of observations between two adjacent breaks is set equal to the smallest possible value depending on the sample size, mostly $h = 0.1$, see Bai and Perron (2003), p.7.

looking at the joint posterior mass function of the number of breaks and the number of lags, depicted in figure D.18. When undertaking a 'Helicopter tour' around the posterior surface it can be observed, that a second mode of this model posterior distribution is at $(p,m) = (1,4)$, see figure D.19.[30] Having determined the most likely model specification the Bayesian approach of section 2.5 is used to test for a unit root. Besides the posterior probabilities of a unit root, also the classical p-values of the ADF test are calculated along with their Bayesian analogues, the posterior tail probabilities of the nonstationary region, i.e. of $\theta \geq 1$. For the latter I use a $N(1, \sqrt{T})$ prior and the approximate Jeffreys prior of Phillips (1991*b*). Table A.6 shows the results of the unit root tests. From there it can be observed that the only country, which is likely to exhibit unit root behavior is Greece with posterior probabilities of 84% and 93%, depending on the prior specification. Note that one arrives at the same conclusion when consulting the Bayesian 'p-values' in order to test the hypothesis of nonstationarity, which are both larger than the usual significance levels. When looking at the lower posterior density plot of Greece in figure D.8 the impact of the Jeffreys prior on the posterior shape can be seen, namely that it puts more weight on the explosive region of θ, compared to a normal prior. In contrast to the classical ADF test, the null is rejected for almost all of the analyzed countries using the Bayes test.[31] Comparing the posterior means under a normal prior with the corresponding model averaged point estimates (see table A.5) it can be recognized that for most of the countries both estimates are quite close to each other. This fact is also reflected in the corresponding risk functions presented in section 2.5. Interestingly from the model averaged half lives it can be observed, that when controlling for possible uncertainty induced through the model selection step, Greece does not have the longest half life with about three years, as would be expected given the unit root results. In fact Spain and Japan both have longer half lives with 5.42 and 5.32 years, respectively. The overall OECD country averages are (standard deviations in brackets): $HL_{Norm} = 3.11$ (3.64) years, $HL_{\overline{M}} = 2.04$ (1.47) years and $HL_{ADF} = 4.73$ (2.72) years.

In sum, the empirical results neither suggest unit root hysteresis nor pronounced persistence of the annual unemployment rates for the majority of the OECD countries. The empirical findings are also in accordance with those of other authors, as for example Papell

[30] The figures were constructed as in Klein (2008).

[31] This can be attributed to the lack of power of the ADF test in small samples and that it does not control for structural breaks.

et al. (2000) and Summers (2004), who find no overall evidence of hysteresis or marked persistence in the unemployment rates of these countries. In the light of economic theory this supports the perception that unemployment rates are best described as transitory fluctuations around an equilibrium path. Thus the results support structural theories of unemployment which imply that unemployment rates follow a (trend)stationary process with possible shifting behavior due to changes in structural factors.

2.8. Summary and conclusion

Most of the existing approaches for model selection in the class of ARMA models with multiple structural breaks are based on information criteria. However these can be difficult to implement when the number of autoregressive lags, the number of breaks, and the associated break dates are unknown and have to be estimated simultaneously. Furthermore, most of the classical approaches do not capture the possible uncertainty induced through a model selection step. In contrast, the presented sampling based method provides the researcher not only with point estimates for the unknown model indicators but with the whole joint posterior probability distribution of these quantities. Using this distribution model averaged point estimates of all quantities of interest can be computed. With the proposed Bayesian approach it is possible to select the most likely model specification for unit root testing in the case of multiple breaks. This is accomplished by using a mixed MCMC sampling strategy, which enables to switch between parameter spaces of different dimensions. The presented simulation results indicate that the sampler performs well in finding the true parameter values of the underlying data generating process. In a next step the posterior probability of a unit root is computed using different prior distributions to construct a Bayesian unit root test. The Bayes test is compared with three classical unit root tests in terms of test power and shows clear superiority especially in small samples. In an empirical application, the unemployment rates of 17 OECD countries are analyzed with respect to possible unemployment hysteresis. The results indicate that the only country with high posterior probabilities for unit root hysteresis is Greece, whereas Japan and Spain show slightly increased levels of persistence. However, by applying model averaging techniques, it is found, that although Greece still shows an increased level of persistence after a shock compared to the OECD average, Spain and Japan both exhibit

higher levels of persistence than Greece. Overall the empirical analysis suggests that the majority of the OECD unemployment rates are likely to follow a trend stationary process with possible level shifts, which is also implied by structuralist theories of employment.

Next attention will be drawn to higher frequency data, namely to quarterly and monthly time series data. In contrast to the aggregate case with annual data of the last chapter here different forms of stochastic trends can appear, where each of these forms is associated with a specific frequency of the power spectrum (cf. Bloomfield (2000), chapter 9). Before introducing some Bayesian approaches to test for a periodic and a seasonal unit root in chapter 4, first some general characteristics related to seasonal time series are reviewed together with two selected frequentist unit root testing procedures as a preliminary.

Analysis of seasonal time series

Concepts related to seasonal time series

3.1. Stochastic seasonality and seasonal integration

Consider a univariate time series y_t, $t = 1 \ldots T$, for $s = 1 \ldots S$ seasons, which is observed during $N = \lfloor T/S \rfloor$ years, where $\lfloor x \rfloor$ denotes the greatest integer part of x. Let the observations be generated by an autoregressive process of order p ('AR(p)'):

$$\phi_p(L) \cdot y_t = \mu + \varepsilon_t \tag{3.1}$$

where μ is an intercept term and $\phi_p(L) = 1 - \phi_1 L - \ldots - \phi_p L^p$ is a polynomial in the lag operator of order p, and the ε_t are generated by a white noise process, i.e. are uncorrelated with zero mean and constant variance σ^2 (cf. Spanos (1999), p.443), henceforth $\varepsilon_t \sim WN(0, \sigma^2)$.

If the above polynomial $\phi_p(L)$ can be factorized as $\phi_p(L) = \phi^{\star}_{p-S}(L) \cdot (1 - L^S)$, the series y_t is said to be *seasonally integrated* (see Ghysels and Osborn (2001), p.43). For example, in the case of quarterly data, i.e. $S = 4$, the filter $(1 - L^4)$ can be factorized as $(1 - L) \cdot (1 + L) \cdot (1 + i) \cdot (1 - i)$, with $i \equiv \sqrt{-1}$ the imaginary part of a complex number. That is, the process y_t has four unit roots, which are illustrated in figure 3.1 (cf. Hylleberg et al. (1990)), and are given by a real-valued nonseasonal unit root and three seasonal unit roots, viz. one real-valued and a conjugate pair of complex-valued roots.

To get the idea of a seasonal unit root consider the following seasonal AR process of order

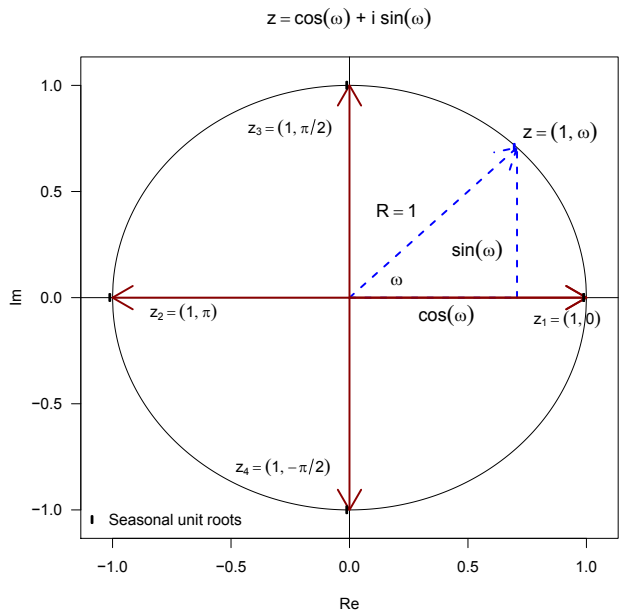


Figure 3.1.: Complex unit circle with seasonal unit roots for $S = 4$.

one for S seasons ('SAR(1)'):

$$y_t = \phi y_{t-S} + \varepsilon_t \ , \quad \varepsilon_t \sim WN(0, \sigma^2) \tag{3.2}$$

with characteristic polynomial $\phi_S(z) \equiv 1 - \phi z^S$, $z \in \mathbb{C}$.

A study of the properties of a stochastic process can be conducted on the frequency domain by analyzing its spectral density (or power spectrum), cf. Bloomfield (2000), Priestley (2004). Therefore the spectral density of the process (3.2) is considered next in more detail. Assuming weak stationarity of the SAR(1) process in (3.2) the corresponding MA(∞) representation is given by

$$y_t = \psi(z) \cdot \varepsilon_t \quad \text{, where } \psi(z) \equiv \left(1 - \phi z^S\right)^{-1} \ , \ z \in \mathbb{C} \tag{3.3}$$

Recall that any complex number can be expressed as $z = R \cdot [\cos(\omega) + i \cdot \sin(\omega)] = R \cdot \exp(i\omega)$, with angle $\omega \in [0, \pi]$ in radian and $R \equiv |z|$ the modulus of z. Furthermore the autocovariance-generating function of this stochastic process is given by $g_y(z) = \sigma_\varepsilon^2 \cdot \psi(z) \cdot \psi(z^{-1})$ and is linked to the spectral density via $s_y(\omega) = g_y(z)/2\pi$ (see Hamilton

(1994), chapter 6 for details), which, in the general case of S seasons (assuming $R = 1$), is given by

$$s_y(\omega) = \frac{\sigma_\varepsilon^2}{2\pi} \cdot \left(1 - \phi z^S\right)^{-1} \cdot \left(1 - \phi z^{-S}\right)^{-1} \tag{3.4a}$$

$$= \frac{\sigma_\varepsilon^2}{2\pi} \cdot \left(1 - \phi \cdot \exp\left(S \cdot i\omega\right)\right)^{-1} \cdot \left(1 - \phi \cdot \exp\left(-S \cdot i\omega\right)\right)^{-1} \tag{3.4b}$$

$$= \frac{\sigma_\varepsilon^2}{2\pi} \cdot \left|1 - \phi \cdot \exp\left(S \cdot i\omega\right)\right|^{-2} = \frac{\sigma_\varepsilon^2}{2\pi} \cdot \left(1 + \phi^2 - 2\phi \cdot \cos\left(S \cdot \omega\right)\right)^{-1} \tag{3.4c}$$

utilizing the fact that $|\psi(z)|^2 = \psi(z) \cdot \psi(\bar{z})$, with $\bar{z} \equiv \exp(-i\omega)$ the complex conjugate number of z (cf. Hamilton (1994), Appendix A.2).

Next the spectral densities of four autoregressive processes of the form (3.2) are compared. In figure 3.2 the power spectrum of a stationary nonseasonal (or annual) AR(1) process, i.e. with $S = 1$ and $|\phi| \ll 1$ (see upper left panel), a nonseasonal integrated AR(1) process ('ARI(1)'), i.e. with $S = 1$ and $\phi = 1$ (see upper right panel), a stationary quarterly SAR(1), i.e. with $S = 4$ and $|\phi| \ll 1$ (see lower left panel), whose spectral density is given in (3.4), and finally the power spectrum of a seasonally integrated AR(1) process ('SI(1)'), with $S = 4$ and $\phi = 1$ (see lower right panel), are depicted.[1]

What can be observed from the upper and lower right spectra in figure 3.2, is that in case of an ARI(1) and SI(1) process, i.e. a (non)seasonal Random Walk, the spectrum gets more concentrated around the frequencies associated with the four unit roots of $1 - z^4 = (1 - z) \cdot (1 + z) \cdot (1 + z^2)$. These are shown in figure 3.1 and are given by $z_1 = 1$ $(= \cos(0) + i \cdot \sin(0))$, $z_2 = -1$ $(= \cos(\pi) + i \cdot \sin(\pi))$ and the complex conjugate pair $z_{3/4} = \pm i$ $(= \cos(\pm\frac{\pi}{2}) + i \cdot \sin(\pm\frac{\pi}{2}))$. Here z_1 is the nonseasonal unit root, which is associated with the zero spectral frequency and is denoted by $z_1 = (R = 1, \omega = 0)$ in figure 3.1. z_2 is the real semiannual unit root, associated with frequency π, implying that in the frequency domain y_t has a component that gives rise to a half-cycle every period, or a full cycle every two periods. Last the complex pair $z_{3/4}$ are the seasonal unit roots, which are associated with the $\omega = \pm\frac{\pi}{2}$ spectral frequencies, that is the associated unit root process contains a full cycle every four periods, see figure 3.1.[2]

[1] The SI(1) process is also called a *seasonal Random Walk*, see Ghysels and Osborn (2001).

[2] However both cycles are indistinguishable in the frequency domain, that is they are *aliases*, because they both correspond to four-period cycles, see Ghysels and Osborn (2001), p.22, Bloomfield (2000), p.21.

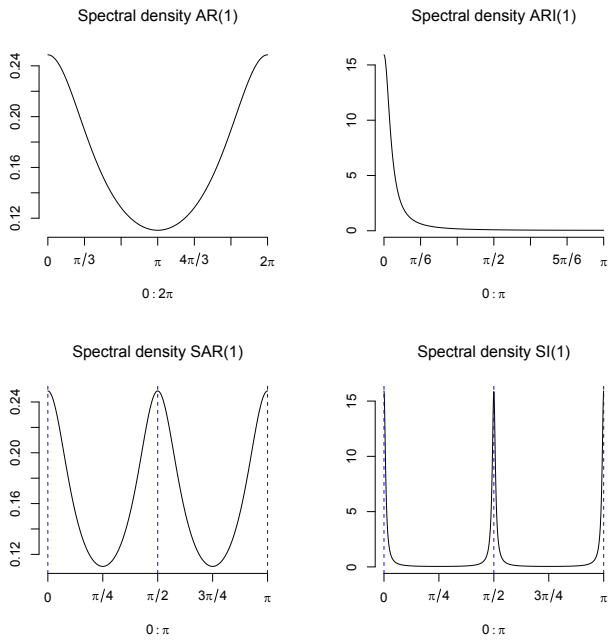


Figure 3.2.: Theoretical spectral densities of quarterly (non)stationary (S)AR(1) processes.

Analogously in the monthly case, the characteristic polynomial $1-z^{12}$ can be factorized according to one real nonseasonal unit root, and 11 seasonal unit roots, viz. one real and five complex conjugate pairs of unit roots (see Beaulieu and Miron (1993) for details). Seasonal unit roots correspond with stochastic trends at the seasonal frequencies and therefore allow for strongly changing seasonality. In this respect, it should be noticed that a seasonal Random Walk implies S separate nonseasonal (annual) Random Walk processes, one relating to each season (see Dickey et al. (1984)). To see this, note that the above SAR(1) process could alternatively be expressed as

$$y_{s,n} = \phi y_{s,n-1} + \varepsilon_{s,n} \quad \text{, where } \varepsilon_t = \varepsilon_{s,n} \tag{3.5}$$

with $s = 1...S$ seasons and $n = 1...N$ years (see Pagano (1978), Tiao and Grupe (1980), Ghysels et al. (2006)).

As noted by Ghysels and Osborn (2001), p.26, this form is useful in emphasizing that the autoregressive relationship for $y_{s,n}$ in season s relates to the same season in the preceding year. Solving the stochastic difference equation in (3.5) by substituting for lagged y on

the right-hand side, assuming $y_{s,0} = 0$, $\forall s$, and setting $\phi = 1$, yields

$$y_{s,n} = \sum_{j=0}^{n-1} \varepsilon_{s,n-j} \tag{3.6}$$

Hence each process $y_{s,n}$,$s = 1...S$, is driven only by shocks relating to the specific season s, see Ghysels and Osborn (2001) for details. The implication of S separate annual Random Walks has been the basis for a large amount of testing procedures for seasonal unit roots starting with the work of Dickey et al. (1984), Osborn et al. (1988), Hylleberg et al. (1990), Franses (1991), inter alia. See Ghysels and Osborn (2001) for a survey of the related literature.

3.2. Periodic processes and periodic integration

One potential drawback of time series models like the SAR model in (3.2) is that seasonal movements are assumed to be constant over the year. Although changing seasonality can simply be modeled in a deterministic fashion, for instance by replacing the intercept μ in (3.2) by seasonally varying intercepts μ_s, $s = 1 \dots S$, the dynamics of y_t are still assumed to be constant over time. A more flexible class of linear models are so called Periodic Autoregressive Moving Average ('PARMA') models which can be considered as a seasonally varying generalization of the class of ARMA models (Box et al. (2008)). PARMA models have the general form (see Ghysels and Osborn (2001), p.140):

$$\phi_s(L) \cdot y_{s,n} = \mu_s + \theta_s(L) \cdot \varepsilon_{s,n} \ , \ s = 1 \dots S \ \text{ and } \ n = 1 \dots N, \tag{3.7}$$

where $\phi_s(L) = 1 - \phi_{s1}(L) - \dots - \phi_{sp}(L^p)$ and $\theta_s(L) = 1 - \theta_{s1}(L) - \dots - \theta_{sq}(L^q)$ are the seasonally varying autoregressive and moving average lag polynomials, respectively, and $\varepsilon_{s,n}$ is an i.i.d. process over both season and year, i.e. $E(\varepsilon_{s,n}\varepsilon_{k,j}) = 0$ unless $s = k$ and $n = j$. Furthermore heteroscedasticity over the seasons is often permitted, so that $E(\varepsilon_{s,n}^2) = \sigma_s^2$. Pioneering works related to PAR(MA) models go back to Gladyshev (1961), Pagano (1978), Troutman (1979), Tiao and Grupe (1980), inter alia, see also Franses and Paap (2006).

A widely used member of this class, which will be used in the next to chapters, is a periodic autoregressive model of order $p = \max_s\{p_s\}$, denoted 'PAR(p)':

$$y_{s,n} = \mu_s + \phi_{s1} \cdot y_{s-1,n} + \ldots + \phi_{sp} \cdot y_{s-p,n} + \varepsilon_{s,n} \quad , \text{ for } s = 1 \ldots S \tag{3.8}$$

which can also be expressed more explicitly as in Boswijk and Franses (1996), Franses and Koop (1997):

$$y_t = \sum_{s=1}^{S} \mu_s \cdot D_{s,t} + \sum_{s=1}^{S} \phi_{s1} \cdot D_{s,t} \cdot y_{t-1} + \ldots + \sum_{s=1}^{S} \phi_{sp} \cdot D_{s,t} \cdot y_{t-p} + \varepsilon_t \tag{3.9}$$

with dummy variable $D_{s,t} = 1$, if $s_t = s$, with $s_t \equiv 1 + (t-1) \text{mod } S$, for $s = 1 \ldots S$, and 0 otherwise. Here s_t denotes the season in which observation t falls, assuming that y_1 is observed in season 1 for simplicity.

For example, with monthly data an observation y_t at date $t = 14$ is observed in season $s_{14} = 1 + (14-1) \text{mod } 12 = 2$, that is in February of year n. The year n in which observation $y_t = y_{s,n}$ falls can be obtained as $n_t = 1 + \text{int}[(t-1)/S]$, where 'int' denotes the integer part, i.e. for example $n_{14} = 1 + \text{int}[13/12] = 2$, see Ghysels and Osborn (2001), p.6.

Dropping the constants μ_s in (3.8) for the moment, note that the PAR(p) model is a special case of

$$y_t = \phi_{1,t} \cdot y_{t-1} + \ldots + \phi_{p,t} \cdot y_{t-p} + \varepsilon_t \ , \ t = 1...T \tag{3.10}$$

which is usually called a random-coefficient autoregression of order p (cf. Franses and Paap (2006), p.7). Assuming $\phi_{s_t} = \phi_t$, with $s_t = 1 + (t-1) \text{mod } S$, then yields a PAR($p$) model. Furthermore the PAR(p) model itself encompasses the nonperiodic AR(p) model

$$y_t = \phi_1 \cdot y_{t-1} + \ldots + \phi_p \cdot y_{t-p} + \varepsilon_t \ , \ t = 1...T \tag{3.11}$$

by assuming $\phi_s = \phi$, $\forall s$.

One important aspect of periodic processes like (3.7) is that they are nonstationary by construction, because their autocorrelation function and hence their spectral density varies with the season (see Troutman (1979), p.222 for more details). This latter observation

suggests an alternative model representation for an analysis of stationarity, unit roots and stochastic trends. These issues are analyzed most conveniently in a multivariate model framework, see Franses (1994), Boswijk and Franses (1996), also Lütkepohl (2007), p.591.

The idea is to stack the seasonal observations $y_{s,n}$ in the annual sequence of $(S \times 1)$ vectors $Y_n = (y_{1,n}, \ldots, y_{S,n})'$, where $y_{s,n} = y_{S\cdot(n-1)+s}$ is the observation in season s of year n, for $n = 1 \ldots N$. This was first proposed by Gladyshev (1961) and adapted by many others in the sequel, see Franses (2003), Franses and Paap (2006) for an overview.

For illustration purposes, I will write out the univariate model in (3.8) for the simplest case with $p = 1$ and $S = 4$, that is a quarterly PAR(1) process, here explicitly:

$$\begin{aligned} y_{1,n} &= \mu_1 + \phi_{11} \cdot y_{4,n-1} + \varepsilon_{1,n} \\ y_{2,n} &= \mu_2 + \phi_{21} \cdot y_{1,n} + \varepsilon_{2,n} \\ y_{3,n} &= \mu_3 + \phi_{31} \cdot y_{2,n} + \varepsilon_{3,n} \\ y_{4,n} &= \mu_4 + \phi_{41} \cdot y_{3,n} + \varepsilon_{4,n} \end{aligned}$$

as $y_{s,n} = y_{S+s,n-1}$, for $s \leq 0$.

Following Boswijk and Franses (1996) the above system of equations can be written as a multivariate model of Y_n:

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ -\phi_{21} & 1 & 0 & 0 \\ 0 & -\phi_{31} & 1 & 0 \\ 0 & 0 & -\phi_{41} & 1 \end{pmatrix} \begin{pmatrix} y_{1,n} \\ y_{2,n} \\ y_{3,n} \\ y_{4,n} \end{pmatrix}$$

$$= \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 & \phi_{11} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} y_{1,n-1} \\ y_{2,n-1} \\ y_{3,n-1} \\ y_{4,n-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,n} \\ \varepsilon_{2,n} \\ \varepsilon_{3,n} \\ \varepsilon_{4,n} \end{pmatrix}$$

or more compactly in matrix notation

$$\Phi_0 \cdot Y_n = \mu + \Phi_1 \cdot Y_{n-1} + E_n \,, \quad n = 1 \ldots N \tag{3.12}$$

where the annual lag operator $L^S Y_n \equiv Y_{n-1}$ is similarly defined as the usual one-period lag operator $L \cdot y_{s,n} \equiv y_{s-1,n}$ and $\mu = (\mu_1, \ldots, \mu_S)'$ is a vector of constants and $E_n = (E_{1,n}, \ldots, E_{S,n})'$ follows an S-dimensional Gaussian vector white noise process with $E_{s,n} = \varepsilon_{S\cdot(n-1)+s}$, respectively. In general, Φ_0 and Φ_k, $k = 1 \ldots P$, are $S \times S$ coefficient matrices with elements

$$\Phi_0[i,j] = \begin{cases} 1 \,, & \text{if } i = j \\ 0 \,, & \text{if } j > i \\ -\phi_{i-j,i} \,, & \text{if } i > j \end{cases} \qquad \Phi_k[i,j] = \phi_{i-j+S\cdot k,i} \tag{3.13}$$

for $i = 1 \ldots S,\ j = 1 \ldots S$. Here the maximum lag order equals $P = 1 + \lfloor (p-1)/S \rfloor$, see Franses and Paap (2006), p.32, Ghysels and Osborn (2001), p.145 f. for details.

Now, in contrast with the univariate form in (3.8) and (3.9), respectively, the multivariate form has constant parameters, which is thus a useful representation to test for the presence of unit roots (see Franses (1994)). Equation (3.12) can finally be expressed as a vector autoregressive process of order P ('VAR(P)') in terms of a matrix polynomial $\Phi(L) = \Phi_0 - \Phi_1 L^S - \ldots - \Phi_P L^{PS}$ (here: with $P = 1$):

$$\Phi(L) \cdot Y_n = \mu + E_n \tag{3.14}$$

The vector system in (3.14) is stable, iff the roots z of the characteristic polynomial

$$\det(\Phi_0 - \Phi_1 \cdot z^S - \ldots - \Phi_P \cdot z^{PS}) = 0, \tag{3.15}$$

are outside the complex unit circle (see Hamilton (1994), p.259). In case of a PAR(1) process the determinant in (3.15) equals

$$\det(\Phi_0 - \Phi_1 \cdot z^S) = 1 - (\phi_{11}\phi_{21}\phi_{31}\phi_{41} \cdot \ldots \cdot \phi_{S1}) \cdot z^S = 0 \tag{3.16}$$

From equation (3.16) it can be seen, that the characteristic polynomial has all solutions

outside the unit circle, iff

$$|\phi_{11} \cdot \ldots \cdot \phi_{S1}| < 1 \tag{3.17}$$

It should be noticed that the condition $|\phi_{s1}| < 1$, $\forall s$, is sufficient but not necessary for the stationarity condition in (3.17) to hold. For this reason, a stationary PAR(1) process can exhibit one or more individual $|\phi_{s1}| \geq 1$. If $\phi_{11} \cdot \ldots \cdot \phi_{S1} = 1$, the system of Y_n has one unit root and the univariate process $y_{s,n}$ is said to be *periodically integrated* (see Franses (1994), p.135). For higher order PAR processes the nonlinear parameter restrictions for stationarity become more complicated. For example, in case of a quarterly PAR(2) process (Boswijk et al. (1995), Franses (1994), p.98) this can be easily verified to be

$$\begin{aligned} \phi_{22}\phi_{31}\phi_{41} + \phi_{22}\phi_{42} + \phi_{12}\phi_{21}\phi_{31} + \phi_{12}\phi_{32} + \phi_{11}\phi_{21}\phi_{31}\phi_{41} + \\ \phi_{11}\phi_{21}\phi_{42} + \phi_{11}\phi_{41}\phi_{32} - \phi_{12}\phi_{22}\phi_{32}\phi_{42} = 1 \end{aligned} \tag{3.18}$$

Since the Bayesian tests presented in the next chapter are constructed for PAR processes of order one, the subsequent discussion mainly focuses on these processes. The concept of first-order periodic integration (PI) can now be formalized according to Ghysels and Osborn (2001), p.155, as follows:

Definition 1. *A nonstationary periodic process $y_{s,,n}$ is said to be periodically integrated of order 1 if there exist quasi-differences $D_s y_{s,n} \equiv y_{s,n} - \phi_{s1} y_{s-1,n}$, with $\prod_{s=1}^{S} \phi_{s1} = 1$ and not all $\phi_{s1} = 1$, such that the VAR representation for the quasi-differences is stationary and invertible.*

The above definition was introduced by Osborn et al. (1988) and is an example of the time-varying parameter definition as given in Granger (1986). It can also be considered as a generalization of the (non)seasonal unit root concept within the class of SARMA models. Following Ghysels and Osborn (2001), p.152, in a PAR context, first-order unit root nonstationarity arises when the characteristic polynomial in (3.15) contains either the seasonal factor $1 - L^S$ or the nonseasonal factor $1 - L$, with all other roots having modulus greater than one. Ghysels and Osborn (2001) distinguish three different types of integration that can induce first-order unit root nonstationarity:

- $y_t \sim I(1)$: When each periodic autoregressive polynomial $\phi_s(L)$ contains the common factor $\Delta_1 = (1 - L)$, but the VAR representation for $\Delta_1 y_{s,n}$ is a stationary

process. This case corresponds to a zero frequency unit root.

- $y_t \sim SI(1)$: When each periodic autoregressive polynomial $\phi_s(L)$ contains the common factor $\Delta_S = (1-L^S)$, but the VAR representation for $\Delta_S y_{s,n}$ is a stationary process. This type of integration is discussed in section 3.1.

- $y_t \sim PI(1)$: When the characteristic polynomial $\det(\Phi(L))$ contains the factor $\Delta_S = (1-L^S)$, see also equation (3.14), but $(1-L^S)$ is not common to each polynomial $\phi_s(L)$, $s = 1...S$, with the VAR for $\Delta_S y_{s,n}$ being stationary.

The first two types of integration are the *nonseasonal* and the *seasonal unit root(s)* considered in the ARMA and SARMA framework, respectively (see Box et al. (2008)). The third case is a specific type, that can arise only in a PARMA context. The definition of periodic integration indicates that periodic integration nests the nonseasonal $(1-L)$, and also the seasonal $(1+L)$ filter, which correspond to the zero and the π-frequency, respectively (see section 3.1 above). For a PAR(1) model this suggests that one should first check the nonlinear restriction $\phi_1 \cdot \ldots \cdot \phi_S$ in (3.17) and then test for $\phi_s = 1, \forall s$, and $\phi_s = -1, \forall s$, see Boswijk and Franses (1996). That is, once it has been established that the periodic unit root null hypothesis

$$H_0 : \ \phi_1 \cdot \ldots \cdot \phi_S = 1 \quad \text{vs.} \quad H_1 : \ |\phi_1 \cdot \ldots \cdot \phi_S| < 1 \tag{3.19}$$

cannot be rejected, the next step is to test whether the hypotheses

$$H_0 : \phi_s = 1 \ \text{, for } s = 1...S-1 \quad \text{vs.} \quad H_1 : \ \phi_s \neq 1 \ , \ \exists s \tag{3.20a}$$

$$H_0 : \phi_s = -1 \ \text{, for } s = 1...S-1 \quad \text{vs.} \quad H_1 : \ \phi_s \neq -1 \ , \ \exists s \tag{3.20b}$$

are valid, which also implies that either $\phi_S = 1$ or $\phi_S = -1$, given $\prod_{s=1}^{S} \phi_s = 1$.

Both hypotheses (3.20) postulate that there is no periodic variation in the autoregressive coefficients, but that the data is generated by a (non)seasonal Random Walk process. Boswijk and Franses (1996) prove that the corresponding null distributions for a test of (3.20a) and (3.20b) asymptotically follow a $\chi^2_{(3)}$- distribution (see Boswijk and Franses (1996), Theorem 2, p.232). When the null in (3.20a) is not rejected the PAR process is said to contain a nonseasonal (i.e. annual) unit root. In the terminology of Franses (2003), p.133, the null in (3.20a) results in a PAR process for an $I(1)$ series, which he

abbreviates as 'PARI'. However when the null in (3.20b) is not rejected, it is said that the PAR process contains a real-valued seasonal unit root, corresponding to a half-year cycle (see Ghysels and Osborn (2001)). In case that the hypotheses in (3.20a) and (3.20b) are both not rejected the PAR model is called a periodically integrated AR model ('PIAR'), see Franses (2003).

3.3. Two classical testing approaches for a periodic unit root

In this section two classical testing approaches for a single periodic unit root are outlined. The first one is based on the vector representation of a PAR process and proceeds within a cointegration framework (Engle and Granger (1987), Johansen (1988)), whereas the second is based on the univariate model representation and will be used as a classical competitor in the next section. More details on periodic unit root and cointegration testing can be found in Franses (2003), Franses and Paap (2006).

The possibility of analyzing periodic unit roots in a cointegration framework stems from the fact that periodic integration must imply cointegration of the nonstationary series $y_{s,n}$ with the seasonally varying differencing filters as the stationary linear combinations (see also definition 1). Moreover the number of unit roots is linked to the number of cointegration relations between the elements of Y_n, see Franses and Paap (2006), p.85.

Premultiplying equation (3.12) with Φ_0^{-1} yields:

$$Y_n = \Phi_0^{-1}\mu + \Phi_0^{-1}\Phi_1 \cdot Y_{n-1} + \Phi_0^{-1}E_n \tag{3.21}$$

with[3]

$$\Phi_0^{-1}\Phi_1 = \begin{pmatrix} 0 & 0 & 0 & \phi_{11} \\ 0 & 0 & 0 & \phi_{11}\phi_{21} \\ 0 & 0 & 0 & \phi_{11}\phi_{21}\phi_{31} \\ 0 & 0 & 0 & \phi_{11}\phi_{21}\phi_{31}\phi_{41} \end{pmatrix} \tag{3.22}$$

By subtracting Y_{n-1} on both sides the model in (3.21) can be written in vector error correction form (cf. Lütkepohl (2007), p.248) as

$$\Delta_1 Y_n = \Phi_0^{-1}\mu + (\Phi_0^{-1}\Phi_1 - \mathbf{I}_4)\cdot Y_{n-1} + \Phi_0^{-1}E_n \tag{3.23a}$$

$$= \mu^\star + \Pi \cdot Y_{n-1} + E_n^\star \tag{3.23b}$$

where Δ_1 is the first-differencing filter for the annual vector series, that is $1 - L^S$ operates on the annual data Y_n (see Franses and Paap (2006), p.65). Hence $\Delta_1 Y_n$ corresponds to $\Delta_4 y_t$ for $S = 4$, or in general to $\Delta_S y_t$. Of primary interest is the matrix Π, which is relevant for the analysis of longrun equilibria among the elements of Y_n:

$$\Pi \equiv \Phi_0^{-1}\Phi_1 - \mathbf{I}_4 = \begin{pmatrix} -1 & 0 & 0 & \phi_{11} \\ 0 & -1 & 0 & \phi_{11}\phi_{21} \\ 0 & 0 & -1 & \phi_{11}\phi_{21}\phi_{31} \\ 0 & 0 & 0 & \phi_{11}\phi_{21}\phi_{31}\phi_{41} - 1 \end{pmatrix} \tag{3.24}$$

The matrix Π contains information on the cointegration relations between the S elements in Y_n. It can be written as $\Pi = \gamma \cdot \phi'$ with loading matrix γ and ϕ the matrix of cointegration vectors both of dimension $4 \times r$, see Franses and Paap (2006), p.64. When there are r stationary linear combinations between the $y_{s,n}$, the matrix Π has rank r with $0 < r < 4$.

[3] This can easily be verified by direct calculation of Φ_0^{-1}, for example via Gauss-Jordan elimination.

In the present case, the cointegration space can be spanned, for example, by

$$\phi' = \begin{pmatrix} -\phi_{21} & 1 & 0 & 0 \\ 0 & -\phi_{31} & 1 & 0 \\ 0 & 0 & -\phi_{41} & 1 \end{pmatrix} \tag{3.25}$$

When the rank r of the matrix Π in (3.23b) is $r = 0$, and hence there are no cointegration relations between the elements of Y_n, this implies that the $1 - L^S$ filter for $y_{s,n}$ is appropriate. The Y_n process is stationary when r equals 4. When $\phi_{11}\phi_{21}\phi_{31}\phi_{41} = 1$ the rank of Π equals $r = 3$, and this implies three cointegrating relationships between the $y_{s,n}$ subseries. Thus the three stationary linear combinations are: $(y_{2,n} - \phi_{21}y_{1,n})$, $(y_{3,n} - \phi_{31}y_{2,n})$ and $(y_{4,n} - \phi_{41}y_{3,n})$, which correspond to three quasi-differences (see definition 1). Note that the fourth stationary variable $(y_{1,n} - \phi_{11}y_{4,n-1})$ is implied by the other three, and thus is not linearly independent.[4] In other words, in case of a single real unit root,[5] there exist $S - 1$ linear combinations $(1 - \phi_{s1}L) \cdot y_{s,n}$, that transform the series y_t to a stationary process, see Franses (2003), p.128, for details. Since a test for the rank of the matrix Π can also be considered as a test for a periodic unit root, Franses (1994) proposes to use the likelihood-based cointegration test developed in Johansen (1988). However as can be seen from equation (3.23a) the cointegration approach by Franses (1994) is highly parameterized[6] compared to the univariate approach considered next.

Boswijk and Franses (1996) propose an alternative testing strategy using the univariate model representation in (3.9) as a starting point (see also Boswijk and Franses (1995)). For the ease of reference and also because this classical test is compared with one of the Bayes tests presented in the next chapter, I will briefly outline the basic testing strategy of the Boswijk and Franses (1996) test.[7]

In the simplest case of a quarterly PAR(1) process without deterministic terms the authors

[4]This can be checked by inserting the first and third of the above linear combinations into the second, imposing the unit root restriction via $\phi_1 = 1/(\phi_2 \cdot \phi_3 \cdot \phi_4)$, and solving for $y_{1,n}$.
[5]For the case of multiple unit roots, see Franses and Paap (2006), p.67.
[6]Here an unrestricted $S \times S$ matrix of coefficients has to be estimated.
[7]This test has been implemented in `R` by the author.

consider estimating the following unrestricted regression under the alternative H_1:

$$y_t = \sum_{s=1}^{4} \phi_s D_{s,t} y_{t-1} + \varepsilon_t \tag{3.26}$$

Assuming independently normally distributed errors ε_t, the maximum likelihood (ML) estimators of the ϕ_s are given by the least squares estimators. Under H_0 the following restricted regression is estimated:

$$y_t = \phi_1 D_{1,t} y_{t-1} + \phi_2 D_{2,t} y_{t-1} + \phi_3 D_{3,t} y_{t-1} + (\phi_1 \phi_2 \phi_3)^{-1} D_{4,t} y_{t-1} + \varepsilon_t \tag{3.27}$$

The authors consider maximizing the restricted log-likelihood via non-linear least squares (see ibid., p.229). A likelihood ratio test statistic can be constructed as

$$LR \equiv T \cdot \ln(RSS_0 / RSS_1)$$

where RSS_0 and RSS_1 denote the residual sums of squares from the regressions under H_0 and H_1, respectively.

A one-sided test may then be constructed from the studentized statistic:

$$LR_\tau = \text{sign}(\widehat{\rho} - 1) \cdot \sqrt{LR} \tag{3.28}$$

Boswijk and Franses (1996), Theorem 1, p.230, show that the statistic in (3.28) has the following asymptotic distribution under the null hypothesis:

$$LR_\tau \xrightarrow{d} \left\{ \int_0^1 W(r)^2 dr \right\}^{-\frac{1}{2}} \int_0^1 W(r) \cdot dW(r) \tag{3.29}$$

where $W(r)$ is a standard Wiener process and $\xrightarrow{d}$ denotes convergence in distribution.

This statistic has the same asymptotic null distribution as Fuller's τ statistic for a nonperiodic AR model without drift and trend and thus the tabulated critical values in Fuller (1996) can be used.

In the next chapter an alternative Bayesian testing approach for a periodic unit root in the presence of a structural break at unknown time is presented.

Bayesian analysis of periodic unit roots with a break

4.1. Introduction

Periodic autoregressive models have been applied to economic time series as an alternative to constant parameter models like seasonal autoregressive moving average models or seasonal means models, see Osborn et al. (1988), Osborn and Smith (1989), Franses (1995), Franses and Koop (1997) among others. PAR processes can for example arise when modeling seasonal decisions of consumers (Osborn et al. (1988)), whereas Hansen and Sargent (1993) suggest that they arise from seasonal technology. In the literature such changing seasonal variation in the data has often been modeled in a deterministic fashion, e.g. through the inclusion of seasonal dummy variables or so called seasonally integrated autoregressive models. However, empirical studies have found evidence for periodic variation in many macroeconomic series. For example, Franses (1995) finds in a business cycle analysis of quarterly US unemployment rates that seasonal fluctuations are not constant across two business cycle stages and further that shocks to the unemployment rate are transitory in expansion and persistent in recession periods. Ooms and Franses (1997) show that quarterly unadjusted OECD unemployment rates of Germany and the US can best be described by a PAR process with a unit root. Within a periodic error-correction framework Birchenhall et al. (1989) provide empirical findings that the long-run income elasticity of consumption and its rate of adjustment to equilibrium both vary seasonally. They take this as evidence in favor of the hypothesis that consumers have seasonal preferences and also seasonally varying degrees of habit persistence. In such a case, in order to remove any stochastic trend, the applied differencing filter has to vary with the season, too. Consequently, this leads to the concept of periodic (co)integration (see Osborn et al.

(1988)) which can be perceived as a time-varying generalization of the concept of seasonal (co)integration, see Hylleberg et al. (1990), Beaulieu and Miron (1993), Franses (1994) and Ghysels and Osborn (2001), Franses (2003) for overviews. Different approaches have been proposed to test for periodic integration in economic time series, see Franses (1994), Boswijk et al. (1995), Boswijk and Franses (1996), Franses and Koop (1997), inter alia. However nearly all of these proceed within a classical (or frequentist) framework. By contrast there is a vast literature on Bayesian hypothesis testing for a (non-periodic) zero frequency unit root, see Sims (1988), Phillips (1991*b*), Sims and Uhlig (1991), Schotman and van Dijk (1991), inter alia, and the related subsequent discussion on these works, also Maddala and Kim (1998) for an overview. For a Bayesian adaption of the Hylleberg et al. (1990) (HEGY) model framework to test for seasonal unit roots, see Franses et al. (1997). One exception is the paper of Franses and Koop (1997), who present a Bayesian framework to test for different kinds of unit roots, namely (non)seasonal unit roots at the zero and π-spectral frequency (cf. Ghysels and Osborn (2001)) but most importantly for periodic unit roots. For the latter they propose a sampling based as well as an analytical, i.e. non-sampling based, testing strategy, whereas for (non)seasonal unit roots they suggest an approximate Bayesian test based on highest posterior density (HPD) regions. Their tests assume one structural break at unknown time and are thus conditional on the occurrence of a break. The authors find widespread evidence for (non)periodic integration using data of nine major UK macroeconomic time series. Franses and Koop (1997) then apply their methods to a large set of quarterly UK macroeconomic time series and find great evidence in favor of periodic integration of order one.

The subsequent analysis aims to extend the work of Franses and Koop (1997) in several directions. First, the approach presented here allows to capture the uncertainty induced by conditioning on a structural break model and therefore assigning zero prior probability to models without a break. This is achieved by using a mixture of discrete and continuous posterior distributions for unit root testing, where the discrete parts are given by the posterior probabilities of a model with and without a structural break. Thus the presented approach essentially consists of averaging over the discrete space of candidate models by using Bayesian model averaging (BMA) techniques (cf. Raftery et al. (1997)). This BMA strategy is pursued in order test for periodic unit roots and also to test for zero and π-frequency unit roots. For the latter two cases a Bayesian F-test is proposed. All presented tests allow a shift to occur in the seasonal means and/or time trends and are applied

to both quarterly and monthly data. To illustrate the performance of the proposed tests, the results of Monte Carlo experiments are presented. In contrast to the existing literature on periodic unit root testing, the Bayesian tests are compared to classical competitors in terms of test power. The results suggest that the Bayes test for a periodic unit root has favorable power also in small samples, whereas the Bayesian F-test for a (non)seasonal unit root has less power in small samples. Overall the Bayes tests outperform their classical competitors, especially in short time series. In an empirical application the proposed unit root tests are finally used to test for periodic integration of order one in monthly unadjusted OECD unemployment rates.

In the economic literature most studies do not consider seasonal forms of non-stationarity in the data but instead work with seasonally adjusted data. The effect of working with seasonally adjusted data on the results of unit root tests has been an object of investigation in the literature. Ghysels and Perron (1993) show that when the data are generated by a stationary ARMA process there is a positive asymptotic upward bias induced by an application of the X-11 filter.[1] Thus one can expect unit root tests performed with filtered data to be less powerful against stationary alternatives, because asymptotically the sample estimate of the AR coefficient for such data is greater than for unfiltered data (see Ghysels (1990)). The empirical results presented in this chapter suggest, that the majority of the analyzed monthly OECD unemployment rates are in fact generated by a nonperiodic unit root process. The presence of periodic variation in the data is further analyzed by a Bayesian test for parameter-constancy. The latter shows that there is not much evidence for stochastic periodicity in the unemployment series. In the light of economic theory, favoring a nonperiodic unit root process in order to describe most of the considered OECD unemployment series can be interpreted as supporting the hypothesis of unemployment hysteresis, namely that labor market shocks have a long-run impact on the level of unemployment, see for example Blanchard and Summers (1986).

The rest of this chapter is organized as follows: in section 4.2 the model is introduced and in section 4.3 the proposed Bayesian testing procedures are presented. Then in section 4.4 the results of Monte Carlo experiments to study the power properties of the tests are described. In section 4.5 the results of an empirical application to OECD unemployment data are discussed and section 4.6 summarizes the main findings and concludes.

[1] For a review of seasonal adjustment methods see Ghysels and Osborn (2001).

4.2. Model and definitions

In the following the focus is on PAR models with at most one structural break in the deterministic component(s) at an unknown point in time. Similar models have been used for example in Franses and Koop (1997) and Franses (2003). In case of a structural break in the periodic intercepts and slopes the model has the following general form

$$y_t = \Big(\sum_{s=1}^{S}\sum_{i=1}^{p_s} \phi_{i,s}\cdot D_{s,t}\cdot y_{t-i}\Big) + \lambda_t + \varepsilon_t \quad , \quad \varepsilon_t \overset{i.i.d.}{\sim} N(0,\sigma_s^2) \tag{4.1a}$$

$$\lambda_t = \sum_{s=1}^{S}\Big((\mu_s + \alpha_s\cdot\tau_t)\cdot D_{s,t} + (\mu_s^\star + \alpha_s^\star\cdot\tau_t)\cdot D_{s,t}^\star\Big) \tag{4.1b}$$

with $t = 1,...,T$ observations and $s = 1,...,S$ seasons.

Let $s_t = 1 + [(t-1) \text{mod } S]$ denote the season of observation t (see section 3.2), and let $T_B \in \,]p_s + \kappa,\ T - \kappa]$ be the unknown break point, where κ denotes the first and last ten percent of observations in the sample, which are truncated in order to avoid possible endpoint problems. Then the seasonal dummy variables in (4.1) are given by $D_{s,t} = 1$, if $s_t = s \ \wedge\ t \leq T_B$, otherwise 0, and $D_{s,t}^\star = 1$, if $s_t = s \ \wedge\ t > T_B$, and 0 otherwise. Last, τ_t in (4.1b) denotes the value of the linear time trend at time t. In the case of no structural break $\alpha_s^\star$ and $\mu_s^\star$ equal zero for all seasons. In the following, $p_s = 1$ and $\sigma^2 = \sigma_s^2, \forall s$, is assumed, i.e. a homogeneous autoregressive lag order of one and no seasonal heteroscedasticity.

The periodic autoregressive structural break model in (4.1) can be expressed more conveniently in matrix form. For this purpose define the following expressions:

$$\begin{aligned}
\underset{p\times 1}{\mathbf{y}_0} &\equiv [\, y_1,\, y_2,\, \dots\, ,\, y_p\,]' \\
\underset{T-p\times 1}{\mathbf{y}} &\equiv [\, y_{p+1},\, y_{p+2},\, \dots\, ,\, y_T\,]' \\
\mathbf{y}_{-1}^{(s)} &\equiv [\, D_{s,p+1}\cdot y_p,\, D_{s,p+2}\cdot y_{p+1},\, \dots\, , D_{s,T}\cdot y_{T-1}\,]' \\
&\vdots \\
\mathbf{y}_{-p}^{(s)} &\equiv [\, D_{s,p+1}\cdot y_1,\, D_{s,p+2}\cdot y_2,\, \dots\, , D_{s,T}\cdot y_{T-p}\,]'
\end{aligned}$$

and

$$
\begin{aligned}
\underset{T-p \times p\cdot S}{\mathbf{X}} &\equiv [\, \mathbf{y}_{-1}^{(1)}, \ldots, \mathbf{y}_{-1}^{(S)}\,,\, \mathbf{y}_{-2}^{(1)}, \ldots, \mathbf{y}_{-2}^{(S)}\,,\, \ldots\,,\, \mathbf{y}_{-p}^{(1)}, \ldots, \mathbf{y}_{-p}^{(S)}\,]' \\
\underset{T-p \times 1}{\varepsilon} &\equiv [\, \varepsilon_{p+1}, \varepsilon_{p+2}, \ldots, \varepsilon_T\,]' \\
\underset{T-p \times 4\cdot S}{\mathbf{Z}} &\equiv [\, (D_1, \ldots, D_S)_{(T-p\times S)}\,,\, (D_1^\star, \ldots, D_S^\star)_{(T-p\times S)}\,, \\
&\quad (D_{1,p+1}\cdot\tau_{p+1}, \ldots, D_{1,T}\cdot\tau_T)'\,, \ldots, (D_{S,p+1}\cdot\tau_{p+1}, \ldots, D_{S,T}\cdot\tau_T)'\,, \\
&\quad (D_{1,p+1}^\star\cdot\tau_{p+1}, \ldots, D_{1,T}^\star\cdot\tau_T)'\,, \ldots, (D_{S,p+1}^\star\cdot\tau_{p+1}, \ldots, D_{S,T}^\star\cdot\tau_T)'\,] \\
\underset{p\cdot S\times 1}{\phi} &\equiv [\, \underbrace{(\phi_{1,1}, \phi_{1,2}, \ldots, \phi_{1,S})}_{\equiv\phi_1'}, \ldots, \underbrace{(\phi_{p,1}, \phi_{p,2}, \ldots, \phi_{p,S})}_{\equiv\phi_p'}\,]' \\
\underset{4\cdot S\times 1}{\delta} &\equiv [\, \overbrace{\mu_1, \ldots, \mu_S}^{\equiv\mu}\,,\, \overbrace{\mu_1^\star, \ldots, \mu_S^\star}^{\equiv\mu^\star}\,,\, \overbrace{\alpha_1, \ldots, \alpha_S}^{\equiv\alpha}\,,\, \overbrace{\alpha_1^\star, \ldots, \alpha_S^\star}^{\equiv\alpha^\star}\,]'
\end{aligned}
$$

Therefore the model in (4.1) can finally be written in more compact form:[2]

$$
\mathbf{y} \;=\; \mathbf{X}\cdot\phi + \mathbf{Z}\cdot\delta + \varepsilon \;=\; \underset{T-p\;\times\;(4+p)\cdot S}{\left(\mathbf{X} \vdots \mathbf{Z}\right)} \;\cdot\; \underset{(4+p)\cdot S\;\times\;1}{\begin{pmatrix}\phi\\ \delta\end{pmatrix}} \;+\; \varepsilon \;=\; \widetilde{\mathbf{X}}\cdot\mathbf{B} + \varepsilon \tag{4.2}
$$

where in the following $d \equiv (4+p)\cdot S$.

4.3. Bayesian testing for a periodic unit root

In order to draw inference from a Bayesian PAR model (BPAR) some prior information with regard to the unknown parameters has to be specified. In contrast to Franses and Koop (1997) I do not assume an informative prior for the regression parameters and the innovation variance, but assume the elements of $\mathbf{B}$ and $\log\sigma^2$ as independently and uniformly distributed, to express prior ignorance with respect to these quantities (cf. Zellner

[2] In anticipation of the analysis in the next chapter, the model is stated here for a general lag order p.

(1971), p.66).[3] To express also a lack of prior knowledge with respect to the unknown break date a discrete uniform prior over the set of possible break dates is chosen. These assumptions lead to the following prior distributions:

$$f(\mathbf{B},\sigma^2) \propto \sigma^{-2}\ , \text{ with } \sigma^2>0\ ,\ \mathbf{B}\in\mathbb{R} \tag{4.3a}$$

$$f(T_B|\ m=1,p=1) = \frac{1}{T-1-2\kappa}\cdot\mathbf{1}_{(\kappa+1<T_B\leq T-\kappa)} \tag{4.3b}$$

where $\mathbf{1}_{(.)}$ denotes an indicator function.

By an application of the Bayes Theorem the joint posterior density function of all unknown quantities, under a structural change model, is given by[4]

$$f(\mathbf{B},\sigma^2,T_B|\ m=1,\ \mathbf{y}) = \frac{f(\mathbf{y}|\ \mathbf{B},\sigma^2,T_B,\ m=1)\cdot f(\mathbf{B},\ \sigma^2|\ T_B,\ m=1)\cdot f(T_B|\ m=1)}{f(\mathbf{y}|\ m=1)} \tag{4.4}$$

where m denotes the unknown number of structural breaks, restricted to $m\in\{0,1\}$.

In the subsequent analysis interest focuses on testing the following hypotheses:

$$H_A:\ \phi_s=1 \quad \text{against} \quad |\phi_s|<1\ ,\ \text{for } s=1...S \tag{4.5a}$$

$$H_B:\ \phi_1\cdot\ldots\cdot\phi_S=1 \quad \text{against} \quad |\phi_1\cdot\ldots\cdot\phi_S|<1 \tag{4.5b}$$

$$H_C:\ \phi_1=\ldots=\phi_S=1 \quad \text{against} \quad \phi_s\neq 1\ ,\ \exists s \tag{4.5c}$$

$$H_D:\ \phi_1=\ldots=\phi_S=-1 \quad \text{against} \quad \phi_s\neq -1\ ,\ \exists s \tag{4.5d}$$

where in the following $\rho\equiv\prod_{s=1}^{S}\phi_s$.

Testing the null hypothesis H_A, i.e. a Random Walk in season s, requires the computation of the marginal posterior density of $\phi_s,\forall s$, whereas testing the hypothesis H_B, i.e. the presence of a periodic unit root, requires the derivation of the marginal posterior of the nonlinear parameter restriction in (4.5b). If H_B can not be rejected, the process is called periodically integrated of order one, abbreviated by $y_t\sim PI(1)$ (see Ghysels and Osborn (2001) for details). Last, the two hypotheses in $H_{(C)D}$ are useful for the analysis of a

[3] It turned out that the empirical results presented in section 4.5 are largely unaffected by this choice.
[4] In the following conditioning on $\tilde{\mathbf{X}}$ and $p=1$ is omitted.

(non)seasonal unit root, which implies that the series is (seasonally) integrated of order one, i.e. $y_t \sim I(1)$ and $y_t \sim SI(1)$. That is under H_C and H_D it is postulated that the series is generated by a zero and π-frequency unit root process, respectively. Note that both H_C and H_D are nested within H_B. If the restriction in H_B could not be rejected and a subsequent test shows that either of the nonperiodic unit root null hypotheses H_C and H_D can not be rejected, too, such a process is called a PAR process for an $I(1)$ series ('PARI'). By contrast, if H_C and H_D could be rejected, but not H_B, then the process is called a periodically integrated AR process ('PIAR'), see section 3.2. Whereas the required densities for H_A and $H_{C(D)}$ can be stated analytically (see appendix G for details), the analysis of H_B is more complex due to the nonlinearity of the null constraint. In order to circumvent any computationally intensive techniques, like Markov chain Monte Carlo techniques, Franses and Koop (1997) propose to linearize the restriction $\rho = \phi_1 \cdot \ldots \cdot \phi_S = 1$ in H_B by taking logs and noting that $\log(1+x) \approx x$ for small x.[5] As pointed out by the authors and also according to own experience, the ϕ_s-coefficients are close to one in practice, so this should yield a good approximation. The periodic unit root constraint in H_B can therefore be approximated by the following linear restriction

$$H_0 : \phi_1 + \ldots + \phi_S = S \quad \text{against} \quad |\phi_1 + \ldots + \phi_S| < S \tag{4.6}$$

where $\theta \equiv \sum_{s=1}^{S} \phi_s$ is written henceforth.

A second approach to test the hypothesis H_B is to generate random draws from the posterior distribution of the vector of PAR coefficients ϕ, which follows a multivariate Student-t distribution with $\nu = T - 1 - d$ degrees of freedom (see appendix G.2), say $M = 10000$ times, and then compute the product ρ for each draw. After utilizing kernel density estimates in order to approximate the posterior ordinates, mean and variance can be computed as well as the 95%-HPD region by using one-dimensional numerical integration.[6] Then a test of H_B can be conducted by rejecting the null hypothesis, if the null restriction $\rho = 1$ lies outside the HPD region. In the empirical analysis of section 4.5, this testing strategy is used for comparison with the approximate approach of testing the linear restriction in (4.6).

