Forecasting U.S. Unemployment Rates Under a Bayesian Framework and Pseudo-Bayesian Small Area Estimation

by

Jiacheng Li

(Under the Direction of Gauri Sankar Datta & T.N. Sriram)

Abstract

This dissertation focuses on two different studies that are motivated by real-life applications. A fully Bayesian approach is proposed to carry out each of the studies and then appropriate comparisons are made to establish the superiority of the proposed methods.

The first study focuses on providing reliable forecasts of the U.S. civilian unemployment rate. To obtain reliable forecasts, we group the states based on similarities in the time series structure or based on the region in which they are located and develop a fully Bayesian framework that borrows strength across states and uses the unemployment insurance claims data as a covariate. The Metropolis-Hastings algorithm and the Gibbs sampling method are applied to generate the posterior distributions of the relevant model parameters. Grouping states according to their similarities enables us to utilize small-area estimation to generate more accurate forecasts of UE rates for states within the same group. In terms of Mean Absolute Prediction Error (MAPE), the Bayesian framework provides more accurate forecasts than a recently developed frequentist framework called **Bigtime**.

In the second study, we develop a Pseudo-Bayesian small area estimation as an alternative to the empirical best linear unbiased prediction method. The empirical best linear unbiased prediction method has dominated the frequentist model-based approach to small-area estimation. This method estimates model parameters based on the marginal distribution of the data. As an alternative to this method, the observed best prediction (OBP) method estimates the parameters by minimizing an objective function that is determined by the total mean squared prediction error. Using this objective function in the well-known Fay-Herriot model, we develop a pseudo-posterior distribution for the model parameters under nearly non-informative priors. Real hospital data, median incomes of four-person households by state produced by the U.S. Census Bureau, and simulation studies show that the pseudo-Bayesian estimators (PBEs) compete favorably with the OBPs and empirical best linear unbiased predictions (EBLUPs). The PBE estimates are robust to mean misspecification and provide excellent frequentist properties. Being Bayesian by construction, they automatically avoid negative estimates of standard errors, enjoy a dual justification, and provide a desirable alternative to practitioners.

Index words: Time series; Bayesian inference; Metropolis Hasting; Gibbs sampling method; Empirical best linear unbiased prediction; Fay-Herriot model; Observed best prediction

Forecasting U.S. Unemployment Rates Under a Bayesian Framework and Pseudo-Bayesian Small Area Estimation

by

Jiacheng Li

B.E., Huazhong University of Science and Technology, 2015

A Dissertation Submitted to the Graduate Faculty of the
University of Georgia in Partial Fulfillment of the Requirements for the Degree.

Doctor of Philosophy

Athens, Georgia

2023

Forecasting U.S. Unemployment Rates Under a Bayesian Framework and Pseudo-Bayesian Small Area Estimation

by

Jiacheng Li

Major Professor: Gauri Sankar Datta
T. N. Sriram

Committee: Liang Liu
Ray (Shuyang) Bai

Electronic Version Approved:

Ron Walcott
Dean of the Graduate School
The University of Georgia
May 2023

Dedication

To my parents Chaofa Li and Xueer Shi

Acknowledgments

I would like to start by thanking my parents, Chaofa Li and Xueer Shi, and my sister Xia Li, for their unwavering support and encouragement throughout my Ph.D. journey. From the very beginning, they have been my source of strength and inspiration. Without their love and dedication, I would not have been able to reach this milestone. I am deeply thankful to them for the sacrifices they have made to help me achieve my dreams.

I am also deeply grateful to my main advisor, Professor Gauri Datta, and my co-advisor, Professor T.N. Sriram, for their guidance, support, and invaluable advice throughout my research. Their expertise and knowledge have been instrumental in helping me shape my research and complete my dissertation. I am extremely grateful to them for their patience and generosity. I would also like to thank Professor Liang Liu and Professor Shuyang Bai for their insights and constructive feedback during my dissertation defense. Their comments have helped me improve my work and make it stronger. I am truly thankful for their valuable contributions.

Finally, I would like to thank my other family members and friends for their help and support along the way. Their encouragement and support have kept me going and I am thankful to them for being there for me.

Contents

List of Figures

List of Tables

Chapter 1

Forecasting U.S. Unemployment Rates Under a Bayesian Framework

1.1 Introduction

The United States (U.S.) civilian unemployment (UE) rate is one of the essential economic indicators that has been studied for many decades. It can be utilized for regional planning and the allocation of funds to states as part of several federal aid programs. In conjunction with other economic data, it plays a vital role in assisting policymakers determine when and what type of assistance should be provided to people impacted by unemployment. Therefore, it is of national importance to build statistical models for UE rates that improve their forecast at the state level.

Unemployment Insurance (UI) programs are administered at the state level and they provide assistance to those who are in search of employment. Although the UI claims data provide useful information, they do not directly give information on the total number of people unemployed because the claims data do not include all unemployed people. For example, self-employed workers and workers in certain not-for-profit organizations are not covered by the UI programs. Since the UI programs do not cover all unemployed, those who lose their jobs have to file applications to determine their eligibility for UI assistance. According to the Bureau of Labor Statistics (BLS), over the past decades, only about one-third of the total unemployed are covered by regular UI benefits. Even though the UI claims data have limitations in measuring the unemployment rate, they are still widely used to indicate labor market conditions. In this dissertation, we will use the UI claims data as an important covariate in predicting the UE rate.

It is possible to build statistical models for the UE rate using the class of Box-Jenkins models, see Box and Jenkins, 1976. For the U.S. UE rate data, Montgomery et al., 1998 conducted a comprehensive study and compared the performance of forecasts for a variety of linear and nonlinear time series models. Tiller, 1992 developed a method to increase the accuracy of the state-level UE rate estimates. Tiller, 1992 represented the observed CPS sample estimates $Y_{i,t}$ as $Y_{i,t} = \theta_{i,t} + \epsilon_{i,t}$ for $i = 1, \cdots, m$, and $t = 1, \cdots, T$, where $\theta_{i,t}$ is the true UE rate for domain i at time t, and $\epsilon_{i,t}$ is the sampling error. For each state i, the BLS modeled $\theta_{i,t}$'s using a structural time series with explanatory variables and the $\epsilon_{i,t}$ as an autoregressive-moving average (ARMA) model, where the time t denotes month. A limitation of the BLS model is that it does not utilize the information across states and does not provide uncertainty measures of the state estimates. In the literature, there are articles that recognize the importance of using Bayesian methods in forecasting UE rates. For example, Chen and Lee, 1995 presented a Bayesian analysis of the threshold autoregressive model with two regimes, and Datta et al., 1999 proposed a hierarchical Bayes method using a time series generalization of a widely used cross-sectional model in small-area estimation.

The Vector Autoregressive Moving Average (VARMA) model is a cornerstone of multivariate time series theory. VARMA models are known to provide a more parsimonious description of a linear time-invariant system than Vector Autoregressive (VAR) models, which is a special case of VARMA without moving average coefficients. As noted in Wilms et al., 2021, despite its advantages over VAR, VARMA has not been very popular among practitioners due to its computational and theoretical challenges in model identification and specification. Wilms et al., 2021 propose a new optimization-based approach to VARMA identification built upon the notion of parsimony. Among all equivalent data-generating models, they use convex optimization to seek the parameterization that is simplest in a certain sense. A user-specified strongly convex penalty is used to measure model simplicity, and that same penalty is then used to define an estimator that can be efficiently computed. They overcome the challenges in using VARMA by theoretically and empirically investigating the large-scale VARMA as a competitive alternative to the VAR.

Model selection and parameter estimation are some of the issues faced by VARMA models. Challenges of aggregating the model selection and parameter estimation are akin to the variable selection challenges in linear regression, where shrinkage methods (e.g., ridge, lasso, elastic net) have been successfully used in combining selection and parameter estimation. More specifically, Wilms et al., 2021 estimated and calculated the degree of parsimony of VARMA parameters using convex regularizers such as the L_1-norm, L_2-norm, the nuclear norm, etc. Due to the fact that the VARMA model incorporates latent lagged errors, they first approximate the unobservable errors and then estimate the VARMA using the approximated lagged errors. They implement their fully-automated VARMA identification and estimation procedure in the *R* package **bigtime**.

In this dissertation, for each of the 51 states in the U.S. (50 states plus District of Columbia), we consider the quarterly UE rate time series data spanning 29 years starting from the first quarter of 1990 to the fourth quarter of 2018. The goal is to produce better forecasts of UE rates for each of the 51 states by borrowing strength across states. With this goal in mind, we consider two possible approaches to grouping the 51 states. In the first approach, for each of the 51 states, we determine the best autoregressive integrated moving average (ARIMA) model for its UE rate series using the built-in R function **arima**. Then, we group the states based on the similarities in the ARIMA structures. This leads to dividing the 51 states into 5 distinct groups such that the ARIMA structures (of UE rates) for states within each group are similar; see Table 1.1. In the second approach, we group the 51 states simply based on the four regions–Northeast, Midwest, South, and West; see Table 1.2 The idea behind grouping the 51 states based on similarities in the ARIMA structures or similarities that may exist within each region is to produce better forecasts of UE rates for each state (within a group) by borrowing strength across other states in that group. This setup naturally enables us to use methodologies from small-area estimation combined with a Bayesian framework to make reliable forecasts of UE rates for each of the 51 states in the U.S.

For each state, we designate the UE rates from the first quarter of 1990 (1990:Q1) to the fourth quarter of 2017 (2017:Q4) as the *training data* and the UE rates for remaining four quarters of 2018 as the *test data*. First, we use Bayesian framework to build time series models for the *training data*, and then compute the forecasts of UE rates for the remaining four quarters in the *test data* based on the fitted models. Finally, using Mean Absolute Prediction Errors (MAPE), we compare the accuracy of the forecasts based on our models and those based on the frequentist method of Wilms et al., 2021, which used the **bigtime** R package.

The first part of the dissertation is structured as follows. In Section 1.2, we propose our model, which is a Bayesian inference methodology for ARIMA processes based on the Markov Chain Monte Carlo (MCMC) and Metropolis-Hastings methods. Detailed derivation and forecasts of the final four unused observations (the year 2018) for states belonging to the same group are also supplied (two scenarios, one by groups defined by similar time series structures and the other by regions designated by the Census Bureau). Regarding MAPE, we compare our forecasting results to those provided by the **Bigtime** R program. In Section 1.3, the results, summary, and conclusions are presented.

1.2 Bayesian Analysis of Time Series models

1.2.1 Literature Review

The U.S. civilian UE rate is the proportion of the labor force that is jobless. The BLS publishes this rate each month for the entire country and its many geographic and demographic sub-domains. For instance, the UE rate estimates are provided for all 51 states (50 states plus the District of Columbia), all metropolitan statistical regions, and all counties, cities, and towns with a population of at least 25,000. The unemployment rate is utilized in regional planning and the allocation of funds to states as part of several federal aid programs. [BLS, 2015]. This section develops methodologies for accurately forecasting the U.S. UE rates for each of the 51 states. Among the important economic data compiled by the BLS, the UE rates for states and local areas are regarded as essential indicators of regional economic conditions. The state workforce agencies produce these estimates as part of the Federal-State cooperative Local Area Unemployment Statistics (LAUS) program.

The statistical models used for developing statewide LAUS estimates have been replaced with time series regression models, featuring real-time benchmarking to monthly national Current Population Survey (CPS) employment totals. The models produce seasonally adjusted estimates within the estimation model, as well as non-seasonally adjusted estimates and measures of errors. To produce seasonally adjusted estimates, the LAUS program applies an X-11 type of seasonal adjustment filter to the benchmarked (not seasonally adjusted) estimates to reduce distortions to seasonality after benchmarking, and then further smooth out variability introduced by the benchmarking process with a trend-type filter. Due to the impact of COVID-19 and to avoid artificial dampening of outlier effects attribution, we only choose the quarterly historical UE rate series from January 1990 to December 2018 and apply the X-11 type of seasonal adjustment filter.

Modeling of the UE rate, including parameter estimation, may be done using the class of Box-Jenkins models; see Box and Jenkins, 1976. For the U.S. quarterly UE rate data, Montgomery et al., 1998 conducted a comprehensive study and compared the performance of forecasts for a variety of linear and nonlinear time series models. They proposed the seasonal ARIMA $(1,1,0) \times (1,0,1)_4$ model as a statistical model for the quarterly UE rate. The overall sample size for the CPS is sufficient to produce reliable estimates of the UE rate at the national level that satisfy certain pre-specified precision requirements. Each state has been classified as a direct-use state or an indirect-use state with respect to the available sample size for each state. For states such as California, Florida, Illinois, Massachusetts, Michigan, New Jersey, New York, North Carolina, Ohio, Pennsylvania, and Texas, the sample sizes are large enough to provide reliable estimates. Those states are classified as direct-use states. However, the remaining states and the District of Columbia, which are classified as indirect-use states, don't provide adequate samples, so standard design-based estimators are not precise enough. Therefore, there is a need to improve efficiency for the states whose sample sizes are insufficient.

Tiller, 1992 developed a method to increase the accuracy of the state-level UE rate estimates. Tiller, 1992 represented the observed CPS sample estimates $Y_{i,t}$ as $Y_{i,t} = \theta_{i,t} + \epsilon_{i,t}$ for $i = 1, \cdots, m$, and $t = 1, \cdots, T$, where $\theta_{i,t}$ is the true UE rate for domain i at time t, and $\epsilon_{i,t}$ is the sampling error, m is the total small areas, and T is the number of time points. The BLS modeled $\theta_{i,t}$'s using a structural time series with explanatory variables and the $\epsilon_{i,t}$ as an ARMA model in order to capture the auto-correlations. Clearly, this method does not utilize the information across states and does not provide uncertainty measures of the state estimates.

Our goal is to produce reliable forecasts of U.S. civilian UE rates for each state by borrowing strength across states. To this end, we consider two possible approaches to grouping the 51 states. In the first approach, for each of the 51 states, we determine the best autoregressive integrated moving average (ARIMA) model for its UE rate series using the built-in R function **arima**. Then, we group the states based on the similarities in the ARIMA structures. This leads to dividing the 51 states into 5 distinct groups such that the ARIMA structures (of UE rates) for states within each group are similar; see Table 1.1. For example, Maine (ME), North Dakota (ND), Delaware (DE), Mississippi (MS), Louisiana (LA), Idaho (ID), New Mexico (NM), Alaska (AK), Hawaii (HI), and Washington (WA) form a single group (Group 1) since they all share the same time series structure ARIMA $(1, 1, 1)$. Even though these states are not geographically connected, the fact that they share a similar time series structure will allow us to borrow strength from the states within the group and produce improved parameter estimates and forecasts.

Table 1.1: Five groups of states with similar ARIMA structure

Group (number of states)	States
Group 1 (10)	ME, ND, DE, MS, LA, ID, NM, AK, HI, WA
Group 2 (13)	NJ, IN, MO, NE, NC, SC, KY, AR, CA, AL, OK, MT, WY
Group 3 (10)	CT, IL, MA, NH, RI, VT, NY, PA, UT, MI
Group 4 (13)	OH, KS, MN, VA, DC, TN, NV, WI, IA, SD, MD, WV, CO
Group 5 (5)	FL, GA, TX, AZ, OR

In the second approach, we group the 51 states simply based on the four regions–Northeast, Midwest, South, and West; see Table 1.2

Table 1.2: States within the four U.S. regions (Census Bureau classification)

Region (number of states)	States
Region 1: Northeast (9)	CT, ME, MA, NH, RI, VT, NJ, NY, PA
Region 2: Midwest (12)	IL, IN, MI, OH, WI, IA, KS, MN, MO, NE, ND, SD
Region 3: South (17)	DE, FL, GA, MD, NC, SC, VA, DC, WV, AL, KY, MS, TN, AR, LA, OK, TX
Region 4: West (13)	AZ, CO, ID, MT, NV, NM, UT, WY, AK, CA, HI, OR, WA

As mentioned earlier, the idea behind grouping the 51 states based on similarities in the ARIMA structures or similarities that may exist within each region is to produce better forecasts of UE rates for each state (within a group) by borrowing strength across other states in that group. The formation of groups of states naturally enables us to use methodologies from small-area estimation combined with a Bayesian framework to produce reliable forecasts of UE rates for each of the 51 states in the U.S. As mentioned in the introduction, we will also use the UI claims data for each state as a covariate to improve the forecast of UE rates. For more details, see Section 1.2.2.

Seasonal AutoRegressive Moving Average (SARIMA) models have been widely used in economics and statistics. There is considerable literature on inference for these models using frequentist approaches, such as least squares or maximum likelihood methods (see Anderson, 1978, and Azzalini, 1981). A Bayesian modeling framework has the advantage of being able to incorporate available prior information in a natural way. Bayesian inference has been facilitated by the emergence of MCMC simulation methods such as the Gibbs sampler (see Tanner and Wong, 1987, Gelfand and Smith, 1990) and Metropolis-Hastings (MH) algorithms (Metropolis et al., 1953, Hastings, 1970, and Tierney, 1994). These methods are powerful tools for simulating intractable joint distribution of interest. A sample of draws is the output of the simulation, and it can be used for various purposes such as computing certain characteristics associated with posterior distributions, e.g., moments.

Bayesian framework for ARMA models was spurred by the approach of Monahan, 1983, and Broemeling and Shaarawy, 1986. Marriott et al., 1993 discussed an approach to the estimation of ARMA models that are based on sampling functions of the partial auto-correlations. The importance of developing Bayesian inference for regression models with ARMA (p,q) errors was recognized early by Chib and Greenberg, 1994, McCulloch and Tsay, 1993, and Albert and Chib, 1993. In fact, Chib and Greenberg, 1994 developed exact methods for regression models with ARMA (p,q) errors in a Bayesian framework by using Gibbs sampling and Metropolis-Hasting algorithms.

Although a Bayesian perspective for time series has been actively pursued, a full treatment for ARMA models is still lacking. In Section 1.2.2, we present a Bayesian inference methodology for ARMA models using MCMC and Metropolis Hasting methods and the assumptions on prior distributions. In Section 1.2.3, we discuss the exact likelihood function for ARMA models. In Section 1.2.4, we describe our modeling of the U.S. civilian UE rates for each of the 51 states that are grouped based on similarities in the ARIMA structures and based on the region, respectively. Also, Bayesian fitting and inference through the Gibbs sampler are discussed in the subsequent sections. Overall results, summary and conclusions are given in Section 1.3 and Section 1.4, respectively.

1.2.2 Bayesian Inference Methodology for ARMA model

Let η_t denote a univariate $ARMA(p, q)$ time series modeled as

$$\phi_p(B)\eta_t = \theta_q(B)a_t, \quad t = p + q + 1, \ldots, T \tag{1.1}$$

where $a_t \overset{i.i.d.}{\sim} N(0, \sigma^2)$, $\sigma^2 > 0$, $\phi_p(B) = (1 - \phi_1 B - \cdots - \phi_p B^p)$, and $\theta_q(B) = (1 + \theta_1 B + \cdots + \theta_q B^q)$ are autoregressive and moving average polynomials, respectively. Here, we assume that $\phi_p(B)$ and $\theta_q(B)$ obey the usual stationarity and invertibility conditions. Then, equivalently, 1.1 can be expressed as:

$$\eta_t = \phi_1 \eta_{t-1} + \cdots + \phi_p \eta_{t-p} + a_t + \theta_1 a_{t-1} + \cdots + \theta_q a_{t-q}. \tag{1.2}$$

Let $\boldsymbol{\phi} = (\phi_1, \cdots, \phi_p)^\intercal$, $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_q)^\intercal$, $\boldsymbol{\eta} = (\eta_1, \cdots, \eta_T)^\intercal$, and $\boldsymbol{a} = (a_1, \cdots, a_T)^\intercal$, where $\intercal$ denotes the transpose of a vector or a matrix.

In small-area estimation of population means, the Fay-Herriot model (Fay and Herriot, 1979) is a popular model to use. This model can be presented as a mixed effects model:

$$y_i = \boldsymbol{x_i}^\intercal \boldsymbol{\beta} + v_i + e_i,\ i = 1, \cdots, N, \tag{1.3}$$

where $\boldsymbol{x_i}$ is a vector of known covariates, $\boldsymbol{\beta}$ is a vector of unknown regression coefficients, v_i's are area-specific random effects following $N(0, A)$, and e_i's are sampling errors following $N(0, D_i)$. Also, v_i are assumed to be independent of e_i ($v_i \perp\!\!\!\perp e_i$). For each i, suppose $y_{i,t}$ is a time series in t and $w_{i,t} = (1 - B)y_{i,t}$ is the first-differenced series. Motivated by the Fay-Herriot model, to utilize the information across areas/states, we can model $w_{i,t}$ series as

$$w_{i,t} = \beta_0 + x_{i,t}\beta_i + \eta_{i,t} + \epsilon_{i,t},\ i = 1, \ldots, N,\ t = 1, \ldots, T, \tag{1.4}$$

where, for each i, $x_{i,t}$ is a first-differenced, observed covariate time series, β_i's are unknown regression coefficients, $\eta_{i,t}$ is an unobserved time series, $\epsilon_{i,t} \overset{i.i.d.}{\sim} N(0, \sigma^2)$, and $\eta_{i,t} \perp\!\!\!\perp \epsilon_{i,t}$. In this dissertation, $w_{i,t}$ denotes the first-differenced U.S. civilian UE rate for state i at time t and $x_{i,t}$

denotes the first-differenced UI claims series. As in 1.2, for each state i, we assume that $\eta_{i,t}$ follows an ARMA(p,q) time series model defined by

$$\eta_{i,t} = \phi_1\eta_{i,t-1} + \cdots + \phi_p\eta_{i,t-p} + a_{i,t} + \theta_1 a_{i,t-1} + \cdots + \theta_q a_{i,t-q}. \tag{1.5}$$

where $a_{i,t} \overset{i.i.d.}{\sim} N(0, \sigma^2_{\eta_i})$, $\sigma^2_{\eta_i} > 0$. Equation 1.4 can also be represented in vector form as:

$$\mathbf{W} = \begin{bmatrix} 1 & \boldsymbol{x_1} & & \\ 1 & & \boldsymbol{x_2} & \text{\Large 0} \\ \vdots & & & \ddots \\ 1 & \text{\Large 0} & & \boldsymbol{x_N} \end{bmatrix}_{NT\times(N+1)} \boldsymbol{\beta} + \begin{bmatrix} \boldsymbol{\eta_1} \\ \boldsymbol{\eta_2} \\ \vdots \\ \boldsymbol{\eta_N} \end{bmatrix}_{NT\times 1} + \begin{bmatrix} \boldsymbol{\epsilon_1} \\ \boldsymbol{\epsilon_2} \\ \vdots \\ \boldsymbol{\epsilon_N} \end{bmatrix}_{NT\times 1} \tag{1.6}$$

where $\mathbf{W} = (\boldsymbol{w_1}, \boldsymbol{w_2}, \cdots, \boldsymbol{w_N})^\intercal$, with $\boldsymbol{w_i} = (w_{i1}, w_{i2}, \cdots, w_{iT})^\intercal$, and $\boldsymbol{x_i} = (x_{i1}, x_{i2}, \cdots, x_{iT})^\intercal$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \cdots, \beta_N)^\intercal$, $\boldsymbol{\eta} = (\boldsymbol{\eta_1}, \cdots, \boldsymbol{\eta_N})^\intercal$ with $\boldsymbol{\eta_i} = (\eta_{i,1}, \eta_{i,2}, \cdots, \eta_{i,T})^\intercal$, and $\boldsymbol{\epsilon_i} = (\epsilon_{i1}, \epsilon_{i2}, \cdots, \epsilon_{iT})^\intercal$.

Let $\boldsymbol{\Phi} = (\boldsymbol{\phi_1}, \cdots, \boldsymbol{\phi_N})^\intercal$ with $\boldsymbol{\phi_i} = (\phi_{i1}, \cdots, \phi_{ip})^\intercal$, $\boldsymbol{\Theta} = (\boldsymbol{\theta_1}, \cdots, \boldsymbol{\theta_N})^\intercal$ with $\boldsymbol{\theta_i} = (\theta_{i1}, \cdots, \theta_{iq})^\intercal$. Given the data and the parameter $\boldsymbol{\Psi} = (\boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\Phi}, \boldsymbol{\Theta}, \boldsymbol{\sigma^2_\eta}, \sigma^2)$, the Bayesian model specification requires a likelihood function $f(\mathbf{W}|\boldsymbol{\Psi})$ and a prior density $\pi(\boldsymbol{\Psi})$. By the Bayes theorem, we obtain the posterior density as $\pi(\boldsymbol{\Psi}|\mathbf{W}) \propto f(\mathbf{W}|\boldsymbol{\Psi})\pi(\boldsymbol{\Psi})$. Given the $\mathbf{W}$, the posterior distribution of the parameter vector $\boldsymbol{\Psi} = (\boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\Phi}, \boldsymbol{\Theta}, \boldsymbol{\sigma^2_\eta}, \sigma^2)$ can be expressed as:

$$\pi(\boldsymbol{\Psi}|\mathbf{W}) \propto \underbrace{f(\mathbf{W}|\boldsymbol{\beta}, \boldsymbol{\eta}, \sigma^2)}_{1} \times \underbrace{\pi(\boldsymbol{\beta})}_{2} \times \underbrace{f(\boldsymbol{\eta}|\boldsymbol{\Phi}, \boldsymbol{\Theta}, \boldsymbol{\sigma^2_\eta})}_{3} \times \underbrace{\prod_{i=1}^{N}\{\pi(\boldsymbol{\phi_i}, \boldsymbol{\theta_i})\pi(\sigma^2_{\eta_i})\}}_{4} \times \underbrace{\pi(\sigma^2)}_{5}. \tag{1.7}$$

For the prior distributions, we make the following assumptions.

Assumption 1 *$f(\boldsymbol{W}|\boldsymbol{\beta},\boldsymbol{\eta},\sigma^2)$ is a $N(\boldsymbol{X\beta}+\boldsymbol{\eta},\Sigma)$ density, where*

$$\Sigma = \begin{bmatrix} \begin{bmatrix} 2\sigma^2 & -\sigma^2 & \\ -\sigma^2 & \ddots & -\sigma^2 \\ & -\sigma^2 & 2\sigma^2 \end{bmatrix} & & \\ & \begin{bmatrix} \ddots & \ddots & \\ & \ddots & \ddots \\ & \ddots & \ddots \end{bmatrix} & \\ & & \begin{bmatrix} 2\sigma^2 & -\sigma^2 & \\ -\sigma^2 & \ddots & -\sigma^2 \\ & -\sigma^2 & 2\sigma^2 \end{bmatrix} \end{bmatrix}.$$

Note that, inside the Σ matrix are the band matrices, and $\boldsymbol{\eta}$ is a $NT\times 1$ vector.

Assumption 2 *$\pi(\boldsymbol{\beta})$ is a $N(\boldsymbol{\beta_0},\Sigma_{\beta_0})$ density, where $\boldsymbol{\beta_0}$ and $\Sigma_{\boldsymbol{\beta_0}}$ are hyperparameters. We assume $\beta_0=(0,1,\cdots,1)^\intercal$, and $\Sigma_{\boldsymbol{\beta_0}}=100\times\boldsymbol{I_{N+1}}$.*

Assumption 3 *$\boldsymbol{\eta}\sim N(\boldsymbol{0},\Sigma_\eta)$, where*

$$\Sigma_\eta = \begin{pmatrix} \sigma^2_{\eta_1}M^{-1}(\boldsymbol{\phi_1},\boldsymbol{\theta_1}) & & & \\ & \sigma^2_{\eta_2}M^{-1}(\boldsymbol{\phi_2},\boldsymbol{\theta_2}) & & \boldsymbol{0} \\ & \boldsymbol{0} & \ddots & \\ & & & \sigma^2_{\eta_N}M^{-1}(\boldsymbol{\phi_N},\boldsymbol{\theta_N}) \end{pmatrix},$$

and M is the covariance matrix of ARMA model in terms of $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$. Here, for $i=1,\cdots,N$, $\boldsymbol{\phi_i}=(\phi_{i1},\cdots,\phi_{ip})$ and $\boldsymbol{\theta_i}=(\theta_{i1},\cdots,\theta_{iq})$ where $p,q\leqslant 2$ and p or q are same for each state within one group. We denote the variance term as Σ_η. Then

$$f(\boldsymbol{\eta}|\boldsymbol{\phi},\boldsymbol{\theta},\boldsymbol{\sigma^2_\eta}) = (2\pi)^{\frac{-NT}{2}}|\Sigma_\eta|^{-\frac{1}{2}}exp\{-\frac{1}{2}\boldsymbol{\eta}^\intercal\Sigma_\eta^{-1}\boldsymbol{\eta}\}.$$

Assumption 4 *Due to stationarity and invertibility constraints, for each element of vector* $\boldsymbol{\phi_i}$, *we assume that* $\phi_{ip} \sim U(-1,1)$, *where* $i = 1, \cdots, N$ *and* $p \leqslant 2$. *Similarly, for each element of vector* $\boldsymbol{\theta_i}$, *we assume that* $\theta_{iq} \sim U(-1,1)$, *where* $i = 1, \cdots, N$ *and* $q \leqslant 2$.

