

Mika Mahosenaho

# Bayesian switching model for forecasting GDP growth rates

Faculty of Information Technology and Communication Sciences (ITC)  
Master's thesis  
June 2020

# Abstract

Mika Mahosenaho: Bayesian switching model for forecasting GDP growth rates  
Master's thesis  
Tampere University  
Master's Degree Programme in Computational Big Data Analytics  
June 2020

---

The forecast properties of econometric models for gross domestic product (GDP) have been of great interest since Nelson (1972) criticized large-scale macroeconomic models. Various approaches have been proposed, yet forecasting over the irregular duration such as recessions remains a bottleneck. The well known Markov switching models take into account the different phases of business cycles, but they are generally unable to provide more accurate forecasts than the conventional autoregressive-moving-average (ARMAX) model. We present a new Bayesian multi-country model, with a random intercept process. The new model features nonstationarity occurring from stochastic breaks in the process and accommodates the multi-country variation with country-specific parameters. The Markov switching, ARMAX and the proposed Bayesian models are compared in an empirical forecast analysis for 13 countries. The proposed Bayesian model provided the most accurate forecasts in terms of RMSFE (root mean squared forecast errors). The developed Bayesian model offers a flexible foundation for further analysis of the model parameters and reliable prediction interval estimates.

**Keywords:** Time series, forecasting, hierarchical Bayes, Gibbs sampling, gross domestic product.

The originality of this thesis has been checked using the Turnitin Originality Check service.

# Tiivistelmä

Mika Mahosenaho: Bayesiläinen vaihtomalli bruttokansantuotteen kasvun ennustamiseksi

Pro gradu -tutkielma

Tampereen yliopisto

Master's Degree Programme in Computational Big Data Analytics

Kesäkuu 2020

---

Ekonomististen mallien ominaisuudet bruttokansantuotteen ennustamiseksi nousivat kiinnostuksen kohteeksi, kun Nelson (1972) asetti suuret ekonometriset makromallit kyseenalaiseksi. Vaikka lukuisia menetelmiä on ehdotettu, ennusteiden tarkkuus varsinkin taantumien aikana on osoittautunut heikoksi. Hyvin tunnetut Markovin vaihtomallit kykenevät ottamaan huomioon suhdannevaihteluiden eri vaiheet, mutta yleisesti ottaen sen ennusteet eivät ole olleet vertailukohteena olevan autoregressiivisen liukuvan keskiarvon (ARMAX) -mallin ennusteita tarkempia. Tässä tutkielmassa esitellään uusi bayesiläinen paneeliaineistoon sovitettava malli, jossa vakiotermi on satunnainen. Uuden mallin piirteitä ovat stokastisista rakennemuutoksista seuraava epästationaarisuus sekä muista maista saatavan tiedon hyödyntäminen maakohtaisten parametrien estimoinnissa. Markovin vaihtomalli, ARMAX-malli ja tutkielmassa kehitelty bayesiläinen malli asetetaan vertailuun ennustuskokeeseen, jonka aineisto koostuu kolmestatoista maasta. Kokeessa bayesiläinen malli tuotti kaikkein tarkimmat ennusteet keskineliövirheellä mitattuna. Malli tarjoaa joustavan ratkaisun ennustevälien laatimiseen ja mallin parametrien tarkempaan analyysiin.

**Keywords:** Aikasarja, ennustaminen, Bayes, Gibbsin näytteistys, bruttokansantuote.

Tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck -ohjelmalla.

## **Acknowledgements**

I would like to thank my supervisor Hyon-Jung Kim-Ollila for her support, guidance, and help through the process of my thesis. I would also like to thank Pekka Pere for his valuable advice and comments.

# Contents

1	Introduction . . . . .	1
2	Methodology . . . . .	4
2.1	ARMAX models . . . . .	4
2.1.1	Estimation of ARMAX model . . . . .	5
2.1.2	Model selection . . . . .	6
2.1.3	Forecasting with ARMAX model . . . . .	6
2.2	Markov switching model . . . . .	6
2.2.1	Estimation of the Markov switching model . . . . .	7
2.2.2	Forecasting with MS model . . . . .	8
2.3	Models in Bayesian framework . . . . .	9
2.3.1	Bayesian time series modeling . . . . .	9
2.3.2	Gibbs sampling . . . . .	10
2.3.3	BS model . . . . .	12
2.3.4	MuB model . . . . .	14
2.3.5	MuBS model . . . . .	15
2.4	Forecast evaluation . . . . .	17
3	Empirical forecast analysis . . . . .	18
3.1	Description of data . . . . .	18
3.1.1	GDP growth rate . . . . .	18
3.1.2	Composite leading indicator . . . . .	19
3.2	Model specification . . . . .	23
3.3	Results . . . . .	24
3.3.1	A case study: US GDP growth rate forecasts . . . . .	26
4	Conclusions . . . . .	32
	References . . . . .	33
	APPENDIX A. The full conditional distributions of BS model . . . . .	34
	APPENDIX B. The full conditional distributions of MuB model . . . . .	37
	APPENDIX C. The full conditional distributions of MuBS model . . . . .	41

# 1 Introduction

Since the influential Bretton Woods conference in 1944, gross domestic product (GDP) has internationally been the most important tool for measuring the size of a nation's economy. The size of the economy, and perhaps more importantly the rate at which it grows, have an impact on various institutions, like the government, companies and households. The anticipated growth rates of economies are closely followed decision making inputs across the world, but often the forecasted growth rates are far from the actual growth rates. For example, in 2008 the US economy sunk into a recession as a result of an unanticipated financial crisis. Not only the econometric models did not warn us about the crisis, but also the GDP forecasts were largely unreliable across 2008-2009, an era that has been called the "Great Recession" (Stekler and Talwar 2013).

The development of econometric models have been of great interest for five decades. The first notable model-based GDP forecasts were obtained from so called large-scale macroeconometric (LSM) models, built mainly in the work of Jan Tinbergen and Lawrence Klein. The results from these models were initially promising, but Nelson (1972) first presented evidence that the forecasts from the LSM models were in many respects worse than the forecasts from random walk model and simple autoregressive-moving-average (ARMA) model propagated by George Box and Gwilym Jenkins. In the 1980s the study of econometric models was divided in two branches: economic theory driven models and empirical models. The ground work for economic theory driven modeling was provided by Kydland and Prescott (1982), who took the first step in the development of dynamic stochastic general equilibrium (DSGE) models. Two years before the first steps of DSGE modeling, Sims (1980) introduced vector autoregressive (VAR) model using the empirical macroeconomic approach. In opposition to DSGE models, VAR models are driven by empirical evidence rather than economic theory.

Various studies have been made where multivariate VAR and DSGE models have been put to test against the simple univariate AR and random walk benchmarks. The multivariate models aim to describe the relationship among the macroeconomic variables like GDP, inflation and interest rate. For example Smets and Wouters (2007) argued that the forecasting ability of DSGE models might be slightly better to those of VAR and AR models. Wickens (2014) tested the forecasting ability of the DSGE model of Smets and Wouters over the turbulent period of 2008-2011. The conclusion was that the DSGE model forecast well only when the preceding forecasts are accurate, and no turning points occur in the economy during the forecasting period.

Beyond the branches focusing on DSGE and VAR approaches, some nonstationary models have also been proposed. As GDP growth rates appear to be subject of abrupt changes, Hamilton (1989) introduced a nonstationary model labelled as Markov switching autoregressive (MS) model. The model was designed to capture the different dynamics during expansions and recessions, and it sparked significant interest among econometricians. The forecasting ability of such models was assessed by Clements and Krolzig (1998). Based on Monte Carlo study and empirical study using data from the US, the authors conclude, that while the model was able to capture the business cycle dynamics in-sample, it did not necessarily produce superior forecasts compared to more simple methods. Clements and Krolzig add that under the two models' assumptions the conventional AR was in fact a fairly robust forecasting device.

In this thesis the empirical forecast analysis comparing MS and ARMA models is further extended to encompass 13 countries from the organisation for economic co-operation and development (OECD). Two variables are used in the analysis for all countries: GDP growth rate as the dependent variable and composite leading indicator (CLI), calculated by OECD, as the exogenous covariate. In addition to the covariance extended MS and ARMA (or ARMAX) models, a novel Bayesian model is proposed, of which three versions are investigated separately.

The first version of the proposed model is Bayesian switching ARMAX (BS) model. Similarly to MS, in BS it is assumed that the observed time series consists of stationary sections. Like Hamilton in his original paper, Clements and Krolzig allowed only the intercept term to switch between recession and expansion sections. Also in BS model the abrupt changes are only allowed to affect the intercept term of the model. The difference is that whereas MS assumes different constant intercepts for recession and expansion, BS assumes that the intercept is drawn from the normal distribution whenever a structural break is detected. The second version of the proposed model is multi-country Bayesian ARMAX (MuB) model, which utilizes the multi-country aspect of the data. The country-specific coefficients are drawn from common distributions, causing shrinkage between the corresponding coefficients of the country-specific submodels. Finally, the third version of the proposed model is multi-country Bayesian switching ARMAX (MuBS) model, which combines the random intercept and the multi-country features of BS and MuB models. The five models compared in the forecast analysis are described in detail in the following methodology Chapter 2.

The empirical forecast analysis is described in Chapter 3. First the data are introduced in detail in Section 3.1, followed by the determinations of the model specifications in Section 3.2. The results of the forecast analysis are discussed in Section 3.3. Averaging over the 13 countries, each version of the proposed Bayesian

model was more accurate than ARMAX and MS approaches in terms of mean absolute error and root mean squared error for all forecast horizons ranging from 1 to 4 steps. The evidence suggests that the multi-country feature of the model might be more useful than the random intercept feature, but on average the full model was the most accurate. Chapter 4 concludes.



## 2 Methodology

### 2.1 ARMAX models

Suppose that a time series

$$\mathbf{y}_T = \{y_1, y_2, \dots, y_T\}$$

of length  $T$  is observed, which is generated by a stochastic process  $Y_t$ . Let

$$\mathbf{x}_T = \{x_1, x_2, \dots, x_T\}$$

be an exogenous covariate series of length  $T$ . The covariance-stationary process needs to meet two requirements:

1.  $E(Y_t) = \mu_t = \mu$  for all  $t$
2.  $\gamma_{jt} = E(Y_t - \mu_t)(Y_{t-j} - \mu_t) = E(Y_t - \mu)(Y_{t-j} - \mu) = \gamma_j$  for all  $t$  and any  $j$ .

ARMAX processes are a class of covariance-stationary processes. They consists of three parts: the autoregressive (AR) part, the moving average (MA) part and the exogenous covariate (X) part. AR terms determine how much the past values of the time series account for the upcoming value, and MA terms determine the impact of the past error terms. The ARMAX model can be written as

$$Y_t = c + \sum_{i=1}^p \phi_i Y_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \sum_{i=1}^r \beta_i x_{t-i} + \varepsilon_t, \quad (2.1)$$

where the error term

$$\varepsilon_t \sim N(0, \sigma^2),$$

with

$$\varepsilon_t \text{ independent of } \varepsilon_\tau \text{ for all } t \neq \tau.$$

The intercept term  $c$ , the MA coefficients  $(\theta_1, \theta_2, \dots, \theta_q)'$  and the coefficients for the covariate  $(\beta_1, \beta_2, \dots, \beta_r)'$  can be any real numbers. The AR coefficients  $(\phi_1, \phi_2, \dots, \phi_p)'$  are restricted so that the roots of  $(1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p)$  lie outside the unit circle, which is the stationarity condition for the AR part. The vector of all unknown parameters is denoted by  $\boldsymbol{\alpha} = (c, \phi_1, \phi_2, \dots, \phi_p, \theta_1, \theta_2, \dots, \theta_q, \beta_1, \beta_2, \dots, \beta_r, \sigma)'$ , where  $\sigma$  is the standard deviation of the error term.

### 2.1.1 Estimation of ARMAX model

The parameter vector  $\boldsymbol{\alpha}$  can be estimated from the data  $\boldsymbol{\mathcal{X}}_T = (\mathbf{y}_T, \mathbf{x}_T)$  using the conditional maximum likelihood estimation (CMLE) method. With CMLE, the values of the parameter vector are estimated by maximizing the conditional likelihood of the data. The conditional likelihood of the data can be written as

$$\begin{aligned} \mathcal{L}(\boldsymbol{\alpha}^*) &= f_{Y_T, Y_{T-1}, \dots, Y_1}(y_T, y_{T-1}, \dots, y_1 | \boldsymbol{\mathcal{X}}_0; \boldsymbol{\alpha}^*) \\ &= \prod_{t=1}^T f_{Y_t}(y_t | \boldsymbol{\mathcal{X}}_{t-1}; \boldsymbol{\alpha}^*) \\ &= \prod_{t=1}^T \frac{1}{\sqrt{2\pi}\sigma^*} \exp \left[ \frac{-(y_t - E(Y_t | \boldsymbol{\mathcal{X}}_{t-1}; \boldsymbol{\alpha}^*))^2}{2(\sigma^*)^2} \right] \\ &= (2\pi)^{-\frac{T}{2}} (\sigma^*)^{-T} \exp \left[ \frac{-\sum_{t=1}^T (y_t - E(Y_t | \boldsymbol{\mathcal{X}}_{t-1}; \boldsymbol{\alpha}^*))^2}{2(\sigma^*)^2} \right], \end{aligned}$$

where  $\boldsymbol{\alpha}^*$  is a candidate parameter vector, and  $\boldsymbol{\mathcal{X}}_0$  holds the chosen initial values for the first observations. Due to computational stability, it is better to calculate conditional log-likelihood instead of conditional likelihood. The conditional log-likelihood is

$$\log \mathcal{L}(\boldsymbol{\alpha}^*) = -\frac{T}{2} \log(2\pi) - T \log \sigma^* - \frac{\sum_{t=1}^T (y_t - E(Y_t | \boldsymbol{\mathcal{X}}_{t-1}; \boldsymbol{\alpha}^*))^2}{2(\sigma^*)^2}. \quad (2.2)$$

The log-likelihood 2.2 is maximized using Broyden–Fletcher–Goldfarb–Shanno (BFGS) optimization algorithm. Generally BFGS algorithm finds a parameter vector  $\mathbf{x} \in \mathcal{R}^n$  that minimizes a function  $f(\mathbf{x})$ . Because the log-likelihood is maximized instead of minimized, the function is set to  $f(\mathbf{x}) = -\log \mathcal{L}(\mathbf{x})$ , where  $\mathbf{x} = \boldsymbol{\alpha}^*$ .

BFGS belongs to the class of quasi-Newton optimization methods, and it can be applied to all non-convex functions with continuous second derivatives. Quasi-Newton methods use the approximated ( $n \times n$ ) Hessian matrix  $\mathbf{H}$ , which includes the approximated second partial derivatives. The Hessian matrix is approximated to gain speed and reliability for the procedure, and it is updated during each  $k$ th iteration of the algorithm. Given the initial guesses  $\boldsymbol{\alpha}_0$  and  $\mathbf{H}_0$ , the BFGS procedure is as follows:

1. Obtain the direction  $\mathbf{p}_k$  for the update by solving  $\mathbf{H}_k \mathbf{p}_k = -\Delta f(\boldsymbol{\alpha}^*)$
2. Find a scalar  $\delta_k > 0$  that minimizes  $f(\boldsymbol{\alpha}^*_k + \delta_k \mathbf{p}_k)$  (line search)
3. Update the parameter vector  $\boldsymbol{\alpha}^*_{k+1} = \boldsymbol{\alpha}^*_k + \mathbf{s}_k$ , where  $\mathbf{s}_k = \delta_k \mathbf{p}_k$
4. Update the approximated Hessian  $\mathbf{H}_{k+1} = \mathbf{H}_k + \frac{\mathbf{y}_k \mathbf{y}_k^\top}{\mathbf{y}_k^\top \mathbf{s}_k} - \frac{\mathbf{H}_k \mathbf{s}_k \mathbf{s}_k^\top \mathbf{H}_k^\top}{\mathbf{s}_k^\top \mathbf{H}_k \mathbf{s}_k}$ , where

$$\mathbf{y}_k = \Delta f(\boldsymbol{\alpha}^*_{k+1}) - \Delta f(\boldsymbol{\alpha}^*_k).$$

These steps are repeated until convergence, which results in  $\hat{\boldsymbol{\alpha}} = \boldsymbol{\alpha}^*_k$ . It is important to note that the parameters are not restricted, which means that the solution may not be stationary as is intended.