In accordance with Franses and Koop (1997), I allow for a single break in the deterministic

[5] This follows from a first order Taylor series expansion of $\log(1+x)$ around zero.
[6] Here for example Simpson's rule could be used, cf. Bauwens et al. (1999), p.69.

components, but in contrast to the latter authors, I additionally allow a break to occur in the seasonal time trend so that abrupt changes in the trending behavior of the series can also be captured. Thus, given the lag order and the specification of deterministics, the set of candidate models, $\mathscr{M}$, contains only two elements M_i, namely a structural change (SC) model ($m = 1$) and a no-structural change (NSC) model ($m = 0$).[7] The key ingredients for BMA techniques are the model posterior probabilities (see Raftery et al. (1997)):

$$f(M_j|\mathbf{y}) = \frac{f(\mathbf{y}|M_j) \cdot f(M_j)}{\sum_{i=1}^{K} f(\mathbf{y}|M_i) \cdot f(M_i)} \tag{4.7}$$

with $f(M_i)$ the prior probability of model M_i, which are chosen as $f(M_i) = 1/K$, for $\forall i$, and $K = 2$. The prior predictive density under the SC model is given by

$$\begin{aligned} f(\mathbf{y}|\ m = 1) \ = \ \sum_{T_B=\kappa+2}^{T-\kappa} \int_{\mathbb{R}^+} \int_{\mathbb{R}^d} f(\mathbf{y}|\mathbf{B}, \sigma^2, T_B,\ m = 1) \cdot \\ f(\mathbf{B},\ \sigma^2|T_B,\ m = 1) \cdot f(T_B|\ m = 1) \cdot d\mathbf{B} \cdot d\sigma^2 \end{aligned} \tag{4.8}$$

where $f(T_B|\ m = 1)$ is given in (4.3b) and $\mathbf{B}$ denotes the vector of regression parameters.

Note that the inner two integrations can be solved analytically. As pointed out by Raftery et al. (1997) all probabilities are implicitly conditional on $\mathscr{M}$, i.e. the set of considered models. In order to construct tests for the above hypotheses, the model specific posterior density functions are required. For the analysis of H_A we need the marginal[8] posterior distribution of $\phi_s,\ s = 1...S$, which can be computed as

$$\begin{aligned} f(\phi_s|\ \mathbf{y}) \ = \ \left[\sum_{T_B=\kappa+2}^{T-\kappa} f(\phi_s|\ T_B,\ m = 1,\ \mathbf{y}) \cdot f(T_B|\ m = 1, \mathbf{y}) \right] \cdot f(m = 1|\ \mathbf{y}) \\ + f(\phi_s|\ m = 0,\ \mathbf{y}) \cdot f(m = 0|\ \mathbf{y}) \end{aligned} \tag{4.9}$$

[7] Of course in the present context another model indicator could be a binary variable, indicating the inclusion of a time trend (or any other exogenous regressor). Such variable selection issues are however not further pursued here, although they would be relatively straightforward to implement, see for example George et al. (1993).

[8] In this context marginalization refers to the predefined model space.

Note that the posterior weight for the NSC model used in (4.9) can be calculated as

$$f(m=0 \mid \mathbf{y}) \;\propto\; \left|\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right|^{-\frac{1}{2}} \cdot \left[\nu \cdot s^2\right]^{-\frac{\nu}{2}} \tag{4.10}$$

with $\nu = T-1-d$ and s^2 the usual sampling estimate of the innovation variance.[9]

Moreover $f(m=1|\ \mathbf{y})$ is proportional to (4.8). The posterior distribution $f(\phi_s|T_B,\ m=1,\ \mathbf{y}),\ s=1...S$, has the form of a univariate Student-t density with $\nu = T-1-d$ degrees of freedom (see appendix G.3 for details). Thus the distribution in (4.9) is a model weighted mixture of t-densities. Here d_1 and d_2, with $d = d_1 + d_2$, denote the dimensions of the following subvectors:

$$\mathbf{B} \;\equiv\; \left[\mathbf{B_1}' \vdots \mathbf{B_2}'\right]' \;=\; \left[\underset{1\times 1}{\phi_s} \vdots (\phi'_{-s}\ ,\ \delta')\right]' \tag{4.11}$$

In appendix G.4 it is shown that the posterior distribution of the linear form θ, under model M_i, follows a univariate Student-t distribution with $\nu = T-1-d$ degrees of freedom, mean $\widehat{\theta} = \iota' \cdot \widehat{\phi}$ and variance $\nu s^2 \cdot c_J/(\nu-2)$, where $c_J = \iota'(\widetilde{\mathbf{X}}_{\mathbf{11}} - \widetilde{\mathbf{X}}_{\mathbf{12}} \cdot \widetilde{\mathbf{X}}_{\mathbf{22}}^{-1} \cdot \widetilde{\mathbf{X}}_{\mathbf{21}})^{-1}\iota$. The marginal posterior $f(\theta|\ \mathbf{y})$ is then obtained in the same way as $f(\phi_s|\mathbf{y})$ in (4.9), and therefore is a model weighted mixture of t-densities.

Let the support of θ be $\Theta = [0,\ S]$ in the following. The posterior probability of the null hypothesis (4.6), i.e. of the singleton $\Theta_0 = \{\theta_0\}$, is obtained as in Berger and Delampady (1987), Berger and Sellke (1987), namely by using a mixed prior density, which assigns positive prior probability mass π_0 to Θ_0,[10] and uses a continuous density $f_1(\theta)$, with prior

[9] This follows from the results of appendix G, by letting $\mathbf{B}_0$ and $\mathbf{M}$ in (6.35) go to zero.
[10] Here $\pi_0 = 0.5$ is used.

weight $1-\pi_0$, on the complementary parameter set $\Theta_1 = \{\theta : 0 \leq \theta < S\}$:[11]

$$P(H_0|\,\mathbf{y}) = \left[1 + \frac{1-\pi_0}{\pi_0} \cdot \frac{f(\mathbf{y}|\,\Theta_1)}{f(\mathbf{y}|\,\theta_0)}\right]^{-1} \tag{4.12a}$$

$$\text{with } f(\mathbf{y}|\,\Theta_1) = \int_{\Theta_1} f(\mathbf{y}|\,\theta) \cdot f_1(\theta) \cdot d\theta \tag{4.12b}$$

where the integral in (4.12b) is computed by Simpson's rule (cf. Bauwens et al. (1999), p.69).

Under a '0-1' loss function then an optimal Bayesian decision rule $\varphi(\mathbf{y})$ is given by (see Robert (2007), proposition 5.2.2, p.225.):

$$\varphi(\mathbf{y}) = \begin{cases} 0\,, & \text{if } P(H_0|\mathbf{y}) \geq 0.5 \\ 1\,, & \text{otherwise} \end{cases} \tag{4.13}$$

where 1 means rejection of the null hypothesis and $P(H_0|\mathbf{y})$ denotes the posterior probability of the latter, which is abbreviated by P_0 henceforth. Further, since the decision rule (4.13) is based on a Student-t posterior density it is called the 'Bayesian t-test' in the following, in order to discriminate it from the F-test introduced next.

Consider testing the linear hypotheses $\mathbf{R}\phi = \mathbf{r}$ in (4.5c) and (4.5d), with $\mathbf{R}$ a $J \times S$ matrix of zeros and ones and $\mathbf{r}$ a J-vector of constants. In the present context, the identity matrix can be used for $\mathbf{R}$ and thus $J = S$. Franses and Koop (1997) propose an approximate Wald-type test of H_C, where they approximate the multivariate Student-t posterior distribution of ϕ by a Normal distribution. Consequently H_C can be tested using the inner product of standardized normal random variables, which follows a $\chi^2_{(S)}$- distribution (see ibid., p.515). However the authors focus in their analysis on quarterly models where there are usually much more degrees of freedom available than in the case of more parameter-intensive monthly PAR models. Since for monthly time series of moderate length, i.e. $T = 100-150$, such an approximation can be problematic, implying $\nu = T - p - d < 100$

[11] Note that in this approach the sets Θ_0 and Θ_1 are treated in a different way, since otherwise Θ_0 would have zero Lebesgue measure. Instead of changing the nature of Θ_0, by assigning positive probability mass to it, Pereira et al. (2008) pursue a different strategy by looking for the 'tangential set', T, of points having posterior density values higher than any in Θ_0. Then H_0 would not be rejected if the posterior probability of T, the Bayesian evidence value against H_0, is small, see Pereira and Stern (1999), Pereira et al. (2008) for details.

degrees of freedom, with $d = (4+p)\cdot S$, an exact Bayesian F-test is used instead. For this purpose define the following standardized inner product:

$$\begin{aligned} Q(\mathbf{r}|\, M_i,\, \mathbf{y}) &\equiv \left[\mathbf{R}(\phi - \widehat{\phi})\right]' \cdot \left[\mathbf{R}(\widetilde{\mathbf{X}}_{\mathbf{11}} - \widetilde{\mathbf{X}}_{\mathbf{12}} \cdot \widetilde{\mathbf{X}}_{\mathbf{22}}^{-1} \cdot \widetilde{\mathbf{X}}_{\mathbf{21}})^{-1}\mathbf{R}'\right]^{-1} \cdot \left[\mathbf{R}(\phi - \widehat{\phi})\right] / (S \cdot s^2) \\ &= (\mathbf{r} - \mathbf{R}\widehat{\phi})' \cdot \left[\mathbf{R}(\widetilde{\mathbf{X}}_{\mathbf{11}} - \widetilde{\mathbf{X}}_{\mathbf{12}} \cdot \widetilde{\mathbf{X}}_{\mathbf{22}}^{-1} \cdot \widetilde{\mathbf{X}}_{\mathbf{21}})^{-1}\mathbf{R}'\right]^{-1} \cdot (\mathbf{r} - \mathbf{R}\widehat{\phi}) / (S \cdot s^2) \end{aligned} \tag{4.14}$$

In appendix G.5 it is shown that the quadratic form follows an $F(\nu_1 = S,\ \nu_2 = T-1-d)$ posterior distribution, with S and $T-1-d$ degrees of freedom. A simple decision rule based on Bayesian 'p-values', pv, for testing the linear hypotheses H_C and H_D, with $\mathbf{R}\phi = \mathbf{r}_0 = \pm 1$, respectively, is to reject the null, whenever

$$pv(\mathbf{r}_0|\, M_i,\, \mathbf{y}) \;\equiv\; 1 - P_F(Q(\mathbf{r}_0|\, M_i,\, \mathbf{y})) \tag{4.15}$$

is smaller than a nominal level of significance, where $P_F(.)$ denotes the cumulative distribution function of the $F(\nu_1 = S,\ \nu_2 = T-1-d)$ posterior distribution.

Since for a selected null hypothesis, i.e. for $\mathbf{r}_0 = \pm 1$, the quadratic form $Q(\mathbf{r}_0,\ T_B|\, m = 1,\ \mathbf{y})$ under a structural break model is only a function of the integer-valued break date T_B, the model averaged quadratic form, denoted by $\overline{Q}_{BMA}$, can be obtained from

$$\begin{aligned} \overline{Q}_{BMA}(\mathbf{r}_0|\, \mathbf{y}) \;=\; & Q(\mathbf{r}_0|\, m = 0,\, \mathbf{y}) \cdot f(m = 0 \mid \mathbf{y}) + \\ & \left[\sum_{T_B=\kappa+2}^{T-\kappa} Q(\mathbf{r}_0, T_B|\, m = 1,\, \mathbf{y}) \cdot f(T_B|\, m = 1,\, \mathbf{y})\right] \cdot f(m = 1|\, \mathbf{y}) \end{aligned} \tag{4.16}$$

with $f(m|\, \mathbf{y})$ the posterior probability mass function of a break occurrence with normalizing constant given by

$$f(\mathbf{y}) \;=\; \sum_{m=0}^{1} \sum_{T_B=\kappa+2}^{T-\kappa} f(\mathbf{y}\,|m,\, T_B) \cdot f(T_B|\, m) \cdot f(m) \tag{4.17}$$

In order to compute the corresponding posterior tail probabilities, the marginalized posterior cumulative distribution function of Q is required. For this purpose, the model specific

$F(\nu_1 = S,\ \nu_2 = T-1-d)$ posterior ordinates are averaged over the two candidate models, i.e.

$$f(Q|\ \mathbf{y})\ =\ f(Q\mid m=0,\ \mathbf{y})\cdot f(m=0\mid \mathbf{y})\ + \\ \left[\sum_{T_B=\kappa+2}^{T-\kappa} f(Q\mid m=1,\ T_B,\ \mathbf{y})\cdot f(T_B|\ m=1,\ \mathbf{y})\right]\cdot f(m=1|\ \mathbf{y}) \tag{4.18}$$

The corresponding Bayesian p-values, i.e. the right tail probabilities of (4.18), and highest posterior density regions can then be obtained by numerical integration.

4.4. Monte Carlo evidence: periodic unit root tests

In this section the results of several Monte Carlo (MC) experiments to study the performances of the Bayes tests of H_B and $H_{(C)D}$, i.e. of a periodic and a real (non)seasonal unit root, are presented. For a test of H_B, or its linear approximation in (4.6), the Bayesian t-test in (4.13) is used, where in (4.12a) the conditional or the model averaged posterior distribution of the linear form θ is utilized. For a test of the hypothesis $H_{(C)D}$ the Bayesian F-test, using the conditional p-value (4.15) or the corresponding model averaged p-value on the basis of (4.18), is computed. In the following simulation study, first the power properties of the conditional tests are compared to those of classical competitors. Then also some simulation evidence concerning the model averaged versions of the Bayes tests is presented.

In the first simulation experiment the Bayesian t-test of H_B, given a PAR(1) model without a break, is compared with a classical competitor, namely the test of Boswijk and Franses (1996) (abbreviated 'BF', see section 3.3 for some details) in terms of test power. For this purpose, trajectories of the following quarterly PAR(1) process without a break are generated:

$$y_t = \sum_{s=1}^{4} D_{s,t}\cdot(\mu_s + \phi_s\cdot y_{t-1}) + \varepsilon_t\ ,\ \varepsilon_t \overset{i.i.d.}{\sim} N(0,1)\ ,\ t = 1...150 \tag{4.19}$$

with seasonal intercepts $\mu_s = 1,\ \forall s$.

In order to estimate the empirical power functions for a given sample size, I generate

trajectories for each $(\phi_1, \phi_2, \phi_3, \phi_4)'$- combination in the $[0.8, 1.1]^4$-space. Here for each ϕ_s, $s = 1...4$, a grid of six values is used, which results in $6^4 = 1296$ parameter constellations. The rejection probabilities of the null hypothesis are approximated by the average number of rejections, i.e.

$$f(\text{'Reject } H_0\text{'} \mid \rho_0, \mathbf{y}) \approx \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}_{(P_0<0.5)} \tag{4.20}$$

where $\mathbf{1}_{(.)}$ denotes an indicator function, N is the number of replications, and ρ_0 denotes a particular value of this parameter.

Note that $(\phi_1, \phi_2, \phi_3, \phi_4) \mapsto \rho = \phi_1 \cdot \phi_2 \cdot \phi_3 \cdot \phi_4$ is not a bijection and thus different ϕ_s-combinations can lead to the same ρ values and hence to similarly persistent processes. For this reason, if there are any ties in the sequence of ρ's at a value ρ_0, the associated rejection frequencies are averaged and this average is taken as the value of the power function at ρ_0.[12] In figure 4.1 the resulting power function for $N = 100$ replications and a sample size of $T = 150$ is shown.[13] It is evident that with the exception of seven ρ-

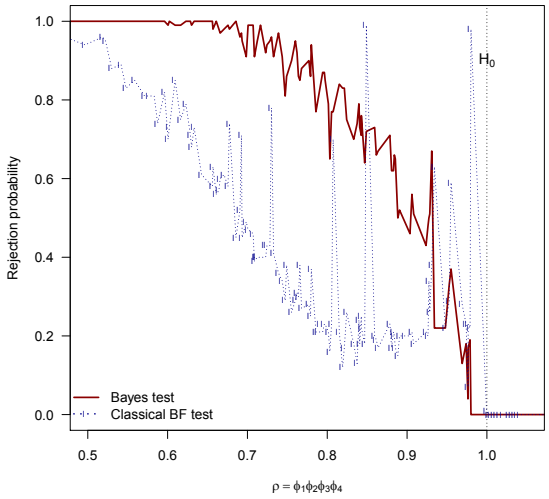


Figure 4.1.: Power functions of the Bayesian t-test and the BF-test for quarterly periodic data ($T = 150$)

[12] Franses (1995) considers a similar MC experiment to analyze the empirical power properties of the BF-test and some other related tests. However the author only simulates two specific data generating processes in order to compute the empirical sizes and powers of the tests. In contrast, computation of the whole power function allows to draw a more comprehensive picture of the test characteristics.

[13] Similar power functions are obtained using larger sample sizes T.

values, the Bayes test outperforms the classical BF-test in terms of power. This is most pronounced if ρ lies in the stationary parameter region. When approaching the null, i.e. $\rho \to 1$, both procedures perform almost equally in terms of power.

In the second MC experiment the power function of the Bayesian F-test of the hypothesis H_D, i.e. a seasonal (biannual) unit root, is simulated.[14] As a classical competitor the test of Hylleberg et al. (1990) (HEGY) for a (real-valued) seasonal unit root at the π-frequency is chosen. The HEGY testing approach uses a t-statistic to test the point null of a real seasonal unit root against a left-sided alternative (see ibid., also Ghysels and Osborn (2001), p.60, for details). For the subsequent results the finite sample critical values in Hylleberg et al. (1990), p.226, have been utilized.[15]

For the computations of the power functions $N = 200$ trajectories of the following simple AR(1) process without a break are generated:

$$y_t = \mu + \phi y_{t-1} + \varepsilon_t \ , \ \varepsilon_t \overset{i.i.d.}{\sim} N(0,1) \ , \ t = 1...100 \tag{4.21}$$

with $-1 \leq \phi < 0$ with $\mu = 0$.

Figure 4.2 shows a realization of the process (4.21) under the null H_D, i.e. for $\phi = -1$, together with the sample (partial) autocorrelation function (S(P)ACF) and the periodogram. Note that for $\phi = -1$ the system in (4.21) implies a 'bounce back' and hence exhibits a half-cycle every period. This oscillating behavior induced by an alternating sign can also be observed when solving the stochastic difference equation for y_t recursively, which yields[16]

$$y_t = \begin{cases} -y_0 + \mu + \sum_{j=0}^{t-1} (-1)^j \cdot \varepsilon_{t-j} & \text{, for } t \text{ odd} \\ y_0 + \sum_{j=0}^{t-1} (-1)^j \cdot \varepsilon_{t-j} & \text{, for } t \text{ even} \end{cases} \tag{4.22}$$

From figure 4.2 it is obvious that most of the variation in y_t can be attributed to the biannual frequency π since the periodogram has its maximum at two observations.[17] As

[14] The simulation results for a nonseasonal unit root (H_C) are nearly identical and are therefore omitted to save space.

[15] All classical tests are conducted on a 5% nominal level of significance.

[16] Note that the intercept does not translate into a linear time trend here as in the case of a nonseasonal Random Walk process.

[17] The periodogram is computed only over the range of $[0, S/2]$ observations, because of the aliasing effect, cf. Bloomfield (2000).

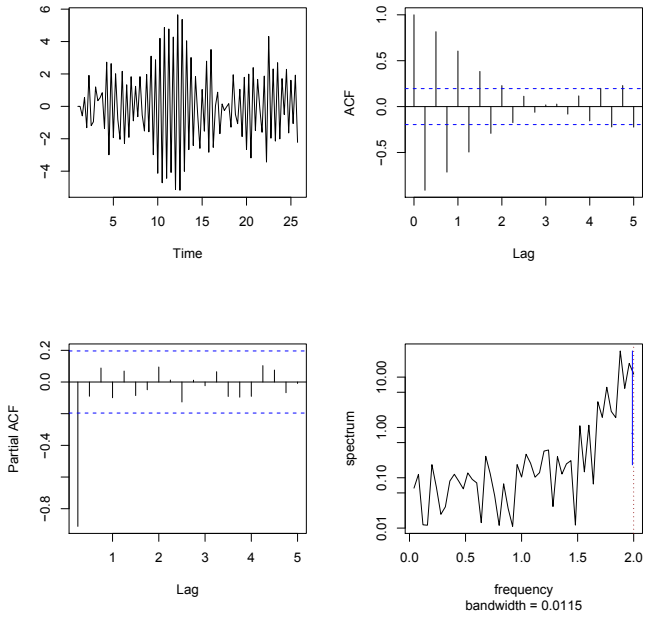


Figure 4.2.: AR(1) trajectory together with S(P)ACF and estimated power spectrum for $T = 100$

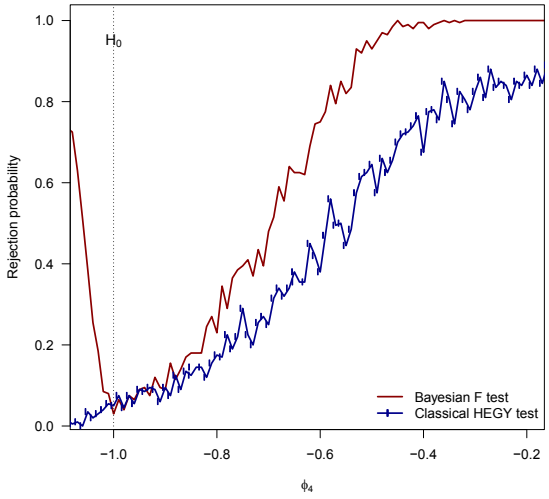


Figure 4.3.: Comparison of power functions for quarterly nonperiodic data ($T = 100$)

can be seen from figure 4.3 the Bayesian F-test for a unit root at the π-frequency, given a PAR(1) model without a break, outperforms the classical HEGY test in terms of power. The increasing rejection frequencies of the F-test for $\phi < -1$ are due to its two-sided

alternative region, whereas the HEGY t-test has a one-sided alternative.[18] Observing the actual sizes of the Bayesian and classical procedure reveals that the former exhibits a rejection frequency of 3.5% under the null, whereas the latter test with 5.5% almost exactly follows its nominal significance level.

Summing up, the two Bayesian unit root tests for a PAR(1) model without a break perform favorable in terms of power when compared to classical competitors. Since the previous results have been obtained for a given model specification M_i, next also some simulation evidence for the model averaged versions of the two Bayes tests is presented. These are used to test the hypothesis H_C, i.e. of a nonseasonal unit root. With regard to the empirical analysis in the next section, the subsequent simulation experiments are conducted using monthly data ($S = 12$). As a DGP, a nonperiodic AR(1) process with a break in the seasonal intercepts at $T_B = T/2$ and without a break is used, where the process with a break has the following form:

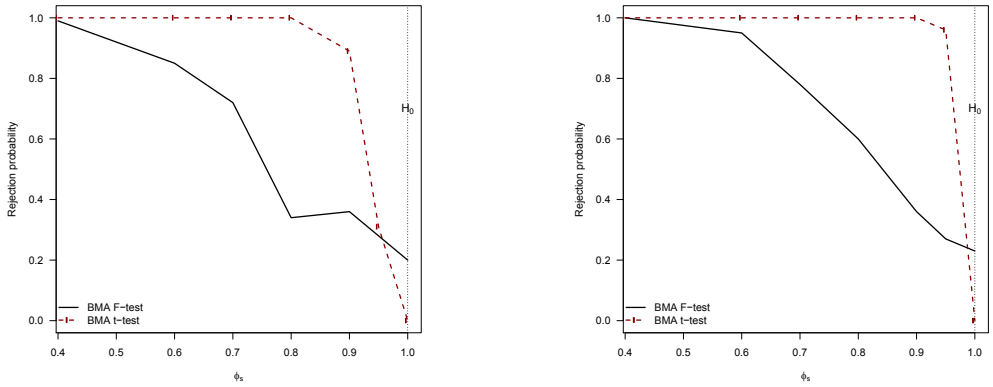
$$y_t = \sum_{s=1}^{12} (\mu_s D_{s,t} + \mu_s^\star D_{s,t}^\star) + \phi \cdot y_{t-1} + \varepsilon_t \ , \ \varepsilon_t \overset{i.i.d.}{\sim} N(0,1) \ , \ t = 1...T \tag{4.23}$$

with $\mu_s = 1.5$, $\mu_s^\star = 0.2, \forall s$, and the dummy variables $D_{s,t}$ and $D_{s,t}^\star$ are defined as in section 4.2. In case of no structural break $\mu_s^\star = 0, \forall s$.

For the simulation of the power functions a grid of ϕ-values, with $\phi \in [0.4, 1.05]$, is used.[19] In order to examine the small sample performance of the two Bayes tests, trajectories of length $T = 100$ are generated from (4.23). This is repeated $N = 100$ times and the rejection probabilities are then approximated by the average rejection frequencies. In figures 4.4(a) and 4.4(b) the power functions of the two tests are shown for a DGP with and without a structural break.

From there it is evident that the BMA F-test exhibits quite distinct test properties compared to the BMA t-test in this simulation experiment. The latter test has favorable power in case of a break and also in case of no break. In particular, the F-test shows an overrejection (with $\approx 20\%$) under the null hypothesis H_C, i.e. an increased type 1 error, in contrary to the t-test (with $\approx 0\%$). The results of other simulation experiments, which are omitted here to save space, indicate that these overrejections can be attributed to a biased estima-

[18] As a consequence rejection of $y_t \sim SI(1)$ does not necessarily imply stationarity.
[19] In order to save computing time the power functions are approximated at seven ϕ-values.



(a) With break (b) Without break

Figure 4.4.: Power functions of the BMA F- and t-test for PAR(1) processes with(out) a break ($T = 100$).

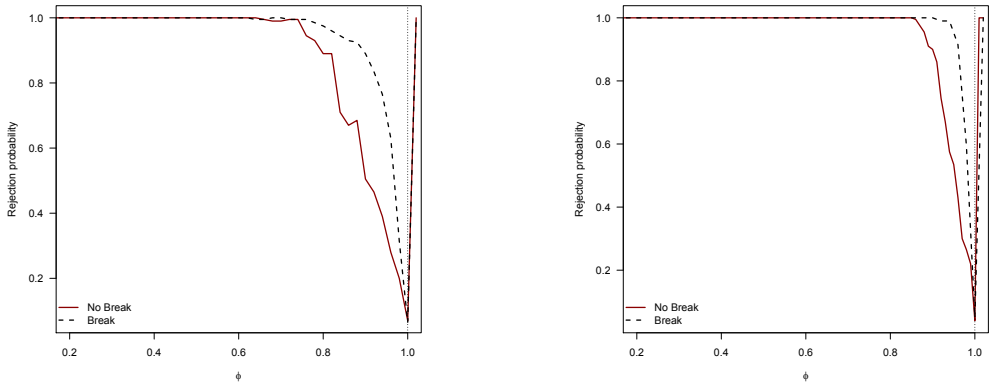
tion of the model probabilities, $f(M_i|\mathbf{y})$, in small samples. This bias mainly depends on the size of the innovation variance σ^2 and to some extent also on the specification of the deterministics in the DGP. To get an impression of this small sample bias, table 4.1 shows the estimated posterior probabilities for the DGP in (4.23) with $m = 1$ and $m = 0$, i.e. with and without a break.

Table 4.1.: Posterior model probabilities

DGP	No Break ($T = 100$)		Break ($T = 100$)		No Break ($T = 300$)	
ϕ	$m = 0$	$m = 1$	$m = 0$	$m = 1$	$m = 0$	$m = 1$
0.40	0.34	0.66	0.19	0.81	0.72	0.28
0.60	0.27	0.73	0.17	0.83	0.77	0.23
0.70	0.38	0.72	0.14	0.86	0.79	0.21
0.80	0.38	0.72	0.15	0.85	0.80	0.20
0.90	0.26	0.74	0.08	0.92	0.75	0.25
0.95	0.24	0.76	0.04	0.96	0.71	0.29
1.00	0.01	0.99	0.01	0.99	0.46	0.54

From table 4.1 it can be observed that the estimated posterior probabilities of $m = 0$ are biased towards a model with break when the sample size is small ($T = 100$), but this bias

is reduced when the sample size increases ($T = 300$).[20] What can also be recognized from the results is that for $\phi \to 1$, a process without a break resembles a process with a break, and thus it becomes harder to discriminate between both. In order to analyze the effect of the estimated model weights on the outcome of the BMA F-test, the power function is simulated for the same DGPs as above (see (4.23)), but now a conditional F-test, given the true model specification, is used. In figures 4.5(a) and 4.5(b) the resulting power functions for $T = 100$ and $T = 200$ are shown. The figures suggest that the conditional



(a) For $T = 100$ (b) For $T = 200$

Figure 4.5.: Power functions of the F-test for PAR(1) processes with(out) a break.

Bayesian F-test has favorable power and size properties and that its power increases with the sample size.

To sum up, the model averaged Bayesian t-test shows favorable power and size properties compared to classical competitors, and is less affected by a (possibly) biased estimation of the model weights in small samples. In contrast, the BMA F-test can have serious size problems in the case of short time series. The latter finding should be kept in mind when interpreting the results of the subsequent empirical analysis.

[20] For larger sample sizes this effect becomes even more pronounced.

4.5. Empirical application to monthly unemployment data

In this section the unit root tests presented above are applied to answer the question, if there is persistent behavior in the unemployment rates of selected OECD countries. Here the harmonized monthly unemployment rates[21] of 17 OECD countries from January 1999 to March 2012 ($T = 159$) are used.[22] The high level of unemployment in countries of the European union compared to other countries of the OECD has been an object of investigation for many years. In the economic literature in principal there are two theoretical explanations of this phenomenon: the non-accelerating inflation rate of unemployment (NAIRU) and unemployment hysteresis. The former theory implies the unemployment rate to follow a trend stationary process, i.e. after an exogenous shock the rate will recover to its long run equilibrium. By contrast hysteresis implies that temporary shocks have permanent effects on the level of unemployment and thus the underlying stochastic process has a unit root. This 'unit root hysteresis' definition is the most common in the literature on the stochastic properties of unemployment rates (cf. Blanchard and Summers (1986)). Persistence can be regarded as a special case of the NAIRU concept in the sense that the unemployment rate follows a stationary process but also has a stochastic component that is nearly integrated of order one, for example in the present case with the product of autoregressive coefficients being close to one ('Quasi Random Walk').

As a preliminary step of the empirical analysis diagnostic tests are applied to check for the presence of periodic variation in the level of the series. Here I follow a similar strategy as used in Boswijk and Franses (1996), p.231, within a classical context. Therefore, first a Bayesian PAR(1) model with seasonal intercepts and without a break is estimated, using the respective posterior means as point estimates and then the null of no periodicity, i.e. $H_0 : \phi_s = \phi$, $\forall s$, is tested against the alternative of periodicity $H_1 : \phi_s \neq \phi$, $\exists s$. This test can readily be conducted by using the conditional Bayesian F-test[23] for the linear hypothesis

[21] All series are analyzed in logarithmic form.

[22] The harmonized unemployment rates give the numbers of unemployed persons as a percentage of the labor force. The labor force consists of employees, the self-employed, unpaid family workers and the unemployed. The used data set was extracted from OECD sources, see www.oecd.org.

[23] Note that in a similar manner one could also construct a Bayes test in order to test for periodicity in the variance of the series, i.e. for seasonal heteroscedasticity, see Franses and Paap (2006), p.40 for a classical approach.

$\mathbf{R} \cdot \phi = \mathbf{0}$ introduced above, with $\mathbf{R} = [\mathbf{I}_{S-1}\ ,\ -\iota]$ an $S-1 \times S$ matrix of contrasts, $\mathbf{I}_{S-1}$ the identity matrix, and ι an $S-1$ vector of ones. Hence a Bayesian F-test of H_0, denoted F_{PAR}, can be based on an $F(\nu_1 = S-1,\ \nu_2 = T-1-d)$ distribution when S seasonal intercepts are included in the model, see also Franses (2003), p.104, and Boswijk and Franses (1996) for a similar result using a classical framework.

Table B.1 (see appendix B) shows the results for the F_{PAR}-test together with the corresponding 'Bayesian p-values'. The results indicate that, even by assuming a liberal significance level of 10%, for only 6 out of 17 countries the null of no periodic variation can be rejected. According to this test there seems to be not much evidence for periodicity in the OECD unemployment series. In order to check the robustness of the results, I pursue a strategy similar to the classical approach in Franses (2003), p.116, by using a recursive testing strategy. Therefore, the Bayesian F_{PAR}-test is computed for rolling subsamples of the original data. The corresponding F-tests when one year of data is added and removed successively are called a Bayesian *forward* and *backward recursive* F_{PAR}-test, respectively. Figures E.20 to E.28 (see appendix E) show the evolution of the corresponding Bayesian p-values. Note that for the forward recursive tests the first time window reaches from 1/1999 to 12/2002 and the final window from 1/1999 to 3/2012, i.e. the whole time period. In contrast, for the backward recursive tests the first time window is from 1/1999 to 3/2012, where the last window includes the period from 1/2008 to 3/2012. The plots indicate that for most countries the sequence of F-tests exhibits some time windows where there seems to be significant periodicity in the estimated coefficients on the 5% significance level with the exception of France, Japan, Norway, Spain and the US.[24] On the grounds of these results the application of a flexible PAR model seems justifiable. Note that there is some analytical as well as empirical evidence (cf. Franses (2003)) for starting with a PAR model, which includes the possibility of having $S-1$ seasonal unit roots as well as the zero frequency unit root, instead of first transforming the data, e.g. by applying a seasonally differencing filter Δ_S, since this already assumes a certain number of unit roots.

In table B.3 the test results for a periodic unit root are shown for each OECD country. In the second column of table B.3 the deterministic model specification is indicated. The inclusion of seasonal intercepts ('Drift') or the inclusion of seasonal intercepts together

[24] In these five cases a more parsimonious SARMA model could also be fitted.

with a seasonal time trend ('Both') is suggested on the grounds of visual inspection of the data. In columns three and four of table B.3 the posterior probabilities, P_0, of H_0 : $\theta = \sum_{s=1}^{12} \phi_s = 12$, using a PAR model with a break and using the BMA version of the t-test are given, respectively. Columns five and six show the posterior means and the HPD regions of $\rho = \prod_{s=1}^{12} \phi_s$, given the structural change model with the break date set equal to the posterior mode T_B^{MAP} of $f(T_B|\ m = 1, p = 1, \mathbf{y})$.[25] The posterior modes are given in table B.2. In table B.3 it can be seen that for 10 out of 17 of the OECD countries the null hypothesis of a periodic unit root can not be rejected on the basis of the BMA posterior probabilities (P_0). Among the four countries having the highest probabilities are Greece, Ireland, Spain and the UK. Almost the same conclusions can be drawn when using the HPD regions of ρ. Here for 11 countries the null constraint (4.5b) lies within the 95%-HPD region, which implies that for these countries the null can not be rejected when assuming a 5% level of significance.[26] The last column also shows the empirical results from an application of the classical BF test. It is obvious that with the exception of Finland the null can not be rejected. This finding can be attributed to the relatively lower power of the test in small samples as noted in section 4.4, and also to the fact that the test does not control for a possible structural break.

Given the results of the periodic unit root tests, I test for a possible unit root at the zero and π-frequency using the model averaged F-test. The results are shown in table B.4 and suggest that for most of the series for which the hypothesis of a periodic unit root could not be rejected also the null of a nonseasonal unit root can not be rejected, namely for France, Germany, Norway, UK and the US. For Belgium, Canada and the series of Japan the null of a zero frequency unit root can not be rejected when assuming the usual levels of significance. For the latter two also a relatively high posterior probability of a periodic unit root can be seen from table B.3 although these probabilities are still below 50%.[27] With the exception of Belgium the findings of table B.4 are in accordance with those in table B.3 and imply that the unemployment series of the mentioned countries are best described by a PARI process, i.e. a PAR process for an $I(1)$ series.

[25] That is, first the estimate of the break date is computed and then the sampling approach of section 4.3 is applied, conditional on the estimated break date.

[26] Note that for Japan the upper bound is only 1.01.

[27] For Belgium the findings are not clear. However given the above MC evidence, i.e. that for short series the F-test exhibits lower power than the test based on P_0, it seems justifiable to draw the inferences using only the latter approach. Then it could be concluded that the series of Belgium is (trend)stationary.

The test results are also supported when analyzing each period separately by means of the HPD regions of the ϕ_s-coefficients given in tables B.5 and B.6, respectively.[28] An examination of these tables reveals widespread evidence for hypothesis $H_A : \phi_s = 1$, i.e. a Random Walk in season s (see (4.5a)), since most series yield posterior densities of ϕ_s that include the null restriction in their 95% confidence regions. Finally, as can be seen from the results of the π-frequency unit root tests (see table B.4), there seems to be no evidence for seasonal integration in the data.

Summing up, indication of a periodic unit root has been found for the unemployment rates of Denmark, Greece, Ireland, Netherlands and Spain, which means that these series are driven by a PIAR process, i.e. a periodically integrated AR process. Note that a periodic unit root implies that the dynamic response of the unemployment rate to a shock also depends on the season. Furthermore, for the series of Canada, France, Germany, Norway, UK and the US the results suggest the existence of a nonseasonal unit root ('PARI'). For Japan there seems to be some evidence for a stochastic trend, too. For the remaining series, i.e. those of Australia, Finland, Italy and Sweden no evidence for a (non)periodic unit root is found. Hence it can be concluded that most of the unadjusted monthly OECD unemployment rates are most likely (in terms of posterior probability) driven by a unit root process. More precisely, the empirical results suggest that this process is in most of the cases associated with a nonseasonal, i.e. zero frequency unit root, which is also in accordance with the estimates of the spectral densities of the series. This suggests that the corresponding time series should be modeled by taking (non)periodic first order differences (see Ghysels and Osborn (2001), p.153). In the light of economic theory, this supports theories of unemployment hysteresis, which imply that labor market shocks have a permanent impact on the level of unemployment.

4.6. Summary and concluding remarks

In this chapter Bayesian testing strategies to test for a periodic unit root with a possible break at an unknown point in time have been presented. On these grounds also an exact Bayesian F-test has been proposed in order to test for a unit root at the zero and the π-

[28] The HPD regions are computed on basis of the model averaged Student-t posterior densities of the ϕ_s coefficients.

spectral frequency. All tests have been based on model averaging techniques so that it is possible to combine evidence from a model with and without a structural break. The results of simulation experiments to study the performance of the tests indicate that the model averaged Bayesian F-tests for a zero and a π-frequency unit root can have low power and increased size in small samples. This feature has been attributed to a small sample bias of the estimated model posterior probabilities, i.e. the weights of the BMA tests. In contrast, the test for a periodic unit root has shown to be less sensitive with respect to the sample size and had better power properties throughout all conducted simulation experiments. In practice the true data generating process is usually unknown and hence conditioning on the true parametrization is not possible. Here one either has to select a particular specification for the analysis or has to resort to model averaging techniques. With the latter approach it is possible to capture the uncertainty induced through a model selection step.

In an empirical application these methods have been used to test for unemployment hysteresis effects in the monthly unadjusted unemployment rates series of 17 OECD countries. The results show that in fact most of the analyzed unemployment rates exhibit unit root behavior. Among the four countries having the highest posterior probabilities of a periodic unit root are Greece, Ireland, Spain and the UK. Moreover many of these series are driven by a nonperiodic stochastic trend, which is implied by a zero frequency unit root. This is also in accordance with a conducted Bayesian pretest for periodicity, which has shown that there is not much periodic variation in the unemployment series. Moreover no evidence for (real-valued) seasonal unit roots has been found. With regard to further inference, e.g. prediction of future data, these empirical findings suggest that the seasonality in the monthly OECD unemployment data should in most cases be modeled by seasonal nonperiodic models, after having applied the relevant differencing filters. In the light of economic theory this suggests that there is evidence for unemployment hysteresis in the respective countries, which implies that labor market shocks have a permanent impact on the level of unemployment. In future research the Bayesian periodic unit root test could be extended to allow for heterogeneity in the lag orders p_s and the residual variances σ_s^2. Another interesting branch of future work would be to extend the above testing procedure in order to capture also skewness and leptokurtic behavior in the data.

In the next chapter the focus will be on forecasting seasonal time series data. For this

purpose, a Bayesian prediction approach based on model averaging will be presented in the following.

Forecasting seasonal time series. A Bayesian model averaging approach

5.1. Introduction

Many time series used in economic research exhibit some form of stochastic or deterministic seasonality. There are two principal ways to deal with such sort of variation, i.e. to adjust the data or to model the variation explicitly using seasonal time series models. In this chapter the latter strategy is pursued with a focus on the prediction of future data. In the nonseasonal time series literature a large number of alternative models has been proposed for this task, see Ghysels and Osborn (2001), Ghysels et al. (2006) for overviews. A distinguishing feature of seasonal time series models is the way in which seasonality is represented. The latter can be modeled to be constant over the different seasons, $s = 1...S$, i.e. quarters, months and so on. Another possibility is to allow for seasonal heterogeneity or periodicity in the stochastic and/or deterministic parts of the data generating process. As noted by Osborn and Smith (1989) many empirical applications give no reason for assuming the model parameters to be constant over the seasons, beyond parsimony and convention. The class of periodic autoregressive moving average (PARMA) models provides a flexible alternative to other seasonal linear time series models, like seasonal autoregressive moving average (SARMA) models, by allowing the coefficients to vary with the seasons, see Ghysels and Osborn (2001), Franses and Paap (2006). In most economic applications periodic models are assumed to take an autoregressive form and are known as periodic autoregressive (PAR) models. The latter can be useful in capturing economic situations where agents show distinct characteristics over the seasons, such as

seasonally varying utilities of consumption, see for example Osborn et al. (1988). Important contributions to the research on PARMA models have been made by Gladyshev (1961), Pagano (1978), Tiao and Grupe (1980), Vecchia (1985), Osborn (1991), Franses (1994), Boswijk and Franses (1996), inter alia.

Nearly all works on PARMA models use a classical frame of reference with the exception of Andel (1983) and Franses and Koop (1997) who choose a Bayesian framework. The latter authors present a Bayesian approach for prediction and unit root testing in PAR(p) models. In the subsequent analysis the forecasting approach of Franses and Koop (1997) is extended in several directions. First, the authors treat the autoregressive lag order p as fixed and known. In contrast, I treat the lag order as a discrete random variable to which a prior probability mass is assigned. Because the required number of autoregressive coefficients is in general $p \cdot S$, lag order selection may be more crucial in the context of PAR models than in constant-parameter AR models, see also Ghysels et al. (2006). Second, Franses and Koop (1997) allow for the possibility of one structural break in the seasonal intercepts at unknown time and then average out this nuisance parameter. In the subsequent analysis I allow for one structural break in the seasonal intercepts and/or seasonal time trends, but treat the occurrence of a break, or the number of breaks $m = 0, 1$, as an additional random parameter. Then the autoregressive lag order and the number of structural breaks are used as model indicators in order to identify different candidate models, M_i, in the model space. Most importantly, instead of resorting to a model selection approach by picking out a single model for inference, as for example in Franses and Koop (1997), a prediction approach for PAR models with an unknown number of lags and an unknown number of structural breaks based on model averaging is presented.

There is a growing literature concerned with model averaging. Seminal contributions to Bayesian model averaging are Madigan and Raftery (1994), Draper (1995), Raftery et al. (1997), and for a review of the literature Hoeting et al. (1999), Raftery and Zheng (2003). Some econometric BMA applications include Fernandez and Steel (2001), Hong and Preston (2012), inter alia. See for example Hjort and Claeskens (2003), Hansen (2007) for frequentist approaches and additional references. As argued by some authors (cf. Raftery et al. (1996)) model selection procedures ignore the uncertainty induced through a model selection step and thus underestimate uncertainty about the quantities of interest, as for example future observations. Hence accounting for model uncertainty by

using a model averaging approach may improve out-of-sample predictive accuracy. With regard to forecasting this seems to be a natural approach to capture model uncertainty by including all models M_i under consideration into a (super) model.

Further, in appendix H it is shown that the joint posterior predictive distribution of a PAR model is given by the product of Student-t densities, viz. the one-step ahead posterior predictive distributions given the preceding forecasts. Since no analytical expressions for the required marginal posterior predictive distributions of the y_{T+k}, $k = 1...K$, exist, and further since the possibility of a structural break at an unknown point in time introduces an additional nuisance parameter, which can not be integrated out analytically, a Markov chain Monte Carlo approach based on data augmentation (see Tanner and Wong (1987)) is presented. By using the model posterior probabilities as weights in order to capture uncertainty a model averaged posterior predictive distribution is computed to forecast future data. The empirical results presented in this chapter indicate that the BMA prediction approach is able to improve forecasting accuracy when compared to a model selection approach.

After having introduced the Bayesian PAR (BPAR) model in section 5.2 the Gibbs sampling algorithm for parameter estimation and sampling future values is presented in section 5.3. In order to examine the forecasting performance under different data generating processes the results of a simulation study are given in section 5.4. Here also a Bayesian test is introduced in order to compare the predictive ability of different forecasting models. In section 5.5 an empirical application of the proposed prediction approach is presented. Here the monthly unemployment rates of the 16 federal states of Germany and of East- and West-Germany are predicted 12-months ahead. Finally, section 5.6 then summarizes the results and concludes.

5.2. Periodic autoregressive models with one break

In the following I focus on PAR models with at most one structural break in the deterministic component(s) at an unknown point in time. Similar models have been used for example in Franses and Koop (1997) and Franses (2003). In case of a structural break in the periodic intercepts and time trends the model has the following general form, which

is reproduced here from section 4.2 for the ease of reference:

$$y_t = \Big(\sum_{s=1}^{S}\sum_{i=1}^{p_s} \phi_{i,s}\cdot D_{s,t}\cdot y_{t-i}\Big) + \lambda_t + \varepsilon_t \;\;,\;\; \varepsilon_t \overset{i.i.d.}{\sim} N(0,\sigma_s^2) \tag{5.1a}$$

$$\lambda_t = \sum_{s=1}^{S}\Big((\mu_s + \alpha_s\cdot\tau_t)\cdot D_{s,t} + (\mu_s^\star + \alpha_s^\star\cdot\tau_t)\cdot D_{s,t}^\star\Big) \tag{5.1b}$$

with $t = 1,...,T$ observations and $s = 1,...,S$ seasons.

As in the analysis of section 4.2, let $s_t = 1 + \big[(t-1)\text{mod}\, S\big]$ denote the season of observation t, and let $T_B \in \,]p_s + \kappa,\; T - \kappa]$ be the unknown break point, where κ denotes the first and last ten percent of observations in the sample, which are truncated in order to avoid possible endpoint problems in the identification of T_B. Then the seasonal dummy variables in (5.1) are given by $D_{s,t} = 1,$ if $s_t = s \;\wedge\; t \leq T_B$, otherwise 0, and $D_{s,t}^\star = 1,$ if $s_t = s \;\wedge\; t > T_B$, and 0 otherwise. Last, τ_t in (5.1b) denotes the value of the linear time trend at time t. Further it is assumed that $p = p_s$ and $\sigma^2 = \sigma_s^2,\; \forall s$, i.e. a homogeneous autoregressive lag order across the seasons and no seasonal heteroscedasticity.

The above periodic autoregressive structural break model can be expressed more conveniently in matrix form (see section 4.2 for definitions):

$$\mathbf{y} \;=\; \mathbf{X}\cdot\phi + \mathbf{Z}\cdot\delta + \varepsilon = \underset{T-p\;\times\;(4+p)\cdot S}{\Big(\mathbf{X}\,\vdots\,\mathbf{Z}\Big)} \cdot \underset{(4+p)\cdot S\;\times\;1}{\begin{pmatrix}\phi\\ \delta\end{pmatrix}} + \varepsilon \;=\; \widetilde{\mathbf{X}}\cdot\mathbf{B} + \varepsilon \tag{5.2}$$

where in the following $d \equiv (4+p)\cdot S$.

In order to draw any inference within a Bayesian framework, the specification of prior distributions for the unknown parameters is required. These are introduced in the next section.

5.2.1. Specification of prior distributions

Within a Bayesian frame of reference a priori knowledge with respect to the quantities of interest can be imposed. For the analysis below I assume a conjugate normal prior for the

regression coefficients and express prior ignorance with respect to all other parameters, i.e.

$$f(\mathbf{B} \mid \sigma^2) = N_d(\mathbf{B}_0 \, , \, \sigma^2 \mathbf{V}) \tag{5.3a}$$

$$f(\sigma^2) \propto 1/\sigma^2 \, , \text{ with } \; \sigma^2 > 0 \tag{5.3b}$$

$$f(T_B) \propto \frac{1}{T-p-2\kappa} \; , \text{ with } \kappa + p < T_B \leq T - \kappa \tag{5.3c}$$

with prior hyperparameters $\mathbf{B}_0 = (\phi_0', \; \delta_0')'$ and $\mathbf{V}$. For the covariance matrix of the regression coefficients

$$\mathbf{V} = \begin{pmatrix} c_1 \cdot \mathbf{I}_{d_1} & 0 \\ 0 & c_2 \cdot \mathbf{I}_{d_2} \end{pmatrix} \tag{5.4}$$

is chosen, where c_1 and c_2 are fixed scalars and $\mathbf{I}_{d_j}, \; j = 1,2$, denote identity matrices of dimensions conformable with the matrices $\mathbf{X}$ and $\mathbf{Z}$ in (5.2), cf. Franses and Koop (1997).[1] Further $\delta_0 = \mathbf{0}$ is imposed to express ignorance with respect to the prior means of the seasonal dummy and trend coefficients. For the prior means of the autoregressive parameters $\phi_0 = \iota$ is assumed, with ι a vector of ones, since experience shows that in many practical applications of PAR models the estimated autoregressive coefficients are close to one. A similar observation has been made by other authors, cf. Franses and Koop (1997). However it should be noticed that for long time series this prior constraint is dominated by the sample information.

Let m_{max} and p_{max} denote the maximum number of structural breaks and the maximum number of autoregressive lags, respectively. In the subsequent analysis the number of breaks, which for simplicity is restricted to $m = \{0,1\}$, and the number of autoregressive lags are used as model indicators in order to identify different candidate models, $M_i = (p,m)$.[2] Thus the discrete space of possible models, denoted by Γ in the following, is given by the cartesian product $\Gamma = \{1,...,p_{max}\} \times \{0,1\}$. In order to express ignorance with respect to the model complexity, i.e. the size of the model, I use the following joint

[1] In the empirical analysis below c_1 and c_2 in (5.4) are both set equal to 100 in order to express lack of prior knowledge with regard to the variation of the regression coefficients.

[2] In principle, the break location, T_B, could also be included in the definition of M_i. Instead it is treated as a nuisance parameter here and is averaged out.

prior distribution for the model indicators

$$f(M_i) \propto 1/\#\Gamma \tag{5.5}$$

where $\#\Gamma$ denotes the cardinality of the set Γ.

One could argue that imposing a flat prior over the space of admissible models is in fact informative, since more complex models are assigned the same prior weight as simpler models and thus one should use a prior that assigns less weight to more complex models, see for example Phillips (1991*b*), Phillips (1991*a*) for a discussion of this issue in the context of unit root testing. However for the analysis below I restrict the model space to $m_{max} = 1$ and $p_{max} = 12$, which implies putting zero prior weight to more complex model combinations. Note that $p_{max} = 12$ implies considering also annual lag structures.

As a specific model M_i can be identified by its (p,m)-combination, and a corresponding break date in case of $m = 1$, I condition on M_i in the following to indicate inference under a particular model specification. Combining the above prior distributions with the likelihood function according to Bayes' Theorem, it is straightforward to show that the joint posterior density function of all unknown quantities, given the break date T_B, has the form of a Normal-Inverse-Gamma-2 distribution, see Bauwens et al. (1999). However, when not conditioning on the discrete-valued break point, the joint posterior becomes a mixture of discrete and continuous densities. Of course this case only arises when $m = 1$.

5.2.2. Model augmentation for prediction

The main objective in a Bayesian approach to forecasting is to derive the predictive density function, $f(\widetilde{\mathbf{y}}_K|\mathbf{y})$, which does not depend on the unknown parameters, and contains all information about the unobserved ('missing') future values in $\widetilde{\mathbf{y}}_K$, given knowledge of the past observations $\mathbf{y}$ (see Zellner (1971), Judge et al. (1985)). In order to predict future values y_{T+k}, $k = 1...K$, I define a K-vector of $\widetilde{\mathbf{y}}_K$-values and their corresponding lags for

$s = 1...S$ and $k = 1...K$:

$$
\begin{aligned}
\widetilde{\mathbf{y}}_K &\equiv [\, y_{T+1}, \ldots, y_{T+K} \,]' \\
\widetilde{\varepsilon}_K &\equiv [\, \varepsilon_{T+1}, \ldots, \varepsilon_{T+K} \,]' \\
\widetilde{\mathbf{y}}^{(s)}_{K-1} &\equiv [\, D_{s,T+1} \cdot y_T,\ D_{s,T+2} \cdot y_{T+1},\ \ldots\ , D_{s,T+K} \cdot y_{T+K-1} \,]' \\
&\ \ \vdots \\
\widetilde{\mathbf{y}}^{(s)}_{K-p} &\equiv [\, D_{s,T+1} \cdot y_{T+1-p},\ D_{s,T+2} \cdot y_{T+2-p},\ \ldots\ , D_{s,T+K} \cdot y_{T+K-p} \,]' \\
\mathbf{X}_K &\equiv [\, \widetilde{\mathbf{y}}^{(1)}_{K-1},\ \ldots\ , \widetilde{\mathbf{y}}^{(1)}_{K-p},\ \ldots,\ \widetilde{\mathbf{y}}^{(12)}_{K-1},\ \ldots\ , \widetilde{\mathbf{y}}^{(12)}_{K-p} \,] \\
\mathbf{Z}_K &\equiv [\, D_K,\ D^{\star}_K,\ D_K \cdot \tau_K\ ,\ D^{\star}_K \cdot \tau_K \,]
\end{aligned}
$$

Finally, stacking all stochastic and deterministic future components as

$$
\underset{K\ \times\ (4+p)\cdot S}{\widetilde{\mathbf{W}}_K} \equiv [\, \mathbf{X}_K \vdots \mathbf{Z}_K \,] \tag{5.6}
$$

yields the following prediction equations

$$
\widetilde{\mathbf{y}}_K = \widetilde{\mathbf{W}}_K \cdot \mathbf{B} + \widetilde{\varepsilon}_K\ ,\ \widetilde{\varepsilon}_K \overset{i.i.d.}{\sim} N_K(\mathbf{0},\ \sigma^2 \mathbf{I}_K) \tag{5.7}
$$

where it is assumed that future observations are generated by a similar process as the observed data.