Assumption 5 *For each* $i = 1, \cdots, N$, $\pi(\sigma^2_{\eta_i}) \sim IG(\frac{v_1}{2}, \frac{\delta_1}{2})$, *and* $\pi(\sigma^2) \sim IG(\frac{v_2}{2}, \frac{\delta_2}{2})$, *where* $IG(\cdot)$ *is the inverse gamma distribution, and* $v_1,\ v_2,\ \delta_1,\ \delta_2$ *are hyperparamters.*

Due to stationarity and invertibility constraints, we set $-1 < \theta_{iq}, \phi_{ip} < 1$. We now perform the Fisher transformation of θ_{iq} *and* ϕ_{ip} defined as $\zeta_{\theta_{iq}} = \frac{1}{2}\ln\frac{1+\theta_{iq}}{1-\theta_{iq}}$ and $\zeta_{\phi_{ip}} = \frac{1}{2}\ln\frac{1+\phi_{ip}}{1-\phi_{ip}}$. Then, $\zeta_{\theta_{iq}}$ and $\zeta_{\phi_{ip}} \in (-\infty, \infty)$. We assume that $\zeta_{\theta_{iq}} \sim N(0, \tau^2)$ and $\zeta_{\phi_{ip}} \sim N(0, \tau^2)$, where $p, q \leqslant 2$. By the normaility assumption on $\zeta_{\theta_{iq}}$, we know that

$$P[-2\tau < \frac{1}{2}\ln\frac{1+\theta_{iq}}{1-\theta_{iq}} < 2\tau] = 0.95.$$

This enables us to find τ^2 using the following arguments:

$$\begin{aligned} 0.95 = P[-2\tau < \frac{1}{2}\ln\frac{1+\theta_{iq}}{1-\theta_{iq}} < 2\tau] &= P[e^{-4\tau} < \frac{1+\theta_{iq}}{1-\theta_{iq}} < e^{4\tau}] \\ &= P[1+e^{-4\tau} < \frac{2}{1-\theta_{iq}} < 1+e^{4\tau}] \\ &= P[1-\frac{2}{1+e^{-4\tau}} < \theta_{iq} < 1-\frac{2}{1+e^{4\tau}}]. \end{aligned} \tag{1.8}$$

Suppose we want to allow a broad range of possible values for θ_{iq}, say $-0.9 < \theta_{iq} < 0.9$, then it can be shown that $4\tau = \ln(19)$. Denote $\tau_0 = \frac{1}{4}\ln(19)$, then $\zeta_{\theta_{iq}} \sim N(0, \tau_0^2)$. Similarly, $\zeta_{\phi_{ip}} \sim N(0, \tau_0^2)$.

Note that the hyperparamters $\boldsymbol{\beta_0}, \Sigma_{\boldsymbol{\beta_0}},\ v_1,\ v_2,\ \delta_1,\ \delta_2, \tau_0^2$ are assumed to be known. Also, each parameter is assumed to be independently distributed with each other. We also denote $\sigma^{2^*} = \ln(\sigma^2)$ and, for each $i = 1, \cdots, N$, ${\sigma^2_{\eta_i}}^* = \ln(\sigma^2_{\eta_i})$, so that σ^{2^*} and ${\sigma^2_{\eta_i}}^* \in (-\infty, \infty)$. Finally, let $\boldsymbol{\Psi}^* = (\boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\zeta_\Phi}, \boldsymbol{\zeta_\Theta}, \boldsymbol{\sigma_\eta^{2^*}}, \sigma^{2^*})$. Thus, based on the prior distribution assumptions and transformations, equation 1.7 can be rewritten as

$$
\begin{aligned}
\pi(\mathbf{\Psi}^*|\mathbf{W}) \propto\ & (2\pi)^{-\frac{NT}{2}}|\Sigma|^{-\frac{1}{2}}exp\{-\frac{1}{2}(\boldsymbol{W}-\boldsymbol{x\beta}-\boldsymbol{\eta})^{\intercal}\Sigma^{-1}(\boldsymbol{W}-\boldsymbol{x\beta}-\boldsymbol{\eta})\} \\
& \times |\Sigma_\beta|^{-\frac{1}{2}}exp\{(\boldsymbol{\beta}-\boldsymbol{\beta_0})^{\intercal}\Sigma_\beta^{-1}(\boldsymbol{\beta}-\boldsymbol{\beta_0})\} \\
& \times (2\pi)^{-\frac{NT}{2}}|\Sigma_\eta|^{-\frac{1}{2}}exp\{-\frac{1}{2}\boldsymbol{\eta}^{\intercal}\boldsymbol{\Sigma_\eta^{-1}}\boldsymbol{\eta}\} \\
& \times \prod_{i=1}^{N}\{exp\{-\frac{1}{2}\zeta_{\boldsymbol{\theta_i}}^{\intercal}\tau_0^{-1}\zeta_{\boldsymbol{\theta_i}}\}exp\{-\frac{1}{2}\zeta_{\boldsymbol{\phi_i}}^{\intercal}\tau_0^{-1}\zeta_{\boldsymbol{\phi_i}}\}(\sigma_{\eta_i}^{2\,*})^{-\frac{\nu_1}{2}-1}exp(-\frac{1}{\sigma_{\eta_i}^{2\,*}}\frac{\delta_1}{2})\} \\
& \times (\sigma^{2*})^{-\frac{\nu_2}{2}-1}exp(-\frac{1}{\sigma^{2*}}\frac{\delta_2}{2}) \times \prod_{i=1}^{N}|J_{\sigma_{\eta_i}^2}| \times |J_{\sigma^2}| \times \prod_{i=1}^{N}|J_{\boldsymbol{\theta_i}}| \times \prod_{i=1}^{N}|J_{\boldsymbol{\phi_i}}|, \qquad (1.9)
\end{aligned}
$$

where $|J_{\sigma_{\eta_i}^2}| = exp(\sigma_{\eta_i}^{2\,*})$, $|J_{\sigma^2}| = exp(\sigma^{2*})$, and $|J_{\boldsymbol{\theta_i}}| = |J_{\boldsymbol{\phi_i}}| = \mathbf{1}$ for $i = 1, \cdots, N$. We use the assumptions that

$$
\begin{aligned}
\mathbf{W}|\mathbf{\Psi} &\sim N(\boldsymbol{X\beta}+\boldsymbol{\eta}, \Sigma), \\
\boldsymbol{\beta} &\sim N(\boldsymbol{\beta_0}, \Sigma_\beta), \\
\boldsymbol{\eta} &\sim N(\mathbf{0}, \Sigma_\eta)
\end{aligned}
$$

to integrate out $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$. Note that

$$
\begin{aligned}
E(\mathbf{W}) &= E[E(\boldsymbol{W}|\mathbf{\Psi})] = E(\boldsymbol{X\beta}+\boldsymbol{\eta}) = \boldsymbol{X\beta_0} \triangleq h \\
V(\mathbf{W}) &= E[V(\boldsymbol{W}|\mathbf{\Psi})] + V[E(\boldsymbol{W}|\mathbf{\Psi})] = \Sigma + X\Sigma_\beta X^{\intercal} + \Sigma_\eta \triangleq \mathbf{\Omega}
\end{aligned}
$$

Therefore, the posterior distribution of $\mathbf{\Psi}^*_{-\boldsymbol{\beta},-\boldsymbol{\eta}}$ is

$$
\begin{aligned}
\pi(\mathbf{\Psi}^*_{-\boldsymbol{\beta},-\boldsymbol{\eta}}|\mathbf{W}) \propto\ & |\Omega|^{-\frac{1}{2}}exp\{-\frac{1}{2}(\mathbf{W\text{-}h})^{\intercal}\Sigma^{-1}(\mathbf{W\text{-}h})\} \\
& \times \prod_{i=1}^{N}\{exp\{-\frac{1}{2}\zeta_{\boldsymbol{\theta_i}}^{\intercal}\tau_0^{-1}\zeta_{\boldsymbol{\theta_i}}\}exp\{-\frac{1}{2}\zeta_{\boldsymbol{\phi_i}}^{\intercal}\tau_0^{-1}\zeta_{\boldsymbol{\phi_i}}\}(\sigma_{\eta_i}^{2\,*})^{-\frac{\nu_1}{2}-1}exp(-\frac{1}{\sigma_{\eta_i}^{2\,*}}\frac{\delta_1}{2})\} \\
& \times (\sigma^{2*})^{-\frac{\nu_2}{2}-1}exp(-\frac{1}{\sigma^{2*}}\frac{\delta_2}{2}) \times \prod_{i=1}^{N}|J_{\sigma_{\eta_i}^2}| \times |J_{\sigma^2}|. \qquad (1.10)
\end{aligned}
$$

Hence,

$$
\begin{aligned}
\ln(\pi(\boldsymbol{\Psi}^{*}_{-\boldsymbol{\beta},-\boldsymbol{\eta}}|\mathbf{W})) \propto & -\frac{1}{2}|\Omega| - \frac{1}{2}(\mathbf{W}\text{-}\mathbf{h})^{\intercal}\Sigma^{-1}(\mathbf{W}\text{-}\mathbf{h}) \\
& -\frac{1}{2}\sum_{i=1}^{N}\left\{\zeta_{\boldsymbol{\theta}_i}^{\intercal}\tau_0^{-1}\zeta_{\boldsymbol{\theta}_i} - \frac{1}{2}\zeta_{\boldsymbol{\phi}_i}^{\intercal}\tau_0^{-1}\zeta_{\boldsymbol{\phi}_i}\right\} - \left\{\frac{\nu_1}{2}+1\right\}\sum_{i=1}^{N}\ln(\sigma^{2\,*}_{\eta_i}) - \sum_{i=1}^{N}\frac{1}{\sigma^{2\,*}_{\eta_i}}\frac{\delta_1}{2} \\
& -\left\{\frac{\nu_2}{2}+1\right\}\ln(\sigma^{2*}) - \frac{1}{\sigma^{2*}}\frac{\delta_2}{2} + \ln|\boldsymbol{J}_{\sigma^2}| + \sum_{i=1}^{N}\ln|J_{\sigma^2_{\eta_i}}|. \qquad (1.11)
\end{aligned}
$$

We know that the joint posterior distribution of $\boldsymbol{\beta},\boldsymbol{\eta}$ is:

$$
\begin{aligned}
f(\boldsymbol{\beta},\boldsymbol{\eta}|\boldsymbol{\Psi}^{*}_{-\boldsymbol{\beta},-\boldsymbol{\eta}}) &\propto f(\mathbf{W}|\boldsymbol{\beta},\boldsymbol{\eta}) \times \pi_1(\boldsymbol{\beta}) \times \pi_2(\boldsymbol{\eta}) \\
&\propto N(\mathbf{W}|\boldsymbol{X\beta}+\boldsymbol{\eta},\boldsymbol{\Sigma}) \times N(\boldsymbol{\beta}|\boldsymbol{\beta_0},\boldsymbol{\Sigma_{\beta_0}}) \times N(\boldsymbol{\eta}|\boldsymbol{\eta_0},\Sigma_{\eta_0}) \\
f(\boldsymbol{\eta}|\boldsymbol{\Psi}^{*}_{-\boldsymbol{\eta}}) &= \int f(\boldsymbol{\beta},\boldsymbol{\eta}|\boldsymbol{\Psi}^{*}_{-\boldsymbol{\beta},\boldsymbol{\eta}})\mathrm{d}\boldsymbol{\beta} \\
&\propto \int N(\mathbf{W}|\boldsymbol{X\beta}+\boldsymbol{\eta},\boldsymbol{\Sigma}) \times N(\boldsymbol{\beta}|\boldsymbol{\beta_0},\boldsymbol{\Sigma_{\beta_0}})\mathrm{d}\boldsymbol{\beta} \times N(\boldsymbol{\eta}|\boldsymbol{\eta_0},\boldsymbol{\Sigma_{\eta_0}}). \qquad (1.12)
\end{aligned}
$$

According to the properties of normal distribution, if $Y|X = x \sim N(x,\tau^2)$, $\tau^2 > 0$ and $X \sim N(\mu,\sigma^2)$, then the joint pdf of X, Y is $f(x,y) = \frac{1}{\sqrt{(2\pi)^2\sigma^2\tau^2}}exp\{-\frac{(x-\mu)^2}{2\sigma^2} - \frac{(y-x)^2}{2\tau^2}\}$. The marginal distribution of Y is $N(\mu,\sigma^2+\tau^2)$. Thus, equation 1.12

$$
\propto N(\mathbf{W}|\boldsymbol{X\beta_0}+\boldsymbol{\eta},\boldsymbol{\Sigma}+\boldsymbol{X\Sigma_{\beta_0}X}^{\intercal}) \times N(\boldsymbol{\eta}|\boldsymbol{\eta_0},\boldsymbol{\Sigma_{\eta_0}}) \qquad (1.13)
$$

We also know that if $f(y|\theta) = N(y|\theta,\sigma^2)$, and $\pi(\theta) = N(\theta|\mu,\tau^2)$, then

$$
\pi(\theta|y) = N\left(\theta \,\middle|\, \frac{\frac{y}{\sigma^2}+\frac{\mu}{\tau^2}}{\frac{1}{\sigma^2}+\frac{1}{\tau^2}}, \left(\frac{1}{\sigma^2}+\frac{1}{\tau^2}\right)^{-1}\right).
$$

Therefore,

$$\begin{aligned}
\boldsymbol{\eta}|\boldsymbol{\Psi}^*_{-\boldsymbol{\eta}} &\sim \mathbf{MVN}(\boldsymbol{E}(\boldsymbol{\eta}|\boldsymbol{\Psi}^*_{-\boldsymbol{\eta}}), \boldsymbol{V}(\boldsymbol{\eta}|\boldsymbol{\Psi}^*_{-\boldsymbol{\eta}})), with \\
\boldsymbol{V}(\boldsymbol{\eta}|\boldsymbol{\Psi}^*_{-\boldsymbol{\eta}}) &= \{\boldsymbol{\Sigma}^{-1}_{\boldsymbol{\eta_0}} + (\boldsymbol{\Sigma} + \boldsymbol{X}\boldsymbol{\Sigma}_{\boldsymbol{\beta_0}}\boldsymbol{X}^{\intercal})^{-1}\}^{-1} \\
\boldsymbol{E}(\boldsymbol{\eta}|\boldsymbol{\Psi}^*_{-\boldsymbol{\eta}}) &= \boldsymbol{V}(\boldsymbol{\eta}|\boldsymbol{\Psi}^*_{-\boldsymbol{\eta}}) \times \boldsymbol{\Sigma}^{-1}_{\boldsymbol{\eta_0}}\boldsymbol{\eta_0} + (\boldsymbol{\Sigma} + \boldsymbol{X}\boldsymbol{\Sigma}_{\boldsymbol{\beta_0}}\boldsymbol{X}^{\intercal})^{-1}(\boldsymbol{W} - \boldsymbol{X}\boldsymbol{\beta_0}),
\end{aligned} \tag{1.14}$$

where the prior density of $\boldsymbol{\eta}$ is as in Assumption 3, $\boldsymbol{\Sigma}$ as defined in Assumption 1, M is the covariance matrix of ARMA model in terms of $\boldsymbol{\phi_i}$ and $\boldsymbol{\theta_i}$, where $\boldsymbol{\phi_i}$ is a vector of length p and $\boldsymbol{\theta_i}$ is a vector of length q, where $p, q \leqslant 2$. $\sigma^2_{\eta_i}$, $\boldsymbol{\phi_i}$ and $\boldsymbol{\theta_i}$ are posterior samples from Metropolis-Hasting method above, and $\boldsymbol{\eta_0} = \mathbf{0}$, $\beta_0 = (0, 1, \cdots, 1)^{\intercal}$, and $\Sigma_{\beta_0} = 100 \times \boldsymbol{I_{N+1}}$. Once we get the marginal distribution of $\boldsymbol{\eta}$, the conditional distribution of $\boldsymbol{\beta}$ is

$$\begin{aligned}
\boldsymbol{f}(\boldsymbol{\beta}|\boldsymbol{\eta}) &\propto \boldsymbol{N}(\boldsymbol{W}|\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\eta}, \boldsymbol{\Sigma}) \times \boldsymbol{N}(\boldsymbol{\beta}|\boldsymbol{\beta_0}, \boldsymbol{\Sigma_{\beta_0}}) \\
&\sim \boldsymbol{MVN}(\boldsymbol{E}(\boldsymbol{\beta}|\boldsymbol{\eta}), \boldsymbol{V}(\boldsymbol{\beta}|\boldsymbol{\eta})), where \\
\boldsymbol{V}(\boldsymbol{\beta}|\boldsymbol{\eta}) &= (\boldsymbol{X}^{\intercal}\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1} \\
\boldsymbol{E}(\boldsymbol{\beta}|\boldsymbol{\eta}) &= \boldsymbol{V}(\boldsymbol{\beta}|\boldsymbol{\eta}) \times (\boldsymbol{X}^{\intercal}\boldsymbol{\Sigma}^{-1}(\boldsymbol{W} - \boldsymbol{\eta})).
\end{aligned} \tag{1.15}$$

1.2.3 Exact likelihood function of ARMA models

Recall from Assumption 3 that the probability density of $\boldsymbol{\eta}$ is

$$f(\boldsymbol{\eta}|\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\sigma^2_\eta}) = (2\pi\sigma^2_\eta)^{-\frac{T}{2}}|M|^{\frac{1}{2}}exp(-\frac{1}{2\sigma^2_\eta}\boldsymbol{\eta}^{\intercal}M\boldsymbol{\eta}) \tag{1.16}$$

which as a function of θ's, ϕ's and σ^2_η for fixed η is the likelihood function, providing a motive for finding M and its determinant. Galbraith and Galbraith, 1974 obtained explicit expressions for M in terms of the $\boldsymbol{\phi}$'s and $\boldsymbol{\theta}$'s where p, q are less than 2. For the special case ARMA(1,1) defined by

$$\eta_t - \phi\eta_{t-1} = z_t - \theta z_{t-1},\ |\phi| < |1|$$

the determinant of M is

$$|M| = \left[\frac{(1-\phi\theta)^2 - (\theta-\phi)^2\theta^{2T}}{(1-\phi^2)(1-\theta^2)}\right]^{-1},$$

where, ϕ and θ are scalars. The following arguments show that the determinant, $|M|$, is positive. Let $g(\theta) = (1-\phi\theta)^2 - (\theta-\phi)^2\theta^{2T}$. Suppose $|\theta| < 1$, then

$$\begin{aligned}
(\theta-\phi)^2\theta^{2T} &< (\theta-\phi)^2 \\
g(\theta) &> (1-\phi\theta)^2 - (\theta-\phi)^2 \\
&= 1 + \phi^2\theta^2 - \theta^2 - \phi^2 \\
&= (1-\theta^2)(1-\phi^2) \\
&> 0.
\end{aligned}$$

Therefore, $\frac{g(\theta)}{(1-\theta^2)(1-\phi^2)} > 1$. Now, suppose $|\theta| > 1$, then

$$\begin{aligned}
g(\theta) &= (1-\phi\theta)^2 - (\theta-\phi)^2\theta^{2T} \\
&= \theta^{2T+2}[(\theta^{-1}-\phi)^2\theta^{-2T} - (1-\phi\theta^{-1})^2] \\
&= \theta^{2T+2}[(\theta^*-\phi)^2(\theta^*)^{2T} - (1-\phi\theta^*)^2]
\end{aligned}$$

where $\theta^{-1} = \theta^*$, and $|\theta^*| < |\theta|^{-1}$. Then

$$\begin{aligned}
|M|^{-1} &= \frac{\theta^{2T+2}}{(1-\theta^2)(1-\phi^2)}[(\theta^*-\phi)^2(\theta^*)^{2T} - (1-\phi\theta^*)^2] \\
&= \frac{\theta^{2T}}{(1-\theta^2)(\phi^{-2}-1)}[(\theta^*-\phi)^2(\theta^*)^{2T} - (1-\phi\theta^*)^2] \\
&= -\theta^{2T}\frac{g(\theta^*)}{(1-\phi^2)(1-(\theta^*)^2} \\
&> \theta^{2T} \\
&> 1,
\end{aligned}$$

since $|\theta| > 1$. Thus, $0 < |M| \leqslant 1$, and the equation holds if and only if $\theta = 1$.

Furthermore, the elements of M matrix are:

$$
\begin{aligned}
m_{rr} &= \frac{(1-\phi\theta)^2(\theta-\phi)^2\left[1-\theta^{2(r-1)}\right]\left[1-\theta^{2(T-r)}\right]+\left(1-\theta^2\right)\left[(1-\phi\theta)^2-(\theta-\phi)^2\theta^{2(T-1)}\right]}{(1-\theta^2)\left[(1-\phi\theta)^2-(\theta-\phi)^2\theta^{2T}\right]}\\
m_{rs} &= \frac{\theta^{s-r-1}(1-\phi\theta)(\theta-\phi)\left[(1-\phi\theta)-(\theta-\phi)\theta^{2r-1}\right]\left[(1-\phi\theta)-(\theta-\phi)\theta^{2(T-s)+1}\right]}{(1-\theta^2)\left[(1-\phi\theta)^2-(\theta-\phi)^2\theta^{2T}\right]}.
\end{aligned}
\tag{1.17}
$$

In 1.17, the expression for m_{rr} is valid for $1 \leqslant r \leqslant T$ and that for m_{rs} is valid for $1 \leqslant r < s \leqslant T$. The explicit formulation of the M matrix and its determinant function in terms of $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$ could be derived from equation 1.17. In our Bayesian analysis, we will derive the moments and other characteristics of the posterior distribution of Ψ based on the given assumptions. Frequently, the posterior density is analytically intractable. In order to estimate the model parameters, a sampling-based method has been implemented. The Gibbs sampling method entails sampling from the whole conditional distribution of each parameter in a systematic manner, contingent on the prior sample values of the other parameters. However, comprehensive conditional distributions can be difficult to derive at times. The Metropolis Hasting algorithm is another excellent alternative.

Our analysis will establish the mean and standard deviation of the posterior distribution of $\boldsymbol{\Psi}$ based on the aforementioned assumptions. We estimate $\boldsymbol{\Psi}$ using the posterior means and evaluate the uncertainty in the estimation of $\boldsymbol{\Psi}$ using the posterior standard deviation. For these computations, we use the Gibbs sampler; see Gelman and Rubin, 1992

The Gibbs sampler is a Monte Carlo Markovian updating technique that generates the marginal, conditional, and joint distributions of random variables. The Gibbs sampling procedure requires sampling from the complete conditional distributions associated elements of $\boldsymbol{\Psi}$ in some regular sequence. In our application, the full conditional distributions of $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$ are standard distributions that can be easily sampled. In contrast, those of $\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\sigma_{\eta}^2}, \sigma^2$ are more intricate, and therefore they are computed in an analogous manner.

To implement the sampling algorithm, we integrate out $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$, as indicated in equation 1.11. We employ draws based on a Metropolis Hastings distribution with the remaining parameters. We use the random walk version of the Metropolis Hastings algorithm with suitable Gaussian proposals to collect samples from the necessary stationary distributions (Hastings, 1970). In the random walk version, new candidates are chosen by drawing from a distribution conditioned on the current parameter value, i.e., by drawing a step away from the current parameter value. More specifically, let U denote the current value of the parameter vectors. We draw V from the proposal centered at U. We calculate the ratio $\alpha(U,V) = f(y|\boldsymbol{\phi}, \boldsymbol{\theta}, \boldsymbol{\sigma}_{\boldsymbol{\eta}}^{\mathbf{2}}, \boldsymbol{\sigma}^{\mathbf{2}}, V)\pi(V)/f(y|\boldsymbol{\phi}, \boldsymbol{\theta}, \boldsymbol{\sigma}_{\boldsymbol{\eta}}^{\mathbf{2}}, \boldsymbol{\sigma}^{\mathbf{2}}, U)\pi(U)$. Then, we move U to V with the probability of $\alpha(U,V)$.

Clearly, successful implementation of the MH algorithm requires a suitable proposal density. Note that, in a Bayesian analysis of ARMA model, Chib and Greenberg, 1994 showed the full conditional distributions of ϕ, σ^2, and a_0 belong to standard families of distributions, and to implement the MH algorithm of θ, they proposed a candidate-generating density as the truncated normal approximation by expanding $a_t(\theta)$ around θ^+ as $a_t \approx a_t(\theta^+) - w_t^{'}(\theta - \theta^+)$, where θ^+ denotes the nonlinear least squares estimate of θ and $w_t^{'}$ is the $i-$th row of $W(\theta^+) = (\partial a(\theta)/\partial \theta^{'})|\theta = \theta^+$. Also note that in a Bayesian analysis of AR or MA model, Marin et al., 2005 represented the polynomials as the factorized quantity $\phi_p(B) = \prod_{i=1}^{p}(1 - \lambda_i B)$ or $\theta_q(B) = \prod_{i=1}^{q}(1 - \lambda_i B)$, and used a reversible jump algorithm that distinguishes between the number of complex roots in the inverse roots, λ_i. Then, they simulated ϕ or θ from a proposal based on a simple random walk. Our Bayesian analysis in ARMA model use the Fisher transformation for $\boldsymbol{\theta}, \ \boldsymbol{\phi}$ shown around equation 1.8.

1.2.4 Bayesian modeling of UE rates data

In this section, for the U.S. civilian UE rates data, we fit the Bayesian ARIMA model proposed in Section 1.2.2, which includes parameter estimation, prediction, model comparison, and forecasting. The data consists of quarterly UE rates for 51 states (50 states plus District of Columbia) spanning 29 years, from January 1990 to December 2018. Suppose $y_{i,t}, \ t = 1, \cdots, T, \ i = 1, \cdots, N$ represents the UE rate for state i at time t. As described in 1.2, our goal is to produce reliable forecasts of U.S. civilian UE rates for each state by borrowing strength across states.

To this end, we consider two possible approaches to grouping the 51 states. In the first approach, we group the 51 states based on the similarities in the ARIMA structures. This leads to dividing the 51 states into 5 distinct groups such that the ARIMA structures (of UE rates) for states within each group are similar; see Table 1.1. In the second approach, we group the 51 states simply based on the four regions–Northeast, Midwest, South, and West; see Table 1.2.

As mentioned in Section 1.2, the idea behind grouping the 51 states based on similarities in the ARIMA structures or similarities that may exist within each region is to produce better forecasts of UE rates for each state (within a group) by borrowing strength across other states in that group. The formation of groups of states naturally enables us to use the methodologies from small-area estimation combined with a Bayesian framework developed in Section 1.2.2 to produce reliable forecasts of UE rates for each of the 51 states in the U.S.

We model the first-differenced series $w_{i,t} = y_{i,t} - y_{i,t-1}$, for $t = 1, \cdots, T$, $i = 1, \cdots, N$, where T is the number of time points and N is the number of states in one group/region. As mentioned in the introduction, we will also use the UI claims data for each state as a covariate to improve the forecast of UE rates; see model (1.18) for more details. For each state, we designate the UE rates from the first quarter of 1990 (1990:Q1) to the fourth quarter of 2017 (2017:Q4) as the *training data* and the UE rates for remaining four quarters of 2018 as the *test data*. We use Bayesian framework to build time series models for the *training data*, and then compute the forecasts of UE rates for the remaining four quarters in the *test data* based on the fitted models. Finally, using the Mean Absolute Prediction Error (MAPE), we compare the accuracy of the forecasts based on our Bayesian models and those based on the frequentist method of Wilms et al., 2021, which used the **bigtime** R package.

1.2.5 Posterior features

We use the posterior means and standard deviations computed by the Gibbs sampler to estimate the parameter space $\mathbf{\Psi}$ and evaluate the uncertainty in the estimates, respectively. The implementation is discussed at the end of Section 1.2.2. To conduct the Gibbs sampler, we choose the values of hyperparameters to be $\boldsymbol{\beta_0} = c(0, 1, \cdots, 1)$, $\Sigma_\beta = diag(100)$, $v_1 = 0.01$, $v_2 = 0.659$, $\delta_1 = 0.01$ and $\delta_2 = 0.09$.

For each group/region, we run MCMC for 100,000 iterations. To limit the influence of starting values on the final results, the first 250,000 replications are discarded as "burning-in" samples. Moreover, to diminish the serial correlation of the run, we retain every 50th sample out of the remaining replications. For each state in a group/region, we fit the proposed Bayesian model to the 111 (first-differenced) UE rate values in the *training data* (1990:Q1 to 2017:Q4). The last four observations in the *test data* (2018:Q1 to 2018:Q4) are utilized for forecast evaluations.