### 2.1.2 Model selection

Akaike's information criterion (AIC) is used to take into account the goodness-of-fit of the model and the model complexity, penalizing the model of the high number of parameters. AIC is defined as follows:

$$AIC(\hat{\boldsymbol{\alpha}}) = -2 \log \mathcal{L}(\hat{\boldsymbol{\alpha}}) + 2K,$$

where  $\log \mathcal{L}(\hat{\boldsymbol{\alpha}})$  is defined as in 2.2, and  $K$  is the number of model parameters. Smaller AIC indicates better fit.

### 2.1.3 Forecasting with ARMAX model

The  $h$ -step point forecast from ARMAX model 2.1 is defined as

$$\hat{Y}_{T+h} = \hat{c} + \sum_{i=1}^p \hat{\phi}_i \hat{Y}_{T+h-i} + \sum_{i=1}^q \hat{\theta}_i \hat{\varepsilon}_{T+h-i} + \sum_{i=1}^r \hat{\beta}_i \hat{x}_{T+h-i},$$

for  $h = 1, 2, \dots$ , where parameters  $\hat{c}, \hat{\phi}_1, \dots, \hat{\phi}_p, \hat{\theta}_1, \dots, \hat{\theta}_q, \hat{\beta}_1, \dots, \hat{\beta}_r$  are obtained after CMLE estimation. Furthermore,  $\hat{Y}_{T+h-i}$  is a known value  $Y_{T+h-i}$  for  $i \geq h$ , and a forecasted value otherwise. The value for the covariate part  $\hat{x}_{T+h-i}$  is a known value  $x_{T+h-i}$  for  $i \geq h$ , but  $\hat{x}_{T+h-i} = x_T$  otherwise. Finally,  $\hat{\varepsilon}_{T+h-i} = Y_{T+h-i} - \hat{Y}_{T+h-i}$  for  $i \geq h$  and 0 otherwise.

## 2.2 Markov switching model

Markov switching model is a nonstationary time series model. The model coefficients switch between a number of regimes according to a latent Markov chain  $S_t$ , which has unknown but constant transition probabilities. The coefficients are constant for each regime, so the model is locally stationary. Following Hamilton (1989), for MS model the number of possible states of the Markov chain  $S_t$  is restricted to two ( $S_t = j$ , where  $j = 0, 1$ ).

In the original paper of Hamilton, the regime shift is only allowed to affect the intercept term. The model has intercept terms  $c_0$  and  $c_1$  for the states 0 and 1. We follow these choices, but also add an exogenous covariate to the model. The model

can be written as

$$Y_t = c_{S_t} + \sum_{i=1}^p \phi_i Y_{t-i} + \sum_{i=1}^r \beta_i x_{t-i} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma^2), \quad (2.3)$$

where the transition probabilities of  $S_t$  are

$$\begin{aligned} P(S_t = 1 | S_{t-1} = 0) &= p_{01} \\ P(S_t = 1 | S_{t-1} = 1) &= p_{11} \\ P(S_t = 0 | S_{t-1} = 0) &= p_{00} = 1 - p_{01} \\ P(S_t = 0 | S_{t-1} = 1) &= p_{10} = 1 - p_{11}. \end{aligned}$$

The parameter vector of the model 2.3 is denoted

$$\boldsymbol{\alpha} = (p_{01}, p_{11}, c_0, c_1, \phi_1, \phi_2, \dots, \phi_p, \beta_1, \beta_2, \dots, \beta_r, \sigma)'$$

### 2.2.1 Estimation of the Markov switching model

The estimation procedure of MS model 2.3 is a combination CMLE (see 2.1.1) and a specific recursive algorithm for deducting the state probabilities of each time point  $t$ . The likelihood of a single observation

$$f_{Y_t}(y_t | S_t = j, \boldsymbol{\mathcal{X}}_{t-1}; \boldsymbol{\alpha}),$$

depends on the random variable  $S_t$  in addition to parameter vector  $\boldsymbol{\alpha}$  and observations  $\boldsymbol{\mathcal{X}}_{t-1}$ . The density conditional on  $S_t = 0$  is

$$f_{Y_t}(y_t | S_t = 0, \boldsymbol{\mathcal{X}}_{t-1}; \boldsymbol{\alpha}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ \frac{-(y_t - c_0 - \sum_{i=1}^p \phi_i Y_{t-i} - \sum_{i=1}^r \beta_i x_{t-i})^2}{2\sigma^2} \right\} \quad (2.4)$$

and the density conditional on  $S_t = 1$  is

$$f_{Y_t}(y_t | S_t = 1, \boldsymbol{\mathcal{X}}_{t-1}; \boldsymbol{\alpha}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ \frac{-(y_t - c_1 - \sum_{i=1}^p \phi_i Y_{t-i} - \sum_{i=1}^r \beta_i x_{t-i})^2}{2\sigma^2} \right\} \quad (2.5)$$

The conditional log-likelihood of the data then becomes:

$$\log \mathcal{L}(\boldsymbol{\alpha}) = \sum_{t=1}^T \log f_{Y_t}(y_t | \boldsymbol{\mathcal{X}}_{t-1}; \boldsymbol{\alpha}),$$

where the density of a single observation is the sum of densities 2.4 and 2.5, multiplied by probabilities  $P(S_t = 0|\boldsymbol{\mathcal{X}}_{t-1}; \boldsymbol{\alpha})$  and  $P(S_t = 1|\boldsymbol{\mathcal{X}}_{t-1}; \boldsymbol{\alpha})$  respectively:

$$f_{Y_t}(y_t|\boldsymbol{\mathcal{X}}_{t-1}; \boldsymbol{\alpha}) = \sum_{j=0}^1 P(S_t = j|\boldsymbol{\mathcal{X}}_{t-1}; \boldsymbol{\alpha}) f_{Y_t}(y_t|S_t = j, \boldsymbol{\mathcal{X}}_{t-1}; \boldsymbol{\alpha}).$$

The probabilities  $P(S_t = j|\boldsymbol{\mathcal{X}}_{t-1}; \boldsymbol{\alpha})$  are estimated from the data using the recursive filter suggested by Hamilton. Each iteration of the filter consists of two steps. During the first step, the probabilities of the two states at  $t$  are calculated with respect to  $\boldsymbol{\alpha}$  and the data up to  $t$ :

$$P(S_t = 0|\boldsymbol{\mathcal{X}}_t; \boldsymbol{\alpha}) = \frac{P(S_t = 0|\boldsymbol{\mathcal{X}}_{t-1}; \boldsymbol{\alpha}) f_{Y_t}(y_t|S_t = 0, \boldsymbol{\mathcal{X}}_{t-1}; \boldsymbol{\alpha})}{\sum_{j=0}^1 P(S_t = j|\boldsymbol{\mathcal{X}}_{t-1}; \boldsymbol{\alpha}) f_{Y_t}(y_t|S_t = j, \boldsymbol{\mathcal{X}}_{t-1}; \boldsymbol{\alpha})}$$

$$P(S_t = 1|\boldsymbol{\mathcal{X}}_t; \boldsymbol{\alpha}) = \frac{P(S_t = 1|\boldsymbol{\mathcal{X}}_{t-1}; \boldsymbol{\alpha}) f_{Y_t}(y_t|S_t = 1, \boldsymbol{\mathcal{X}}_{t-1}; \boldsymbol{\alpha})}{\sum_{j=0}^1 P(S_t = j|\boldsymbol{\mathcal{X}}_{t-1}; \boldsymbol{\alpha}) f_{Y_t}(y_t|S_t = j, \boldsymbol{\mathcal{X}}_{t-1}; \boldsymbol{\alpha})}.$$

This step requires the predicted probabilities of the states from the previous iteration. During the next step the predicted probabilities are calculated for the following iteration:

$$P(S_{t+1} = 0|\boldsymbol{\mathcal{X}}_t; \boldsymbol{\alpha}) = p_{00}P(S_t = 0|\boldsymbol{\mathcal{X}}_t; \boldsymbol{\alpha}) + p_{10}P(S_t = 1|\boldsymbol{\mathcal{X}}_t; \boldsymbol{\alpha}) \quad (2.6)$$

$$P(S_{t+1} = 1|\boldsymbol{\mathcal{X}}_t; \boldsymbol{\alpha}) = p_{01}P(S_t = 0|\boldsymbol{\mathcal{X}}_t; \boldsymbol{\alpha}) + p_{11}P(S_t = 1|\boldsymbol{\mathcal{X}}_t; \boldsymbol{\alpha}). \quad (2.7)$$

In the beginning of the iterative algorithm, the predicted probabilities for the states need to be initiated. Hamilton proposes  $P(S_1 = 0|\boldsymbol{\mathcal{X}}_0; \boldsymbol{\alpha}) = \frac{1}{k}$  and  $P(S_1 = 1|\boldsymbol{\mathcal{X}}_0; \boldsymbol{\alpha}) = \frac{1}{k}$  for the initiative step, where  $k = 2$  is equal to the number of the possible states in the Markov chain.

## 2.2.2 Forecasting with MS model

When the state of the Markov chain for  $T + h$  is known, the  $h$ -step ahead forecast

$$\hat{Y}_{T+h|T} = E(Y_{T+h}|S_{T+h} = j, \boldsymbol{\mathcal{X}}_T, \hat{\boldsymbol{\alpha}}) = \hat{c}_j + \sum_{i=1}^p \hat{\phi}_i \hat{Y}_{T+h-i} + \sum_{i=1}^r \hat{\beta}_i \hat{x}_{T+h-i},$$

is derived iteratively, where  $\hat{Y}_{T+h-i}$  is a known value  $Y_{T+h-i}$  for  $i \geq h$ , and a forecasted value otherwise. The value for the covariate part  $\hat{x}_{T+h-i}$  is a known value  $x_{T+h-i}$  for  $i \geq h$ , but  $\hat{x}_{T+h-i} = x_T$  otherwise. Since the state of the Markov chain

for the next period is not known, the forecast becomes a mixture of two forecast for the two states, multiplied with their probabilities

$$\begin{aligned}\hat{Y}_{T+h|T} &= E(Y_{T+h}|\mathbf{x}_T, \hat{\boldsymbol{\alpha}}) \\ &= P(S_{T+h} = 0|\mathbf{x}_T; \hat{\boldsymbol{\alpha}}) \left( \hat{c}_0 + \sum_{i=1}^p \hat{\phi}_i \hat{Y}_{T+h-i} + \sum_{i=1}^r \hat{\beta}_i x_{T+h-i} \right) \\ &\quad + P(S_{T+h} = 1|\mathbf{x}_T; \hat{\boldsymbol{\alpha}}) \left( \hat{c}_1 + \sum_{i=1}^p \hat{\phi}_i \hat{Y}_{T+h-i} + \sum_{i=1}^r \hat{\beta}_i x_{T+h-i} \right),\end{aligned}$$

where  $P(S_{T+h} = 0|\mathbf{x}_T; \hat{\boldsymbol{\alpha}})$  and  $P(S_{T+h} = 1|\mathbf{x}_T; \hat{\boldsymbol{\alpha}})$  are obtained from the equations 2.6 and 2.7 of the algorithm.

## 2.3 Models in Bayesian framework

Three versions of the proposed Bayesian model are constructed in this section. The first version is Bayesian switching ARMAX (BS) model, which has an intercept term that shifts according to a Gaussian random variable whenever a stochastic break occurs. The second version is multi-country Bayesian ARMAX (MuB) model, which utilizes multi-country aspect of the data, as the country-specific coefficients are drawn from common distributions. The full version is multi-country Bayesian switching ARMAX (MuBS) model, which combines the features of the two prior versions. The models are conditionally conjugate hierarchical models, so for each model a Gibbs sampler is developed for the purpose of posterior analysis.

### 2.3.1 Bayesian time series modeling

Classical statistics views the data as random, and the model parameters as unknown deterministic constants. In Bayesian statistics, however, the unknown model parameters are random variables. In a single-parameter model, the model parameter  $\theta$  is assumed to follow a distribution based on some previous knowledge, which is called the prior distribution. The distribution of model parameters conditional on the data is called the posterior distribution. The density of the posterior distribution can be obtained using the Bayes' rule:

$$p(\theta|\mathbf{y}_T) = \frac{p(\theta, \mathbf{y}_T)}{p(\mathbf{y}_T)} = \frac{p(\theta)p(\mathbf{y}_T|\theta)}{p(\mathbf{y}_T)} = \frac{p(\theta)p(\mathbf{y}_T|\theta)}{\int p(\theta)p(\mathbf{y}_T|\theta)d\theta},$$

where  $p(\theta)$  is the prior density,  $p(\mathbf{y}_T|\theta)$  the likelihood and  $p(\mathbf{y}_T)$  is the marginal density of  $\mathbf{y}_T$ . Assuming conditional independence, the likelihood of the data  $p(\mathbf{y}_T|\theta)$  is a joint density  $p(y_1, \dots, y_T|\theta)$ , which due to conditional exchangeability can be

represented as  $\prod_{t=1}^T p(y_t|\theta, \mathbf{y}_{t-1})$ . Calculating  $p(\mathbf{y}_T)$  is not necessary, but evaluating

$$p(\theta|\mathbf{y}_T) \propto p(\theta)p(\mathbf{y}_T|\theta)$$

is sufficient.

Forecasting future time series observations in Bayesian framework can be based on the posterior predictive distribution

$$p(y_{T+1}|\mathbf{y}_T) = \int p(y_{T+1}|\theta, \mathbf{y}_T)p(\theta|\mathbf{y}_T)d\theta.$$

Because of dependence between the observations, predictions for further forecast horizons are conditioned on the previous forecasts, so that  $h$ -step forecast is

$$\begin{aligned} p(y_{T+h}|\mathbf{y}_T) &= \int p(y_{T+h}|\theta, y_{T+h-1}, \dots, y_{T+1}, \mathbf{y}_T) \times \\ &\quad p(y_{T+h-1}|\theta, y_{T+h-2}, \dots, y_{T+1}, \mathbf{y}_T) \cdots \times \\ &\quad p(y_{T+1}|\theta, \mathbf{y}_T)p(\theta|\mathbf{y}_T)dy_{T+h-1} \cdots dy_{T+1}d\theta. \end{aligned}$$

### 2.3.2 Gibbs sampling

A Gibbs sampler is a Markov chain Monte Carlo (MCMC) algorithm for sampling from the posterior distribution. In Gibbs sampling, full conditional distributions of the model parameters are determined, from which the samples of unknown parameters are sequentially drawn from. The sequence of draws forms a Markov chain, which ultimately converges towards the true posterior distribution. The sample draws from the converged distributions can be used to summarize the properties of the posterior distribution. Furthermore, posterior predictive simulations can be obtained by conditioning the likelihood of new observations on the posterior parameter draws.

Gibbs sampling is particularly useful for hierarchical models, which are structured in stages. In hierarchical models, the model parameters of the higher stage depend on hyperparameters, which are also random variables generated in the lower stage. The distributions of hyperparameters are called hyperpriors.

Let  $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_P)$  be the parameter vector for all stages of a conditionally conjugate hierarchical model. The joint posterior density of the parameter vector  $\boldsymbol{\theta}$  is

$$p(\boldsymbol{\theta}|\mathbf{y}_T) = \frac{p(\boldsymbol{\theta})p(\mathbf{y}_T|\boldsymbol{\theta})}{\int p(\boldsymbol{\theta})p(\mathbf{y}_T|\boldsymbol{\theta})d\boldsymbol{\theta}} \propto p(\boldsymbol{\theta})p(\mathbf{y}_T|\boldsymbol{\theta}),$$

where  $p(\boldsymbol{\theta})$  is the joint prior distribution. Due to conjugacy, the full conditional

distributions  $P(\theta_i|\boldsymbol{\theta}_{-i}, \mathbf{y}_T)$  are easily derived from the joint posterior density, and the parameter samples can be directly drawn from those distributions. Before drawing samples from the full conditional distributions, the model parameters are initiated with any values on the space of their possible values.