The joint probability density function (pdf) of all unknown quantities, given the data, can be factorized as follows:[3]

[3] In the following the first p observations are used as initial values $\mathbf{y}_0$. Thus I work with the approximate instead of the exact likelihood function, see Bauwens et al. (1999), p.134 f. The conditioning on $\mathbf{y}_0$ is suppressed subsequently.

$$
\begin{aligned}
f(\widetilde{\mathbf{y}}_K, \mathbf{B}, \sigma^2, T_B |\, M_i,\, \widetilde{\mathbf{X}},\, \widetilde{\mathbf{W}}_K,\, \mathbf{y}) \;&=\; f(\mathbf{B}, \sigma^2, T_B |\, M_i,\, \widetilde{\mathbf{X}},\, \mathbf{y}) \cdot f(\widetilde{\mathbf{y}}_K |\, M_i, \mathbf{B}, \sigma^2, T_B, \widetilde{\mathbf{W}}_K,\, \mathbf{y}) \\
&\propto\; f(\mathbf{B} |\, \sigma^2, T_B,\, M_i) \cdot f(\sigma^2) \cdot f(T_B |\, M_i) \cdot \\
&\quad\; f(\mathbf{y} |\, M_i, \mathbf{B}, \sigma^2, T_B,\, \widetilde{\mathbf{X}}) \cdot f(\widetilde{\mathbf{y}}_K |\, M_i,\, \mathbf{B}, \sigma^2, T_B, \widetilde{\mathbf{W}}_K,\, \mathbf{y})
\end{aligned}
\tag{5.8}
$$

Henceforth conditioning on $\widetilde{\mathbf{X}}$ and $\widetilde{\mathbf{W}}_K$ is omitted. As already noted, the major task here is to obtain the marginal predictive posterior distributions $f(y_{T+k} |\, M_i,\, \mathbf{y})$ for conducting k-step ahead forecasts, $k = 1...K$. Let for convenience be $\xi_k \equiv (\mathbf{B}',\, \sigma^2,\, T_B,\, \widetilde{\mathbf{y}}_K \setminus \{\, y_{T+k}\})' \in \Xi$, with $\Xi = \mathbb{R}^d \times \mathbb{R}^+ \times \mathbb{N}_{]\kappa+p;\, T-\kappa]} \times \mathbb{R}^{K-1}$. Then the posterior predictive density of the scalar random variable $y_{T+k},\ k = 1...K$, under M_i, is obtained by integrating out all other parameters ξ_k from the joint posterior (see Geweke and Whiteman (2006)):[4]

$$
f(y_{T+k} \mid M_i,\, \mathbf{y}) \;=\; \int_{\Xi} f(\xi_k |\, M_i,\, \mathbf{y}) \cdot f(y_{T+k} \mid \xi_k,\, M_i,\, \mathbf{y}) \cdot d\xi_k \tag{5.9}
$$

As a point estimator of the unknown future value $y_{T+k},\ k = 1...K$, the posterior predictive mean of (5.9), $E(y_{T+k} |\, M_i,\, \mathbf{y})$, is chosen. At this point it should be noted that the joint posterior predictive density of a periodic autoregressive model is not the standard K-variate Student-t density as in the case of a normal (nonrecursive) linear regression model, discussed in Zellner (1971), p.73, also Judge et al. (1985), p.123. This has been emphasized by Broemeling and Land (1984) for the case of a nonperiodic autoregressive model of order p. Moreover, the latter authors establish that for $K \geq 1$ the joint posterior predictive density is given by the product of K univariate t-densities, viz. a marginal for y_{T+1} and $K-1$ conditional t-densities. In appendix H it is shown that this fact also applies to the more general case of periodic autoregressive models of order p. Further the marginal posterior predictive densities (5.9) are not known analytically for $K > 1$ (see Broemeling and Land (1984), p.1319, Koop et al. (1995), Bauwens et al. (1999), p.138). This essentially follows from the fact that the joint posterior predictive density of $\widetilde{\mathbf{y}}_K$, after integrating out all other parameters, is not a multivariate Student-t density (see Broemel-

[4] In case of the discrete break date T_B the corresponding integration is in fact a summation.

ing and Land (1984), p.1319). Since allowing for a structural break at an unknown point in time introduces an additional nuisance parameter in the likelihood function, which can not be integrated out analytically, an MCMC approach in order to obtain these densities and also to get point estimates for the unknown parameters is presented in the following.

5.3. Markov chain Monte Carlo approach

Generating multistep forecasts by means of Markov chain Monte Carlo techniques allows to directly exploit the structure of the recursive prediction equations (5.7). In the following a Gibbs sampler for generating random draws from (5.8) and for obtaining the predictive densities in (5.9) is presented. Since the future values y_{T+k} are treated as latent variables (or missing data), the subsequent MCMC algorithm can be perceived as a data augmentation approach, see Tanner and Wong (1987). The basic sampling scheme is outlined in the following.

Algorithm 2 : Data augmentation

Step 1: *Set iterations on* $j = 1$ *and initialize* $T_B^{(0)}, \mathbf{B}^{(0)}, \sigma^{2(0)}, \widetilde{\mathbf{y}}_K^{(0)}$ *randomly or deterministically.*

Step 2: *Draw a new break date* $T_B^{(j)}$ *from a multinomial posterior distribution*

$f(T_B^{(j)} \mid \mathbf{B}^{(j-1)}, \sigma^{2(j-1)}, \widetilde{\mathbf{y}}_K^{(j-1)}, M_i, \mathbf{y})$ *on the sample space* $T_B \in]\kappa + p,\ T - \kappa]$

Step 3: *Draw a new random vector* $\mathbf{B}^{(j)}$ *from a multivariate normal posterior distribution* $f(\mathbf{B}^{(j)} \mid T_B^{(j)}, \sigma^{2(j-1)}, \widetilde{\mathbf{y}}_K^{(j-1)}, M_i, \mathbf{y})$

Step 4: *If stationarity should be imposed, accept a candidate draw* $\phi^{(j)}$ *in* $\mathbf{B}^{(j)}$, *if* $|\det(\Phi_0 - \Phi_1 \cdot z)| > \omega$ *for* $z = 1$, $\omega > 0$, *otherwise return to step 3*

Step 5: *Draw new* $\sigma^{2(j)}$ *from an inverse gamma posterior distribution*

$f(\sigma^{2(j)} \mid T_B^{(j)}, \mathbf{B}^{(j)}, \widetilde{\mathbf{y}}_K^{(j-1)}, M_i, \mathbf{y})$

Step 6: *Draw a new* ${y_{T+k}}^{(j)}$ *from a univariate normal posterior distribution*

$f(y_{T+k}^{(j)} \mid \sigma^{2(j)}, T_B^{(j)}, \mathbf{B}^{(j)}, \widetilde{\mathbf{y}}_K^{(j-1)} \setminus \{y_{T+k}\}, M_i, \mathbf{y})$, k=1...K

Step 7: *Compute the marginal posterior predictive densities*

$f(y_{T+k}^{(g)} \mid M_i, \mathbf{y}) \simeq \frac{1}{J} \sum_{j=1}^{J} f(y_{T+k}^{(g)} \mid \xi_k^{(j)}, M_i, \mathbf{y})$, $k = 1...K$, $g = 1...G$

Step 8: *Set* $j = j + 1$, *return to step 2.*

with $j = 1, ..., J$ Gibbs runs and $\xi_k^{(j)} \equiv (\mathbf{B}'^{(j)}, \sigma^{2(j)}, T_B{}^{(j)}, \widetilde{\mathbf{y}}_K^{(j)} \setminus \{y_{T+k}\})'$.

The sampling scheme shows how to draw from the joint posterior distribution (5.8). In step 1, initial values for the Markov chain have to be chosen, where for $\mathbf{B}$ and σ^2, the corresponding least squares estimates could be used; for the break date T_B an initial value can be drawn from the uniform prior distribution and the unknown future values, y_{T+k}, $k = 1...K$, could be initialized with the sample mean of $\mathbf{y}$. In step 2, generate a new break date from a full conditional multinomial posterior distribution $M(n =$

$T-p-2\kappa;\ \theta_{p+\kappa+1},...,\theta_{T-\kappa})$ with n classes and class probabilities given by

$$\theta_t = \frac{\exp\left\{-\frac{1}{2\sigma^2}(y_t - \widetilde{x}_t \cdot \mathbf{B})^2\right\}}{\sum_{t=p+1+\kappa}^{T-\kappa} \exp\left\{-\frac{1}{2\sigma^2}(y_t - \widetilde{x}_t \cdot \mathbf{B})^2\right\}} \quad , \text{for } t = p+\kappa+1,...,T-\kappa$$

where $\widetilde{x}_t$ and y_t denote the t-th row of the matrix $\widetilde{\mathbf{X}}$ and the t-th element of the vector $\mathbf{y}$, respectively.

Then in the next step, given a new break date T_B, a new candidate vector $\mathbf{B}$ is generated from a full conditional multivariate normal distribution with mean vector $\mu_B = \mathbf{R}^{-1}(\widetilde{\mathbf{X}}' \cdot \mathbf{y} + \mathbf{V}^{-1}\mathbf{B}_0 + \widetilde{\mathbf{W}}_K' \cdot \widetilde{\mathbf{y}}_K)$ and covariance matrix $\mathbf{\Sigma}_B = \sigma^2 \cdot \mathbf{R}^{-1}$, where $\mathbf{R} \equiv \mathbf{V}^{-1} + \widetilde{\mathbf{X}}'\widetilde{\mathbf{X}} + \widetilde{\mathbf{W}}_K'\widetilde{\mathbf{W}}_K$. It is important to note here that since the matrix $\widetilde{\mathbf{X}}$ and also some parts of the matrix $\widetilde{\mathbf{W}}_K$ depend on the unknown break point through their deterministic components, these matrices have to be updated accordingly after having drawn a new break date in sampling step 2.

Next in step 5, the innovation variance σ^2 is updated by drawing from a full conditional $IG_2(a^\star, b^\star)$ density, given all other parameters, with shape parameter $a^\star = T-p+d+K+2$ and scale parameter $b^\star = (\mathbf{y} - \widetilde{\mathbf{X}} \cdot \mathbf{B})' \cdot (\mathbf{y}_K - \widetilde{\mathbf{X}} \cdot \mathbf{B}) + (\widetilde{\mathbf{y}}_K - \widetilde{\mathbf{W}}_K \cdot \mathbf{B})' \cdot (\widetilde{\mathbf{y}} - \widetilde{\mathbf{W}}_K \cdot \mathbf{B}) + (\mathbf{B} - \mathbf{B_0})'\mathbf{V}^{-1}(\mathbf{B} - \mathbf{B_0})$. After having drawn a new value for the innovation variance, a new trajectory $\{y_{T+1},...,y_{T+K}\}$ is generated. Therefore, in step 6, new candidates for the unknown future values y_{T+k}, $k = 1...K$, are successively drawn from the K full conditional univariate normal distributions, with mean $\widetilde{\mu}_k = W_{(k)} \cdot \mathbf{B}$, and variance σ^2 given from step 5, where $W_{(k)}$ denotes the k-th row of the matrix $\widetilde{\mathbf{W}}_K$.

In order to facilitate the integrations in (5.9) additional Monte Carlo integration steps, for $k = 1...K$, using the Gibbs sampling draws $\xi_k^{(j)},\ j = 1...J$, are conducted. For this purpose first define a grid of $\{y_{T+k}^{(g)},\ g = 1...G\}$ values over the support of y_{T+k}, for $k = 1...K$. Then the posterior ordinates can be approximated as in step 7 above, see Chib (1995). Note that the marginalized posterior ordinates are needed for the subsequent model averaging.

Next consider a finite set of candidate models $\mathscr{M} = \{M_1, \ldots, M_I\}$. As already noted, each candidate model M_i is associated with a vector of model indicators (p,m). Attaching prior probability mass, $f(M_i)$, to each candidate model we can average over the model space

as (see Hong and Preston (2012)):

$$f(y_{T+k}|\,\mathbf{y}) \;=\; \sum_{i=1}^{I} f(y_{T+k}\mid M_i,\,\mathbf{y})\cdot f(M_i|\,\mathbf{y}) \quad ,\, k=1...K \tag{5.10}$$

with (5.10) being approximated on the same grid of $y_{T+k}^{(g)}$ values, $g=1...G$, as used in sampling algorithm 2.

Further if one wants to impose stationarity on the PAR(p) model in (5.2), the posterior draws of the subvector ϕ in $\mathbf{B}$ can be constrained by using an additional accept-reject step in step 4. Note, that although PAR models are nonstationary by construction,[5] a stationarity condition can be stated by using a multivariate model representation as in Tiao and Grupe (1980). By writing the univariate PAR(p) model (5.2) as an S-dimensional vector autoregressive model of order P ('VAR(P)'), with $P=1+\lfloor(p-1)/S\rfloor$, the usual stability condition of VAR models in terms of the characteristic polynomial with roots z, can be used (see section 3.2, also Hamilton (1994), p.259, for details). By restricting the lag order to $p_{max}=S$ in the following, the stationarity condition of a VAR(1) model is given by

$$\det(\Phi_0-\Phi_1\cdot z^S)\neq 0 \quad \text{for } |z|\leq 1 \tag{5.11}$$

see section 3.2 for the definition of the Φ matrices.

Since for a unit root process, i.e. with $z=1$, the determinant in step 4 will be close to zero in practice, an arbitrarily small value $\omega>0$, e.g. $\omega=0.05$ can be chosen here as a benchmark value.

The major focus of the presented forecasting approach lies on the model averaged predictive distributions in (5.10). It can be shown that the means of these mixture distributions minimize the prediction mean squared error and are in this sense optimal, see Raftery and Zheng (2003), Theorem 4, p.5. In order to compute the BMA predictive distributions in (5.10) the posterior probability of model M_i, $i=1...I$, is needed, which can be obtained

[5] This can easily be checked from the first two moments of a PAR process, see Ghysels and Osborn (2001), Franses and Paap (2006) for details.

from Bayes' Theorem:

$$f(M_i|\,\mathbf{y}) \;=\; f(\mathbf{y}|\,M_i)\cdot f(M_i)/f(\mathbf{y}) \tag{5.12}$$

In case of a structural change model, i.e. with $M_i = (p_i,\ m_i = 1)$, the first expression in (5.12) is the marginal likelihood under model M_i:

$$f(\mathbf{y}|\,M_i) \;=\; \sum_{T_B=p+\kappa+1}^{T-\kappa} \int \ldots \int f(\mathbf{y}|\,\mathbf{B},\sigma^2,T_B,\,M_i)\cdot f(\mathbf{B},\,\sigma^2,T_B|\,M_i)\cdot d\mathbf{B}\cdot d\sigma^2 \tag{5.13}$$

and the normalizing constant in (5.12) is equal to

$$f(\mathbf{y}) \;=\; \sum_{i=1}^{I} f(M_i)\cdot f(\mathbf{y}|\,M_i) \tag{5.14}$$

Note that all $d+1$ integrals in (5.13) can be solved analytically due to the conjugate framework. In case of a model without a structural break, i.e. with $M_i = (p_i,\ m_i = 0)$, the summation over all possible change points in (5.13) can of course be omitted and the marginal likelihood, under model M_i, then has the well-known form of a multivariate Student-t density of dimension $T-p$, see Judge et al. (1985), p.128, also Hamilton (1994), p.368.

5.4. Monte Carlo analysis

Next the results of four Monte Carlo (MC) experiments are presented in order to compare the forecasting performance of a conditional BPAR model with other candidate forecasting models. In order to save computing time the computations are conducted for a specific model M_i and are in this sense conditional. To infer if the reported differences in predictive accuracy are also statistically significant, a simple Bayesian test is presented. The simulated out-of-sample forecasts of the BPAR model are compared to a nonseasonal AR model, a seasonal AR (SAR) model (see Box et al. (2008), Ghysels and Osborn (2001))

and a seasonal (or periodic) means model (abbreviated by 'PMEANS' in the following).[6] A nonseasonal AR model is chosen, because PAR models can be perceived as seasonally varying generalizations of AR models. In this sense an AR model serves as a benchmark with constant parameters. Similarly, a PMEANS model serves as a benchmark for the case of deterministic seasonality in the following, cf. Osborn and Smith (1989).

The comparison of forecasting models is usually done on the basis of a risk function (see Diebold and Mariano (1995), Inoue and Kilian (2006), also Geweke and Whiteman (2006), p.20). The most prominent measures of forecasting accuracy in this context are the prediction mean squared error (PMSE), assuming a quadratic loss structure, and the mean absolute percentage error (MAPE),[7] assuming an absolute valued loss function, see Meese and Rogoff (1983), Stock and Watson (1999), among others, for applications. As a Bayesian forecasting rule the posterior predictive expectation, $\widehat{y}_{T+k} \equiv E(y_{T+k}|\, M_i,\, \mathbf{y})$, is used in the following. When expressed as a function of the forecasting horizon the (cumulated) PMSE and MAPE of a simulated out-of-sample K-step ahead forecast are given by

$$PMSE(K) = \sum_{k=1}^{K} E\left(\widehat{y}_{t+k} - y_{T+k}\right)^2 \tag{5.15a}$$

$$MAPE(K) = \sum_{k=1}^{K} E\left(\left|\left(\widehat{y}_{t+k} - y_{T+k}\right)/y_{T+k}\right|\right) \tag{5.15b}$$

Here y_{T+k} is the actual realization of the process at time $T+k$, the k-step ahead forecast is denoted by $\widehat{y}_{T+k}$, and the expectation is taken with respect to $\mathbf{y}$.

5.4.1. A Bayesian sign test for comparing predictive accuracy

To be able to test if the realized loss differences measured by a loss function $g(y_t, \widehat{y}_t)$, e.g. a quadratic loss function $g(y_t, \widehat{y}_t) = (\widehat{y}_t - y_t)^2$, are also statistically significant some authors propose tests to evaluate predictive accuracy, cf. Diebold and Mariano (1995), Giacomini and White (2006), Clark and West (2007). To be consistent here, a Bayesian

[6] Here the variable of interest is simply regressed on a set of S dummy variables $D_{s,t}$, which equal one if observation t is associated with season s.

[7] Note that the MAPE for a specific horizon k does not depend on the scale or dimension.

analogue to the sign test proposed in Diebold and Mariano (1995) is introduced. For this test the latter authors report favorable power and size properties on the basis of MC experiments. Furthermore this test is applicable not only in case of quadratic loss structures, but also to non-quadratic and relative loss functions as assumed for the calculation of the MAPE in (5.15) and thus matches the requirements here. Given a realized sequence of loss-differentials, $d_t \equiv [g(\widehat{\varepsilon}_{it}) - g(\widehat{\varepsilon}_{jt})]$, under forecasting model $i, j = 1...3$, with prediction error $\widehat{\varepsilon}_{it} = y_t - \widehat{y}_{it}$ under model i, a test for comparing predictive accuracy can be constructed on the signs of a sequence of loss-differentials $\{d_t\}_{t=1}^{T}$, with $\mathbf{1}_{(d_t>0)}(d_t) = 1$ in case of a positive sign, and 0 otherwise. Therefore the null hypothesis of a zero median loss-differential, i.e.

$$H_0 : \ Med(g(\widehat{\varepsilon}_{it}) - g(\widehat{\varepsilon}_{jt})) = 0 \tag{5.16}$$

can be tested, which implies $P(g(\widehat{\varepsilon}_{it}) > g(\widehat{\varepsilon}_{jt})) = P(g(\widehat{\varepsilon}_{it}) < g(\widehat{\varepsilon}_{jt}))$ or equivalently $P(\mathbf{1}_{(d_t>0)}(d_t) = 1) = 0.5$, see Diebold and Mariano (1995).[8]

A Bayesian analogue can be constructed by assuming that each observation $\mathbf{1}_{(d_t>0)}(d_t)$, $t = 1...T$, is generated according to a Bernoulli distribution with unknown ('success') probability for a positive loss-differential sign, $\pi \equiv P(\mathbf{1}_{(d_t>0)}(d_t) = 1)$, and that π has a Beta prior density, $\pi \sim Be(\alpha,\ \beta)$, with prior mean equal to $\alpha/(\alpha+\beta)$. Let $x \equiv \sum_{t=1}^{T} \mathbf{1}_{(d_t>0)}(d_t)$ denote the number of positive signs in the sequence,[9] which is a sufficient statistic for π, therefore conditioning on x instead of the data does not change the validity of the subsequent probability statements, see Monahan and Boos (1992), p.272. By conjugation it immediately follows that the posterior density of π, that is the probability for a positive loss-differential sign, also behaves according to a Beta distribution, i.e. $\pi|x \sim Be(\alpha+x;\ \beta+T-x)$, with posterior mean equal to $(\alpha+x)/(\alpha+\beta+T)$. Of course the null could also be tested against more informative alternatives like $H_1 : \pi > 0.5$, which postulates the dominance of model j over model i in $d_t \equiv [g(\widehat{\varepsilon}_{it}) - g(\widehat{\varepsilon}_{jt})]$, and vice versa.[10]

[8] Note that the sign test presumes i.i.d. observations, an assumption that needs to be checked in practice.

[9] This is the S_2-test statistic of Diebold and Mariano (1995), p.255, which follows a Binomial distribution with parameters T and $\pi = 0.5$ under the null hypothesis.

[10] A more elaborated Bayesian approach for the analysis of regression errors is presented by Zellner (1975) (see also Chaloner and Brant (1988) for a similar approach), which has been adopted by Lubrano (2001) in order to test for ARCH(1) effects and non-linearities in time series.

Note that in the subsequent computations of the multistep ahead forecasts, the considered prediction horizons are at most two years. Thus we have to deal with quite short loss-differential sequences.[11] For this reason the hyperparameters of the Beta prior have to be chosen with some care, see Robert (2007), p.124. Therefore four different parameterizations of the Beta prior are stated next, which will be used throughout the following simulation study in order to check the robustness of the posterior results with respect to the prior specification.

If prior ignorance with respect to the unknown probability π should be expressed, noninformative priors can be chosen. Here, for example, Haldane's prior $f(\pi) \propto [\pi \cdot (1-\pi)]^{-1}$ can be used, which appears to be the limit of an unnormalized Beta prior when α and β go to zero.[12] The latter choice then leads to a $Be(x,\ T-x)$ posterior distribution with mean x/T, which is also the maximum likelihood estimator of π. Another popular noninformative prior is the Jeffreys prior $f(\pi) \propto \sqrt{I(\pi)}$, with $I(\pi)$ the Fisher information, which in the present case can be shown to be $f(\pi) \propto [\pi \cdot (1-\pi)]^{-1/2}$, i.e. proportional to a $Be(0.5, 0.5)$ density. The third prior, used in the following, is the 'Bayes' or 'Bayes-Laplace' prior, $f(\pi) = Be(1,1)$, which assigns a density of one to each value of π, cf. Berger (1980), p.89, Robert (2007), p.127. These priors are shown in figure F.29 of appendix F. As can be observed from figure F.29, Jeffreys' and Haldane's prior both assign the largest density values to the end points zero and one. However, since we are interested in testing $H_0 : \pi = 0.5$, this could lead to an increase in the type I or type II error of the test and therefore in addition a more informative $Be(1.01, 1.01)$ prior is used (see the lower right panel in figure F.29), which gives more weight to moderate π values.[13] Note that the null restriction is imposed a priori by choosing $\alpha = \beta$, which implies a prior mean of π equal to 0.5. Now an optimal Bayesian decision rule $\varphi(x)$ under a '0-1' loss function (see Robert (2007), proposition 5.2.2, p.225.) is given by

$$\varphi(x) = \begin{cases} 0\,, & \text{if } P(H_0|x) \geq 0.5 \\ 1\,, & \text{otherwise} \end{cases} \tag{5.17}$$

[11] For example, in the MC experiments presented below, 2-years ahead forecasts using quarterly data are conducted and thus $T = 8$, whereas in the empirical application of section 5.5, 1-year ahead forecasts using monthly data are considered and thus the length of the realized loss-differential sequences is $T = 12$.

[12] Here for all computations $\alpha = \beta = 10^{-10}$ is used.

[13] Note that in general according to the Bernstein von Mises theorem the posterior distribution for an unknown quantity is effectively independent of the prior distribution once the amount of information supplied by the data is large enough, see van der Vaart (1998), p.140.

where 1 means rejection of the null hypothesis and $P(H_0|x)$ denotes the posterior probability of the latter.

Since the hypothesis of interest is a precise hypothesis, i.e. with zero prior probability mass, the continuous Beta prior has to be modified in order to assign positive prior probability mass to the singleton under H_0 (see Berger and Sellke (1987), Berger and Delampady (1987), also Pereira and Stern (1999) for a different approach). Therefore the following mixed prior is utilized

$$f(\pi) = \omega_0 \cdot \mathbf{1}_{H_0}(\pi) + (1-\omega_0) \cdot f_1(\pi) \,, \tag{5.18}$$

with $\omega_0 > 0$, $\mathbf{1}_{H_0}(\pi)$ an indicator function and $f_1(\pi)$ one of the priors discussed above.[14]

By an application of Bayes' Theorem using the mixed prior in (5.18) the posterior probability of the null hypothesis can be calculated as

$$P(H_0|x) = \frac{f(x|\pi_0) \cdot \omega_0}{\int_0^1 f(x|\pi) f(\pi) d\pi} = \frac{f(x|\pi_0) \cdot \omega_0}{\omega_0 \cdot f(x|\pi_0) + (1-\pi_0) \cdot m_1(x)} \tag{5.19a}$$

$$\text{with } \; m_1(x) = \int_{\{\pi:\pi \neq 0.5\}} f(x|\pi) \cdot f_1(\pi) \cdot d\pi \tag{5.19b}$$

In order to check the performance of the Bayes test in (5.17) the posterior probability $P(H_1|x) = 1 - P(H_0|x)$ is computed for each of the four Beta priors. To examine further the influence of the sample size T on the subsequent posterior results, $P(H_1|x)$ is calculated for a sequence of $x \in \{0, ..., T\}$ values, assuming $T = 8$ and $T = 60$.

For a given sample size T the posterior probability $P(H_1|x)$ can be represented as a function in x, or x/T, i.e. $p(x_0) \equiv P(H_1|x = x_0; T)$, for $x_0 = 0, ..., T$. Under $H_0 : \pi_0 = 0.5$ we would expect that, for given T, the posterior probability of H_1 would decrease when $\frac{x}{T} \to \pi_0$ and increase otherwise. The results shown in figures F.30 and F.31 (see appendix F) confirm that this is the case for most of the considered priors. The figures suggest that, with the exception of the results under Haldane's prior, all specifications yield quite comparable results. Moreover even for sequences of length $T = 60$ the posterior probabilities of H_1 using Haldane's prior are still equal to 0 within a large subset of the domain of $p(x)$, which would lead us to favor the null even for extreme, i.e. very small (or large) numbers

[14] Here $\omega_0 = 0.5$ is chosen.

of positive signs, x, implying a large type II error. Therefore, in the applications below all conclusions are drawn on the basis of one of the other three posteriors.

As pointed out by Diebold and Mariano (1995) comparing the predictive accuracy of different models constitutes a multiple hypotheses testing problem. In the frequentist case the overall size of the test, $\alpha^\star$, has to be adjusted by a Bonferonni correction, $\alpha^\star/n$, where n denotes the number of tests. However the Bayesian sign test has not been derived under the assumption of a fixed type I error, as in a Neyman-Pearson testing framework, hence another strategy is pursued in the following. Note that in the subsequent analysis interest mainly centers on testing joint hypotheses of the form

$$H_0 : \pi_{1,2} = \pi_{1,3} = 0.5 \tag{5.20}$$

given that model 1 has exhibited the lowest relative loss among, for example, three competing models.

Here $\pi_{i,j}$ denotes the parameter of interest in a pairwise comparison of the models i and j, respectively. Hence first the model with the lowest PMSE is detected ('Model 1') and then (5.20) is tested accordingly. This a union-intersection testing problem (see Casella and Berger (2002), p.380) and therefore (5.20) is rejected, if any of the single hypotheses $H_{0,\gamma} : \pi_\gamma = 0.5$, with $\gamma \in \{(1,2),(1,3)\}$, is rejected using the decision rule (5.17). This testing strategy is pursued to evaluate the results of the following MC experiments and also the empirical results of the next section.

5.4.2. Simulation evidence

For the subsequent simulation study four Monte Carlo experiments using different data generating processes (DGP) are conducted. Here the predictive performances of seasonal models in 2-years ahead forecasts are examined using quarterly data. In order to reduce computing times, the BPAR predictions are conducted using the conditional posterior predictive densities, i.e. given values for the autoregressive lag order p and the number of breaks m, instead of the model averaged predictive densities. Further, because the posterior expectation is used as a Bayesian forecasting rule, I focus on the PMSE in the analysis. The corresponding MAPE values are reported for reasons of comparison.

In the first two experiments the predictive accuracies of the considered models are examined, when the DGP is a stationary PAR(1) and a periodically integrated AR(1) (abbreviated by 'PIAR(1)') process, respectively. Here the forecasting accuracy of a PAR(1) model is compared with those of an AR(1) model, both having seasonally varying intercepts, and with those of a deterministic PMEANS model. In the third and fourth simulation design the forecasting performances of a SAR(1) and a SARMA$(1,0) \times (1,1)$ process both with seasonally varying intercepts are analyzed, given that the data are generated by a nonperiodic stochastic process. In the latter two experiments the forecasting models are: a PAR(4) model, a quarterly SAR(1) and a SARMA$(1,0) \times (1,1)$ model, respectively, all three having periodically varying intercepts, and finally a PMEANS model. Note, that all considered PAR and (S)ARMA models encompass the PMEANS model.

In the first simulation experiment ('design 1') trajectories are generated from a stationary, quarterly PAR(1) process without a break:

$$y_t = \sum_{s=1}^{4} D_{s,t} \cdot (\mu_s + \phi_s \cdot y_{t-1} + u_t) \ , \ u_t \overset{i.i.d.}{\sim} N(0,1) \ , \ t = 1...T \tag{5.21}$$

with $\phi = (0.85,\ 0.67,\ 0.92,\ 1.1)'$ and $\mu = (0.85,\ 0.95,\ 0.92,\ 1)'$.

For the second experiment ('design 2') the same parametrization as in design 1 is used, but a periodic unit root is imposed by setting $\phi_4 = 1/(\phi_1 \cdot \phi_2 \cdot \phi_3)$, see Osborn et al. (1988), Boswijk et al. (1995), Franses and Paap (2006) for details on periodic integration. Note, that for ϕ_1, ϕ_2 and ϕ_3 values close to one, which is often the case in practice, this kind of integration coincides with a (nonseasonal) unit root at the zero spectral frequency.

For all simulations $T = 800$ draws from the respective processes are generated, discarding the first 500 draws due to burn-in and then the remaining $T = 300$ draws are used to conduct k-step ahead forecasts for $k = 1,...,8$.[15] This procedure is repeated $N = 100$ times in order to approximate the expectations in (5.15). For each of the 100 trajectories the above sampling algorithm is run 5500 times, discarding the first 500 draws, and then the respective posterior predictive means of y_{T+k}, $k = 1...K$ are calculated.[16] For the classical forecasts of the (S)AR model and the PMEANS model the Kalman filter and the

[15] All initial values are chosen to be fixed and equal to zero.

[16] The MC integration steps to obtain the marginal posterior predictive distributions of the y_{T+k}, $k = 1...8$, are conducted on a grid of 100 points.

least squares projections are used, respectively.

In figure 5.1 the (cumulated) PMSEs for each of the three models are shown for design 1 (left panel) and design 2 (right panel). As can be recognized from the left panel

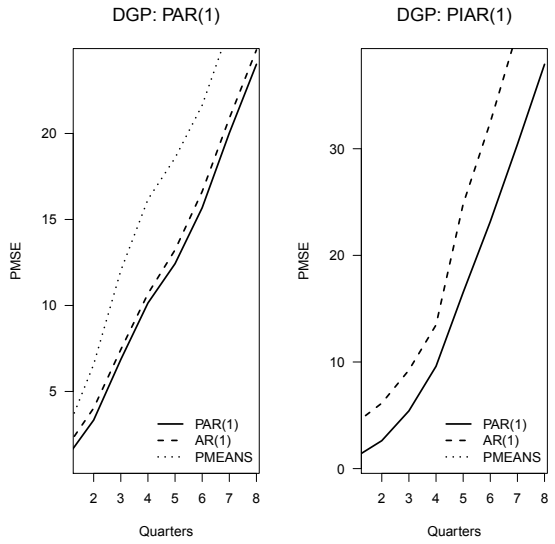


Figure 5.1.: (Cumulated) PMSEs for 2-years ahead forecasts (Design 1 and 2).

the PAR(1) model slightly dominates the AR(1) model and clearly beats the PMEANS model for each considered forecasting period. The associated average PMSEs (with average MAPEs in parentheses) of the three models are 1.73 (0.20) for the PAR(1) model, 1.76 (0.21) for the AR(1) model and 1.94 (0.23) for the PMEANS model. To check if the median loss-differentials are significantly different from zero the Bayesian sign test is applied using the prior distributions discussed in section 5.4.1. Further the classical sign test and the classical Wilcoxon-Signed Rank (WSR) test as proposed in Diebold and Mariano (1995), are applied here for reasons of comparison with the Bayesian testing results.[17] Table 5.1 shows that the median loss-differentials of the PAR-AR and AR-PMEANS comparisons are not significantly different from zero according to the Bayesian sign tests and also according to the classical tests shown in the last two columns of table 5.1. For the latter two tests a 10% nominal level of significance is assumed due to the short loss-sequences ($T = 12$).[18] In contrast, the PAR-PMEANS differentials are always

[17] Where it is shown in Diebold and Mariano (1995) that the latter test has slightly more power in small samples than the former.

[18] For each loss differential series a Runs test for randomness is conducted, where rejection of the null of

significant. However the classical sign test indicates the weakest evidence against the null, which could be a consequence of its lack of power as reported by Diebold and Mariano (1995). Looking at the results of the Bayes tests we recognize a similar behavior of the test in terms of power as described in section 5.4.1, i.e. that the results under the three prior distributions are quite comparable if not identical.[19]

Further the joint hypothesis $H_0 : \pi_{1,2} = \pi_{1,3} = 0.5$ is tested, i.e. a multiple comparison of models is conducted, see Berry and Hochberg (1999). Following the strategy outlined in section 5.4.1, the overall null is rejected if at least one of the single tests rejects the null. According to the results of simulation experiment 1 it can be concluded that in case of a stationary PAR(1) process, the three models differ significantly in terms of forecasting accuracy. However there is no significant deterioration in accuracy when using a more parsimonious AR model instead of a PAR model.[20]

Table 5.1.: Test results - Design 1 / DGP: PAR(1)

Comparisons	Jeffreys'	Bayes'	$Be(1.01, 1.01)$	Sign test (pv.)	WSR test (pv.)
1-2	0.7852	0.7110	0.7110	1.0000	0.3125
1-3	0.2298	0.2195	0.2195	0.0703	0.0546
2-3	0.5638	0.4961	0.4961	0.2891	0.1094

'1': PAR(1), '2': AR(1), '3': PMEANS, 'pv.': p-value

Next we turn our attention to the results of simulation experiment 2 in which the forecasting performances in case of nonstationary PAR(1) data are analyzed. From the right panel of figure 5.1 it can be recognized that in case of PIAR(1) data, the PAR(1) model clearly dominates the other two competitors. Note that the PMEANS model can not be seen in figure 5.1 due to very large PMSE values. The associated average PMSEs (and MAPEs) of the three models in this case are: 2.18 (0.42) for the PAR(1) model, 2.47 (0.42) for the AR(1) model and 9.58 (1.48) for the PMEANS model.

In table 5.2 the test results for comparing predictive accuracy are summarized. All loss-

'randomness' would be problematic with regard to the iid-assumption of the used tests, see Diebold and Mariano (1995). Here no further evidence for nonrandomness of the sequences has been found.

[19] Under Haldane's prior all posterior probabilities are equal to one.

[20] Similar results have been obtained for other parameterizations of the DGP in (5.21) and also for a periodic moving average process of order one as DGP. The average PMSEs in the latter case are 1.22 for a PAR(1) model, 1.23 for an AR(1) model and 1.26 for a PMEANS model.

differentials of the pairwise comparisons are significant.[21] Hence the results suggest, that in contrast to the results of design 1, there is a significant deterioration in forecasting accuracy when incorrectly using a nonperiodic AR model or a deterministic PMEANS model in the case of periodic nonstationarity.

Table 5.2.: Test results - Design 2 / DGP: PIAR(1)

Comparisons	Jeffreys'	Bayes'	$Be(1.01, 1.01)$	Sign test (pv.)	WSR test (pv.)
1-2	0.0195	0.0339	0.0342	0.0078	0.0078
1-3	0.0195	0.0339	0.0342	0.0078	0.0078
2-3	0.0195	0.0339	0.0342	0.0078	0.0078

'1': PAR(1), '2': AR(1), '3': PMEANS, 'pv.': p-value

In simulation design 3 trajectories are generated according to a quarterly SAR(1) process in order to simulate also annual lag structures. Finally, with regard to the empirical analysis of the next section also more general dynamics, given by a SARMA$(1,0) \times (1,1)$ process, are simulated in design 4. In design 3 and 4 the nonperiodic SAR(MA) models are identically specified as the respective DGPs and compared with a PAR(4) model and a PMEANS model. The stochastic process used in design 4 is parameterized as follows:

$$y_t = \sum_{s=1}^{4} \mu_s D_{s,t} + \phi_1 y_{t-1} + \phi_4 y_{t-4} + u_t - \theta_4 u_{t-4} \ , \ u_t \overset{i.i.d.}{\sim} N(0,1) \tag{5.22}$$

with $\phi_1 = 0.35$, $\phi_4 = 0.45$, $\theta_4 = 0.35$ and $\mu_s = 0.5$, $\forall s$, where for simulation design 3 the same specification is chosen except that ϕ_1 and θ_4 are set equal to zero.

For the simulation experiments 3 and 4 the corresponding PMSEs are shown in figure 5.2. From there it can be observed that in case of quarterly data, generated by a constant-parameter SAR(1) process, a PAR(4) predicts the data almost equally well as the (true) SAR(1) model. The PMSEs (MAPEs) of experiment 3 are 1.22 (0.81) for the PAR(4) model, 1.21 (0.83) for the SAR(1) model, and 1.35 (0.93) for the PMEANS model. Table 5.3 summarizes the test results. The test findings are similar to those of design 1, namely that there is no significant deterioration in predictive accuracy when estimating a PAR(4) model instead of the (correct) quarterly SAR(1) model. However estimating a purely

[21] The corresponding results under Haldane's prior are 0.0077 for all three comparisons.

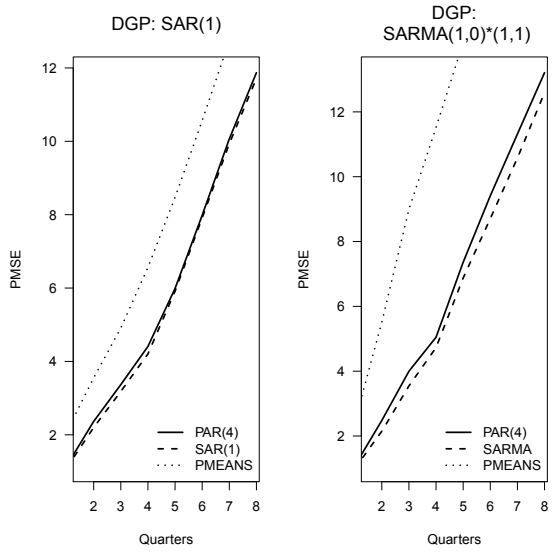


Figure 5.2.: (Cumulated) PMSEs for 2-years ahead forecasts (Design 3 and 4).

Table 5.3.: Test results - Design 3 / DGP: SAR(1)

Comparisons	Jeffreys'	Bayes'	$Be(1.01, 1.01)$	Sign test (pv.)	WSR test (pv.)
1-2	0.5638	0.4961	0.4953	0.2890	0.1953
1-3	0.0195	0.0339	0.0343	0.0078	0.0078
2-3	0.0195	0.0339	0.0343	0.0078	0.0078

'1': PAR(4), '2': SAR(1), '3': PMEANS, 'pv.': p-value

deterministic model and thus neglecting the stochastic structure of the data results in a significant loss in accuracy, which would lead to a rejection of the overall null of predictive concordance of the three models. Finally, from the right panel of figure 5.2, it can be recognized that if the data are generated by a quarterly SARMA$(1,0) \times (1,1)$ process, a PAR(4) model produces slightly higher mean squared losses than the true SARMA model. However the median loss differences between the PAR and the SARMA model are not significant. The corresponding average PMSEs (MAPEs) of this experiment are 1.29 (2.24) for the PAR model, 1.26 (2.09) for the SARMA model and 1.64 (3.25) for the PMEANS model.

In sum, both a PAR model and a constant-parameter SAR(MA) model perform similarly in terms of squared error loss. Both clearly outperform a deterministic PMEANS model

Table 5.4.: Test results - Design 4 / DGP: SARMA$(1,0) \times (1,1)$

Comparisons	Jeffreys'	Bayes'	$Be(1.01, 1.01)$	Sign test (pv.)	WSR test (pv.)
1-2	0.5638	0.4961	0.4953	0.2890	0.2500
1-3	0.0195	0.0339	0.0343	0.0078	0.0078
2-3	0.0195	0.0339	0.0343	0.0078	0.0078

'1': PAR(4), '2': SARMA$(1,0) \times (1,1)$, '3': PMEANS, 'pv.': p-value

when the data exhibit stochastic seasonality. In the context of periodically integrated data a PAR model provides significantly more accurate forecasts than its competitors. So far, this suggests the usefulness of the proposed periodic forecasting model even in the case of nonperiodic data.

In the next section the model averaged PAR (BMA-PAR) model with a possible break is used to predict monthly unadjusted unemployment rates of Germany.

5.5. Forecasting German monthly unemployment data

Next the simulated out-of-sample forecasts using the three candidate models of the last section are conducted to evaluate the accuracy in predicting monthly unemployment rates of the 16 German federal states and of East- and West-Germany. Furthermore the predictive performance of a BPAR model, using the model averaged predictive distribution in (5.10) for prediction, is compared to a BPAR model using the conditional predictive distribution (5.9), where M_i is obtained through a model selection step. Finally the BMA-PAR model is used to conduct 12-months ahead out-of-sample forecasts of these series. In brief, for the German monthly unemployment rates the empirical results show that the BPAR model outperforms a SARMA and a PMEANS model in terms of squared and absolute error loss. Furthermore, the subsequent findings suggest superiority of the BMA approach over a (conditional) model selection approach and are thus in accordance with the theoretical results in the literature, cf. Raftery and Zheng (2003).

The data set consists of monthly unadjusted unemployment rates of the 16 federal states

of Germany as well as the aggregated series for West and East Germany for the sample period 01/1991 to 02/2013. As in the simulation exercises of the last section the Bayesian forecasting results are compared to a seasonal AR model and seasonal (or periodic) means model. The untransformed series together with their sample (partial) autocorrelation functions and spectral densities are depicted in figures F.32-F.49 (see appendix F). From there it can be observed that most of the series exhibit strong autocorrelation, which for some series declines periodically, or shows up seasonally varying swings. Further, when looking at the periodograms in the lower right panels, it can be recognized that for most series a great amount of variation can be attributed to the zero spectral frequency, which indicates the existence of a nonseasonal stochastic trend component. Also for many of the series we observe clear peaks in the spectral density around monthly frequencies, which in addition suggests the existence of a seasonal component. Due to these results and the results of unreported unit root tests, the original series are transformed by applying first differences. In order to get a clearer picture of the seasonal variation in the data, the monthly boxplots for each country given in figures F.50-F.58 are examined (see appendix F). From a visual inspection of the boxplots it can be concluded that most of the conditional, i.e. monthly, sampling distributions are relatively homogenous, which would advocate the use of constant-parameter models. As counterexamples the series of Lower Saxony and of Bavaria can be considered.

In order to check the data for the presence of periodic forms of serial dependency of order one, the Bayesian *forward* and *backward* recursive F-tests for no periodicity, introduced in section 4.5, are applied. For an ease of reference these procedures are restated in the following.[22] First estimate an unrestricted PAR(1) model with seasonal intercepts, using the respective posterior means[23] as point estimates and then use these estimates to test the null of no periodicity, i.e. $H_0 : \phi_s = \phi, \forall s$, against the alternative of periodicity $H_1 : \phi_s \neq \phi, \exists s$. Note that the null hypothesis implies a nonperiodic AR(1) model, whereas under the alternative a PAR(1) model is assumed. The null can equivalently be tested by considering $H_0 : \mathbf{R} \cdot \phi = \mathbf{0}$, with $\mathbf{R} = [\mathbf{I}_{S-1},\ -\iota]$ an $(S-1) \times S$ matrix of linear contrasts, $\mathbf{I}_{S-1}$ the identity matrix and ι an $(S-1)$-vector of ones, where $d = dim(\mathbf{B})$. Given normality of the innovations and the prior assumptions stated in section 5.2.1, a Bayesian test can be conducted based on an $F(\nu_1 = S-1,\ \nu_2 = T-p-d)$ posterior distribution,

[22] See also Boswijk and Franses (1996) for a similar strategy within a classical framework.
[23] These can be obtained analytically.

see appendix G.5 for details. To check the robustness of the results a recursive testing procedure is adopted, using the Bayesian F-test, denoted by F_{PAR}, for rolling subsamples of the original data. The F-test when one year of data is added and removed successively is then called a Bayesian *forward* and *backward* recursive F_{PAR} test, respectively. Note that for the forward recursive test the first time window reaches from 1/1991 to 12/1994 and the final window from 1/1991 to 2/2013, i.e. the whole time period. In contrast, for the backward recursive test the first time window is from 1/1991 to 2/2013, where the last window includes the period from 1/2009 to 2/2013. The corresponding test results for each region are shown in table C.1. The null of no periodicity is rejected in the majority of cases. In contrast to the descriptive results from the boxplots, these findings suggest the appropriateness of a periodic model for the prediction of German unemployment data.

In the MCMC algorithm used to generate the forecasts of the model averaged BPAR(p) model the admissible parameter region of the model indicators are restricted to $m_{max} = 1$ and $p_{max} = 12$, where the latter is chosen in order to capture also annual dynamics. Analogously to the BPAR models, one structural break in the deterministic components of the SARMA and the periodic means model is allowed. The identification of the break date is accomplished by computing the Bayesian information criterion (BIC) (see Schwarz (1978)) for all possible break points omitting the first and last ten percent of the sample in order to preclude possible end point problems, and then the minimizing date is used as a change point.

What becomes evident from a first visual inspection of the original series (see figures F.32-F.49) is an instantaneous increase in the level of most of the series around the year 2005. This shift is a consequence of the so called 'Hartz IV' labor market reform, which took effect by January 2005.[24] A fixed step-dummy variable in order to control for this event is not used here, because the considered models only allow for one structural break and it could be possible, and in fact this has been the case for some series, that there is another date that is associated with a lower information criterion. However for the majority of the federal states the date 01/2005 minimizes the BIC and hence is taken as the final break date. The order specification of the SARMA$(p,\ q) \times (P,\ Q)_{12}$ model

[24]This reform brought together the former unemployment benefits for long term unemployed ('Arbeitslosenhilfe') and the former welfare benefits ('Sozialhilfe'). That is, since January 2005 these two groups have both been considered as 'unemployed'. This simple change in 'measurement' of the unemployment rate induced the instantaneous level shift for most of the series.

is conducted by computing the BIC for all possible model combinations $(p,\ q,\ P,\ Q)$, where $p_{max} = q_{max} = P_{max} = Q_{max} = 2$ is chosen in order to obtain more parsimonious specifications. For the SARMA model specifications the usual iterative three-stage strategy: identification, estimation, diagnostic checking is employed. Estimation is done by conditional maximum likelihood and the multistep forecasts are conducted by using the Kalman filter.

In table C.2 and C.3 (see appendix C) the average PMSEs and MAPEs of the 1-year ahead in-sample predictions are reported. For the computations the last 12 observations are saved for comparison and the preceding observations are then used to calculate the forecasts for each model. The results support the PAR model in terms of forecasting accuracy. More precisely, for 9 (10) out of the 18 series the BMA-PAR model has the lowest average loss in terms of the PMSE (MAPE). In only 4 (4) out of the 18 series the more parsimonious SARMA model is superior, whereas the deterministic periodic means model in only 4 (3) cases provides more accurate forecasts with respect to the PMSE (MAPE) than the other forecasting models. In order to check if the loss-differentials are significantly different from each other, the Bayesian sign test of section 5.4.1 with a $Be(1.01, 1.01)$ prior is used. Here for each of the 18 series pairwise comparisons using the Bayesian sign test are conducted. Given model 1 has exhibited the lowest average loss, the overall null $H_0 : \pi_{1,2} = \pi_{1,3} = \pi_{1,4} = 0.5$ is rejected, if at least one of the three pairwise comparisons is rejected. In advance a Runs test for randomness and also a Box-Ljung test for serial correlation of order one are applied. Here for most of the considered loss-differential sequences the iid-assumption of the sign test is invalidated. For this reason, the strategy outlined in Diebold and Mariano (1995), p.255, is adopted, namely to partition the loss-differential sequence of two models i and j into k subsequences $\{d_{ij,1}, d_{ij,1+k}, d_{ij,1+2k}, ...\}$, $\{d_{ij,2}, d_{ij,2+k}, d_{ij,2+2k}, ...\}, ..., \{d_{ij,k}, d_{ij,2k}, d_{ij,3k}, ...\}$ and then to conduct the sign test on each of these subsequences. In the present case of four competing models, $H_0 : \pi_{ij} = 0.5$ is rejected if any of the three (sub-) tests could reject the null. In the case of serial dependence of order $(k-1)$ these 'thinned out' subsequences should not exhibit any further serial correlation. For the German unemployment data $k = 2$ has shown to be sufficient to achieve this goal.

As already emphasized at the beginning of this section, interest centers here on capturing the effect of model averaging on the forecasting accuracy compared to a model selection

strategy. Therefore, in the last columns of tables C.2 and C.3, also the PMSE and MAPE values of a conditional BPAR model, which is selected by using BIC, are shown.[25] From the tables it can be observed that for 6 (6) out of the 18 series a conditional BPAR model is superior, however only if the model averaged BPAR model is not considered.[26] When the BMA-PAR model is also taken as a competitor, the model selected BPAR model has only in 1 (1) out of the 18 cases the (statistically significant) lowest average loss compared to the other three forecasting models. Hence this finding is in accordance with the theoretical results stated in the BMA literature, cf. Raftery and Zheng (2003), namely that BMA point estimates minimize the predictive MSE.

Finally, the BMA-PAR model is used to conduct 12-months ahead out-of-sample predictions for the period 03/2013 - 02/2014. In appendix C the BMA posterior means of y_{T+k}, for $k = 1...12$, together with the corresponding Bayesian standard errors and the 95% HPD prediction error intervals are shown (see tables C.4-C.12). For illustration purposes, the original unemployment series of West-Germany together with the predicted values (denoted by '$\widetilde{y}_{T+k}$') is shown in figure F.59. What can be seen from the figure is that the predicted values as well as the HPD intervals favorably render the seasonal variation in the data. Furthermore in figures F.60-F.62 the underlying model averaged marginal posterior predictive densities of y_{T+k}, $k = 1...12$, for West-Germany are depicted, where these densities are computed on a grid of $\{y_{T+k}^{(g)},\ g = 1...100\}$ values. For the prediction first differences are applied to the original series in order to remove any (stochastic or deterministic) trends, and then the predicted values are transformed back to levels. It should be noted, that for nearly all federal states the corresponding model posterior probability mass function, $f(M_i|\mathbf{y})$, assigns decreasing posterior weight to models with higher lag orders, with or without a structural break.[27] This is in accordance with the results of other authors, see Osborn and Smith (1989), Franses and Koop (1997), inter alia, and also with own experience, namely that PAR models of low order often provide a good fit to many economic data sets.

[25] In this context, note the following useful approximate relationship between the BIC and the posterior probability mass function of model M_i: $f(M_i|\ data) \approx \exp(-1/2\ BIC_i) / \sum_{j=1}^{I} \exp(-1/2\ BIC_j)$, which can be derived by applying a Laplace approximation (see Tierney and Kadane (1986), Tierney et al. (1989)) to the joint posterior density.

[26] Most of these six loss-differences are however not statistically significant.

[27] The results for the 18 series are omitted here.

5.6. Concluding remarks

A Bayesian forecasting approach for the class of periodic autoregressive models has been presented. Since the model admits one structural break in the process mean at an unknown point in time, this introduces an additional nuisance parameter in the posterior distribution, which can not be integrated out analytically. Therefore an MCMC approach, based on data augmentation, is proposed in order to sample from the joint posterior predictive density, under a specific model. Where the latter can be characterized by the number of autoregressive lags and the number of structural breaks. In order to capture possible uncertainty induced through a model selection step, a model averaging approach for prediction has been presented. Instead of working with conditional, i.e. model-specific, predictive distributions, the model posterior probability mass function has been utilized to obtain model averaged posterior predictive densities. Where the posterior means of these mixture distributions have been used as point forecasts of the unknown future values.

In a Monte Carlo study, Bayesian PAR models have been compared with SARMA and seasonal means models in terms of forecasting accuracy. In particular, the simulation results lend support to the use of PAR models in the case of periodic unit roots. In order to test if two competing forecasting models differ significantly with respect to their predictive accuracy, a Bayesian sign test has been proposed. In an empirical application the model averaged BPAR model has been used to forecast monthly unemployment rates of the 16 federal states of Germany and of East- and West-Germany for one year ahead. In simulated-out-of-sample forecasts the BMA-PAR model has clearly outperformed a SARMA and a periodic means model. In addition, a comparison of the forecasting accuracy of the BMA-PAR model and a model-selected BPAR model has demonstrated how model averaging can improve predictive accuracy.

Overall the results suggest that periodic autoregressive models provide a flexible alternative to commonly used seasonal models like SARMA and seasonal means models, particularly in the case of periodic unit roots, i.e. when seasonality might change over time. Moreover it has been demonstrated that combining evidence from different forecasting models by means of a mixture posterior predictive distribution helps to improve forecasting accuracy. However here the predefinition of the model space, i.e. the choice of p_{max} and m_{max}, might be crucial. One drawback of PAR(MA) models is their great

number of parameters, which gets more inflated when choosing a model averaging approach. For future work the above model specification could be made more parsimonious by allowing for heterogenous autoregressive lag orders p_s, $s = 1...S$. Further, by treating each individual lag $D_{s,t} \cdot y_{t-i}$, $s = 1...S$, $i = 1..p_s$, as a variable whose inclusion is controlled, for example by a stochastic search variable selection approach (cf. George et al. (1993), So et al. (2006), Chen et al. (2011)), the number of parameters could substantially be reduced. Since the inclusion of irrelevant lags introduces additional noise into the forecasting process (cf. Clark and West (2007)), this approach might not only reduce the number of parameters but also increase the forecasting accuracy.

6. Final summary and discussion

This dissertation presents new evidence on the use of Bayesian methods for unit root testing and forecasting of (non)seasonal time series data. In the context of nonseasonal data (see chapter 2), a Bayesian unit root testing approach for the case of multiple structural breaks is presented. Here the number of structural breaks, their location, and also the number of autoregressive lags in the test regression are treated as random variables. The Bayes test proceeds by first identifying the most likely model, and then unit root inference is drawn conditional on this model specification. For the model selection a mixed MCMC sampling approach is proposed, which allows to jump between parameter spaces of varying dimension. As a sensitivity check with respect to the prior choice, the frequentist risk functions for the posterior Bayes estimator of the long-run impact coefficient, are simulated using different prior specifications. With the presented model selection approach it is possible to identify the most likely candidate model through the approximated joint posterior distribution of the number of autoregressive lags and the number of structural breaks. This distribution captures the uncertainty induced by picking out a particular model for inference and can further be utilized to compute model averaged point estimates of the parameters of interest, e.g. the half lives of a shock. The results of an empirical application, using annual unemployment rates of 17 OECD countries, indicate that the only country with high posterior probability for unemployment hysteresis is Greece, whereas Japan and Spain show slightly increased levels of persistence. Overall the empirical analysis suggests that the majority of the considered OECD unemployment rates are likely to follow a trend stationary process with possible level shifts.