1.2.6 Forecasting

Forecasts of UE rates are determined via the predictive density. For each state, recall that our proposed model in 1.18 for the first-differenced UE rate series is given by

$$w_{i,t} = \beta_0 + x_{i,t}\beta_i + \eta_{i,t} + \epsilon_{i,t},\ i = 1, \ldots, N,\ t = 1, \ldots, T.$$

Let us consider Group 1 in Table 1.1. For each state in Group 1, suppose $\eta_t = a_t - \boldsymbol{\theta} a_{t-1} + \boldsymbol{\phi}\eta_{t-1}$, where $a_t \sim N(0, \sigma_\eta^2)$. First, we obtain the posteriors of $(\boldsymbol{\theta}, \boldsymbol{\phi}, \sigma^2, \sigma_\eta^2)$, using the MH algorithm. Then, using these estimates, we generate $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$ from

$$\pi(\boldsymbol{\beta}, \boldsymbol{\eta} | \boldsymbol{\Theta}, \boldsymbol{\Phi}, \boldsymbol{\sigma_\eta^2}, \boldsymbol{W}, \sigma^2) \propto f(\boldsymbol{W} | \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\Theta}, \boldsymbol{\Phi}, \boldsymbol{\sigma_\eta^2}, \sigma^2) \times f(\boldsymbol{\eta} | \boldsymbol{\Theta}, \boldsymbol{\Phi}, \boldsymbol{\sigma_\eta^2}) \times \pi(\boldsymbol{\beta}). \tag{1.18}$$

We know that

$$\begin{aligned} f(\boldsymbol{W} | \boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\Theta}, \boldsymbol{\Phi}, \boldsymbol{\sigma_\eta^2}, \sigma^2) &\sim MVN(\boldsymbol{W} | \boldsymbol{X\beta} + \boldsymbol{N_2\eta}, \Sigma) \\ f(\boldsymbol{\eta} | \boldsymbol{\Theta}, \boldsymbol{\Phi}, \boldsymbol{\sigma_\eta^2}) &\sim MVN(\boldsymbol{\eta} | \boldsymbol{0}, \boldsymbol{\sigma_\eta^2 M^{-1}}(\boldsymbol{\theta}, \boldsymbol{\phi})), \end{aligned}$$

where MVN denotes multivariate normal distribution. Then, the forecast at time $T + l$ is

$$\boldsymbol{w_{i,T+l}} = \boldsymbol{x_{i,T+l}^{\intercal}}\boldsymbol{\beta} + \eta_{i,T+l},\ \text{for } l = 1, \cdots, 4. \tag{1.19}$$

Let $\boldsymbol{\eta}^{\boldsymbol{A}} = \begin{pmatrix} \boldsymbol{\eta} \\ \eta_{T+1,T+L} \end{pmatrix}$ denote the augmented vector with

$$\begin{aligned} f(\boldsymbol{\eta}^{\boldsymbol{A}}|\boldsymbol{W}, \boldsymbol{\Theta}, \boldsymbol{\Phi}, \boldsymbol{\beta}, \boldsymbol{\sigma}_{\boldsymbol{\eta}}^{\mathbf{2}}, \sigma^2) &\propto f(\boldsymbol{W}, \boldsymbol{\eta}^{\boldsymbol{A}}|\boldsymbol{\Theta}, \boldsymbol{\Phi}, \boldsymbol{\beta}, \boldsymbol{\sigma}_{\boldsymbol{\eta}}^{\mathbf{2}}, \sigma^2) \\ &= f(\boldsymbol{W}|\boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\Theta}, \boldsymbol{\Phi}, \boldsymbol{\sigma}_{\boldsymbol{\eta}}^{\mathbf{2}}, \sigma^2) f(\boldsymbol{\eta}^{\boldsymbol{A}}|\boldsymbol{\Theta}, \boldsymbol{\Phi}, \boldsymbol{\beta}, \boldsymbol{\sigma}_{\boldsymbol{\eta}}^{\mathbf{2}}, \sigma^2) \\ &= MVN(\boldsymbol{W}|\boldsymbol{X\beta} + \boldsymbol{N_2\eta}, \Sigma) \times MVN(\boldsymbol{\eta}^{\boldsymbol{A}}|\mathbf{0}, \boldsymbol{\sigma}_{\boldsymbol{\eta}}^{\mathbf{2}}\boldsymbol{M}^{-\mathbf{1}}(\boldsymbol{\theta}, \boldsymbol{\phi})) \\ &\propto \pi(\eta_{T+1,T+L}|\boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\Theta}, \boldsymbol{\Phi}, \boldsymbol{\sigma}_{\boldsymbol{\eta}}^{\mathbf{2}}, \sigma^2) \end{aligned}$$

which implies that

$$\pi(\eta_{T+1,T+L}|\boldsymbol{\beta}, \boldsymbol{\eta}, \boldsymbol{\Theta}, \boldsymbol{\Phi}, \boldsymbol{\sigma}_{\boldsymbol{\eta}}^{\mathbf{2}}, \sigma^2, \boldsymbol{w}) \propto f(\boldsymbol{\eta}^{\boldsymbol{A}}|\boldsymbol{W}, \boldsymbol{\Theta}, \boldsymbol{\Phi}, \boldsymbol{\beta}, \boldsymbol{\sigma}_{\boldsymbol{\eta}}^{\mathbf{2}}, \sigma^2).$$

Note that the conditional pdf of $\eta_{T+1,T+L}$ is $MVN(\boldsymbol{\eta}^{\boldsymbol{A}}|\mathbf{0}, \boldsymbol{\sigma}_{\boldsymbol{\eta}}^{\mathbf{2}}\boldsymbol{M}^{-\mathbf{1}}(\boldsymbol{\theta}, \boldsymbol{\phi}))$, and $\boldsymbol{\eta}$ is a vector of length T and $\eta_{T+1,T+L}$ is a vector of length L. Let us denote the variance-covariance matrix as

$$\Sigma^{-1} = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22}. \end{pmatrix}$$

where M_{11} is a matrix with dimension of T by T and M_{22} is a matrix with dimension L by L. Then,

$$E(\eta_{T+1,T+L}|\boldsymbol{\eta}) = -M_{22}^{-1}M_{21}\boldsymbol{\eta} \tag{1.20}$$

$$V(\eta_{T+1,T+L}|\boldsymbol{\eta}) = \sigma_\eta^2 M_{22}^{-1}. \tag{1.21}$$

Returning to unemployment rates, we examine the posteriors of the four observations held out for prediction. For example, Table 1.8, 1.9, 1.10 and Figures 1.1, 1.2 give the forecasting features of the future data of Group 1 under our proposed model. The detailed results and conclusion for the rest of the four groups and for four regions are given in Chapter 1.3.

1.3 UE Rate Data Analysis by Group and Region

For time series model selection, many criteria exist. For each fitted model, we report the Akaike Information Criterion (AIC; Akaike, 1998), Bayesian Information Criterion (BIC; Schwarz, 1978) defined by

$$AIC = (-2) \times \text{log maximum likelihood} + 2 \times (\text{number of parameters}) \tag{1.22}$$

$$BIC = (-2)\times\text{log maximum likelihood}+\text{number of parameters}\times log(\text{the number of observations}) \tag{1.23}$$

as the measure of the fit of a model constructed with parameters determined using the maximum likelihood approach. For each group/region, N states share the same β_0, and each state has its own unique covariate coefficient β_i, $i = 1, \cdots, N$, p autoregressive coefficients, and q moving average coefficients. The number of parameters, in this case, is then $N \times (p + q + 1) + 1$.

To provide the most accurate forecast for future observations, it is crucial to select the optimal model from among all candidate models. We compare the forecasting accuracy between our Bayesian method and the frequentist method of Wilms et al., 2021 using **Bigtime** R package in terms of Mean Absolute Prediction Errors (MAPE) for the year 2018.

In fact, for each state i within a Group (or a Region) and for each time series model fitted to the UE rate series of state i, based on our Bayesian framework we compute the forecast of UE rates for the time points $t = T + 1, T + 2, T + 3,$ and $T + 4$, where $T = 112$ (2017:Q4). More specifically, suppose N is the number of states in a Group (or a Region), then, for each state $i = 1, \cdots, N$ within a Group (or a Region) whose UE rates are modeled by a time series we compute

$$\text{MAPE}_i = \frac{1}{4} \sum_{t=T+1}^{T+4} \frac{|\hat{y}_{i,t} - y_{i,t}|}{y_{i,t}} \times 100 \tag{1.24}$$

where $\hat{y}_{i,t}$ denotes the forecasted UE rate and $y_{i,t}$ denotes the observed UE rate for state i at time t. Then, for each Group (or Region) and for each time series model fitted to the UE rate series of states within the Group or Region, we compute the Total MAPE defined by

$$\text{Total MAPE} = \sum_{i=1}^{N} \text{MAPE}_i, \tag{1.25}$$

where N is the number of states within the Group or the Region. Incidentally, the number of states in one Group or Region need not be same as those in another Group or Region. While Tables 1.3 to 1.7 and Tables 1.23 to 1.26 report the Total MAPE values for the Groups and Regions, respectively, the Tables 1.8 to 1.22 give the MAPE_i values for each state i within a Group that is modeled by a time series. Similar such values are given for each stte within a Region.

1.3.1 Data Analysis by Group

For each Group, after fitting each time series model we compute the AIC and BIC values. In addition, we also determine the forecasts of the UE rates for 2018:Q1 to 2018:Q4 and compute the Total MAPE values. These are reported in Tables 1.3, 1.4, 1.5, 1.6 and 1.7, where the second columns are the acceptance rates for the MH algorithm. Gelman et al., 1997 published theoretical results about the optimal scaling problem for MH algorithms with Gaussian proposals. They suggested that the acceptance rate optimizing the efficiency of the process approaches 0.234. The acceptance rates of six different time series structures of five groups range from 0.144 to 0.419, which are acceptable. The third and fourth columns are AIC and BIC values for each time series structure, which are used to evaluate and make comparisons among the six models. The smaller AIC/BIC value indicates a better model fit. The last column of the tables display the Total MAPE values for N states within each group, where the UE rates of the states within the Group are modeled by a time series model. In general, the Total MAPE values based on our Bayesian framework perform better than those based on the **Bigtime** method.

Table 1.3: Group 1: Information Criteria and Total MAPE for six time series models and MAPE for Bigtime

Group 1 (N=10)	Accept rate	AIC	BIC	Total MAPE
AR(0)	0.225	850.47	905.60	65.33 (4)
AR(1)	0.230	177.15	237.30	70.41 (6)
MA(1)	0.326	56.95	117.09	61.49 (1)*
AR(2)	0.295	-81.39	-16.24	62.82 (2)
MA(2)	0.419	-137.02	-71.86	63.17 (3)
ARMA(1,1)	0.230	-142.80	-77.65	66.63 (5)**
Bigtime				81.30 [RM = 32.22%]

Table 1.4: Group 2: Information Criteria and Total MAPE for six time series models and MAPE for Bigtime

Group 2 (N=13)	Accept rate	AIC	BIC	Total MAPE
AR(0)	0.172	996.83	1070.68	85.00 (6)
AR(1)	0.174	200.57	279.68	55.65 (1)*
MA(1)	0.263	95.61	174.73	70.86 (5)
AR(2)	0.236	41.26	125.65	62.28 (3)
MA(2)	0.188	-37.90	46.49	65.16 (4)
ARMA(1,1)	0.333	-75.42	8.98	60.86 (2) **
Bigtime				78.29 [RM = 40.68%]

Table 1.5: Group 3: Information Criteria and Total MAPE for six time series models and MAPE for Bigtime

Group 3 (N=10)	Accept rate	AIC	BIC	Total MAPE
AR(0)	0.229	905.24	960.37	60.69 (6)
AR(1)	0.232	-29.74	30.40	47.74 (1)*
MA(1)	0.322	14.31	74.45	52.98 (5)
AR(2)	0.298	-176.99	-111.83	49.60 (2)
MA(2)	0.413	-223.61	-158.45	52.35 (4)
ARMA(1,1)	0.230	-320.60	-255.44	51.04 (3)**
Bigtime				64.01 [RM = 34.08%]

Table 1.6: Group 4: Information Criteria and Total MAPE for six time series models and MAPE for Bigtime

Group 4 (N=13)	Accept rate	AIC	BIC	Total MAPE
AR(0)	0.171	1010.40	1084.24	92.14 (6)
AR(1)	0.177	-71.58	7.54	55.68 (2)
MA(1)	0.261	-133.91	-54.79	71.84 (5)
AR(2)	0.236	-381.03	-296.64	66.41 (4)
MA(2)	0.369	-424.06	-339.67	63.37 (3)
ARMA(1,1)	0.175	-495.81	-411.41	54.36 (1)*/**
Bigtime				77.66 [RM = 42.86%]

Table 1.7: Group 5: Information Criteria and Total MAPE for six time series models and MAPE for Bigtime

Group 5 (N=5)	Accept rate	AIC	BIC	Total MAPE
AR(0)	0.267	544.27	570.18	38.90 (6)
AR(1)	0.144	184.76	215.00	26.85 (2)
MA(1)	0.145	146.78	177.01	33.38 (5)
AR(2)	0.182	14.86	49.41	26.13 (1)*
MA(2)	0.187	29.62	64.17	30.42 (4)
ARMA(1,1)	0.389	0.37	34.99	27.19 (3)**
Bigtime				34.74 [RM=32.95%]

In Tables 1.3 to 1.7, the orders of time series models based on the AIC and BIC values are consistent. Also, the ARMA(1,1) model seems to provide the best AIC/BIC value for each group. However, ARMA (1,1) does not always produce the lowest Total MAPE value. In the last column of the tables, we indicate the smallest AIC/BIC value by '**' and the smallest Total MAPE value by '*'. In terms of the Total MAPE values, except for AR(0) model, which indicates that no dependence structure exists in the model, the remaining time series models (see Tables 1.3 to 1.7) outperform the **Bigtime** method.

The last row of the Tables 1.3, 1.4, 1.5, 1.6 and 1.7 provides a Relative Measure (RM) in % that is based on the ratio of MAPE value for the **Bigtime** method to the smallest Total MAPE value among the six time series models. More specifically, the Relative Measure of performance is defined as

$$\text{Relative Measure (RM)} = \left(\frac{\text{MAPE for Bigtime}}{\text{Smallest Total MAPE value}} - 1 \right) \times 100. \tag{1.26}$$

Among the five groups, the smallest RM is 32.22% for Group 1 with 10 states, and the largest RM is 42.86% for Group 4 with 13 states. This says that the MAPE value for the **Bigtime** method for each Group is always a significant percentage larger than the smallest Total MAPE for Bayesian ARMA model. Thus, in terms of the forecast accuracy as measured by the MAPE values, our Bayesian approach outperforms **Bigtime** by a large margin.

Figures 1.1 to 1.9 depict the four-step (2018:Q1 to 2018:Q4) UE rate forecasts for each state under various time series models compared to the actual UE rate for the last four quarters, where the solid black line represents the actual UE rate. In these figures, the red dotted line represents the AR(0) model, the green dotted line represents the AR(1) model, the blue dotted line represents the MA(1) model, the light-blue dotted line represents the AR(2) model, the purple dotted line represents the MA(2) model, the yellow dotted line represents the ARMA(1,1) model, and the grey dotted line represents the **bigtime** model. We additionally marked the model with the least MAPE and AIC/BIC values with a '*' and a '**', respectively. The tables from 1.8 to 1.22 provide the absolute forecast error relative to the actual value for each step for each state and the average $MAPE_i$ value across four steps for each state. Details are given below.

Detailed results for Group 1

For Group 1, MA(1) fits the best in terms of Total MAPE value across the 10 states within the group. For each individual state, the best model differs. For example, for state 1, Maine (ME), in Group 1, MA(2) model has the smallest MAPE value (3.213) among the six models; see Table 1.10. However, MA(2) does not perform as well in forecasting UE rates of Mississippi (MS) and Hawaii (HI) as the MAPE values are 14.02 and 13.66, respectively. For the state of Mississippi (MS), the MAPE value for AR(0) model is the smallest among all the six models. A closer look at Figure 1.1 shows that the actual UE rate of Mississippi increases from the first quarter of 2018 to the last quarter of 2018, and AR(0) without any dependence structure seems to produce the closest forecast of the actual UE rate. **Bigtime** method results are summarized in Appendix Tables from A.1 to A.5 for comparison purposes.

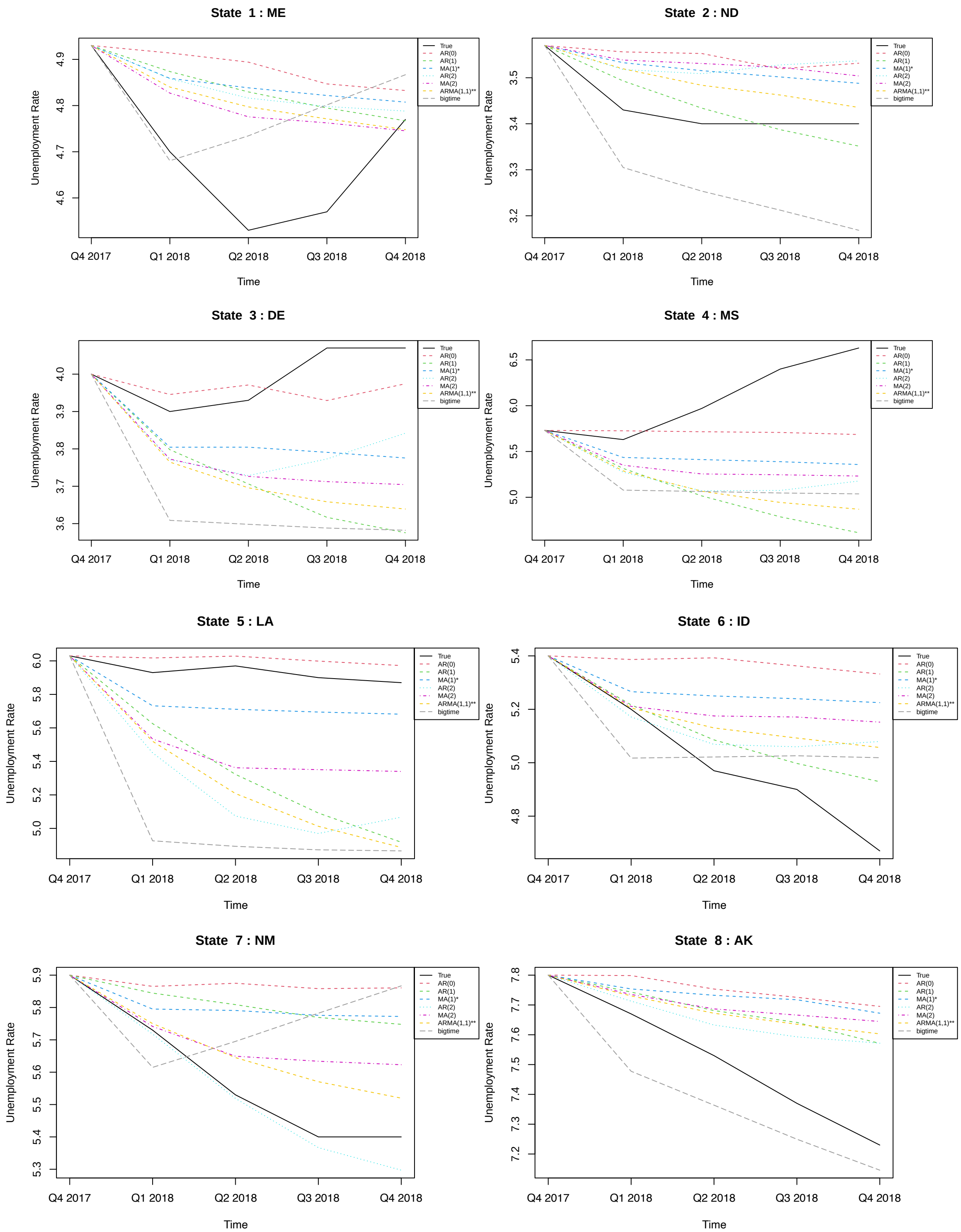


Figure 1.1: Group 1: Quarterly Forecasts vs actual UE rates (1/2)

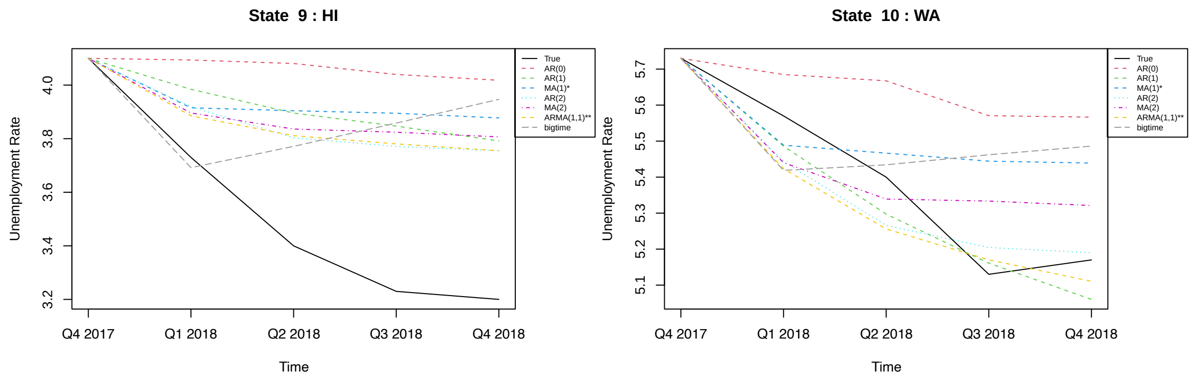


Figure 1.2: Group 1: Quarterly Forecasts vs actual UE rates (cont. 2/2)

Table 1.8: Four steps ahead Forecast Results for Group 1 (1/3)

	AR(0)					AR(1)				
state	step1	step2	step3	step4	MAPE	step1	step2	step3	step4	MAPE
1:ME	4.550	8.036	6.059	1.312	4.989	3.695	6.614	4.917	0.062	3.822
2:ND	3.679	4.489	3.514	3.869	3.888	1.836	0.999	0.380	1.430	1.161
3:DE	1.169	1.043	3.455	2.351	2.005	2.620	5.693	11.135	12.148	7.899
4:MS	1.704	4.260	10.815	14.249	7.757	5.597	15.994	25.211	30.426	19.307
5:LA	1.478	0.980	1.674	1.740	1.468	5.137	10.845	13.700	16.227	11.477
6:ID	3.585	8.503	9.436	14.179	8.926	0.315	2.322	1.986	5.545	2.542
7:NM	2.361	6.237	8.487	8.530	6.404	1.998	5.042	6.838	6.444	5.080
8:AK	1.677	2.966	4.824	6.434	3.975	0.954	2.008	3.682	4.715	2.840
9:HI	9.751	20.017	25.071	25.561	20.100	6.819	14.577	19.098	18.489	14.746
10:WA	2.064	4.954	8.590	7.670	5.819	1.510	1.904	0.605	2.122	1.535
Sum	32.018	61.485	81.925	85.895	65.331	30.481	65.998	87.552	97.608	70.409

Table 1.9: Four steps ahead Forecast Results for Group 1 (2/3)

	MA(1)					AR(2)				
state	step1	step2	step3	step4	MAPE	step1	step2	step3	step4	MAPE
1:ME	3.381	6.809	5.511	0.787	4.122	3.363	6.318	4.990	0.379	3.763
2:ND	2.994	3.395	2.992	2.587	2.992	2.559	3.200	3.760	4.024	3.386
3:DE	2.448	3.191	6.860	7.232	4.933	3.301	5.125	7.296	5.601	5.331
4:MS	3.464	9.353	15.803	19.191	11.953	6.329	15.057	20.706	21.911	16.001
5:LA	3.341	4.334	3.480	3.205	3.590	8.077	15.011	15.768	13.678	13.134
6:ID	1.270	5.634	6.930	11.883	6.429	0.553	1.985	3.258	8.768	3.641
7:NM	1.138	4.715	6.958	6.895	4.926	0.212	0.201	0.615	1.904	0.733
8:AK	1.091	2.689	4.715	6.118	3.653	0.557	1.363	3.026	4.727	2.418
9:HI	4.966	14.832	20.587	21.167	15.388	5.077	11.867	16.752	17.338	12.758
10:WA	1.460	1.233	6.128	5.202	3.506	2.275	2.484	1.452	0.388	1.650
Sum	25.553	56.185	79.964	84.267	61.492	32.303	62.611	77.623	78.718	62.815

Table 1.10: Four steps ahead Forecast Results for Group 1 (3/3)

	MA(2)					ARMA(1,1)				
state	step1	step2	step3	step4	MAPE	step1	step2	step3	step4	MAPE
1:ME	2.702	5.422	4.213	0.516	3.213	2.980	5.899	4.403	0.463	3.436
2:ND	3.163	3.851	3.611	3.055	3.420	2.605	2.460	1.835	1.042	1.986
3:DE	3.264	5.180	8.785	8.982	6.553	3.472	5.959	10.113	10.589	7.533
4:MS	4.980	11.977	18.034	21.087	14.020	6.080	15.144	22.755	26.552	17.633
5:LA	6.683	10.185	9.308	9.028	8.801	6.913	12.765	15.041	16.758	12.869
6:ID	0.208	4.127	5.533	10.321	5.047	0.093	3.224	3.925	8.288	3.883
7:NM	0.199	2.159	4.331	4.131	2.705	0.353	2.081	3.159	2.219	1.953
8:AK	0.850	2.081	4.017	5.734	3.171	0.787	1.888	3.599	5.160	2.859
9:HI	4.448	12.825	18.394	18.955	13.656	4.169	12.094	17.047	17.344	12.664
10:WA	2.331	1.129	3.966	2.922	2.587	2.635	2.666	0.786	1.148	1.809
Sum	28.828	58.936	80.192	84.731	63.173	30.087	64.180	82.663	89.563	66.625

Detailed results for Group 2

For Group 2, AR(1) is the best model based on Total MAPE value across the 13 states in the group. For each individual state, however, the best model differs. Detail MAPEs are shown in Tables 1.11, 1.12 and 1.13.

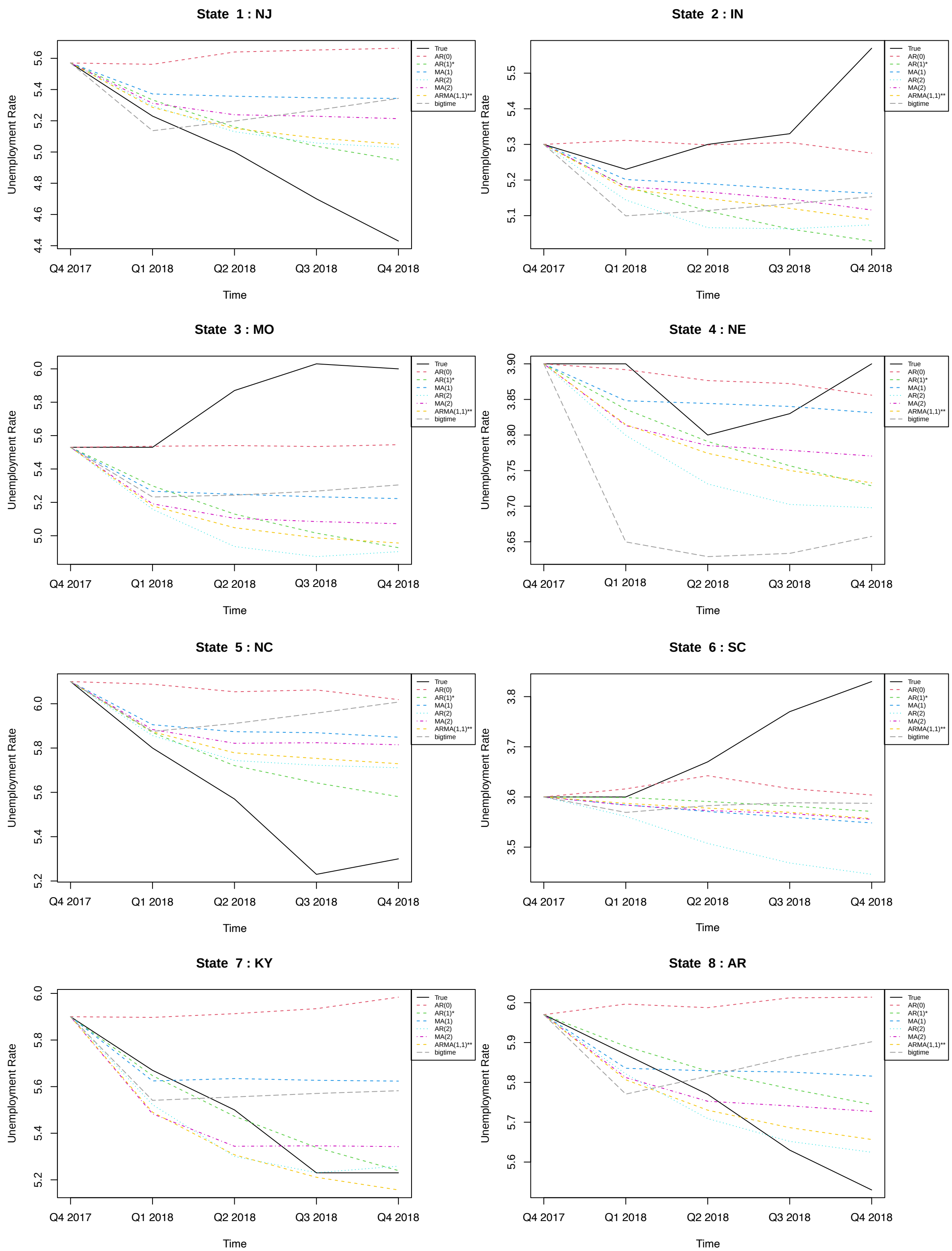


Figure 1.3: Group 2: Quarterly Forecasts vs actual UE rates (1/2)

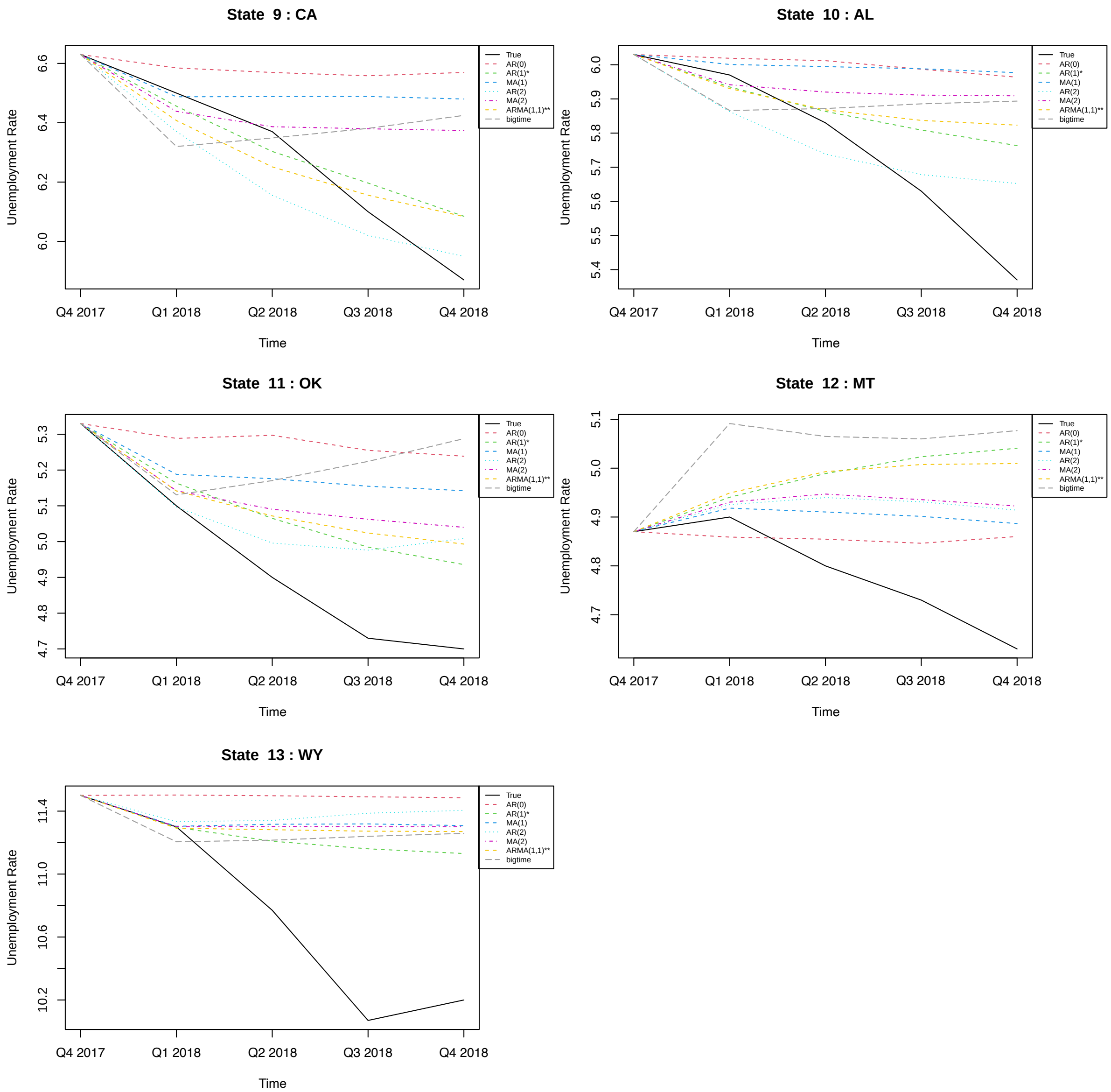


Figure 1.4: Group 2: Quarterly Forecasts vs actual UE rates (cont. 2/2)