The period preceding the convergence of the Markov chain is called the burn-in period. The sample draws after the burn-in period might be slightly correlated, which is why those sample draws are often thinned, i.e. only every  $d$ th sample draw is kept. The sample draws from the burn-in period are discarded together with the thinned sample draws. An important choice is to decide the length of the burn-in period. There are a few techniques to assessing the convergence. A good practice is to run separate chains with various initial values, and compare those to each other.

The Gibbs sampling algorithm is outlined as follows:

1. Initialize the algorithm

- 1) Set the initial values  $\boldsymbol{\theta}^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_p^{(0)})$ .
- 2) Find the burn-in length  $n_{burn}$ , which seems to achieve convergence of the Markov's chain.
- 3) Choose  $M$ , which is the desired final sample size.
- 4) If necessary, choose a suitable thinning factor  $d$  to reduce the autocorrelation of the Markov's chain.

2. Set  $n = n_{burn} + d * M$ .

3. Set  $t = 1$ .

4. If  $t = n$  go to step 6, otherwise draw parameter values from their full conditional distributions:

- 1)  $\theta_1^{(t)}$  from  $p(\theta_1|\theta_2^{(t-1)}, \theta_3^{(t-1)}, \dots, \theta_p^{(t-1)}; \mathbf{y}_T)$
- 2)  $\theta_2^{(t)}$  from  $p(\theta_2|\theta_1^{(t)}, \theta_3^{(t-1)}, \dots, \theta_p^{(t-1)}; \mathbf{y}_T)$
- ...
- p)  $\theta_p^{(t)}$  from  $p(\theta_p|\theta_1^{(t)}, \theta_2^{(t)}, \dots, \theta_{p-1}^{(t)}; \mathbf{y}_T)$

and then draw from the predictive distributions

- 1)  $y_{T+1}^{(t)}$  from  $p(y_{T+1}|\theta_1^{(t)}, \theta_2^{(t)}, \dots, \theta_p^{(t)})$
- 2)  $y_{T+2}^{(t)}$  from  $p(y_{T+2}|\theta_1^{(t)}, \theta_2^{(t)}, \dots, \theta_p^{(t)}, y_{T+1}^{(t)})$
- ...
- h)  $y_{T+h}^{(t)}$  from  $p(y_{T+h}|\theta_1^{(t)}, \theta_2^{(t)}, \dots, \theta_p^{(t)}, y_{T+1}^{(t)}, \dots, y_{T+h-1}^{(t)})$



5. Set  $t = t + 1$  and go back to step 4.
6. Discard the first  $n_{burn}$  drawn parameter and forecast values. Pick every  $d$ th draws of the remaining samples and discard the rest.

Once the samples  $\{y_{T+h}^{(1)}, y_{T+h}^{(2)}, \dots, y_{T+h}^{(M)}\}$  for each  $h = 1, \dots, H$  are obtained, the point forecasts can be derived. The goal of point forecasts is to minimize the MSE, which can be achieved by approximating the true expectations of the forecast distributions. The expectations can be approximated simply by calculating the sample averages, based on the strong law of large numbers

$$\hat{Y}_{T+h} = \frac{\sum_{i=1}^M y_{T+h}^{(i)}}{M} \rightarrow E[y_{T+h}] \text{ as } M \rightarrow \infty.$$

### 2.3.3 BS model

In this section the Bayesian switching ARMAX model is developed for a single-country. The model consists of three stages:

Stage 1:

$$Y_t = c_t + \sum_{i=1}^p \phi_i Y_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \sum_{i=1}^r \beta_i x_{t-i} + \varepsilon_t, \text{ where } \varepsilon_t \sim N(0, \sigma^2)$$

$$c_t = (1 - \gamma_t)c_{t-1} + \gamma_t \delta_t$$

Stage 2:

$$\gamma_t \sim \text{Bern}(\eta)$$

$$\delta_t \sim N(\zeta, \tau^2)$$

Stage 3:

$$\phi_i \sim N(0, 100^2) \text{ for } i = 1, \dots, p$$

$$\theta_i \sim N(0, 100^2) \text{ for } i = 1, \dots, q$$

$$\beta_i \sim N(0, 100^2) \text{ for } i = 1, \dots, r$$

$$\zeta \sim N(0, 100^2)$$

$$\sigma^2 \sim \text{Inv-Gamma}(\text{shape} = 0.0001, \text{scale} = 0.0001)$$

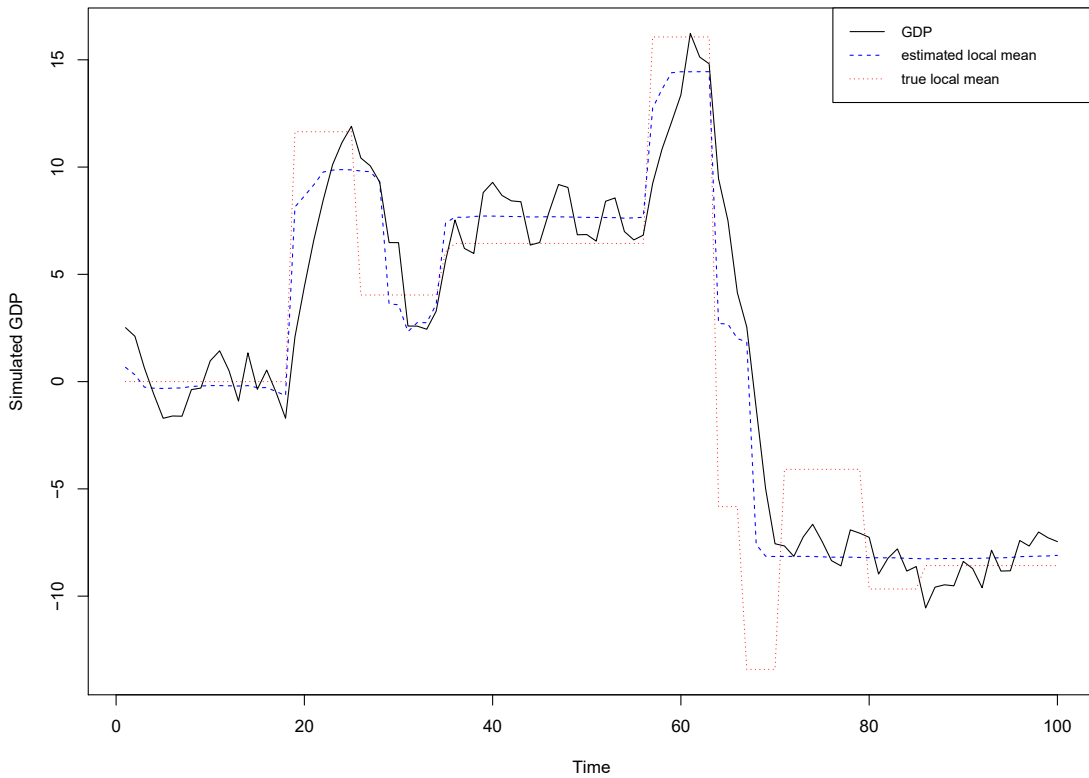
$$\tau^2 \sim \text{Inv-Gamma}(\text{shape} = 0.0001, \text{scale} = 0.0001)$$

$$\eta \sim \text{Uniform}(0, 1)$$

In addition to the time varying intercept term, the observation depends on the lagged observations of the dependent series  $Y$ , the external covariate  $x$ , and the error term  $\varepsilon$ , and their coefficients  $\boldsymbol{\phi} = (\phi_1, \phi_2, \dots, \phi_p)'$ ,  $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_r)'$  and  $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_q)'$ . The variance of the error term  $\varepsilon_t$  is  $\sigma^2$ . This is essentially the same model as the ARMAX in ( 2.1), except for the time varying intercept term.

The components  $\gamma_t$  and  $\delta_t$  of the time varying intercept term  $c_t$  are generated at Stage 2. The first component  $\gamma_t$  is a Bernoulli distributed random variable, which determines whether or not the intercept term shifts. The probability of a shift is determined by  $\eta$ . The second component  $\delta_t$  is a Gaussian variable, which is the new value of the intercept in case of a shift. The parameters of  $\delta_t$  are mean  $\zeta$  and variance  $\tau^2$ . If no shift occurs, the intercept term remains stagnant.

Parameters  $\phi, \theta, \beta, \sigma^2, \eta, \zeta$  and  $\tau^2$  are generated at the third stage. Any prior knowledge is not incorporated at this stage, so noninformative hyperparameters are chosen for the distributions of those parameters. All Gaussian parameters  $\phi, \theta, \beta$  and  $\zeta$  have mean 0 and variance  $100^2$ . The variance parameters  $\sigma^2$  and  $\tau^2$  of the model are assumed to follow the inverse gamma distribution, with both shape and scale parameters having value 0.0001. Finally, the probability parameter  $\eta$  of the intercept shift is drawn from  $\text{Uniform}(0,1)$ . The posterior distributions of the model parameters are derived in detail in Appendix A.



**Figure 2.1** A simulated example from the process matching to the proposed Bayesian switching model. The figure shows the true local means and estimated local means as well as the observed data.

### An example of BS model

100 observations are generated from BS model to demonstrate the process for local means. The example process is a simple AR(1), with  $\phi = 0.8$  and  $\sigma^2 = 1$ :

$$\begin{aligned} y_t &= c_t + 0.8y_{t-1} + \varepsilon_t, \text{ where } \varepsilon_t \sim N(0, 1) \\ c_t &= (1 - \gamma_t)c_{t-1} + \gamma_t\delta_t, \end{aligned}$$

and

$$\begin{aligned} \gamma_t &\sim \text{Bern}(0.1) \\ \delta_t &\sim N(0, 1). \end{aligned}$$

Figure 2.1 shows, that the time series locally behaves like an AR(1), but the process might suddenly experience a drift to another local mean  $\left(\frac{\delta_t}{1-\phi}\right)$  instead of  $\left(\frac{c_{t-1}}{1-\phi}\right)$ , due to  $\gamma_t$  getting value 1 instead of 0. The posterior analysis enables us to roughly estimate the true local means in this example.

### 2.3.4 MuB model

In the multi-country Bayesian ARMAX the time points are  $t = 1, \dots, T$ , as before, but now a country index  $n = 1, \dots, N$  is added. The observations from the dependent series are collected in a  $(T \times N)$  matrix

$$\mathbf{y}_{t,n} = \begin{bmatrix} y_{1,1} & y_{1,2} & \cdots & y_{1,N} \\ y_{2,1} & y_{2,2} & \cdots & y_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ y_{T,1} & y_{T,2} & \cdots & y_{T,N} \end{bmatrix},$$

where rows present time points of observations and columns represent the series, from which variable  $y$  has been measured. Similarly, the exogenous covariate series are collected in a  $(T \times N)$  matrix

$$\mathbf{x}_{t,n} = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,N} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ x_{T,1} & x_{T,2} & \cdots & x_{T,N} \end{bmatrix}.$$

The model can be described in 3 stages:

Stage 1:

$$Y_{t,n} = c_n + \sum_{i=1}^p \phi_{i,n} Y_{t-i,n} + \sum_{i=1}^q \theta_{i,n} \varepsilon_{t-i,n} + \sum_{i=1}^r \beta_{i,n} x_{t-i,n} + \varepsilon_t,$$

where  $\varepsilon_t \sim N(0, \sigma_n^2)$

Stage 2:

$$\begin{aligned} c_n &\sim N(\lambda_c, \psi_c^2) \\ \phi_{i,n} &\sim N(\lambda_{\phi_i}, \psi_{\phi_i}^2), \text{ for } i = 1, \dots, p \\ \theta_{i,n} &\sim N(\lambda_{\theta_i}, \psi_{\theta_i}^2), \text{ for } i = 1, \dots, q \\ \beta_{i,n} &\sim N(\lambda_{\beta_i}, \psi_{\beta_i}^2), \text{ for } i = 1, \dots, r \\ \sigma_n^2 &\sim \text{Inv-Gamma}(\text{shape} = 0.0001, \text{scale} = 0.0001) \end{aligned}$$

Stage 3:

$$\begin{aligned} \lambda_c &\sim N(0, 100^2) \\ \lambda_{\phi_i} &\sim N(0, 100^2), \text{ for } i = 1, \dots, p \\ \lambda_{\theta_i} &\sim N(0, 100^2), \text{ for } i = 1, \dots, q \\ \lambda_{\beta_i} &\sim N(0, 100^2), \text{ for } i = 1, \dots, r \\ \psi_c^2 &\sim \text{Inv-Gamma}(\text{shape} = 0.0001, \text{scale} = 0.0001) \\ \psi_{\phi_i}^2 &\sim \text{Inv-Gamma}(\text{shape} = 0.0001, \text{scale} = 0.0001), \text{ for } i = 1, \dots, p \\ \psi_{\theta_i}^2 &\sim \text{Inv-Gamma}(\text{shape} = 0.0001, \text{scale} = 0.0001), \text{ for } i = 1, \dots, q \\ \psi_{\beta_i}^2 &\sim \text{Inv-Gamma}(\text{shape} = 0.0001, \text{scale} = 0.0001), \text{ for } i = 1, \dots, r \end{aligned}$$

The intercept term  $c_n$  is constant for each country  $n$  in MuB model. They are drawn from a Gaussian distribution in Stage 2, with a global mean  $\lambda_c$  and a global variance  $\psi_c^2$ . The global hyperparameters themselves are drawn in Stage 3 from global noninformative hyperpriors. The rest of the model's coefficients  $\phi_{1,n}, \dots, \phi_{p,n}, \theta_{1,n}, \dots, \theta_{q,n}, \beta_{1,n}, \dots, \beta_{r,n}$  are generated in a similar manner. The Gibbs sampler for this model is given in Appendix A.

The core idea here is that each country-wise submodel has mutually corresponding terms that might share similar characteristics. The coefficients, that appear to be have the similar mean (denoted by  $\lambda$ ) according to posterior analysis, get small values for the the variance terms (denoted by  $\psi$ ). Conversely, the coefficients that are not very similar among the countries, get larger values for the variance terms.

### 2.3.5 MuBS model

The full multi-country Bayesian switching ARMAX (MuBS) model combines the random intercept and the multi-country aspects of the two previous BS and MuB

models. The model can be described in 4 stages:

Stage 1:

$$Y_{t,n} = c_{t,n} + \sum_{i=1}^p \phi_{i,n} Y_{t-i,n} + \sum_{i=1}^q \theta_{i,n} \varepsilon_{t-i,n} + \sum_{i=1}^r \beta_{i,n} x_{t-i,n} + \varepsilon_t,$$

where  $\varepsilon_t \sim N(0, \sigma_n^2)$

$$c_{t,n} = (1 - \gamma_{t,n})c_{t-1,n} + \gamma_{t,n}\delta_{t,n}$$

Stage 2:

$$\gamma_{n,t} \sim \text{Bern}(\eta_n)$$

$$\delta_{n,t} \sim N(\zeta_n, \tau_n^2)$$

Stage 3:

$$\phi_{i,n} \sim N(\lambda_{\phi_i}, \psi_{\phi_i}^2), \text{ for } i = 1, \dots, p$$

$$\theta_{i,n} \sim N(\lambda_{\theta_i}, \psi_{\theta_i}^2), \text{ for } i = 1, \dots, q$$

$$\beta_{i,n} \sim N(\lambda_{\beta_i}, \psi_{\beta_i}^2), \text{ for } i = 1, \dots, r$$

$$\eta_n \sim \text{Uniform}(0, 1)$$

$$\zeta_n \sim N(0, \omega^2)$$

$$\sigma_n^2 \sim \text{Inv-Gamma}(\text{shape} = 0.0001, \text{scale} = 0.0001)$$

$$\tau_n^2 \sim \text{Inv-Gamma}(\text{shape} = 0.0001, \text{scale} = 0.0001)$$

Stage 4:

$$\lambda_{\phi_i} \sim N(0, 100^2), \text{ for } i = 1, \dots, p$$

$$\lambda_{\theta_i} \sim N(0, 100^2), \text{ for } i = 1, \dots, q$$

$$\lambda_{\beta_i} \sim N(0, 100^2), \text{ for } i = 1, \dots, r$$

$$\psi_{\phi_i}^2 \sim \text{Inv-Gamma}(\text{shape} = 0.0001, \text{scale} = 0.0001), \text{ for } i = 1, \dots, p$$

$$\psi_{\theta_i}^2 \sim \text{Inv-Gamma}(\text{shape} = 0.0001, \text{scale} = 0.0001), \text{ for } i = 1, \dots, q$$

$$\psi_{\beta_i}^2 \sim \text{Inv-Gamma}(\text{shape} = 0.0001, \text{scale} = 0.0001), \text{ for } i = 1, \dots, r$$

$$\omega^2 \sim \text{Inv-Gamma}(\text{shape} = 0.0001, \text{scale} = 0.0001)$$

The coefficients of MuBS are generated similarly to MuB as well as the observations' variance parameters  $\sigma_n^2$ . The time varying intercept terms  $c_{t,n}$  are similar to the one in BS model, except the distribution of  $\delta_{t,n}$  are drawn from  $N(\zeta_n, \tau_n^2)$  instead of  $N(0, 100^2)$ . Here,  $\zeta_n$  are drawn from Gaussian distribution, where the mean equals to 0 and  $\omega^2$  is the variance. The distributions of  $\omega^2$  and  $\tau^2$  are both inverse gammas, with shape and scale being 0.0001. The Gibbs sampler for this model is given in Appendix A.