In the second part of the thesis the focus is on unit root testing and forecasting of seasonal time series data. Since in practice often time series of higher than annual frequency are used, seasonal forms of nonstationarity can occur, and thus testing procedures for seasonal unit roots are required. In this regard, a flexible class of seasonal time series models, able

to capture changing seasonality, is the class of periodic autoregressive models. In chapter 4, a Bayesian testing strategy to test for a periodic unit root with a possible mean break at an unknown point in time is presented. For this test the posterior density of the product of periodic autoregressive coefficients is required. Here an approximation, based on a first order Taylor series expansion, and a direct sampling approach, based on the joint posterior density of the PAR coefficients, are presented. Further the marginal posterior densities of the PAR(1) coefficients, ϕ_s, $s = 1...S$, are derived and used to test the hypothesis of a Random Walk in season s. In addition, in order to test for a real-valued seasonal and a nonseasonal unit root, a Bayesian F-test is proposed. All unit root tests presented in chapter 4 are based on model averaging techniques so that it is possible to combine evidence from different models, here: a PAR(1) model with and without a structural break. As a pretest a Bayesian recursive F-test to test for the presence of periodic variation of order one in the data is proposed. In an empirical analysis these methods are applied to test for unemployment hysteresis in the monthly unadjusted unemployment rates of the 17 OECD countries used in the empirical section of chapter 2. The results show that most of the monthly unemployment rates exhibit unit root behavior. Among the four countries having the highest posterior probabilities of a periodic unit root are Greece, Ireland, Spain and the UK. Moreover many of these series are driven by a nonperiodic stochastic trend, which is implied by a zero frequency unit root.

In contrast to the results of chapter 2, the empirical results of chapter 4 suggest the presence of a stochastic trend in the OECD unemployment rates. However it should be noted that although the same countries have been considered, the analysis of chapter 4 has been performed with monthly data, whereas in chapter 2 yearly data have been used. More importantly, the statistical models in both chapters are quite different, since in chapter 2 nonperiodic autoregressive models of possibly high order and multiple breaks have been considered, whereas in chapter 4 periodic autoregressive models of order one with at most one break have been used. In this respect, there is a trade-off often experienced in practice between allowing for a (possibly large) number of structural breaks and the unit-root-evidence found in the data. In other words, by allowing for breaks in the deterministic part of a stochastic process some of the sample variation is captured by the additional deterministic terms and thus the variation attributed to the stochastic part of the process is reduced.

In chapter 5, a novel Bayesian forecasting approach for the class of periodic autoregressive models is presented. For this class the joint posterior predictive distribution of the multistep ahead forecasts is derived. Similar to the testing approach of chapter 2, the periodic forecasting model in chapter 5 treats the number of autoregressive lags, the occurrence of a break and the corresponding break date as unknown random parameters. However, instead of working with conditional, i.e. model-specific, predictive distributions, the model posterior probability distribution is utilized to obtain model averaged posterior predictive densities. The posterior means of these mixture distributions are used as point forecasts of the unknown future values. In an empirical application, using monthly unadjusted unemployment rates of the 16 federal states of Germany and of East- and West-Germany, it is demonstrated that model averaging helps to improve forecasting accuracy compared to a conditional approach. To test if two forecasting models differ significantly with respect to their predictive accuracy, a Bayesian sign test, using several prior specifications as a sensitivity check, is proposed. Since the model admits one structural break at an unknown point in time, this introduces an additional nuisance parameter in the joint posterior distribution which can not be integrated out analytically. Therefore an MCMC approach, based on data augmentation, is proposed in order to sample from the joint posterior predictive density, under a specific model.

In each of the three articles the presented Bayesian methods have been compared with classical competitors using simulated and real data. Overall the results suggest that Bayesian methods may often provide useful alternatives to classical methods, as for example in the case of unit root testing with an unknown number of structural breaks and associated break dates. This may particularly be the case when some regularity conditions under which the classical tests are derived become invalid, e.g. due to short time series, whereas Bayesian probability statements stay the same irrespective of the sample size. For example, under a flat prior the marginal posterior distribution of the long-run impact coefficient in an Augmented Dickey-Fuller regression is a Student-t density, irrespective of the sample size. Moreover, as discussed in section 1.2, this distribution stays the same for stationary and also nonstationary data. By using a Bayesian frame of reference it is straightforward to assign positive prior probabilities to the members of a predefined set of candidate models and then to obtain the (approximate) model posterior probability distribution. The latter can then be used for model selection or model averaging. With regard to the prediction of future data, using monthly unemployment data of Germany, it is shown

that model averaging can improve forecasting accuracy.

In a Bayesian framework the influence of the prior distribution becomes more pronounced in small samples and thus problems of prior elicitation have to be taken more seriously. Throughout this dissertation normally distributed innovations have been assumed in order to utilize some analytical results based on conjugate prior distributions. This might be restrictive for many types of data, as for example financial data, which exhibit some well known stylized facts like heavy tails, leptokurtosis and volatility clustering. Another critical assumption concerns the functional form of the considered models, namely the linearity, and also the assumed form of the structural breaks. In this thesis only breaks in the process mean have been allowed. However controlling for breaks in the variance may also be important in the context of unit root testing (cf. Kim et al. (2002), Cavaliere (2005*b*)).

One route for future work is to focus on Bayesian unit root testing in nongaussian, nonlinear time series models. Here the multi-regime heteroscedastic SETAR(1) model presented in So et al. (2006) and used in Chen et al. (2012) in order to test for local nonstationarity could be extended to capture higher order dynamics in the mean equation. Moreover the presented methods could be extended to allow for variance breaks at unknown points in time and their effects on the outcomes of Bayesian unit root tests could be examined. In this dissertation point hypotheses have been tested by using a mixed prior distribution, which assigns a positive prior probability to the singleton under the null (cf. Berger and Delampady (1987)). As a robustness check of the presented results, the Full Bayesian Significance Test introduced by Pereira and Stern (1999) (see also Pereira et al. (2008)) for testing precise hypotheses could be used.

Acknowledgements:

The author would like to thank Cathy W.S. Chen (Feng Chia University, Taiwan), Uwe Blien (Institute for Employment Research (IAB) and University of Bamberg), Enzo Weber (IAB and University of Regensburg), Stefan Fuchs (IAB), and Christian Assmann (University of Bamberg) for many helpful comments.

Bibliography

Andel, J. (1983), "Statistical Analysis of Periodic Autoregression," *Aplikace matematiky*, 28(5), 364–385.

Arestis, P., and Biefang-Frissancho Mariscal, I. (1999), "Unit Roots and structural breaks in OECD unemployment," *Economic Letters*, 65, 149–156.

Bai, J., and Perron, P. (1998), "Estimating and Testing Linear Models with Multiple Structural Changes," *Econometrica*, 66(1), 47–78.

Bai, J., and Perron, P. (2003), "Computation and analysis of multiple structural change models," *Journal of Applied Econometrics*, 18(1), 1–22.

Banerjee, A., Dolado, J., Galbraith, J. W., and Hendry, D. F. (1993), *Co-Integration, Error-Correction and the Econometric Analysis of Non-Stationary Data*, Oxford: Oxford University Press.

Bauwens, L., Lubrano, M., and Richard, J. F. (1999), *Bayesian Inference in Dynamic Econometric Models*, 1. edn, Oxford: Oxford University Press.

Beaulieu, J., and Miron, J. (1993), "Seasonal unit roots in aggregate U.S. data," *Journal of Econometrics*, 54, 305–328.

Berger, J. O. (1980), *Statistical Decision Theory and Bayesian Analysis*, 2. edn, New York: Springer.

Berger, J. O., and Delampady, M. (1987), "Testing precise hypotheses," *Statistical Science*, 2(3), 317–335.

Berger, J. O., and Sellke, T. (1987), "Testing a point null hypothesis: the irreconcilability of P values and evidence," *Journal of the American Statistical Association*, 82(397), 112–122.

Berger, T., and Everaert, G. (2008), "Unemployment Persistence and the Nairu: A Bayesian Approach," *Scottish Journal of Political Economy*, 55(3), 281–299.

Berry, D., and Hochberg, Y. (1999), "Bayesian perspectives on multiple comparisons," *Journal of Statistical Planning and Inference*, 82(1), 215–227.

Besag, J. (1989), "A Candidate's Formula: A Curious Result in Bayesian Prediction," *Biometrika*, 76(1), 183.

Birchenhall, C. R., Bladen-Hovell, R., Chui, A., Osborn, D. R., and Smith, J. (1989), "A Seasonal Model of Consumption," *The Economic Journal*, 99(397), 837–843.

Blanchard, O. J., Katz, L. F., Hall, R. E., and Eichengreen, B. (1992), "Regional Evolutions," *Brookings Papers on Economic Activity*, (1), 1–75.

Blanchard, O. J., and Summers, L. H. (1986), "Hysteresis and the European Unemployment Problem," *NBER Macroeconomics Annual*, 1, 15–78.

Blanchard, O. J., and Summers, L. H. (1987), "Hysteresis in Unemployment," *European Economic Review*, 31, 288–295.

Bloomfield, P. (2000), *Fourier Analysis of Time Series*, 2. edn, New York: Wiley & Sons.

Boswijk, H., and Franses, P. H. (1995), "Testing for periodic integration," *Economics Letters*, 48, 241–248.

Boswijk, H., and Franses, P. H. (1996), "Unit roots in periodic autoregressions," *Journal of Time Series Analysis*, 17(3), 221–245.

Boswijk, H., Franses, P. H., and Haldrup, N. (1995), "Multiple unit roots in periodic autoregression," *Journal of Econometrics*, 80, 167–193.

Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. (2008), *Time Series Analysis Forecasting and Control*, 4 edn, New Jersey: Wiley.

Broemeling, L., and Land, M. (1984), "On forecasting with univariate autoregressive processes: A Bayesian approach," *Communications in Statistics - Theory and Methods*, 13(11), 1305–1320.

Casella, G. (1985), "An Introduction to Empirical Bayes Data Analysis," *The American Statistician*, 39(2), 83–87.

Casella, G., and Berger, R. L. (2002), *Statistical Inference*, 2. edn, Pacific Grove: Duxbury.

Casella, G., and George, E. I. (1992), "Explaining the Gibbs Sampler," *The American Statistician*, 46(3), 167–174.

Cavaliere, G. (2005*a*), "Limited Time Series with a Unit Root," *Econometric Theory*, 21(5), 907–945.

Cavaliere, G. (2005*b*), "Unit Root Tests under Time-Varying Variances," *Econometric Reviews*, 23(3), 259–292.

Chaloner, K., and Brant, R. (1988), "A Bayesian approach to outlier detection and residual analysis," *Biometrika*, 75, 651–659.

Chen, C., Chen, S., and Lee, S. (2012), "Bayesian unit root test in double threshold heteroskedastic models," *Computational Economics*, 35(4), 1–18.

Chen, C., Liu, F., and Gerlach, R. (2011), "Bayesian subset selection for threshold autoregressive moving-average models," *Computational Statistics*, 26, 1–30.

Chib, S. (1995), "Marginal Likelihood from the Gibbs Output," *Journal of the American Statistical Association*, 90(432), 1313–1321.

Chib, S. (1998), "Estimation and comparison of multiple change-point models," *Journal of Econometrics*, 86, 221–241.

Chib, S., and Greenberg, E. (1995), "Understanding the Metropolis-Hastings Algorithm," *The American Statistician*, 49(4), 327–335.

Chipman, H., George, E. I., and McCulloch, R. E. (2001), "The Practical Implementation of Bayesian Model Selection," *IMS Lecture Notes - Monograph Series*, 38.

Christiano, L. J. (1992), "Searching for a Break in GNP," *Journal of Business Economics and Statistics*, 10, 237–250.

Clark, T., and West, K. (2007), "Approximately normal tests for equal predictive accuracy in nested models," *Journal of Econometrics*, 138, 291–311.

DeGroot, M. H. (1970), *Optimal Statistical Decisions*, 1. edn, New York: Wiley and Sons.

DeJong, D. N. (1996), "A Bayesian Search for Structural Breaks in US GNP," in *Advances in Econometrics: Bayesian Methods Applied to Time Series Analysis*, ed. T. Fomby, Vol. B, Greenwich and Conn.: JAI Press.

Dellaportas, P., Forster, J. J., and Ntzoufras, I. (2002), "On Bayesian model and variable selection using MCMC," *Statistics and Computing*, 12(1), 27–36.

Dickey, D. A., and Fuller, W. A. (1979), "Distribution of the Estimates for Autoregressive Time Series with a Unit Root," *Journal of the American Statistical Association*, (74), 427–431.

Dickey, D. A., Hasza, D. P., and Fuller, W. A. (1984), "Testing for Unit Roots in Seasonal Time Series," *Journal of the American Statistical Association*, 79(386), 355–367.

Diebold, F. X., and Mariano, R. (1995), "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics*, 15(3), 253–263.

Draper, D. (1995), "Assessment and propagation of model uncertainty," *Journal of Royal Statistical Society, Series B*, 57(1), 45–70.

Dreze, J. H. (1977), "Bayesian regression analysis using poly-t densities," *Journal of Econometrics*, 6(3), 329–354.

Ehlers, R. S., and Brooks, S. P. (2002), "Efficient Construction of Reversible Jump MCMC Proposals for Autoregressive Time Series Models," *Working Paper*, .

Ehlers, R. S., and Brooks, S. P. (2004), "Bayesian Analysis of Order Uncertainty in ARIMA Models," *Technical Report*, .

Elliot, G., Rothenberg, T., and Stock, J. H. (1996), "Efficient Tests for an Autoregressive Unit Root," *Econometrica*, 64, 813–836.

Engle, R. F., and Granger, C. W. J. (1987), "Co-Integration and Error Correction: Representation, Estimation, and Testing," *Econometrica*, 55(2), 251–276.

Fernandez, C. L., and Steel, M. (2001), "Benchmark priors for Bayesian model averaging," *Journal of Econometrics*, 100, 381–427.

Franses, P. H. (1991), "Seasonality, non-stationarity and the forecasting of monthly time series," *International Journal of Forecasting*, 7, 199–208.

Franses, P. H. (1994), "A multivariate approach to modeling univariate seasoanl time series," *Journal of Econometrics*, 63, 133–151.

Franses, P. H. (1995), "Quarterly US unemployment: Cycles, seasons and asymmetries," *Empirical Economics*, 20(4), 717–725.

Franses, P. H. (2003), *Periodicity and stochastic trends in economic time series*, 2. edn, New York: Oxford University Press.

Franses, P. H., Hoek, H., and Paap, R. (1997), "Bayesian analysis of seasonal unit roots and seasonal mean shifts," *Journal of Econometrics*, 78, 359–380.

Franses, P. H., and Koop, G. (1997), "A Bayesian Analysis of Periodic Integration," *Journal of Forecasting*, 16, 509–532.

Franses, P. H., and Paap, R. (2006), *Periodic Time Series Models*, 2. edn, New York: Oxford University Press.

Fuller, W. A. (1996), *Introduction to statistical time series*, 2. edn, New York: Wiley.

Garcia-Donato, G., and Martinez-Beneito, M. (2012), "On sampling strategies in Bayesian variable selection problems with large model spaces (forthcoming)," *Journal of the Royal Statistical Society*, .

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (1995), *Bayesian Data Analysis*, 1. edn, London: Chapman & Hall.

George, E. I., McCulloch, and R.E. (1993), "Variable Selection Via Gibbs Sampling," *Journal of the American Statistical Association*, 88(423), 881–889.

Geweke, J., and Whiteman, C. H. (2006), "Bayesian Forecasting," *Handbook of Economic Forecasting*, 1, 3–80.

Ghysels, E. (1990), "Unit root tests and the statistical pitfalls of seasonal adjustment: The case of U.S. postwar Real Gross National Product," *Journal of Business and Economic Statistics*, 8(2), 145–152.

Ghysels, E., and Osborn, D. R. (2001), *The Econometric Analysis of Seasonal Time Series*, 1. edn, Cambridge: Cambridge University Press.

Ghysels, E., Osborn, D., and Rodrigues, P. (2006), *Forecasting Seasonal Time Series*, Vol. 1 of *Handbook of Economic Forecasting*, Amsterdam: Elsevier.

Ghysels, E., and Perron, P. (1993), "The effect of seasonal adjustment filters on tests for a unit root," *Journal of Econometrics*, 55, 57–98.

Giacomini, R., and White, H. (2006), "Tests of Conditional Predictive Ability," *Econometrica*, 74, 1545–1578.

Gladyshev, E. (1961), "Periodically correlated random sequences," *Soviet Mathematics*, 2, 385–388.

Godsill, S. J. (2001), "On the Relationship Between Markov Chain Monte Carlo Methods for Model Uncertainty," *Journal of Computational and Graphical Statistics*, 10(2), 230–248.

Granger, C. W. J. (1986), "Developments in the study of cointegrated economic variables," *Oxford Bulletin of Economics and Statistics*, 48(3), 213–228.

Greene, W. H. (2003), *Econometric Analysis*, 5. edn, New Jersey: Prentice Hall.

Hall, A. (1994), "Testing for a Unit Root in Time Series with Pretest Data-Based Model Selection," *Journal of Business and Economic Statistics*, 12, 461–470.

Hall, R. (1978), "Stochastic implication of the cycle-permanent income hypothesis: theory and evidence," *Journal of Political Economy*, 86, 971–987.

Hamilton, J. D. (1994), *Time Series Analysis*, 1. edn, New York: Cambridge University Press.

Hansen, B. E. (2007), "Least Squares Model Averaging," *Econometrica*, 75(4), 1175–1189.

Hansen, L., and Sargent, T. (1993), "Seasonality and Approximation Errors in Rational Expectations Models," *Journal of Econometrics*, 55, 21–56.

Hassler, U., and Wolters, J. (2009), "Hysteresis in Unemployment Rates? A Comparison between Germany and the US," in *Jahrbücher für Nationalökonomie und Statistik*, ed. P. Winker, Vol. 229, Stuttgart: Lucius.

Hastings, W. K. (1970), "Monte Carlo sampling methods using Markov chains and their applications," *Biometrika*, 57(1), 97–109.

Hjort, N., and Claeskens, G. (2003), "Frequentist Model Average Estimators," *Journal of the American Statistical Association*, 98, 879–899.

Hoeting, J., Madigan, D., Raftery, A. E., and Volinsky, C. (1999), "Bayesian model averaging: A tutorial," *Statistical Science*, 14, 382–417.

Hong, H., and Preston, B. (2012), "Bayesian averaging, prediction and nonnested model selection," *Journal of Econometrics*, 167, 358–369.

Huerta, G., and West, M. (1999), "Bayesian inference on periodicities and component spectral structure in time series," *Journal of Time Series Analysis*, 20(4), 401–416.

Hylleberg, S., Engle, R. F., Granger, C. W. J., and Yoo, B. S. (1990), "Seasonal Integration and Cointegration," *Journal of Econometrics*, 44, 215–238.

Inoue, A., and Kilian, L. (2006), "On the selection of forecasting models," *Journal of Econometrics*, (130), 273–306.

Johansen, S. (1988), "Statistical analysis of cointegration vectors," *Journal of Economic Dynamics and Control*, 12(2), 231–254.

Johnson, N., and Kotz, S. (1970), *Continuous Univariate Distributions 2*, 1. edn, New Jersey: John Wiley & Sons, Inc.

Judge, G., Griffiths, W., Hill, R., Lütkepohl, H., and Lee, T. (1985), *The Theory and Practice of Econometrics*, 2. edn, New York: John Wiley & Sons, Inc.

Kadane, J. B., Chan, N. H., and Wolfson, L. J. (1996), "Priors for unit root models," *Journal of Econometrics*, 75(1), 99–111.

Kass, R. E., and Raftery, A. E. (1995), "Bayes Factors," *Journal of the American Statistical Association*, 90(430), 773–795.

Kendall, M. G., and Stuart, A. (1969), *The Advanced Theory of Statistics - Distribution Theory*, Vol. 1, 3. edn, London: Charles Griffin and Company Limited.

Kim, T. H., Leybourne, S., and Newbold, P. (2002), "Unit root tests with a break in innovation variance," *Journal of Econometrics*, 109(2), 365–387.

King, R., Plosser, C., and Rebelo, S. (1988), "Production, growth and business cycles: I. The basic neoclassical model," *Journal of Monetary Economics*, 21, 195–232.

Klein, I. (2008), "Some R graphics for bivariate distributions," *Discussion Paper - University of Nuremberg*, (83).

Koop, G., Oseiwalski, J., and Steel, M. F. J. (1995), "Bayesian long-run prediction in time series models," *Journal of Econometrics*, 69, 61–80.

Koop, G., and Potter, S. (1999), "Dynamic Asymmetries in U.S. Unemployment," *Journal of Business & Economic Statistics*, 3(17), 298–312.

Koop, G., and Potter, S. (2004), "Forecasting and estimating multiple change-point models with an unknown number of change points," *Review of Economic Studies*, 74(3), 763–789.

Koop, G., and Steel, M. F. J. (1994), "A Decision-Theoretic Analysis of the Unit-Root Hypothesis using Mixtures of Elliptical Models," *Journal of Business and Economic Statistics*, 12, 95–107.

Layard, R., Nickell, S., and Jackman, R. (1991), *Unemployment: macroeconomic performance and the labour market*, 1. edn, Oxford: Oxford University Press.

Liang, F., Paulo, R., Molina, G., Clyde, M., and Berger, J. (2008), "Mixtures of g-Priors for Bayesian Variable Selection," *Journal of the American Statistical Association*, 103(481), 410–423.

Lopes, H. F. (2006), "A note on Reversible Jump Markov Chain Monte Carlo," *Discussion Paper*, .

Lubrano, M. (2001), "Smooth Transition Garch Models: a Bayesian Perspective," *Recherches Economiques de Louvain / Louvain Economic Review*, 67(3), 257–287.

Lütkepohl, H. (2007), *New Introduction to Multiple Time Series Analysis*, 2. edn, Berlin: Springer.

Maddala, G. S., and Kim, I. M. (1998), *Unit Roots, Cointegration and Structural Change*, 6. edn, New York: Oxford University Press.

Madigan, D., and Raftery, A. E. (1994), "Model selection and accounting for model uncertainty in graphical models using Occam's window," *Journal of the American Statistical Association*, 89(428).

Marin, J. M., and Robert, C. P. (2010), *Bayesian Core*, 1. edn, New York: Springer.

Marriott, J., and Newbold, P. (2000), "The strength of evidence for unit autoregressive roots and structural breaks: A Bayesian perspective," *Journal of Econometrics*, 98(1), 1–25.

Meese, R., and Rogoff, K. (1983), "Empirical exchange rate models of the seventies. Do they fit out of sample?," *Journal of International Economics*, (14), 3–24.

Mikhail, O., Eberwein, C., and Handa, J. (2006), "Estimating persistence in Canadian unemployment: evidence from a Bayesian ARFIMA," *Applied Economics*, 38(15), 1809–1819.

Mitchell, W. F. (1993), "Testing for unit roots and persistence in OECD unemployment rates," *Applied Economics*, 25(12), 1489–1501.

Monahan, J., and Boos, D. (1992), "Proper likelihoods for Bayesian analysis," *Biometrika*, 79(2), 271–278.

Nelson, C. R., and Plosser, C. I. (1982), "Trends and Random Walks in Macroeconomic Time Series," *Journal of Monetary Economics*, 10, 139–162.

Ng, S., and Perron, P. (2001), "Lag length selection and the construction of unit root tests with good size and power," *Econometrica*, 69(6), 1519–1554.

Ooms, M., and Franses, P. H. (1997), "On periodic correlations between estimated seasonal and nonseasonal components in German and U.S. unemployment," *Journal of Business and Economic Statistics*, 15(4), 470–481.

Osborn, D. R. (1991), "The Implications of Periodically Varying Coefficients for Seasonal Time-Series Processes," *Journal of Econometrics*, 48, 373–384.

Osborn, D. R., Chui, A., Smith, J., and Birchenhall, C. R. (1988), "Seasonality and the Order of Integration for Consumption," *Oxford Bulletin of Economics and Statistics*, 50, 361–377.

Osborn, D. R., and Smith, J. (1989), "The Performance of Periodic Autoregressive Models in Forecasting Seasonal U.K. Consumption," *Journal of Business and Economic Statistics*, 7(1), 117–128.

Pagano, M. (1978), "On periodic and multiple autoregressions," *The Annals of Statistics*, 6(6), 1310–1317.

Papell, D. H., Murray, C. J., and Ghiblawi, H. (2000), "The Structure of Unemployment," *Review of Economics and Statistics*, 82, 309–315.

Papell, D. H., and Prodan, R. (2004), "The Uncertain Unit Root in U.S. Real GDP: Evidence with Restricted and Unrestricted Structural Change," *Journal of Money, Credit and Banking*, 36(3), 423–427.

Pascalau, R. (2007), "Unit Root Tests with Smooth Breaks: An Application to the Nelson-Plosser Data Set," *Working Paper*, .

Pereira, C., and Stern, J. (1999), "Evidence and credibility: full Bayesian significance test for precise hypotheses.," *Entropy*, 1, 69–80.

Pereira, C., Stern, J., and Wechsler, S. (2008), "Can a Significance Test be genuinely Bayesian?," *Bayesian Analysis*, 3, 79–100.

Perron, P. (1989), "The great crash, the oil price shock and the unit root hypothesis.," *Econometrica*, 57, 1361–1401.

Perron, P., and Vogelsang, T. J. (1992), "Nonstationarity and Level Shifts with an Application to Purchasing Power Parity," *Journal of Business and Economic Statistics*, 10, 301–320.

Phelps, E. S. (1994), *Structural Slumps: The Modern Equilibrium Theory of Unemployment, Interest and Assests*, Cambridge: Harvard University Press.

Philippe, A. (2006), "Bayesian analysis of autoregressive moving average processes with unknown orders," *Computational Statistics & Data Analysis*, 51(3), 1904–1923.

Phillips, P. C. B. (1991*a*), "Bayesian Routes and Unit Roots: De Rebus Prioribus Semper Est Disputandum," *Journal of Applied Econometrics*, 6(4), 435–473.

Phillips, P. C. B. (1991*b*), "To Criticize the Critics: An Objective Bayesian Analysis of Stochastic Trends," *Journal of Applied Econometrics*, 6(4), 333–364.

Phillips, P. C. B., and Perron, P. (1988), "Testing for a Unit Root in Time Series Regression," *Biometrika*, 88(75), 335–346.

Poole, D. (2006), *Linear Algebra. A Modern Introduction*, 2. edn, Belmont: Thomson Brooks/Cole.

Priestley, M. (2004), *Spectral Analysis and Time Series*, Probability and Mathematical Statistics, 2. edn, London: Elsevier.

Raftery, A. E. (1995), "Bayesian Model Selection in Social Research," *Sociological Methodology*, 25, 111–163.

Raftery, A. E., Madigan, D., and Hoeting, J. (1997), "Bayesian Model Averaging for Linear Regression Models," *Journal of the American Statistical Association*, 92, 179–191.

Raftery, A. E., and Zheng, Y. (2003), "Performance of Bayesian Model Averaging," *Journal of the American Statistical Association*, 98, 931–938.

Raftery, A., Madigan, D., and Volinsky, C. (1996), "Accounting for model uncertainty in survival analysis improves predictive performance (with discussion)," in *Bayesian Statistics*, eds. J. Berger, J. Bernardo, A. Dawid, D. Lindley, and A. Smith, Vol. 5, London: Oxford University Press, pp. 323–349.

Raiffa, H., and Schlaifer, R. (2000), *Applied Statistical Decision Theory*, 1. edn, New York: Wiley and Sons.

Richard, J., and Tompa, H. (1980), "On the evaluation of Poly-t density functions," *Journal of Econometrics*, 12, 335–351.

Robert, C. P. (2007), *The Bayesian Choice*, 1. edn, New York: Springer.

Roed, K. (1996), "Unemployment Hysteresis - Macro Evidence from 16 OECD Countries," *Empirical Economics*, 21, 589–600.

Said, S. E., and Dickey, D. A. (1984), "Testing for Unit Roots in autoregressive moving average models of unknown order," *Biometrika*, 71(3), 599–607.

Schotman, P., and van Dijk, H. K. (1991), "On Bayesian Routes to Unit Roots," *Journal of Applied Econometrics*, 6(4), 387–401.

Schwarz, G. (1978), "Estimating the dimension of a model," *The Annals of Statistics*, 6(2), 461–464.

Shafer, G. (1982), "Lindley's paradox," *Journal of the American Statistical Association*, 77(378), 325–334.

Sims, C. A. (1988), "Bayesian Skepticism on Unit Root Econometrics," *Journal of Economic Dynamics and Control*, 12, 436–474.

Sims, C. A., and Uhlig, H. (1991), "Understanding Unit Rooters: A Helicopter Tour," *Econometrica*, 59(6), 1591–1599.

So, M., Chen, C., and Liu, F. (2006), "Best subset selection of autoregressive models with exogenous variables and generalized autoregressive conditional heteroscedasticity errors," *Journal of the Royal Statistical Society: Series C*, 55, 201–224.

Spanos, A. (1999), *Probability Theory and Statistical Inference*, 1. edn, Cambridge: Cambridge University Press.

Stock, J. H. (1994), "Deciding between I(1) and I(0)," *Journal of Econometrics*, 63, 105–131.

Stock, J. H., and Watson, M. W. (1999), "Forecasting inflation," *Journal of Monetary Economics*, 44(2), 293–335.

Summers, P. M. (2004), "Bayesian Evidence on the Structure of Unemployment," *Economics Letters*, 83, 299–306.

Tanner, M., and Wong, W. (1987), "The Calculation of Posterior Distributions by Data Augmentation," *Journal of the American Statistical Association*, 82(398), 528–540.

Tiao, G. C., and Grupe, M. (1980), "Hidden periodic autoregressive-moving average models in time series data," *Biometrika*, 67(2), 365–373.

Tierney, L., and Kadane, J. B. (1986), "Accurate Approximations for Posterior Moments and Marginal Densities," *Journal of the American Statistical Association*, 81(393), 82–86.

Tierney, L., Kass, R., and Kadane, J. (1989), "Fully Exponential Laplace Approximations to Expectations and Variances of Nonpositive Functions," *Journal of the American Statistical Association*, 84(407), 710–716.

Troughton, P. T., and Godsill, S. J. (1997*a*), "Bayesian Model Selection for Time Series using Markov Chain Monte Carlo," *Technical Report*, pp. 3733–3736.

Troughton, P. T., and Godsill, S. J. (1997*b*), "A Reversible Jump Sampler for Autoregressive Time Series, Employing Full Conditionals to Achieve Efficient Model Space Moves," *Technical Report*, pp. 1–13.

Troutman, B. (1979), "Results in Periodic Autoregression," *Biometrika*, 66(2), 219–228.

Uhlig, H. (1994), "What Macroeconomists should know about unit roots - A bayesian perspective," *Econometric Theory*, 10, 645–671.

van der Vaart, A. (1998), *Asymptotic Statistics*, Cambridge: Cambridge University Press.

Vecchia, A. (1985), "Maximum Likelihood Estimation for Periodic Autoregressive Moving Average Models," *Technometrics*, 27(4), 375–384.

Vermaak, J., Andrieu, C., Doucet, A., and Godsill, S. J. (2004), "Reversible Jump Markov Chain Monte Carlo Strategies for Bayesian Model Selection in Autoregressive Processes," *Journal of Time Series Analysis*, 25(6), 785–809.

Vogelsang, T. J., and Perron, P. (1992), "Nonstationarity and Level Shifts with an Application to Purchasing Power Parity," *Journal of Business and Economic Statistics*, 10(3), 301–320.

Wallis, K. (1987), "Time Series Analysis of bounded economic variables," *Journal of Time Series Analysis*, 8(1), 115–123.

Wang, J., and Zivot, E. (2000), "A Bayesian Time Series Model of Multiple Structural Changes in Level, Trend, and Variance," *Journal of Business and Economic Statistics*, 18(3), 374–386.

Zellner, A. (1971), *Introduction to Bayesian Inference in Econometrics*, 1. edn, New York: Wiley and Sons.

Zellner, A. (1975), "Bayesian Analysis of Regression Error Terms," *Journal of the American Statistical Association*, 70(349), 138–144.

Zellner, A. (1986), "On assessing prior distributions and Bayesian regression analysis with g-prior distributions," in *Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finietti*, eds. P. K. Goel, and A. Zellner, Amsterdam: North-Holland, pp. 233–243.

Zellner, A., and Siow, A. (1980), "Posterior Odds Ratios for Selected Regression Hypotheses," in *Bayesian Statistics*, eds. J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith, Valencia: University Press, pp. 583–603.

Zivot, E., and Phillips, P. C. B. (1994), "A Bayesian Analysis of Trend Determination of Economic Time Series," *Econometric Reviews*, 13(3), 291–336.

A. Tables - chapter 2

Table A.1.: Posterior probabilities of the number of structural breaks

		Number of breaks							
Country	*Years*	0	1	2	3	4	5	AIC	BIC
Australia	1966-2010	0.194	0.077	0.117	0.238	0.254	0.120	4	4
Belgium	1983-2010	0.047	0.214	0.136	0.173	0.308	0.122	4	4
Canada	1976-2010	0.098	0.165	0.136	0.178	0.266	0.158	5	3
Denmark	1983-2010	0.091	0.050	0.124	0.254	0.325	0.156	5	4
Finland	1963-2010	0.051	0.041	0.360	0.217	0.221	0.111	5	4
France	1983-2010	0.062	0.043	0.405	0.254	0.171	0.066	5	4
Germany	1970-2010	0.188	0.109	0.259	0.241	0.142	0.062	4	3
Greece	1983-2010	0.061	0.143	0.203	0.255	0.239	0.101	4	4
Ireland	1983-2010	0.096	0.037	0.145	0.305	0.286	0.132	5	5
Italy	1970-2010	0.018	0.035	0.191	0.416	0.260	0.080	3	3
Japan	1968-2010	0.041	0.115	0.187	0.425	0.055	0.181	4	5
Netherlands	1971-2010	0.086	0.037	0.367	0.239	0.171	0.101	4	3
Norway	1972-2010	0.121	0.070	0.234	0.258	0.219	0.100	4	4
Spain	1972-2010	0.039	0.089	0.144	0.244	0.316	0.168	4	4
Sweden	1963-2010	0.034	0.070	0.167	0.366	0.218	0.145	4	3
UK	1984-2010	0.053	0.086	0.090	0.211	0.355	0.206	4	5
US	1960-2010	0.155	0.117	0.106	0.300	0.232	0.090	4	3

'AIC' and 'BIC': Akaike's and Schwarz's information criterion.

Table A.2.: Posterior probabilities of the autoregressive lag order

	Number of lags							
Country	1	2	3	4	5	AIC	BIC	PACF
Australia	0.816	0.081	0.031	0.019	0.054	2	2	1
Belgium	0.699	0.128	0.047	0.027	0.099	1	1	2
Canada	0.196	0.323	0.194	0.188	0.099	1	2	2
Denmark	0.809	0.044	0.031	0.032	0.084	5	1	1
Finland	0.151	0.368	0.207	0.069	0.205	3	2	2
France	0.825	0.069	0.030	0.017	0.059	2	1	1
Germany	0.158	0.421	0.172	0.103	0.147	4	2	1
Greece	0.769	0.052	0.032	0.028	0.119	5	1	1
Ireland	0.719	0.084	0.046	0.025	0.126	1	1	1
Italy	0.835	0.041	0.027	0.022	0.075	1	1	1
Japan	0.719	0.076	0.045	0.026	0.134	4	1	1
Netherlands	0.201	0.422	0.201	0.071	0.105	5	2	2
Norway	0.474	0.238	0.097	0.066	0.126	2	2	2
Spain	0.094	0.319	0.235	0.237	0.116	2	2	2
Sweden	0.631	0.144	0.072	0.043	0.110	3	2	2
UK	0.663	0.070	0.080	0.077	0.111	5	1	1
US	0.612	0.195	0.071	0.035	0.087	2	2	2

'AIC', 'BIC' and 'PACF': Akaike's and Schwarz's information criterion and the (sample) partial autocorrelation function.

Table A.3.: Bayesian break date estimates

		Break dates				
Country	T	1st	2nd	3rd	4th	5th
Australia	45	1974	1981	1990	1993	-
Belgium	28	1988	1992	1999	2001	-
Canada	35	1981	1985	1990	1997	-
Denmark	28	1988	1993	2005	2008	-
Finland	48	1990	1993	-	-	-
France	28	1992	1999	-	-	-
Germany	41	1980	1990	-	-	-
Greece	28	1992	1999	2008	-	-
Ireland	28	1994	2008	2009	-	-
Italy	41	1976	1982	1999	-	-
Japan	43	1992	2002	2008	-	-
Netherlands	40	1980	1983	-	-	-
Norway	39	1987	1993	2005	-	-
Spain	39	1991	1993	2007	2009	-
Sweden	48	1991	1997	2008	-	-
UK	27	1987	1990	1993	2008	-
US	51	1974	1983	2008	-	-

Maximum a-posteriori (MAP) point estimator of T_B used.

Table A.4.: Classical Bai and Perron (2003) break date estimates

		Break dates				
Country	*T*	1st	2nd	3rd	4th	5th
Australia	45	1974	1981	1990	1998	-
Belgium	28	1988	1993	1999	2001	-
Canada	35	1981	1986	1991	1997	-
Denmark	28	1985	1988	1993	1996	2008
Finland	48	1975	1991	1998	-	-
France	28	1992	1999	-	-	-
Germany	41	1975	1981	1992	-	-
Greece	28	1992	1997	2000	2005	2008
Ireland	28	1994	1998	2006	-	-
Italy	41	1976	1983	2001	-	-
Japan	43	1974	1992	1998	2004	-
Netherlands	40	1981	1987	1996	-	-
Norway	39	1981	1988	1996	2005	-
Spain	39	1978	1981	1992	1998	2007
Sweden	48	1991	1998	-	-	-
UK	27	1987	1991	1994	1997	2008
US	51	1974	1986	2006	-	-

Table A.5.: Posterior point estimates $\widehat{\theta}$ and half lives under different priors

Country	$\widehat{\theta}_{Jeff}$	$\widehat{\theta}_{Norm}$	$\widehat{\theta}_{\overline{M}}$	HL_{Norm}	$HL_{\overline{M}}$	HL_{ADF}
Australia	0.59	0.59	0.65	1.30	1.63	5.05
Belgium	0.36	0.36	0.43	0.67	0.83	1.36
Canada	0.51	0.51	0.43	1.03	0.82	2.21
Denmark	0.60	0.60	0.61	1.35	1.41	2.89
Finland	0.77	0.76	0.78	2.58	2.78	6.58
France	0.33	0.32	0.36	0.61	0.67	3.14
Germany	0.72	0.71	0.70	2.05	1.94	2.64
Greece	1.07	0.96	0.81	15.43	3.33	2.92
Ireland	0.80	0.80	0.65	3.02	1.61	5.87
Italy	0.43	0.43	0.44	0.82	0.85	4.73
Japan	0.89	0.88	0.88	5.60	5.32	12.44
Netherlands	0.88	0.87	0.79	5.11	2.93	4.83
Norway	0.60	0.60	0.47	1.34	0.93	5.84
Spain	0.91	0.90	0.88	6.58	5.42	7.06
Sweden	0.62	0.61	0.61	1.41	1.41	7.49
UK	0.76	0.76	0.64	2.48	1.54	3.15
US	0.64	0.63	0.56	1.50	1.20	2.25

'Jeff': Jeffreys prior, 'Norm': Normal prior, '$\overline{M}$': Model averaged, 'ADF': Classical ADF test.

Table A.6.: Posterior probabilities of a unit root and tail probabilities

	$P(\theta = 1 \mid \widehat{\gamma}, \widehat{\mathbf{k}}, \mathbf{y})$		$P(\theta \geq 0.975 \mid \widehat{\gamma}, \widehat{\mathbf{k}}, \mathbf{y})$		$P(T \leq t_{adf} \mid \theta = 1)$
Country	Normal	Jeffreys	Normal	Jeffreys	ADF test
Australia	0.00	0.00	0.00	0.00	0.92
Belgium	0.00	0.00	0.00	0.00	0.39
Canada	0.00	0.00	0.00	0.00	0.22
Denmark	0.00	0.00	0.00	0.00	0.43
Finland	0.01	0.01	0.00	0.00	0.55
France	0.00	0.00	0.00	0.00	0.60
Germany	0.00	0.00	0.00	0.00	0.47
Greece	0.84	0.93	0.45	0.81	0.33
Ireland	0.03	0.03	0.00	0.01	0.97
Italy	0.00	0.00	0.00	0.00	0.74
Japan	0.25	0.25	0.02	0.05	0.35
Netherlands	0.19	0.19	0.02	0.03	0.33
Norway	0.00	0.00	0.00	0.00	0.57
Spain	0.05	0.05	0.01	0.01	0.48
Sweden	0.00	0.00	0.00	0.00	0.32
UK	0.01	0.01	0.00	0.00	0.57
US	0.00	0.00	0.00	0.00	0.43

B. Tables - chapter 4

Table B.1.: Testing for no periodicity

Country	F_{PAR} - stat.	p - val.
Australia	0.90	0.54
Belgium	0.96	0.48
Canada	1.16	0.32
Denmark	2.07	0.03
Finland	0.51	0.90
France	0.67	0.77
Germany	1.71	0.08
Greece	3.62	0.00
Ireland	0.67	0.77
Italy	7.90	0.00
Japan	1.01	0.44
Netherlands	1.73	0.07
Norway	0.86	0.58
Spain	1.16	0.32
Sweden	2.02	0.03
UK	1.07	0.39
US	0.54	0.87

'F_{PAR}': F-statistic to test the null of no periodicity,
i.e. $H_0 : \phi_s = \phi_S$, for $s = 1...S-1$.

Table B.2.: Posterior probabilities of the number of breaks and the break dates

	Number of breaks			Number of breaks	
Country	$f(m=0\|\mathbf{y})$	$f(m=1\|\mathbf{y})$	T_B^{MAP}	BIC ($m=0$)	BIC ($m=1$)
Australia	0.57	0.43	2009(1)	-2.72	-2.35
Belgium	0.99	0.01	2001(7)	-2.12	-1.79
Canada	0.99	0.01	2008(12)	-2.31	-1.93
Denmark	0.99	0.01	2008(11)	-2.65	-2.35
Finland	0.38	0.62	2009(1)	-0.75	-0.41
France	1.00	0.00	2005(3)	-3.19	-2.94
Germany	0.01	0.99	2006(12)	-2.35	-2.08
Greece	0.03	0.97	2010(8)	-0.62	-0.25
Ireland	0.00	1.00	2008(4)	-2.31	-2.06
Italy	0.00	1.00	2003(11)	-1.13	-0.82
Japan	1.00	0.00	2011(1)	-3.08	-2.71
Netherlands	0.99	0.01	2005(7)	-3.25	-2.93
Norway	1.00	0.00	2006(5)	-3.51	-3.18
Spain	0.00	1.00	2008(3)	-1.85	-1.64
Sweden	0.87	0.13	2000(12)	-1.06	-0.76
UK	0.99	0.01	2008(4)	-3.69	-3.36
US	0.99	0.01	2009(5)	-2.64	-2.29

'$f()$': Posterior probability of a model with(out) break. 'BIC': Bayesian information criterion.
'T_B': Maximum a-posteriori (MAP) point estimate of the break date.

Table B.3.: Bayesian and classical test results for a periodic unit root

Country	Determ.	P_0 cond.	P_0 BMA	Post. mean ρ	ρ - 95% HPD	τ_{BF} - stat.
Australia	Both	0.47	0.43	0.20	[0.02 ; 0.64]	-1.73
Belgium	Drift	0.23	0.18	0.36	[0.17 ; 0.66]	-2.37
Canada	Both	0.52	0.47	0.40	[0.15 ; 0.85]	-2.37
Denmark	Drift	0.86	0.78	1.09	[0.84 ; 1.40]	0.67
Finland	Both	0.01	0.00	0.15	[0.02 ; 0.40]	-3.66
France	Drift	0.72	0.58	0.74	[0.54 ; 1.01]	-1.90
Germany	Drift	0.82	0.86	0.84	[0.71 ; 1.00]	0.46
Greece	Drift	0.94	0.95	0.70	[0.33 ; 1.23]	1.97
Ireland	Drift	0.99	0.99	1.16	[1.04 ; 1.28]	2.87
Italy	Drift	0.25	0.33	0.47	[0.23 ; 0.80]	-2.53
Japan	Drift	0.33	0.32	0.60	[0.31 ; 1.01]	-1.85
Netherlands	Both	0.80	0.79	0.90	[0.66 ; 1.19]	-0.83
Norway	Drift	0.66	0.62	0.79	[0.55 ; 1.13]	-1.37
Spain	Drift	0.99	0.99	1.23	[1.09 ; 1.38]	3.76
Sweden	Drift	0.15	0.14	0.37	[0.13 ; 0.81]	-2.56
UK	Drift	0.98	0.98	1.13	[0.95 ; 1.35]	1.55
US	Drift	0.88	0.86	0.96	[0.79 ; 1.15]	-0.48

Notations: 'Determ.' denotes the deterministic specification of the seasonal components included in the model, i.e. periodic drifts, trends or both. 'P_0 cond.': denotes the posterior probability of the periodic unit root null (conditional on a certain model). 'P_0 BMA': denotes the BMA result of the posterior probability of the periodic unit root null (given the MAP estimate T_B^{MAP} for the break point). 'ρ - 95% HPD': denotes the HPD interval of the sampled ρ-values. 'τ_{BF}-stat.': denotes the τ test statistic of the classical Boswijk and Franses (1996) test with -2.86 the 5% critical value.

Table B.4.: Results of (non)seasonal unit root tests

Country	F - stat. (+1)	F - stat. (-1)	p - val. (+1)	p - val. (-1)
Australia	1.99	373.62	0.03	0.00
Belgium	1.30	512.10	0.22	0.00
Canada	1.28	807.19	0.23	0.00
Denmark	1.94	2759.99	0.04	0.00
Finland	4.44	43.62	0.00	0.00
France	0.93	1512.57	0.51	0.00
Germany	1.57	2030.56	0.11	0.00
Greece	4.20	541.72	0.00	0.00
Ireland	2.06	6929.18	0.03	0.00
Italy	4.58	490.83	0.00	0.00
Japan	1.17	570.45	0.30	0.00
Netherlands	1.64	2485.01	0.09	0.00
Norway	0.93	1394.06	0.51	0.00
Spain	1.95	9594.88	0.04	0.00
Sweden	2.36	280.79	0.01	0.00
UK	1.14	6094.98	0.33	0.00
US	0.52	5861.31	0.89	0.00

The reported F-statistics and corresponding p-values are obtained from BMA versions of the tests. '+1' and '-1' denote the tests for a unit root at the zero and π-frequency, respectively.

Table B.5.: 95% HPD intervals of the ϕ_s coefficients

Country	ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6
Australia	[0.77; 1.32]	[0.73; 1.23]	[0.47; 1.14]	[0.65; 1.01]	[0.82; 1.29]	[0.77; 1.19]
Belgium	[0.81; 1.23]	[0.91; 1.26]	[0.83; 1.15]	[0.76; 1.08]	[0.80; 1.14]	[0.69; 1.04]
Canada	[0.86; 1.17]	[0.84; 1.13]	[0.91; 1.18]	[0.83; 1.10]	[0.93; 1.20]	[0.73; 0.98]
Denmark	[1.04; 1.21]	[0.87; 1.03]	[0.97; 1.13]	[0.88; 1.05]	[0.88; 1.06]	[0.93; 1.10]
Finland	[0.55; 1.30]	[0.40; 1.15]	[0.35; 1.10]	[0.45; 1.25]	[0.30; 1.00]	[0.20; 0.95]
France	[0.92; 1.15]	[0.89; 1.08]	[0.86; 1.06]	[0.82; 1.03]	[0.83; 1.04]	[0.84; 1.07]
Germany	[0.81; 0.99]	[0.97; 1.17]	[0.95; 1.13]	[0.93; 1.14]	[0.93; 1.13]	[0.86; 1.06]
Greece	[0.85; 1.10]	[0.86; 1.21]	[0.65; 1.09]	[0.84; 1.22]	[1.04; 1.35]	[0.71; 1.01]
Ireland	[0.75; 1.15]	[0.75; 1.10]	[0.75; 1.05]	[0.80; 1.00]	[0.85; 1.05]	[0.90; 1.05]
Italy	[0.55; 1.00]	[0.95; 1.35]	[0.95; 1.30]	[0.70; 1.05]	[0.55; 0.95]	[0.80; 1.25]
Japan	[0.73; 1.09]	[0.67; 1.06]	[0.97; 1.36]	[0.89; 1.22]	[0.87; 1.18]	[0.84; 1.14]
Netherl.	[0.98; 1.16]	[0.92; 1.08]	[0.96; 1.12]	[0.84; 1.00]	[0.89; 1.06]	[0.93; 1.10]
Norway	[0.83; 1.06]	[0.83; 1.06]	[0.88; 1.12]	[0.91; 1.15]	[0.90; 1.13]	[0.88; 1.10]
Spain	[0.90; 1.12]	[0.90; 1.06]	[0.88; 1.06]	[0.88; 1.02]	[0.92; 1.02]	[0.92; 1.04]
Sweden	[0.56; 1.04]	[0.64; 1.16]	[0.64; 1.14]	[0.90; 1.36]	[0.72; 1.16]	[0.66; 1.12]
UK	[0.97; 1.09]	[0.98; 1.12]	[0.92; 1.04]	[0.96; 1.09]	[0.97; 1.09]	[0.94; 1.07]
US	[0.95; 1.08]	[0.93; 1.05]	[0.94; 1.06]	[0.89; 1.02]	[0.95; 1.08]	[0.95; 1.08]

All HPD regions were computed using the respective model averaged posteriors of the ϕ_s coefficients.

Table B.6.: 95% HPD intervals of the ϕ_s coefficients (Cont.)

Country	ϕ_7	ϕ_8	ϕ_9	ϕ_{10}	ϕ_{11}	ϕ_{12}
Australia	[0.73; 1.14]	[0.83; 1.24]	[0.81; 1.18]	[0.80; 1.17]	[0.62; 0.98]	[0.71; 1.12]
Belgium	[0.73; 1.11]	[0.77; 1.13]	[0.72; 1.09]	[0.53; 0.92]	[0.74; 1.20]	[0.80; 1.25]
Canada	[0.86; 1.14]	[0.84; 1.12]	[0.70; 0.98]	[0.83; 1.16]	[0.75; 1.07]	[0.84; 1.17]
Denmark	[0.83; 1.01]	[0.91; 1.09]	[0.96; 1.14]	[0.95; 1.12]	[0.93; 1.10]	[0.92; 1.08]
Finland	[0.51; 1.25]	[0.40; 1.15]	[0.15; 0.95]	[0.50; 1.35]	[0.40; 1.25]	[0.25; 1.10]
France	[0.83; 1.07]	[0.92; 1.16]	[0.81; 1.04]	[0.87; 1.12]	[0.92; 1.16]	[0.84; 1.07]
Germany	[0.87; 1.06]	[0.95; 1.15]	[0.99; 1.18]	[0.94; 1.12]	[0.86; 1.03]	[0.86; 1.04]
Greece	[0.89; 1.22]	[1.02; 1.34]	[0.76; 1.02]	[0.94; 1.25]	[1.05; 1.37]	[0.77; 1.04]
Ireland	[0.90; 1.10]	[0.90; 1.05]	[0.90; 1.00]	[0.90; 1.05]	[0.90; 1.05]	[0.90; 1.10]
Italy	[0.75; 1.25]	[0.60; 0.90]	[0.25; 1.15]	[0.70; 1.20]	[0.70; 1.15]	[0.75; 1.20]
Japan	[0.79; 1.09]	[0.74; 1.04]	[0.82; 1.14]	[0.71; 1.02]	[0.79; 1.13]	[0.71; 1.06]
Netherl.	[0.80; 0.97]	[0.93; 1.11]	[0.91; 1.09]	[0.92; 1.09]	[0.93; 1.11]	[0.88; 1.06]
Norway	[0.94; 1.16]	[0.90; 1.11]	[0.77; 0.98]	[0.83; 1.06]	[0.88; 1.12]	[0.87; 1.10]
Spain	[0.94; 1.06]	[0.94; 1.04]	[0.96; 1.06]	[0.90; 1.04]	[0.91; 1.04]	[0.86; 1.06]
Sweden	[0.38; 0.90]	[0.66; 1.18]	[0.68; 1.22]	[0.62; 1.12]	[0.52; 1.04]	[0.88; 1.36]
UK	[0.92; 1.04]	[0.94; 1.07]	[0.97; 1.10]	[0.93; 1.05]	[0.91; 1.03]	[0.97; 1.10]
US	[0.92; 1.05]	[0.94; 1.07]	[0.92; 1.05]	[0.93; 1.06]	[0.91; 1.04]	[0.96; 1.09]

All HPD regions were computed using the respective model averaged posteriors of the ϕ_s coefficients.

C. Tables - chapter 5

Table C.1.: Testing for no periodicity

Series	F_{PAR} - stat.	p - val.
East-Germany	2.69	0.00
West-Germany	2.34	0.00
Baden-Wuerttemberg	0.93	0.51
Bavaria	6.22	0.00
Berlin	1.17	0.31
Brandenburg	1.79	0.06
Bremen	0.48	0.91
Hamburg	0.38	0.96
Hesse	2.48	0.01
Lower Saxony	2.93	0.00
Mecklenburg-Western Pomerania	2.15	0.02
North Rhine-Westphalia	1.76	0.06
Rhineland-Palatinate	2.64	0.00
Saarland	1.97	0.03
Saxony	2.89	0.00
Saxony-Anhalt	2.72	0.00
Schleswig-Holstein	2.11	0.02
Thuringia	2.03	0.03

'F_{PAR}': F-statistic to test the null of no periodicity,
i.e. $H_0 : \phi_s = \phi_S$, for $s = 1...S-1$.

Table C.2.: Evaluation of 12-months ahead forecasts

	BMA-PAR	SARMA$(p,\, q) \times (P,\, Q)_{12}$		PMEANS	BPAR
Series	*PMSE*	*PMSE*	*Model order*	*PMSE*	*PMSE*
East-Germany	0.3238	0.2932	$(2,\, 0) \times (2,\, 0)$	0.2497	0.2586
West-Germany	0.1995*	0.3173	$(2,\, 0) \times (1,\, 0)$	0.3547	0.3207
Baden-Wuerttemberg	0.2205	0.2451	$(2,\, 1) \times (1,\, 0)$	0.1883	0.1370*
Bavaria	0.3035*	0.4067	$(1,\, 1) \times (2,\, 0)$	0.7351	0.3958
Berlin	0.3835	0.3652*	$(1,\, 1) \times (1,\, 0)$	0.7728	0.3982
Brandenburg	0.2021	0.3349	$(1,\, 1) \times (2,\, 0)$	0.2514	0.3336
Bremen	0.2833*	0.4024	$(2,\, 0) \times (1,\, 0)$	0.3497	0.3367
Hamburg	0.0965*	0.2436	$(1,\, 1) \times (1,\, 0)$	0.1393	0.1484
Hesse	0.2897	0.3830	$(1,\, 1) \times (1,\, 0)$	0.2217*	0.2537
Lower Saxony	0.2826*	0.4021	$(1,\, 1) \times (1,\, 0)$	0.3944	0.3934
Mecklenburg-Western Pom.	0.2348	0.1408*	$(1,\, 1) \times (1,\, 0)$	0.5287	0.3998
North Rhine-Westphalia	0.3141*	0.3758	$(1,\, 1) \times (1,\, 0)$	0.3611	0.3557
Rhineland-Palatinate	0.1034*	0.4190	$(1,\, 1) \times (1,\, 1)$	0.4069	0.4026
Saarland	0.6974	0.6813	$(1,\, 1) \times (1,\, 0)$	0.5794*	0.5906
Saxony	0.2835	0.5534	$(1,\, 1) \times (2,\, 0)$	0.2822*	0.3013
Saxony-Anhalt	0.4367	0.2162*	$(1,\, 0) \times (2,\, 0)$	0.2446	0.2591
Schleswig-Holstein	0.1282	0.1165*	$(2,\, 0) \times (1,\, 0)$	0.2021	0.2117
Thuringia	0.4691*	0.7103	$(1,\, 0) \times (2,\, 1)$	0.6316	0.6545

'PMSE': Predictive Mean Squared Error , '$\star$': significant

Table C.3.: Evaluation of 12-months ahead forecasts (Cont.)