Table 1.11: Four steps ahead Forecast Results for Group 2 (1/3)

	AR(0)					AR(1)				
state	step1	step2	step3	step4	MAPE	step1	step2	step3	step4	MAPE
1:NJ	6.341	12.807	20.279	27.869	16.824	1.978	3.215	7.170	11.703	6.016
2:IN	1.558	0.028	0.460	5.286	1.833	0.907	3.533	5.019	9.716	4.794
3:MO	0.112	5.618	8.217	7.574	5.380	4.188	12.610	16.823	17.861	12.870
4:NE	0.209	2.013	1.104	1.129	1.114	1.637	0.239	1.912	4.394	2.046
5:NC	4.966	8.689	15.910	13.559	10.781	1.210	2.704	7.891	5.298	4.276
6:SC	0.446	0.754	4.065	5.904	2.792	0.031	2.153	4.987	6.756	3.482
7:KY	4.004	7.510	13.471	14.414	9.850	0.406	0.495	2.080	0.183	0.791
8:AR	2.156	3.771	6.789	8.750	5.367	0.351	1.012	2.741	3.881	1.996
9:CA	1.305	3.131	7.514	11.918	5.967	0.679	1.049	1.585	3.657	1.742
10:AL	0.822	3.122	6.340	11.055	5.335	0.587	0.574	3.177	7.326	2.916
11:OK	3.696	8.105	11.101	11.463	8.591	1.242	3.369	5.383	5.016	3.752
12:MT	0.838	1.139	2.459	4.970	2.352	0.825	3.937	6.200	8.873	4.959
13:WY	1.790	6.758	14.111	12.599	8.815	0.027	4.075	10.830	9.123	6.014
Sum	28.243	63.445	111.820	136.490	85.001	14.068	38.965	75.798	93.787	55.654

Table 1.12: Four steps ahead Forecast Results for Group 2 (2/3)

	MA(1)					AR(2)				
state	step1	step2	step3	step4	MAPE	step1	step2	step3	step4	MAPE
1:NJ	2.713	7.143	13.772	20.613	11.060	1.239	2.601	7.602	13.497	6.235
2:IN	0.545	2.083	2.912	7.313	3.213	1.645	4.409	5.009	8.903	4.992
3:MO	4.767	10.589	13.210	12.956	10.380	6.712	15.921	19.158	18.257	15.012
4:NE	1.330	1.167	0.266	1.759	1.131	2.576	1.809	3.330	5.183	3.224
5:NC	1.815	5.455	12.220	10.359	7.462	0.968	3.118	9.411	7.754	5.313
6:SC	0.439	2.698	5.578	7.355	4.018	1.073	4.431	8.000	10.032	5.884
7:KY	0.805	2.446	7.593	7.527	4.593	2.599	3.630	0.010	0.551	1.698
8:AR	0.590	1.025	3.482	5.169	2.566	0.830	1.056	0.389	1.711	0.997
9:CA	0.192	1.855	6.370	10.392	4.702	1.994	3.369	1.309	1.354	2.006
10:AL	0.517	2.824	6.363	11.300	5.251	1.796	1.572	0.860	5.258	2.372
11:OK	1.726	5.629	8.980	9.411	6.436	0.061	1.958	5.203	6.571	3.448
12:MT	0.371	2.299	3.624	5.542	2.959	0.519	2.911	4.252	6.134	3.454
13:WY	0.032	5.075	12.401	10.863	7.093	0.298	5.300	13.076	11.813	7.622
Sum	15.842	50.288	96.771	120.559	70.864	22.310	52.085	77.609	97.018	62.257

Table 1.13: Four steps ahead Forecast Results for Group 2 (3/3)

	MA(2)					ARMA(1,1)				
state	step1	step2	step3	step4	MAPE	step1	step2	step3	step4	MAPE
1:NJ	1.566	4.771	11.253	17.686	8.819	1.073	3.032	8.287	13.974	6.592
2:IN	0.932	2.522	3.439	8.155	3.762	1.054	2.865	3.930	8.634	4.121
3:MO	6.127	13.041	15.672	15.465	12.576	6.352	14.002	17.292	17.394	13.760
4:NE	2.222	0.391	1.344	3.322	1.820	2.189	0.676	2.082	4.288	2.309
5:NC	1.441	4.514	11.359	9.711	6.756	1.266	3.743	10.004	8.100	5.778
6:SC	0.451	2.642	5.388	7.169	3.912	0.355	2.513	5.318	7.128	3.828
7:KY	3.292	2.836	2.218	2.154	2.625	3.168	3.502	0.366	1.405	2.110
8:AR	0.985	0.302	1.972	3.566	1.706	1.081	0.689	1.003	2.291	1.266
9:CA	0.948	0.261	4.584	8.581	3.593	1.424	1.865	0.913	3.641	1.961
10:AL	0.476	1.547	4.994	10.038	4.264	0.665	0.651	3.679	8.439	3.358
11:OK	0.824	3.887	7.026	7.232	4.742	0.813	3.505	6.217	6.237	4.193
12:MT	0.622	3.063	4.347	6.314	3.586	1.004	4.018	5.866	8.197	4.771
13:WY	0.008	4.945	12.228	10.799	6.995	0.073	4.753	11.943	10.492	6.815
Sum	19.894	44.722	85.824	110.192	65.156	20.517	45.814	76.900	100.220	60.862

Detailed results for Group 3

For Group 3, AR(1) fits the best in terms of the Total MAPE value (47.74) across 10 states. For each individual state, the best model differs. For instance, for state 1, Connecticut (CT), AR(2) model has a smallest MAPE value (0.538) among all other models; see Table 1.15). However, AR(2) does not perform as well in predicting the UE rates for Utah (UT). For the state of Michigan (MI), AR(0) prediction is the closest to the actual UE rate. From Figure 1.6, the actual UE rate of Michigan goes up from the first quarter of 2018 to the last quarter of 2018, and AR(0), without any dependence structure, is the closest to the actual UE rate. **Bigtime** method results are summarized in Appendix Table A.3.

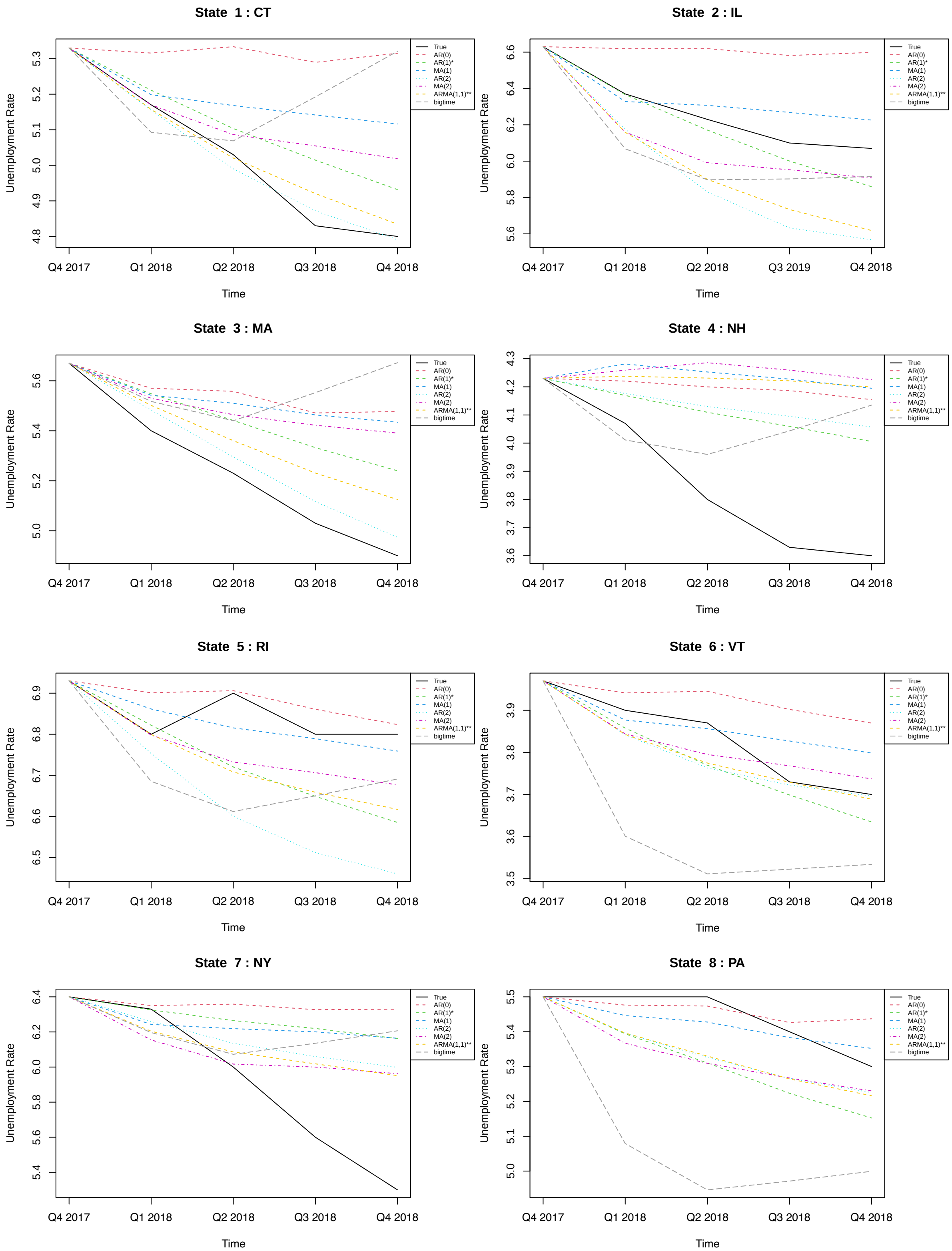
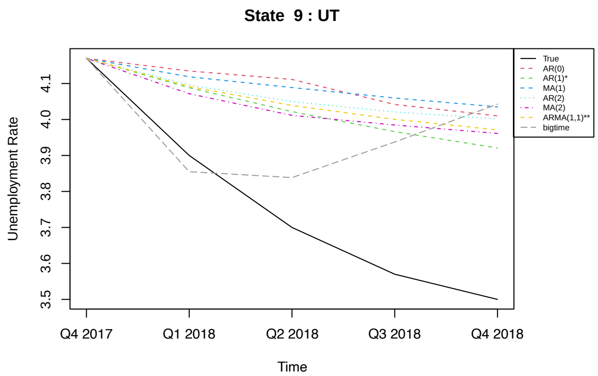


Figure 1.5: Group 3: Quarterly Forecasts vs actual UE rates (1/2)



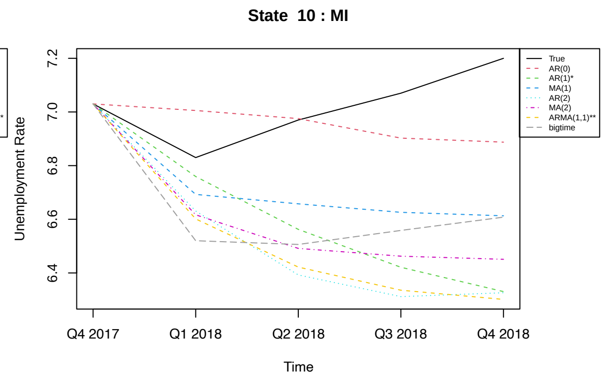


Figure 1.6: Group 3: Quarterly Forecasts vs actual UE rates (cont. 2/2)

Table 1.14: Four steps ahead Forecast Results for Group 3 (1/3)

	AR(0)					AR(1)				
state	step1	step2	step3	step4	MAPE	step1	step2	step3	step4	MAPE
1:CT	2.818	6.036	9.516	10.731	7.275	0.795	1.464	3.822	2.745	2.207
2:IL	3.915	6.251	7.893	8.705	6.691	0.059	0.947	1.625	3.466	1.524
3:MA	3.153	6.265	8.771	11.777	7.492	2.739	4.040	6.012	6.937	4.932
4:NH	3.695	10.524	15.337	15.405	11.240	2.440	8.147	11.841	11.273	8.425
5:RI	1.484	0.088	0.895	0.346	0.703	0.322	2.603	2.223	3.156	2.076
6:VT	1.060	1.945	4.614	4.584	3.051	1.085	2.595	0.836	1.760	1.569
7:NY	0.322	5.972	12.984	19.424	9.675	0.070	4.414	11.066	16.265	7.954
8:PA	0.431	0.480	0.487	2.585	0.996	1.928	3.449	3.271	2.785	2.858
9:UT	6.021	11.117	13.215	14.572	11.231	4.823	8.700	11.097	12.026	9.162
10:MI	2.569	0.076	2.368	4.341	2.338	1.035	5.848	9.178	12.077	7.035
Sum	25.468	48.754	76.080	92.470	60.692	15.296	42.207	60.971	72.490	47.742

Table 1.15: Four steps ahead Forecast Results for Group 3 (2/3)

	MA(1)					AR(2)				
state	step1	step2	step3	step4	MAPE	step1	step2	step3	step4	MAPE
1:CT	0.552	2.746	6.450	6.588	4.084	0.298	0.791	0.870	0.193	0.538
2:IL	0.660	1.240	2.759	2.572	1.808	3.058	6.405	7.661	8.276	6.350
3:MA	2.623	5.356	8.615	10.900	6.874	1.532	1.249	1.720	1.505	1.502
4:NH	5.192	11.915	16.457	16.527	12.523	2.611	8.683	12.825	12.699	9.204
5:RI	0.904	1.224	0.158	0.602	0.722	0.688	4.341	4.235	4.996	3.565
6:VT	0.594	0.353	2.604	2.663	1.553	1.490	2.762	0.202	0.102	1.139
7:NY	1.376	3.647	10.724	16.298	8.011	1.112	2.260	8.212	13.172	6.189
8:PA	0.977	1.315	0.314	0.984	0.897	1.902	3.157	2.475	1.410	2.236
9:UT	5.604	10.508	13.717	15.288	11.279	5.027	9.454	12.636	14.339	10.364
10:MI	2.011	4.482	6.281	8.154	5.232	2.916	8.286	10.725	12.136	8.516
Sum	20.493	42.786	68.079	80.576	52.983	20.634	47.388	61.561	68.828	49.603

Table 1.16: Four steps ahead Forecast Results for Group 3 (3/3)

	MA(2)					ARMA(1,1)				
state	step1	step2	step3	step4	MAPE	step1	step2	step3	step4	MAPE
1:CT	0.002	1.123	4.649	4.547	2.580	0.250	0.189	1.874	0.718	0.758
2:IL	3.312	3.827	2.417	2.671	3.057	3.319	5.299	5.999	7.440	5.514
3:MA	2.401	4.480	7.793	10.015	6.172	1.878	2.493	3.994	4.586	3.238
4:NH	4.646	12.781	17.336	17.371	13.034	4.115	11.345	16.293	16.645	12.099
5:RI	0.031	2.427	1.374	1.811	1.411	0.006	2.782	2.075	2.689	1.888
6:VT	1.441	1.936	1.022	1.001	1.350	1.470	2.461	0.025	0.301	1.064
7:NY	2.774	0.281	7.139	12.497	5.673	1.987	1.433	7.470	12.283	5.793
8:PA	2.424	3.466	2.461	1.319	2.418	1.891	3.089	2.509	1.581	2.268
9:UT	4.400	8.419	11.612	13.180	9.403	4.914	9.160	12.055	13.463	9.898
10:MI	3.124	6.866	8.593	10.405	7.247	3.349	7.865	10.387	12.481	8.520
Sum	24.555	45.606	64.396	74.817	52.345	23.179	46.116	62.681	72.187	51.040

Detailed results for Group 4

For Group 4, ARMA(1,1) is the best model in terms of both the Total MAPE value and AIC/BIC value across 13 states. For each individual state, the best model differs. For instance, although, ARMA(1,1) is the best model for Kansas (KS), Minnesota (MN), and West Virginia (WV), it is not the best fit for the remaining states.

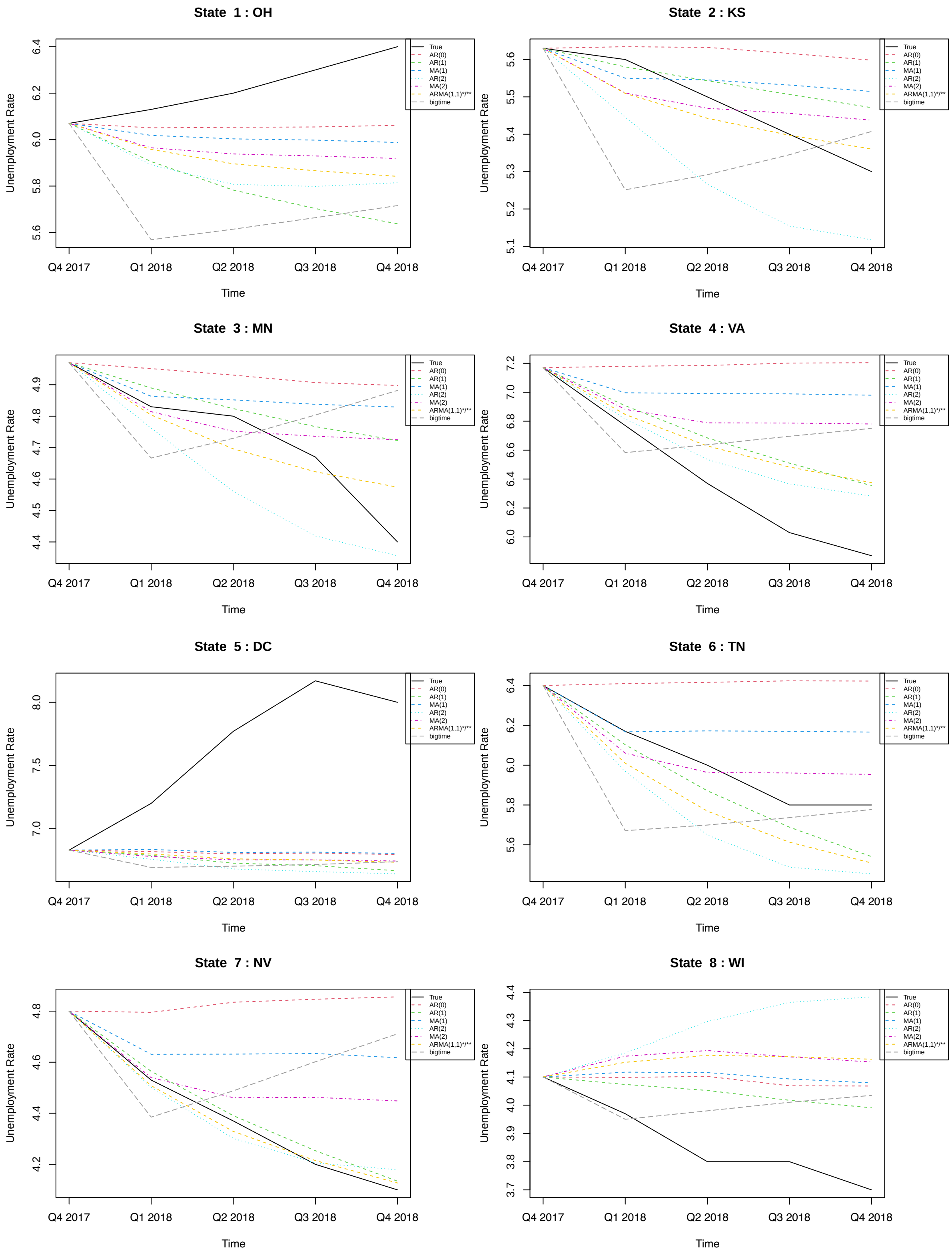


Figure 1.7: Group 4: Quarterly Forecasts vs actual UE rates (1/2)

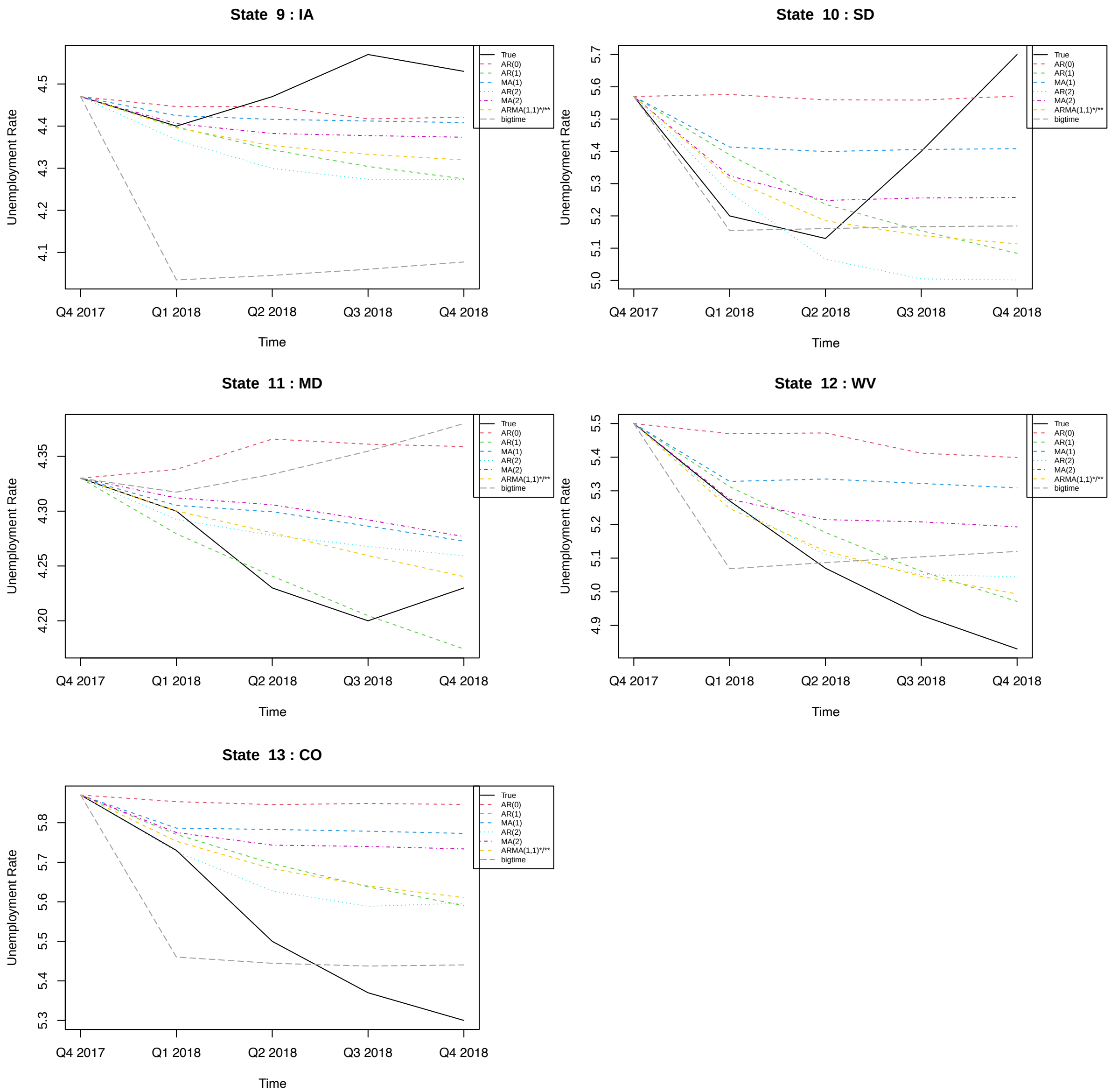


Figure 1.8: Group 4: Quarterly Forecasts vs actual UE rates(cont. 2/2)

Table 1.17: Four steps ahead Forecast Results for Group 4 (1/3)

	AR(0)					AR(1)				
state	step1	step2	step3	step4	MAPE	step1	step2	step3	step4	MAPE
1:OH	1.291	2.371	3.897	5.289	3.212	3.662	6.729	9.472	11.910	7.943
2:KS	0.613	2.407	4.006	5.628	3.164	0.352	0.785	1.972	3.229	1.584
3:MN	2.508	2.721	5.070	11.309	5.402	1.236	0.501	2.072	7.334	2.786
4:VA	6.050	12.797	19.430	22.732	15.252	2.025	4.939	7.962	8.258	5.796
5:DC	5.308	12.482	16.684	15.068	12.386	5.708	13.428	17.911	16.665	13.428
6:TN	3.884	6.934	10.752	10.732	8.076	1.092	2.142	1.918	4.471	2.406
7:NV	5.855	10.635	15.396	18.437	12.581	0.757	0.476	1.272	0.832	0.834
8:WI	3.235	7.951	7.075	9.945	7.052	2.609	6.641	5.723	7.863	5.709
9:IA	1.056	0.524	3.346	2.402	1.832	0.047	2.826	5.815	5.634	3.581
10:SD	7.235	8.374	2.943	2.260	5.203	3.637	2.063	4.556	10.805	5.265
11:MD	0.886	3.209	3.835	3.048	2.744	0.480	0.254	0.112	1.316	0.540
12:WV	3.794	7.929	9.771	11.782	8.319	0.804	2.094	2.653	2.923	2.118
13:CO	2.151	6.286	8.913	10.306	6.914	0.705	3.578	4.987	5.478	3.687
Sum	43.866	84.620	111.118	128.938	92.137	23.114	46.456	66.425	86.718	55.677

Table 1.18: Four steps ahead Forecast Results for Group 4 (2/3)

	MA(1)					AR(2)				
state	step1	step2	step3	step4	MAPE	step1	step2	step3	step4	MAPE
1:OH	1.824	3.173	4.797	6.443	4.059	3.898	6.333	7.959	9.149	6.835
2:KS	0.893	0.826	2.439	4.047	2.051	2.752	4.253	4.553	3.440	3.749
3:MN	0.687	1.078	3.593	9.750	3.777	1.404	4.991	5.375	0.996	3.192
4:VA	3.347	9.750	15.902	18.906	11.976	0.679	2.598	5.588	7.024	3.972
5:DC	5.067	12.337	16.600	14.955	12.240	6.142	14.018	18.484	16.977	13.905
6:TN	0.040	2.869	6.380	6.315	3.901	3.266	5.851	5.385	5.964	5.117
7:NV	2.232	5.987	10.333	12.626	7.795	0.638	1.566	0.132	1.925	1.065
8:WI	3.709	8.306	7.715	10.242	7.493	5.418	13.065	14.849	18.486	12.954
9:IA	0.563	1.208	3.454	2.686	1.978	0.749	3.817	6.485	5.669	4.180
10:SD	4.105	5.251	0.110	5.118	3.646	1.400	1.247	7.318	12.252	5.554
11:MD	0.122	1.641	2.056	1.011	1.208	0.176	1.134	1.616	0.694	0.905
12:WV	1.100	5.229	7.949	9.913	6.048	0.190	0.799	2.461	4.439	1.972
13:CO	0.983	5.145	7.608	8.923	5.665	0.056	2.320	4.070	5.597	3.011
Sum	24.672	62.800	88.936	110.935	71.837	26.768	61.992	84.275	92.612	66.411

Table 1.19: Four steps ahead Forecast Results for Group 4 (3/3)

	MA(2)					ARMA(1,1)				
state	step1	step2	step3	step4	MAPE	step1	step2	step3	step4	MAPE
1:OH	2.696	4.221	5.879	7.519	5.079	2.809	4.904	6.887	8.721	5.830
2:KS	1.598	0.554	1.038	2.595	1.446	1.618	1.042	0.050	1.142	0.963
3:MN	0.326	1.005	1.417	7.386	2.534	0.543	2.163	1.003	3.960	1.917
4:VA	1.654	6.571	12.557	15.513	9.074	1.135	4.068	7.514	8.605	5.330
5:DC	5.837	13.094	17.349	15.706	12.996	5.566	12.988	17.362	15.830	12.936
6:TN	1.775	0.609	2.773	2.648	1.951	2.595	3.845	3.218	5.022	3.670
7:NV	0.209	2.084	6.243	8.498	4.258	0.486	0.942	0.352	0.659	0.610
8:WI	5.131	10.361	9.773	12.247	9.378	4.585	9.918	9.787	12.515	9.201
9:IA	0.136	1.959	4.213	3.452	2.440	0.095	2.605	5.187	4.643	3.132
10:SD	2.384	2.299	2.672	7.769	3.781	2.210	1.074	4.835	10.290	4.602
11:MD	0.282	1.793	2.192	1.107	1.343	0.005	1.188	1.415	0.244	0.713
12:WV	0.096	2.846	5.634	7.512	4.022	0.432	1.008	2.346	3.385	1.793
13:CO	0.788	4.424	6.888	8.183	5.071	0.404	3.338	5.035	5.857	3.659
Sum	22.912	51.820	78.628	100.135	63.373	22.483	49.083	64.991	80.873	54.356

Detailed results for Group 5

For Group 5, AR(2) fits the best in terms of the Total MAPE value (26.13) across 5 states. For each individual state, the best model differs. For instance, for state 1, Florida (FL), MA(2) model has a smallest MAPE value (1.538); see Table 1.22. However, MA(2) does not perform as well in predicting the states of Georgia (GA). **Bigtime** method results are summarized in Appendix Table A.5.