## 2.4 Forecast evaluation

An out-of-sample forecast analysis requires that a part of the data is used solely to model estimation, and the forecasts from that model are compared to data that has not been utilized in model estimation. The data are divided in two parts: a training set and a test set. The end of the training set is the forecast origin  $T$ , and the forecasted data point at  $T + h$  lies in the test set. The models are recursively re-estimated as the forecast analysis progresses, so that all data up to the forecast origin  $T$  is used to estimate the models.

Suppose the length of the complete data set is  $N$  and the task is to test the forecast accuracy of a time series model on observations  $M + h, M + h + 1, \dots, N - 1, N$ . Here  $M$  is referred to as the first forecast origin and  $N$  is the length of the full data set. The model is first estimated using observations  $\{y_1, y_2, \dots, y_M\}$ , from which the forecast  $\hat{Y}_{M+h|M}$  is obtained. Then, the forecast error  $e_{M+h}$  can be calculated as

$$e_{T+h} = y_{T+h} - \hat{Y}_{T+h|T},$$

substituting  $T = M$ . After calculating the first forecast error, the model is re-estimated using observations 1 to  $M + 1$ . Now the forecast origin of the model is updated to  $T = M + 1$ , and the estimated model is used to obtain forecast  $\hat{Y}_{M+1+h|M+1}$ . After obtaining the forecast, the next forecast error  $e_{M+1+h}$  can be calculated. In this manner forecasts  $\{\hat{Y}_{M+h|M}, \hat{Y}_{M+1+h|M+1}, \dots, \hat{Y}_{N-h+h|N-h}\}$ , and the corresponding forecast errors  $\{e_{M+h}, e_{M+h+1}, \dots, e_N\}$  are obtained.

There are many approaches to evaluate the forecast accuracy using the sequence of forecast errors  $\{e_{M+h}, e_{M+h+1}, \dots, e_N\}$ . In the forecast analysis

$$\text{Mean absolute forecast error: MAFE} = \frac{\sum_{t=M+h}^N |e_t|}{N},$$

and

$$\text{Root mean squared forecast error: RMSFE} = \sqrt{\frac{\sum_{t=M+h}^N e_t^2}{N}}$$

are compared between the models. For both measures, the smaller number is better. The main difference between them is that RMSFE emphasizes large forecast errors in comparison to MAFE.

### 3 Empirical forecast analysis

The forecasting characteristics of the models listed in Table 3.1 are contrasted on the empirical GDP growth rates from 13 countries. The table summarizes the main properties of the models.

Model	Random intercept	Multi-country	AR	MA-term	Covariate
ARMAX			✓	✓	✓
MS	✓		✓		✓
BS	✓		✓	✓	✓
MuB		✓	✓	✓	✓
MuBS	✓	✓	✓	✓	✓

**Table 3.1** *The main properties of the five models compared in the forecast analysis.*

Quarterly observations, for both GDP growth rates and CLIs, span from the first quarter of 1969 to the fourth quarter of 2018. This 50-year period comprehends 200 observations per country. The details of the variables are described in the following section. Prior to the forecast analysis, the observations up to the first forecast origin, the fourth quarter of 1999, are used to determine the model specifications ( $p$ ,  $q$  and  $r$ ). The details of the model specifications are outlined in 3.2. The results of the forecast analysis are explored in 3.3.

#### 3.1 Description of data

The empirical forecast analysis focuses on 13 OECD countries. In the analysis, the CLI is as an exogenous covariate to forecasting GDP growth rates. Among all the countries in the world, the selected 13 countries are the ones with enough data available to cover the period from 1969 to 2018. The data is gathered by OECD, and it is downloaded from the data bank of the Federal Reserve.

##### 3.1.1 GDP growth rate

GDP measures the market value of final goods and services produced in a country within a year, minus the value of imports. The intermediate consumption required to make the products is not part of GDP. The data from the OECD are inflation as well as seasonally adjusted year-on-year growth rates of GDP. The GDP processing by OECD is as follows:

Given a set  $C$ , containing all goods and services in the economy, nominal GDP

at period  $t$  is ideally calculated as

$$\text{nominal GDP}_t = \sum_{c \in C} P_{c,t} Q_{c,t},$$

where  $P_{c,t}$  is the price of product  $c$  in period  $t$ , and  $Q_{c,t}$  is the quantity of product  $c$  in period  $t$ . The nominal GDP is inflation adjusted to real terms using chain-weighting, a method that measures GDP in terms of current and previous period prices (Steindel 1995). After obtaining the real GDP, it is purified from seasonal effects, such as within-year pattern, using X-12-ARIMA adjustment program. See Findley et al. (1998) for detailed explanation of the adjustment program. Finally, the year-on-year growth rate can be calculated from seasonally adjusted real GDP:

$$\text{Year-on-year growth rate}_t = \frac{\text{adjusted GDP}_t - \text{adjusted GDP}_{t-4}}{\text{adjusted GDP}_{t-4}}.$$

The governments represented in the forecasting experiment are committed to compile their GDP series according to the System of National Accounts (SNA) 2008 standard. The standard aims to make different countries' economies comparable to each other. Compiling GDP is not a simple task, and GDP series go through frequent revisions as new information is revealed. The data used in this thesis is revised at the release of the fourth quarter of 2018. Usually revisions are based on additional information received during the year, and the changes are modest. Despite the common standard, the impact of the revisions might vary between countries, because the governments may face unique data collecting issues.

Table 3.2 summarizes the GDP growth rates of the countries under examination. The US economy was clearly the largest economy in 1968, as it was in 2017. Over the 49-year period the US economy grew more than 20-fold. During the same time frame, the Australian economy experienced the fastest growth, resulting in more than 40-fold increase. The Japanese economy had the most uneven growth in terms of standard deviation (SD).

### 3.1.2 Composite leading indicator

CLI is an indicator variable designed by OECD to give early signals of turning points in business cycles. Business cycle is perceived as fluctuation in GDP around its long term trend. OECD publishes CLI monthly, but here the CLI series are converted to quarterly frequency by averaging out the within quarter data points. According to OECD, the leading time varies, but should be around two to three quarters ahead of GDP (see Gyomai and Guidetti 2012).

---

<sup>1</sup>Total GDP values are in 2017 US dollars. Data provided by World Bank.



country	Total GDP in billions of \$US <sup>1</sup>			Year-on-year growth rates			
	1968	2017	fold change	Mean	Min	Max	SD
US	<b>943</b>	<b>19391</b>	20.6	2.768	-3.924	8.578	2.150
Japan	147	4872	33.1	2.847	-8.682	12.844	<b>3.241</b>
Germany	198	3677	18.6	2.181	-6.935	7.712	2.284
UK	105	2622	25.0	2.255	-6.082	9.748	2.267
France	130	2583	19.9	2.299	-3.780	<b>15.486</b>	1.947
Italy	88	1935	22.0	1.879	-7.154	9.851	2.672
Canada	71	1653	23.3	2.781	-4.070	8.316	2.214
Australia	33	1323	<b>40.1</b>	<b>3.246</b>	-3.406	8.389	1.923
Netherlands	28	826	29.5	2.454	-4.357	8.186	2.237
Switzerland	18	679	37.7	1.874	<b>-8.898</b>	9.730	2.399
Belgium	21	493	23.5	2.323	-3.808	7.279	2.060
Austria	12	416	34.7	2.554	-5.143	9.334	2.158
Denmark	13	324	24.9	1.956	-6.168	7.368	2.324

**Table 3.2** Description of the countries' GDP data. The economies are sorted by 2017 total GDP in descending order.

CLI combines measurements of several variables from OECD's main economic indicator database. The component series have an economically justified leading relationship to GDP. Other common features include monthly frequency, lack of significant revisions and quick availability whenever the period occurs. Examples of common component series include stock indices, manufacturing order books, the number of newly issued dwelling permits, consumer confidence indicators and interest rate spreads. The component series for each country's CLI are listed in Table 3.3 (further details in OECD 2020).

Country	List of component series
US	Work started for dwellings sa (number), Net new orders - durable goods sa (USD), Share prices: NYSE composite (2015=100), Consumer - Confidence indicator sa (normal = 100), Weekly hours worked: manufacturing sa (hours), Manufacturing - Industrial confidence indicator (% balance), Spread of interest rates (% p.a.)

- Japan Ratio of inventories to shipments sa (2015=100) inverted, ITS Volume of imports/volume of exports sa (2015=100), Ratio loans to deposits sa (%) inverted, Monthly hours worked: manufacturing sa (2015=100), Work started for dwellings sa (2015=100), Share prices: TOPIX index (2015=100), Spread of interest rates (% p.a.), Small Business Survey: Sales tendency (% balance)
- Germany IFO business climate indicator (normal=100), Orders inflow/demand (manuf.): tendency (% balance), Export order books (manuf.): expectation (% balance), New orders in manuf. industry (2010 = 100), Finished goods stocks (manuf.): level (% balance) inverted, Spread of interest rates (% p.a.), Services – Demand evolution: future tendency (% balance), Consumer - Confidence indicator sa (% balance)
- UK Services – Demand evolution: future tendency (% balance), Passenger car registrations sa (number), Consumer - Confidence indicator sa (% balance), Great British Pound Interbank LIBOR 3 Months Delayed (% p.a.) inverted, Manufacturing - Production: future tendency sa (% balance), Share prices: FTSE LOCAL UK (£) index (2015=100)
- France Passenger car registrations sa (number), Consumer - Confidence indicator sa (% balance), Manufacturing - Production: future tendency sa (% balance), Share prices: CAC All-tradable index (2015=100), CPI HICP All items (2015=100) inverted, Manufacturing - Export order books: level sa (% balance), Construction - Selling prices: future tendency sa (% balance), Permits issued for dwellings sa (2015=100)
- Italy Consumer - Confidence indicator sa (% balance), Manufacturing - Order books: level sa (% balance), Deflated orders for total manufactured goods (Value) sa (2010 = 100), Manufacturing - Production: future tendency sa (% balance), CPI All items (2010=10) inverted, Imports from Germany c.i.f. (USD)
- Canada Deflated Monetary aggregate M1 sa (2010 cad), Manufacturing - Industrial confidence indicator (USA - PMI) sa, Consumer - Confidence indicator (2015=100), Spread of interest rates (% p.a.), Ratio of inventories to shipments inverted, Share prices: S&P/TSX composite index (2015=100)

Australia	Permits issued for dwellings sa (number), Manufacturing - Orders inflow: tendency sa (% balance), Manufacturing - Production: tendency sa (% balance), Manufacturing - Employment: tendency sa (% balance), Share prices: S&P/ASX 200 index (2015=100), Terms of trade, Yield 10-year commonwealth government bonds (% p.a.) inverted
Netherlands	Manufacturing - Order books: level sa (% balance), Manufacturing - Production: future tendency sa (% balance), Manufacturing - Finished goods stocks: level sa (% balance) inverted, Manufacturing - Business situation of Germany: present sa (normal=100), Services - Demand evolution: future tendency (% balance), Consumer - Confidence indicator sa (% balance), Share prices: AEX index (2015=100)
Switzerland	Manufacturing - Finished goods stocks: level (% balance) inverted, Manufacturing - Orders inflow: tendency sa (% balance), Manufacturing - Production: future tendency sa (% balance), Share prices: UBS-100 index (2015=100), Consumer - Expected economic situation sa (% balance), Silver prices (CHF/kJ)
Belgium	Passenger car registrations sa (number), Manufacturing - Employment: future tendency sa (% balance), Manufacturing - Export orders inflow: tendency sa (% balance), Manufacturing - Demand: future tendency sa (% balance), Manufacturing - Production: tendency sa (% balance), Consumer - Confidence indicator sa (% balance)
Austria	Manufacturing - Production: future tendency sa (% balance), Manufacturing - Order books: level sa (% balance), Manufacturing - Business situation of Germany: present sa (normal=100), Consumer - Confidence indicator sa (% balance), Job vacancies: unfilled sa (persons), Spread of interest rates (% p.a.)
Denmark	Total retail trade (Volume) sa (2015=100), Passenger car registrations sa (number), Manufacturing - Employment: future tendency sa (% balance), Manufacturing - Production: future tendency sa (% balance), Central bank official discount rate (% per annum) inverted, Deflated monetary aggregate M1 (dkk), ITS Mineral fuels, lubricants and related materials exports, deflated CPI energy sa (dkk), Consumer - Confidence indicator sa (% balance)

---

**Table 3.3** *The models selected to the simulated forecasting experiment*

The construction of CLI series from the component series is a two step process. The first step is filtering and normalizing the individual component series. During filtering, factors such as seasonal patterns, outliers, high frequency noise and trends are removed. Then, the filtered series are normalized by subtracting the mean and dividing by the mean absolute deviation of the series. The second step is aggregation, which combines the component series to one. It is done by averaging the growth rates of the processed component series.

A new CLI observation is published as soon as 60 percent of the required component series observations become available. The series are published with two months lag, which can be cumbersome to a forecaster. CLI observations M1:M3 correspond to the first quarter of GDP. Similarly M4:M6 correspond to the second quarter of GDP, M7:M9 to the third quarter of GDP and M10:M12 to the fourth quarter of GDP. Given the two month lag, at least one monthly CLI can be obtained before the release of GDP estimate, which can be utilized in prediction.

### 3.2 Model specification

The assumptions of the models in Table 3.1 require different strategies for determining specifications of  $p, q$  and  $r$ . ARMAX and BS models allow unique specifications for each country, whereas MuB and MuBS models demand a common specification among countries. MS allows unique specifications for each country, but they have to be determined separately from ARMAX and BS specifications, because the MA term is excluded.

The AIC of ARMAX is assessed for every combination of  $p, q, r \in \{0, 1, 2, 3, 4\}$ . Hence, the AIC of 125 models are compared for each country. The first part of Table 3.4 lists the specifications of ARMAX model that produced the smallest AIC values. These specifications are used for both ARMAX and BS models. The second part of the table shows the specifications of MS model providing the smallest AIC values.