	BMA-PAR	SARMA$(p,\, q) \times (P,\, Q)_{12}$		PMEANS	BPAR
Series	*MAPE*	*MAPE*	*Model order*	*MAPE*	*MAPE*
East-Germany	0.0203	0.0220	$(2,\, 0) \times (2,\, 0)$	0.0153	0.0180
West-Germany	0.0229*	0.0452	$(2,\, 0) \times (1,\, 0)$	0.0510	0.0454
Baden-Wuerttemberg	0.0427	0.0461	$(2,\, 1) \times (1,\, 0)$	0.0379	0.0248*
Bavaria	0.0648*	0.0784	$(1,\, 1) \times (2,\, 0)$	0.1683	0.0877
Berlin	0.0240	0.0239*	$(1,\, 1) \times (1,\, 0)$	0.0498	0.0251
Brandenburg	0.0121	0.0266	$(1,\, 1) \times (2,\, 0)$	0.0133	0.0183
Bremen	0.0149*	0.0305	$(2,\, 0) \times (1,\, 0)$	0.0259	0.0232
Hamburg	0.0088*	0.0259	$(1,\, 1) \times (1,\, 0)$	0.0140	0.0151
Hesse	0.0429	0.0574	$(1,\, 1) \times (1,\, 0)$	0.0330*	0.0379
Lower Saxony	0.0312*	0.0517	$(1,\, 1) \times (1,\, 0)$	0.0480	0.0479
Mecklenburg-Western Pom.	0.0149	0.0077*	$(1,\, 1) \times (1,\, 0)$	0.0377	0.0271
North Rhine-Westphalia	0.0337*	0.0407	$(1,\, 1) \times (1,\, 0)$	0.0392	0.0385
Rhineland-Palatinate	0.0139*	0.0692	$(1,\, 1) \times (1,\, 1)$	0.0652	0.0635
Saarland	0.0816	0.0809	$(1,\, 1) \times (1,\, 0)$	0.0683*	0.0702
Saxony	0.0176	0.0486	$(1,\, 1) \times (2,\, 0)$	0.0209	0.0228
Saxony-Anhalt	0.0299	0.0159*	$(1,\, 0) \times (2,\, 0)$	0.0166	0.0179
Schleswig-Holstein	0.0149	0.0132*	$(2,\, 0) \times (1,\, 0)$	0.0241	0.0249
Thuringia	0.0456*	0.0691	$(1,\, 0) \times (2,\, 1)$	0.0630	0.0655

'MAPE': Mean Absolute Percentage Error, '$\star$': significant

Table C.4.: Bayesian one-year ahead forecasts (1)

	East Germany		*West Germany*	
Dates	BMA (s.e.)	95% - HPD	BMA (s.e.)	95% - HPD
2013(3)	12.49 (0.29)	[11.92; 13.10]	6.99 (0.20)	[6.61; 7.39]
2013(4)	11.92 (0.32)	[11.25; 12.54]	6.79 (0.23)	[6.35; 7.27]
2013(5)	11.34 (0.40)	[10.52; 12.15]	6.49 (0.22)	[6.06; 6.95]
2013(6)	11.02 (0.42)	[10.18; 11.89]	6.38 (0.22)	[5.94; 6.85]
2013(7)	11.36 (0.49)	[10.34; 12.39]	6.61 (0.28)	[6.06; 7.16]
2013(8)	11.23 (0.33)	[10.57; 11.90]	6.62 (0.27)	[6.11; 7.17]
2013(9)	10.78 (0.43)	[9.91; 11.66]	6.38 (0.32)	[5.73; 7.03]
2013(10)	10.47 (0.43)	[9.62; 11.33]	6.30 (0.44)	[5.44; 7.21]
2013(11)	10.50 (0.40)	[9.68; 11.31]	6.36 (0.44)	[5.46; 7.29]
2013(12)	11.11 (0.68)	[9.71; 12.48]	6.64 (0.52)	[5.56; 7.74]
2014(1)	12.89 (0.72)	[11.42; 14.35]	7.37 (0.52)	[6.19; 8.45]
2014(2)	13.05 (0.47)	[12.10; 14.00]	7.44 (0.49)	[6.39; 8.59]

'BMA': Mean of model averaged posterior predictive density, standard error in brackets.
'HPD': Highest posterior density region.

Table C.5.: Bayesian one-year ahead forecasts (2)

	Baden-Wuerttemberg		*Bavaria*	
Dates	BMA (s.e.)	95% - HPD	BMA (s.e.)	95% - HPD
2013(3)	4.66 (0.19)	[4.29; 5.03]	4.89 (0.26)	[4.37; 5.38]
2013(4)	4.53 (0.22)	[4.11; 4.96]	4.49 (0.28)	[3.96; 5.08]
2013(5)	4.30 (0.22)	[3.87; 4.72]	4.08 (0.27)	[3.58; 4.64]
2013(6)	4.18 (0.27)	[3.64; 4.73]	3.94 (0.27)	[3.42; 4.49]
2013(7)	4.34 (0.32)	[3.70; 4.99]	4.05 (0.32)	[3.43; 4.69]
2013(8)	4.47 (0.27)	[3.92; 5.02]	4.27 (0.28)	[3.73; 4.83]
2013(9)	4.26 (0.26)	[3.74; 4.79]	4.08 (0.34)	[3.38; 4.78]
2013(10)	4.16 (0.34)	[3.49; 4.85]	3.98 (0.45)	[3.13; 4.94]
2013(11)	4.16 (0.36)	[3.43; 4.93]	4.11 (0.51)	[3.10; 5.21]
2013(12)	4.30 (0.36)	[3.56; 5.04]	4.56 (0.77)	[2.95; 6.22]
2014(1)	4.71 (0.34)	[3.95; 5.37]	5.60 (0.73)	[3.92; 7.04]
2014(2)	4.71 (0.37)	[3.93; 5.51]	5.62 (0.52)	[4.59; 6.85]

'BMA': Mean of model averaged posterior predictive density, standard error in brackets.
'HPD': Highest posterior density region.

Table C.6.: Bayesian one-year ahead forecasts (3)

	Berlin		*Brandenburg*	
Dates	BMA (s.e.)	95% - HPD	BMA (s.e.)	95% - HPD
2013(3)	14.17 (0.27)	[13.62; 14.71]	11.80 (0.33)	[11.16; 12.49]
2013(4)	14.05 (0.31)	[13.44; 14.66]	11.24 (0.40)	[10.42; 12.04]
2013(5)	13.62 (0.30)	[12.99; 14.21]	10.56 (0.41)	[9.73; 11.39]
2013(6)	13.30 (0.38)	[12.50; 14.02]	10.37 (0.50)	[9.38; 11.36]
2013(7)	13.57 (0.31)	[12.95; 14.19]	10.79 (0.53)	[9.71; 11.90]
2013(8)	13.51 (0.31)	[12.89; 14.13]	10.57 (0.39)	[9.77; 11.35]
2013(9)	13.12 (0.38)	[12.34; 13.85]	10.11 (0.44)	[9.23; 11.01]
2013(10)	12.89 (0.41)	[12.08; 13.70]	9.88 (0.44)	[9.01; 10.75]
2013(11)	12.75 (0.41)	[11.89; 13.56]	9.91 (0.42)	[9.06; 10.76]
2013(12)	12.98 (0.41)	[12.17; 13.79]	10.61 (0.80)	[9.01; 12.24]
2014(1)	14.01 (0.74)	[12.41; 15.61]	12.42 (0.71)	[11.01; 13.88]
2014(2)	14.05 (0.35)	[13.33; 14.72]	12.59 (0.46)	[11.64; 13.54]

'BMA': Mean of model averaged posterior predictive density, standard error in brackets.
'HPD': Highest posterior density region.

Table C.7.: Bayesian one-year ahead forecasts (4)

	Bremen		*Hamburg*	
Dates	BMA (s.e.)	95% - HPD	BMA (s.e.)	95% - HPD
2013(3)	12.69 (0.41)	[11.91; 13.53]	8.65 (0.37)	[7.92; 9.71]
2013(4)	12.67 (0.50)	[11.67; 13.65]	8.61 (0.49)	[7.63; 9.69]
2013(5)	12.36 (0.45)	[11.45; 13.24]	8.39 (0.48)	[7.57; 9.79]
2013(6)	12.23 (0.48)	[11.24; 13.15]	8.26 (0.37)	[7.55; 9.21]
2013(7)	12.51 (0.49)	[11.54; 13.48]	8.49 (0.42)	[7.68; 9.71]
2013(8)	12.43 (0.56)	[11.30; 13.57]	8.39 (0.42)	[7.55; 9.62]
2013(9)	12.10 (0.62)	[10.88; 13.39]	8.19 (0.36)	[7.54; 9.07]
2013(10)	12.02 (0.67)	[10.72; 13.50]	8.09 (0.39)	[7.21; 8.86]
2013(11)	11.93 (0.65)	[10.64; 13.31]	8.03 (0.44)	[7.14; 9.25]
2013(12)	12.09 (0.74)	[10.62; 13.68]	8.15 (0.49)	[7.07; 9.20]
2014(1)	13.06 (1.39)	[10.51; 16.68]	8.72 (0.41)	[7.79; 9.58]
2014(2)	13.16 (0.70)	[11.71; 14.71]	8.70 (0.36)	[7.94; 9.46]

'BMA': Mean of model averaged posterior predictive density, standard error in brackets.
'HPD': Highest posterior density region.

Table C.8.: Bayesian one-year ahead forecasts (5)

	Hesse		*Lower Saxony*	
Dates	BMA (s.e.)	95% - HPD	BMA (s.e.)	95% - HPD
2013(3)	6.77 (0.19)	[6.39; 7.16]	7.58 (0.25)	[6.94; 7.96]
2013(4)	6.66 (0.22)	[6.22; 7.09]	7.25 (0.30)	[6.51; 7.69]
2013(5)	6.37 (0.21)	[5.96; 6.78]	6.79 (0.28)	[6.07; 7.20]
2013(6)	6.28 (0.22)	[5.85; 6.71]	6.71 (0.35)	[5.80; 7.21]
2013(7)	6.53 (0.26)	[6.02; 7.04]	7.00 (0.29)	[6.26; 7.39]
2013(8)	6.44 (0.23)	[5.98; 6.91]	6.86 (0.31)	[6.16; 7.38]
2013(9)	6.18 (0.28)	[5.61; 6.74]	6.55 (0.33)	[5.73; 7.06]
2013(10)	6.12 (0.35)	[5.39; 6.83]	6.45 (0.37)	[5.52; 7.04]
2013(11)	6.08 (0.32)	[5.41; 6.72]	7.54 (0.41)	[5.51; 7.21]
2013(12)	6.28 (0.37)	[5.49; 7.02]	6.95 (0.60)	[5.43; 7.94]
2014(1)	6.89 (0.42)	[6.01; 7.78]	7.81 (0.72)	[5.94; 9.04]
2014(2)	6.96 (0.36)	[6.26; 7.75]	7.84 (0.45)	[6.59; 8.49]

'BMA': Mean of model averaged posterior predictive density, standard error in brackets.
'HPD': Highest posterior density region.

Table C.9.: Bayesian one-year ahead forecasts (6)

	Mecklenburg-Western Pomerania		*North Rhine-Westphalia*	
Dates	BMA (s.e.)	95% - HPD	BMA (s.e.)	95% - HPD
2013(3)	14.37 (0.36)	[13.66; 15.08]	9.24 (0.19)	[8.87; 9.61]
2013(4)	13.42 (0.41)	[12.62; 14.26]	9.10 (0.21)	[8.69; 9.51]
2013(5)	12.49 (0.43)	[11.63; 13.36]	8.87 (0.21)	[8.46; 9.28]
2013(6)	11.89 (0.47)	[10.93; 12.84]	8.79 (0.21)	[8.38; 9.20]
2013(7)	12.10 (0.54)	[11.06; 13.20]	9.04 (0.27)	[8.51; 9.59]
2013(8)	12.60 (0.52)	[11.49; 13.71]	8.94 (0.25)	[8.45; 9.46]
2013(9)	11.44 (0.50)	[10.41; 12.41]	8.70 (0.27)	[8.27; 9.25]
2013(10)	11.21 (0.47)	[10.24; 12.15]	8.64 (0.31)	[8.03; 9.25]
2013(11)	11.65 (0.43)	[10.78; 12.52]	8.63 (0.33)	[7.96; 9.28]
2013(12)	12.53 (0.60)	[11.31 ; 13.73]	8.51 (0.39)	[8.00; 9.62]
2014(1)	14.51 (0.90)	[12.69; 16.39]	9.36 (0.39)	[8.56; 10.16]
2014(2)	14.68 (0.55)	[13.54; 15.78]	9.49 (0.55)	[8.39; 10.67]

'BMA': Mean of model averaged posterior predictive density, standard error in brackets.
'HPD': Highest posterior density region.

Table C.10.: Bayesian one-year ahead forecasts (7)

	Rhineland-Palatinate		*Saarland*	
Dates	BMA (s.e.)	95% - HPD	BMA (s.e.)	95% - HPD
2013(3)	6.26 (0.26)	[5.73; 6.71]	8.09 (0.23)	[7.64; 8.53]
2013(4)	5.99 (0.30)	[5.35; 6.55]	7.92 (0.29)	[7.35; 8.50]
2013(5)	5.65 (0.27)	[5.11; 6.17]	7.64 (0.24)	[7.17; 8.12]
2013(6)	5.54 (0.30)	[5.00; 6.12]	7.54 (0.25)	[7.04; 8.03]
2013(7)	6.79 (0.32)	[5.16; 6.44]	7.79 (0.32)	[7.15; 8.44]
2013(8)	5.69 (0.33)	[5.03; 6.36]	7.71 (0.33)	[7.07; 8.35]
2013(9)	5.40 (0.37)	[4.66; 6.17]	7.41 (0.27)	[6.88; 7.96]
2013(10)	5.32 (0.43)	[4.48; 6.32]	7.28 (0.35)	[6.56; 7.98]
2013(11)	5.38 (0.48)	[4.39; 6.41]	7.28 (0.31)	[6.65; 7.90]
2013(12)	5.74 (0.56)	[4.54; 6.89]	7.50 (0.41)	[6.66; 8.35]
2014(1)	6.47 (0.52)	[5.27; 7.58]	8.10 (0.41)	[7.20; 8.92]
2014(2)	6.47 (0.44)	[5.56; 7.41]	8.13 (0.46)	[7.15; 9.09]

'BMA': Mean of model averaged posterior predictive density, standard error in brackets.
'HPD': Highest posterior density region.

Table C.11.: Bayesian one-year ahead forecasts (8)

	Saxony		*Saxony-Anhalt*	
Dates	BMA (s.e.)	95% - HPD	BMA (s.e.)	95% - HPD
2013(3)	11.56 (0.36)	[10.82; 12.28]	13.15 (0.35)	[12.44; 13.84]
2013(4)	10.90 (0.38)	[10.10; 11.65]	12.55 (0.39)	[11.78; 13.32]
2013(5)	10.26 (0.45)	[9.35; 11.20]	11.98 (0.44)	[11.10; 12.88]
2013(6)	9.80 (0.50)	[8.76; 10.78]	11.59 (0.49)	[10.58; 12.61]
2013(7)	10.10 (0.63)	[8.85; 11.43]	11.91 (0.52)	[10.92; 12.99]
2013(8)	10.00 (0.40)	[9.19; 10.78]	11.69 (0.39)	[10.91; 12.49]
2013(9)	9.50 (0.50)	[8.49; 10.51]	11.08 (0.46)	[10.15; 12.02]
2013(10)	9.09 (0.46)	[8.17; 10.02]	10.57 (0.54)	[9.47; 11.66]
2013(11)	9.12 (0.41)	[8.31; 9.94]	10.53 (0.41)	[9.70; 11.35]
2013(12)	9.77 (0.76)	[8.26; 11.30]	11.19 (0.81)	[9.58; 12.81]
2014(1)	11.55 (0.83)	[9.90; 13.33]	13.14 (0.68)	[11.82; 14.51]
2014(2)	11.77 (0.57)	[10.56; 12.92]	13.26 (0.47)	[12.31; 14.17]

'BMA': Mean of model averaged posterior predictive density, standard error in brackets.
'HPD': Highest posterior density region.

Table C.12.: Bayesian one-year ahead forecasts (9)

	Schleswig-Holstein		*Thuringia*	
Dates	BMA (s.e.)	95% - HPD	BMA (s.e.)	95% - HPD
2013(3)	8.04 (0.23)	[7.58; 8.49]	9.78 (0.40)	[8.99; 10.60]
2013(4)	7.54 (0.28)	[6.98; 8.11]	9.00 (0.44)	[8.11; 9.88]
2013(5)	7.07 (0.25)	[6.58; 7.58]	8.38 (0.49)	[7.39; 9.36]
2013(6)	6.90 (0.33)	[6.23; 7.54]	7.89 (0.56)	[6.77; 9.04]
2013(7)	7.15 (0.26)	[6.61; 7.68]	8.28 (0.57)	[7.19; 9.51]
2013(8)	7.05 (0.29)	[6.48; 7.65]	8.19 (0.41)	[7.32; 9.04]
2013(9)	6.75 (0.34)	[6.06; 7.43]	7.64 (0.47)	[6.67; 8.59]
2013(10)	6.69 (0.40)	[5.86; 7.51]	7.26 (0.58)	[6.07; 8.45]
2013(11)	6.93 (0.47)	[5.95; 7.87]	7.28 (0.51)	[6.25; 8.32]
2013(12)	7.32 (0.52)	[6.26; 8.36]	8.07 (0.95)	[6.17; 9.95]
2014(1)	8.14 (0.80)	[6.38; 9.87]	9.95 (1.22)	[7.71; 12.41]
2014(2)	8.18 (0.41)	[7.32; 9.02]	10.10 (0.62)	[8.79; 11.37]

'BMA': Mean of model averaged posterior predictive density, standard error in brackets.
'HPD': Highest posterior density region.

D. Figures - chapter 2

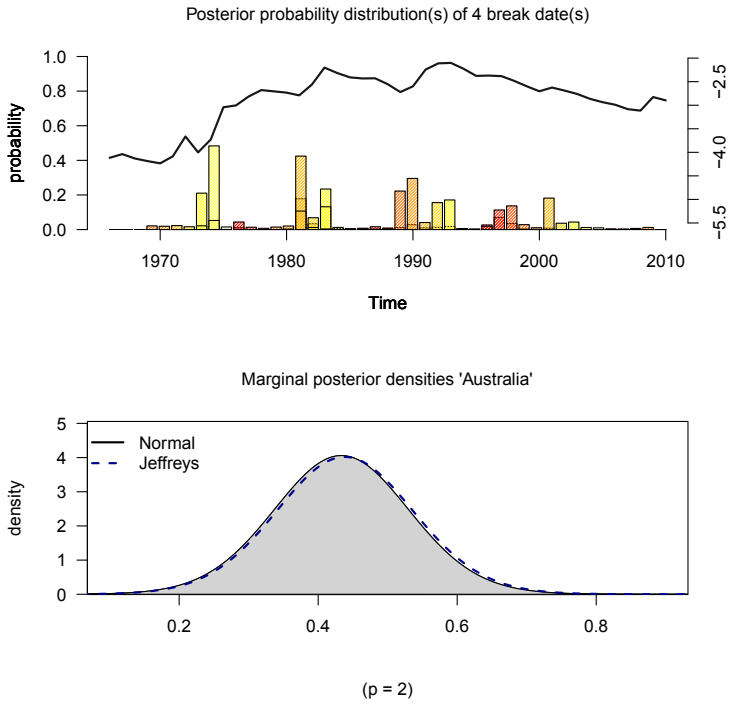


Figure D.1.: Series Australia: posterior densities for the break dates and the long run coefficient.

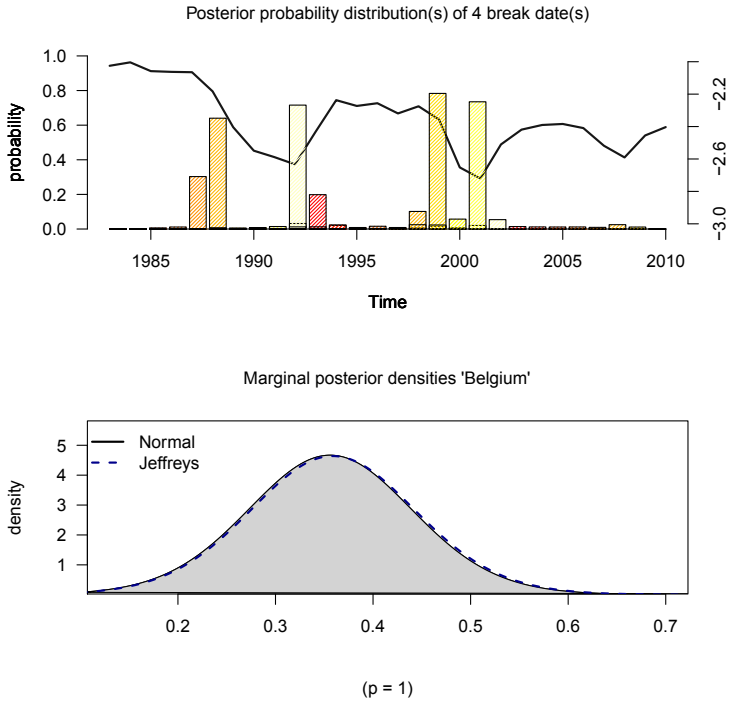


Figure D.2.: Series Belgium: posterior densities for the break dates and the long run coefficient.

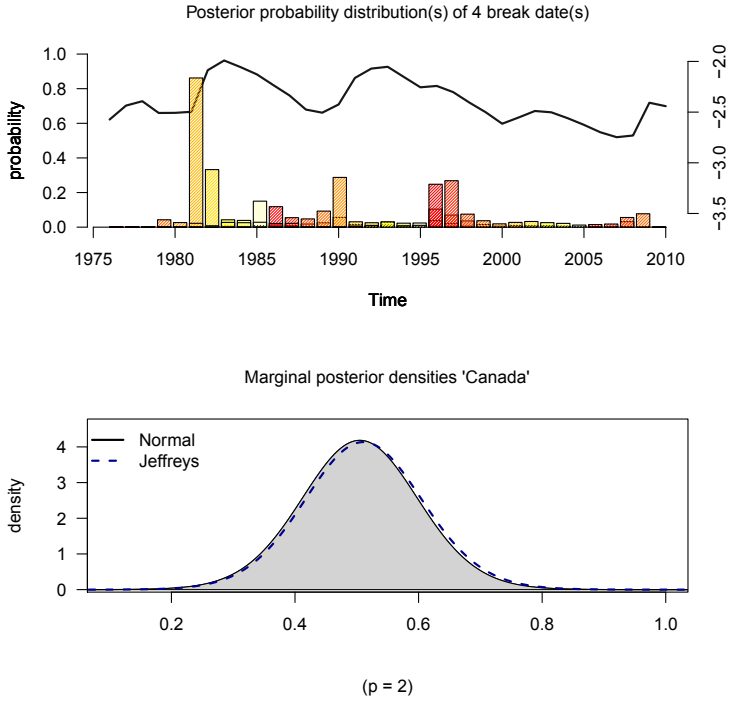


Figure D.3.: Series Canada: posterior densities for the break dates and the long run coefficient.

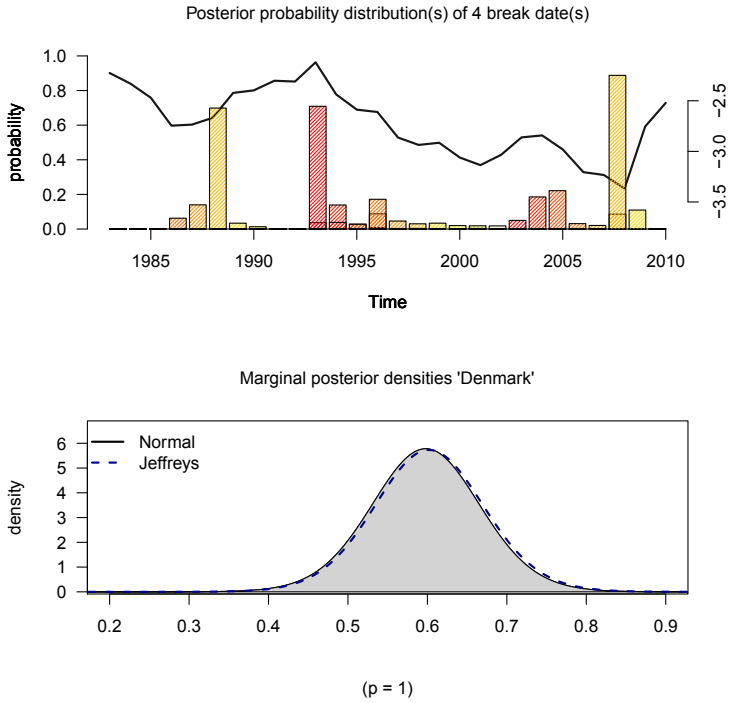


Figure D.4.: Series Denmark: posterior densities for the break dates and the long run coefficient.

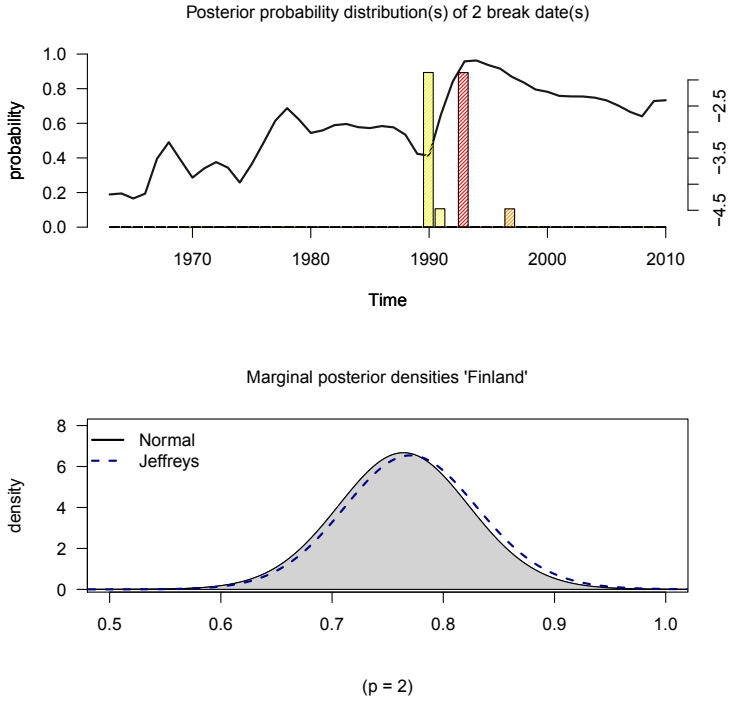


Figure D.5.: Series Finland: posterior densities for the break dates and the long run coefficient.

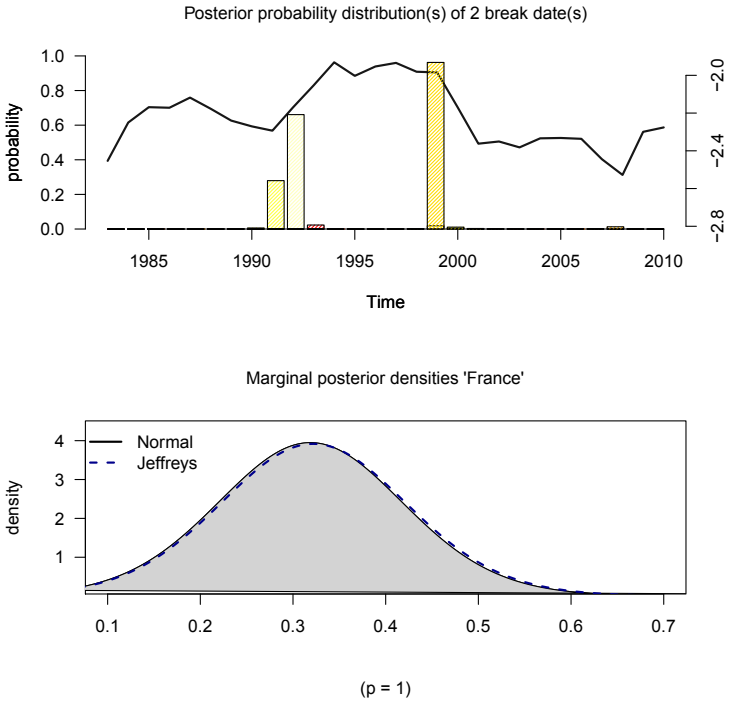


Figure D.6.: Series France: posterior densities for the break dates and the long run coefficient.

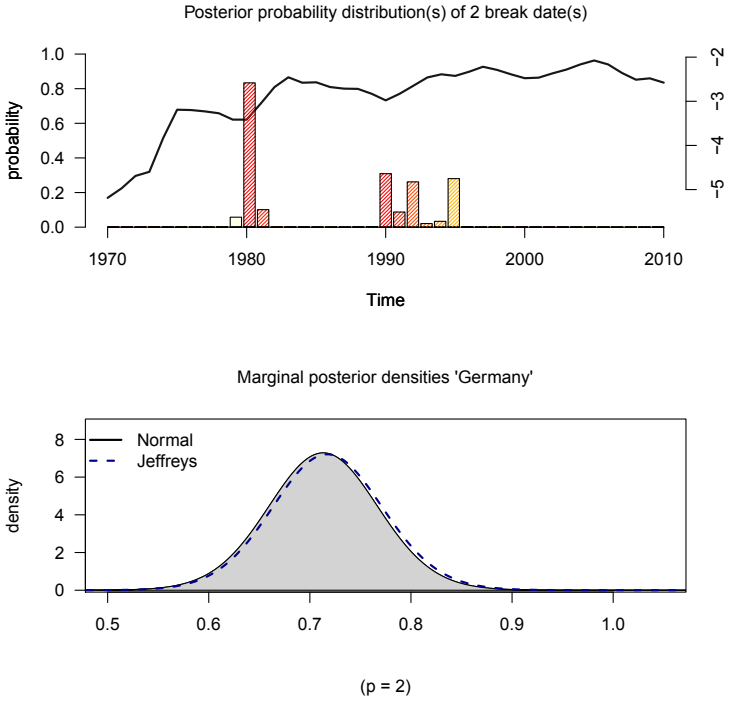


Figure D.7.: Series Germany: posterior densities for the break dates and the long run coefficient.

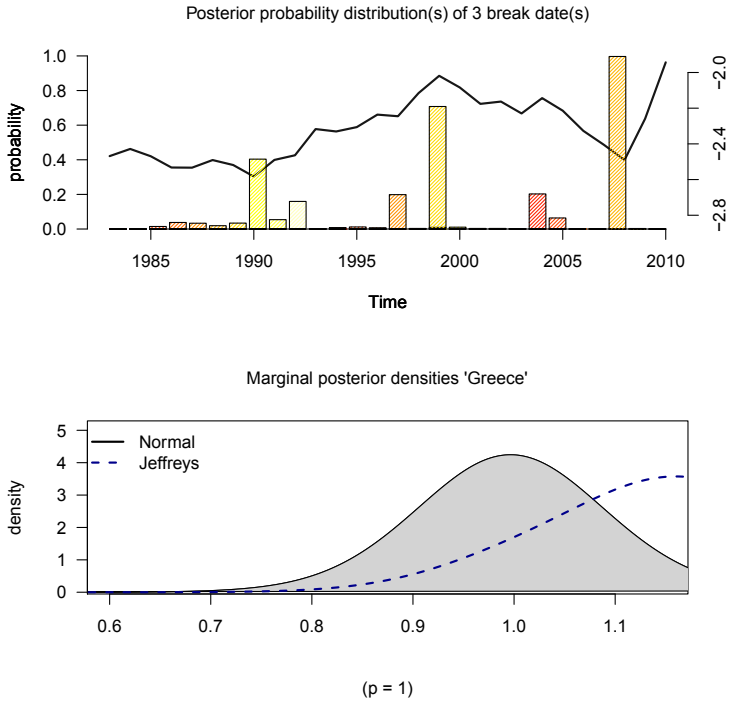


Figure D.8.: Series Greece: posterior densities for the break dates and the long run coefficient.

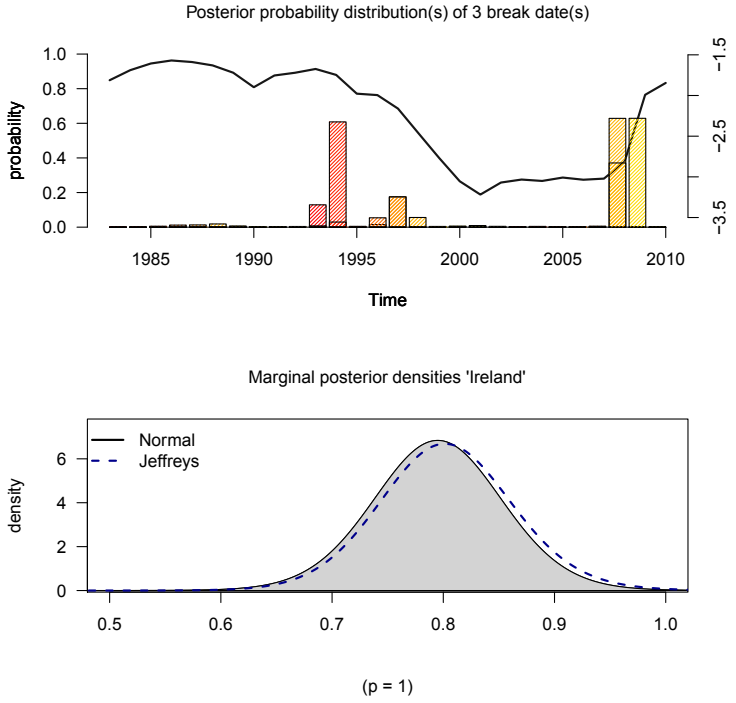


Figure D.9.: Series Ireland: posterior densities for the break dates and the long run coefficient.

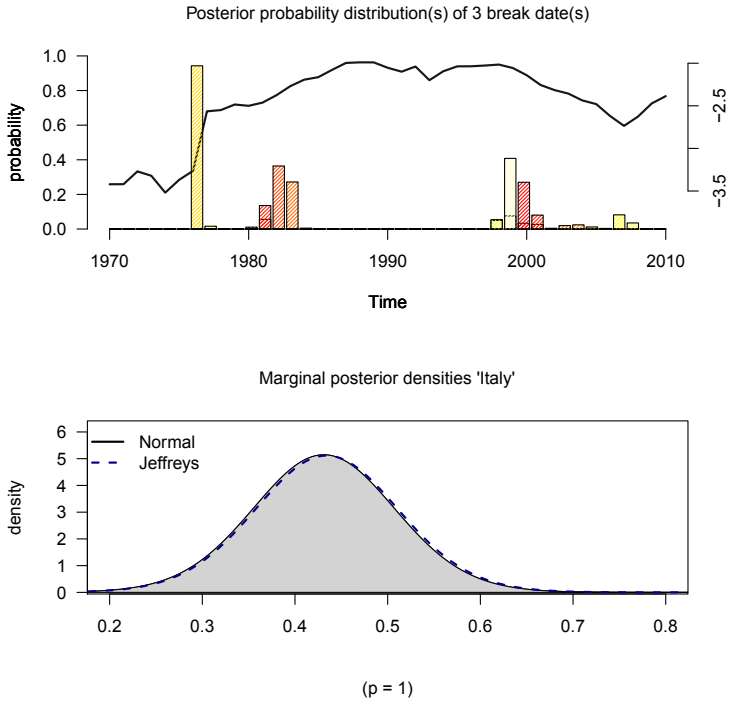


Figure D.10.: Series Italy: posterior densities for the break dates and the long run coefficient.

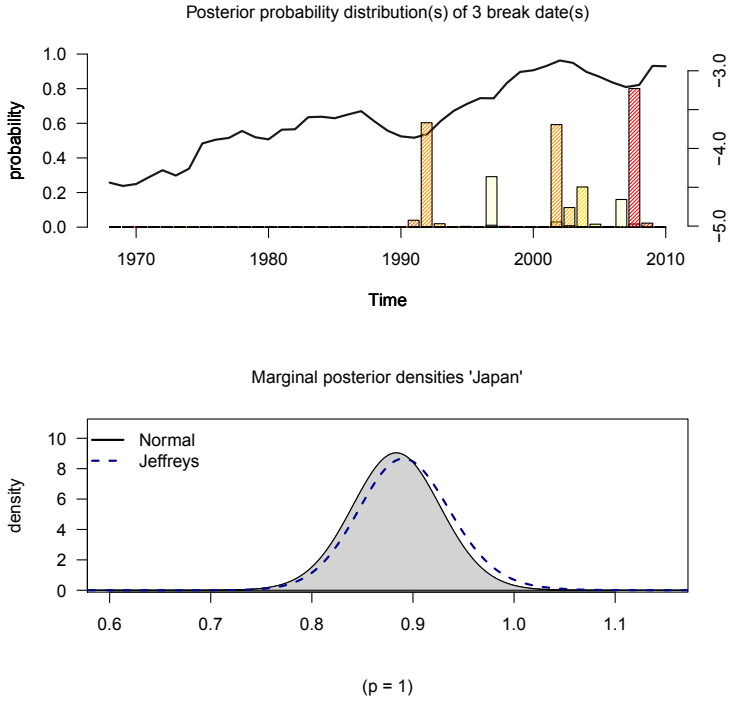


Figure D.11.: Series Japan: posterior densities for the break dates and the long run coefficient.

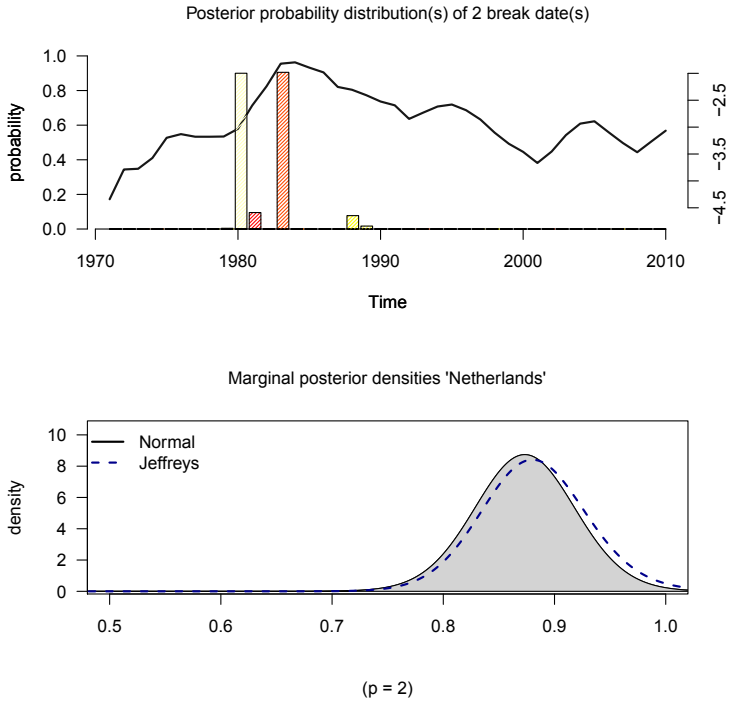


Figure D.12.: Series Netherlands: posterior densities for the break dates and the long run coefficient.

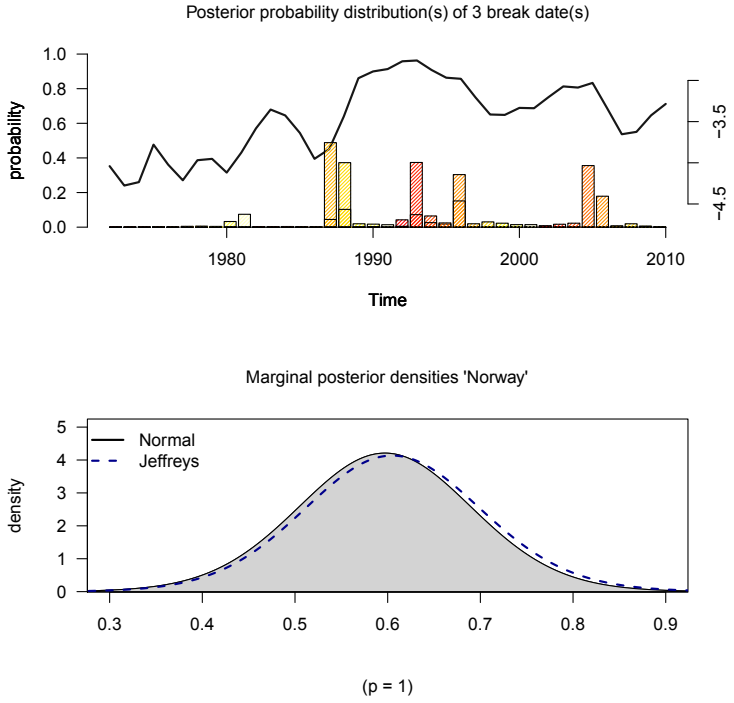


Figure D.13.: Series Norway: posterior densities for the break dates and the long run coefficient.

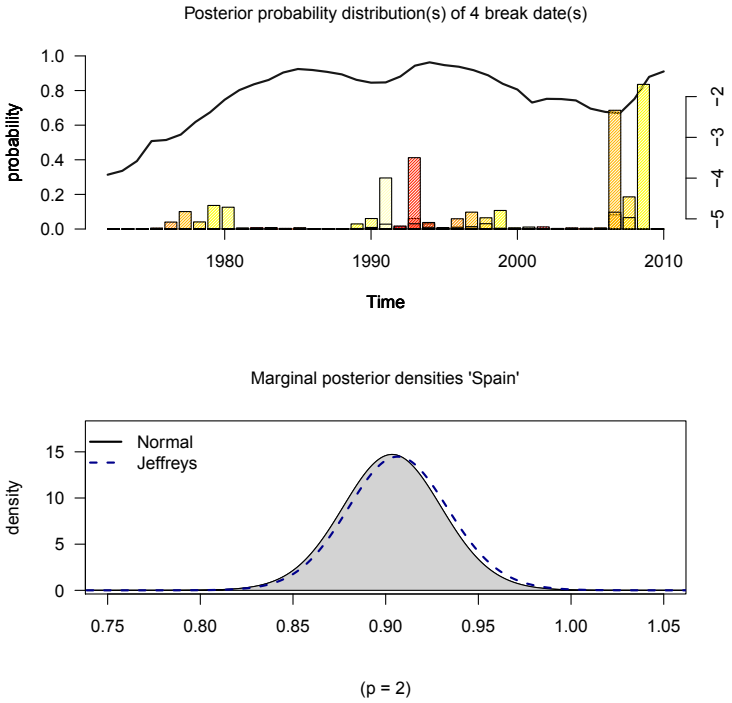


Figure D.14.: Series Spain: posterior densities for the break dates and the long run coefficient.

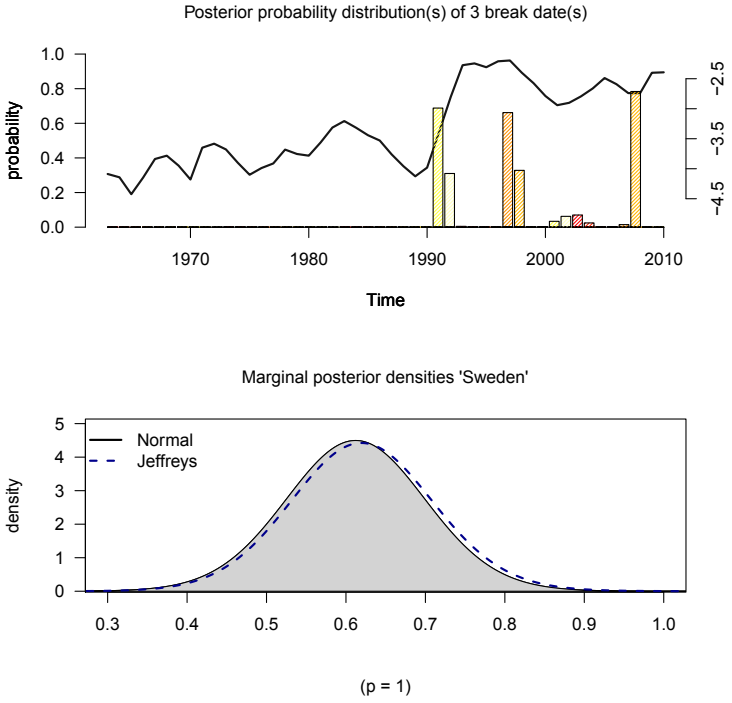


Figure D.15.: Series Sweden: posterior densities for the break dates and the long run coefficient.

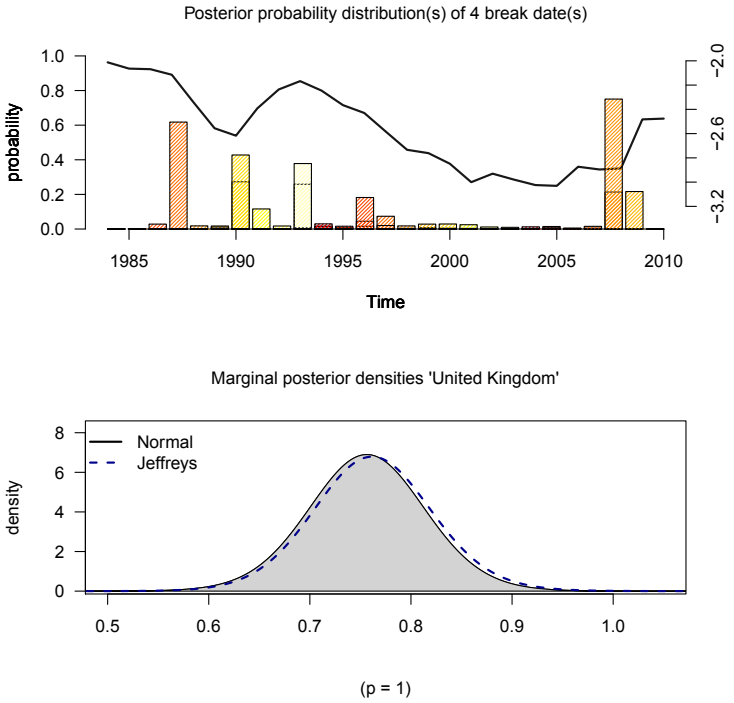


Figure D.16.: Series UK: posterior densities for the break dates and the long run coefficient.

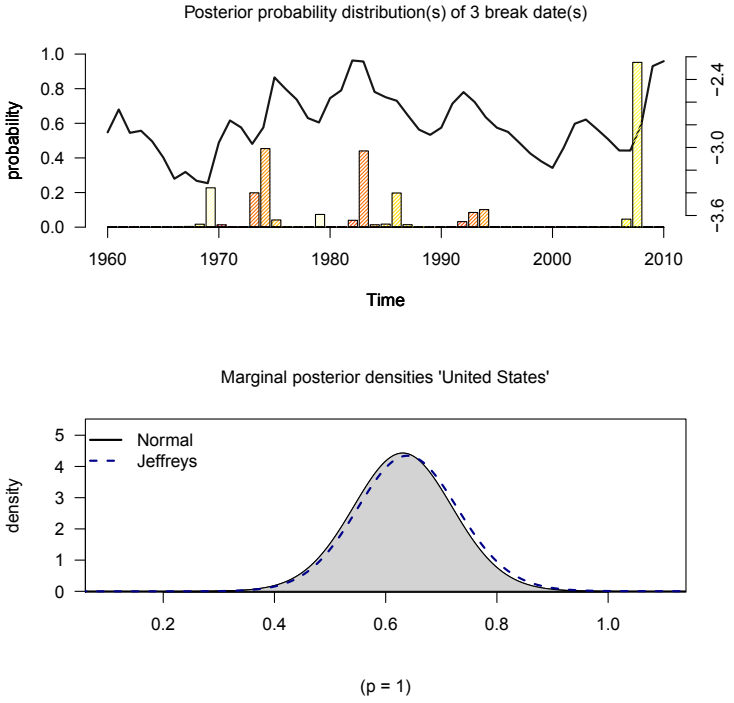


Figure D.17.: Series US: posterior densities for the break dates and the long run coefficient.

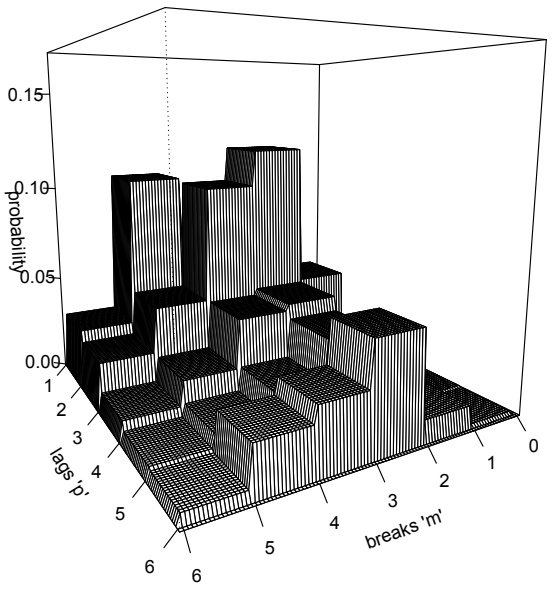


Figure D.18.: Helicopter tour Germany: joint posterior mass function of break number and lags.

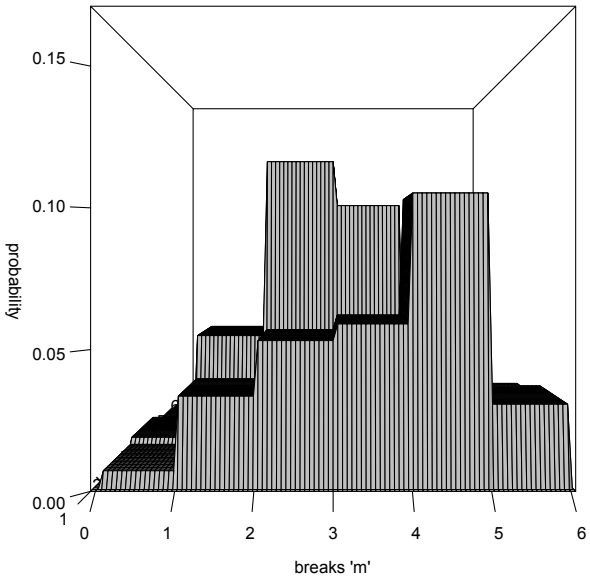
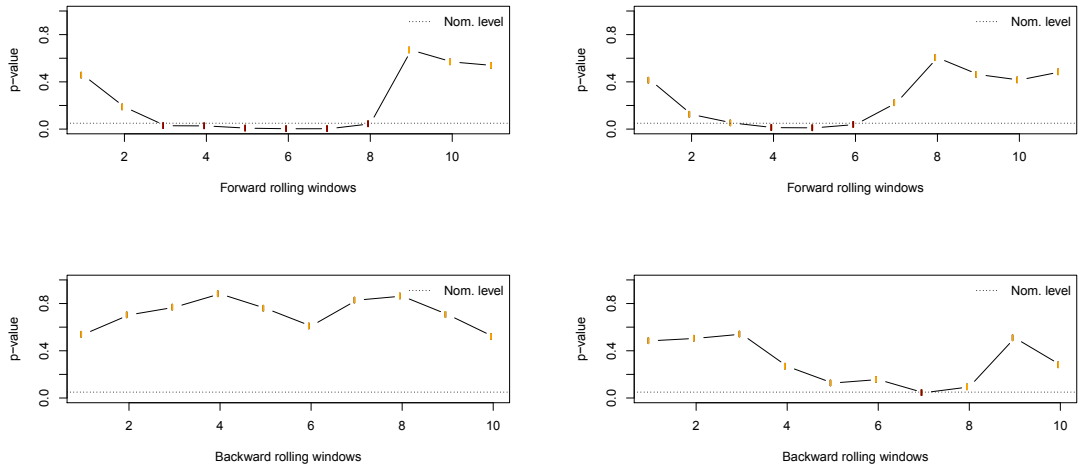


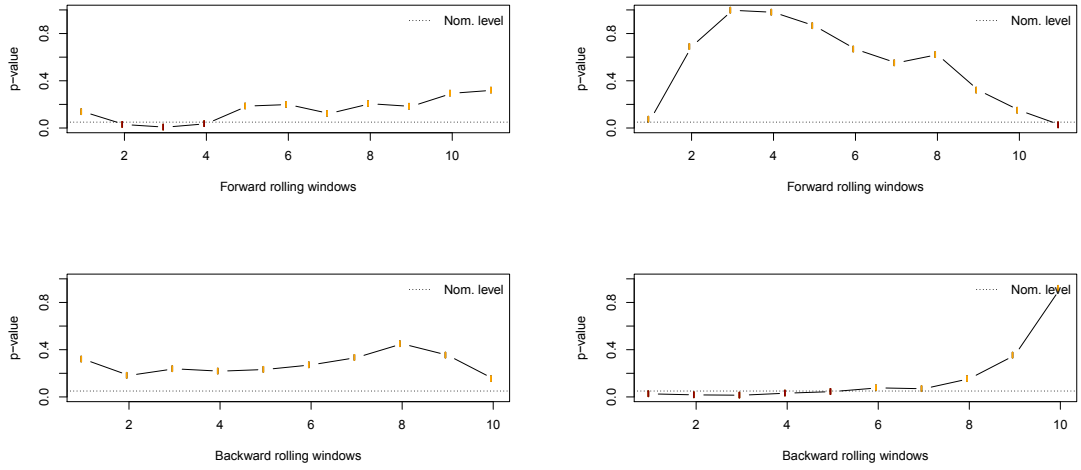
Figure D.19.: Helicopter tour Germany (2): joint posterior mass function of break number and lags.