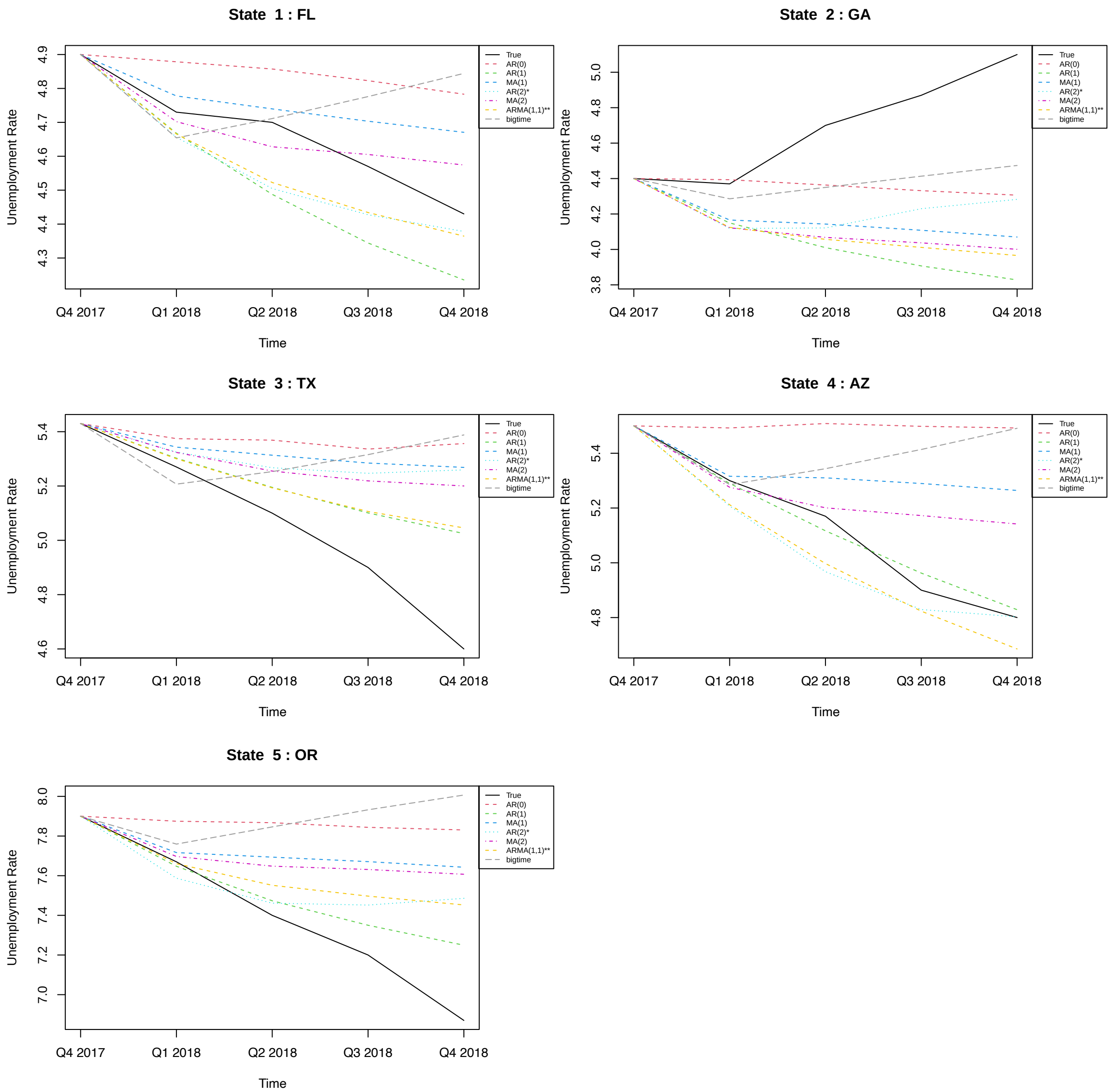


Figure 1.9: Group 5: Quarterly Forecasts vs actual UE rates

Table 1.20: Four steps ahead Forecast Results for Group 5 (1/3)

	AR(0)					AR(1)				
state	step1	step2	step3	step4	MAPE	step1	step2	step3	step4	MAPE
1:FL	3.141	3.347	5.535	7.968	4.998	1.314	4.518	4.940	4.394	3.792
2:GA	0.536	7.159	11.053	15.562	8.578	5.044	14.675	19.776	24.943	16.110
3:TX	1.981	5.270	8.898	16.440	8.147	0.616	1.862	4.099	9.238	3.954
4:AZ	3.631	6.548	12.212	14.422	9.203	0.190	1.030	1.273	0.597	0.772
5:OR	2.669	6.313	8.943	13.982	7.977	0.294	0.992	2.083	5.522	2.223
Sum	11.958	28.637	46.641	68.374	38.903	7.458	23.077	32.171	44.694	26.851

Table 1.21: Four steps ahead Forecast Results for Group 5 (2/3)

	MA(1)					AR(2)				
state	step1	step2	step3	step4	MAPE	step1	step2	step3	step4	MAPE
1:FL	1.010	0.844	2.917	5.431	2.550	1.566	4.153	3.116	1.168	2.501
2:GA	4.667	11.846	15.643	20.194	13.088	5.753	12.313	13.144	16.028	11.810
3:TX	1.388	4.186	7.848	14.532	6.988	1.039	3.281	7.077	14.335	6.433
4:AZ	0.300	2.712	7.951	9.672	5.159	1.750	3.915	1.442	0.054	1.790
5:OR	0.607	3.967	6.540	11.246	5.590	1.078	0.831	3.497	8.964	3.593
Sum	7.972	23.555	40.899	61.075	33.375	11.186	24.493	28.276	40.549	26.127

Table 1.22: Four steps ahead Forecast Results for Group 5 (3/3)

	MA(2)					ARMA(1,1)				
state	step1	step2	step3	step4	MAPE	step1	step2	step3	step4	MAPE
1:FL	0.590	1.532	0.773	3.255	1.538	1.361	3.774	2.974	1.477	2.397
2:GA	5.671	13.446	17.101	21.554	14.443	5.619	13.682	17.627	22.237	14.791
3:TX	1.017	3.039	6.506	13.047	5.902	0.571	1.824	4.208	9.689	4.073
4:AZ	0.458	0.595	5.561	7.120	3.434	1.661	3.342	1.564	2.388	2.239
5:OR	0.348	3.352	5.990	10.730	5.105	0.124	2.048	4.124	8.480	3.694
Sum	8.084	21.964	35.931	55.706	30.422	9.336	24.670	30.497	44.271	27.194

1.3.2 Data Analysis by Region

As in Section 1.3.1,for each Region, after fitting each time series model we compute the AIC and BIC values. In addition, we also determine the forecasts of the UE rates for 2018:Q1 to 2018:Q4 and compute the Total MAPE values. These are reported in Tables 1.23, 1.24, 1.25 and 1.26, where the second columns are the acceptance rate for the MH algorithm. The acceptance rates of six different time series structures of five groups range from 0.119 to 0.345, which are acceptable. The third and fourth columns are AIC and BIC values for each time series structure, which are used to evaluate and make comparisons among six models.

Table 1.23: Region 1: Information Criteria and Total MAPE for six time series models and MAPE for Bigtime

Region 1 (N=9)	Accept rate	AIC	BIC	Total MAPE
AR(0)	0.248	715.67	764.74	58.61 (6)
AR(1)	0.256	-130.37	-76.40	38.17 (3)
MA(1)	0.345	-76.63	-22.65	47.48 (5)
AR(2)	0.322	-251.40	-192.52	32.72 (1)*
MA(2)	0.441	-283.63	-224.75	42.99 (4)
ARMA(1,1)	0.322	-378.56	-319.68	36.38 (2)**
Bigtime				52.73 [RM = 51.16%]

Table 1.24: Region 2: Information Criteria and Total MAPE for six time series models and MAPE for Bigtime

Region 2 (N=12)	Accept rate	AIC	BIC	Total MAPE
AR(0)	0.248	944.92	1012.45	51.07 (2)
AR(1)	0.190	133.95	206.68	54.36 (3)
MA(1)	0.283	-11.41	61.32	48.10 (1)*
AR(2)	0.253	-100.65	-22.74	75.76 (6)
MA(2)	0.212	-188.16	-110.24	55.37 (4)
ARMA(1,1)	0.188	-220.64	-142.72	60.21 (5)**
Bigtime				74.67 [RM = 55.25%]

Table 1.25: Region 3: Information Criteria and Total MAPE for six time series models and MAPE for Bigtime

Region 3 (N=17)	Accept rate	AIC	BIC	Total MAPE
AR(0)	0.119	1351.09	1450.86	121.55 (6)
AR(1)	0.120	159.52	264.83	103.17 (2)
MA(1)	0.203	40.98	146.29	108.50 (5)
AR(2)	0.177	-207.90	-97.05	99.94 (1)*
MA(2)	0.300	-250.09	-139.24	104.52 (4)
ARMA(1,1)	0.121	-310.90	-200.04	103.81 (3)**
Bigtime				137.81 [RM = 37.89%]

Table 1.26: Region 4: Information Criteria and Total MAPE for six time series models and MAPE for Bigtime

Region 4 (N=13)	Accept rate	AIC	BIC	Total MAPE
AR(0)	0.171	1256.47	1330.32	106.04 (6)
AR(1)	0.173	251.41	330.53	53.51 (3)
MA(1)	0.264	167.89	247.00	78.92 (5)
AR(2)	0.233	-75.28	9.11	51.18 (1)*
MA(2)	0.345	-130.27	-45.88	63.45 (4)
ARMA(1,1)	0.174	-184.09	-99.70	53.45 (2)**
Bigtime				77.40 [RM = 51.23%]

Detailed results for Region 1

For Region 1, Northeast, unlike Section 1.3.1, was defined by the Census Bureau geographically. AR(2) fits the best in terms of Total MAPE value (32.72). The total MAPE value of **Bigtime** is 51.16% larger than the smallest total MAPE value of our Bayesian method. For each individual state, the best model differs. For instance, for state 1, Connecticut, AR(2) model has the smallest MAPE value (0.524); see Table 1.28). However, AR(2) does not perform as well in predicting UE rates of New Hampshire (NH). For the state of Pennsylvania (PA), AR(0) prediction is the closest to the actual unemployment rate. From Figure 1.11, the actual unemployment rate for Pennsylvania flattens from the first quarter of 2018 to the second quarter of 2018, and AR(0) without any dependence is the closest to the actual unemployment rate. **Bigtime** method results are summarized in Appendix Tables from A.6 to A.9 for comparison purposes.

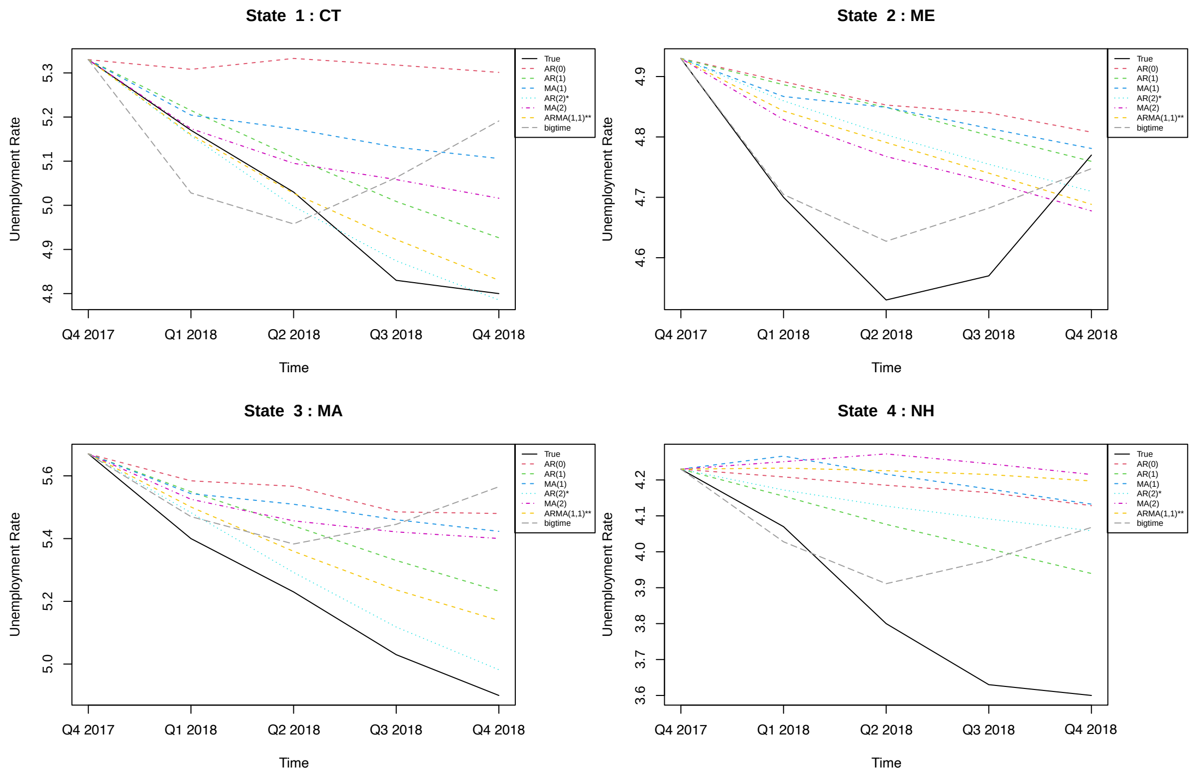


Figure 1.10: Region 1: Quarterly Forecasts vs actual UE rates (1/2)

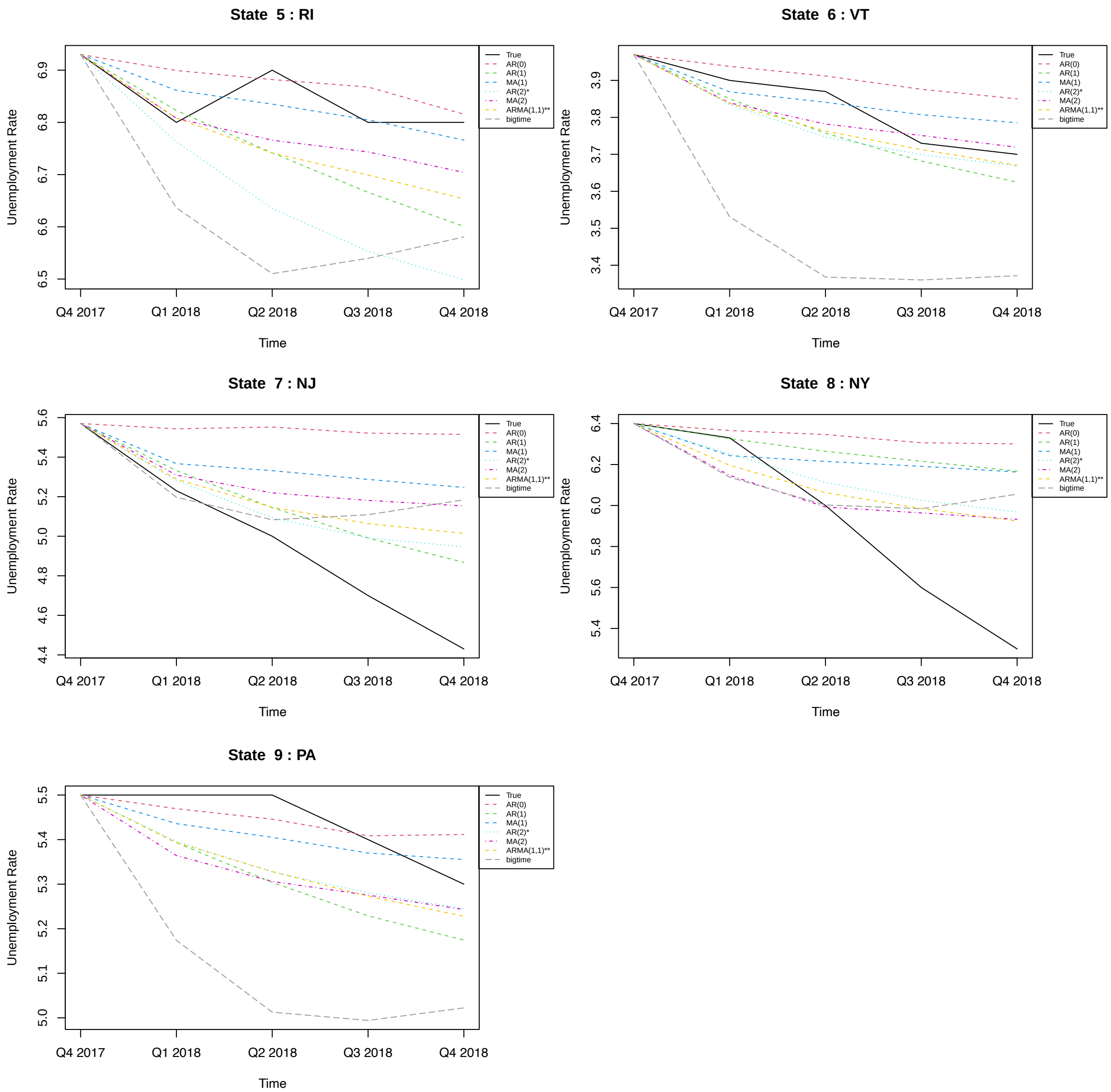


Figure 1.11: Region 1: Quarterly Forecasts vs actual UE rates (cont. 2/2)

Table 1.27: Four steps ahead Forecast Results for Region 1 (1/3)

	AR(0)					AR(1)				
state	step1	step2	step3	step4	MAPE	step1	step2	step3	step4	MAPE
1:CT	2.670	6.020	10.100	10.441	7.308	0.869	1.557	3.701	2.635	2.191
2:ME	4.077	7.124	5.910	0.795	4.476	3.967	7.044	5.085	0.218	4.078
3:MA	3.412	6.436	9.047	11.837	7.683	2.765	4.030	5.970	6.786	4.888
4:NH	3.401	10.139	14.734	14.681	10.739	2.086	7.273	10.426	9.430	7.304
5:RI	1.462	0.260	0.996	0.231	0.737	0.337	2.296	1.970	2.929	1.883
6:VT	0.972	1.084	3.914	4.057	2.507	1.287	2.910	1.290	2.034	1.880
7:NJ	5.989	11.045	17.481	24.493	14.752	1.954	2.889	6.197	9.877	5.229
8:NY	0.569	5.772	12.611	18.881	9.458	0.058	4.403	11.002	16.369	7.958
9:PA	0.559	0.986	0.158	2.103	0.952	1.948	3.562	3.168	2.368	2.761
Sum	23.111	48.866	74.951	87.519	58.612	15.271	35.964	48.809	52.646	38.172

Table 1.28: Four steps ahead Forecast Results for Region 1 (2/3)

	MA(1)					AR(2)				
state	step1	step2	step3	step4	MAPE	step1	step2	step3	step4	MAPE
1:CT	0.664	2.846	6.245	6.376	4.033	0.237	0.642	0.914	0.302	0.524
2:ME	3.553	7.043	5.345	0.229	4.042	3.400	6.038	4.043	1.261	3.686
3:MA	2.654	5.344	8.550	10.677	6.806	1.531	1.189	1.751	1.672	1.536
4:NH	4.825	10.949	15.007	14.794	11.394	2.506	8.609	12.712	12.737	9.141
5:RI	0.903	0.942	0.072	0.498	0.604	0.546	3.834	3.631	4.437	3.112
6:VT	0.789	0.756	2.071	2.309	1.481	1.675	3.184	0.821	0.854	1.634
7:NJ	2.621	6.635	12.498	18.436	10.047	1.001	1.864	6.197	11.660	5.181
8:NY	1.369	3.583	10.561	16.290	7.951	1.246	1.880	7.592	12.618	5.834
9:PA	1.166	1.726	0.559	1.042	1.123	1.896	3.130	2.210	1.039	2.069
Sum	18.544	39.824	60.908	70.651	47.481	14.038	30.370	39.871	46.580	32.717

Table 1.29: Four steps ahead Forecast Results for Region 1 (3/3)

	MA(2)					ARMA(1,1)				
state	step1	step2	step3	step4	MAPE	step1	step2	step3	step4	MAPE
1:CT	0.072	1.295	4.727	4.502	2.649	0.188	0.055	1.916	0.631	0.698
2:ME	2.738	5.245	3.411	1.939	3.333	3.041	5.758	3.722	1.711	3.558
3:MA	2.320	4.331	7.781	10.212	6.161	1.865	2.474	4.106	4.881	3.332
4:NH	4.443	12.435	16.948	17.086	12.728	4.006	11.204	16.119	16.591	11.980
5:RI	0.115	1.944	0.832	1.408	1.075	0.124	2.300	1.483	2.153	1.515
6:VT	1.559	2.272	0.564	0.518	1.228	1.597	2.766	0.459	0.797	1.405
7:NJ	1.517	4.382	10.242	16.324	8.116	1.125	2.932	7.745	13.190	6.248
8:NY	2.886	0.121	6.498	11.935	5.360	2.120	1.049	6.873	11.774	5.454
9:PA	2.471	3.524	2.307	1.074	2.344	1.941	3.125	2.352	1.359	2.194
Sum	18.121	35.549	53.310	64.998	42.994	16.007	31.663	44.775	53.087	36.384

Detailed results for Region 2

For Region 2, Midwest, MA(1) is the best model in terms of the Total MAPE value (48.10). The total MAPE value from **Bigtime** method is 55.25% larger than the smallest MAPE value based on our Bayesian method. For each individual state, the best model differs. For instance, AR(1) fits Illinois (IL) with MAPE value as low as 0.702 (compared to the highest value of 8.024 by AR(0).

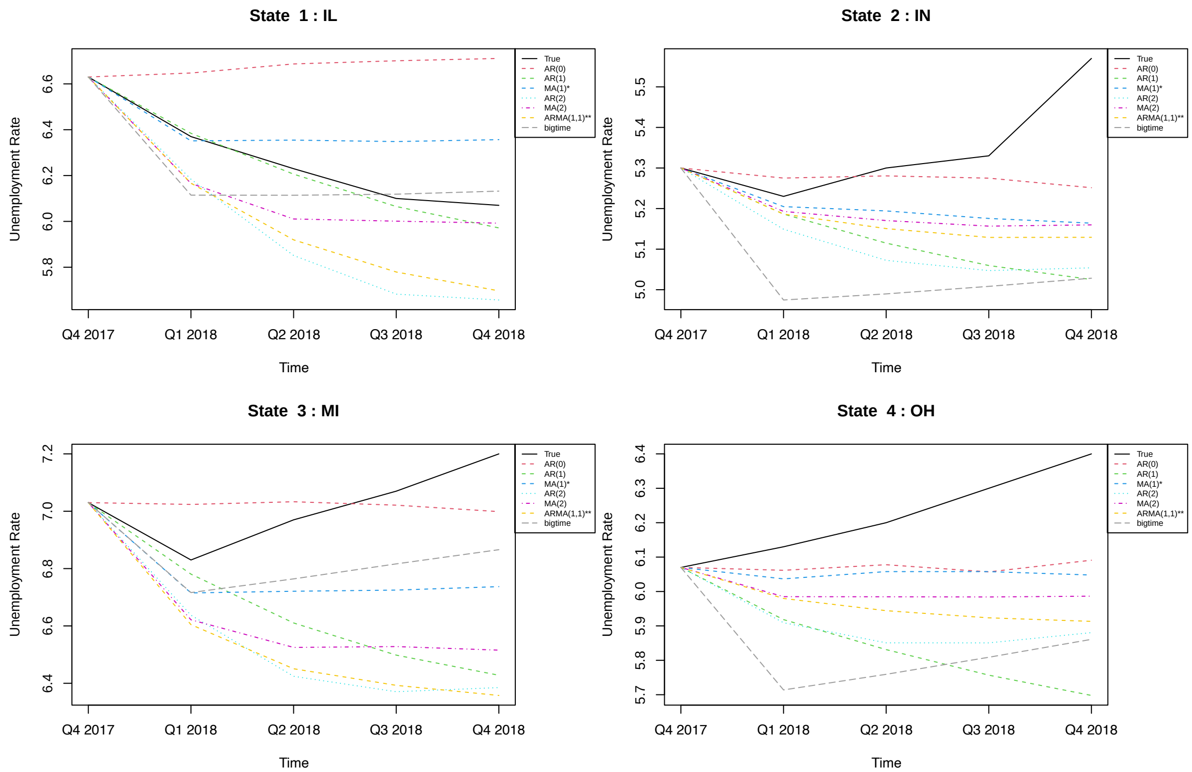


Figure 1.12: One-Year ahead Forecasting for Region2 (cont. 1/2)

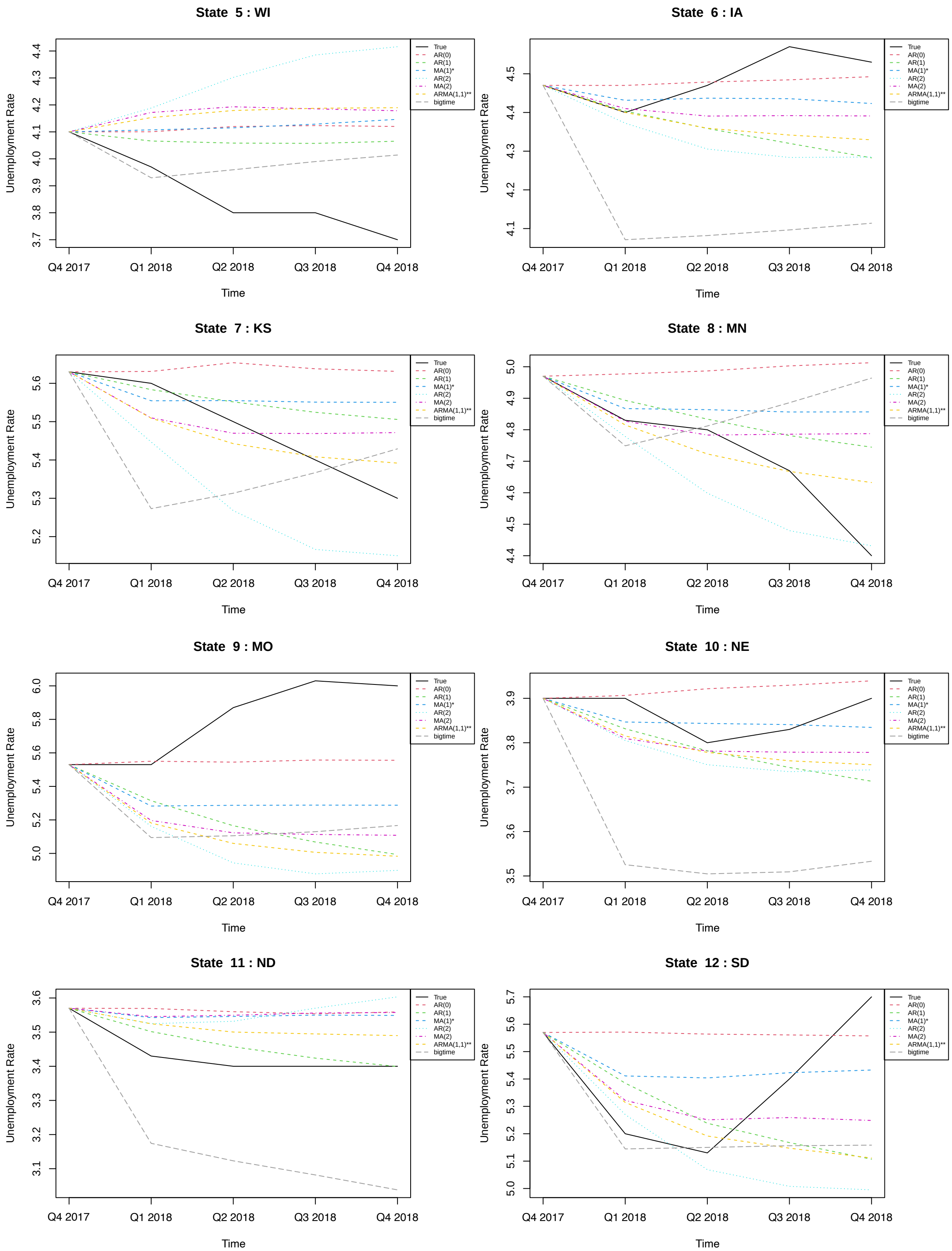


Figure 1.13: One-Year ahead Forecasting for Region2 (cont. 1/2)

Table 1.30: Four steps ahead Forecast Results for Region 2 (1/3)

	AR(0)					AR(1)				
state	step1	step2	step3	step4	MAPE	step1	step2	step3	step4	MAPE
1:IL	4.354	7.334	9.848	10.560	8.024	0.224	0.384	0.575	1.623	0.702
2:IN	0.864	0.367	1.033	5.720	1.996	0.832	3.494	5.073	9.783	4.795
3:MI	2.837	0.906	0.691	2.796	1.808	0.723	5.159	8.087	10.723	6.173
4:OH	1.115	1.971	3.857	4.829	2.943	3.453	5.950	8.621	10.971	7.249
5:WI	3.285	8.415	8.508	11.353	7.890	2.422	6.801	6.770	9.878	6.468
6:IA	1.589	0.192	1.874	0.828	1.121	0.080	2.491	5.468	5.459	3.374
7:KS	0.554	2.796	4.406	6.247	3.501	0.294	0.938	2.303	3.882	1.854
8:MN	3.049	3.891	7.125	13.929	6.998	1.311	0.701	2.388	7.835	3.059
9:MO	0.365	5.537	7.839	7.403	5.286	3.900	12.008	15.953	16.781	12.160
10:NE	0.168	3.200	2.597	1.010	1.744	1.759	0.478	2.239	4.787	2.316
11:ND	4.064	4.701	4.523	4.677	4.491	2.086	1.677	0.701	0.022	1.121
12:SD	7.139	8.460	2.979	2.502	5.270	3.552	2.110	4.302	10.404	5.092
Sum	29.383	47.770	55.280	71.854	51.072	20.636	42.191	62.480	92.148	54.363

Table 1.31: Four steps ahead Forecast Results for Region 2 (2/3)

	MA(1)					AR(2)				
state	step1	step2	step3	step4	MAPE	step1	step2	step3	step4	MAPE
1:IL	0.292	2.000	4.070	4.723	2.771	2.922	6.080	6.842	6.803	5.662
2:IN	0.482	1.997	2.895	7.288	3.166	1.546	4.294	5.307	9.265	5.103
3:MI	1.682	3.573	4.880	6.429	4.141	2.819	7.828	9.893	11.319	7.965
4:OH	1.522	2.293	3.844	5.502	3.290	3.602	5.632	7.133	8.117	6.121
5:WI	3.462	8.289	8.650	12.070	8.118	5.500	13.211	15.399	19.350	13.365
6:IA	0.713	0.740	2.941	2.357	1.688	0.630	3.680	6.263	5.412	3.996
7:KS	0.813	0.999	2.795	4.726	2.333	2.733	4.222	4.323	2.828	3.527
8:MN	0.770	1.329	3.992	10.378	4.117	1.063	4.192	4.076	0.717	2.512
9:MO	4.472	9.931	12.300	11.870	9.643	6.659	15.783	19.106	18.355	14.976
10:NE	1.367	1.144	0.283	1.679	1.118	2.433	1.303	2.485	4.130	2.588
11:ND	3.279	4.293	4.407	4.380	4.090	2.765	3.874	5.005	5.981	4.406
12:SD	4.060	5.340	0.423	4.689	3.628	1.347	1.193	7.261	12.369	5.543
Sum	22.914	41.928	51.480	76.091	48.103	34.019	71.292	93.093	104.646	75.764

Table 1.32: Four steps ahead Forecast Results for Region 2 (3/3)

	MA(2)					ARMA(1,1)				
state	step1	step2	step3	step4	MAPE	step1	step2	step3	step4	MAPE
1:IL	3.207	3.522	1.623	1.280	2.408	3.193	4.984	5.259	6.146	4.896
2:IN	0.707	2.445	3.251	7.364	3.442	0.833	2.818	3.773	7.915	3.835
3:MI	3.054	6.384	7.665	9.507	6.652	3.301	7.451	9.579	11.701	8.008
4:OH	2.365	3.473	5.014	6.462	4.328	2.450	4.126	5.977	7.607	5.040
5:WI	5.089	10.344	10.141	12.926	9.625	4.611	9.980	10.200	13.236	9.507
6:IA	0.228	1.774	3.898	3.068	2.242	0.001	2.477	4.997	4.430	2.976
7:KS	1.610	0.549	1.277	3.234	1.668	1.627	1.043	0.152	1.733	1.139
8:MN	0.059	0.352	2.481	8.810	2.926	0.309	1.601	0.052	5.282	1.811
9:MO	6.027	12.729	15.205	14.862	12.206	6.305	13.802	16.975	16.948	13.508
10:NE	2.317	0.493	1.339	3.121	1.818	2.177	0.572	1.845	3.834	2.107
11:ND	3.367	4.399	4.588	4.636	4.248	2.754	2.950	2.788	2.642	2.784
12:SD	2.342	2.356	2.609	7.920	3.807	2.200	1.204	4.680	10.326	4.602
Sum	30.372	48.820	59.091	83.190	55.370	29.761	53.008	66.277	91.800	60.213

Detailed results for Region 3

For Region 3 (South region), AR(2) fits the best in terms of the total MAPE value of 99.94 across 17 states. For each individual state, the best model differs. For instance, for state 7, West Virginia (WV), AR(2) model has a smallest MAPE value (3.521); see Table 1.34). However, AR(2) does not perform as well in predicting the states of Washington DC (DC). For the state of Delaware (DE), AR(0) prediction is the closest to the actual unemployment rate. From Figure 1.10, the actual unemployment rate of Michigan goes up from the first quarter of 2018 to the last quarter of 2018, and AR(0), without time structure involved, is the closest to the actual unemployment rate.