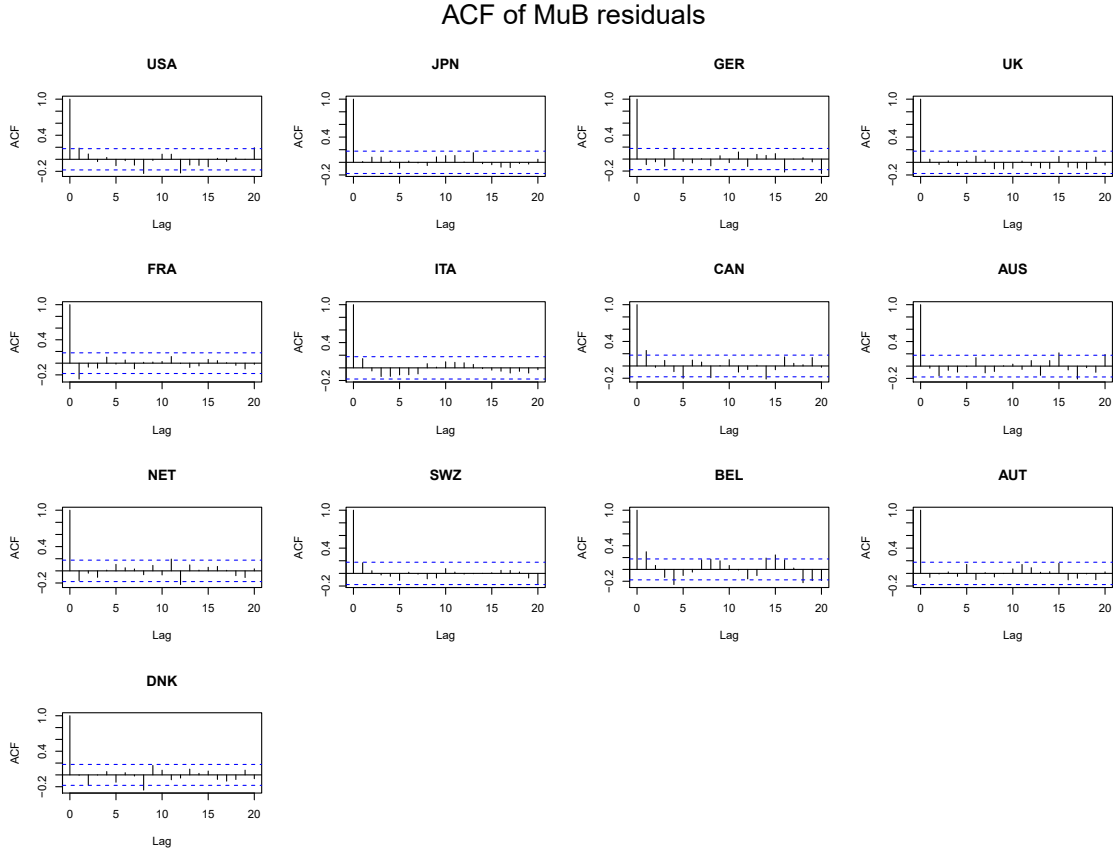
As the country-specific parameters of MuB and MuBS models are drawn from common distributions, the model specifications are common among countries. We choose the specifications of the multi-country models rather general, so that sufficient description of every country is achieved. Figure 3.1 shows that selecting  $p = 4$ ,  $q = 4$  and  $r = 4$  for MuB model produces seemingly independent model residuals, and hence, this specification is used for both multi-country models in the forecast analysis.

		AR term (p)	MA term (q)	Covariate (r)
ARMA and BN-ARMA	USA	0	3	4
	JPN	4	3	0
	GER	1	1	2
	UK	4	3	0
	FRA	3	4	0
	ITA	0	4	0
	CAN	1	3	3
	AUS	0	3	4
	NET	0	3	4
	SWZ	4	3	0
	BEL	1	4	0
	AUT	0	4	3
	DNK	0	3	4
MS-AR	USA	1		4
	JPN	4		0
	GER	3		4
	UK	4		0
	FRA	4		0
	ITA	1		4
	CAN	1		4
	AUS	2		4
	NET	2		2
	SWZ	1		3
	BEL	1		4
	AUT	4		0
	DNK	1		4

**Table 3.4** The model specifications for the 13 countries. The upper section lists the model specifications of ARMAX and BS models, and the bottom section of MS model.

### 3.3 Results

Tables 3.5 and 3.6 summarize the overall forecasting accuracies of the models on each of four forecast horizons. The values on the table are each model's RMSFE (and MAFE) averaged over the 13 countries. For each forecast horizon MuB model had the greatest accuracy in terms of RMSFE and MAFE. Interestingly, BS model performs better compared to ARMAX (model specifications according to Table 3.4), but there does not appear a great difference between MuBS and MuB models. MS model is slightly more accurate than ARMAX on every forecast horizon, according



**Figure 3.1** Sample autocorrelations of the means of the posterior model residuals from MuB, with  $p = 4$ ,  $q = 4$  and  $r = 4$ .

to the average RMSFE and MAFE values, but not as accurate as BS, MuB or MuBS.

	ARMAX	MS	BS	MuB	MuBS
1-step	0.944	0.903	0.763	0.714	<b>0.712</b>
2-step	1.537	1.426	1.282	1.215	<b>1.205</b>
3-step	1.986	1.850	1.740	1.671	<b>1.649</b>
4-step	2.333	2.218	2.117	2.077	<b>2.049</b>

**Table 3.5** Average RMSFE over the 13 countries

Tables 3.7 and 3.8 provide country-specific insight to the forecasting accuracies. The values are averages of RMSFE (and MAFE) statistics over the four forecast horizons. In terms of RMSFE, MuBS was the most accurate model in 6 cases of 13, and MuB in 5 cases, and BS in the remaining 2 cases. When assessing the performance in terms of MAFE, MuBS was the most accurate in 6 cases, MuB in 5 cases, BS in 1 case and ARMAX in 1 case. MS was more accurate than ARMAX in 10 and 9 cases in terms of RMSFE and MAFE respectively.

	ARMAX	MS	BS	MuB	MuBS
1-step	0.676	0.695	0.543	0.520	<b>0.511</b>
2-step	1.075	1.058	0.899	0.852	<b>0.830</b>
3-step	1.394	1.346	1.215	1.171	<b>1.136</b>
4-step	1.655	1.596	1.487	1.468	<b>1.423</b>

*Table 3.6 Average MAFE over the 13 countries*

These results suggest that the random intercept and multi-country properties both provide possible improvements to forecasting GDP growth rates, compared to ARMAX or MS models. It appears that most accuracy improvements can be gained from the multi-country property.

	ARMAX	MS	BS	MuB	MuBS
USA	1.489	1.443	1.326	1.172	<b>1.170</b>
JPN	2.202	2.809	2.305	2.288	<b>2.148</b>
GER	2.607	2.005	2.017	<b>1.784</b>	1.874
UK	1.541	1.575	<b>1.258</b>	1.266	1.271
FRA	1.535	1.352	1.236	<b>1.061</b>	1.238
ITA	1.939	1.756	1.745	1.763	<b>1.583</b>
CAN	1.500	1.291	1.285	<b>1.231</b>	1.265
AUS	0.915	1.028	0.870	0.809	<b>0.787</b>
NET	1.712	1.560	1.585	<b>1.442</b>	1.471
SWZ	1.546	1.449	<b>1.230</b>	1.395	1.298
BEL	1.632	1.293	1.220	1.260	<b>1.125</b>
AUT	1.713	1.511	1.439	1.348	<b>1.292</b>
DNK	1.770	1.719	1.665	<b>1.632</b>	1.727

*Table 3.7 Average RMSFE over the forecast horizons*

### 3.3.1 A case study: US GDP growth rate forecasts

To further illustrate the differences between the models, we inspect the forecasts of US GDP growth rates visually in more detail. Figures 3.2 and 3.3 show the forecasts from ARMAX, MS and MuBS models against the actual data. The first figure demonstrates the 1-step ahead forecasts from MS and ARMAX models. In terms of RMSFE, the forecast accuracy of MS is 1.443, providing a slight improvement to ARMAX's 1.489. However, the two models appear to have different advantages. The ARMAX forecasts tend to be more accurate during the expansionary times,

	ARMAX	MS	BS	MuB	MuBS
USA	1.054	1.080	0.894	0.799	<b>0.798</b>
JPN	1.435	2.054	1.622	1.692	<b>1.454</b>
GER	1.707	1.410	1.352	<b>1.213</b>	1.311
UK	0.994	1.010	0.764	0.755	<b>0.754</b>
FRA	1.099	1.079	0.907	<b>0.799</b>	0.862
ITA	1.305	1.231	1.215	1.265	<b>1.080</b>
CAN	1.141	1.016	0.972	<b>0.881</b>	0.892
AUS	0.752	0.852	0.711	0.624	<b>0.623</b>
NET	1.224	1.143	1.090	0.962	<b>1.034</b>
SWZ	1.169	1.088	<b>0.908</b>	1.023	0.920
BEL	1.162	0.989	0.865	0.950	<b>0.791</b>
AUT	1.296	1.072	0.995	0.919	<b>0.879</b>
DNK	1.258	1.237	1.172	<b>1.156</b>	1.278

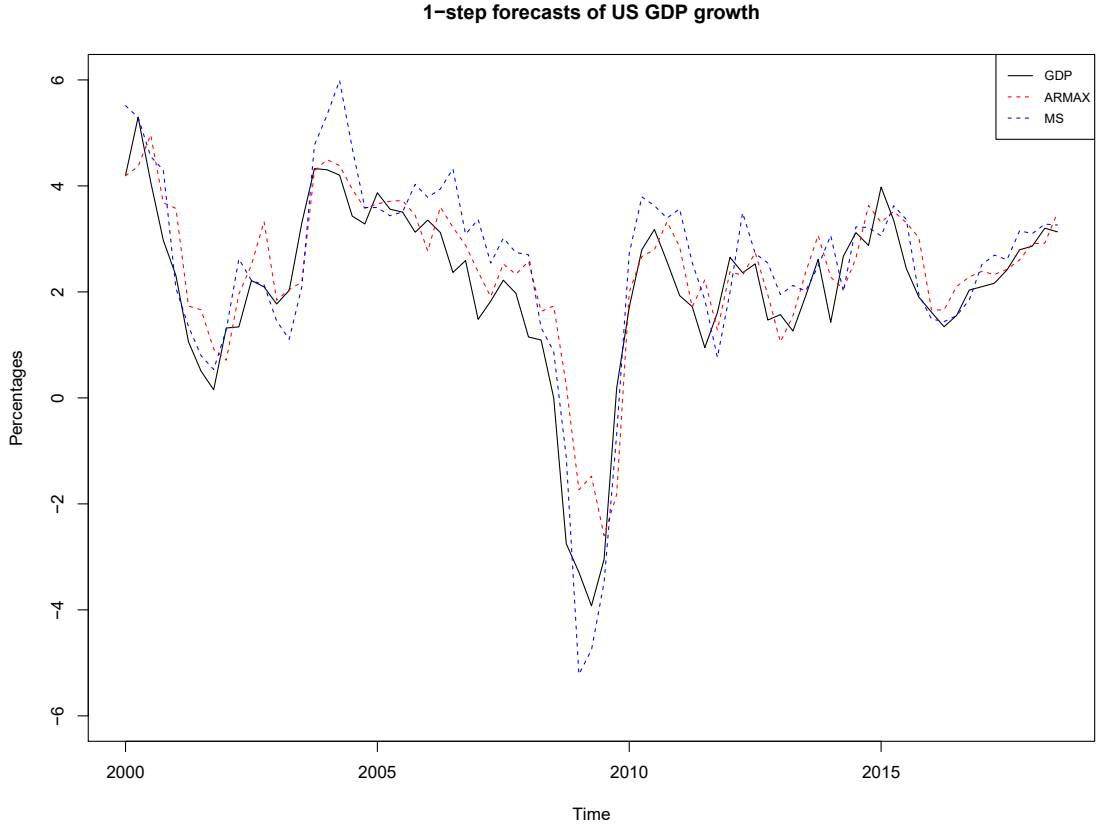
**Table 3.8** Average MAFE over the forecast horizons

but during the sharp declines, for example soon after the financial crisis of 2008, the forecasts tend to overestimate the growth rate. At that time, MS forecasts are more accurate, but also MS model tends to overestimate the growth at turning points.

Figure 3.3 contrast the forecasts from MuBS and ARMAX models. MuBS had the most accurate forecasts of all models for the US GDP growth rate, with RMSFE of 1.17. The model appears to work better than ARMAX during the declines, but also maintaining good accuracy during the expansionary times. The model also avoids overestimating the growth at turning points better than MS. In other words, MuBS could potentially offer better overall forecasts compared to ARMAX and MS models.

Another interesting aspect is the differences in the parameter estimates of the models. Table 3.9 compares the parameters from ARMAX and BS models, which share the same model specifications (shown in Table 2.1.2). The model parameters include 4 covariate coefficients, 4 MA coefficients, the intercept term and the standard deviation of the error term. BS model has a dynamic intercept term  $c_t$ , whereas ARMAX has a constant intercept term  $c$ . Observations from the second quarter of 1969 to the second quarter of 2018 were used to estimate the models demonstrated in the table. There appears to be major differences between the parameters estimates between the models. The covariate coefficients  $\beta_1, \dots, \beta_4$  are close to 0 for BS model, unlike for ARMAX model. The MA coefficients  $\theta_1, \dots, \theta_4$  are all negative for BS and positive for ARMAX. The standard deviation of BS model is much smaller

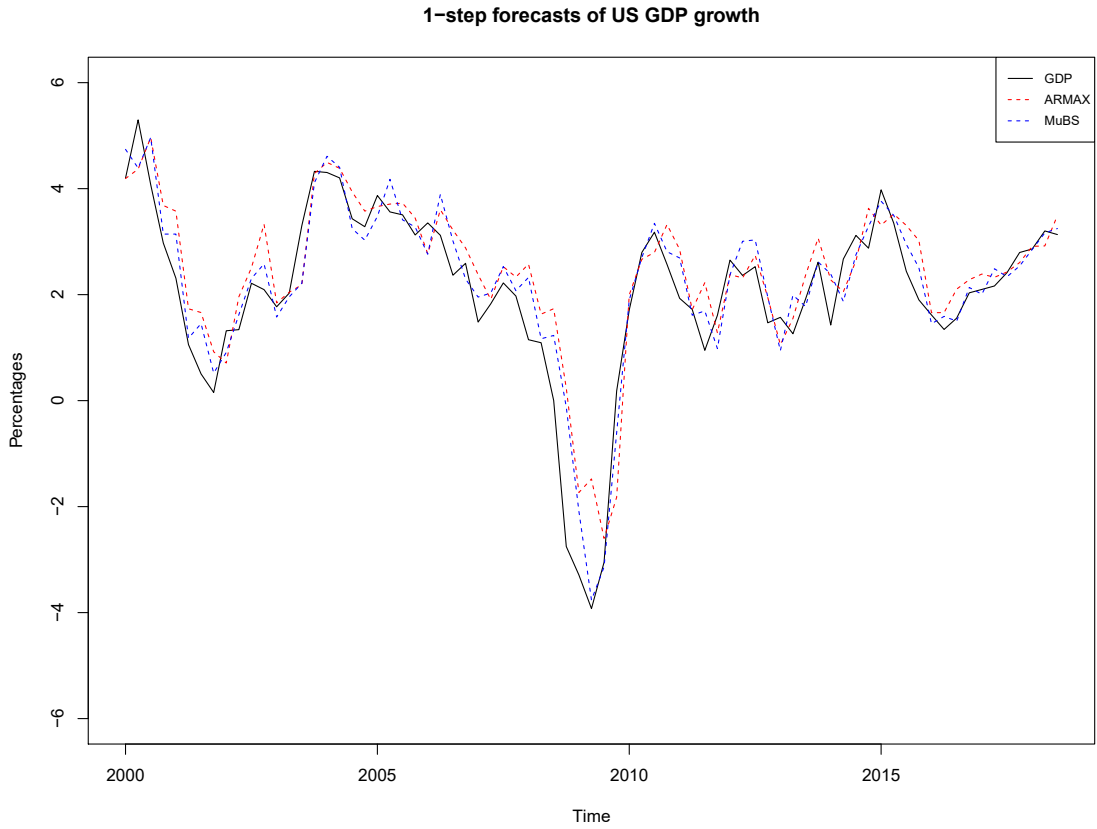




**Figure 3.2** 1-step forecasts of USA GDP growth rate from ARMAX (RMSFE: 1.489) and MS (RMSFE: 1.443) models

compared to the one from ARMAX model. At the end of the estimation period the intercept term  $c_T$  of BS model lies on the negative side (the mean of the distribution being -0.215), whereas the constant intercept term  $c$  of ARMAX model is estimated at 0.293.

Table 3.10 shows the US-specific mean, minimum, maximum, as well as 25%, 50% and 75% quartiles of the posterior parameter distributions for MuBS model. The model was estimated using observations from the second quarter of 1969 to the second quarter of 2018. The global parameters governing the country-specific parameters are listed in Table 3.11. From the table, the magnitude and the similarity of the country-specific parameters among all countries can be assessed. Looking at the standard deviation parameters in the bottom half of the table, we can observe, that coefficients  $\beta_1, \beta_2, \beta_3, \beta_4, \theta_1$  and  $\theta_2$  tend to be close to each other for each country, and most of the country-specific variation in the coefficients come from other parameters.