E. Figures - chapter 4



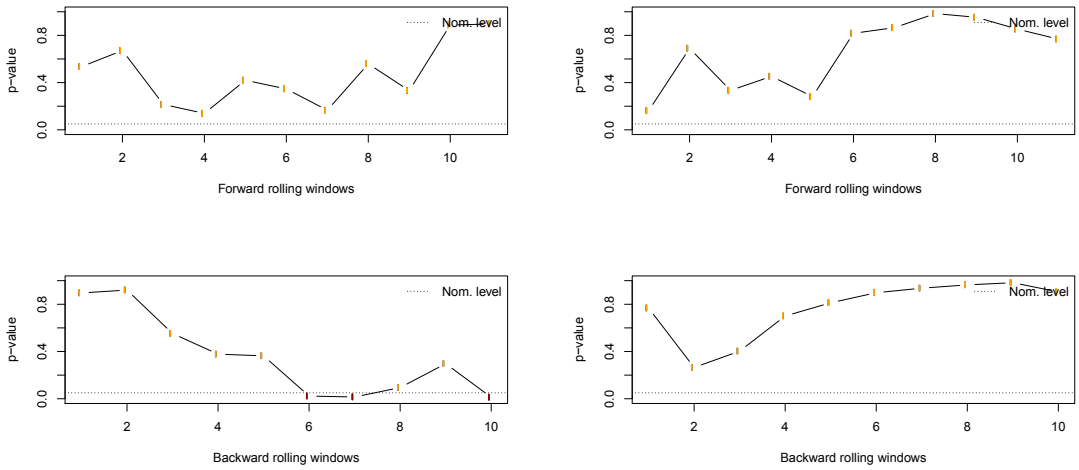
(a) Australia (b) Belgium

Figure E.20.: Bayesian p-values of recursive F-tests of non-periodicity: Australia and Belgium.



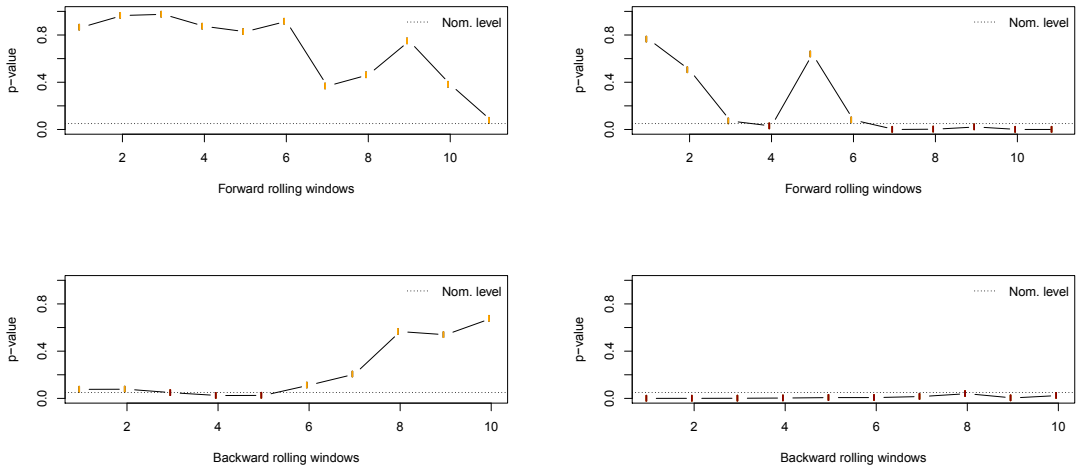
(a) Canada (b) Denmark

Figure E.21.: Bayesian p-values of recursive F-tests of non-periodicity: Canada and Denmark.



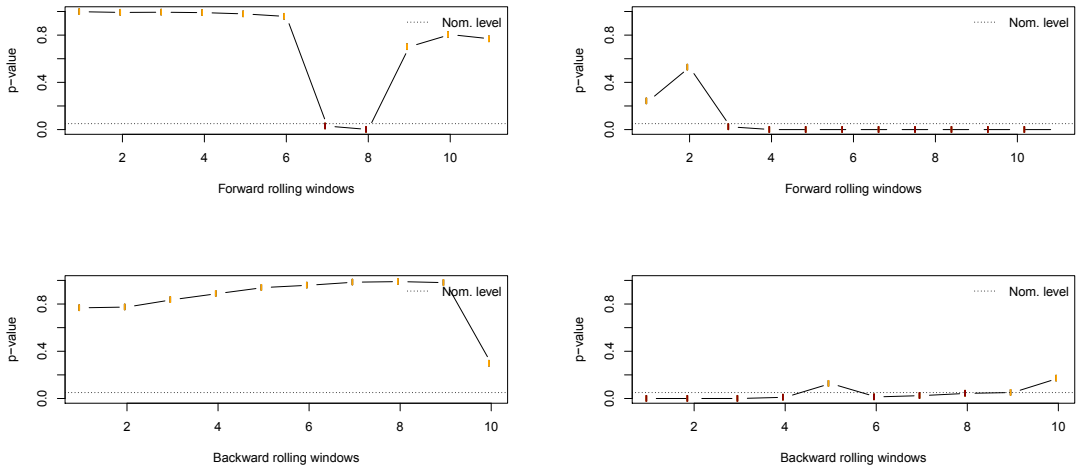
(a) Finland (b) France

Figure E.22.: Bayesian p-values of recursive F-tests of non-periodicity: Finland and France.



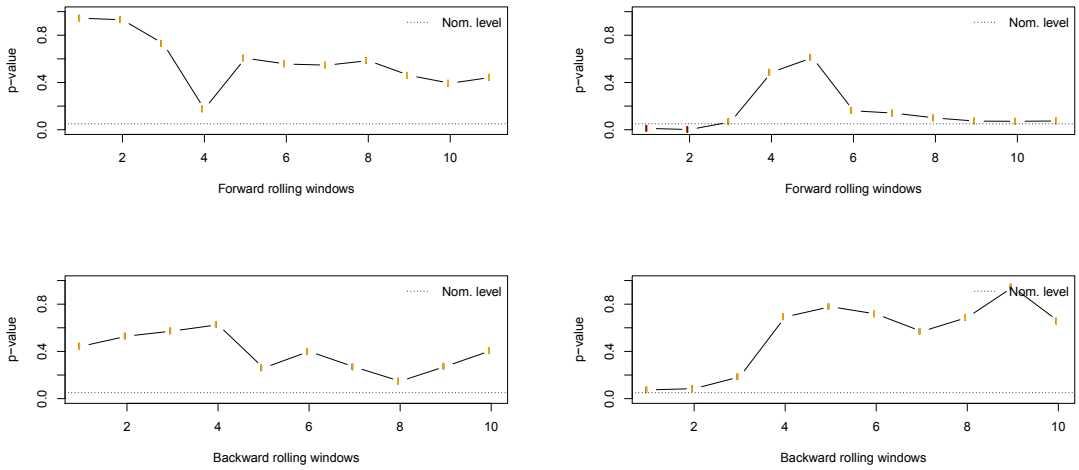
(a) Germany (b) Greece

Figure E.23.: Bayesian p-values of recursive F-tests of non-periodicity: Germany and Greece.



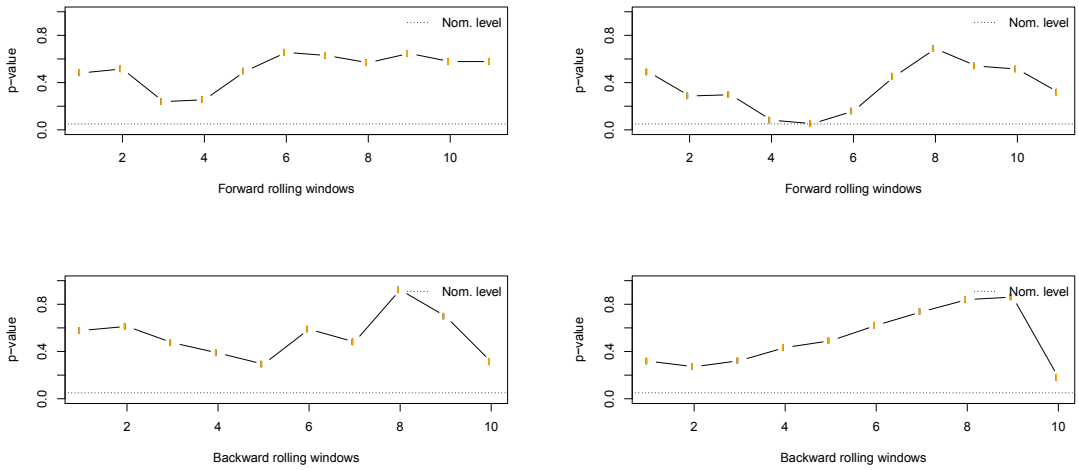
(a) Ireland (b) Italy

Figure E.24.: Bayesian p-values of recursive F-tests of non-periodicity: Ireland and Italy.



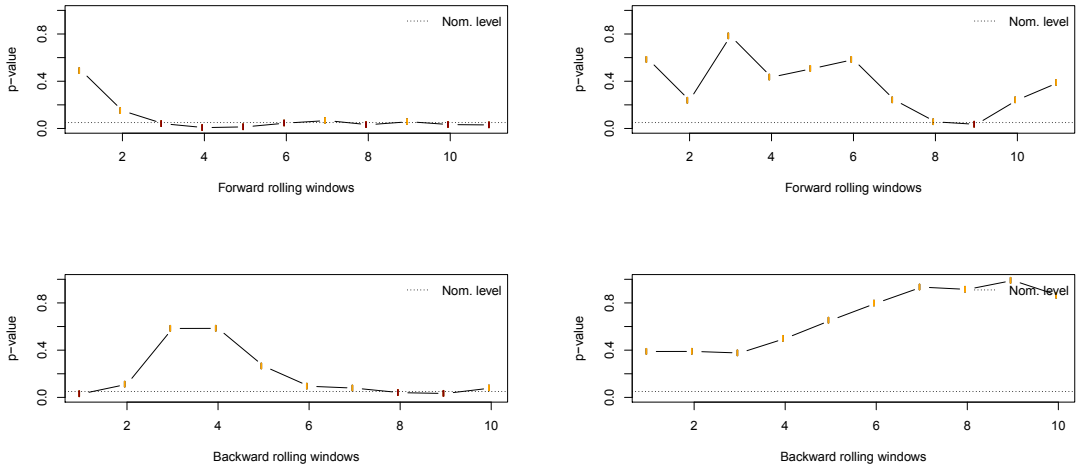
(a) Japan (b) Netherlands

Figure E.25.: Bayesian p-values of recursive F-tests of non-periodicity: Japan and Netherlands.



(a) Norway (b) Spain

Figure E.26.: Bayesian p-values of recursive F-tests of non-periodicity: Norway and Spain.



(a) Sweden (b) Great Britain

Figure E.27.: Bayesian p-values of recursive F-tests of non-periodicity: Sweden and Great Britain.

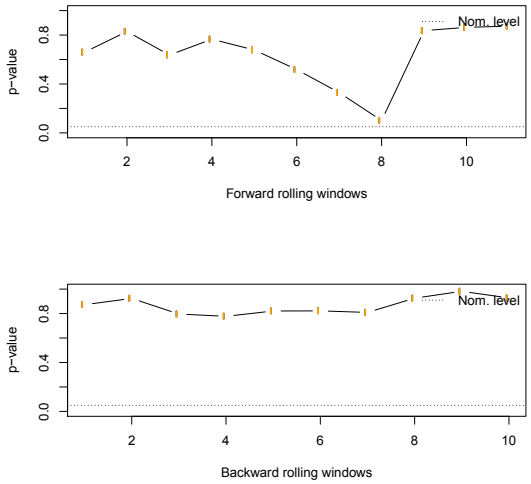


Figure E.28.: Bayesian p-values of recursive F-tests of non-periodicity: USA.

F. Figures - chapter 5

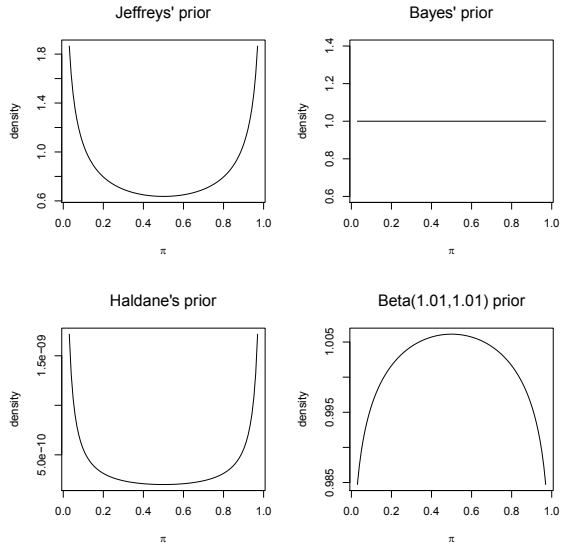


Figure F.29.: Used prior distributions for the Bayesian sign test.

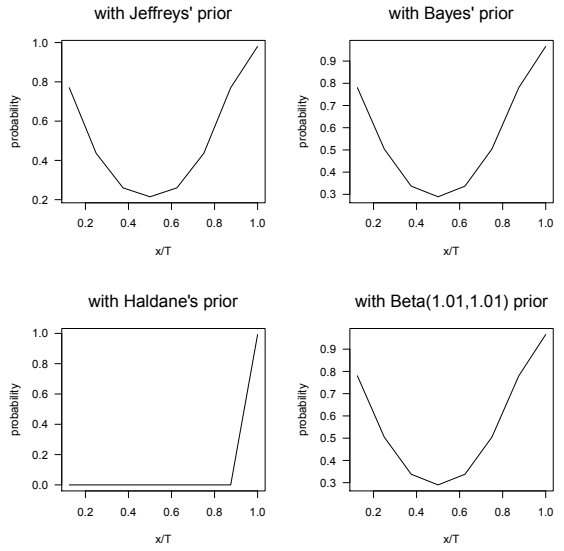


Figure F.30.: Posterior probability of H_1 as a function of x for $T = 8$.

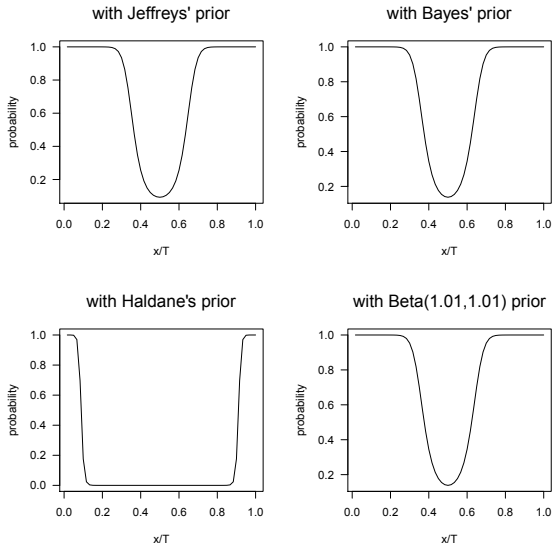


Figure F.31.: Posterior probability of H_1 as a function of x for $T = 60$.

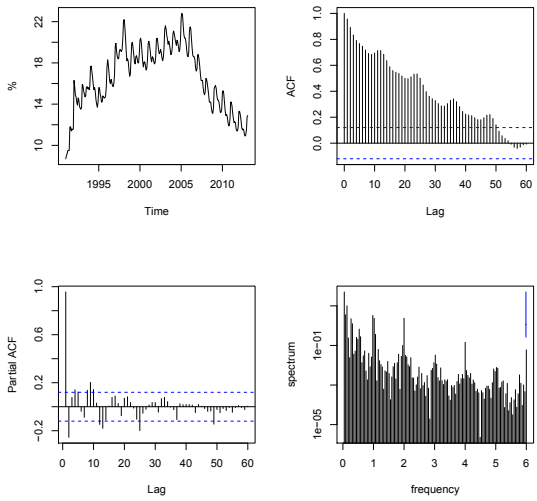


Figure F.32.: Series East-Germany (01/1991-02/2013) with S(P)ACF and periodogram.

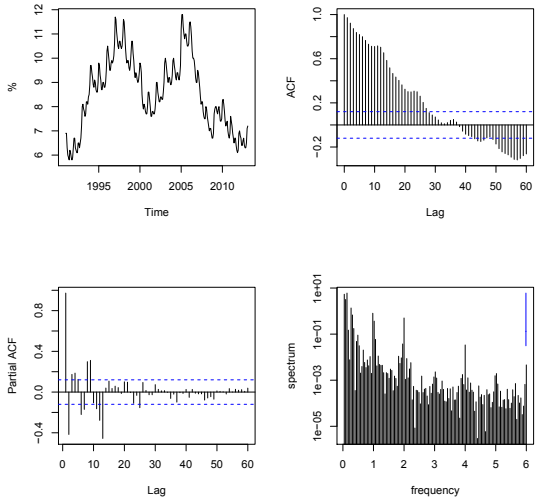


Figure F.33.: Series West-Germany (01/1991-02/2013) with S(P)ACF and periodogram.

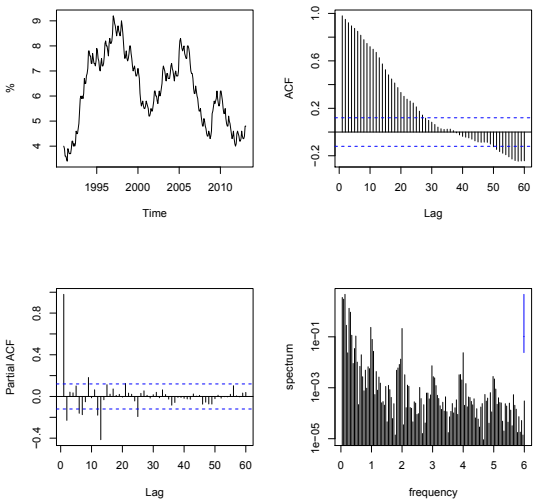


Figure F.34.: Series Baden-Wuerttemberg (01/1991-02/2013) with S(P)ACF and periodogram.

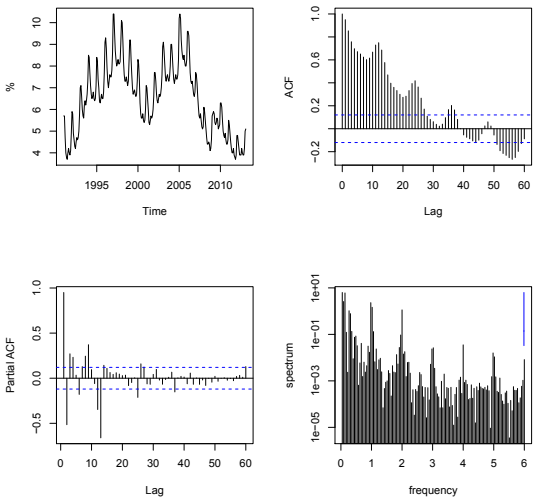


Figure F.35.: Series Bavaria (01/1991-02/2013) with S(P)ACF and periodogram.

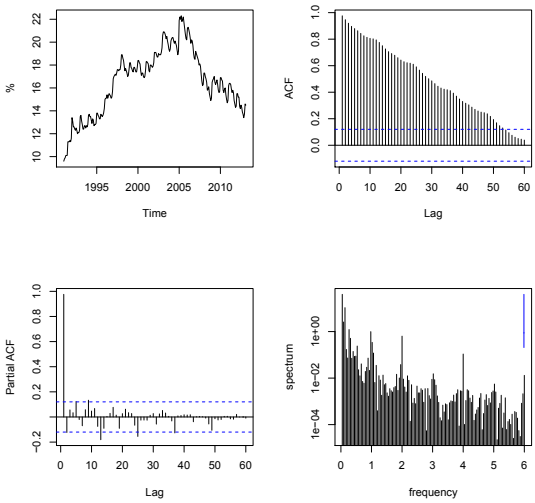


Figure F.36.: Series Berlin (01/1991-02/2013) with S(P)ACF and periodogram.

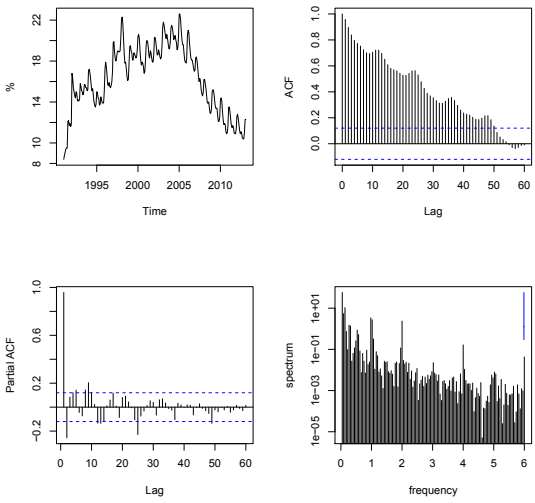


Figure F.37.: Series Brandenburg (01/1991-02/2013) with S(P)ACF and periodogram.

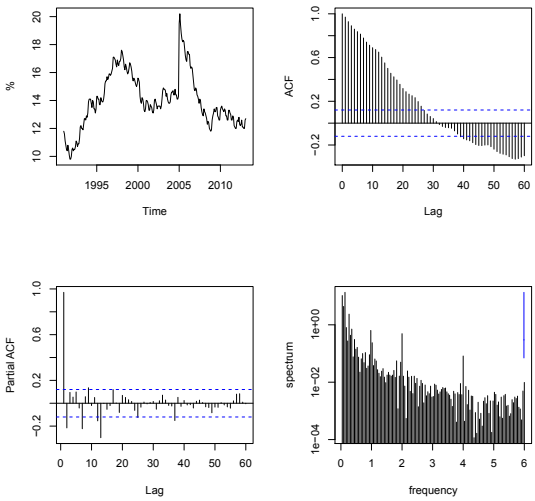


Figure F.38.: Series Bremen (01/1991-02/2013) with S(P)ACF and periodogram.

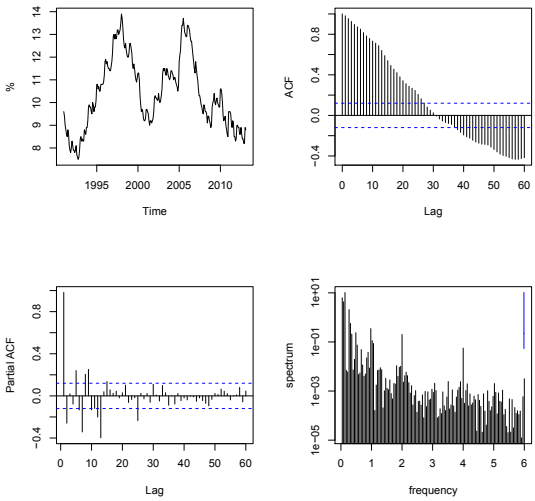


Figure F.39.: Series Hamburg (01/1991-02/2013) with S(P)ACF and periodogram.

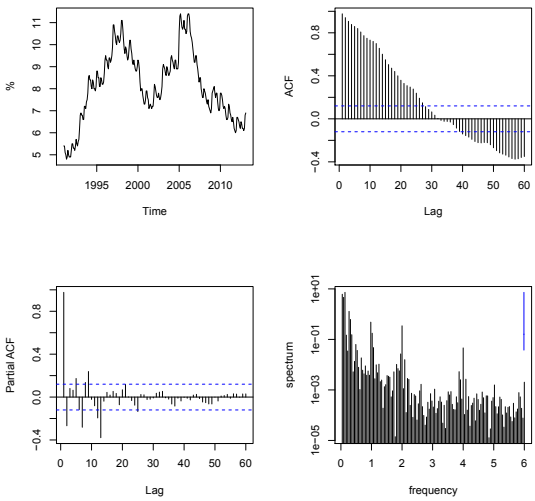


Figure F.40.: Series Hesse (01/1991-02/2013) with S(P)ACF and periodogram.

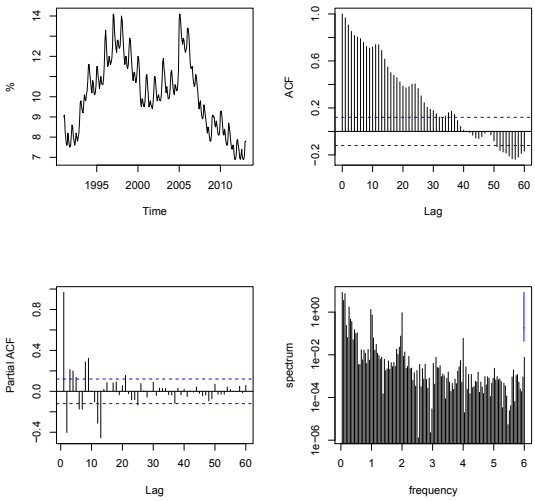


Figure F.41.: Series Lower Saxony (01/1991-02/2013) with S(P)ACF and periodogram.

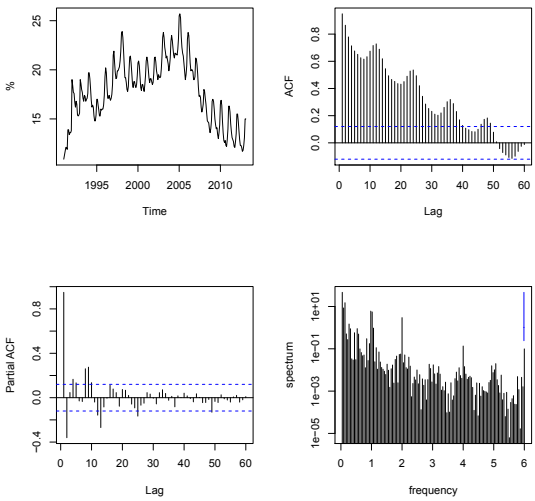


Figure F.42.: Series Mecklenburg-Western Pom. (01/1991-02/2013) with S(P)ACF and periodogram.

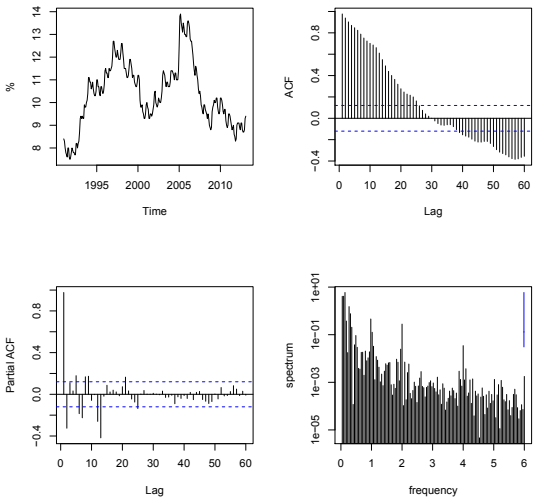


Figure F.43.: Series North Rhine-Westphalia (01/1991-02/2013) with S(P)ACF and periodogram.

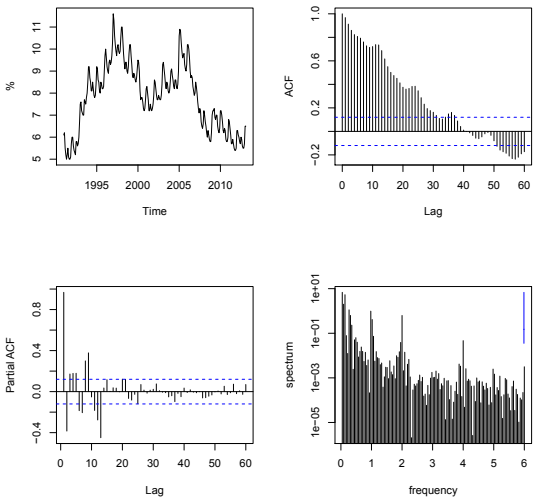


Figure F.44.: Series Rhineland-Palatinate (01/1991-02/2013) with S(P)ACF and periodogram.

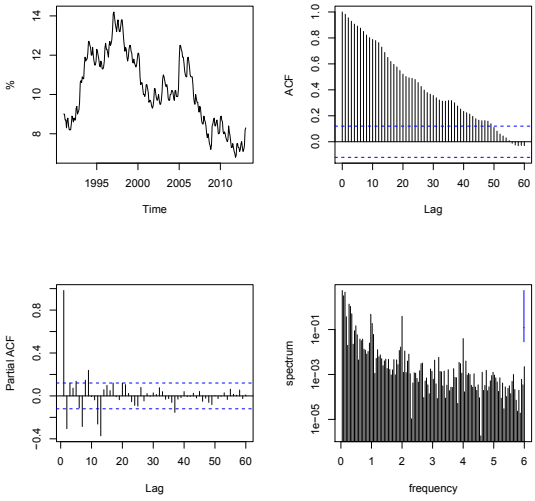


Figure F.45.: Series Saarland (01/1991-02/2013) with S(P)ACF and periodogram.

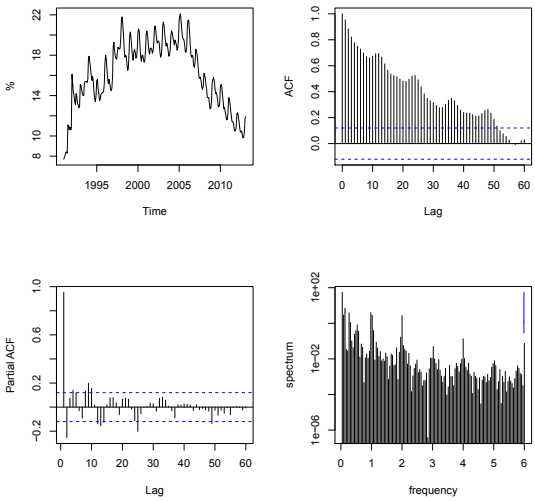


Figure F.46.: Series Saxony (01/1991-02/2013) with S(P)ACF and periodogram.

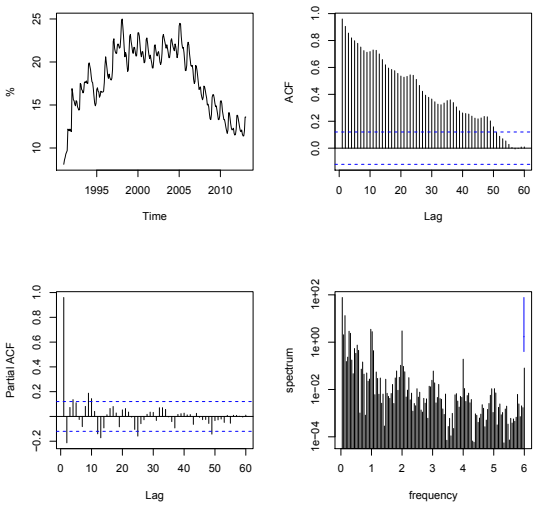


Figure F.47.: Series Saxony-Anhalt (01/1991-02/2013) with S(P)ACF and periodogram.

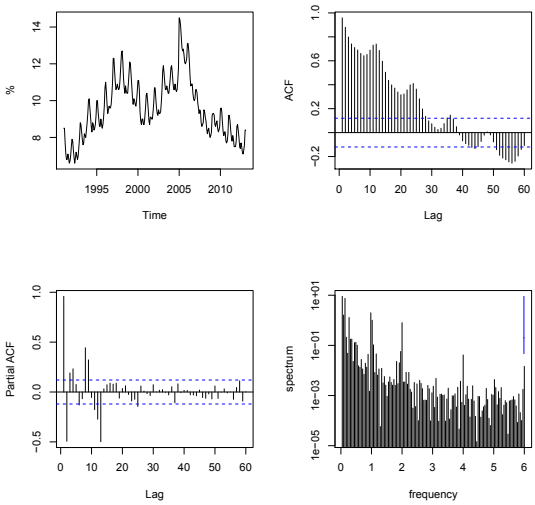


Figure F.48.: Series Schleswig-Holstein (01/1991-02/2013) with S(P)ACF and periodogram.

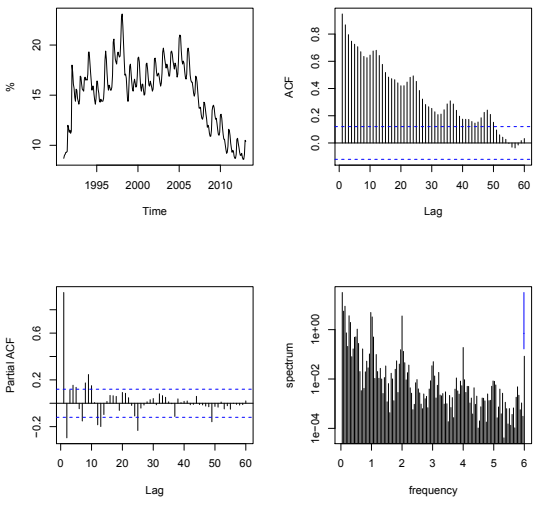
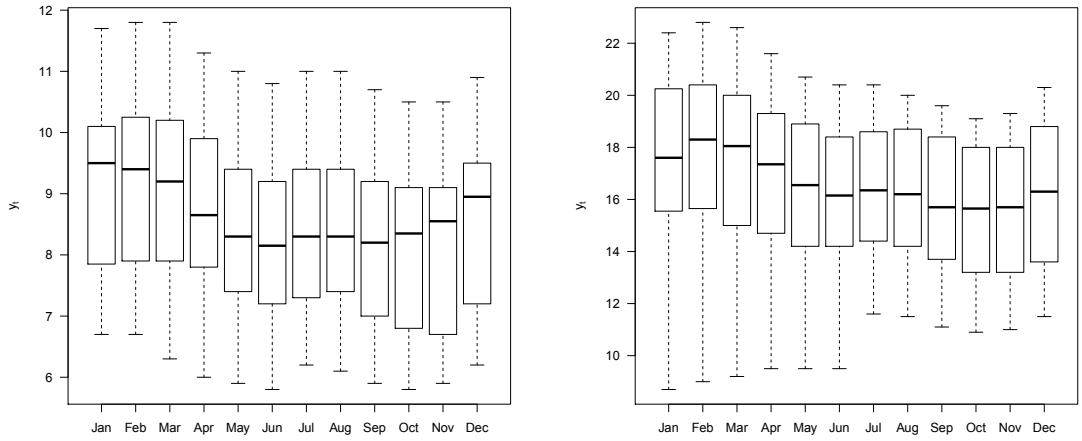
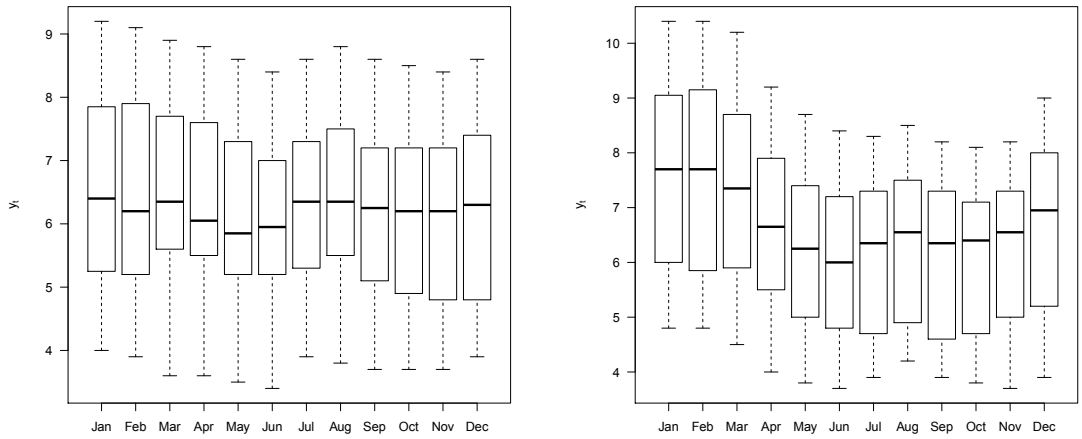


Figure F.49.: Series Thuringia (01/1991-02/2013) with S(P)ACF and periodogram.



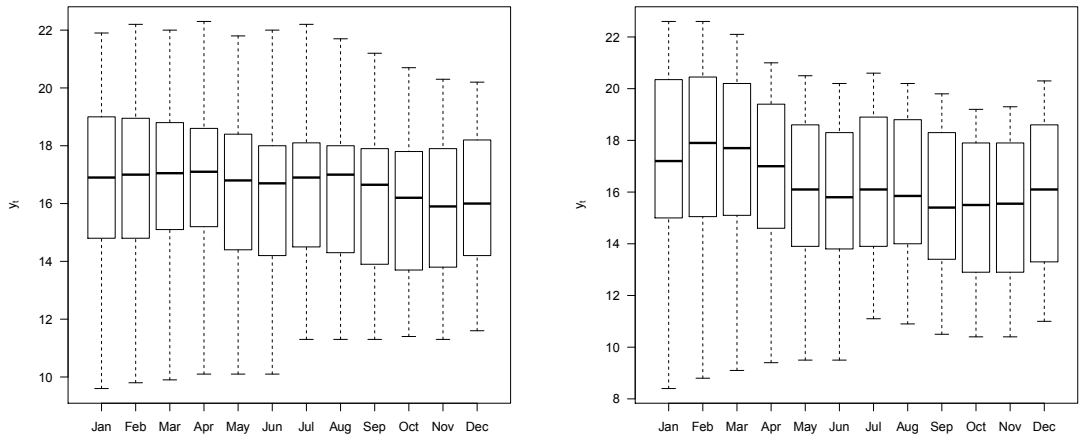
(a) West-Germany (b) East-Germany

Figure F.50.: Seasonal boxplots: West- and East-Germany.



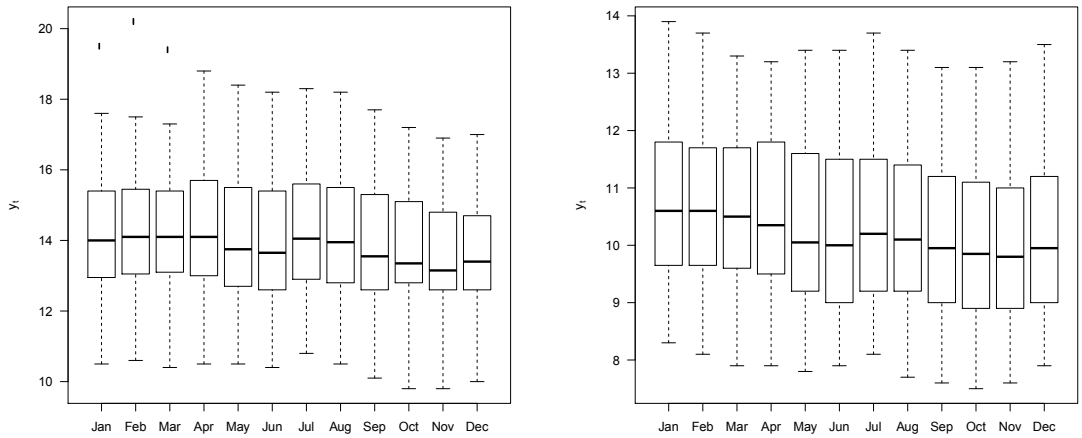
(a) Baden-Wuerttemberg (b) Bavaria

Figure F.51.: Seasonal boxplots: Baden-Wuerttemberg and Bavaria.



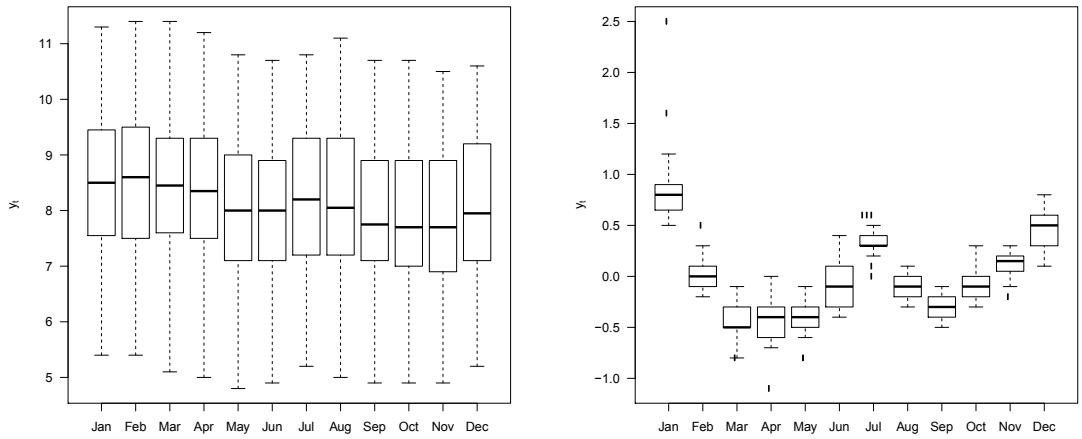
(a) Berlin (b) Brandenburg

Figure F.52.: Seasonal boxplots: Berlin and Brandenburg.



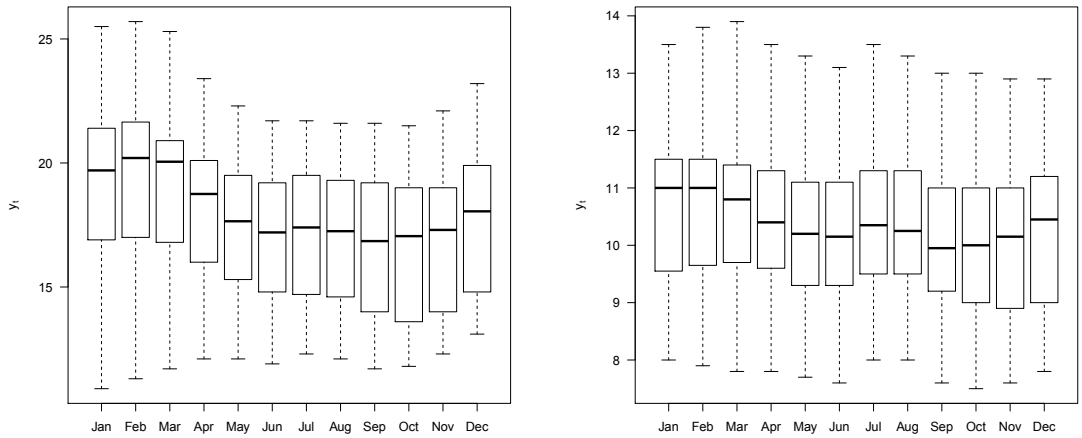
(a) Bremen (b) Hamburg

Figure F.53.: Seasonal boxplots: Bremen and Hamburg.



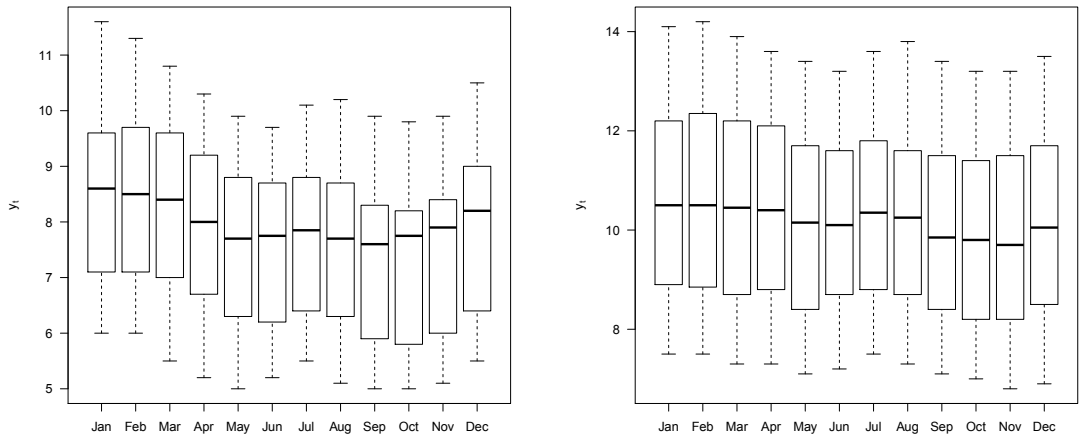
(a) Hesse (b) Lower Saxony

Figure F.54.: Seasonal boxplots: Hesse and Lower Saxony.



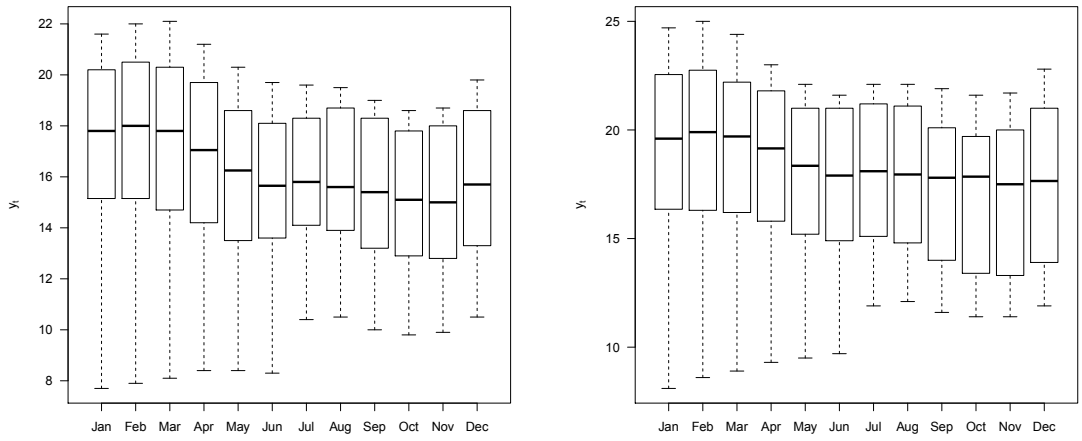
(a) Mecklenburg-Western Pomerania (b) North Rhine-Westphalia

Figure F.55.: Seasonal boxplots: Mecklenburg-Western Pomerania and North Rhine-Westphalia.



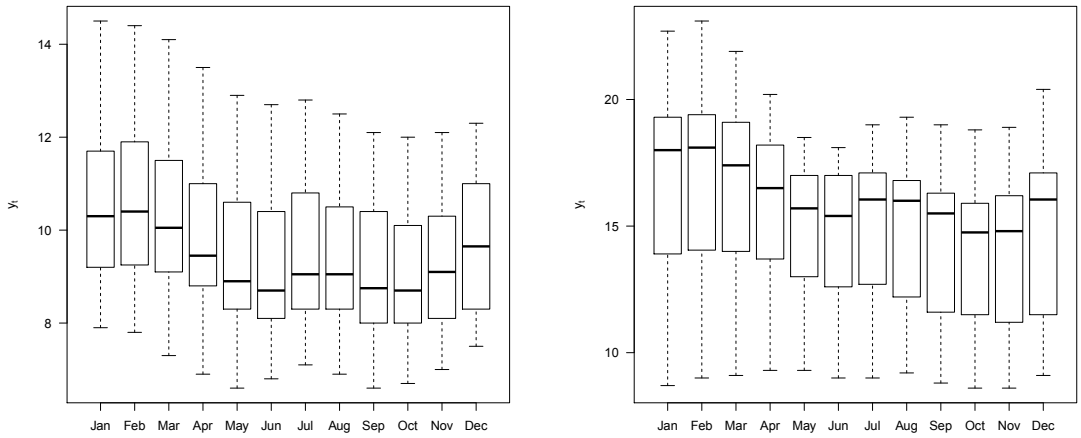
(a) Rhineland-Palatinate (b) Saarland

Figure F.56.: Seasonal boxplots: Rhineland-Palatinate and Saarland.



(a) Saxony (b) Saxony-Anhalt

Figure F.57.: Seasonal boxplots: Saxony and Saxony-Anhalt.



(a) Schleswig-Holstein (b) Thuringia

Figure F.58.: Seasonal boxplots: Schleswig-Holstein and Thuringia.

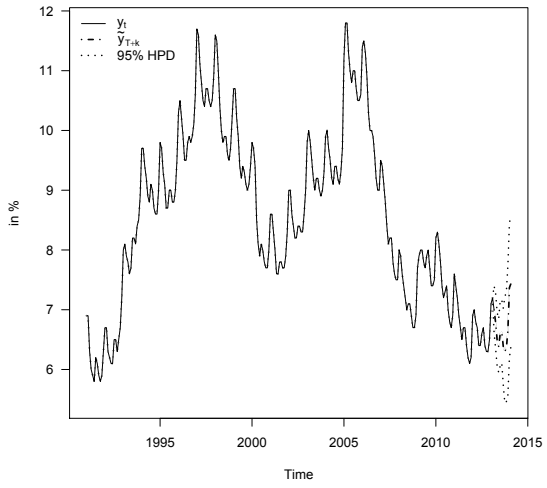


Figure F.59.: One-year ahead forecasts of the unemployment rates of West-Germany.

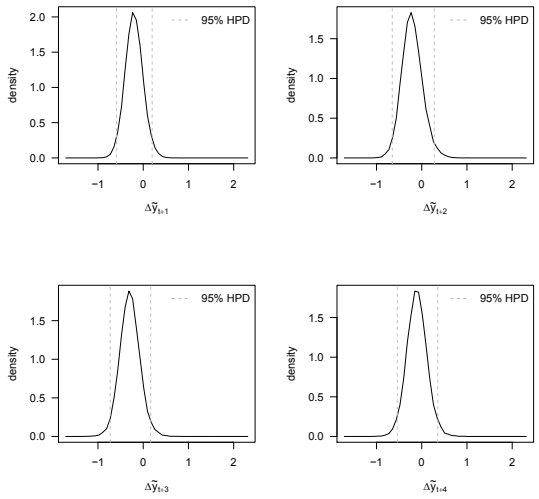


Figure F.60.: Model averaged posterior predictive densities of West-Germany (1).

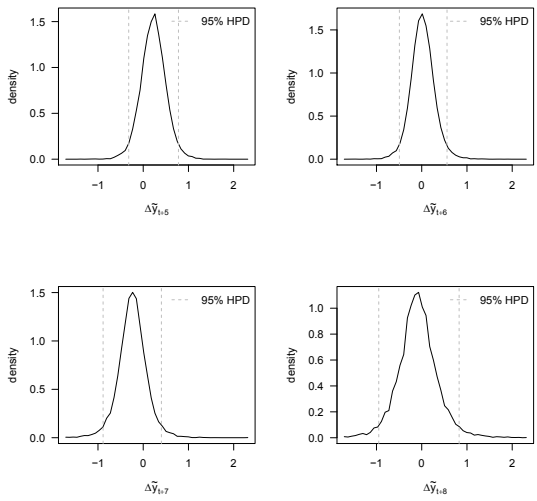


Figure F.61.: Model averaged posterior predictive densities of West-Germany (2).

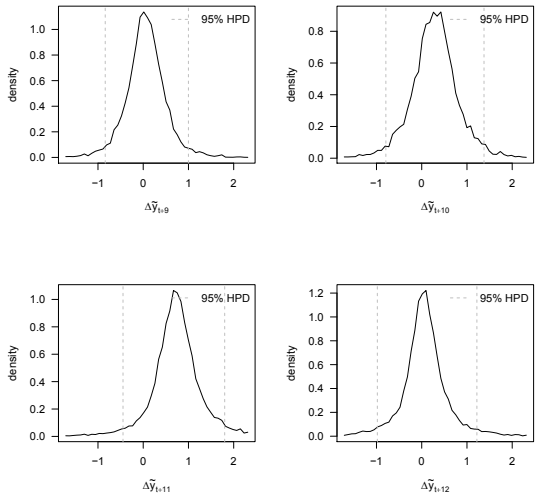


Figure F.62.: Model averaged posterior predictive densities of West-Germany (3).

G. Technical details - chapter 4

In the following the derivations of the distributions, used in chapter 4, are outlined. For this purpose, in section G.1, some basic expressions required in the subsequent sections are derived. Then in section G.2 and section G.3 the posterior density of the subvector $\mathbf{B}_1$ and ϕ_s are established, respectively, where in section G.4 the posterior distribution of the linear form θ is derived. Finally, in section G.5 some details on the derivation of the posterior density of the quadratic form Q, used for the Bayesian F-test, are outlined. To provide more general posterior results a conjugate Normal-Inverse-Gamma-2 (NIG_2) prior for $(\mathbf{B},\ \sigma^2)$ will be used. Note that a diffuse prior as in chapter 4 can easily be obtained as a special case of a NIG_2 prior, where the latter also serves as a starting point for other conjugate priors, e.g. Zellner's g-prior (see Zellner (1986)). In section G.6 some additional details on these prior issues will be given.

G.1. Preliminaries - posterior analysis

By an application of the Bayes Theorem the joint probability density function of all model parameters and the data can be factorized according to

$$\begin{aligned} f(\mathbf{B},\sigma^2,\mathbf{y}|\ M_i) &= f(\mathbf{y}|\ \mathbf{B},\ \sigma^2,\ M_i)\cdot f(\mathbf{B},\ \sigma^2|\ M_i) && (6.1)\\ &= f(\mathbf{B},\sigma^2|\ \mathbf{y},\ M_i)\cdot f(\mathbf{y}|\ M_i) && (6.2)\end{aligned}$$

where M_i is a model indicator for a particular model in the discrete model space $\mathscr{M} = \{M_1,...,M_K\}$. To express ignorance with respect to $\mathscr{M}$, $f(M_i) = 1/K$ is chosen. Utilizing the above assumptions about the data and the parameters, and applying a variance

decomposition to the quadratic form in the likelihood function in (6.1) leads to

$$
\begin{aligned}
f(\mathbf{B},\sigma^2,\mathbf{y}|\, M_i) \;=\; & C_N^{-1}(\sigma^2\mathbf{I}_{T-p}\,;\,T-p)\cdot\exp\left\{-\frac{1}{2\sigma^2}[\nu s^2+(\mathbf{B}-\widehat{\mathbf{B}})'\cdot\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\cdot(\mathbf{B}-\widehat{\mathbf{B}})]\right\}\cdot \\
& C_N^{-1}(\sigma^2\mathbf{M}^{-1};\,d)\cdot\exp\left\{-\frac{1}{2\sigma^2}[(\mathbf{B}-\mathbf{B_0})'\cdot\mathbf{M}\cdot(\mathbf{B}-\mathbf{B_0})]\right\}\cdot \\
& C_g^{-1}\left(\frac{a}{2},\frac{2}{b}\right)\cdot(\sigma^2)^{-\frac{a+2}{2}}\cdot\exp\left\{-\frac{b}{2\sigma^2}\right\}
\end{aligned}
\tag{6.3}
$$

with $d\equiv dim(\mathbf{B})$. Further let $C_N(\mathbf{\Sigma}\,;\,k)=(2\pi)^{\frac{k}{2}}\cdot|\mathbf{\Sigma}|^{\frac{1}{2}}$ and $C_g(a,b)=\Gamma(a)\cdot b^a$, $a,b>0$ be the normalizing constant of the Normal and the Gamma distribution, respectively. Here the first argument of $C_N(;)$ denotes the covariance matrix and the second argument denotes the dimension of the respective random vector (see Bauwens et al. (1999), p.293). Let $\widehat{\mathbf{B}}$ be the least squares estimator of $\mathbf{B}$ and $\nu s^2\equiv(\mathbf{y}-\widetilde{\mathbf{X}}\widehat{\mathbf{B}})'\cdot(\mathbf{y}-\widetilde{\mathbf{X}}\widehat{\mathbf{B}})$ the sum of squared residuals, with $\nu=T-p-d$ degrees of freedom.

Writing this together yields

$$
\begin{aligned}
&=(2\pi)^{-\frac{T-p+d+a}{2}}\cdot(\sigma^2)^{-\frac{\overbrace{T-p+d+a}^{\equiv a_\star}+2}{2}}\cdot|\mathbf{M}|^{\frac{1}{2}}\cdot\frac{b^{\frac{a}{2}}}{\Gamma(\frac{a}{2})\cdot 2^{\frac{a}{2}}} \\
&\cdot\exp\left\{-\frac{1}{2\sigma^2}[\underbrace{b+\nu s^2+(\mathbf{B}-\widehat{\mathbf{B}})'\cdot\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\cdot(\mathbf{B}-\widehat{\mathbf{B}})+(\mathbf{B}-\mathbf{B_0})'\cdot\mathbf{M}\cdot(\mathbf{B}-\mathbf{B_0})}_{\equiv b_\star}]\right\}
\end{aligned}
\tag{6.4}
$$

which has the form of a NIG_2 distribution, see Bauwens et al. (1999), p.302.