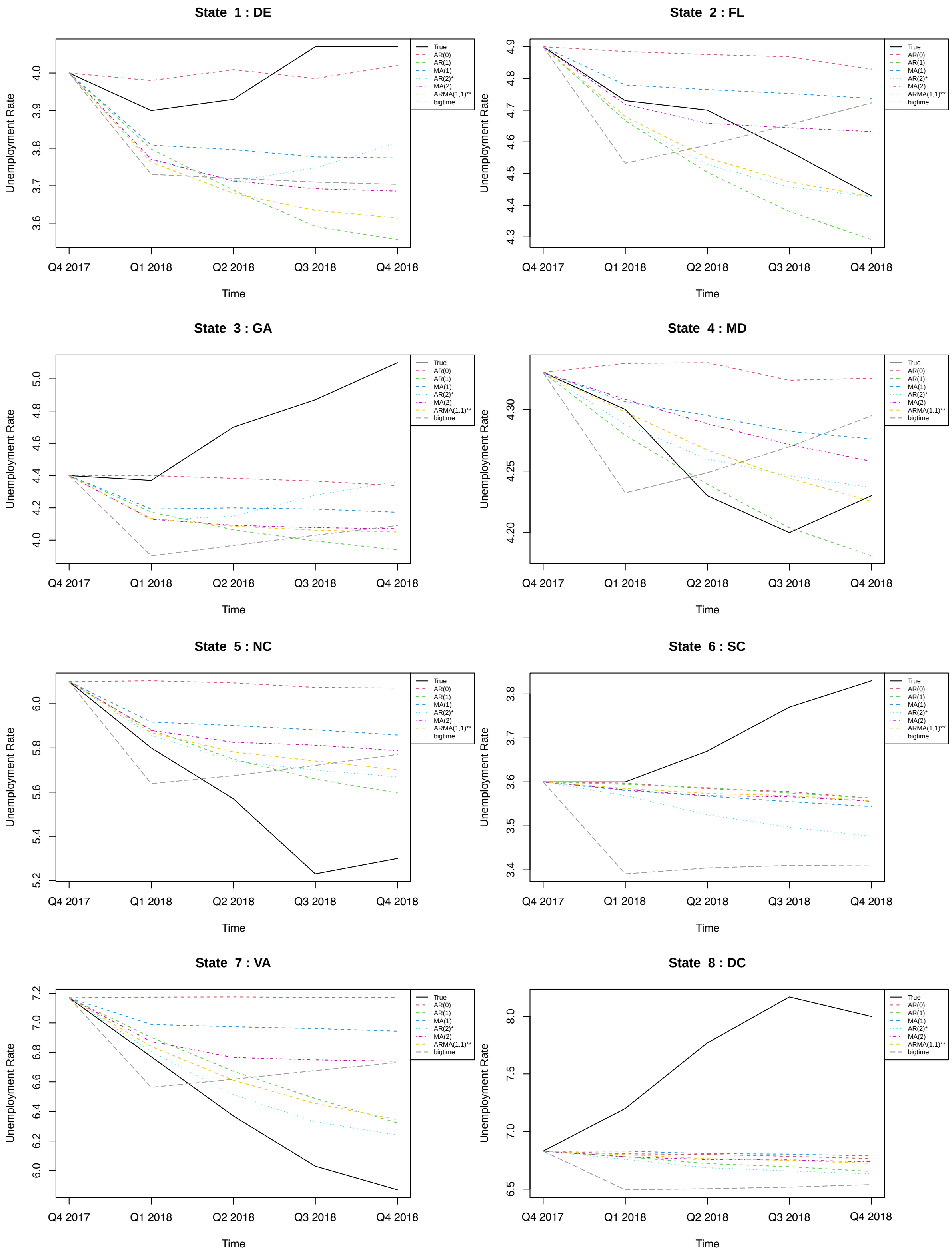


Figure 1.14: Region 3: Quarterly Forecasts vs actual UE rates (cont. 1/3)

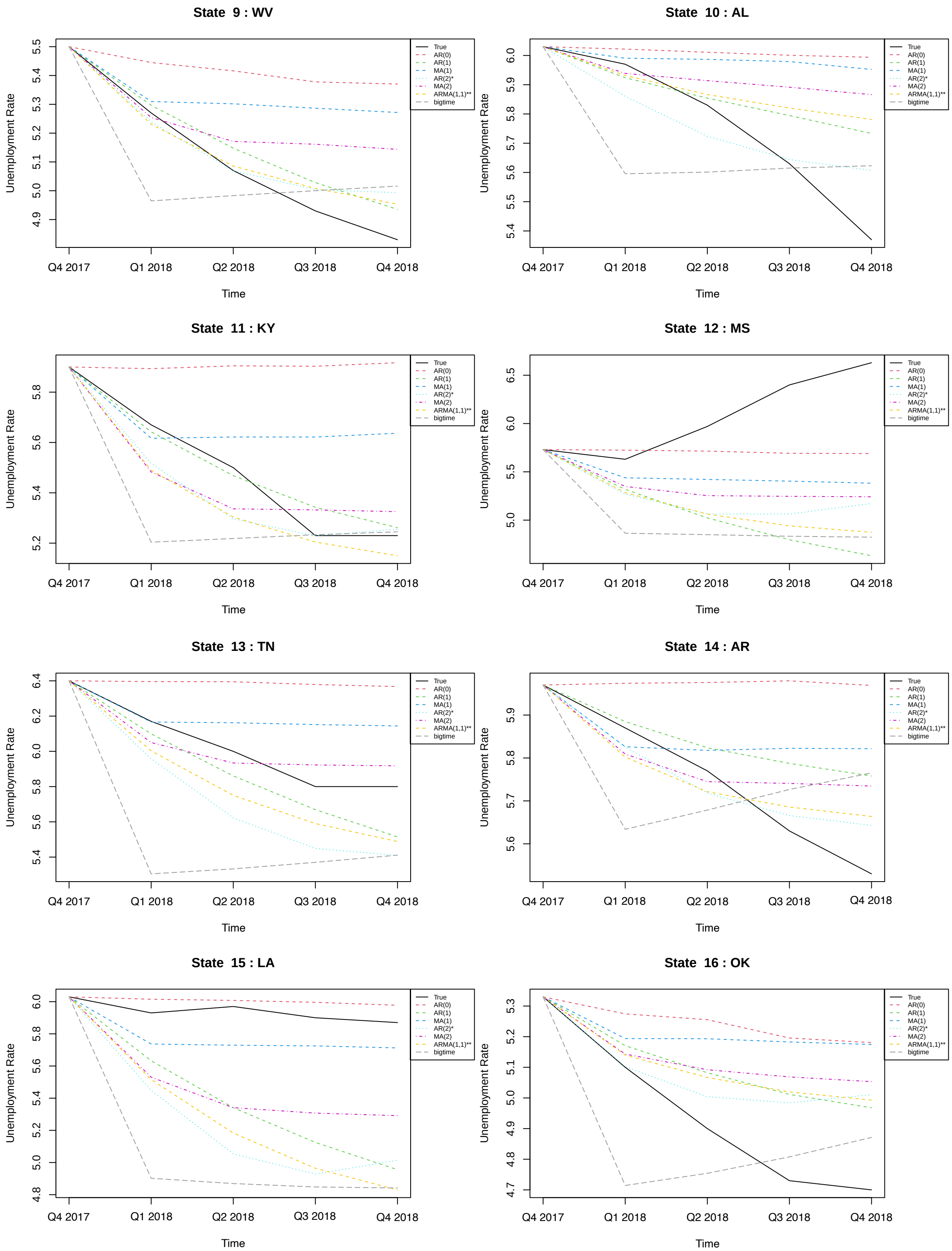


Figure 1.15: Region 3: Quarterly Forecasts vs actual UE rates (cont. 2/3)

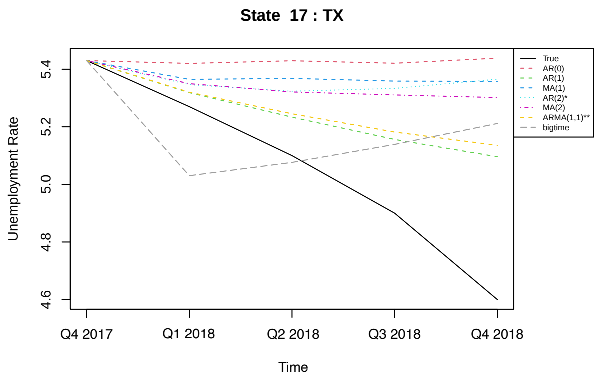


Figure 1.16: Region 3: Quarterly Forecasts vs actual UE rates (cont. 3/3)

Table 1.33: Four steps ahead Forecast Results for Region 3 (1/3)

	AR(0)					AR(1)				
state	step1	step2	step3	step4	MAPE	step1	step2	step3	step4	MAPE
1:DE	2.052	2.008	2.081	1.234	1.844	2.623	6.145	11.759	12.625	8.288
2:FL	3.276	3.732	6.527	9.020	5.639	1.342	4.173	4.145	3.130	3.197
3:GA	0.683	6.741	10.349	14.946	8.180	4.492	13.522	17.974	22.760	14.687
4:MD	0.869	2.553	2.946	2.254	2.156	0.489	0.222	0.094	1.154	0.490
5:NC	5.241	9.414	16.133	14.543	11.333	1.413	3.220	8.203	5.589	4.606
6:SC	0.078	2.324	5.098	6.973	3.618	0.153	2.266	5.182	6.973	3.644
7:VA	5.972	12.650	18.945	22.189	14.939	1.958	4.744	7.592	7.702	5.499
8:DC	5.467	12.473	16.955	15.438	12.583	5.819	13.491	18.059	16.836	13.551
9:WV	3.321	6.826	9.083	11.188	7.604	0.508	1.535	1.993	2.178	1.554
10:AL	0.868	3.105	6.585	11.615	5.543	0.770	0.415	2.921	6.765	2.718
11:KY	3.938	7.355	12.866	13.134	9.323	0.488	0.578	2.153	0.604	0.956
12:MS	1.679	4.281	11.062	14.187	7.802	5.534	15.864	25.052	30.148	19.150
13:TN	3.660	6.580	9.985	9.786	7.503	1.151	2.325	2.257	4.917	2.663
14:AR	1.774	3.567	6.212	7.941	4.874	0.252	0.934	2.796	4.123	2.026
15:LA	1.437	0.634	1.626	1.836	1.383	5.014	10.531	13.139	15.577	11.065
16:OK	3.419	7.255	9.846	10.226	7.687	1.386	3.721	5.957	5.697	4.190
17:TX	2.850	6.459	10.629	18.227	9.541	0.925	2.618	5.226	10.784	4.888
Sum	46.584	97.957	156.928	184.737	121.552	34.317	86.304	134.502	157.562	103.172

Table 1.34: Four steps ahead Forecast Results for Region 3 (2/3)

	MA(1)					AR(2)				
state	step1	step2	step3	step4	MAPE	step1	step2	step3	step4	MAPE
1:DE	2.355	3.404	7.203	7.274	5.059	3.355	5.549	7.918	6.249	5.768
2:FL	1.041	1.381	3.991	6.932	3.336	1.305	3.649	2.446	0.074	1.868
3:GA	4.061	10.639	13.915	18.186	11.700	5.622	11.738	12.154	14.436	10.988
4:MD	0.144	1.538	1.958	1.089	1.182	0.282	0.702	1.101	0.158	0.561
5:NC	2.022	5.949	12.463	10.537	7.743	0.969	3.053	8.972	6.962	4.989
6:SC	0.542	2.786	5.701	7.472	4.125	0.865	3.944	7.242	9.231	5.320
7:VA	3.246	9.486	15.455	18.301	11.622	0.538	2.257	4.977	6.311	3.521
8:DC	5.135	12.382	16.724	15.150	12.348	6.133	13.961	18.510	17.104	13.927
9:WV	0.753	4.570	7.232	9.144	5.425	0.589	0.014	1.497	3.382	1.370
10:AL	0.345	2.687	6.202	10.841	5.019	1.832	1.824	0.253	4.410	2.080
11:KY	0.940	2.213	7.489	7.777	4.605	2.707	3.691	0.022	0.494	1.728
12:MS	3.407	9.191	15.565	18.818	11.745	6.413	15.187	20.882	21.993	16.119
13:TN	0.062	2.708	6.067	5.933	3.692	3.431	6.288	6.044	6.742	5.626
14:AR	0.746	0.827	3.419	5.274	2.566	0.760	0.897	0.638	2.040	1.084
15:LA	3.259	4.029	2.966	2.682	3.234	8.129	15.343	16.472	14.569	13.628
16:OK	1.841	5.985	9.573	10.096	6.874	0.038	2.124	5.369	6.612	3.536
17:TX	1.798	5.259	9.362	16.474	8.223	1.452	4.384	8.845	16.638	7.830
Sum	31.697	85.034	145.285	171.980	108.498	44.420	94.605	123.342	137.405	99.943

Table 1.35: Four steps ahead Forecast Results for Region 3 (3/3)

	MA(2)					ARMA(1,1)				
state	step1	step2	step3	step4	MAPE	step1	step2	step3	step4	MAPE
1:DE	3.307	5.524	9.285	9.440	6.889	3.525	6.356	10.711	11.214	7.952
2:FL	0.256	0.884	1.634	4.571	1.836	1.064	3.190	2.102	0.039	1.599
3:GA	5.491	12.941	16.279	20.187	13.725	5.387	13.059	16.603	20.584	13.908
4:MD	0.191	1.381	1.703	0.659	0.984	0.047	0.875	1.053	0.108	0.521
5:NC	1.373	4.588	11.138	9.202	6.575	1.190	3.807	9.766	7.575	5.585
6:SC	0.496	2.763	5.401	7.155	3.954	0.416	2.622	5.334	7.127	3.875
7:VA	1.529	6.198	11.920	14.819	8.616	1.034	3.792	7.039	8.081	4.986
8:DC	5.827	13.042	17.339	15.790	12.999	5.572	12.949	17.387	15.957	12.966
9:WV	0.298	1.998	4.694	6.494	3.371	0.750	0.313	1.583	2.560	1.302
10:AL	0.523	1.438	4.645	9.240	3.962	0.641	0.622	3.379	7.659	3.075
11:KY	3.300	2.977	1.951	1.821	2.512	3.179	3.572	0.493	1.530	2.193
12:MS	4.999	12.000	18.027	20.938	13.991	6.120	15.203	22.808	26.486	17.654
13:TN	1.944	1.104	2.117	2.029	1.798	2.714	4.165	3.614	5.373	3.966
14:AR	1.037	0.436	1.967	3.700	1.785	1.157	0.842	0.990	2.419	1.352
15:LA	6.738	10.535	10.045	9.858	9.294	6.989	13.178	15.861	17.718	13.437
16:OK	0.847	3.921	7.161	7.517	4.861	0.783	3.404	6.125	6.235	4.137
17:TX	1.514	4.334	8.380	15.251	7.370	0.950	2.849	5.750	11.648	5.299
Sum	39.670	86.064	133.686	158.671	104.522	41.518	90.798	130.598	152.313	103.807

Detailed results for Region 4

For Region 4 (West region), AR(2) is the best model in terms of the total MAPE value across 13 states. For each individual state, the best model differs. Detail MAPE values are presented in Tables 1.17, 1.18.

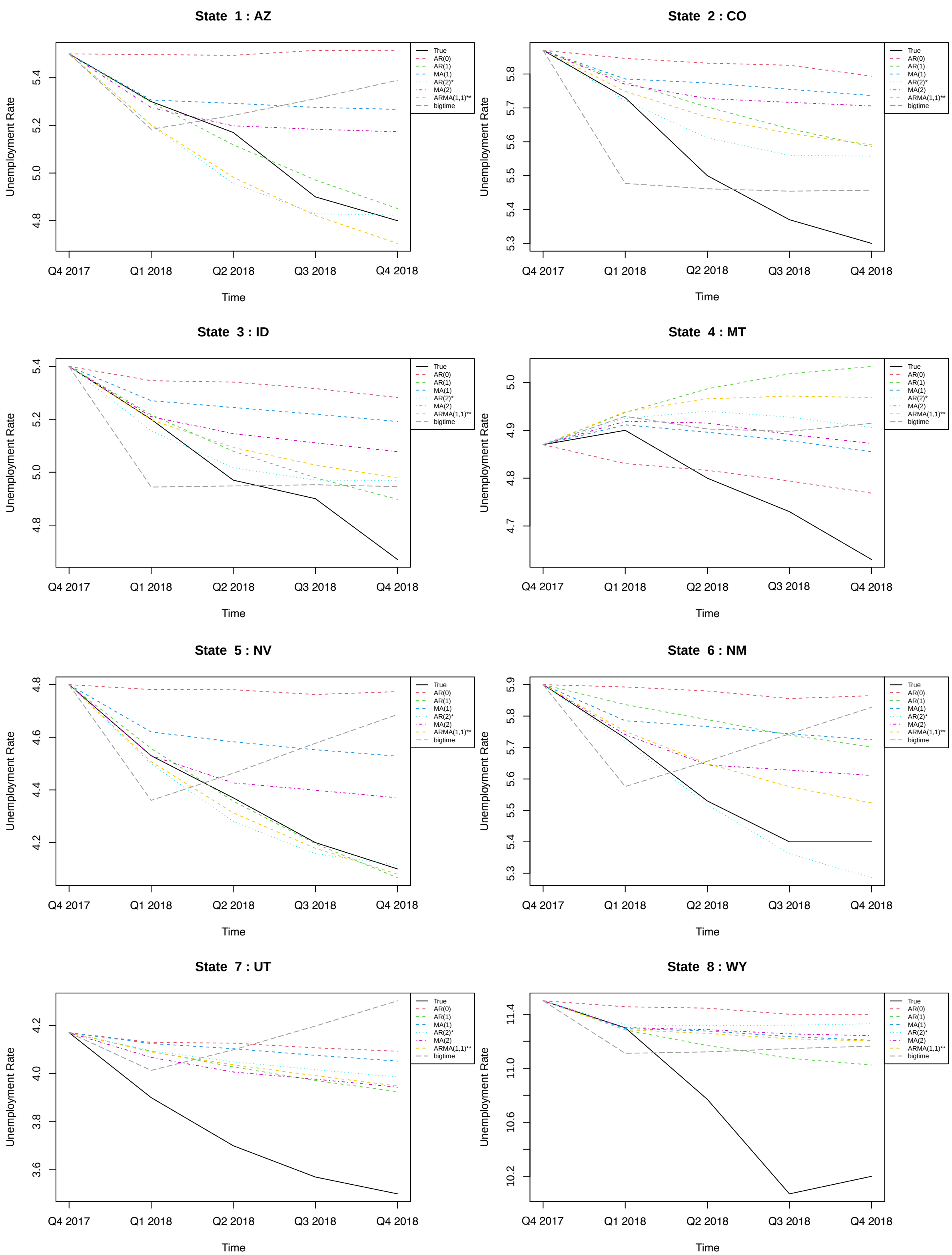


Figure 1.17: Region 4: Quarterly Forecasts vs actual UE rates (cont. 1/2)

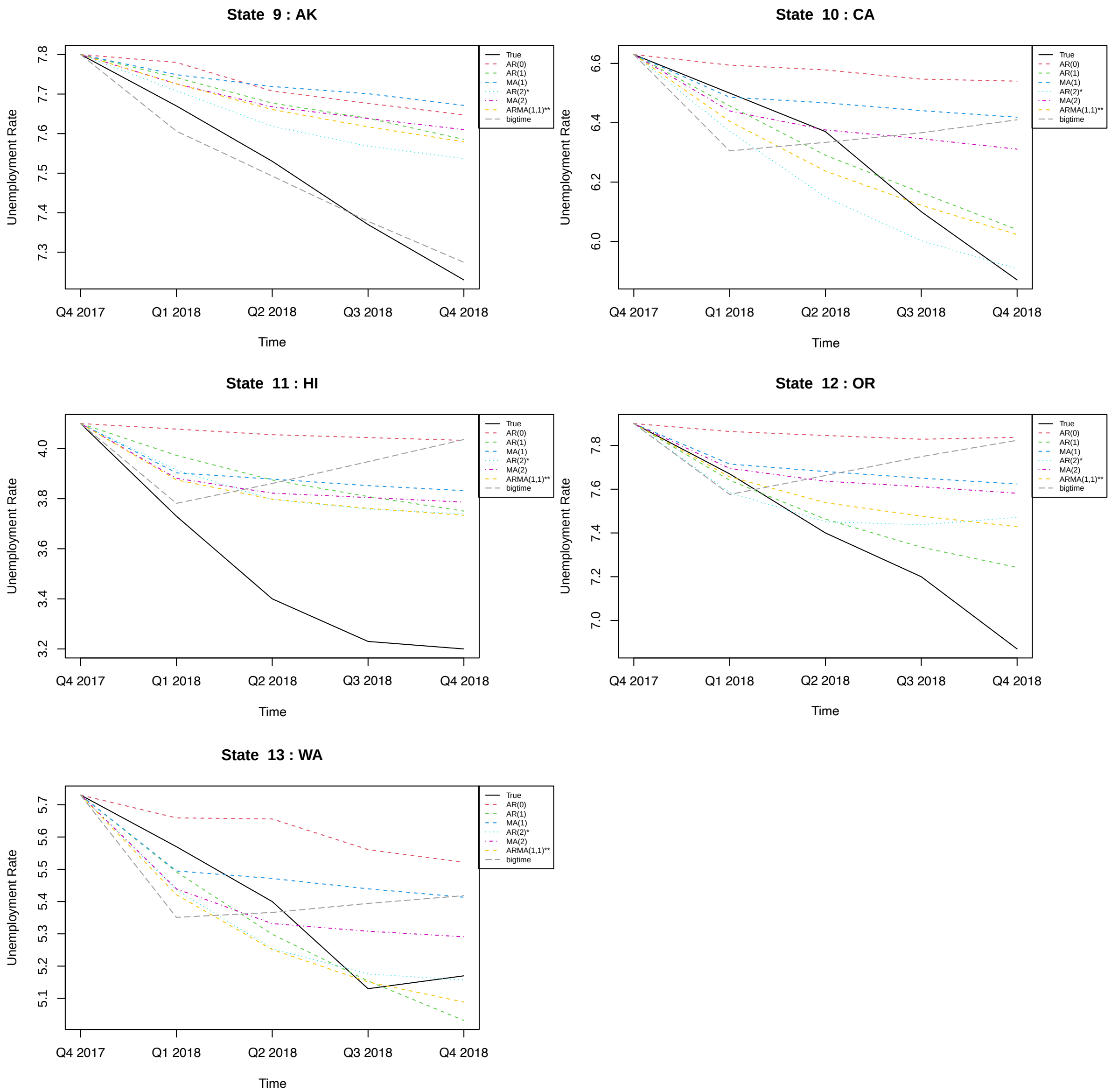


Figure 1.18: Region 4: Quarterly Forecasts vs actual UE rates (cont. 2/2)

Table 1.36: Four steps ahead Forecast Results for Region 4 (1/3)

	AR(0)					AR(1)				
state	step1	step2	step3	step4	MAPE	step1	step2	step3	step4	MAPE
1:AZ	3.715	6.266	12.541	14.895	9.354	0.068	0.998	1.446	1.062	0.893
2:CO	2.027	6.039	8.489	9.317	6.468	0.809	3.682	5.004	5.387	3.720
3:ID	2.807	7.458	8.497	13.113	7.969	0.338	2.201	1.614	4.871	2.256
4:MT	1.420	0.341	1.357	2.995	1.528	0.758	3.890	6.090	8.715	4.863
5:NV	5.551	9.401	13.396	16.426	11.194	0.614	0.280	0.092	0.817	0.451
6:NM	2.838	6.328	8.440	8.609	6.554	1.862	4.678	6.288	5.582	4.602
7:UT	5.885	11.523	15.039	16.939	12.346	4.923	8.856	11.240	12.124	9.286
8:WY	1.378	6.269	13.210	11.761	8.155	0.140	3.701	9.972	8.083	5.474
9:AK	1.434	2.359	4.160	5.774	3.432	0.933	1.955	3.648	4.915	2.863
10:CA	1.449	3.265	7.329	11.418	5.865	0.663	1.249	1.051	2.891	1.464
11:HI	9.326	19.278	25.196	26.024	19.956	6.516	13.958	17.912	17.200	13.896
12:OR	2.526	6.020	8.727	14.075	7.837	0.379	0.853	1.874	5.422	2.132
13:WA	1.601	4.739	8.400	6.802	5.386	1.400	1.877	0.468	2.674	1.605
Sum	41.957	89.286	134.781	158.148	106.044	19.403	48.178	66.699	79.743	53.505

Table 1.37: Four steps ahead Forecast Results for Region 4 (2/3)

	MA(1)					AR(2)				
state	step1	step2	step3	step4	MAPE	step1	step2	step3	step4	MAPE
1:AZ	0.118	2.369	7.665	9.732	4.971	1.874	4.135	1.451	0.491	1.988
2:CO	0.966	4.976	7.166	8.234	5.335	0.151	2.023	3.541	4.858	2.643
3:ID	1.353	5.528	6.515	11.175	6.143	0.834	0.938	1.431	6.394	2.399
4:MT	0.234	2.000	3.132	4.864	2.558	0.510	2.912	4.180	5.952	3.389
5:NV	1.993	4.872	8.398	10.445	6.427	0.683	2.057	1.003	0.351	1.023
6:NM	0.963	4.281	6.362	6.018	4.406	0.189	0.215	0.703	2.116	0.806
7:UT	5.763	10.898	14.169	15.769	11.650	4.987	9.430	12.494	13.880	10.198
8:WY	0.050	4.697	11.570	9.871	6.547	0.184	5.064	12.403	11.075	7.181
9:AK	1.030	2.511	4.490	6.105	3.534	0.499	1.177	2.693	4.242	2.153
10:CA	0.214	1.525	5.587	9.346	4.168	1.974	3.461	1.604	0.639	1.920
11:HI	4.639	14.064	19.258	19.753	14.428	4.944	11.665	16.355	16.865	12.457
12:OR	0.586	3.793	6.254	10.971	5.401	1.108	0.692	3.302	8.751	3.463
13:WA	1.350	1.324	6.031	4.708	3.353	2.394	2.700	0.891	0.250	1.559
Sum	19.259	62.838	106.597	126.991	78.921	20.331	46.469	62.051	75.864	51.179

Table 1.38: Four steps ahead Forecast Results for Region 4 (3/3)

	MA(2)					ARMA(1,1)				
state	step1	step2	step3	step4	MAPE	step1	step2	step3	step4	MAPE
1:AZ	0.473	0.542	5.787	7.775	3.644	1.843	3.634	1.601	1.985	2.266
2:CO	0.690	4.137	6.451	7.656	4.733	0.324	3.138	4.741	5.490	3.423
3:ID	0.196	3.539	4.305	8.732	4.193	0.041	2.481	2.597	6.596	2.929
4:MT	0.384	2.400	3.414	5.240	2.860	0.791	3.451	5.109	7.310	4.165
5:NV	0.009	1.301	4.736	6.599	3.161	0.503	1.301	0.517	0.488	0.702
6:NM	0.195	2.071	4.225	3.911	2.600	0.375	2.148	3.248	2.292	2.016
7:UT	4.287	8.283	11.393	12.672	9.159	4.876	9.100	11.795	12.811	9.646
8:WY	0.008	4.792	11.766	10.212	6.694	0.131	4.533	11.405	9.829	6.474
9:AK	0.726	1.830	3.637	5.257	2.862	0.728	1.738	3.353	4.834	2.663
10:CA	0.926	0.089	4.033	7.512	3.140	1.463	2.088	0.356	2.604	1.628
11:HI	4.061	12.413	17.789	18.317	13.145	3.887	11.705	16.453	16.674	12.180
12:OR	0.329	3.194	5.713	10.354	4.898	0.176	1.869	3.849	8.129	3.506
13:WA	2.347	1.265	3.473	2.337	2.356	2.672	2.756	0.408	1.576	1.853
Sum	14.631	45.856	86.722	106.574	63.445	17.810	49.942	65.432	80.618	53.451

1.4 Conclusions

To provide reliable forecasts of the U.S. civilian unemployment rate for each of the 51 states (50 states plus the District of Columbia), we have developed a time series model under the Bayesian framework that borrows strength across states and uses the unemployment insurance claims data as a covariate. We have conducted our comprehensive study by adopting two different approaches to grouping the 51 states. In the first approach, we grouped the states according to similarities in the ARMA structure, whereas in the second approach, we created groups simply based on the four national regions defined by the Census Bureau.