**Figure 3.3** 1-step forecasts of USA GDP growth rate from ARMAX (RMSFE: 1.489) an MuBS (RMSFE: 1.170) models

	ARMAX			BS			
	CMLE	mean	min	25%	50%	75%	max
$c/c_T$	0.293	-0.215	-2.499	-0.865	-0.338	0.445	3.344
$\beta_1$	0.489	-0.0214	-0.046	-0.033	-0.029	-0.004	0.020
$\beta_2$	0.594	0.0259	0.008	0.024	0.026	0.028	0.040
$\beta_3$	-1.087	0.0308	0.019	0.024	0.029	0.039	0.045
$\beta_4$	0.031	-0.007	-0.027	-0.017	-0.012	0.006	0.013
$\theta_1$	0.758	-1.289	-1.416	-1.334	-1.293	-1.254	-1.040
$\theta_2$	0.215	-1.279	-1.404	-1.329	-1.293	-1.248	-1.037
$\theta_3$	0.699	-1.271	-1.441	-1.344	-1.271	-1.206	-1.031
$\theta_4$	0.734	-0.300	-0.457	-0.377	-0.292	-0.234	-0.103
$\sigma$	1.813	0.710	0.597	0.680	0.708	0.738	0.886

**Table 3.9** The parameter estimates for ARMAX and BS models of US GDP growth rate. The first column shows CMLE estimates for ARMAX model and the following columns show the mean and the quartiles of posterior distributions of BS model.

	mean	min	25%	50%	75%	max
$c_T$	-0.028	-2.499	-0.865	-0.338	0.445	3.344
$\phi_1$	0.595	0.357	0.555	0.594	0.639	0.819
$\phi_2$	0.091	-0.124	0.043	0.085	0.140	0.383
$\phi_3$	0.022	-0.170	-0.025	0.020	0.074	0.252
$\phi_4$	-0.157	-0.436	-0.199	-0.157	-0.114	0.070
$\beta_1$	0.034	0.007	0.027	0.037	0.040	0.051
$\beta_2$	0.017	0.014	0.016	0.016	0.018	0.025
$\beta_3$	0.005	0.005	0.005	0.005	0.005	0.005
$\beta_4$	-0.045	-0.069	-0.053	-0.046	-0.033	-0.030
$\theta_1$	-0.585	-0.652	-0.594	-0.588	-0.572	-0.522
$\theta_2$	-0.546	-0.56	-0.550	-0.547	-0.542	-0.522
$\theta_3$	-0.517	-0.551	-0.519	-0.516	-0.514	-0.501
$\theta_4$	0.417	0.415	0.417	0.417	0.417	0.418
$\sigma$	0.702	0.565	0.668	0.703	0.734	0.905

**Table 3.10** The US-specific posterior mean and the quartiles of MuBS model parameters.

	mean	min	25%	50%	75%	max
$\lambda_{\phi_1}$	0.365	0.184	0.329	0.365	0.403	0.626
$\lambda_{\phi_2}$	0.090	-0.052	0.065	0.090	0.114	0.221
$\lambda_{\phi_3}$	0.0910	-0.052	0.069	0.091	0.116	0.212
$\lambda_{\phi_4}$	-0.179	-0.643	-0.218	-0.180	-0.143	0.081
$\lambda_{\beta_1}$	0.006	0.003	0.006	0.006	0.007	0.010
$\lambda_{\beta_2}$	0.004	0.004	0.004	0.004	0.004	0.004
$\lambda_{\beta_3}$	-0.009	-0.010	-0.009	-0.009	-0.009	-0.008
$\lambda_{\beta_4}$	0.017	0.017	0.017	0.017	0.017	0.017
$\lambda_{\theta_1}$	-0.797	-0.800	-0.797	-0.797	-0.797	-0.796
$\lambda_{\theta_2}$	-0.749	-0.749	-0.749	-0.749	-0.749	-0.749
$\lambda_{\theta_3}$	-0.716	-0.739	-0.723	-0.713	-0.710	-0.696
$\lambda_{\theta_4}$	0.198	0.166	0.197	0.197	0.199	0.239
$\psi_{\phi_1}$	0.191	0.096	0.158	0.185	0.213	0.426
$\psi_{\phi_2}$	0.124	0.051	0.099	0.120	0.143	0.268
$\psi_{\phi_3}$	0.099	0.012	0.076	0.097	0.119	0.264
$\psi_{\phi_4}$	0.211	0.094	0.174	0.203	0.239	0.603
$\psi_{\beta_1}$	1e-4	1e-10	7e-9	2e-7	4e-5	0.002
$\psi_{\beta_2}$	3e-8	2e-13	2e-10	2e-9	1e-8	1e-6
$\psi_{\beta_3}$	5e-5	2e-12	1e-8	1e-5	7e-5	3e-4
$\psi_{\beta_4}$	2e-6	3e-18	1e-14	4e-12	8e-7	5e-5
$\psi_{\theta_1}$	5e-5	3e-9	2e-7	2e-6	2e-5	6e-4
$\psi_{\theta_2}$	5e-9	3e-15	3e-13	1e-12	1e-11	2e-7
$\psi_{\theta_3}$	0.004	1e-5	0.001	0.004	0.007	0.012
$\psi_{\theta_4}$	0.006	4e-7	3e-5	4e-4	0.008	0.059

**Table 3.11** The posterior mean and the quartiles of global model parameters from MuBS.

## 4 Conclusions

In this thesis five methods for GDP growth rate forecasting have been evaluated on 13 OECD countries. CLI was chosen to be an explanatory covariate for the GDP growth rate in all models. The out-of-sample forecasting period spans from the first quarter of 2000 to the fourth quarter of 2018, covering turbulent eras like 2008-2012.

ARMAX and Markov switching models (Hamilton, 1989) were employed first to explore the forecasting performance of the GDP growth rates. We present a new Bayesian model with a random intercept process and a multi-country property to describe the GDP growth rates. The model assumes stochastic breaks in the process similarly to MS model. The difference is that in case of a stochastic break, the Bayesian model assumes a new intercept term to be drawn from a Gaussian distribution, whereas the post-break intercept term of MS is distributed between  $k$  constants. The coefficients of the Bayesian model are drawn, for each country, from common distributions sharing global hyperparameters. This causes shrinkage between the corresponding coefficients of the country-specific submodels. Three versions of the proposed Bayesian model are examined. The first version focuses on the random intercept process without considering the multi-country variation. Conversely, the second model adopts the multi-country variation imposing the stationarity. The third version combines both features.

Point forecasts from the proposed Bayesian model were derived by taking the means of the posterior predictive distributions. Comparing the derived point forecasts to the point forecasts of the frequentist ARMAX and MS models, the proposed model gave very comparable results in terms of RMSFE. It may be argued that comparing the Bayesian prediction points and frequentist estimates directly is not straightforward, but means or medians are used quite often as point summaries in Bayesian approach.

The empirical analysis shows that the forecast results of MS are fairly close to the one of ARMAX, which goes along with the conclusions of Clements and Krolzig (1998). Averaging each model's RMSFE over the 13 countries, all versions of the Bayesian model were more accurate than either ARMAX or MS. The full version of the Bayesian model was the most accurate, but the stationary multi-country version comes very close. Due to simplicity, the stationary multi-country version might be a better choice from practical point of view. However, the Bayesian model offers a flexible foundation for further analysis of the model parameters and reliable prediction interval estimates with easily comprehensible interpretation.

## References

- [1] Clements, M. and Krolzig, H. “A comparison of the forecast performance of Markov-switching and threshold autoregressive models of US GNP”. *The Econometrics Journal* Vol. 1 (1998), pp. 47–75.
- [2] Findley, D., Monsell, B., Bell, W., Otto, M., and Chen, B. “New capabilities and methods of the X-12-ARIMA seasonal-adjustment program”. *Journal of Business and Economic Statistics* Vol. 16 (1998), pp. 127–152.
- [3] Gyomai, G. and Guidetti, E. *OECD system of composite leading indicators*. Available: <http://www.oecd.org/sdd/leading-indicators/41629509.pdf>. 2012.
- [4] Hamilton, J. “A new approach to the economic analysis of nonstationary time series and the business cycle”. *Econometrica* Vol. 57 (1989), pp. 357–384.
- [5] Kydland, F. and Prescott, E. “Time to build and aggregate fluctuations”. *Econometrica* Vol. 50 (1982), pp. 1345–1370.
- [6] Nelson, C. “The prediction performance of the FRB-MIT-PENN model of the U.S. economy”. *American Economic Review* Vol. 62 (1972), pp. 902–917.
- [7] OECD. *OECD composite leading indicators: turning points of reference series and component series*. Available: <https://www.oecd.org/sdd/leading-indicators/CLI-components-and-turning-points.pdf>. 2020.
- [8] Sims, C. “Macroeconomics and reality”. *Econometrica* Vol. 48 (1980), pp. 1–48.
- [9] Smets, F. and Wouters, R. “Shocks and frictions in US business cycles: a Bayesian DSGE approach”. *American Economic Review* Vol. 97 (2007), pp. 586–606.
- [10] Steindel, C. “Chain-weighting: the new approach to measuring GDP”. *Current Issues in Economics and Finance* Vol. 1 (1995).
- [11] Stekler, H. and Talwar, R. “Forecasting the downturn of the great recession”. *Business Economics* Vol. 48 (2013), pp. 113–120.
- [12] Wickens, M. “How useful are DSGE macroeconomic models for forecasting?”. *Open Economies Review* Vol. 25 (2014), pp. 171–193.

## APPENDIX A. The full conditional distributions of BS model

The parameter vectors for the joint probability distributions and the full conditional distributions for BS model are denoted as follows:

$$\begin{aligned}
\boldsymbol{\phi} &= (\phi_1, \dots, \phi_p)' \\
\boldsymbol{\theta} &= (\theta_1, \dots, \theta_q)' \\
\boldsymbol{\beta} &= (\beta_1, \dots, \beta_r)' \\
\boldsymbol{\delta} &= (\delta_1, \dots, \delta_T)' \\
\boldsymbol{\gamma} &= (\gamma_1, \dots, \gamma_T)'
\end{aligned}$$

The joint probability distribution of BS model:

$$\begin{aligned}
& p(\boldsymbol{\Theta}|\mathbf{y}, \mathbf{x}) \\
&= p(\boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{\phi}, \boldsymbol{\theta}, \boldsymbol{\beta}, \zeta, \eta, \sigma, \tau|\mathbf{y}, \mathbf{x}) \\
&\propto p(\mathbf{y}|\mathbf{x}, \boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{\phi}, \boldsymbol{\theta}, \boldsymbol{\beta}) \times p(\boldsymbol{\gamma}|\eta) \times p(\boldsymbol{\delta}|\zeta, \tau) \times p(\boldsymbol{\phi}) \times \\
&\quad p(\boldsymbol{\theta}) \times p(\boldsymbol{\beta}) \times p(\zeta) \times p(\eta) \times p(\sigma) \times p(\tau) \\
&\propto \left\{ \prod_{t=m}^T (\sigma^2)^{-\frac{1}{2}} \exp \left[ \frac{-(y_t - \mu_t)^2}{2\sigma^2} \right] \right\} \times \\
&\quad \left\{ \prod_{t=m}^T \eta^{\gamma_t} (1 - \eta)^{(1-\gamma_t)} \right\} \times \left\{ \prod_{t=m}^T (\tau^2)^{-\frac{1}{2}} \exp \left[ \frac{-(\delta_t - \zeta)^2}{2\tau^2} \right] \right\} \times \\
&\quad \left\{ \prod_{i=1}^p (100^2)^{-\frac{1}{2}} \exp \left[ \frac{-(\phi_i)^2}{2 \times 100^2} \right] \right\} \times \left\{ \prod_{i=1}^q (100^2)^{-\frac{1}{2}} \exp \left[ \frac{-(\theta_i)^2}{2 \times 100^2} \right] \right\} \times \\
&\quad \left\{ \prod_{i=1}^r (100^2)^{-\frac{1}{2}} \exp \left[ \frac{-(\beta_i)^2}{2 \times 100^2} \right] \right\} \times (100^2)^{-\frac{1}{2}} \exp \left[ \frac{-(\zeta)^2}{2 \times 100^2} \right] \times \\
&\quad (\sigma^2)^{-0.0001-1} \exp \left[ -\frac{0.0001}{\sigma^2} \right] \times (\tau^2)^{-0.0001-1} \exp \left[ -\frac{0.0001}{\tau^2} \right],
\end{aligned}$$

where

$$\begin{aligned}
m &= \max(p, q, r) + 1 \text{ and} \\
\mu_t &= c_t + d_t \text{ with} \\
c_t &= (1 - \gamma_t)c_{t-1} + \gamma_t\delta_t, \\
d_t &= \sum_{i=1}^p \phi_i y_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \sum_{i=1}^r \beta_i x_{t-i}
\end{aligned}$$

The full conditional distributions BS model, derived from the joint probability function:

$$\begin{aligned}
\gamma_t | \Theta_{-\gamma_t}, \mathbf{y}, \mathbf{x} &\sim \text{Bern} \left( \frac{\eta \times \exp \left[ -\frac{a}{2\sigma^2} \right]}{\eta \times \exp \left[ -\frac{a}{2\sigma^2} \right] + (1 - \eta) \times \exp \left[ -\frac{b}{2\sigma^2} \right]} \right), \text{ where} \\
a &= \left( \sum_{i=t+1}^T (y_i - d_i - \delta_t)^2 \prod_{j=t+1}^i (1 - \gamma_j) \right) + (y_t - d_t - \delta_t)^2, \\
b &= \left( \sum_{i=t+1}^T (y_i - d_i - c_{t-1})^2 \prod_{j=t+1}^i (1 - \gamma_j) \right) + (y_t - d_t - c_{t-1})^2 \\
\delta_t | \Theta_{-\delta_t}, \mathbf{y}, \mathbf{x} &\sim N \left( \frac{a\gamma_t}{\frac{\sigma^2}{\tau^2} + b\gamma_t}, \frac{\sigma^2}{\left(\frac{\sigma^2}{\tau^2} + b\gamma_t\right)} \right), \text{ where} \\
a &= \left\{ \left[ \sum_{i=t+1}^T (y_i - d_i) \prod_{j=t+1}^i (1 - \gamma_j) \right] + (y_t - d_t) \right\} \text{ and} \\
b &= \left\{ \left[ \sum_{i=t+1}^T \prod_{j=t+1}^i (1 - \gamma_j) \right] + 1 \right\} \\
\phi_i | \Theta_{-\phi_i}, \mathbf{y}, \mathbf{x} &\sim N \left( \frac{\sum_{t=m}^T y_{t-i} (y_t - \tilde{\mu}_t)}{\frac{\sigma^2}{100^2} + \sum_{t=m}^T y_{t-i}^2}, \frac{1}{\frac{1}{100^2} + \frac{\sum_{t=m}^T y_{t-i}^2}{\sigma^2}} \right), \text{ where} \\
\tilde{\mu}_t &= \mu_t - \phi_i y_{t-i}, \\
&\text{for } i = 1, \dots, p \\
\theta_i | \Theta_{-\theta_i}, \mathbf{y}, \mathbf{x} &\sim N \left( \frac{\sum_{t=m}^T \varepsilon_{t-i} (y_t - \tilde{\mu}_t)}{\frac{\sigma^2}{100^2} + \sum_{t=m}^T \varepsilon_{t-i}^2}, \frac{1}{\frac{1}{100^2} + \frac{\sum_{t=m}^T \varepsilon_{t-i}^2}{\sigma^2}} \right), \text{ where} \\
\tilde{\mu}_t &= \mu_t - \theta_i \varepsilon_{t-i}, \\
&\text{for } i = 1, \dots, q \\
\beta_i | \Theta_{-\beta_i}, \mathbf{y}, \mathbf{x} &\sim N \left( \frac{\sum_{t=m}^T x_{t-i} (y_t - \tilde{\mu}_t)}{\frac{\sigma^2}{100^2} + \sum_{t=m}^T x_{t-i}^2}, \frac{1}{\frac{1}{100^2} + \frac{\sum_{t=m}^T x_{t-i}^2}{\sigma^2}} \right), \text{ where} \\
\tilde{\mu}_t &= \mu_t - \beta_i x_{t-i}, \\
&\text{for } i = 1, \dots, r
\end{aligned}$$



$$\begin{aligned}
\eta | \Theta_{-\eta}, \mathbf{y}, \mathbf{x} &\sim \text{Beta} \left( \sum_{t=m}^T \gamma_t + 1, T - m + 1 - \sum_{t=m}^T \gamma_t + 1 \right) \\
\zeta | \Theta_{-\zeta}, \mathbf{y}, \mathbf{x} &\sim \text{N} \left( \frac{\sum_{t=m}^T \delta_t}{T - m + 1 + \frac{\tau^2}{100^2}}, \frac{1}{\frac{1}{100^2} + \frac{T-m+1}{\tau^2}} \right) \\
\sigma^2 | \Theta_{-\sigma^2}, \mathbf{y}, \mathbf{x} &\sim \text{Inv-Gamma}(\text{shape} = s, \text{scale} = r), \text{ where} \\
&\quad s = \frac{T - m + 1}{2} + 0.0001 \text{ and} \\
&\quad r = \frac{1}{2} \sum_{t=m}^T (y_t - \mu_t)^2 + 0.0001 \\
\tau^2 | \Theta_{-\tau^2}, \mathbf{y}, \mathbf{x} &\sim \text{Inv-Gamma}(\text{shape} = s, \text{scale} = r), \text{ where} \\
&\quad s = \frac{T - m + 1}{2} + 0.0001 \text{ and } r = \frac{1}{2} \sum_{t=m}^T (\delta_t - \zeta)^2 + 0.0001
\end{aligned}$$