First it is to show, that $b_\star$ in (6.4) can be written as $b_\star=b+\nu s^2+QF(\mathbf{B_0})+QF(\mathbf{B})$,

with

$$\mu_B \equiv (\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}} + \mathbf{M})^{-1} \cdot (\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\widehat{\mathbf{B}} + \mathbf{M}\mathbf{B_0}) \tag{6.5a}$$

$$\mathbf{H} \equiv \widetilde{\mathbf{X}}'\widetilde{\mathbf{X}} + \mathbf{M} \tag{6.5b}$$

$$\delta \equiv (\mathbf{B} - \mu_B) \tag{6.5c}$$

$$QF(\mathbf{B}) \equiv \delta' \cdot \mathbf{H} \cdot \delta \tag{6.5d}$$

$$QF(\mathbf{B_0}) \equiv (\mathbf{B_0} - \widehat{\mathbf{B}})' \cdot [\mathbf{M}^{-1} + (\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})^{-\mathbf{1}}]^{-1} \cdot (\mathbf{B_0} - \widehat{\mathbf{B}}) \tag{6.5e}$$

Next the derivation of the quadratic forms in (6.5d) and (6.5e) is outlined. The quadratic form in (6.5d) can be obtained by completing the square for $\mathbf{B}$ in:

$$(\mathbf{B} - \widehat{\mathbf{B}})' \cdot \widetilde{\mathbf{X}}'\widetilde{\mathbf{X}} \cdot (\mathbf{B} - \widehat{\mathbf{B}}) + (\mathbf{B} - \mathbf{B_0})'\mathbf{M}(\mathbf{B} - \mathbf{B_0}) \tag{6.6a}$$

$$= \mathbf{B}' \cdot \mathbf{H} \cdot \mathbf{B} - 2\mathbf{B}' \cdot (\widetilde{\mathbf{X}}'\mathbf{y} + \mathbf{M}\mathbf{B}_0) + \widehat{\mathbf{B}}'\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\widehat{\mathbf{B}} + \mathbf{B_0'}\mathbf{M}\mathbf{B_0} \tag{6.6b}$$

Apply a Cholesky decomposition to the positive definite matrix $\mathbf{H}$ in (6.5b) as $\mathbf{H} = \mathbf{L} \cdot \mathbf{L}'$ with $\mathbf{L}$ the resulting lower triangular matrix. Then by defining $\mathbf{A} \equiv \mathbf{L}' \cdot \mathbf{B} \Leftrightarrow \mathbf{B} = (\mathbf{L}')^{-1} \cdot \mathbf{A}$, equation (6.6) can equivalently be expressed as

$$= \mathbf{A}'\mathbf{A} - 2 \cdot \mathbf{A}' \cdot \underbrace{\mathbf{L}^{-\mathbf{1}} \cdot (\widetilde{\mathbf{X}}'\mathbf{y} + \mathbf{M}\mathbf{B}_0)}_{\equiv \mu_A} + \mathbf{y}' \cdot \mathbf{P}_x \cdot \mathbf{y} + \mathbf{B_0'}\mathbf{M}\mathbf{B_0} \tag{6.7a}$$

$$= (\mathbf{A} - \mu_A)' \cdot (\mathbf{A} - \mu_A) - \mu'_A \cdot \mu_A + \mathbf{y}'\mathbf{P}_x\mathbf{y} + \mathbf{B_0'}\mathbf{M}\mathbf{B_0} \tag{6.7b}$$

with $\mathbf{P}_x \equiv \widetilde{\mathbf{X}} \cdot (\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})^{-\mathbf{1}} \cdot \widetilde{\mathbf{X}}'$ the usual linear projection matrix.[1] Writing out the last equation explicitly this results in

$$\begin{aligned} &= \underbrace{\left(\mathbf{B} - \mathbf{H}^{-1}[\widetilde{\mathbf{X}}'\mathbf{y} + \mathbf{M}\mathbf{B_0}]\right)' \cdot \mathbf{H} \cdot \left(\mathbf{B} - \mathbf{H}^{-1}[\widetilde{\mathbf{X}}'\mathbf{y} + \mathbf{M}\mathbf{B_0}]\right)}_{\equiv QF(\mathbf{B}) \,=\, \delta' \cdot \mathbf{H} \cdot \delta} \\ &\quad - (\mathbf{M}\mathbf{B_0} + \widetilde{\mathbf{X}}'\mathbf{y})'\mathbf{H}^{-1}(\mathbf{M}\mathbf{B_0} + \widetilde{\mathbf{X}}'\mathbf{y}) + \mathbf{y}'\mathbf{P}_x\mathbf{y} + \mathbf{B_0}'\mathbf{M}\mathbf{B_0} \end{aligned} \tag{6.8}$$

Next construct $QF(\mathbf{B}_0)$ in a similar fashion. For this reason, first notice that by utilizing

[1] Note that due to possible singularities in the $\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}$ matrix the Moore-Penrose pseudoinverse is used throughout, see Poole (2006), p.611, for details.

(6.8) together with the above definition of $\mathbf{P}_x$, equation (6.6) can be written as

$$\begin{aligned}&(\mathbf{B}-\widehat{\mathbf{B}})'\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}(\mathbf{B}-\widehat{\mathbf{B}})+(\mathbf{B}-\mathbf{B_0})'\mathbf{M}(\mathbf{B}-\mathbf{B_0})\\&=QF(\mathbf{B})-(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\widehat{\mathbf{B}}+\mathbf{M}\mathbf{B_0})'\mathbf{H}^{-1}(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\widehat{\mathbf{B}}+\mathbf{M}\mathbf{B_0})+\widehat{\mathbf{B}}'\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\widehat{\mathbf{B}}+\mathbf{B_0'}\mathbf{M}\mathbf{B_0}\end{aligned} \tag{6.9}$$

Now multiplying out the second term in equation (6.9) yields:

$$\begin{aligned}&(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\widehat{\mathbf{B}}+\mathbf{M}\mathbf{B_0})'(\mathbf{M}+\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})^{-1}(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\widehat{\mathbf{B}}+\mathbf{M}\mathbf{B_0})\\&=\mathbf{B_0'}\mathbf{M}'(\mathbf{M}+\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})^{-\mathbf{1}}\mathbf{M}\mathbf{B_0}+2\mathbf{B'}_0\mathbf{M}'(\mathbf{M}+\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\widehat{\mathbf{B}}\\&\quad+\widehat{\mathbf{B}}'\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}(\mathbf{M}+\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\widehat{\mathbf{B}}\end{aligned} \tag{6.10}$$

Next utilize the Woodbury matrix identity (cf. Lütkepohl (2007), p.660):[2]

$$(\mathbf{M}+\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})^{-1}=\mathbf{M}^{-1}-\mathbf{M}^{-1}(\mathbf{M}^{-1}+(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})^{-1})^{-1}\mathbf{M}^{-1} \tag{6.11a}$$

$$=(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})^{-1}-(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})^{-1}(\mathbf{M}^{-1}+(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})^{-1})^{-1}(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})^{-1} \tag{6.11b}$$

Substituting the first identity (6.11a) in the first and second summand of (6.10) and the second identity (6.11b) in the last summand of (6.10), respectively, yields after some algebra:

$$\begin{aligned}&=\mathbf{B}_0'\mathbf{M}\mathbf{B}_0-\mathbf{B}_0'(\mathbf{M}^{-1}+(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})^{-1})^{-1}\mathbf{B}_0\\&\quad+2\widehat{\mathbf{B}}'\mathbf{M}\mathbf{B}_0-2\widehat{\mathbf{B}}'(\mathbf{M}^{-1}+(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})^{-1})^{-1}(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})^{-1}\mathbf{M}\mathbf{B}_0\\&\quad+\widehat{\mathbf{B}}'\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\widehat{\mathbf{B}}-\widehat{\mathbf{B}}'(\mathbf{M}^{-1}+(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})^{-1})^{-1}\widehat{\mathbf{B}}\end{aligned} \tag{6.12}$$

For convenience let $\mathbf{K}\equiv(\mathbf{M}^{-1}+(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})^{-1})^{-1}$ and multiply the second identity in (6.11b) by $\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}$:

$$\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}(\mathbf{M}+\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})^{-1}=\mathbf{I}-\mathbf{K}(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})^{-1} \tag{6.13}$$

[2] In more general form this matrix identity can be stated as $(\mathbf{A}+\mathbf{U}\cdot\mathbf{C}\cdot\mathbf{V})^{-1}=\mathbf{A}^{-1}-\mathbf{A}^{-1}\cdot\mathbf{U}\cdot(\mathbf{C}^{-1}+\mathbf{V}\cdot\mathbf{A}^{-1}\cdot\mathbf{U})^{-1}\cdot\mathbf{V}\cdot\mathbf{A}^{-1}$, with $\mathbf{A},\mathbf{U},\mathbf{C}$ and $\mathbf{V}$ matrices of conformable dimension. Where the identities (6.11) can then be deduced by appropriate definitions of these matrices.

Then the third and fourth summands in (6.12) are equivalently expressed as

$$\begin{aligned} & 2\widehat{\mathbf{B}}'\mathbf{M}\mathbf{B}_0 - 2\widehat{\mathbf{B}}'(\mathbf{M}^{-1} + (\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})^{-1})^{-1}(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})^{-1}\mathbf{M}\mathbf{B}_0 \\ & = 2\widehat{\mathbf{B}}'[\mathbf{I} - \mathbf{K}(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})^{-\mathbf{1}}]\mathbf{M}\mathbf{B}_0 \end{aligned} \tag{6.14}$$

or, by using (6.13),

$$= 2\widehat{\mathbf{B}}' \cdot \widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}(\mathbf{M} + \widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})^{-1}\mathbf{M} \cdot \mathbf{B}_0 \tag{6.15}$$

Note that the middle matrix factor term in equation (6.15) can be written as

$$\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}(\mathbf{M} + \widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})^{-1}\mathbf{M} = \left[\mathbf{M}^{-1}(\mathbf{M} + \widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})^{-1}\right]^{-1} \tag{6.16a}$$

$$= \underbrace{\left(\mathbf{M}^{-1} + (\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})^{-1}\right)^{-1}}_{= \mathbf{K}} \tag{6.16b}$$

by multiplying out the right-hand side of equation (6.16a).

After utilizing the matrix identities (6.11), (6.13), (6.16) and some lengthy algebra, which is omitted here in order to save space, the left hand side of (6.10) can finally be written as

$$(\mathbf{M}\mathbf{B}_\mathbf{0} + \widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\widehat{\mathbf{B}})'(\mathbf{M} + \widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})^{-1}(\mathbf{M}\mathbf{B}_\mathbf{0} + \widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\widehat{\mathbf{B}}) \tag{6.17a}$$

$$= \mathbf{B}_0'\mathbf{M}\mathbf{B}_0 + \widehat{\mathbf{B}}'\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\widehat{\mathbf{B}} - \mathbf{B}_0'\mathbf{K}\mathbf{B}_0 + 2\widehat{\mathbf{B}}'\mathbf{K}\mathbf{B}_0 - \widehat{\mathbf{B}}'\mathbf{K}\widehat{\mathbf{B}} \tag{6.17b}$$

$$= \mathbf{B}_0'\mathbf{M}\mathbf{B}_0 + \widehat{\mathbf{B}}'\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\widehat{\mathbf{B}} - \underbrace{(\mathbf{B}_0 - \widehat{\mathbf{B}})'(\mathbf{M}^{-1} + (\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})^{-1})^{-1}(\mathbf{B}_0 - \widehat{\mathbf{B}})}_{\equiv QF(\mathbf{B}_0)} \tag{6.17c}$$

$$= \mathbf{B}_0'\mathbf{M}\mathbf{B}_0 + \widehat{\mathbf{B}}'\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\widehat{\mathbf{B}} - (\mathbf{B}_0 - \widehat{\mathbf{B}})'\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}(\mathbf{M} + \widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})^{-1}\mathbf{M}(\mathbf{B}_0 - \widehat{\mathbf{B}}) \tag{6.17d}$$

where the last equation follows from (6.16). Note that for the representation of diffuse prior information, i.e. with a precision of $\mathbf{M} \to 0$, as in the case of a flat prior, this implies $QF(\mathbf{B}_0) \to 0$.

After substituting (6.17c) in (6.9), the left-hand side of (6.9) equals

$$(\mathbf{B} - \widehat{\mathbf{B}})'(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})(\mathbf{B} - \widehat{\mathbf{B}}) + (\mathbf{B} - \mathbf{B}_\mathbf{0})'\mathbf{M}(\mathbf{B} - \mathbf{B}_\mathbf{0}) \;=\; QF(\mathbf{B}) + QF(\mathbf{B}_0) \tag{6.18}$$

and thus the exponent in (6.4) becomes $b_\star = b + \nu s^2 + QF(\mathbf{B_0}) + QF(\mathbf{B})$ as stated above.

Next, 1.) the marginal likelihood (or prior predictive distribution) under model M_i, 2.) the conditional posterior density of a subvector of $\mathbf{B}$ as well as 3.) the marginal posterior density of σ^2 under model M_i are derived.

For this purpose consider the following partitioning of δ in (6.5c) above:

$$\underset{d_1 \times 1}{\delta_1} \equiv (\mathbf{B_1} - \mu_{1B}) = \underset{d_1 \times d}{\mathbf{E}_1} \cdot (\mathbf{B} - \mu_B) \tag{6.19a}$$

$$\underset{d_2 \times 1}{\delta_2} \equiv (\mathbf{B_2} - \mu_{2B}) = \underset{d_2 \times d}{\mathbf{E}_2} \cdot (\mathbf{B} - \mu_B) \tag{6.19b}$$

$$\underset{d \times 1}{\delta} \equiv \left(\delta_1' \vdots \delta_2' \right)' \tag{6.19c}$$

with $d = d_1 + d_2$ and $\mathbf{E}_1 \equiv (\mathbf{I}_{d_1} \vdots \mathbf{0}_{d_2})$ and $\mathbf{E}_2 \equiv (\mathbf{0}_{d_1} \vdots \mathbf{I}_{d_2})$ two transformation matrices that eliminate the lower d_2 and d_1 components of the vector μ_B (see (6.5)), respectively.

Further partition the matrix $\mathbf{H}$ in (6.5b) conformably as follows:

$$\mathbf{H} \quad = \quad \begin{pmatrix} \underset{d_1 \times d_1}{\mathbf{H_{11}}} & | & \underset{d_1 \times d_2}{\mathbf{H_{12}}} \\ - - - - - & & - - - - - \\ \underset{d_2 \times d_1}{\mathbf{H_{21}}} & | & \underset{d_2 \times d_2}{\mathbf{H_{22}}} \end{pmatrix} \tag{6.20}$$

Using the above definitions the quadratic form in (6.8) can be expressed as

$$QF(\mathbf{B}) = \delta' \cdot \mathbf{H} \cdot \delta = \left(\delta_1' \vdots \delta_2' \right) \cdot \begin{pmatrix} \mathbf{H_{11}} & \mathbf{H_{12}} \\ \mathbf{H_{21}} & \mathbf{H_{22}} \end{pmatrix} \cdot \begin{pmatrix} \delta_1 \\ \delta_2 \end{pmatrix} \tag{6.21a}$$

Multiplying out this expression then leads to

$$= \delta_1' \cdot \mathbf{H_{11}} \cdot \delta_1 + 2 \cdot \delta_1' \cdot \mathbf{H_{12}} \cdot \delta_2 + \delta_2' \cdot \mathbf{H_{22}} \cdot \delta_2 \tag{6.22a}$$

Let $\mathbf{H_{22}} = \mathbf{L}\mathbf{L}'$ with $\mathbf{L}$ again the lower triangular matrix from a Cholesky decomposition and define $\delta_2^\star \equiv \mathbf{L}' \cdot \delta_2 \Leftrightarrow \delta_2 = (\mathbf{L}')^{-1} \cdot \delta_2^\star$. Then after completing the square with respect

to δ_2 we have:

$$= \underbrace{(\delta_2+\mu_2)' \cdot \mathbf{H_{22}} \cdot (\delta_2+\mu_2)}_{\equiv QF(\delta_2)} - \mu_2' \cdot \mathbf{H_{22}} \cdot \mu_2 + \delta_1' \cdot \mathbf{H_{11}} \cdot \delta_1 \tag{6.23a}$$

$$= QF(\delta_2) + \underbrace{\delta_1' \cdot (\mathbf{H_{11}} - \mathbf{H_{12}}\mathbf{H_{22}^{-1}}\mathbf{H_{21}}) \cdot \delta_1}_{\equiv QF(\delta_1)} \tag{6.23b}$$

with $\mu_2 \equiv \mathbf{H_{22}^{-1}} \cdot \mathbf{H_{21}} \cdot \delta_1$.

Utilizing the above expressions the joint density in (6.4) can be expressed as

$$\begin{aligned} f(\mathbf{B},\, \sigma^2, \mathbf{y}|\, M_i) = (2\pi)^{-\frac{a_\star}{2}} \cdot (\sigma^2)^{-\frac{a_\star}{2}} \cdot |\mathbf{M}|^{\frac{1}{2}} \cdot \frac{b^{\frac{a}{2}}}{\Gamma(\frac{a}{2}) \cdot 2^{\frac{a}{2}}} \cdot \exp\left\{-\frac{QF(\delta_1)}{2\sigma^2}\right\} \\ \cdot \exp\left\{-\frac{1}{2\sigma^2}[b + \nu s^2 + QF(\mathbf{B}_0) + QF(\delta_2)]\right\} \end{aligned} \tag{6.24}$$

Next integrate out δ_2 from (6.24), since in the next section we want to draw our attention to the analysis of the subvector δ_1:

$$\begin{aligned} f(\delta_1,\, \sigma^2, \mathbf{y}|\, M_i) = (2\pi)^{-\frac{a_\star}{2}} \cdot (\sigma^2)^{-\frac{a_\star}{2}} \cdot |\mathbf{M}|^{\frac{1}{2}} \cdot \frac{b^{\frac{a}{2}}}{\Gamma(\frac{a}{2}) \cdot 2^{\frac{a}{2}}} \cdot \exp\left\{-\frac{QF(\delta_1)}{2\sigma^2}\right\} \\ \cdot \exp\left\{-\frac{1}{2\sigma^2}[b + \nu s^2 + QF(\mathbf{B}_0)]\right\} \\ \cdot C_{N_2} \cdot \int_{\mathbb{R}^{d_2}} C_{N_2}^{-1} \cdot \exp\left\{-\frac{QF(\delta_2)}{2\sigma^2}\right\} \cdot d\delta_2 \end{aligned} \tag{6.25}$$

where $C_{N_2}(\mathbf{\Sigma}_2;\, d_2) = (2\pi)^{\frac{d_2}{2}} \cdot |\mathbf{\Sigma}_2|^{\frac{1}{2}}$ and $\mathbf{\Sigma}_2 = \sigma^2 \cdot \mathbf{H_{22}^{-1}}$.

$$
\begin{aligned}
&= (2\pi)^{-\frac{T-p}{2}} \cdot (2\pi)^{-\frac{d_1}{2}} \cdot (\sigma^2)^{-\frac{\overbrace{T-p+a}^{\equiv a_{\star\star}}+2}{2}} \cdot (\sigma^2)^{-\frac{d_1}{2}} \cdot |\mathbf{M}|^{\frac{1}{2}} \cdot \frac{b^{\frac{a}{2}}}{\Gamma(\frac{a}{2}) \cdot 2^{\frac{a}{2}}} \cdot |\mathbf{H_{22}}|^{-\frac{1}{2}} \\
&\quad \cdot \exp\left\{-\frac{QF(\delta_1)}{2\sigma^2}\right\} \cdot \exp\{-\frac{1}{2\sigma^2}[\underbrace{b + \nu s^2 + QF(\mathbf{B}_0)}_{\equiv b_{\star\star}}]\} \\
&= (2\pi)^{-\frac{T-p}{2}} \cdot (2\pi)^{-\frac{d_1}{2}} \cdot (\sigma^2)^{-\frac{d_1}{2}} \cdot |\mathbf{M}|^{\frac{1}{2}} \cdot \frac{b^{\frac{a}{2}}}{\Gamma(\frac{a}{2}) \cdot 2^{\frac{a}{2}}} \cdot |\mathbf{H_{22}}|^{-\frac{1}{2}} \cdot \exp\left\{-\frac{QF(\delta_1)}{2\sigma^2}\right\} \\
&\quad \cdot C_g\left(\frac{a_{\star\star}}{2}; \frac{2}{b_{\star\star}}\right) \cdot \underbrace{C_g^{-1}\left(\frac{a_{\star\star}}{2}; \frac{2}{b_{\star\star}}\right) \cdot (\sigma^2)^{-\frac{a_{\star\star}+2}{2}} \cdot \exp\left\{-\frac{b_{\star\star}}{2\sigma^2}\right\}}_{= IG_2(a_{\star\star}\,,\, b_{\star\star}) \text{ density}} \\
&= (2\pi)^{-\frac{T-p}{2}} \cdot (2\pi)^{-\frac{d_1}{2}} \cdot (\sigma^2)^{-\frac{d_1}{2}} \cdot |\mathbf{M}|^{\frac{1}{2}} \cdot \frac{b^{\frac{a}{2}}}{\Gamma(\frac{a}{2}) \cdot 2^{\frac{a}{2}}} \cdot |\mathbf{H_{22}}|^{-\frac{1}{2}} \\
&\quad \cdot C_g\left(\frac{a_{\star\star}}{2}; \frac{2}{b_{\star\star}}\right) \cdot IG_2 \cdot C_{N_1}(\mathbf{\Sigma}_1;\, d_1) \cdot \underbrace{C_{N_1}^{-1}(\mathbf{\Sigma}_1;\, d_1) \cdot \exp\left\{-\frac{QF(\delta_1)}{2\sigma^2}\right\}}_{= N_{d_1}(\mu_{1B}\,,\, \mathbf{\Sigma}_1) \text{ density}}
\end{aligned}
\tag{6.26}
$$

with $a_{\star\star} = T - p + a$ and $b_{\star\star} = b + \nu s^2 + QF(\mathbf{B_0})$ and normalizing constants $C_g\left(\frac{a_{\star\star}}{2}; \frac{2}{b_{\star\star}}\right)$, $C_{N_1}(\mathbf{\Sigma}_1;\, d_1)$ as defined above.

Hence, given a model specification M_i and a value for σ^2, the random vector δ_1, or equivalently $\mathbf{B}_1$, follows a d_1-dimensional normal posterior distribution with first and second moments μ_{1B} (see (6.19)) and

$$
\mathbf{\Sigma}_1 = \sigma^2 \cdot (\mathbf{H_{11}} - \mathbf{H_{12}}\mathbf{H_{22}^{-1}}\mathbf{H_{21}})^{-1}\,, \tag{6.27}
$$

respectively, and the conditional posterior density of σ^2, under model M_i, is of the $IG_2\,(a_{\star\star}\,;\, b_{\star\star})$

form. By using (6.27) the last line of (6.26) becomes

$$
\begin{aligned}
&= (2\pi)^{-\frac{T-p}{2}} \cdot |\mathbf{M}|^{\frac{1}{2}} \cdot \underbrace{|\mathbf{H_{22}}|^{-\frac{1}{2}} \cdot \left|(\mathbf{H_{11}} - \mathbf{H_{12}}\mathbf{H_{22}^{-1}}\mathbf{H_{21}})\right|^{-\frac{1}{2}}}_{= |\mathbf{H}|^{-\frac{1}{2}}} \\
&\quad \cdot C_g^{-1}\left(\frac{a}{2}; \frac{2}{b}\right) \cdot C_g\left(\frac{a_{\star\star}}{2}; \frac{2}{b_{\star\star}}\right) \cdot \underbrace{N_{d_1} \cdot IG_2}_{= NIG_2 \text{ density}}
\end{aligned}
\tag{6.28}
$$

where the fact is utilized that $|\mathbf{H}| = |\mathbf{H_{22}}| \cdot |\mathbf{H_{11}} - \mathbf{H_{12}}\mathbf{H_{22}^{-1}}\mathbf{H_{21}}|$, cf. Greene (2003), Appendix A.5.2, p.823.

Next rewrite the product of determinants in (6.28) as follows:

$$
|\mathbf{H}|^{-\frac{1}{2}} \cdot |\mathbf{M}|^{\frac{1}{2}} = \left|\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}} + \mathbf{M}\right|^{-\frac{1}{2}} \cdot \left|\mathbf{M}^{-1}\right|^{-\frac{1}{2}} \tag{6.29a}
$$

$$
= \left|\mathbf{I}_d + \widetilde{\mathbf{X}}'\widetilde{\mathbf{X}} \cdot \mathbf{M}^{-1}\right|^{-\frac{1}{2}} \tag{6.29b}
$$

From Sylvester's determinant theorem it is known that for any two matrices $\underset{d\times T-p}{\mathbf{A}}$ and $\underset{T-p\times d}{\mathbf{B}}$ the following identity holds:

$$
|\mathbf{I}_d + \mathbf{AB}| = \left|\mathbf{I}_{T-p} + \mathbf{BA}\right|
$$

and thus equation (6.29b) can equivalently be written as

$$
\left|\mathbf{I}_d + \widetilde{\mathbf{X}}'\widetilde{\mathbf{X}} \cdot \mathbf{M}^{-1}\right|^{-\frac{1}{2}} = \left|\mathbf{I}_{T-p} + \widetilde{\mathbf{X}}\mathbf{M}^{-1}\widetilde{\mathbf{X}}'\right|^{-\frac{1}{2}} \tag{6.30}
$$

Hence equation (6.28) becomes

$$
\begin{aligned}
f(\mathbf{B}_1,\, \sigma^2, \mathbf{y}|\, M_i) = (2\pi)^{-\frac{T-p}{2}} &\cdot \left|\mathbf{I}_{T-p} + \widetilde{\mathbf{X}}\mathbf{M}^{-1}\widetilde{\mathbf{X}}'\right|^{-\frac{1}{2}} \cdot \\
&C_g^{-1}\left(\frac{a}{2}; \frac{2}{b}\right) \cdot C_g\left(\frac{a_{\star\star}}{2}; \frac{2}{b_{\star\star}}\right) \cdot NIG_2
\end{aligned}
\tag{6.31}
$$

or

$$= (2\pi)^{-\frac{T-p}{2}} \cdot \left|\mathbf{I}_{T-p} + \widetilde{\mathbf{X}}\mathbf{M}^{-1}\widetilde{\mathbf{X}}'\right|^{-\frac{1}{2}} \cdot \frac{b^{\frac{a}{2}}}{\Gamma(\frac{a}{2}) \cdot 2^{\frac{a}{2}}} \cdot \frac{\Gamma(\frac{a_{\star\star}}{2}) \cdot 2^{\frac{T-p+a}{2}}}{\left[b_{\star\star}\right]^{\frac{a_{\star\star}}{2}}} \cdot NIG_2 \tag{6.32a}$$

$$= \frac{\Gamma(\frac{a_{\star\star}}{2})}{\pi^{(\frac{T-p}{2})} \cdot \Gamma(\frac{a}{2})} \cdot b^{\frac{a}{2}} \cdot \left|\mathbf{I}_{T-p} + \widetilde{\mathbf{X}}\mathbf{M}^{-1}\widetilde{\mathbf{X}}'\right|^{-\frac{1}{2}} \cdot \left[b + \nu s^2 + QF(\mathbf{B}_0)\right]^{-\frac{a_{\star\star}}{2}} \cdot NIG_2 \tag{6.32b}$$

Now let $QF(\mathbf{y}) \equiv \nu s^2 + QF(\mathbf{B}_0)$ and note the following identities:

$$\begin{aligned} QF(\mathbf{y}) &= (\mathbf{y} - \widetilde{\mathbf{X}}\widehat{\mathbf{B}})' \cdot (\mathbf{y} - \widetilde{\mathbf{X}}\widehat{\mathbf{B}}) + (\mathbf{B}_0 - \widehat{\mathbf{B}})' \cdot \left(\mathbf{M}^{-\mathbf{1}} + (\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})^{-\mathbf{1}}\right)^{-1} \cdot (\mathbf{B}_0 - \widehat{\mathbf{B}}) \\ &= (\mathbf{y} - \widetilde{\mathbf{X}}\widehat{\mathbf{B}})' \cdot (\mathbf{y} - \widetilde{\mathbf{X}}\widehat{\mathbf{B}}) + (\mathbf{B}_0 - \widehat{\mathbf{B}})' \cdot \widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}(\mathbf{M} + \widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})^{-1}\mathbf{M} \cdot (\mathbf{B}_0 - \widehat{\mathbf{B}}) \\ &= (\mathbf{y} - \widetilde{\mathbf{X}}\mathbf{B}_0)' \cdot (\mathbf{I}_{T-p} + \widetilde{\mathbf{X}}\mathbf{M}^{-\mathbf{1}}\widetilde{\mathbf{X}}')^{-1} \cdot (\mathbf{y} - \widetilde{\mathbf{X}}\mathbf{B}_0) \\ &= (\mathbf{y} - \widetilde{\mathbf{X}}\mathbf{B}_0)' \cdot (\mathbf{I}_{T-p} - \widetilde{\mathbf{X}}(\mathbf{M} + \widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}') \cdot (\mathbf{y} - \widetilde{\mathbf{X}}\mathbf{B}_0) \end{aligned} \tag{6.33}$$

where the first equality follows by using equation (6.17c) and the second equality follows from (6.16). The last but one equality can be obtained after some lengthy algebra, essentially by completing the square with respect to $\mathbf{y}$, see Hamilton (1994), Appendix 12.A., p.368, for more details, and the last identity follows from an application of the Woodbury matrix identity, see footnote 2 on page 185 above.

Hence the joint probability density in (6.32b) can be written as

$$\begin{aligned} f(\mathbf{B}_1,\, \sigma^2,\, \mathbf{y}|\, M_i) &= \frac{\Gamma(\frac{T-p+a}{2})}{\pi^{(\frac{T-p}{2})} \cdot \Gamma(\frac{a}{2})} \cdot b^{-\frac{T-p}{2}} \cdot \left|\mathbf{I}_{T-p} + \widetilde{\mathbf{X}}\mathbf{M}^{-1}\widetilde{\mathbf{X}}'\right|^{-\frac{1}{2}} \\ &\qquad \cdot \left[\, 1 + QF(\mathbf{y})/b \,\right]^{-\frac{T-p+a}{2}} \cdot NIG_2 \\ &= f(\mathbf{y}|\, M_i) \cdot f(\mathbf{B}_1|\, \sigma^2,\, M_i,\, \mathbf{y}) \cdot f(\sigma^2|\, M_i,\, \mathbf{y}) \end{aligned} \tag{6.34}$$

From (6.34) it can be observed that, under model M_i, the conditional posterior distribution of $\mathbf{B}_1$, given σ^2, is a multivariate Normal density with first and second moments μ_{1B} and $\mathbf{\Sigma}_1$, respectively, and the marginal posterior distribution of σ^2 is of the $IG_2\,(a_{\star\star}\,;\, b_{\star\star})$

form, with mean $E(\sigma^2|\, M_i,\, \mathbf{y}) = a_{\star\star}/\,(b_{\star\star}-2)$, for $b_{\star\star} > 2$, and variance $Var(\sigma^2|M_i,\, \mathbf{y}) = \frac{2}{b_{\star\star}-4} \cdot [E(\sigma^2|\, M_i,\, \mathbf{y})]^2$, for $b_{\star\star} > 4$, see Bauwens et al. (1999), p.292.

Moreover from (6.34) it can be recognized that the joint data density, $f(\mathbf{y}|\, M_i)$, under model M_i, is given by

$$\begin{aligned} f(\mathbf{y}|\, M_i) &= \frac{\Gamma(\frac{T-p+a}{2})}{\pi^{(\frac{T-p}{2})} \cdot \Gamma(\frac{a}{2})} \cdot b^{-\frac{T-p}{2}} \cdot |\mathbf{P}|^{-\frac{1}{2}} \cdot [\, 1 + QF(\mathbf{y})\,/\,b\,]^{-\frac{T-p+a}{2}} \\ &= C_t^{-1}\,(\mathbf{P};\, a;\, T-p) \cdot \Big[\, 1 + (\mathbf{y} - \widetilde{\mathbf{X}}\mathbf{B}_0)'\,(b\cdot\mathbf{P})^{-1}(\mathbf{y} - \widetilde{\mathbf{X}}\mathbf{B}_0)\,\Big]^{-\frac{T-p+a}{2}} \end{aligned} \tag{6.35}$$

with $\mathbf{P} \equiv \mathbf{I}_{T-p} + \widetilde{\mathbf{X}}\mathbf{M}^{-1}\widetilde{\mathbf{X}}'$ and

$$C_t\,(\mathbf{P};\, a;\, T-p) = \left[\Gamma\left(\frac{a}{2}\right)/\Gamma\left(\frac{T-p+a}{2}\right)\right] \cdot \pi^{\frac{T-p}{2}} \cdot |b\cdot\mathbf{P}|^{\frac{1}{2}}$$

the normalizing constant of a multivariate Student-t density (cf. Bauwens et al. (1999), p.303).

Hence the sample density (6.35) is a $(T-p)$-dimensional multivariate Student-t density with a degrees of freedom, mean vector $\widetilde{\mathbf{X}}\mathbf{B}_0$ and scale matrix $b\cdot(\mathbf{I}_{T-p} + \widetilde{\mathbf{X}}\mathbf{M}^{-1}\widetilde{\mathbf{X}}')$, see also Hamilton (1994), Appendix 12.A., p.368, for a similar result. For given data, (6.35) is the marginal likelihood function of M_i, and is thus the (unnormalized) probability mass function of the discrete-valued random vector of model indicators M_i, which is needed for the construction of posterior model probabilities.

G.2. Derivation of the posterior density of $\mathbf{B}_1$

Next the (model specific) marginal posterior density $f(\mathbf{B}_1|M_i,\, \mathbf{y})$ will be derived. For this purpose the first equation in (6.26) of section G.1 is utilized, which will be restated here

for the ease of reference:

$$f(\delta_1,\, \sigma^2, \mathbf{y}|\, M_i) = (2\pi)^{-\frac{T-p+d_1}{2}} \cdot (\sigma^2)^{-\frac{\overbrace{d_1+a_{\star\star}}^{\equiv\, \widetilde{a}}+2}{2}} \cdot |\mathbf{M}|^{\frac{1}{2}} \cdot \frac{b^{\frac{a}{2}}}{\Gamma(\frac{a}{2}) \cdot 2^{\frac{a}{2}}} \cdot \\ |\mathbf{H_{22}}|^{-\frac{1}{2}} \cdot \exp\{-\frac{1}{2\sigma^2}[\underbrace{b_{\star\star}+QF(\delta_1)}_{\equiv\, \widetilde{b}}]\} \tag{6.36}$$

with $\widetilde{a} = d_1 + a_{\star\star}$ and $\widetilde{b} = b_{\star\star} + QF(\delta_1)$, and $QF(\delta_1) = \delta_1' \cdot (\mathbf{H_{11}} - \mathbf{H_{12}}\mathbf{H_{22}^{-1}}\mathbf{H_{21}}) \cdot \delta_1$.

Also recall from above that $a_{\star\star} = T - p + a$ and $b_{\star\star} = b + \nu s^2 + QF(\mathbf{B_0}) = b + QF(\mathbf{y})$, with $QF(\mathbf{y}) = (\mathbf{y} - \widetilde{\mathbf{X}}\mathbf{B}_0)' \cdot (\mathbf{I}_{T-p} + \widetilde{\mathbf{X}}\mathbf{M^{-1}}\widetilde{\mathbf{X}}')^{-1} \cdot (\mathbf{y} - \widetilde{\mathbf{X}}\mathbf{B}_0)$.

Next integrate (6.36) with respect to σ^2 by using the properties of the IG_2 density:

$$\begin{aligned} f(\delta_1,\, \mathbf{y}|\, M_i) \;=\; & (2\pi)^{-\frac{T-p+d_1}{2}} \cdot \frac{b^{\frac{a}{2}}}{\Gamma(\frac{a}{2}) \cdot 2^{\frac{a}{2}}} \cdot |\mathbf{H_{22}}|^{-\frac{1}{2}} \cdot |\mathbf{M}|^{\frac{1}{2}} \cdot C_g\left(\frac{\widetilde{a}}{2}; \frac{2}{\widetilde{b}}\right) \cdot \\ & \int_{\mathbb{R}^+} C_g^{-1}\left(\frac{\widetilde{a}}{2}; \frac{2}{\widetilde{b}}\right) \cdot (\sigma^2)^{-\frac{\widetilde{a}+2}{2}} \cdot \exp\left\{-\frac{\widetilde{b}}{2\sigma^2}\right\} \cdot d\sigma^2 \\ =\; & (2\pi)^{-\frac{T-p+d_1}{2}} \cdot \frac{b^{\frac{a}{2}}}{\Gamma(\frac{a}{2}) \cdot 2^{\frac{a}{2}}} \cdot |\mathbf{H_{22}}|^{-\frac{1}{2}} \cdot |\mathbf{M}|^{\frac{1}{2}} \cdot C_g\left(\frac{\widetilde{a}}{2}; \frac{2}{\widetilde{b}}\right) \end{aligned} \tag{6.37}$$

with the normalizing constant of the IG_2 density equal to (see Bauwens et al. (1999), p.292):

$$C_g\left(\frac{\widetilde{a}}{2}; \frac{2}{\widetilde{b}}\right) = \Gamma\left(\frac{d_1 + a_{\star\star}}{2}\right) \cdot 2^{\frac{d_1+a_{\star\star}}{2}} \cdot [b_{\star\star} + QF(\delta_1)]^{-\frac{d_1+a_{\star\star}}{2}} \tag{6.38}$$

Substituting (6.38) in the last equation of (6.37) yields after some rearrangements:

$$\begin{aligned} f(\delta_1,\, \mathbf{y}|\, M_i) \;&=\; \pi^{-\frac{T-p+d_1}{2}} \cdot \frac{b^{\frac{a}{2}}}{\Gamma(\frac{a}{2})} \cdot |\mathbf{H_{22}}|^{-\frac{1}{2}} \cdot |\mathbf{M}|^{\frac{1}{2}} \cdot \Gamma\left(\frac{d_1+a_{\star\star}}{2}\right) \cdot \\ &\qquad (b_{\star\star})^{-\frac{d_1+a_{\star\star}}{2}} \cdot \left[1+QF(\delta_1)/b_{\star\star}\,\right]^{-\frac{d_1+a_{\star\star}}{2}} \\ &=\; \pi^{-\frac{T-p+d_1}{2}} \cdot \frac{b^{\frac{a}{2}}}{\Gamma(\frac{a}{2})} \cdot |\mathbf{H_{22}}|^{-\frac{1}{2}} \cdot |\mathbf{M}|^{\frac{1}{2}} \cdot \Gamma\left(\frac{d_1+a_{\star\star}}{2}\right) \cdot \\ &\qquad [b+QF(\mathbf{y})]^{-\frac{d_1+a_{\star\star}}{2}} \cdot \left[1+\delta_1' \cdot \mathbf{P}_1 \cdot \delta_1\right]^{-\frac{d_1+a_{\star\star}}{2}} \end{aligned} \tag{6.39}$$

with $\mathbf{P}_1 \equiv (\mathbf{H_{11}} - \mathbf{H_{12}}\mathbf{H_{22}^{-1}}\mathbf{H_{21}})/b_{\star\star}$.

$$\begin{aligned} =\; & \pi^{-\frac{T-p+d_1}{2}} \cdot \frac{b^{\frac{a}{2}}}{\Gamma(\frac{a}{2})} \cdot |\mathbf{H_{22}}|^{-\frac{1}{2}} \cdot |\mathbf{M}|^{\frac{1}{2}} \cdot \Gamma\left(\frac{d_1+a_{\star\star}}{2}\right) \cdot \\ & [b+QF(\mathbf{y})]^{-\frac{d_1+a_{\star\star}}{2}} \cdot \underbrace{C_t^{-1}(\mathbf{P}_1,\, a_{\star\star};\, d_1) \cdot \left[1+\delta_1' \cdot \mathbf{P}_1 \cdot \delta_1\,\right]^{-\frac{d_1+a_{\star\star}}{2}}}_{=\mathscr{T}_{d_1}(\mu_{1B},\, \mathbf{P}_1,\, a_{\star\star})} \cdot C_t(\mathbf{P}_1,\, a_{\star\star};\, d_1) \\ =\; & \pi^{-\frac{T-p+d_1}{2}} \cdot \frac{b^{\frac{a}{2}}}{\Gamma(\frac{a}{2})} \cdot |\mathbf{H_{22}}|^{-\frac{1}{2}} \cdot |\mathbf{M}|^{\frac{1}{2}} \cdot \Gamma\left(\frac{d_1+a_{\star\star}}{2}\right) \cdot \mathscr{T}_{d_1}(\mu_{1B},\, \mathbf{P}_1,\, a_{\star\star}) \cdot b^{-\frac{d_1+a_{\star\star}}{2}} \cdot \\ & \left[1+(\mathbf{y}-\widetilde{\mathbf{X}}\mathbf{B}_0)' \cdot \mathbf{P}_y \cdot (\mathbf{y}-\widetilde{\mathbf{X}}\mathbf{B}_0)\right]^{-\frac{d_1+a_{\star\star}}{2}} \cdot C_t(\mathbf{P}_1,\, a_{\star\star};\, d_1) \end{aligned} \tag{6.40}$$

with $\mathbf{P}_y \equiv \left[b \cdot (\mathbf{I}_{T-p} + \widetilde{\mathbf{X}}\mathbf{M^{-1}}\widetilde{\mathbf{X}}')\right]^{-1}$ and $C_t(\mathbf{P}_1,\, a_{\star\star};\, d_1)$ the normalizing constant of the d_1-dimensional Student-t density, $\mathscr{T}_{d_1}(\mu_{1B},\, \mathbf{P}_1,\, a_{\star\star})$, with mean μ_{1B}, scale matrix $\mathbf{P}_1$ and $a_{\star\star}$ degrees of freedom (see Bauwens et al. (1999), Appendix A.2).

For a p-dimensional t-density with scale matrix $\mathbf{P}$ and ν degrees of freedom this normalizing constant has the general form:

$$C_t(\mathbf{P},\, \nu;\, p) = \left[\Gamma\left(\frac{\nu}{2}\right)/\Gamma\left(\frac{\nu+p}{2}\right)\right] \cdot \pi^{\frac{1}{2}p} \cdot |\mathbf{P}|^{-\frac{1}{2}} \tag{6.41}$$

which is stated here for the ease of reference, ibid., p.303.

Using this (6.40) becomes

$$
\begin{aligned}
f(\delta_1,\, \mathbf{y}|\, M_i) \;=\; & \pi^{-\frac{T-p}{2}} \cdot b^{-\frac{T-p+d_1}{2}} \cdot \frac{\Gamma\left(\frac{a_{\star\star}}{2}\right)}{\Gamma\left(\frac{a}{2}\right)} \cdot |\mathbf{P}_1|^{-\frac{1}{2}} \cdot |\mathbf{H_{22}}|^{-\frac{1}{2}} \cdot |\mathbf{M}|^{\frac{1}{2}} \cdot \mathscr{T}_{d_1} \cdot \\
& \left[1 + (\mathbf{y} - \widetilde{\mathbf{X}}\mathbf{B}_0)' \cdot \mathbf{P}_y \cdot (\mathbf{y} - \widetilde{\mathbf{X}}\mathbf{B}_0)\right]^{-\frac{d_1}{2}} \cdot C_t(\mathbf{P}_y,\, a;\, T-p) \\
& \underbrace{C_t^{-1}(\mathbf{P}_y,\, a;\, T-p) \cdot \left[1 + (\mathbf{y} - \widetilde{\mathbf{X}}\mathbf{B}_0)' \cdot \mathbf{P}_y \cdot (\mathbf{y} - \widetilde{\mathbf{X}}\mathbf{B}_0)\right]^{-\frac{a_{\star\star}}{2}}}_{=\mathscr{T}_{T-p}(\widetilde{\mathbf{X}}\mathbf{B}_0,\, \mathbf{P}_y,\, a)} \\
= \; & b^{-\frac{T-p+d_1}{2}} \cdot |\mathbf{P}_1|^{-\frac{1}{2}} \cdot \left|\mathbf{P}_y\right|^{-\frac{1}{2}} \cdot |\mathbf{H_{22}}|^{-\frac{1}{2}} \cdot |\mathbf{M}|^{\frac{1}{2}} \cdot \\
& \left[1 + (\mathbf{y} - \widetilde{\mathbf{X}}\mathbf{B}_0)' \cdot \mathbf{P}_y \cdot (\mathbf{y} - \widetilde{\mathbf{X}}\mathbf{B}_0)\right]^{-\frac{d_1}{2}} \cdot \mathscr{T}_{d_1} \cdot \mathscr{T}_{T-p}
\end{aligned}
\tag{6.42}
$$

Now consider the product of determinants in (6.42) in some more detail:

$$
\begin{aligned}
& |\mathbf{P}_1|^{-\frac{1}{2}} \cdot |\mathbf{H_{22}}|^{-\frac{1}{2}} \cdot |\mathbf{M}|^{\frac{1}{2}} \cdot \left|\mathbf{P}_y\right|^{-\frac{1}{2}} \\
= & \left|\frac{\mathbf{H_{11}} - \mathbf{H_{12}}\mathbf{H_{22}^{-1}}\mathbf{H_{21}}}{b_{\star\star}}\right|^{-\frac{1}{2}} \cdot |\mathbf{H_{22}}|^{-\frac{1}{2}} \cdot |\mathbf{M}|^{\frac{1}{2}} \cdot \left|b^{-1} \cdot (\mathbf{I}_{T-p} + \widetilde{\mathbf{X}}\mathbf{M^{-1}}\widetilde{\mathbf{X}}')^{-1}\right|^{-\frac{1}{2}} \\
= & (b_{\star\star})^{\frac{d_1}{2}} \cdot |\mathbf{H_{22}}|^{-\frac{1}{2}} \cdot \left|\mathbf{H_{11}} - \mathbf{H_{12}}\mathbf{H_{22}^{-1}}\mathbf{H_{21}}\right|^{-\frac{1}{2}} \cdot |\mathbf{M}|^{\frac{1}{2}} \cdot b^{\frac{T-p}{2}} \cdot \left|\mathbf{I}_{T-p} + \widetilde{\mathbf{X}}\mathbf{M^{-1}}\widetilde{\mathbf{X}}'\right|^{\frac{1}{2}} \\
= & \left(b + QF(\mathbf{y})\right)^{\frac{d_1}{2}} \cdot |\mathbf{H}|^{-\frac{1}{2}} \cdot |\mathbf{M}|^{\frac{1}{2}} \cdot b^{\frac{T-p}{2}} \cdot \left|\mathbf{I}_{T-p} + \widetilde{\mathbf{X}}\mathbf{M^{-1}}\widetilde{\mathbf{X}}'\right|^{\frac{1}{2}}
\end{aligned}
\tag{6.43}
$$

utilizing again the fact that $|\mathbf{H}| = |\mathbf{H_{22}}| \cdot |\mathbf{H_{11}} - \mathbf{H_{12}}\mathbf{H_{22}^{-1}}\mathbf{H_{21}}|$, cf. Greene (2003), Appendix A.5.2, p.823, in the last line. Furthermore from (6.29) and (6.30) it is already known that $|\mathbf{H}|^{-\frac{1}{2}} \cdot |\mathbf{M}|^{\frac{1}{2}} = |\mathbf{I}_{T-p} + \widetilde{\mathbf{X}}\mathbf{M^{-1}}\widetilde{\mathbf{X}}'|^{-\frac{1}{2}}$.

Hence the last equation in (6.43) finally simplifies to

$$\begin{aligned} |\mathbf{P}_1|^{-\frac{1}{2}} \cdot |\mathbf{H_{22}}|^{-\frac{1}{2}} \cdot |\mathbf{M}|^{\frac{1}{2}} \cdot \left|\mathbf{P}_y\right|^{-\frac{1}{2}} &= \left(b + QF(\mathbf{y})\right)^{\frac{d_1}{2}} \cdot b^{\frac{T-p}{2}} \\ &= \left(1 + QF(\mathbf{y})/b\right)^{\frac{d_1}{2}} \cdot b^{\frac{T-p+d_1}{2}} \end{aligned} \tag{6.44}$$

Substituting the last result (6.44) into (6.42) finally gives

$$f(\delta_1,\, \mathbf{y}|\, M_i) \;=\; f(\delta_1|\, \mathbf{y},\, M_i) \cdot f(\mathbf{y}|\, M_i) = \mathscr{T}_{d_1} \cdot \mathscr{T}_{T-p} \tag{6.45}$$

That is the joint density $f(\delta_1,\, \mathbf{y}|\, M_i)$ equals the product of two multivariate Student-t densities.

Hence, given the data, the marginal posterior density under model M_i is proportional to

$$f(\mathbf{B_1}|\, M_i,\, \mathbf{y}) \;\propto\; \left[1 + (\mathbf{B_1} - \mu_{1B})' \cdot \frac{\left(\mathbf{H_{11}} - \mathbf{H_{12}} \cdot \mathbf{H_{22}}^{-1} \cdot \mathbf{H}_{21}\right)}{b_{\star\star}} \cdot (\mathbf{B_1} - \mu_{1B})\right]^{-\frac{a_{\star\star}+d_1}{2}} \tag{6.46}$$

which is the kernel of a d_1-dimensional Student-t density with $a_{\star\star} = T - p + a$ degrees of freedom.

In chapter 4, the vector $\mathbf{B}_1$ corresponds to the vector of PAR coefficients ϕ. Under the Jeffreys prior, used in chapter 4, this kernel then becomes (see also section G.6 for details):

$$f(\phi|\, M_i,\, \mathbf{y}) \;\propto\; \left[1 + (\phi - \widehat{\phi})' \cdot \frac{\left(\mathbf{H_{11}} - \mathbf{H_{12}} \cdot \mathbf{H_{22}}^{-1} \cdot \mathbf{H}_{21}\right)}{\nu s^2} \cdot (\phi - \widehat{\phi})\right]^{-\frac{T-p-d_2}{2}} \tag{6.47}$$

since $T - p - d + d_1 = T - p - d_2$ (see Zellner (1971), p.69, for a similar expression).

G.3. Derivation of the posterior density of ϕ_s

The derivation of the (model specific) marginal posterior density of ϕ_s proceeds along the lines of section G.2, except that now one has to partition

$$\mathbf{B} \;=\; [\mathbf{B_1}' \vdots \mathbf{B_2}']' \;=\; \left[\underset{1\times 1}{\phi_s} \;\vdots\; \phi'_{-s}\,,\; \delta'\right]' \tag{6.48a}$$

This leads to

$$\begin{aligned} f(\phi_s|\, M_i,\, \mathbf{y}) \;&\propto\; \left[1 + (\phi_s - \widehat{\phi}_s)^2 \cdot \underbrace{(H_{11} - \mathbf{H_{12}} \cdot \mathbf{H_{22}^{-1}} \cdot \mathbf{H_{21}})/b_{\star\star}}_{\equiv\, (h^{11})^{-1}}\right]^{-\frac{a_{\star\star}+1}{2}} \\ &\propto\; \left[1 + \frac{(\phi_s - \widehat{\phi}_s)^2}{h^{11}}\right]^{-\frac{a_{\star\star}+1}{2}} \end{aligned} \tag{6.49}$$

which has the form of a univariate Student-t density with $a_{\star\star}$ degrees of freedom, location parameter $\widehat{\phi}_s$ and scale parameter h^{11},where H_{11} denotes the scalar $(1,1)$-element of the matrix $\mathbf{H}$ in (6.20).

G.4. Derivation of the marginal posterior of θ

Next the derivation of the marginal posterior of θ, i.e. the linear approximation of the product of periodic autoregressive coefficients is outlined. First define $\theta \equiv \iota' \cdot \phi$, with $\underset{1\times S}{\iota'} = (1,...,1)$. From section G.1 we know that the conditional posterior distribution of the subvector $\mathbf{B}_1$, or more precisely ϕ, given σ^2, follows a multivariate Normal distribution $N_{d_1}(\mu_{1B},\, \mathbf{\Sigma}_1)$, with mean μ_{1B}, see (6.19), and covariance matrix $\mathbf{\Sigma}_1$, see (6.27).

Given these results, it is straightforward to check that, conditional on σ^2 and a model specification M_i, the linear form θ follows a univariate Normal posterior distribution with first and second moments, under a NIG_2 prior, given by

$$E(\theta|\, \sigma^2,\, M_i,\, \mathbf{y}) = \iota' \cdot \mu_{1B} \tag{6.50a}$$

$$Var(\theta|\, \sigma^2,\, M_i,\, \mathbf{y}) = \sigma^2 \cdot \underbrace{\iota' \cdot \left(\mathbf{H_{11}} - \mathbf{H_{12}} \cdot \mathbf{H_{22}}^{-1} \cdot \mathbf{H}_{21}\right)^{-1} \cdot \iota}_{\equiv\, c} \tag{6.50b}$$

and, under a Jeffreys prior, given by

$$E(\theta|\ \sigma^2,\ M_i,\ \mathbf{y}) = \iota' \cdot \widehat{\phi} \tag{6.51a}$$

$$Var(\theta|\ \sigma^2,\ M_i,\ \mathbf{y}) = \sigma^2 \cdot \underbrace{\iota' \cdot (\widetilde{\mathbf{X}}_{\mathbf{11}} - \widetilde{\mathbf{X}}_{\mathbf{12}} \cdot \widetilde{\mathbf{X}}_{\mathbf{22}}^{-1} \cdot \widetilde{\mathbf{X}}_{\mathbf{21}})^{-1} \cdot \iota}_{\equiv c_J} \tag{6.51b}$$

The marginal posterior distribution of the linear form θ, under model M_i, can be obtained by integrating the joint posterior

$$f(\theta, \sigma^2|\ M_i,\ \mathbf{y}) = f(\theta|\ \sigma^2,\ M_i,\ \mathbf{y}) \cdot f(\sigma^2|\ M_i,\ \mathbf{y})$$

with respect to σ^2, using properties of the IG_2 distribution, which yields[3]

$$f(\theta|\ M_i,\ \mathbf{y}) = \frac{\Gamma\left(\frac{a_{\star\star}+1}{2}\right)}{\Gamma\left(\frac{a_{\star\star}}{2}\right) \cdot \sqrt{\pi}} \cdot (b_{\star\star} \cdot c)^{-\frac{1}{2}} \cdot \left[1 + \frac{(\theta - \iota' \cdot \mu_{1B})^2}{b_{\star\star} \cdot c}\right]^{-\frac{a_{\star\star}+1}{2}} \tag{6.52}$$

which is a univariate Student-t density with $a_{\star\star}$ degrees of freedom, mean $\iota' \cdot \mu_{1B}$ and variance $b_{\star\star} \cdot c/(a_{\star\star} - 2)$, see Bauwens et al. (1999), Appendix A.1.4, p.294.

Whereas under a Jeffreys prior, omitting for convenience all terms independent of θ, this simplifies to

$$f(\theta|\ M_i,\ \mathbf{y}) \propto \left[1 + \frac{(\theta - \widehat{\theta})^2}{\nu s^2 \cdot c_J}\right]^{-\frac{T-p-d+1}{2}} \tag{6.53}$$

which is the kernel of a univariate t-density with $T - p - d$ degrees of freedom, mean $\widehat{\theta} = \iota' \cdot \widehat{\phi}$, i.e. the ordinary least squares estimate of θ, and variance $\nu s^2 \cdot c_J/(T - p - d - 2)$, see Zellner (1971), p.70, for a similar expression.