For each state within a Group or a Region, we have designated the UE rates from 1990:Q1 to 2017:Q4 as the *training data* and the remaining four quarters of 2018 as the *test data*. We have built our time series models based on a fully Bayesian framework for the training data, obtained forecasts of UE rates in the test data, and compared the performance of the forecasts with an existing method called **Bigtime**.

Regardless of how we group the states (by similarities in the time series structure or by region), we have demonstrated that our Bayesian time series modeling method outperforms the **Bigtime** method significantly.

A spatial model is commonly used to data with a spatial location, in which the geographic component is an important factor in the analysis. In our data analysis examples, we observe that the RM (relative measure) values by Region are significantly greater than by Group, suggesting that the location may play a significant influence in forecasting the small area means. In future work, we may therefore investigate incorporating a geographic adjacency matrix into the model suggested in this dissertation.

Chapter 2

Pseudo-Bayesian Small Area Estimation

2.1 Introduction

Estimation of suitable population characteristics, say, means or proportions, of many related or similar subpopulations has many applications. Some applications are animal breeding, image analysis, fMRI study, disease mapping and small area estimation. In survey sampling, small areas are partitions of a population in m subpopulations that are defined by demography and/or geography. Estimates of useful subpopulation characteristics θ_i's are urgently needed to implement many government policies for the benefit of the country. However, a traditional estimate Y_i, also called a direct estimate, of θ_i based only on the sample from an individual subpopulation will be highly variable if the corresponding sample size is small. To improve low accuracy of direct estimates of small area means, researchers proposed model-based small area estimation by using an appropriate model for the θ_i's based on some available covariates.

Stein (1956) pioneered simultaneous estimation of the mean vector $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_m)^T$ of a multivariate normal response vector $\mathbf{Y} = (Y_1, \cdots, Y_m)^T$, where the components measure the same characteristic of m related subpopulations. Under the sum of squared error loss function $L(\hat{\boldsymbol{\theta}}, \boldsymbol{\theta}) = \sum_{i=1}^{m}(\hat{\theta}_i - \theta_i)^2$ to estimate $\boldsymbol{\theta}$, James and Stein, 1961 developed an alternative estimator $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \cdots, \hat{\theta}_m)^T$, dominating the classical estimator $\mathbf{Y}$ of $\boldsymbol{\theta}$ in terms of the frequentist (for fixed

$\boldsymbol{\theta}$) risk function. Subsequently, the problem received unprecedented attention from researchers. For example, Lindley and Smith, 1972 developed a hierarchical Bayes (HB) estimator of $\boldsymbol{\theta}$, and Efron and Morris, 1972 presented an empirical Bayes (EB) interpretation and a heuristic justification of dominance of James-Stein estimator.

Fay and Herriot, 1979 proposed a model which is widely used in small area estimation. The Fay-Herriot model implicitly connects direct estimator Y_i of θ_i with related covariates $\mathbf{x}_i$. It is a mixed effects model given by

$$Y_i = \theta_i + e_i,\ \theta_i = \mathbf{x}_i{}^T\boldsymbol{\beta} + v_i,\ i = 1, \ldots, m, \tag{2.1}$$

where Y_i is related to θ_i via the sampling model, and θ_i is related to the covariates $\boldsymbol{x_i}$ via the linking model. It assumes that the sampling error e_i satisfies $E_P(e_i|\boldsymbol{x_i}) = 0$ and $V_P(e_i|\boldsymbol{x_i}) = D_i$, where the subscript P refers to the sampling design. Sampling errors are assumed to be independent and normally distributed, and are independent of linking errors, v_i, the area-specific random effects, which are assumed independent and normally distributed with zero mean and a common variance A. The sampling variances, D_i's, which are usually estimated, are treated as known. We denote model parameters $(\boldsymbol{\beta^T}, A)^T$ by $\boldsymbol{\psi}$, where $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown regression coefficients. In the Fay-Herriot model estimation of means, for a $p \times 1$ vector of known covariates $\boldsymbol{x_i}$, becomes prediction of the mixed effects $\theta_i = \boldsymbol{x_i}^\top\boldsymbol{\beta} + v_i,\ i = 1, \ldots, m$.

For sum of squared error loss when the model parameters are known, the best predictor (BP) of θ_i is

$$\tilde{\theta}_{i,B} = Y_i - \gamma_i(Y_i - \boldsymbol{x_i}^\top\boldsymbol{\beta}), \tag{2.2}$$

where $\gamma_i = D_i/(D_i + A)$. But since $\boldsymbol{\psi}$ is unknown, the predictor $\tilde{\theta}_{i,B}$ cannot be used. If A were known, then the regression coefficient $\boldsymbol{\beta}$ is estimated by the maximum likelihood estimator based on the distribution of the data. This is equivalent to the estimator

$$\tilde{\boldsymbol{\beta}}_M = [\sum_{i=1}^{m}\boldsymbol{x_i}\boldsymbol{x_i}^\top/(D_i + A)]^{-1}[\sum_{i=1}^{m}\boldsymbol{x_i}Y_i/(D_i + A)],$$

which minimizes the weighted sum of squared errors $\sum_{i=1}^{m}(Y_i - \boldsymbol{x_i}^T\boldsymbol{\beta})^2/(D_i + A)$ with respect to $\boldsymbol{\beta}$. (We assume that the matrix $[\boldsymbol{x_1}, \cdots, \boldsymbol{x_m}]$ is of rank p.)

This objective function puts greater weight on more accurate Y_is. Replacing $\boldsymbol{\beta}$ in $\tilde{\boldsymbol{\theta}}_{i,B}$ by $\tilde{\boldsymbol{\beta}}_M$ results in the best linear unbiased predictor (BLUP) of θ_i. In $\tilde{\boldsymbol{\beta}}_M$ (assuming known A), the direct estimator Y_i is weighted inversely proportional to $D_i + A$, which means that a less accurate Y_i will have less influence on $\tilde{\boldsymbol{\beta}}_M$, and also on the BLUP. Similarly, a less accurate direct estimator Y_i will have less importance when A is estimated by maximizing the likelihood or residual likelihood. The implication is that the BLUP/EBLUP is a hybrid of optimal prediction (i.e., BP) and optimal estimation (e.g., MLE). Actually, for a given set of positive weights $w_1, \cdots, w_m$ with $\sum_{i=}^{m} w_i = 1$, the weighted least squares estimator

$$\tilde{\boldsymbol{\beta}}_W = [\sum_{i=1}^{m} w_i \boldsymbol{x_i}\boldsymbol{x_i}^\top]^{-1}(\sum_{i=1}^{m} w_i \boldsymbol{x_i} Y_i) \tag{2.3}$$

includes $\tilde{\boldsymbol{\beta}}_M$ as a special case for $w_i(D_i + A) \propto 1$.

Jiang et al., 2011 argued that the EBLUP is not suitable for prediction of θ_i since areas with less accurate Y_is have less influence in determining the model parameters. Since prediction of θ_is is the main interest, they argued for a purely predictive procedure where both the predictor of $\boldsymbol{\theta}$ and the estimators of model parameters are derived by minimization of predictive mean squared error. They proposed that these parameters should be estimated by minimizing an estimate of the total mean squared error (MSE) of prediction.

Rao and Molina, 2015, p. 166 construct Jiang et al., 2011's objective function $Q(\psi)$ by deriving the total sampling MSE of $\tilde{\theta}_{i,B}, i = 1, \cdots, m$, $MSE_P(\tilde{\theta}_B) = \sum_{i=1}^{m} E_P[\tilde{\theta}_{i,B} - \theta_i]^2$. Using equation (6.4.52) in their book, they simplify $MSE_P(\tilde{\theta}_B) = E_P[\sum_{i=1}^{m}\{\gamma_i^2(Y_i - \boldsymbol{x_i}^T\boldsymbol{\beta})^2 - 2D_i\gamma_i + D_i\}]$ to $E_P[\sum_{i=1}^{m}\{\gamma_i^2(Y_i - \boldsymbol{x_i}^T\boldsymbol{\beta})^2 + 2A\gamma_i - D_i\}]$ by using $D_i - 2D_i\gamma_i = 2A\gamma_i - D_i$, then dropping the last term. Thus,

$$\begin{aligned} Q(\psi) &\overset{def}{=} Q(\boldsymbol{\beta}, A, \mathbf{Y}) = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^T\boldsymbol{\Gamma}^2(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) + 2A \cdot tr(\boldsymbol{\Gamma}) \\ &\overset{def}{=} Q_0(\boldsymbol{\beta}, A, \mathbf{Y}) + Q_1(A). \end{aligned} \tag{2.4}$$

In (2.4), $\mathbf{X} = [\boldsymbol{x_1}, \cdots, \boldsymbol{x_m}]^T$, $\boldsymbol{\Gamma} = diag(\gamma_1, \cdots, \gamma_m)$, $\mathbf{Y} = (Y_1, \cdots, Y_m)^T$ and $Q_1(A) = 2A \cdot tr(\boldsymbol{\Gamma})$.

The resulting estimator $\widetilde{\boldsymbol{\psi}} = (\widetilde{\boldsymbol{\beta}}^T, \widetilde{A})^T$, obtained by minimizing $Q(\boldsymbol{\psi})$, is referred to as the best predictive estimator (BPE) of $\boldsymbol{\psi}$. If A is known, the BPE of $\boldsymbol{\beta}$ can be obtained by minimizing $Q_0(\boldsymbol{\beta}, A, \mathbf{Y})$, the first term on the right-hand side of (2.4). This yields a closed-form solution

$$\widetilde{\boldsymbol{\beta}_O}(A) = (\mathbf{X}^T\boldsymbol{\Gamma}^2\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\Gamma}^2\mathbf{Y} = \left\{\sum_{i=1}^{m}\left(\frac{D_i}{A+D_i}\right)^2 \mathbf{x}_i\mathbf{x}_i^T\right\}^{-1}\sum_{i=1}^{m}\left(\frac{D_i}{A+D_i}\right)^2 \mathbf{x}_i Y_i.$$

The $\widetilde{\boldsymbol{\beta}}_O(A)$ is a special case of (2.3) with $w_i \propto (D_i/(D_i + A))^2$. Unlike in $\widetilde{\boldsymbol{\beta}_M}$, a less accurate Y_i receives a higher weight in $\widetilde{\boldsymbol{\beta}_O}$.

Now, $\tilde{A}$, the BPE of A, is obtained by minimizing $Q_0(\widetilde{\boldsymbol{\beta}}_O(A), A, \mathbf{y}) + Q_1(A)$ with respect to A, subject to $A > 0$. Then $\widetilde{\boldsymbol{\beta}}$, the BPE of $\boldsymbol{\beta}$ when A is unknown, which is usually the case, is $\widetilde{\boldsymbol{\beta}}_O(\tilde{A})$. We denote the BPE of $\boldsymbol{\psi}$ by $\widetilde{\boldsymbol{\psi}} = (\widetilde{\boldsymbol{\beta}}^\top, \widetilde{A})^T$. Finally, Jiang et al., 2011 proposed a predictor of the mixed effects $\boldsymbol{\theta}$ by replacing $\boldsymbol{\psi}$ in the BP (2.2) with its BPE $\widetilde{\boldsymbol{\psi}}$. Since the BPE minimizes the expression inside the expectation in MSPE$[\widetilde{\boldsymbol{\theta}}(\boldsymbol{\psi})]$, that is, the "observed" MSPE, the predictor is referred to as the observed best predictor (OBP).

Jiang et al., 2011 showed that the OBP can significantly outperform EBLUP in terms of the total MSPE if the underlying model is misspecified. On the other hand, if the underlying model is correctly specified, Jiang et al., 2011 show that the overall predictive performance of the OBP is similar to that of the EBLUP if the number of small areas is large. In Section 2.5, we evaluate the robustness of the OBP, EBLUP or other predictors to mean misspecification via two realistic applications, demonstrating that all considered methods incur large but comparable MSEs of prediction when an important covariate is omitted.

2.2 A Pseudo-Bayesian Alternative to OBP

In the frequentist paradigm, both the EBLUP and the OBP of θ_i are developed in two steps. First, the best predictor of θ_i, depending on $\boldsymbol{\psi}$, is derived by treating $\boldsymbol{\psi}$ as known. Next, when $\boldsymbol{\psi}$ is replaced by the BPE in the best predictor of θ_i, one gets the OBP. But if $\boldsymbol{\psi}$ is replaced by its traditional estimators, one gets the EBLUP. More specifically, first $\boldsymbol{\beta}$ is estimated by some appropriate $\boldsymbol{\beta}^*(A)$, as if A were known. Then, $\boldsymbol{\beta}^*(A)$ is used to estimate A by $\hat{A}$ (say), and finally, $\boldsymbol{\beta}$ is estimated by $\boldsymbol{\beta}^*(\hat{A})$.

In a Bayesian paradigm, under sum of squared error loss, the optimal predictor is the Bayes predictor, which is the posterior mean of θ_i based on the joint posterior distribution of $\theta_1, \cdots, \theta_m$, and $\boldsymbol{\psi}$. This posterior mean is actually the iterated expectation of $\tilde{\theta}_{i,B}$, and there is no scope to decouple the prediction of θ_i from Bayesian estimation of the model parameters $\boldsymbol{\psi}$. This Bayes predictor based on popular noninformative priors for $\boldsymbol{\psi}$ was studied by Ghosh, 1992 and Datta et al., 2005. These priors produce posteriors that are nearly equivalent to the likelihood of $\boldsymbol{\psi}$ based on the marginal distribution of the data under the Fay-Herriot model in (2.1). Consequently, the resulting Bayes predictors of the θ_i's are qualitatively similar to the EBLUPs of the θ_i's.

Since OBP is popular, we seek a Bayesian version of this frequentist procedure. Also a Bayesian OBP may avoid the complexity of the estimated MSEs in frequentist OBPs, which are sometimes negative Y. Liu et al., 2022, eqn. (3.2). Indeed, derivation of the MSE estimator requires clever but tedious calculations Y. Liu et al., 2022; see the complexity of the derivation of eqn. (3.2). On the other hand, the uncertainty estimates of Bayesian predictions are posterior variances, which are always positive and much simpler to compute. Finally, Bayesian methods that integrate out all unknown random quantities take better account of the full uncertainty associated with estimation/prediction.

In order to develop a Bayesian analog to the OBP, one needs to construct a sensible posterior distribution for $\boldsymbol{\psi}$ which allows small areas with larger variances to have more influence on the Bayes estimate of $\boldsymbol{\psi}$, just as the BPE of $\boldsymbol{\psi}$ does. In our pseudo-Bayes approach, we couple the conditional posterior distribution of θ_i given $\boldsymbol{\psi}$, $\pi(\theta_i|\boldsymbol{\beta}, A, \mathbf{y})$, which is obtained from (2.1) with the pseudo-posterior distribution of $\boldsymbol{\psi}$, which we denote by $\pi_{ps}(\boldsymbol{\beta}, A|\mathbf{y})$ (in the subscript, "ps" refers to pseudo). Analogous to frequentist method two-step estimation of $\boldsymbol{\beta}$ and A, we first obtain the conditional pseudo-posterior distribution of $\boldsymbol{\beta}$ given A, $\pi_{ps}(\boldsymbol{\beta}|A, \mathbf{y})$, and then derive the marginal pseudo-posterior distribution of A, $\pi_{ps}(A|\mathbf{y})$. We develop these components in the following subsections.

Recall the decomposition of the objective function $Q(\boldsymbol{\psi})$ into $Q_0(\boldsymbol{\beta}, A, \mathbf{Y})$ and $Q_1(A)$ in (2.4). The first step in deriving the pseudo-posterior of $\boldsymbol{\psi}$ obtains the conditional pseudo-posterior of

$\boldsymbol{\beta}$ given A using $Q_0(\boldsymbol{\beta}, A, \mathbf{Y})$. Next, get the pseudo-posterior of A based on the loss function $Q_*(A) = Q_0(\tilde{\boldsymbol{\beta}}(A), A, \mathbf{Y}) + Q_1(A)$.

2.2.1 Update of a prior by a loss function: a pseudo-posterior

To derive a pseudo-posterior pdf for $\boldsymbol{\psi}$, we mimic a framework for general Bayesian inference put forward by Bissiri et al., 2016. They claim the "information" about the parameter $\boldsymbol{\beta}$ in the loss function and in the prior should be carefully calibrated to update the prior distribution using the loss function $Q_0(\boldsymbol{\beta}, A, \mathbf{y})$. We update the uniform prior on $\boldsymbol{\beta}$ by creating an "ad hoc likelihood" for $\boldsymbol{\beta}$ from this loss function. Since the loss function depends on the *scale* of the data $\mathbf{y}$, the ad hoc likelihood needs to be calibrated so that both the ad hoc likelihood and the prior pdf of $\boldsymbol{\beta}$ become "scale-free".

For data d with the density $f(d|\phi)$ and prior pdf $\pi(\phi)$ for the parameter ϕ, Bayes' rule updates the prior pdf $\pi(\cdot)$ to the posterior pdf determined by normalization of the kernel $f(d|\phi)\pi(\phi)$. This kernel is equivalent to $\exp[-ws(\phi, d) - \log(\pi(\phi^*)/\pi(\phi))]$, where $s(\phi, d) = -\log(f(d|\phi))$, $w = 1$ and ϕ^* is a suitably chosen value of ϕ. Bissiri et al., 2016 and Bissiri and Walker, 2019 develop a general Bayesian method which generalizes the Bayes' rule to facilitate updating a prior pdf by loss information on the parameter ϕ. We review their solution.

If a nonnegative function $l(\phi, d)$ is the loss information on ϕ based on data d, Bissiri et al., 2016 suggests a "general Bayes" method to update a prior pdf $\pi(\phi)$ by the normalized version of the kernel $\exp[-wl(\phi, d) - \log(\pi(\phi^*)/\pi(\phi))]$, where $w > 0$ is an appropriate constant to calibrate the loss information so that the components $wl(\phi, d)$ and $-\log(\pi(\phi))$ are on a comparable scale. When $l(\phi, d)$ is $s(\phi, d)$, the *self-information loss*, $w = 1$ is the natural choice to combine the data loss with the prior loss $-\log(\pi(\phi))$ to update the prior distribution. In our application, the loss function $Q_0(\boldsymbol{\beta}, A, \mathbf{y})$ is not a self-information loss. Bissiri et al., 2016 (cf. Section 3) argues that when different types of loss functions are combined, calibration is important.

One way to combine two losses $l(\phi, d)$ and $-\log(\pi(\phi))$ is based on a prior evaluation of the expected value of $l(\phi, d)$ to determine w for the calibration Bissiri et al., 2016, Section 3.2. To pick w, solve

$$wE[l(\phi, d)] = E_\pi[\log(\pi(\check{\phi})/\pi(\phi))], \tag{2.5}$$

where the left hand expectation $E[\cdot]$ is with respect to a joint distribution for d and ϕ, say $m(d,\phi)$, and the resulting marginal for ϕ is $\pi(\phi)$. Specifically, if d given ϕ has mean and variance ϕ and σ^2, respectively, and if $\pi(\phi)$ is normal with mean η and variance τ^2, then for squared error loss $l(\phi,d) = (\phi - d)^2$ and the scale w is $1/(2\sigma^2)$. This calibration relates the loss $wl(\phi,d)$ to the self-information loss for normally distributed data. The scaling constant w is independent of both η and τ^2. Hence, by taking a very large τ^2, one can approach a uniform prior loss for ϕ and calibrate the squared error loss function relative to this approximate prior.

2.2.2 Calibration for $\boldsymbol{\beta}$: pseudo-posterior of $\boldsymbol{\beta}$ given A

We calibrate $Q_0(\boldsymbol{\beta}, A, \mathbf{Y})$, the first component of our loss function. The i^{th} component of $Q_0(\boldsymbol{\beta}, A, \mathbf{Y})$ is $Q_i(\boldsymbol{\beta}, A, Y_i) = \gamma_i^2(Y_i - \boldsymbol{x_i}^T\boldsymbol{\beta})^2$. For a multivariate normal prior for $\boldsymbol{\beta}$ with mean η and variance $\boldsymbol{\Omega}$, and using an induced prior for $\boldsymbol{x_i}^T\boldsymbol{\beta}$, we find that the scaling component w_i corresponding to $Q_i(\boldsymbol{\beta}, A, Y_i)$ satisfies $w_i\gamma_i^2(D_i + A) = 1/2$. Assuming that each $Q_i(\boldsymbol{\beta}, A, Y_i)$ has approximately equal weight w_β, then an approximation is

$$w_\beta = \frac{m}{2}[\sum_{i=1}^{m}\gamma_i^2(D_i + A)]^{-1}$$

Bissiri et al., 2016.

This choice of w_β is free from p, the dimension of $\boldsymbol{\beta}$. It is also free of η and $\boldsymbol{\Omega}$. By taking $\boldsymbol{\Omega}$ “very large”, we can use the traditional uniform prior for $\boldsymbol{\beta}$. Thus the posterior for $\boldsymbol{\beta}$ is

$$\pi_{ps}(\boldsymbol{\beta}|A, \mathbf{y}) \propto \exp[-w_\beta Q_0(\boldsymbol{\beta}, A, \mathbf{y})], \tag{2.6}$$

see Bissiri et al., 2016. The above posterior is multivariate normal with mean $\tilde{\boldsymbol{\beta}}(A)$ and variance $(2w_\beta\mathbf{X}^T\boldsymbol{\Gamma}^2\mathbf{X})^{-1}$. For the special case of a balanced application, the w_β is $(D + A)/[2D^2]$, which reduces $wQ_0(\boldsymbol{\beta}, A, \mathbf{Y})$ to the self-information loss and finds the usual Bayesian solution for $\boldsymbol{\beta}$ given A.

2.2.3 Approximate calibration of loss function for A

We saw that two-step minimization of the loss function $Q(\boldsymbol{\beta}, A, \mathbf{y})$ gives the loss function $Q_*(A) = Q(\tilde{\boldsymbol{\beta}}(A), A, \mathbf{Y})$. Since $Q_*(A)$ does not go to ∞ as $A \to \infty$, the pseudo-likelihood of A does not go to zero as $A \to \infty$. One thus needs a proper prior for A to update it to a proper pseudo-posterior. For a prior $\pi_A(A)$ on A, one must calibrate the loss function $Q_*(A)$ by multiplying it by a suitable weight w_A.

Let $\bar{D} = \sum_{i=1}^{m} D_i/m$. Following Bissiri et al., 2016, for a prior $\pi_A(A) = r\bar{D}^r(A + \bar{D})^{-(r+1)}$ with a suitable $r > 0$, one sees that $\int_0^\infty \log[\pi_A(\check{A})/\pi_A(A)]\pi_A(A)dA = (r+1)/r$. We remark that this value depends on this specific prior. For another proper prior $\pi_A(A)$, we need to derive the corresponding value, an we may evaluate it by the Monte Carlo method. We must compute $E[Q_*(A)/m]$, which we do by iterative expectation. Using the first two moments of $\mathbf{Y}$ (given $\boldsymbol{\beta}$, A), we calculate $E[Q(\tilde{\boldsymbol{\beta}}(A), A, \mathbf{Y})|\boldsymbol{\beta}, A]$, which depends only on A. Then, we find the Monte Carlo expectation of this function with respect to the prior on A by drawing 10000 values of A from this prior distribution. Using this value and $(r+1)/r$, we get the calibration coefficient w_A using equation (2.5). Similar analysis for other priors on A will find the corresponding calibration weight w_A.

For the loss function $Q_*(A)$, an update of the prior $\pi_A(A)$ is given by the pseudo-posterior

$$\pi_{ps}(A|\mathbf{y}) \propto (A + \bar{D})^{-(r+1)} \exp[-w_A Q(\tilde{\boldsymbol{\beta}}(A), A, \mathbf{y})]. \tag{2.7}$$

Again, see Bissiri et al., 2016. For the pseudo-posterior variance of θ_i to be finite, one needs only that r be positive. But, one needs $r > 1$ to ensure that a finite posterior pseudo-variance for $\boldsymbol{\beta}$. Finally, one takes $r > 2$ to ensure that the variance of this pseudo-posterior for A is finite.

To make pseudo-Bayesian inference on θ_i, generate A from its pseudo-posterior in (2.7), then use the generated A to get $\boldsymbol{\beta}$ from its conditional pseudo-posterior in (2.6), so that θ_i has a normal distribution with mean $\tilde{\theta}_{i,B}$ in (2.2) and variance $AD_i/(A + D_i)$.

2.3 Implementing Pseudo-Bayesian Inference

We want to estimate θ_i, which, under the Fay-Herriot model 2.1, is $\theta_i = \mathbf{x}_i^T \boldsymbol{\beta} + v_i$. Pseudo-Bayesian inference on θ_i first makes S independent draws $A^{(s)}$, $s = 1, \ldots S$, from the pseudo-posterior $\pi_{\text{ps}}(A|\mathbf{y})$ in (2.7) using rejection sampling. Given $A^{(s)}$, draw $\boldsymbol{\beta}^{(s)}$ from the conditional posterior in (2.6). Use $(\{\boldsymbol{\beta}^{(s)}\}^\top, A^{(s)})^\top$ to get the sth simulation of the ith small-area mean $\theta_i^{(s)}$ as a draw from $\text{N}\left(y_i - \gamma_i^{(s)} \cdot (y_i \mathbf{x}_i^T \boldsymbol{\beta}^{(s)}), D_i(1-\gamma_i^{(s)})\right)$, where $\gamma_i^{(s)} = D_i/(A^{(s)} + D_i)$.

For a suitably large S, the average $\widetilde{\theta}_i^{\text{PB}} = \frac{1}{S}\sum_{s=1}^{S} \theta_i^{(s)}$ is the pseudo-Bayesian estimate of θ_i. One can similarly compute the posterior variance and posterior quantiles to find a credible interval for θ_i.

2.4 Applications: Two Examples

We give two examples and compare the performance of the PBE to the OBP in terms of point estimates and measures of uncertainty.

2.4.1 Hospital Data

Morris and Christiansen, 1996 had data on 23 hospitals, each of which had performed at least 50 kidney transplants during a 27-month period. Jiang et al., 2011 used this data to illustrate the OBP method. Even though the hospital data are observational data, and not survey data, to estimate transplant failure rate for each hospital, the Fay-Herriot model can still be used. Indeed, Jiang et al., 2011 treated the hospitals as small areas and the responses y_i, graft failure rates for kidney transplant operations, can be viewed as direct estimates. The variance for the graft failure rate, D_i, is approximated by $0.2 \times 0.8 = 0.16/n_i$, where n_i is the number of kidney transplants at the i^{th} hospital (0.2 is the pooled failure rate).

A severity index x_i is available for each hospital. It is the fraction of females, blacks, children, and extremely ill kidney recipients at hospital i. These variables are shown in Table 2.1. The table also includes $\sqrt{D_i}$. Figure 2.1 is scatter plot of y_i vs. x_i, with an overlay of some fitted regression models.

Table 2.1: The PBE and OBP predictions with various measures of uncertainty for the hospital data.