## APPENDIX B. The full conditional distributions of MuB model

The parameter vectors for the joint probability distributions and the full conditional distributions for MuB model are denoted as follows:

$$\begin{aligned}
\boldsymbol{\sigma} &= (\sigma_1, \dots, \sigma_N)' \\
\mathbf{c} &= (c_1, \dots, c_N)' \\
\boldsymbol{\phi} &= (\phi_{1,1}, \dots, \phi_{p,N})' \\
\boldsymbol{\theta} &= (\theta_{1,1}, \dots, \theta_{q,N})' \\
\boldsymbol{\beta} &= (\beta_{1,1}, \dots, \beta_{r,N})' \\
\boldsymbol{\lambda}_\phi &= (\lambda_{\phi,1}, \dots, \lambda_{\phi,p})' \\
\boldsymbol{\lambda}_\theta &= (\lambda_{\theta,1}, \dots, \lambda_{\theta,q})' \\
\boldsymbol{\lambda}_\beta &= (\lambda_{\beta,1}, \dots, \lambda_{\beta,r})' \\
\boldsymbol{\psi}_\phi &= (\psi_{\phi,1}, \dots, \psi_{\phi,p})' \\
\boldsymbol{\psi}_\theta &= (\psi_{\theta,1}, \dots, \psi_{\theta,q})' \\
\boldsymbol{\psi}_\beta &= (\psi_{\beta,1}, \dots, \psi_{\beta,r})'
\end{aligned}$$

The joint probability distribution of MuB model:

$$\begin{aligned}
& p(\boldsymbol{\Theta} | \mathbf{y}, \mathbf{x}) \\
&= p(\mathbf{c}, \boldsymbol{\phi}, \boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\lambda}_\phi, \boldsymbol{\lambda}_\theta, \boldsymbol{\lambda}_\beta, \boldsymbol{\psi}_\phi, \boldsymbol{\psi}_\theta, \boldsymbol{\psi}_\beta, \boldsymbol{\sigma}, \lambda_c, \psi_c | \mathbf{y}, \mathbf{x}) \\
&\propto p(\mathbf{y} | \mathbf{x}, \mathbf{c}, \boldsymbol{\phi}, \boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\sigma}) \times p(\mathbf{c} | \lambda_c, \psi_c) \times p(\boldsymbol{\phi} | \boldsymbol{\lambda}_\phi, \boldsymbol{\psi}_\phi) \times p(\boldsymbol{\theta} | \boldsymbol{\lambda}_\theta, \boldsymbol{\psi}_\theta) \times \\
&\quad p(\boldsymbol{\beta} | \boldsymbol{\lambda}_\beta, \boldsymbol{\psi}_\beta) \times p(\boldsymbol{\sigma}) \times p(\boldsymbol{\lambda}_\phi) \times p(\boldsymbol{\lambda}_\theta) \times p(\boldsymbol{\lambda}_\beta) \times p(\boldsymbol{\psi}_\phi) \times p(\boldsymbol{\psi}_\theta) \times \\
&\quad p(\boldsymbol{\psi}_\beta) \times p(\lambda_c) \times p(\psi_c) \\
&\propto \left\{ \prod_{t=m}^T \prod_{n=1}^N (\sigma_n^2)^{-\frac{1}{2}} \exp \left[ \frac{-(y_{t,n} - \mu_{t,n})^2}{2\sigma_n^2} \right] \right\} \times \\
&\quad \left\{ \prod_{n=1}^N (\psi_c^2)^{-\frac{1}{2}} \exp \left[ \frac{-(c_n - \lambda_c)^2}{2\psi_c^2} \right] \right\} \times \\
&\quad \left\{ \prod_{i=1}^p \prod_{n=1}^N (\psi_{\phi,i}^2)^{-\frac{1}{2}} \exp \left[ \frac{-(\phi_{i,n} - \lambda_{\phi,i})^2}{2\psi_{\phi,i}^2} \right] \right\} \times \\
&\quad \left\{ \prod_{i=1}^q \prod_{n=1}^N (\psi_{\theta,i}^2)^{-\frac{1}{2}} \exp \left[ \frac{-(\theta_{i,n} - \lambda_{\theta,i})^2}{2\psi_{\theta,i}^2} \right] \right\} \times
\end{aligned}$$

$$\begin{aligned}
& \left\{ \prod_{i=1}^r \prod_{n=1}^N (\psi_{\beta,i}^2)^{-\frac{1}{2}} \exp \left[ -\frac{(\beta_{i,n} - \lambda_{\beta,i})^2}{2\psi_{\beta,i}^2} \right] \right\} \times \\
& \left\{ \prod_{n=1}^N (\sigma_n^2)^{-0.0001-1} \exp \left[ -\frac{0.0001}{\sigma_n^2} \right] \right\} \times \\
& \left\{ \prod_{i=1}^p \exp \left[ \frac{-(\lambda_{\phi,i})^2}{2 \times 100^2} \right] \right\} \times \left\{ \prod_{i=1}^q \exp \left[ \frac{-(\lambda_{\theta,i})^2}{2 \times 100^2} \right] \right\} \times \\
& \left\{ \prod_{i=1}^r \exp \left[ \frac{-(\lambda_{\beta,i})^2}{2 \times 100^2} \right] \right\} \times \\
& \left\{ \prod_{i=1}^p (\psi_{\phi,i}^2)^{-0.0001-1} \exp \left[ -\frac{0.0001}{\psi_{\phi,i}^2} \right] \right\} \times \\
& \left\{ \prod_{i=1}^q (\psi_{\theta,i}^2)^{-0.0001-1} \exp \left[ -\frac{0.0001}{\psi_{\theta,i}^2} \right] \right\} \times \\
& \left\{ \prod_{i=1}^r (\psi_{\beta,i}^2)^{-0.0001-1} \exp \left[ -\frac{0.0001}{\psi_{\beta,i}^2} \right] \right\} \times \exp \left[ \frac{-(\lambda_c)^2}{2 \times 100^2} \right] \times \\
& (\psi_c^2)^{-0.0001-1} \exp \left[ -\frac{0.0001}{\psi_c^2} \right],
\end{aligned}$$

where

$$\begin{aligned}
m &= \max(p, q, r) + 1 \text{ and} \\
\mu_{t,n} &= c_n + \sum_{i=1}^p \phi_{i,n} y_{t-i,n} + \sum_{i=1}^q \theta_{i,n} \varepsilon_{t-i,n} + \sum_{i=1}^r \beta_{i,n} x_{t-i,n}
\end{aligned}$$

The full conditional distributions MuB model, derived from the joint probability function:

$$\begin{aligned}
c_n | \Theta_{-c_n}, \mathbf{y}, \mathbf{x} &\sim \text{N} \left( \frac{\sum_{t=m}^T (y_{t,n} - \tilde{\mu}_{t,n}) + \frac{\sigma_n^2}{\psi_c^2} \lambda_c}{T - m + 1 + \frac{\sigma_n^2}{\psi_c^2}}, \frac{1}{\frac{\sigma_n^2}{\psi_c^2} + \frac{1}{\psi_c^2}} \right), \\
&\text{where } \tilde{\mu}_{t,n} = \mu_{t,n} - c_n \\
\phi_{i,n} | \Theta_{-\phi_{i,n}}, \mathbf{y}, \mathbf{x} &\sim \text{N} \left( \frac{\sum_{t=m}^T (y_{t,n} y_{t-i,n} - \tilde{\mu}_{t,n} y_{t-i,n}) + \frac{\sigma_n^2}{\psi_{\phi,i}^2} \lambda_{\phi,i}}{\sum_{t=m}^T y_{t-i,n}^2 + \frac{\sigma_n^2}{\psi_{\phi,i}^2}}, \frac{1}{\frac{\sum_{t=m}^T y_{t-i,n}^2}{\sigma_n^2} + \frac{1}{\psi_{\phi,i}^2}} \right), \\
&\text{where } \tilde{\mu}_{t,n} = \mu_{t,n} - \phi_{i,n} y_{t-i,n} \\
&\text{for } i = 1, \dots, p \\
\theta_{i,n} | \Theta_{-\theta_{i,n}}, \mathbf{y}, \mathbf{x} &\sim \text{N} \left( \frac{\sum_{t=m}^T (y_{t,n} \varepsilon_{t-i,n} - \tilde{\mu}_{t,n} \varepsilon_{t-i,n}) + \frac{\sigma_n^2}{\psi_{\theta,i}^2} \lambda_{\theta,i}}{\sum_{t=m}^T \varepsilon_{t-i,n}^2 + \frac{\sigma_n^2}{\psi_{\theta,i}^2}}, \frac{1}{\frac{\sum_{t=m}^T \varepsilon_{t-i,n}^2}{\sigma_n^2} + \frac{1}{\psi_{\theta,i}^2}} \right), \\
&\text{where } \tilde{\mu}_{t,n} = \mu_{t,n} - \theta_{i,n} \varepsilon_{t-i,n}
\end{aligned}$$

$$\beta_{i,n} | \Theta_{-\beta_{n,i}}, \mathbf{y}, \mathbf{x} \sim \text{N} \left( \frac{\sum_{t=m}^T (y_{t,n} x_{t-i,n} - \tilde{\mu}_{t,n} x_{t-i,n}) + \frac{\sigma_n^2}{\psi_{\beta,i}^2} \lambda_{\beta,i}}{\sum_{t=m}^T x_{t-i,n}^2 + \frac{\sigma_n^2}{\psi_{\beta,i}^2}}, \frac{1}{\frac{\sum_{t=m}^T x_{t-i,n}^2}{\sigma_n^2} + \frac{1}{\psi_{\beta,i}^2}} \right),$$

for  $i = 1, \dots, q$

where  $\tilde{\mu}_{t,n} = \mu_{t,n} - \beta_{i,n} x_{t-i,n}$

for  $i = 1, \dots, r$

$$\sigma_n^2 | \Theta_{-\sigma_n^2}, \mathbf{y}, \mathbf{x} \sim \text{Inv-Gamma}(\text{shape} = s, \text{scale} = r), \text{ where}$$

$$s = \frac{T - m + 1}{2} + 0.0001 \text{ and}$$

$$r = \frac{1}{2} \sum_{t=m}^T (y_{t,n} - \mu_{t,n})^2 + 0.0001$$

$$\lambda_{\phi,i} | \Theta_{-\lambda_{\phi,i}}, \mathbf{y}, \mathbf{x} \sim \text{N} \left( \frac{\sum_{n=1}^N \phi_{i,n}}{N + \frac{\psi_{\phi,i}^2}{100^2}}, \frac{1}{\frac{N}{\psi_{\phi,i}^2} + \frac{1}{100^2}} \right)$$

for  $i = 1, \dots, p$

$$\lambda_{\theta,i} | \Theta_{-\lambda_{\theta,i}}, \mathbf{y}, \mathbf{x} \sim \text{N} \left( \frac{\sum_{n=1}^N \theta_{i,n}}{N + \frac{\psi_{\theta,i}^2}{100^2}}, \frac{1}{\frac{N}{\psi_{\theta,i}^2} + \frac{1}{100^2}} \right)$$

for  $i = 1, \dots, q$

$$\lambda_{\beta,i} | \Theta_{-\lambda_{\beta,i}}, \mathbf{y}, \mathbf{x} \sim \text{N} \left( \frac{\sum_{n=1}^N \beta_{i,n}}{N + \frac{\psi_{\beta,i}^2}{100^2}}, \frac{1}{\frac{N}{\psi_{\beta,i}^2} + \frac{1}{100^2}} \right)$$

for  $i = 1, \dots, r$

$$\lambda_c | \Theta_{-\lambda_c}, \mathbf{y}, \mathbf{x} \sim \text{N} \left( \frac{\sum_{n=1}^N c_n}{N + \frac{\psi_c^2}{100^2}}, \frac{1}{\frac{N}{\psi_c^2} + \frac{1}{100^2}} \right)$$

$$\psi_{\phi,i}^2 | \Theta_{-\psi_{\phi,i}^2}, \mathbf{y}, \mathbf{x} \sim \text{Inv-Gamma}(\text{shape} = s, \text{scale} = r), \text{ where}$$

$$s = \frac{N}{2} + 0.0001 \text{ and } r = \frac{1}{2} \sum_{n=1}^N (\phi_{i,n} - \lambda_{\phi,i})^2 + 0.0001$$

for  $i = 1, \dots, p$

$$\psi_{\theta,i}^2 | \Theta_{-\psi_{\theta,i}^2}, \mathbf{y}, \mathbf{x} \sim \text{Inv-Gamma}(\text{shape} = s, \text{scale} = r), \text{ where}$$

$$s = \frac{N}{2} + 0.0001 \text{ and } r = \frac{1}{2} \sum_{n=1}^N (\theta_{i,n} - \lambda_{\theta,i})^2 + 0.0001$$

for  $i = 1, \dots, q$

$$\psi_{\beta,i}^2 | \Theta_{-\psi_{\beta,i}^2}, \mathbf{y}, \mathbf{x} \sim \text{Inv-Gamma}(\text{shape} = s, \text{scale} = r), \text{ where}$$

$$s = \frac{N}{2} + 0.0001 \text{ and } r = \frac{1}{2} \sum_{n=1}^N (\beta_{i,n} - \lambda_{\beta,i})^2 + 0.0001$$

for  $i = 1, \dots, r$

$$\psi_c^2 | \Theta_{-\psi_c^2}, \mathbf{y}, \mathbf{x} \sim \text{Inv-Gamma}(\text{shape} = s, \text{scale} = r), \text{ where}$$

$$s = \frac{N}{2} + 0.0001 \text{ and } r = \frac{1}{2} \sum_{n=1}^N (c_n - \lambda_c)^2 + 0.0001$$

## APPENDIX C. The full conditional distributions of MuBS model

The parameter vectors for the joint probability distributions and the full conditional distributions of MuBS model are denoted as follows:

$$\begin{aligned}
\boldsymbol{\tau} &= (\tau_1, \dots, \tau_N)' \\
\boldsymbol{\sigma} &= (\sigma_1, \dots, \sigma_N)' \\
\boldsymbol{\zeta} &= (\zeta_1, \dots, \zeta_N)' \\
\boldsymbol{\eta} &= (\eta_1, \dots, \eta_N)' \\
\boldsymbol{\phi} &= (\phi_{1,1}, \dots, \phi_{p,N})' \\
\boldsymbol{\theta} &= (\theta_{1,1}, \dots, \theta_{q,N})' \\
\boldsymbol{\beta} &= (\beta_{1,1}, \dots, \beta_{r,N})' \\
\boldsymbol{\delta} &= (\delta_{1,1}, \dots, \delta_{T,N})' \\
\boldsymbol{\gamma} &= (\gamma_{1,1}, \dots, \gamma_{T,N})' \\
\boldsymbol{\lambda}_\phi &= (\lambda_{\phi,1}, \dots, \lambda_{\phi,p})' \\
\boldsymbol{\lambda}_\theta &= (\lambda_{\theta,1}, \dots, \lambda_{\theta,q})' \\
\boldsymbol{\lambda}_\beta &= (\lambda_{\beta,1}, \dots, \lambda_{\beta,r})' \\
\boldsymbol{\psi}_\phi &= (\psi_{\phi,1}, \dots, \psi_{\phi,p})' \\
\boldsymbol{\psi}_\theta &= (\psi_{\theta,1}, \dots, \psi_{\theta,q})' \\
\boldsymbol{\psi}_\beta &= (\psi_{\beta,1}, \dots, \psi_{\beta,r})'
\end{aligned}$$