G.5. Derivation of the marginal posterior of Q

Next consider the linear form $\mathbf{R}\phi = \mathbf{r}$, with $\mathbf{R}$ a $J \times S$ matrix of linear contrasts and $\mathbf{r}$ a J-vector of constants, where $J = S$ in the following. Utilizing some definitions and arguments used in the derivation of the posterior of θ, it is known that under a NIG_2 prior

[3]Details are omitted to save space.

$f(\phi|\ \sigma^2,\ M_i,\ \mathbf{y}) = N_{d_1}(\mu_{1B},\ \mathbf{\Sigma}_1)$. Hence, conditional on the data and all other parameters, it follows that

$$\mathbf{R}\phi|\ \sigma^2,\ M_i\ ,\mathbf{y}\ \sim\ N_S(\mathbf{R}\cdot\mu_{1B}\ ,\ \underbrace{\sigma^2\cdot\mathbf{R}(\mathbf{H}_{11}-\mathbf{H}_{12}\cdot\mathbf{H}_{22}^{-1}\cdot\mathbf{H}_{21})^{-1}\mathbf{R}'}_{\equiv\ \mathbf{\Omega}}) \tag{6.54}$$

or under a Jeffreys prior:

$$\mathbf{R}\phi|\ \sigma^2,\ M_i\ ,\mathbf{y}\ \sim\ N_S(\mathbf{R}\cdot\widehat{\phi},\ \sigma^2\cdot\mathbf{R}(\widetilde{\mathbf{X}}_{\mathbf{11}}-\widetilde{\mathbf{X}}_{\mathbf{12}}\cdot\widetilde{\mathbf{X}}_{\mathbf{22}}^{-1}\cdot\widetilde{\mathbf{X}}_{\mathbf{21}})^{-1}\mathbf{R}')$$

Further it is known, using (6.54), that the quadratic form

$$Z\equiv[\mathbf{R}(\phi-\mu_{1B})]'\cdot\mathbf{\Omega}^{-1}\cdot[\mathbf{R}(\phi-\mu_{1B})] \tag{6.55}$$

is $\chi^2_{(S)}$- distributed with S degrees of freedom, cf. Kendall and Stuart (1969), chapter 15.

Following Hamilton (1994), p.369, I make the simple change of variables

$$Q\equiv Z\cdot\sigma^2\cdot a_{\star\star}/(S\cdot b_{\star\star}) \tag{6.56}$$

with Jacobian equal to $\left|S\cdot b_{\star\star}/(\sigma^2\cdot a_{\star\star})\right|$.

Since $Z|\sigma^2,\mathbf{y}\sim\chi^2_{(S)}$, or equivalently $Z|\sigma^2,\mathbf{y}\sim G(\frac{S}{2},\ 2)$, it is easy to show that $Q|\sigma^2,\mathbf{y}$ follows a Gamma-2 distribution, $G_2\left(S,\ d\cdot b_{\star\star}/(\sigma^2\cdot a_{\star\star})\right)$, given all other parameters and the data.[4] The posterior $f(Q|\ M_i,\ \mathbf{y})$ can then be obtained by integrating the joint posterior distribution

$$f(Q,\ \sigma^2|\ M_i,\ \mathbf{y}) = f(Q|\ \sigma^2,\ M_i,\ \mathbf{y})\cdot f(\sigma^2|\ M_i,\ \mathbf{y}) \tag{6.57}$$

with respect to σ^2, using properties of the IG_2 density, where as above $f(\sigma^2|\ M_i,\ \mathbf{y}) = IG_2(a_{\star\star},\ b_{\star\star})$.

After some algebra, which is omitted here to save space (see Hamilton (1994), Appendix 12.A, p.370, for some details) it can be shown that the marginal posterior distribution of the quadratic form Q in (6.56), under model M_i, follows an F-distribution with S and

[4]Note that if $X\sim G_2(\nu,s)\Leftrightarrow X\sim G(\frac{\nu}{2},\frac{2}{s})$, see Bauwens et al. (1999), A.15, for details.

$a_{\star\star} = T - p + a$ degrees of freedom, i.e.

$$Q|\, M_i, \mathbf{y} \sim F(\nu_1 = S,\ \nu_2 = a_{\star\star})$$

Whereas under a Jeffreys prior, i.e. if $(b,\ \mathbf{M}) \to 0$ and $a \to -d$, then $Q|\, M_i, \mathbf{y} \sim F(\nu_1 = S,\ \nu_2 = T - p - d)$, which therefore yields the same result as within a classical framework.

G.6. Some comments on the prior distribution

As already mentioned briefly at the very beginning of this appendix the assumed (non-informative) Jeffreys prior of chapter 4 can be obtained as a special case of the natural conjugate NIG_2 prior (see (6.3) above). Recall that the kernels of the multivariate Normal and the IG_2 density, respectively, are given by

$$k(\mathbf{B}|\ \sigma^2) = \sigma^{-d} \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{B}-\mathbf{B_0})'\mathbf{M}(\mathbf{B}-\mathbf{B_0})\right\} \tag{6.58a}$$

$$k(\sigma^2) = \sigma^{-(a+2)} \cdot \exp\left\{-\frac{b}{2\sigma^2}\right\} \tag{6.58b}$$

By letting the scale matrix $\mathbf{M}$ and the shape parameter b both go to zero, we find the noninformative priors by computing the limits of the respective kernels (not the limit of the densities which are trivially equal to zero through their respective normalizing constants). Letting $\mathbf{M} \to 0$, which corresponds to a zero precision, the kernel in (6.58a) becomes

$$k(\mathbf{B}|\ \sigma^2) = \sigma^{-d} \tag{6.59}$$

In addition, letting $b \to 0$ the kernel of the IG_2 prior in (6.58b) becomes

$$k(\sigma^2) = \sigma^{-(a+2)} \tag{6.60}$$

which has different interpretations depending on the choice for the parameter a, see Bauwens et al. (1999), p.114, for a discussion. Obviously from (6.59) and (6.60) the kernel of the joint diffuse prior is equal to

$$k(\mathbf{B},\ \sigma^2) = \sigma^{-(a+d+2)} \tag{6.61}$$

Now by letting $a \to -d$, Jeffreys' prior can be obtained, see ibid.:

$$k(\mathbf{B},\ \sigma^2) = \sigma^{-2} \tag{6.62}$$

As already mentioned in the previous sections G.2-G.5 the corresponding posterior expressions under a noninformative Jeffreys prior can be easily obtained by letting the IG_2

hyperparameters: $b \to 0$, in $b_{\star\star}$, and $a \to -d$, in $a_{\star\star}$, and letting $\mathbf{M} \to 0$ for the precision of $\mathbf{B}$.

Another prior specification, used in an earlier version of the paper presented in chapter 4, is Zellner's g-prior (Zellner (1986)), which has become quite popular because of its analytical convenience, cf. Zellner and Siow (1980), Chipman et al. (2001), Liang et al. (2008), Garcia-Donato and Martinez-Beneito (2012), among others. Zellner's g-prior allows the experimenter to introduce information about the location parameter of a regression, but to bypass the most difficult part of the prior specification, namely the specification of the prior correlation structure. This structure is data-dependent and thus fixed in Zellner's approach as (see Judge et al. (1985), Marin and Robert (2010))

$$\mathbf{B}|\ \sigma^2,\ \widetilde{\mathbf{X}} \sim N_d(\mathbf{B}_0,\ g \cdot \sigma^2 (\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})^{-1}) \qquad \text{and} \qquad f(\sigma^2|\ \widetilde{\mathbf{X}}) \propto \sigma^{-2} \tag{6.63}$$

All the posterior results of the previous sections could easily be adapted for the case of a g-prior simply by substituting $\mathbf{M} = g^{-1} \cdot (\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})$. For example, the marginal likelihood in (6.35), under model M_i, then becomes

$$f(\mathbf{y}|\ M_i) = \frac{\Gamma(\frac{T-p+a}{2})}{\pi^{(\frac{T-p}{2})} \cdot \Gamma(\frac{a}{2})} \cdot b^{-\frac{T-p}{2}} \cdot (g+1)^{-\frac{d}{2}} \cdot [\ 1 + QF(\mathbf{y})/b\ ]^{-\frac{T-p+a}{2}}$$

with

$$QF(\mathbf{y}) = (\mathbf{y} - \widetilde{\mathbf{X}}\mathbf{B}_0)' \left(\mathbf{I}_{T-p} - \frac{g}{g+1} \widetilde{\mathbf{X}} \cdot (\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})^{-1} \cdot \widetilde{\mathbf{X}}' \right) \cdot (\mathbf{y} - \widetilde{\mathbf{X}}\mathbf{B}_0)$$

However one potential drawback with this approach is that it introduces the additional unknown hyperparameter g, which has to be integrated out numerically. Furthermore assuming a g-prior did not yield better results in the empirical application presented in chapter 4 compared to the results using a Jeffreys prior.

H. Technical details - chapter 5

H.1. Derivation of the posterior predictive distribution of $\widetilde{\mathbf{y}}_K$

In the following it will be shown that the results in Broemeling and Land (1984), stated for a *nonperiodic* AR(p) model, also apply to the more general case of a *periodic* AR(p) model of the form (5.1) as in section 5.2.[5] The subsequent results apply to the case of a model without a structural break or with a break at known date T_B.

The joint probability density function (pdf) of $\mathbf{B}$, σ^2, $\widetilde{\mathbf{y}}_K$, given the data $\mathbf{y}$ and a particular model specification M_i, can be factorized as (cf. Zellner (1971), p.72):

$$f(\widetilde{\mathbf{y}}_K, \mathbf{B}, \sigma^2 |\, M_i,\, \widetilde{\mathbf{X}},\, \widetilde{\mathbf{W}}_K,\, \mathbf{y}) = f(\mathbf{B}, \sigma^2 |\, M_i,\, \widetilde{\mathbf{X}},\, \mathbf{y}) \cdot f(\widetilde{\mathbf{y}}_K |\, M_i,\, \mathbf{B},\, \sigma^2, \widetilde{\mathbf{W}}_K,\, \mathbf{y}) \quad (6.64)$$

In the subsequent conditioning on $\widetilde{\mathbf{X}}$ and M_i will be omitted in order to simplify the notation. It is important to keep in mind that the matrix $\widetilde{\mathbf{W}}_K$ contains deterministic future values, but more importantly, also lagged future values y_{T+K-p}, for $p = 1...p_{max}$, when a K-step ahead forecast is considered. In the following it will be useful to partition the vector of unknown future observations according to

$$\widetilde{\mathbf{y}}_K = \begin{pmatrix} \widetilde{\mathbf{y}}_{K-1} \\ y_{T+K} \end{pmatrix} \quad (6.65)$$

where $\widetilde{\mathbf{y}}_{K-1}$ is a subvector of dimension $K-1$ and y_{T+K} the unknown future value at $t = T+K$.

Further partition the right-hand side variables of the prediction equation (5.7) (see section 5.2.2, p.82), contained in the matrix $\widetilde{\mathbf{W}}_K$, conformably to (6.65):

$$\widetilde{\mathbf{W}}_K = \begin{pmatrix} \widetilde{\mathbf{W}}_{K-1} \\ W_{(K)} \end{pmatrix} \quad (6.66)$$

with $\widetilde{\mathbf{W}}_{K-1}$ a submatrix of dimension $K-1 \times d$ and $W_{(K)}$ a vector of dimension $1 \times d$ which is just the K-th row of the matrix $\widetilde{\mathbf{W}}_K$.

[5] Note that a nonperiodic AR model is a special case of a periodic AR model, namely for $\phi_s = \phi$, $\forall s$.

Note that for $K = 1$, $\widetilde{\mathbf{y}}_1 = y_{T+1}$ and $\widetilde{\mathbf{W}}_1 = W_{(1)}$. The main results are summarized in the following proposition, which generalizes Theorem 1, in Broemeling and Land (1984), to the periodic case.

Proposition 1. *If $\{y_t : t = 1...T\}$ is a trajectory of a Gaussian PAR(p) process with unknown parameters $\mathbf{B} = (\phi',\ \delta')' \in \mathbb{R}^d$ and $\sigma^2 \in \mathbb{R}^+$, where the vector $\mathbf{y}$ contains the last $T - p$ observations and $\mathbf{y}_0$ contains the first p observations, and if unknown future values $\{y_{T+k} : k = 1...K\}$, contained in the vector $\widetilde{\mathbf{y}}_K$, are generated by the same process as the observations, and $\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}$ and $\widetilde{\mathbf{W}}'_K\widetilde{\mathbf{W}}_K$ are symmetric positive definite matrices, and the joint prior distribution of $(\mathbf{B},\ \sigma^2)$ is of the Normal-Inverse-Gamma-2 form with parameters $\mathbf{B}_0 \in \mathbb{R}^d$, $\mathbf{V}$ a symmetric positive definite scale matrix and $a, b > 0$, then the posterior predictive distribution of $\widetilde{\mathbf{y}}_K$ given $\mathbf{y}$ and $\mathbf{y}_0$, can be expressed as the product K univariate predictive densities, namely the marginal density of y_{T+1}, the conditional predictive density of y_{T+2} given y_{T+1} and so on, where these densities have the following form:*[6]

- *If $K = 1$, the predictive density of y_{T+1} is a Student-t density with $\nu^\star = T - p + a$ degrees of freedom, mean*

$$E(y_{T+1} |\ \mathbf{y}) = D_1^{-1} \cdot E_1 \tag{6.67}$$

and variance

$$Var(y_{T+1} |\ \mathbf{y}) = \frac{F_1 - E_1' \cdot D_1^{-1} \cdot E_1}{(\nu^\star - 2) \cdot D_1} \tag{6.68}$$

where

$$\begin{aligned} D_1 &= 1 - W_{(1)} \cdot \mathbf{R}^{-1} \cdot W'_{(1)} \\ E_1 &= W_{(1)} \cdot \mathbf{R}^{-1} \cdot \mathbf{F}_1 \\ F_1 &= \mathbf{S} - y_{T+1}^2 - \mathbf{F}_1' \cdot \mathbf{R}^{-1} \cdot \mathbf{F}_1 \end{aligned} \tag{6.69}$$

[6] Conditioning on $\mathbf{y}_0$ and $\widetilde{\mathbf{X}}$ is omitted.

with

$$\mathbf{R} = \mathbf{V}^{-1} + \widetilde{\mathbf{X}}' \cdot \widetilde{\mathbf{X}} + \widetilde{\mathbf{W}}'_K \cdot \widetilde{\mathbf{W}}_K \tag{6.70a}$$

$$\mathbf{S} = \mathbf{y}' \cdot \mathbf{y} + \widetilde{\mathbf{y}}'_K \cdot \widetilde{\mathbf{y}}_K + \mathbf{B}'_0 \cdot \mathbf{V}^{-1} \cdot \mathbf{B}_0 + b \tag{6.70b}$$

$$\mathbf{F}_1 = \mathbf{V}^{-1}\mathbf{B}_0 + \widetilde{\mathbf{X}}'\mathbf{y} \tag{6.70c}$$

- *In general, if $K > 1$, the conditional predictive density of y_{T+K} given $\widetilde{\mathbf{W}}_K$ (see* (5.6) *in section 5.2.2, p.82) is a Student-t density with $\nu^\star = T - p + a + K - 1$ degrees of freedom, mean*

$$E(y_{T+K} | \, \widetilde{\mathbf{W}}_K, \, \mathbf{y}) = D_K^{-1} \cdot E_K \tag{6.71}$$

and variance

$$Var(y_{T+K} | \, \widetilde{\mathbf{W}}_K, \, \mathbf{y}) = \frac{F_K - E'_K \cdot D_K^{-1} \cdot E_K}{(\nu^\star - 2) \cdot D_K} \tag{6.72}$$

where

$$\begin{aligned} D_K &= 1 - W_{(K)} \cdot \mathbf{R}^{-1} \cdot W'_{(K)} \\ E_K &= W_{(K)} \cdot \mathbf{R}^{-1} \cdot \left(\mathbf{F}_1 + \widetilde{\mathbf{W}}'_{K-1} \cdot \widetilde{\mathbf{y}}_{K-1} \right) \\ F_K &= \mathbf{S} - y^2_{T+K} - \left(\mathbf{F}_1 + \widetilde{\mathbf{W}}'_{K-1} \cdot \widetilde{\mathbf{y}}_{K-1} \right)' \mathbf{R}^{-1} \left(\mathbf{F}_1 + \widetilde{\mathbf{W}}'_{K-1} \cdot \widetilde{\mathbf{y}}_{K-1} \right) \end{aligned} \tag{6.73}$$

Proof:

Writing out (6.64) more explicitly (omitting the normalizing constants for convenience) the kernel of this joint pdf is given by

$$\begin{aligned} & f(\widetilde{\mathbf{y}}_K, \mathbf{B}, \, \sigma^2 \mid M_i, \, \mathbf{y}) \\ & \propto (\sigma^2)^{-\frac{T-p}{2}} \cdot \exp\left\{ -\frac{1}{2\sigma^2} (\mathbf{y} - \widetilde{\mathbf{X}} \cdot \mathbf{B})' \cdot (\mathbf{y} - \widetilde{\mathbf{X}} \cdot \mathbf{B}) \right\} \cdot (\sigma^2)^{-\frac{a+2}{2}} \cdot \exp\left\{ -\frac{b}{2\sigma^2} \right\} \cdot \\ & \quad (\sigma^2)^{-\frac{d}{2}} \cdot \exp\left\{ -\frac{1}{2\sigma^2} (\mathbf{B} - \mathbf{B_0})' \cdot \mathbf{V}^{-1} \cdot (\mathbf{B} - \mathbf{B_0}) \right\} \cdot \\ & \quad (\sigma^2)^{-\frac{K}{2}} \cdot \exp\left\{ -\frac{1}{2\sigma^2} (\widetilde{\mathbf{y}}_K - \widetilde{\mathbf{W}}_K \cdot \mathbf{B})' \cdot (\widetilde{\mathbf{y}}_K - \widetilde{\mathbf{W}}_K \cdot \mathbf{B}) \right\} \end{aligned}$$

(6.74)

with $d = (4+p)\cdot S\ (= dim(\mathbf{B}))$.

First consider only the exponent of (6.74), which is equal to

$$\begin{aligned}(\mathbf{y}-\widetilde{\mathbf{X}}\cdot\mathbf{B})'\cdot(\mathbf{y}-\widetilde{\mathbf{X}}\cdot\mathbf{B})+(\widetilde{\mathbf{y}}_K-\widetilde{\mathbf{W}}_K\cdot\mathbf{B})'\cdot(\widetilde{\mathbf{y}}_K-\widetilde{\mathbf{W}}_K\cdot\mathbf{B})\\ +(\mathbf{B}-\mathbf{B_0})'\cdot\mathbf{V}^{-1}\cdot(\mathbf{B}-\mathbf{B_0})+b\end{aligned} \tag{6.75}$$

Next complete the square with respect to the vector $\mathbf{B}$. By using the matrix $\mathbf{R}$ given in (6.70a), the exponent (6.75) can be expressed more compactly as

$$=\mathbf{B}'\cdot\mathbf{R}\cdot\mathbf{B}-2\cdot\mathbf{B}'(\underbrace{\widetilde{\mathbf{X}}'\mathbf{y}+\mathbf{V}^{-1}\mathbf{B}_0+\widetilde{\mathbf{W}}'_K\cdot\widetilde{\mathbf{y}}_K}_{\equiv\,\mathbf{F}})+\underbrace{\mathbf{y}'\mathbf{y}+\widetilde{\mathbf{y}}'_K\widetilde{\mathbf{y}}_K+\mathbf{B}'_0\mathbf{V}^{-1}\mathbf{B}_0+b}_{\equiv\,\mathbf{S},\text{ see (6.70b)}} \tag{6.76}$$

Applying a Cholesky decomposition to the matrix $\mathbf{R}=\mathbf{L}\mathbf{L}'$, with $\mathbf{L}$ the corresponding lower triangular matrix, and defining $\widetilde{\mathbf{B}}\equiv\mathbf{L}'\mathbf{B}\ \Leftrightarrow\ \mathbf{B}=(\mathbf{L}')^{-1}\widetilde{\mathbf{B}}$, then after completing the square with respect to $\widetilde{\mathbf{B}}$, equation (6.76) can be written as

$$=(\widetilde{\mathbf{B}}-\mathbf{L}^{-1}\cdot\mathbf{F})'\cdot(\widetilde{\mathbf{B}}-\mathbf{L}^{-1}\cdot\mathbf{F})-\mathbf{F}'\cdot\mathbf{R}^{-1}\cdot\mathbf{F}+\mathbf{S} \tag{6.77}$$

and by substituting $\widetilde{\mathbf{B}}=\mathbf{L}'\mathbf{B}$ and some algebra, (6.77) is seen to be

$$=(\mathbf{B}-\mathbf{R}^{-1}\cdot\mathbf{F})'\cdot\mathbf{R}\cdot(\mathbf{B}-\mathbf{R}^{-1}\cdot\mathbf{F})+\mathbf{S}-\mathbf{F}'\cdot\mathbf{R}^{-1}\cdot\mathbf{F} \tag{6.78}$$

where in the following the quadratic form in $\mathbf{B}$ is denoted by $QF(\mathbf{B})$.

Plugging (6.78) into the exponent of the joint pdf (6.74) the latter becomes

$$f(\widetilde{\mathbf{y}}_K,\,\mathbf{B},\,\sigma^2\mid\mathbf{y})\propto(\sigma^2)^{-\frac{T-p+K+d+a+2}{2}}\cdot\exp\left\{-\frac{1}{2\sigma^2}[\,QF(\mathbf{B})+\mathbf{S}-\mathbf{F}'\cdot\mathbf{R}^{-1}\cdot\mathbf{F}\,]\right\} \tag{6.79}$$

Now integrate out $\mathbf{B}$ from (6.79) using properties of the multivariate Normal density

$$f(\widetilde{\mathbf{y}}_K,\ \sigma^2 \mid \mathbf{y}) \propto \int_{\mathbb{R}^d} |\mathbf{R}|^{\frac{1}{2}} \cdot (\sigma^2)^{-\frac{d}{2}} \cdot \exp\left\{-\frac{1}{2\sigma^2} \cdot QF(\mathbf{B})\right\} \cdot d\mathbf{B}$$
$$\cdot |\mathbf{R}|^{-\frac{1}{2}} \cdot (\sigma^2)^{-\frac{T-p+K+a+2}{2}} \cdot \exp\left\{-\frac{1}{2\sigma^2}[\ \mathbf{S} - \mathbf{F}' \cdot \mathbf{R}^{-1} \cdot \mathbf{F}\ ]\right\} \tag{6.80}$$

and write $b^\star \equiv \mathbf{S} - \mathbf{F}' \cdot \mathbf{R}^{-1} \cdot \mathbf{F}$ henceforth.

Next integrate out σ^2 from (6.80) by using the properties of the IG_2 density:

$$f(\widetilde{\mathbf{y}}_K \mid \mathbf{y}) \ \propto\ |\mathbf{R}|^{-\frac{1}{2}} \cdot C_g \int_{\mathbb{R}^+} C_g^{-1} \cdot (\sigma^2)^{-\frac{a^\star+2}{2}} \cdot \exp\left\{-\frac{b^\star}{2\sigma^2}\right\} \cdot d\sigma^2 \tag{6.81}$$

with $a^\star \equiv T - p + K + a$ and $C_g(\frac{a^\star}{2};\ \frac{2}{b^\star})$ the normalizing constant of the Gamma-2 distribution, see Bauwens et al. (1999), p.292.

In the following the main task is to obtain the posterior predictive distribution of $\widetilde{\mathbf{y}}_K$ under model M_i. By subsuming all terms independent of $\widetilde{\mathbf{y}}_K$ into the proportionality sign the expression in (6.81) is proportional to

$$\begin{aligned} f(\widetilde{\mathbf{y}}_K \mid \mathbf{y}) &\propto |\ \mathbf{R}\ |^{-\frac{1}{2}} \cdot (b^\star)^{-\frac{a^\star}{2}} \\ &= |\ \mathbf{R}\ |^{-\frac{1}{2}} \cdot\ (\ \mathbf{S} - \mathbf{F}' \cdot \mathbf{R}^{-1} \cdot \mathbf{F}\ )^{-\frac{a^\star}{2}} \end{aligned} \tag{6.82}$$

Furthermore by proceeding as in Zellner (1971), p.72 f., we can express the last factor of (6.82) in terms of quadratic forms, by completing the square with respect to $\widetilde{\mathbf{y}}_K$, $\mathbf{y}$ and $\mathbf{B}_0$, which after some algebra[7] finally results in

$$f(\widetilde{\mathbf{y}}_K \mid \mathbf{y}) \propto |\ \mathbf{R}\ |^{-\frac{1}{2}} \cdot\ [\ QF(\mathbf{B}_0) + QF(\mathbf{y}) + QF(\widetilde{\mathbf{y}}_K)\ ]^{-\frac{a^\star}{2}} \tag{6.83a}$$

$$= |\ \mathbf{R}\ |^{-\frac{1}{2}} \cdot c^{-\frac{a^\star}{2}} \cdot\ [\ 1 + QF(\widetilde{\mathbf{y}}_K)/c\ ]^{-\frac{a^\star}{2}} \tag{6.83b}$$

with $QF(.)$ the respective quadratic forms and $c \equiv QF(\mathbf{y}) + QF(\mathbf{B}_0)$.

[7] Details are omitted here in order to save space, cf. ibid.

For an ease of reference $QF(\widetilde{\mathbf{y}}_K)$ is stated here explicitly without a proof:

$$QF(\widetilde{\mathbf{y}}_K) \;=\; \left(\widetilde{\mathbf{y}}_K - \widetilde{\mathbf{W}}_K \cdot \widehat{\beta}_B\right)' \cdot \mathbf{E}_\star \cdot \left(\widetilde{\mathbf{y}}_K - \widetilde{\mathbf{W}}_K \cdot \widehat{\beta}_B\right) \tag{6.84}$$

with scale matrix[8]

$$\mathbf{E}_\star \equiv \mathbf{I}_K - \widetilde{\mathbf{W}}_K \cdot \mathbf{R}^{-1} \cdot \widetilde{\mathbf{W}}'_K \tag{6.85a}$$

$$= \left(\mathbf{I}_K + \widetilde{\mathbf{W}}_K \cdot (\mathbf{V}^{-1} + \widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})^{-1} \cdot \widetilde{\mathbf{W}}'_K\right)^{-1} \tag{6.85b}$$

and

$$\widehat{\beta}_B \;\equiv\; \left(\mathbf{V}^{-1} + \widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1} \cdot \left(\mathbf{V}^{-1}\mathbf{B}_0 + \widetilde{\mathbf{X}}'\mathbf{y}\right) \tag{6.86}$$

the usual Bayes estimate of $\mathbf{B}$ under quadratic loss, cf. Bauwens et al. (1999), p.58.[9]

At first sight the expression in (6.83b) resembles that given in Zellner (1971), p.73, which has the form of a multivariate t-density (see also Bauwens et al. (1999), Theorem 2.25, p.61). However when examining (6.83b) in some more detail we recognize that the matrix $\mathbf{R} = \mathbf{V}^{-1} + \widetilde{\mathbf{X}}'\widetilde{\mathbf{X}} + \widetilde{\mathbf{W}}'_K\widetilde{\mathbf{W}}_K$, and hence the determinant, depends on lagged future values y_t, $t \leq T + K - 1$, through the matrix $\widetilde{\mathbf{W}}_K$. As a consequence the last factor in (6.83b) is not the kernel of a multivariate t-density, since a subset of the values in $\widetilde{\mathbf{W}}_K$ contains lagged future values of y_t. Hence the posterior predictive distribution of a PAR model is not a multivariate t-density as in the case of a standard (non-autoregressive) linear regression model, see for example Judge et al. (1985), p.122, Gelman et al. (1995), p.239, among others.

Next the posterior predictive distribution $f(\widetilde{\mathbf{y}}_k|\ \mathbf{y})$ is examined in more detail. As noted by Broemeling and Land (1984), p.1309 (henceforth abbreviated by 'BL'), for the case of a nonperiodic AR(p) model, and as can also be seen from (6.82), the Bayesian predictive

[8]The last equation (6.85b) follows from an application of the Woodbury matrix identity by setting $\mathbf{A} = \mathbf{I}_K$, $\mathbf{U} = \widetilde{\mathbf{W}}_K$, $\mathbf{V} = \widetilde{\mathbf{W}}'_K$ and $\mathbf{C} = (\mathbf{V}^{-1} + \widetilde{\mathbf{X}}'\widetilde{\mathbf{X}})^{-1}$ in $(\mathbf{A} + \mathbf{U} \cdot \mathbf{C} \cdot \mathbf{V})^{-1}$, see footnote 2 on page 185 of section G.1.

[9]A similar expression as that in (6.84) can be found in Bauwens et al. (1999), Theorem 2.25, p.61., using (6.85b), or in Zellner (1971), p.73, when using a joint diffuse prior for $\mathbf{B}$ and σ^2.

distribution of $\widetilde{\mathbf{y}}_K$, for $K \geq 1$, can be factorized according to:

$$f(\widetilde{\mathbf{y}}_K \mid \mathbf{y}) = g_1(\widetilde{\mathbf{y}}_{K-1},\, \mathbf{y}) \cdot g_2(\widetilde{\mathbf{y}}_K\,,\, \mathbf{y}) \quad , \qquad \widetilde{\mathbf{y}}_K \in \mathbb{R}^K \tag{6.87}$$

with

$$g_1(\widetilde{\mathbf{y}}_{K-1},\, \mathbf{y}) \propto \mid \mathbf{R} \mid^{-\frac{1}{2}} \tag{6.88a}$$

$$g_2(\widetilde{\mathbf{y}}_K\,,\, \mathbf{y}) \propto (\, \mathbf{S} - \mathbf{F}' \cdot \mathbf{R}^{-1} \cdot \mathbf{F}\,)^{-\frac{T-p+K+a}{2}} \tag{6.88b}$$

For the sake of clarity consider next the matrix $\widetilde{\mathbf{W}}_K$ in more detail. Note that this matrix consists of two submatrices, see (5.6) in chapter 5. Below only the first of these two submatrices, namely $\mathbf{X}_K$, is indicated here for $s = 1,2$ explicitly. This matrix is given by

$$\begin{bmatrix}
D_{1,T+1} \cdot y_T & \cdots & D_{1,T+1} \cdot y_{T+1-p} & D_{2,T+1} \cdot y_T & \cdots & D_{2,T+1} \cdot y_{T+1-p} & \cdots \\
D_{1,T+2} \cdot y_{T+1} & \cdots & D_{1,T+2} \cdot y_{T+2-p} & D_{2,T+2} \cdot y_{T+1} & \cdots & D_{2,T+2} \cdot y_{T+2-p} & \cdots \\
D_{1,T+3} \cdot y_{T+2} & \cdots & D_{1,T+3} \cdot y_{T+3-p} & D_{2,T+3} \cdot y_{T+2} & \cdots & D_{2,T+3} \cdot y_{T+3-p} & \cdots \\
\vdots & & \vdots & \vdots & & \vdots & \\
D_{1,T+K} \cdot y_{T+K-1} & \cdots & D_{1,T+K} \cdot y_{T+K-p} & D_{2,T+K} \cdot y_{T+K-1} & \cdots & D_{2,T+K} \cdot y_{T+K-p} & \cdots
\end{bmatrix}$$

with dimension $K \times d$ and $D_{s,t}$ equals one if observation t falls in season s, and equals zero otherwise.[10]

For $K = 1$ and a sample $\mathbf{y}$, the first term $g_1(\widetilde{\mathbf{y}}_0,\, \mathbf{y})$ in (6.87) does not depend on y_{T+1} since $\widetilde{\mathbf{W}}_1$ contains only y_t-terms with $t < T+1$, and $g_2(\widetilde{\mathbf{y}}_1,\, \mathbf{y})$ only depends on $\widetilde{\mathbf{y}}_1 = y_{T+1}$, given $\mathbf{y}$. Note that if $K = 1$, then $\widetilde{\mathbf{y}}_{K-1} = \widetilde{\mathbf{y}}_0 = \mathbf{0}$ and $\widetilde{\mathbf{W}}_K = W_{(K)}$, since $\widetilde{\mathbf{W}}_{K-1} = \mathbf{0}$, see expressions (6.65) and (6.66). Hence the marginal posterior predictive pdf for a one-step ahead forecast is

$$f(y_{T+1} \mid \mathbf{y}) \propto g_2(y_{T+1},\, \mathbf{y}) \tag{6.89}$$

and will be shown to have the form of a univariate Student-t density. In this case, g_2 equals the kernel and g_1 equals the normalizing constant of a univariate t-density.

For $K = 2$, the first term $g_1(\widetilde{\mathbf{y}}_1,\, \mathbf{y})$ does not depend on y_{T+2}, since both $\widetilde{\mathbf{y}}_1$ and $\widetilde{\mathbf{W}}_2$ contain

[10] Where the subsequent discussion follows that in Broemeling and Land (1984) for the nonperiodic case.

only y_t-terms with $t < T+2$, and the second term, $g_2(\widetilde{\mathbf{y}}_2,\ \mathbf{y})$, depends on both y_{T+1} and y_{T+2} through $\widetilde{\mathbf{y}}_2$. Hence for given values $y_1,...,y_T,y_{T+1}$, the conditional posterior predictive distribution of a two-step ahead forecast, $f(y_{T+2}|\ y_{T+1},\ \mathbf{y})$, is $g_2(\widetilde{\mathbf{y}}_2,\ \mathbf{y})$, which will be seen to have the form of a conditional t-density. Since the joint posterior predictive pdf for $K = 2$ can be factorized as

$$f(\widetilde{\mathbf{y}}_2|\ \mathbf{y}) = f(y_{T+2}|\ y_{T+1},\ \mathbf{y}) \cdot f(y_{T+1}|\ \mathbf{y}) \tag{6.90}$$

it follows that this joint pdf is given by the product of two univariate t-densities (a marginal and a conditional).

In general, for $K \geq 2$ the first term $g_1(\widetilde{\mathbf{y}}_{K-1},\ \mathbf{y})$ in (6.87) does not depend on y_{T+K} since $\widetilde{\mathbf{y}}_{K-1}$ and $\widetilde{\mathbf{W}}_K$ contain only y_t-terms with $t \leq T+K-1$, and $g_2(\widetilde{\mathbf{y}}_K,\ \mathbf{y})$ depends on $y_{T+1},...,y_{T+K}$ through $\widetilde{\mathbf{y}}_K$. Hence for given values $\widetilde{\mathbf{y}}_{K-1}$ and $\mathbf{y}$, the conditional posterior predictive pdf for a K-step ahead forecast, $f(y_{T+K}|\ \widetilde{\mathbf{W}}_K,\ \mathbf{y})$, is equal to $g_2(\widetilde{\mathbf{y}}_K,\ \mathbf{y})$, which will be shown below to be proportional to a t-density kernel. Since the joint posterior predictive pdf for $K \geq 1$ can be factorized as

$$f(\widetilde{\mathbf{y}}_K|\ \mathbf{y}) = \prod_{k=1}^{K} f(y_{T+k}|\ \widetilde{\mathbf{y}}_{k-1},\ \mathbf{y}) \quad \text{, with } \widetilde{\mathbf{y}}_0 = \mathbf{0} \tag{6.91}$$

it follows that the posterior predictive pdf of $\widetilde{\mathbf{y}}_K$ can be expressed as the product of K univariate t-densities, viz. a marginal t-density for y_{T+1} and $K-1$ conditional t-densities. However the crucial point here is that this pdf is not the standard K-variate Student-t density, as in Zellner (1971), p.73, for the case of a (nonautoregressive) linear regression model. The same result has been established by Broemeling and Land (1984) for the nonperiodic case.

Next it will be shown that for $K = 1$ the marginal posterior predictive density of y_{T+1} is a Student-t density with $T-p+a$ degrees of freedom and that for $K > 1$ the posterior predictive density of y_{T+K}, given $\widetilde{\mathbf{W}}_K$, is a conditional Student-t density with $T-p+a+K-1$ degrees of freedom.

First write out the second term of (6.82) explicitly:

$$\begin{aligned}\mathbf{S}-\mathbf{F}'\cdot\mathbf{R}^{-1}\cdot\mathbf{F} \;=\; & \mathbf{y}'\mathbf{y}+\widetilde{\mathbf{y}}'_K\cdot\widetilde{\mathbf{y}}_K+\mathbf{B}'_0\mathbf{V}^{-1}\mathbf{B}_0 \\ & -(\widetilde{\mathbf{X}}'\mathbf{y}+\mathbf{V}^{-1}\mathbf{B}_0+\widetilde{\mathbf{W}}'_K\cdot\widetilde{\mathbf{y}}_K)'\cdot(\mathbf{V}^{-1}+\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}+\widetilde{\mathbf{W}}'_K\widetilde{\mathbf{W}}_K)^{-1}\cdot(\widetilde{\mathbf{X}}'\mathbf{y}+\mathbf{V}^{-1}\mathbf{B}_0+\widetilde{\mathbf{W}}'_K\cdot\widetilde{\mathbf{y}}_K)\end{aligned} \tag{6.92}$$

and use some notation to simplify the subsequent algebra

$$\begin{aligned}&\mathbf{S}-\mathbf{F}'\cdot\mathbf{R}^{-1}\cdot\mathbf{F} \\ &=\overbrace{\mathbf{B}'_0\mathbf{V}^{-1}\mathbf{B}_0}^{\equiv\mathbf{S}_1}+\overbrace{\mathbf{y}'\mathbf{y}}^{\equiv\mathbf{S}_2}+\overbrace{\widetilde{\mathbf{y}}'_K\cdot\widetilde{\mathbf{y}}_K}^{\equiv\mathbf{S}_3} \\ &\qquad-(\overbrace{\mathbf{V}^{-1}\mathbf{B}_0+\widetilde{\mathbf{X}}'\mathbf{y}}^{\equiv\mathbf{F}_1}+\overbrace{\widetilde{\mathbf{W}}'_K\cdot\widetilde{\mathbf{y}}_K}^{\equiv\mathbf{F}_2})'\cdot(\mathbf{V}^{-1}+\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}+\widetilde{\mathbf{W}}'_K\widetilde{\mathbf{W}}_K)^{-1}\cdot(\mathbf{F}_1+\mathbf{F}_2) \\ &=\widetilde{\mathbf{y}}'_K\cdot\widetilde{\mathbf{y}}_K-\left[\widetilde{\mathbf{W}}'_K\cdot\widetilde{\mathbf{y}}_K+\mathbf{F}_1\right]'\cdot\mathbf{R}^{-1}\cdot\left[\widetilde{\mathbf{W}}'_K\cdot\widetilde{\mathbf{y}}_K+\mathbf{F}_1\right]+\mathbf{S}_1+\mathbf{S}_2 \\ &=\widetilde{\mathbf{y}}'_K\cdot\underbrace{[\mathbf{I}_K-\widetilde{\mathbf{W}}_K\mathbf{R}^{-1}\widetilde{\mathbf{W}}'_K]}_{\equiv\underset{K\times K}{\mathbf{H}}}\cdot\widetilde{\mathbf{y}}_K-2\cdot\widetilde{\mathbf{y}}'_K\cdot\underbrace{\widetilde{\mathbf{W}}_K\mathbf{R}^{-1}\mathbf{F}_1}_{\equiv\underset{K\times 1}{\mathbf{D}}}+\underbrace{\mathbf{S}_1+\mathbf{S}_2-\mathbf{F}_1\mathbf{R}^{-1}\mathbf{F}_1}_{\equiv\text{ const}}\end{aligned} \tag{6.93}$$

Next complete the square with respect to y_{T+K}, i.e. the K-th element in the vector $\widetilde{\mathbf{y}}_K$ (see (6.65)). Further partition the matrix $\mathbf{H}$ in (6.93) as follows:

$$\mathbf{H} \;=\; \left(\begin{array}{c|c} \underset{K-1\times K-1}{\mathbf{H_{11}}} & \underset{K-1\times 1}{\mathbf{H_{12}}} \\ \hline \underset{1\times K-1}{\mathbf{H_{21}}} & \underset{1\times 1}{\mathbf{H_{22}}} \end{array}\right) \tag{6.94}$$

and also the vector $\mathbf{D}$, defined in the last line of (6.93):

$$\mathbf{D}=\left(\begin{array}{c}\mathbf{D}_{K-1}\\ D_{(K)}\end{array}\right)=\left(\begin{array}{c}\widetilde{\mathbf{W}}_{K-1}\cdot\mathbf{R}^{-1}\cdot\mathbf{F}_1\\ W_{(K)}\cdot\mathbf{R}^{-1}\cdot\mathbf{F}_1\end{array}\right) \tag{6.95}$$

with $\mathbf{D}_{K-1}$ a vector of dimension $K-1$ and $D_{(K)}$ the K-th element of $\mathbf{D}$.

Using these expressions the last equation in (6.93) can be written more compactly as

$$= (\,\widetilde{\mathbf{y}}'_{K-1}\,,\,y_{T+K}\,)\cdot\begin{pmatrix}\mathbf{H_{11}} & \mathbf{H_{12}}\\ \mathbf{H_{21}} & \mathbf{H_{22}}\end{pmatrix}\cdot\begin{pmatrix}\widetilde{\mathbf{y}}_{K-1}\\ y_{T+K}\end{pmatrix} - 2\cdot(\,\widetilde{\mathbf{y}}'_{K-1}\,,\,y_{T+K}\,)\cdot\begin{pmatrix}\mathbf{D}_{K-1}\\ D_{(K)}\end{pmatrix} + \text{ const} \tag{6.96}$$

Multiplying out (6.96) and completing the square with respect to y_{T+K} yields

$$\begin{aligned}
&\mathbf{S}-\mathbf{F}'\cdot\mathbf{R}^{-1}\cdot\mathbf{F}\\
&= \mathbf{H}_{22}\cdot\overbrace{\left[y_{T+K}-\mathbf{H}_{22}^{-1}\cdot(D_{(K)}-\mathbf{H}_{21}\cdot\widetilde{\mathbf{y}}_{K-1})\right]^2}^{\equiv QF(y_{T+K})} - \mathbf{H}_{22}^{-1}\cdot\left[D_{(K)}-\mathbf{H}_{21}\cdot\widetilde{\mathbf{y}}_{K-1}\right]^2\\
&\quad + \underbrace{\widetilde{\mathbf{y}}'_{K-1}\mathbf{H}_{11}\widetilde{\mathbf{y}}_{K-1}-2\cdot\widetilde{\mathbf{y}}'_{K-1}\mathbf{D}_{K-1}+\text{ const}}_{\equiv\,\mathbf{q}}\\
&= \mathbf{H}_{22}\cdot QF(y_{T+K}) - \mathbf{H}_{22}^{-1}\cdot\left[D_{(K)}-\mathbf{H}_{21}\cdot\widetilde{\mathbf{y}}_{K-1}\right]^2+\mathbf{q}
\end{aligned} \tag{6.97}$$

To get explicit expressions for the mean and variance of y_{T+K} recall from (6.93) that the matrix $\mathbf{H}$ is given by

$$\begin{aligned}
\mathbf{H} &= \mathbf{I}_K - \widetilde{\mathbf{W}}_K\mathbf{R}^{-1}\widetilde{\mathbf{W}}'_K\\
&= \mathbf{I}_K - \begin{pmatrix}
\overbrace{\widetilde{\mathbf{W}}_{K-1}\mathbf{R}^{-1}\widetilde{\mathbf{W}}'_{K-1}}^{\equiv\mathbf{V}_{11}} & | & \overbrace{\widetilde{\mathbf{W}}_{K-1}\mathbf{R}^{-1}W'_{(K)}}^{\equiv\mathbf{V}_{12}}\\
---------- & & ----------\\
\underbrace{W_{(K)}\mathbf{R}^{-1}\widetilde{\mathbf{W}}'_{K-1}}_{\equiv\mathbf{V}_{21}} & | & \underbrace{W_{(K)}\mathbf{R}^{-1}W'_{(K)}}_{\equiv\mathbf{V}_{22}}
\end{pmatrix}\\
&= \begin{pmatrix}
\mathbf{I}_{K-1}-\mathbf{V}_{11} & | & -\mathbf{V}_{12}\\
----- & & -----\\
-\mathbf{V}_{21} & | & 1-\mathbf{V}_{22}
\end{pmatrix}
\end{aligned} \tag{6.98}$$

From $QF(y_{T+K})$ in (6.97) the conditional posterior expectation of y_{T+K} can be seen to

equal

$$E(y_{T+K}|\,\widetilde{\mathbf{W}}_K,\,\mathbf{y}) \;=\; \mathbf{H}_{22}^{-1}\cdot(D_{(K)}-\mathbf{H}_{21}\cdot\widetilde{\mathbf{y}}_{K-1}) \tag{6.99}$$

or by inserting the expressions of the second equation of (6.98):

$$=\left(1-W_{(K)}\cdot\mathbf{R}^{-1}\cdot W'_{(K)}\right)^{-1}\cdot\left(D_{(K)}+W_{(K)}\cdot\mathbf{R}^{-1}\cdot\widetilde{\mathbf{W}}'_{K-1}\cdot\widetilde{\mathbf{y}}_{K-1}\right) \tag{6.100}$$

From the definition of $D_{(K)}$ in (6.95), the conditional expectation in (6.100) can be written as

$$\begin{aligned}&E(y_{T+K}|\,\widetilde{\mathbf{W}}_K,\,\mathbf{y})\\&=\left(1-W_{(K)}\cdot\mathbf{R}^{-1}\cdot W'_{(K)}\right)^{-1}\cdot W_{(K)}\cdot\mathbf{R}^{-1}\cdot\left(\mathbf{F}_1+\widetilde{\mathbf{W}}'_{K-1}\cdot\widetilde{\mathbf{y}}_{K-1}\right)\end{aligned} \tag{6.101}$$

verifying (6.71) of proposition 1, see BL, p.1310 (26), for a similar expression.

To derive the conditional posterior variance of y_{T+K}, stated in (6.72), first write out the last equation of (6.97), utilizing

$$\mathbf{q}=\widetilde{\mathbf{y}}'_{K-1}\cdot\mathbf{H}_{11}\cdot\widetilde{\mathbf{y}}_{K-1}-2\cdot\widetilde{\mathbf{y}}'_{K-1}\mathbf{D}_{K-1}+\mathbf{S}_1+\mathbf{S}_2-\mathbf{F}'_1\mathbf{R}^{-1}\mathbf{F}_1$$

together with the definitions of (6.98) to yield

$$\begin{aligned}\mathbf{S}-\mathbf{F}'\cdot\mathbf{R}^{-1}\cdot\mathbf{F} \;&=\; \mathbf{H}_{22}\cdot QF(y_{T+K})\\&-\mathbf{H}_{22}^{-1}\cdot\left(D_{(K)}+W_{(K)}\mathbf{R}^{-1}\widetilde{\mathbf{W}}'_{K-1}\cdot\widetilde{\mathbf{y}}_{K-1}\right)^2+\widetilde{\mathbf{y}}'_{K-1}\cdot(\mathbf{I}_{K-1}-\mathbf{V}_{11})\cdot\widetilde{\mathbf{y}}_{K-1}\\&-2\cdot\widetilde{\mathbf{y}}'_{K-1}\mathbf{D}_{K-1}+\mathbf{S}_1+\mathbf{S}_2-\mathbf{F}'_1\mathbf{R}^{-1}\mathbf{F}_1\end{aligned} \tag{6.102}$$

For an ease of comparison with the results in BL define:

$$E_K \;\equiv\; D_{(K)}+W_{(K)}\cdot\mathbf{R}^{-1}\cdot\widetilde{\mathbf{W}}'_{K-1}\cdot\widetilde{\mathbf{y}}_{K-1} \tag{6.103}$$

$$=\; W_{(K)}\cdot\mathbf{R}^{-1}\cdot\left(\mathbf{F}_1+\widetilde{\mathbf{W}}'_{K-1}\cdot\widetilde{\mathbf{y}}_{K-1}\right) \tag{6.104}$$

where the last line follows from the definition of $D_{(K)}$, see (6.95).

Then (6.102) can be written as

$$\begin{aligned}\mathbf{S}-\mathbf{F}'\cdot\mathbf{R}^{-1}\cdot\mathbf{F} \;=\; & \mathbf{H}_{22}\cdot QF(y_{T+K})-E_K'\cdot\mathbf{H}_{22}^{-1}\cdot E_K+\mathbf{S}-y_{T+K}^2 \\ & -\mathbf{F}_1'\mathbf{R}^{-1}\mathbf{F}_1-\widetilde{\mathbf{y}}_{K-1}'\cdot\mathbf{V}_{11}\cdot\widetilde{\mathbf{y}}_{K-1}-2\cdot\widetilde{\mathbf{y}}_{K-1}'\cdot\mathbf{D}_{K-1}\end{aligned} \tag{6.105}$$

by utilizing the fact that $\mathbf{S}_1+\mathbf{S}_2=\mathbf{S}-\mathbf{S}_3=\mathbf{S}-\widetilde{\mathbf{y}}_{K-1}\cdot\widetilde{\mathbf{y}}_{K-1}-y_{T+K}^2$, see equation (6.93).

Further by using the definitions of $\mathbf{D}_{K-1}$ and $\mathbf{V}_{11}$, given in (6.95) and (6.98), equation (6.105) becomes

$$\begin{aligned}&=\mathbf{H}_{22}\cdot QF(y_{T+K})-E_K'\cdot\mathbf{H}_{22}^{-1}\cdot E_K+\mathbf{S}-y_{T+K}^2-[\,\underbrace{\widetilde{\mathbf{y}}_{K-1}'\cdot\widetilde{\mathbf{W}}_{K-1}}_{\equiv C'}\cdot\mathbf{R}^{-1}\cdot\widetilde{\mathbf{W}}_{K-1}'\cdot\widetilde{\mathbf{y}}_{K-1} \\ &\quad+2\cdot\widetilde{\mathbf{y}}_{K-1}'\cdot\widetilde{\mathbf{W}}_{K-1}\cdot\mathbf{R}^{-1}\mathbf{F}_1+\mathbf{F}_1'\mathbf{R}^{-1}\mathbf{F}_1\,] \\ &=\mathbf{H}_{22}\cdot QF(y_{T+K})-E_K'\cdot\mathbf{H}_{22}^{-1}\cdot E_K+\mathbf{S}-y_{T+K}^2-[\,\mathbf{C}'\mathbf{R}^{-1}\mathbf{C}+2\cdot\mathbf{C}'\cdot\mathbf{R}^{-1}\mathbf{F}_1+\mathbf{F}_1'\mathbf{R}^{-1}\mathbf{F}_1\,] \\ &=\mathbf{H}_{22}\cdot QF(y_{T+K})-E_K'\cdot\mathbf{H}_{22}^{-1}\cdot E_K+\underbrace{\mathbf{S}-y_{T+K}^2-(\,\mathbf{C}+\mathbf{F}_1)'\cdot\mathbf{R}^{-1}\cdot(\,\mathbf{C}+\mathbf{F}_1)}_{\equiv F_K}\end{aligned} \tag{6.106}$$

Note that since $\mathbf{H}_{22}$ corresponds to D_K in (6.73) of proposition 1, $\mathbf{H}_{22}=D_K$ is used in the following. By using the last equation of (6.106) the posterior predictive distribution in (6.82) can be written as:

$$\begin{aligned}f(\widetilde{\mathbf{y}}_K|\,\mathbf{y}) \;&=\; g_1(\widetilde{\mathbf{y}}_{K-1},\,\mathbf{y})\cdot g_2(\widetilde{\mathbf{y}}_K,\,\mathbf{y}) \\ &\propto\; |\mathbf{R}|^{-\frac{1}{2}}\cdot[\,F_K-E_K'\cdot D_K^{-1}\cdot E_K+D_K\cdot QF(y_{T+K})\,]^{-\frac{a^\star}{2}} \\ &=\; |\mathbf{R}|^{-\frac{1}{2}}\cdot\left(F_K-E_K'\cdot D_K^{-1}\cdot E_K\right)^{-\frac{a^\star}{2}}\cdot\left[\,1+\left(\frac{D_K}{F_K-E_K'\cdot D_K^{-1}\cdot E_K}\right)\cdot QF(y_{T+K})\,\right]^{-\frac{\nu^\star+1}{2}}\end{aligned} \tag{6.107}$$

with $\nu^\star\equiv T-p+a+K-1$ and $QF(y_{T+K})=\left(y_{T+K}-E(y_{T+K}|\,\widetilde{\mathbf{W}}_K,\,\mathbf{y})\right)^2$, where the expectation is given in (6.101).

First from (6.107) it can be observed that the last factor is the kernel of the proposed conditional Student-t density of y_{T+K} with $\nu^\star$ posterior degrees of freedom, given $\widetilde{\mathbf{y}}_{K-1} = (y_{T+1},...,y_{T+K-1})'$, future deterministic values and the data $\mathbf{y}$, contained in the matrix $\widetilde{\mathbf{W}}_K$. From the expressions in (6.73) of proposition 1 it can be seen that the first two factors in the last line of (6.107) do not depend on y_{T+K}, but only on $\widetilde{\mathbf{y}}_{K-1}$. With regard to the conditional Student-t density of y_{T+K}, for $K > 1$, this means that given $\widetilde{\mathbf{y}}_{K-1}$ the first two factors can be subsumed into the normalizing constant of $f(y_{T+K}|\, \widetilde{\mathbf{W}}_K,\, \mathbf{y})$. Hence, given $\widetilde{\mathbf{y}}_{K-1}$, for $K > 1$ the factor g_2 in (6.107) is equal to a conditional Student-t density with $\nu^\star$ posterior degrees of freedom. From (6.107) the variance of the conditional predictive pdf of y_{T+K} (see Bauwens et al. (1999), A.35, p.294) is given by:

$$Var(y_{T+K}|\, \widetilde{\mathbf{W}}_K,\, \mathbf{y}) = \frac{F_K - E_K' \cdot D_K^{-1} \cdot E_K}{(\nu^\star - 2) \cdot D_K} \tag{6.108}$$

verifying (6.72) of proposition 1.[11]

To obtain the conditional posterior predictive density of y_{T+K-1} given values $\widetilde{\mathbf{y}}_{K-2} = (y_{T+1},...,y_{T+K-2})'$ and $\mathbf{y}$, complete the square with respect to y_{T+K-1} in the second factor of (6.107), which then can be shown to have the form of a univariate conditional t-density, and so on.

The expressions of the first two moments for $K = 1$, proposed in (6.67) and (6.68) above, can be obtained from (6.101) and (6.108), respectively, by noting that for $K = 1$, $\widetilde{\mathbf{y}}_{K-1} = \mathbf{0}$ and $\widetilde{\mathbf{W}}_{K-1} = \mathbf{0}$, and thus $\widetilde{\mathbf{y}}_K = y_{T+1}$ and $\widetilde{\mathbf{W}}_K = W_{(1)}$, where $W_{(1)}$ only depends on the observed data $\mathbf{y}$. To obtain the corresponding expressions under the diffuse prior for σ^2, used in section 5.2.1, the hyperparameters of the IG_2 prior have to be chosen according to $a \to -d$ and $b \to 0$.

□

[11] Note that the precision of y_{T+K} is equal to $Var(y_{T+K}|\, \widetilde{\mathbf{W}}_K,\, \mathbf{y})^{-1}$ in (6.108), see BL, p.1311 (27), for a similar result.