The joint probability distribution of MuBS model:

$$\begin{aligned}
& p(\boldsymbol{\Theta} | \mathbf{y}, \mathbf{x}) \\
&= p(\omega, \boldsymbol{\psi}_\phi, \boldsymbol{\psi}_\theta, \boldsymbol{\psi}_\beta, \boldsymbol{\lambda}_\phi, \boldsymbol{\lambda}_\theta, \boldsymbol{\lambda}_\beta, \boldsymbol{\tau}, \boldsymbol{\sigma}, \boldsymbol{\zeta}, \boldsymbol{\eta}, \boldsymbol{\phi}, \boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\delta}, \boldsymbol{\gamma} | \mathbf{y}, \mathbf{x}) \\
&\propto p(\mathbf{y} | \mathbf{x}, \boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{\phi}, \boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\sigma}) \times p(\boldsymbol{\gamma} | \boldsymbol{\eta}) \times p(\boldsymbol{\delta} | \boldsymbol{\zeta}, \boldsymbol{\tau}) \times p(\boldsymbol{\phi} | \boldsymbol{\lambda}_\phi, \boldsymbol{\psi}_\phi) \times \\
&\quad p(\boldsymbol{\theta} | \boldsymbol{\lambda}_\theta, \boldsymbol{\psi}_\theta) \times p(\boldsymbol{\beta} | \boldsymbol{\lambda}_\beta, \boldsymbol{\psi}_\beta) \times p(\boldsymbol{\eta}) \times p(\boldsymbol{\zeta} | \omega) \times p(\boldsymbol{\sigma}) \times p(\boldsymbol{\tau}) \times \\
&\quad p(\boldsymbol{\lambda}_\phi) \times p(\boldsymbol{\lambda}_\theta) \times p(\boldsymbol{\lambda}_\beta) \times p(\boldsymbol{\psi}_\phi) \times p(\boldsymbol{\psi}_\theta) \times p(\boldsymbol{\psi}_\beta) \times p(\omega) \\
&\propto \left\{ \prod_{t=m}^T \prod_{n=1}^N (\sigma_n^2)^{-\frac{1}{2}} \exp \left[ \frac{-(y_{t,n} - \mu_{t,n})^2}{2\sigma_n^2} \right] \right\} \times \\
&\quad \left\{ \prod_{t=m}^T \prod_{n=1}^N \eta_n^{\gamma_{t,n}} (1 - \eta_n)^{(1 - \gamma_{t,n})} \right\} \times \left\{ \prod_{t=m}^T \prod_{n=1}^N (\tau_n^2)^{-\frac{1}{2}} \exp \left[ \frac{-(\delta_{t,n} - \zeta_n)^2}{2\tau_n^2} \right] \right\} \times
\end{aligned}$$

$$\begin{aligned}
& \left\{ \prod_{i=1}^p \prod_{n=1}^N (\psi_{\phi,i}^2)^{-\frac{1}{2}} \exp \left[ \frac{-(\phi_{i,n} - \lambda_{\phi,i})^2}{2\psi_{\phi,i}^2} \right] \right\} \times \\
& \left\{ \prod_{i=1}^q \prod_{n=1}^N (\psi_{\theta,i}^2)^{-\frac{1}{2}} \exp \left[ \frac{-(\theta_{i,n} - \lambda_{\theta,i})^2}{2\psi_{\theta,i}^2} \right] \right\} \times \\
& \left\{ \prod_{i=1}^r \prod_{n=1}^N (\psi_{\beta,i}^2)^{-\frac{1}{2}} \exp \left[ \frac{-(\beta_{i,n} - \lambda_{\beta,i})^2}{2\psi_{\beta,i}^2} \right] \right\} \times \\
& \left\{ \prod_{n=1}^N (\omega^2)^{-\frac{1}{2}} \exp \left[ \frac{-(\zeta_n)^2}{2\omega^2} \right] \right\} \times \left\{ \prod_{n=1}^N (\sigma_n^2)^{-0.0001-1} \exp \left[ -\frac{0.0001}{\sigma_n^2} \right] \right\} \times \\
& \left\{ \prod_{n=1}^N (\tau_n^2)^{-0.0001-1} \exp \left[ -\frac{0.0001}{\tau_n^2} \right] \right\} \times \left\{ \prod_{i=1}^p \exp \left[ \frac{-(\lambda_{\phi,i})^2}{2 \times 100^2} \right] \right\} \times \\
& \left\{ \prod_{i=1}^q \exp \left[ \frac{-(\lambda_{\theta,i})^2}{2 \times 100^2} \right] \right\} \times \left\{ \prod_{i=1}^r \exp \left[ \frac{-(\lambda_{\beta,i})^2}{2 \times 100^2} \right] \right\} \times \\
& \left\{ \prod_{i=1}^p (\psi_{\phi,i}^2)^{-0.0001-1} \exp \left[ -\frac{0.0001}{\psi_{\phi,i}^2} \right] \right\} \times \\
& \left\{ \prod_{i=1}^q (\psi_{\theta,i}^2)^{-0.0001-1} \exp \left[ -\frac{0.0001}{\psi_{\theta,i}^2} \right] \right\} \times \\
& \left\{ \prod_{i=1}^r (\psi_{\beta,i}^2)^{-0.0001-1} \exp \left[ -\frac{0.0001}{\psi_{\beta,i}^2} \right] \right\} \times \\
& (\omega^2)^{-0.0001-1} \exp \left[ -\frac{0.0001}{\omega^2} \right],
\end{aligned}$$

where

$$\begin{aligned}
m &= \max(p, q, r) + 1 \text{ and} \\
\mu_{t,n} &= c_{t,n} + d_{t,n} \text{ with} \\
c_{t,n} &= (1 - \gamma_{t,n})c_{t-1,n} + \gamma_{t,n}\delta_{t,n} \text{ and} \\
d_{t,n} &= \sum_{i=1}^p \phi_{i,n}y_{t-i,n} + \sum_{i=1}^q \theta_{i,n}\varepsilon_{t-i,n} + \sum_{i=1}^r \beta_{i,n}x_{t-i,n}
\end{aligned}$$

The full conditional distributions MuBS model, derived from the joint probability function:

$$\begin{aligned}
\gamma_{t,n} | \Theta_{-\gamma_{t,n}}, \mathbf{y}, \mathbf{x} &\sim \text{Bern} \left( \frac{\eta_n \times \exp \left[ -\frac{a}{2\sigma_n^2} \right]}{\eta_n \times \exp \left[ -\frac{a}{2\sigma_n^2} \right] + (1 - \eta_n) \times \exp \left[ -\frac{b}{2\sigma_n^2} \right]} \right), \text{ where} \\
a &= \left( \sum_{i=t+1}^T (y_{i,n} - d_{i,n} - \delta_{t,n})^2 \prod_{j=t+1}^i (1 - \gamma_{j,n}) \right) + (y_{t,n} - d_{t,n} - \delta_{t,n})^2,
\end{aligned}$$

$$\begin{aligned}
b &= \left( \sum_{i=t+1}^T (y_{i,n} - d_{i,n} - c_{t-1,n})^2 \prod_{j=t+1}^i (1 - \gamma_{j,n}) \right) + \\
&\quad (y_{t,n} - d_{t,n} - c_{t-1,n})^2 \\
\delta_{t,n} | \Theta_{-\delta_{t,n}}, \mathbf{y}, \mathbf{x} &\sim N \left( \frac{a\gamma_{t,n}}{\frac{\sigma_n^2}{\tau_n^2} + b\gamma_{t,n}}, \frac{\sigma_n^2}{\left(\frac{\sigma_n^2}{\tau_n^2} + b\gamma_{t,n}\right)} \right), \text{ where} \\
a &= \left\{ \left[ \sum_{i=t+1}^T (y_{i,n} - d_{i,n}) \prod_{j=t+1}^i (1 - \gamma_{j,n}) \right] + (y_{t,n} - d_{t,n}) \right\} \text{ and} \\
b &= \left\{ \left[ \sum_{i=t+1}^T \prod_{j=t+1}^i (1 - \gamma_{j,n}) \right] + 1 \right\} \\
\phi_{i,n} | \Theta_{-\phi_{i,n}}, \mathbf{y}, \mathbf{x} &\sim N \left( \frac{\sum_{t=m}^T (y_{t,n} y_{t-i,n} - \tilde{\mu}_{t,n} y_{t-i,n}) + \frac{\sigma_n^2}{\psi_{\phi,i}^2} \lambda_{\phi,i}}{\sum_{t=m}^T y_{t-i,n}^2 + \frac{\sigma_n^2}{\psi_{\phi,i}^2}}, \frac{1}{\frac{\sum_{t=m}^T y_{t-i,n}^2}{\sigma_n^2} + \frac{1}{\psi_{\phi,i}^2}} \right), \\
&\quad \text{where } \tilde{\mu}_{t,n} = \mu_{t,n} - \phi_{i,n} y_{t-i,n} \\
&\quad \text{for } i = 1, \dots, p \\
\theta_{i,n} | \Theta_{-\theta_{i,n}}, \mathbf{y}, \mathbf{x} &\sim N \left( \frac{\sum_{t=m}^T (y_{t,n} \varepsilon_{t-i,n} - \tilde{\mu}_{t,n} \varepsilon_{t-i,n}) + \frac{\sigma_n^2}{\psi_{\theta,i}^2} \lambda_{\theta,i}}{\sum_{t=m}^T \varepsilon_{t-i,n}^2 + \frac{\sigma_n^2}{\psi_{\theta,i}^2}}, \frac{1}{\frac{\sum_{t=m}^T \varepsilon_{t-i,n}^2}{\sigma_n^2} + \frac{1}{\psi_{\theta,i}^2}} \right), \\
&\quad \text{where } \tilde{\mu}_{t,n} = \mu_{t,n} - \theta_{i,n} \varepsilon_{t-i,n} \\
&\quad \text{for } i = 1, \dots, q \\
\beta_{i,n} | \Theta_{-\beta_{i,n}}, \mathbf{y}, \mathbf{x} &\sim N \left( \frac{\sum_{t=m}^T (y_{t,n} x_{t-i,n} - \tilde{\mu}_{t,n} x_{t-i,n}) + \frac{\sigma_n^2}{\psi_{\beta,i}^2} \lambda_{\beta,i}}{\sum_{t=m}^T x_{t-i,n}^2 + \frac{\sigma_n^2}{\psi_{\beta,i}^2}}, \frac{1}{\frac{\sum_{t=m}^T x_{t-i,n}^2}{\sigma_n^2} + \frac{1}{\psi_{\beta,i}^2}} \right), \\
&\quad \text{where } \tilde{\mu}_{t,n} = \mu_{t,n} - \beta_{i,n} x_{t-i,n} \\
&\quad \text{for } i = 1, \dots, r \\
\sigma_n^2 | \Theta_{-\sigma_n^2}, \mathbf{y}, \mathbf{x} &\sim \text{Inv-Gamma}(\text{shape} = s, \text{scale} = r), \text{ where} \\
s &= \frac{T - m + 1}{2} + 0.0001 \text{ and} \\
r &= \frac{1}{2} \sum_{t=m}^T (y_{t,n} - \mu_{t,n})^2 + 0.0001 \\
\eta_n | \Theta_{-\eta_n}, \mathbf{y}, \mathbf{x} &\sim \text{Beta} \left( \sum_{t=m}^T \gamma_{t,n} + 1, T - m + 1 - \sum_{t=m}^T \gamma_{t,n} + 1 \right) \\
\zeta_n | \Theta_{-\zeta_n}, \mathbf{y}, \mathbf{x} &\sim N \left( \frac{\sum_{t=m}^T \delta_{t,n}}{T - m + 1 + \frac{\tau_n^2}{\omega^2}}, \frac{1}{\frac{T - m + 1}{\tau_n^2} + \frac{1}{\omega^2}} \right) \\
\tau_n^2 | \Theta_{-\tau_n^2}, \mathbf{y}, \mathbf{x} &\sim \text{Inv-Gamma}(\text{shape} = s, \text{scale} = r), \text{ where} \\
s &= \frac{T - m + 1}{2} + 0.0001 \text{ and } r = \frac{1}{2} \sum_{t=m}^T (\delta_{t,n} - \zeta_n)^2 + 0.0001
\end{aligned}$$



$$\lambda_{\phi,i} | \Theta_{-\lambda_{\phi,i}}, \mathbf{y}, \mathbf{x} \sim \text{N} \left( \frac{\sum_{n=1}^N \phi_{i,n}}{N + \frac{\psi_{\phi,i}^2}{100^2}}, \frac{1}{\frac{N}{\psi_{\phi,i}^2} + \frac{1}{100^2}} \right)$$

for  $i = 1, \dots, p$

$$\lambda_{\theta,i} | \Theta_{-\lambda_{\theta,i}}, \mathbf{y}, \mathbf{x} \sim \text{N} \left( \frac{\sum_{n=1}^N \theta_{i,n}}{N + \frac{\psi_{\theta,i}^2}{100^2}}, \frac{1}{\frac{N}{\psi_{\theta,i}^2} + \frac{1}{100^2}} \right)$$

for  $i = 1, \dots, q$

$$\lambda_{\beta,i} | \Theta_{-\lambda_{\beta,i}}, \mathbf{y}, \mathbf{x} \sim \text{N} \left( \frac{\sum_{n=1}^N \beta_{i,n}}{N + \frac{\psi_{\beta,i}^2}{100^2}}, \frac{1}{\frac{N}{\psi_{\beta,i}^2} + \frac{1}{100^2}} \right)$$

for  $i = 1, \dots, r$

$$\psi_{\phi,i}^2 | \Theta_{-\psi_{\phi,i}^2}, \mathbf{y}, \mathbf{x} \sim \text{Inv-Gamma}(\text{shape} = s, \text{scale} = r), \text{ where}$$

$$s = \frac{N}{2} + 0.0001 \text{ and } r = \frac{1}{2} \sum_{n=1}^N (\phi_{i,n} - \lambda_{\phi,i})^2 + 0.0001$$

for  $i = 1, \dots, p$

$$\psi_{\theta,i}^2 | \Theta_{-\psi_{\theta,i}^2}, \mathbf{y}, \mathbf{x} \sim \text{Inv-Gamma}(\text{shape} = s, \text{scale} = r), \text{ where}$$

$$s = \frac{N}{2} + 0.0001 \text{ and } r = \frac{1}{2} \sum_{n=1}^N (\theta_{i,n} - \lambda_{\theta,i})^2 + 0.0001$$

for  $i = 1, \dots, q$

$$\psi_{\beta,i}^2 | \Theta_{-\psi_{\beta,i}^2}, \mathbf{y}, \mathbf{x} \sim \text{Inv-Gamma}(\text{shape} = s, \text{scale} = r), \text{ where}$$

$$s = \frac{N}{2} + 0.0001 \text{ and } r = \frac{1}{2} \sum_{n=1}^N (\beta_{i,n} - \lambda_{\beta,i})^2 + 0.0001$$

for  $i = 1, \dots, r$

$$\omega^2 | \Theta_{-\omega^2}, \mathbf{y}, \mathbf{x} \sim \text{Inv-Gamma}(\text{shape} = s, \text{scale} = r), \text{ where}$$

$$s = \frac{N}{2} + 0.0001 \text{ and } r = \frac{1}{2} \sum_{n=1}^N (\zeta_n)^2 + 0.0001$$