Hospitals	x	y	PBEc	OBPc	$\sqrt{D}$	Post SDc	Root MSPEc	Length CrIc	Length CIc	PBEo	OBPo	Post SDo	Root MSPEo	Length CrIo	Length CIo
1	0.112	0.302	0.248	0.246	0.055	0.023	0.027	0.0904	0.1036	0.237	0.239	0.023	0.030	0.0899	0.1157
2	0.206	0.140	0.1821	0.183	0.053	0.021	0.028	0.0849	0.1039	0.180	0.180	0.022	0.030	0.0862	0.1157
3	0.104	0.203	0.230	0.230	0.052	0.021	0.028	0.0860	0.1041	0.220	0.220	0.022	0.030	0.0835	0.1158
4	0.168	0.333	0.241	0.238	0.052	0.022	0.028	0.0862	0.1041	0.247	0.249	0.022	0.030	0.0853	0.1158
5	0.337	0.347	0.347	0.348	0.047	0.041	0.028	0.1596	0.1052	0.349	0.347	0.043	0.030	0.1672	0.1159
6	0.169	0.216	0.229	0.229	0.046	0.020	0.028	0.0784	0.1055	0.234	0.234	0.021	0.030	0.0838	0.1159
7	0.211	0.156	0.177	0.178	0.046	0.021	0.028	0.0839	0.1055	0.171	0.172	0.022	0.030	0.0848	0.1159
8	0.195	0.143	0.193	0.194	0.046	0.019	0.028	0.0770	0.1055	0.198	0.197	0.019	0.030	0.0759	0.1159
9	0.221	0.220	0.174	0.173	0.044	0.023	0.028	0.0921	0.1061	0.161	0.162	0.028	0.030	0.1073	0.1159
10	0.077	0.205	0.173	0.173	0.044	0.025	0.028	0.0973	0.1061	0.177	0.181	0.026	0.030	0.0991	0.1159
11	0.195	0.209	0.201	0.200	0.042	0.019	0.028	0.0744	0.1067	0.204	0.206	0.020	0.030	0.0811	0.1158
12	0.185	0.266	0.219	0.217	0.041	0.019	0.028	0.0753	0.1070	0.229	0.228	0.019	0.030	0.0733	0.1158
13	0.202	0.240	0.197	0.195	0.041	0.020	0.028	0.0777	0.1070	0.200	0.201	0.021	0.030	0.0828	0.1158
14	0.108	0.262	0.242	0.241	0.036	0.021	0.028	0.0816	0.1090	0.233	0.235	0.022	0.029	0.0846	0.1152
15	0.204	0.144	0.181	0.182	0.036	0.019	0.028	0.0763	0.1090	0.180	0.180	0.020	0.029	0.0729	0.1152
16	0.072	0.116	0.146	0.147	0.035	0.026	0.029	0.1020	0.1094	0.153	0.154	0.026	0.029	0.1020	0.1150
17	0.142	0.201	0.242	0.243	0.033	0.020	0.029	0.0803	0.1103	0.236	0.236	0.020	0.029	0.0805	0.1144
18	0.136	0.212	0.244	0.245	0.032	0.021	0.029	0.0806	0.1108	0.238	0.238	0.020	0.029	0.0756	0.1140
19	0.172	0.189	0.219	0.220	0.031	0.017	0.029	0.0689	0.1112	0.223	0.223	0.018	0.029	0.0706	0.1135
20	0.202	0.212	0.195	0.194	0.029	0.018	0.029	0.0707	0.1120	0.198	0.199	0.017	0.029	0.0667	0.1122
21	0.087	0.166	0.192	0.193	0.029	0.019	0.029	0.0749	0.1120	0.186	0.187	0.021	0.029	0.0806	0.1122
22	0.177	0.173	0.209	0.211	0.027	0.017	0.029	0.0699	0.1126	0.212	0.211	0.017	0.028	0.0658	0.1103
23	0.072	0.165	0.155	0.155	0.025	0.023	0.029	0.0905	0.1128	0.163	0.165	0.023	0.028	0.0893	0.1077

[a]

[a] The PBEs and OBPs, with estimated uncertainty, for the hospital data with cubic and the Q-O models. Here, column names x, y and $\sqrt{D}$ are obvious. The other columns, for example, PBEc, Post SDc and length CrIc are point estimate, posterior SD and length of credible intervals for the pseudo-Bayes method for the cubic model. The OBPc, Root MSPEc and CIc are point estimate, square root of estimated MSPE and length of confidence intervals for the OBP method for the cubic model. Analogously, the columns with a suffix o correspond to similar quantities from the quadratic outlying model.

As in Jiang et al., 2011, we assume the sampling error distribution is normal. The scatter plot in Figure 2.1 suggests a quadratic model, but with the fifth observation (in the upper right corner) being an outlier. To handle that datum, one can use either the cubic Fay-Herriot model, with $\mathbf{x}_i^T\boldsymbol{\beta} = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3$, or the quadratic-outlying (Q-O) model $\mathbf{x}_i^T\boldsymbol{\beta} = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + d \cdot I(x_i > 0.3)$, where $I(\cdot)$ is the indicator function Jiang et al., 2011. The Q-O model has a "jump" in the response when x_i is greater than 0.3 which is only for the fifth hospital.

In both models, the fifth observation is highly influential. From the design matrices, the leverage diagnostic is 1 for the Q-O model, and 0.997 for cubic regression. Table 2.1 shows the OBP estimates, their root mean squared prediction errors (RMSPE) and lengths of approximate 95% prediction intervals for both the cubic and the Q-O models, with the *c* suffix referring to the cubic model estimates (e.g., OBPc) and the *o* suffix referring the Q-O model. The table includes the PBE estimates, their posterior SDs and lengths of 95% equitailed credible intervals for both models. The 95% PBE credible intervals are obtained from the 0.025 and 0.975 quantiles of the posterior predictive distributions. These intervals are nearly identical to counterparts obtained as the *posterior mean* $\pm$ 1.96$\times$ *posterior SD*.

This paper uses $r = 1$, so the prior median of A is $\bar{D}$. In the absence of other information, it is a reasonable choice. We investigated another prior for A, and the PBE results are materially the same for $r = 0.1$.

We derive Bayesian estimates from 5000 samples of A drawn from its pseudo-posterior distribution. Figure 2.2 shows the histogram. The PBE of A is the posterior median. For the cubic model, the PBE is $\widetilde{A}^{\text{PB}} = 2.28 \times 10^{-4}$, which is a bit larger than $\widetilde{A}^{\text{BPE}} = 1.93 \times 10^{-4}$, the BPE from Jiang et al., 2011. The 5000 samples of A lead to 5000 sets $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)^\top$. We show posterior histograms of components of $\boldsymbol{\beta}$ in Figure 2.3. The histograms show that each regression parameter is nonzero with a high probability. The PBE of $\boldsymbol{\beta}$ is $\widetilde{\boldsymbol{\beta}}^{\text{PB}} = (-0.376, 11.462, -65.639, 112.741)^\top$, which is close to the BPE estimate $\widetilde{\boldsymbol{\beta}}^{\text{BPE}} = (-0.366, 11.268, -64.595, 111.112)^\top$.

The 5000 samples of $(\boldsymbol{\beta}^T, A)^T$ generate 5000 independent samples of $\boldsymbol{\theta} = (\theta_i)_{1 \leqslant i \leqslant m}$ from its conditional posterior. From those simulated values, we obtain $\widetilde{\theta}_i^{\text{PB}}$, the PBE of θ_i, and a 95% credible interval on it (the mean and the 0.025 and 0.975 quantiles). The PBE and posterior standard deviation of θ_i are given in Table 2.1: the left side is the cubic model, the right is the Q-O model.

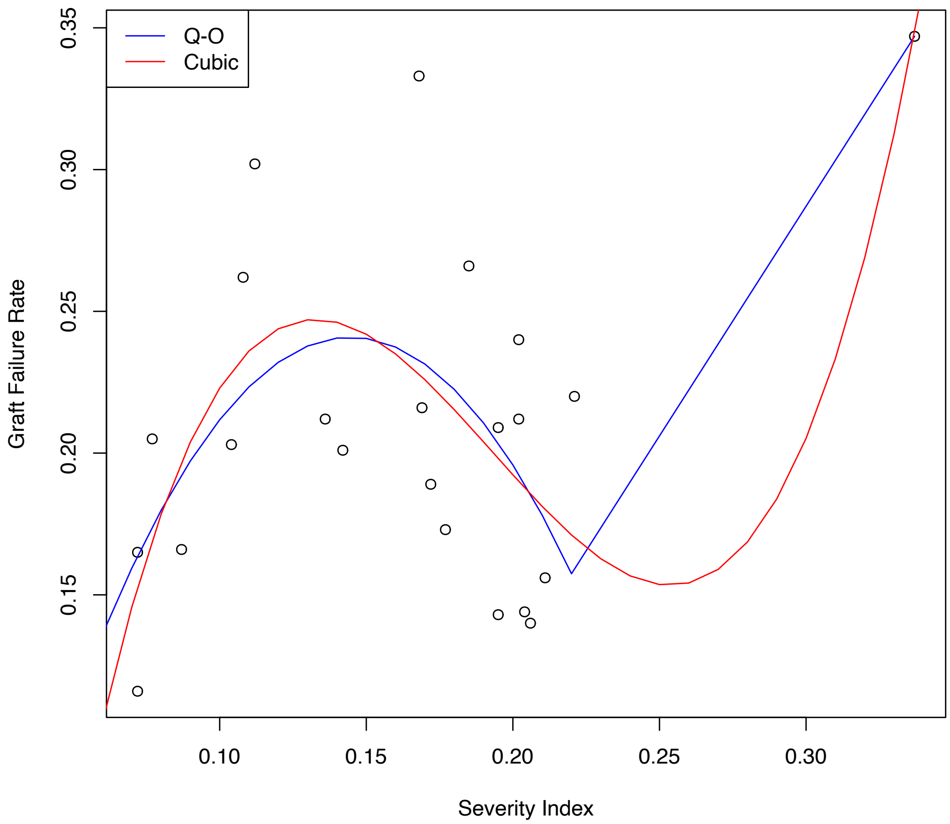


Figure 2.1: Scatter plot of graft failure rate vs. severity index is overlaid with the cubic and Q-O fitted models. In the figure, the black dots are data points, the red line is the cubic fitted model, and the blue line is the Q-O fitted model.

The 95% credible intervals from the cubic model are shown in Figure 2.4. Table 2.1 also gives $\widetilde{\theta}_i^{\text{OBP}}$, the OBP of θ_i. Again, the cubic and the Q-O models have PBE and the OBP estimates that are essentially the same. Such small differences suggest that the OBP and the PBE estimators have similar biases. Simulations in Section 2.5 find that both estimators are actually unbiased under the correct model, and are similarly but only moderately biased for a misspecified model.

Jiang et al., 2011, eq. (46) uses a second-order unbiased estimator of the MSPE of $\widetilde{\theta}_i^{\text{OBP}}$ to quantify uncertainty in the OBP. We found it satisfactory. It is highly variable, often underestimates the truth, and is sometimes negative. To correct these flaws, X. Liu et al., 2022, eq.3.2) develops

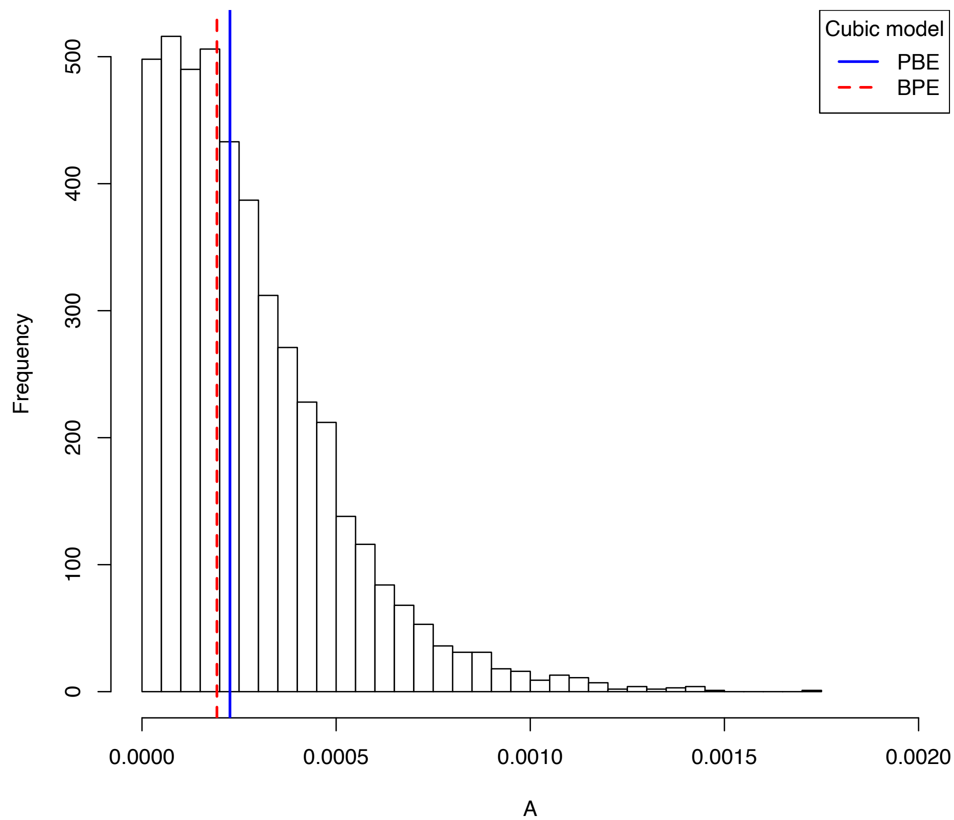


Figure 2.2: Posterior histogram of of A for the cubic model. A dotted red vertical line is drawn at BPE of A, and a solid blue vertical line is drawn at the PBE of A.

a new MSPE estimator. Let $\hat{\boldsymbol{\beta}}$ and $\hat{A}$ denote the BPE estimators. Define $\hat{e}_i = y_i - \boldsymbol{x_i}^\top \hat{\boldsymbol{\beta}}$, and $\hat{\delta}_i = (\hat{A} + D_i)^{-1}$. Denoting $D_i/(D_i + \hat{A})$ by $\hat{B}_i$, define $\hat{t}_m = \sum_{i=1}^m \hat{\delta}_i^2$, $\hat{u}_{1m} = p \sum_{i=1}^m \hat{B}_i^4$, $\hat{T}_m = \sum_{i=1}^m (\hat{\delta}_i^2 \hat{e}_i^4 - 3)$, $\hat{s}_{km} = \sum_{i=1}^m \hat{\delta}_i^k \hat{B}_i^2$, $k = 0, 1, 2$, and $\hat{V}_{km} = \sum_{i=1}^m \hat{\delta}_i^k \hat{B}_i^4 (\hat{\delta}_i^2 \hat{e}_i^4 - 1)$, $k = 0, 1$. Further define

$$\hat{a}_i = \frac{\hat{T}_m \hat{B}_i^4 \hat{\delta}_i^2}{\hat{s}_{1m} \hat{t}_m}, \quad \hat{b}_i = \hat{B}_i^2 \{ \frac{2\hat{u}_{1m}}{\hat{s}_{0m} \hat{s}_{1m}} + \frac{3}{\hat{s}_{1m}^3} (\hat{s}_{1m} \hat{V}_{1m} - \hat{s}_{2m} \hat{V}_{0m}) \} + \frac{2\hat{B}_i^2 \hat{\delta}_i \hat{V}_{0m}}{\hat{s}_{1m}^2}.$$

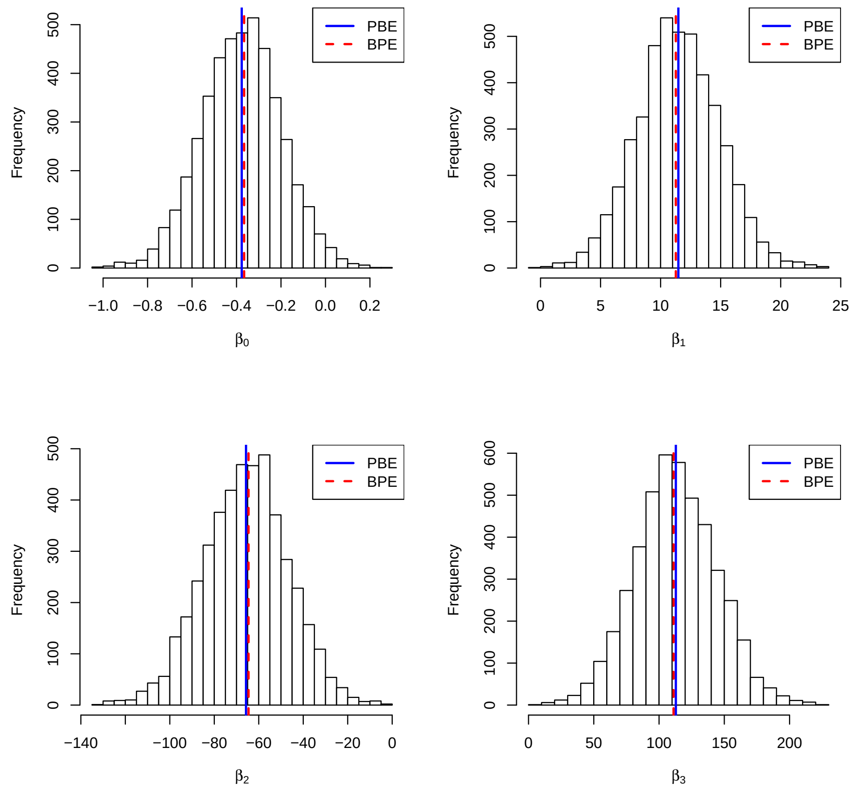


Figure 2.3: Posterior histograms of $\beta_0, \beta_1, \beta_2$ and β_3 (cubic model). On each histogram a dotted red vertical line is drawn at BPE, and a solid blue vertical line is drawn at PBE.

The new estimator from X. Liu et al., 2022 is

$$\widehat{MSPE}(\widetilde{\theta}_i^{\text{OBP}}) = \hat{A}\hat{B}_i - 2\hat{a}_i + \hat{b}_i. \tag{2.8}$$

Like other MSE estimators (maximum likelihood (ML), residual maximum likelihood (REML), the EBLUPs of Fay and Herriot, 1979 and Prasad and Rao, 1990), the new estimator of the MSE of the OBP does not depend on an individual observation y_i. Unlike these frequentist estimators, the posterior variance of θ_i does depend on y_i.

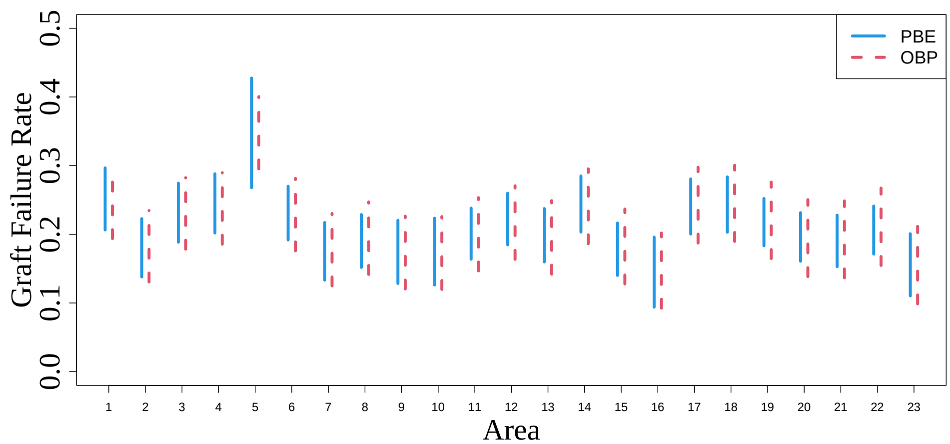


Figure 2.4: 95% credible/prediction intervals of $\boldsymbol{\theta}$ under the cubic model. Blue solid lines are PBE intervals and red dashed lines are OBP intervals.

Given the OBP $\widetilde{\theta}_i^{\text{OBP}}$ and its RMSPE estimator $\widetilde{\text{RMSPE}}(\widetilde{\theta}_i^{\text{OBP}})$ from equation (2.8), an approximate $100(1-\alpha)\%$ prediction interval for θ_i is

$$\left(\widetilde{\theta}_i^{\text{OBP}} - z_{\alpha/2} \cdot \widetilde{\text{RMSPE}}(\widetilde{\theta}_i^{\text{OBP}}),\ \widetilde{\theta}_i^{\text{OBP}} + z_{\alpha/2} \cdot \widetilde{\text{RMSPE}}(\widetilde{\theta}_i^{\text{OBP}})\right), \tag{2.9}$$

where $z_{\alpha/2}$ is the $\alpha/2$ quantile of standard normal distribution. For the hospital data, OBP 95% prediction intervals of the form (2.9) are shown in Figure 2.4. The PBE and OBP point estimates of $\boldsymbol{\theta}$ are very similar, but the 95% credible intervals are generally shorter than the 95% prediction intervals. The average length of the OBP intervals is 0.1078, 27% longer than the average length of the credible intervals.

Even though both OBP and PBE intervals have similar centers, their intervals are dissimilar, because of their different measures of uncertainty. The posterior standard deviations are generally smaller than the RMSPEs of the OBP predictors (except for the fifth observation). Unusually, the RMSPEs for the last four small areas are at least as large as their sampling standard deviations (SDs); consequently, there is no gain from using the OBP predictors. On the other hand, the RMSPE for the fifth hospital is 0.027, unexpectedly and unreasonably smaller than both the posterior SD (0.041) and the sampling SD (0.043). There are a few outliers in the scatter plot. But the fifth observation is highly influential. Its leverage diagnostic is 0.9973, and it substantially attracted the estimated regression function to the observed value of $y_5 = 0.347$. Thus the OBP estimate of θ_5 is 0.348, which is hardly any different from y_5. This makes us wonder if the estimated RMSPE, 0.027, is unrealistically small.

In the Q-O model, the row vector $\boldsymbol{x_5}^\top$ for this model is linearly independent of the other rows of the design matrix, so the corresponding leverage diagnostic is 1. For any $m \times m$ positive definite diagonal matrix $\boldsymbol{W}$, it holds that $\boldsymbol{x_5}^\top(\boldsymbol{X}^\top \boldsymbol{W}\boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{W} = (0, \cdots, 0, 1, 0, \cdots, 0)$; the fifth element is 1. Thus the OBP of θ_5 must be y_5 and so its MSPE will be D_5, which is known. However, the estimated MSPE from equation (2.8) for $i = 5$ does not simplify to D_5. For the PBE, it can be shown that $\theta_5|A, \sim N(y_5, \tau_5)$, where $\tau_5 = \frac{AD_5}{A+D_5} + \frac{1}{m}\sum_{i=1}^{m} \frac{D_i^2}{D_i+A}$, which is free of the design matrix . Thus, the posterior mean or the median of θ_5 is also y_5. Our PBE estimate of θ_5 is slightly different from y_5 because we approximate the posterior by simulation. Moreover, the posterior variance of θ_5 is the posterior mean of τ_5, which is not necessarily equal to D_5. Since the OBP and PBE of θ_5 are exactly or nearly equal to y_5, it would be unreasonable to have a sharp disagreement among their uncertainty estimates. While the posterior SD 0.043 is nearly equal to the sampling SD 0.047, the estimated RMSPE value is only 0.030 (certainly a substantial underestimate of the correct RMSPE). The reason for this underestimation is perhaps due to the inadequacy of the asymptotic MSPE estimator in eqn. (3.2) of Y. Liu et al., 2022. While the estimated RMSPEs for the fifth and the last 4 hospitals appear to be unusual, the posterior SDs appear to be more reasonable measures of uncertainty.

2.4.2 Median Incomes of 4-person Families by State

The second example is to estimate median incomes of four-person families for the 50 U.S. states and the District of Columbia ($m = 51$ small areas). These estimates are used by the U.S. Department of Health and Human Services to run an energy assistance program for low-income families. The U.S. Census Bureau produces these estimates annually from the Current Population Survey (CPS). However, CPS sample sizes are small, so direct estimates are too variable except for ten states, which have bigger samples and are known as *direct use states*.

We estimate the 1989 median income of four-person families using the 1990 CPS data and compare the estimates with the values obtained from the 1990 decennial population census, which also collected income data in 1989 for a large number families. The census median incomes are based on very large samples collected from about 15% families receiving the *long version of the census questionnaire to report income*; these estimates have negligible standard errors. We consider the Census data as ground truth, that is *true* θ_i when evaluating the accuracy of the small area estimates that we study.

Let $y_i,\ i = 1, \ldots, m$, be the direct estimate of the 1989 four-person family median income of the ith state from the CPS. As before, we fit a Fay-Herriot model; see 2.1. We use two covariates that were suggested by Fay, 1987:

(1) x_{i1}: 1979 four-person family median income of the ith state from the 1980 census,

(2) $x_{i2} = (\mathrm{PCI}_{i,1989}/\mathrm{PCI}_{i,1979}) \cdot x_{i1}$: 1989 adjusted census four-person family median income of the ith state,

where PCI is the per capita income from the U.S. Bureau of Economic Analysis. The mean function of the Fay-Herriot model is $\mathbf{x}_i^T \boldsymbol{\beta} = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}$.

To use our pseudo-Bayesian method, take the prior for A proportional to $(A + \bar{D})^{-2}$. Then make 5000 independent draws of $(\boldsymbol{\beta}^\top, A)^\top = (\beta_0, \beta_1, \beta_2, A)^T$ from its posterior distribution. Histograms of $(\beta_0, \beta_1, \beta_2, A)$ are shown in Figure 2.5. The PBE of $(\boldsymbol{\beta}^T, A)^\top$ is $(12894, 0.334, 0.784, 5517221)^\top$. The BPE of $(\boldsymbol{\beta}^T, A)^\top$ is $(13018, 0.335, 0.783, 5645033)^T$. Histograms in Fig. 2.5 show that the differences between the two sets of estimates are insignificant.

Finally, using the draws of $(\boldsymbol{\beta}^T, A)^T$, 5000 posterior simulations of $\boldsymbol{\theta} = (\theta_i)_{1 \leqslant i \leqslant m}$ are drawn independently from the conditional posterior of $\boldsymbol{\theta}$. We use these draws to compute PBE point estimates and 95% credible intervals.

In addition to the PBE and OBP, we compute the EBLUP of $\boldsymbol{\theta}$. There are four widely used EBLUPs in small area estimation, depending on the estimator of A: the ML, REML, FH, and PR EBLUPs. We denote these EBLUPs by $\widehat{\boldsymbol{\theta}}^{\mathrm{ML}} = (\widehat{\theta}_i^{\mathrm{ML}})_{1 \leqslant i \leqslant m}$, $\widehat{\boldsymbol{\theta}}^{\mathrm{REML}} = (\widehat{\theta}_i^{\mathrm{REML}})_{1 \leqslant i \leqslant m}$, $\widehat{\boldsymbol{\theta}}^{\mathrm{FH}} = (\widehat{\theta}_i^{\mathrm{FH}})_{1 \leqslant i \leqslant m}$, and $\widehat{\boldsymbol{\theta}}^{\mathrm{PR}} = (\widehat{\theta}_i^{\mathrm{PR}})_{1 \leqslant i \leqslant m}$, respectively. See Prasad and Rao, 1990, Datta and

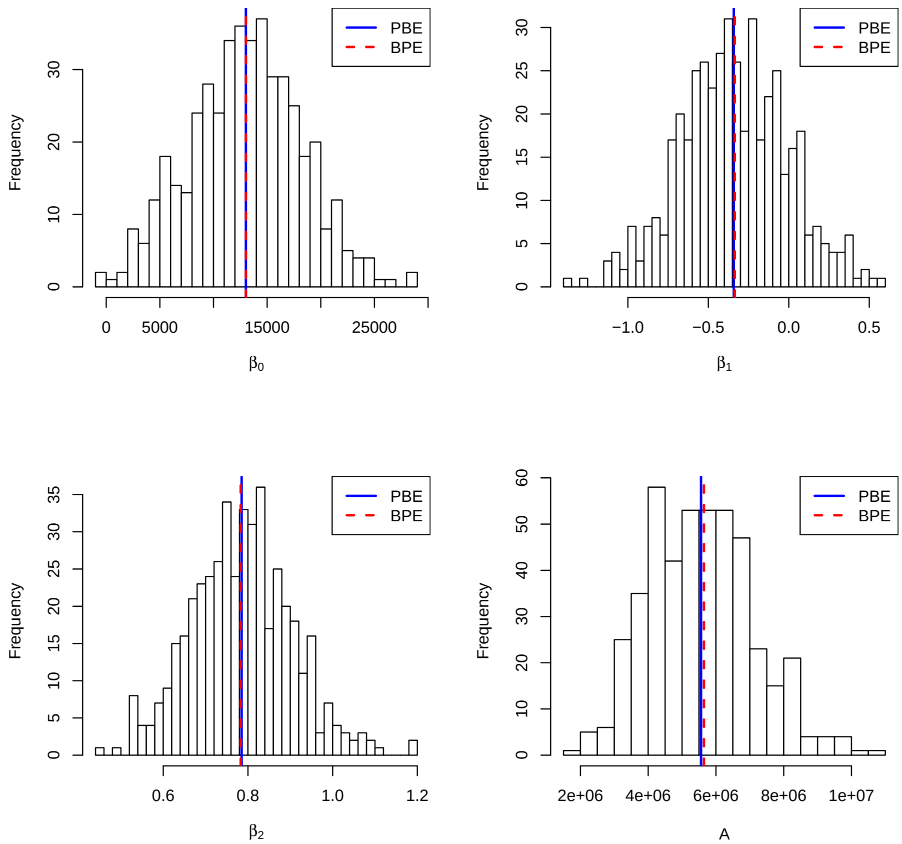


Figure 2.5: Posterior histograms of $\beta_0, \beta_1, \beta_2$ and A for median incomes example. On each histogram a dotted red vertical line is drawn at BPE, and a solid blue vertical line at PBE.

Lahiri, 2000, and Datta et al., 2005 for second-order unbiased estimators of the MSPEs of these EBLUPs.

Let $\widehat{\text{MSPE}}(\widehat{\theta}_i)$ denote those MSPE estimators, where $\widehat{\theta}_i$ is one of the four EBLUPs of θ_i. Let $\widehat{\text{RMSPE}}(\widehat{\theta}_i) = \sqrt{\widehat{\text{MSPE}}(\widehat{\theta}_i)}$. Given $\widehat{\theta}_i$ and $\widehat{\text{RMSPE}}(\widehat{\theta}_i)$, an approximate $100(1-\alpha)\%$ prediction interval for θ_i is $(\widehat{\theta}_i - z_{\alpha/2} \cdot \widehat{\text{RMSPE}}(\widehat{\theta}_i),\ \widehat{\theta}_i + z_{\alpha/2} \cdot \widehat{\text{RMSPE}}(\widehat{\theta}_i))$. To construct the OBP confidence intervals we use the estimated MSPE from equation (2.8) (which is equation (3.2) of Y. Liu et al., 2022).

We compare the direct estimates and the model based estimates from four EBLUPs, the OBP and the PBE methods, in terms of their proximity to the truth, treating the 1990 census median incomes as the true values of $\theta_i,\ i = 1, \dots, m$. For each estimator $\breve{\boldsymbol{\theta}} = (\breve{\theta}_i)_{1 \leqslant i \leqslant m}$ of $\boldsymbol{\theta} = (\theta_i)_{1 \leqslant i \leqslant m}$, we compute four deviation measures:

$$\begin{cases} \text{average absolute deviation:AAD}(\breve{\boldsymbol{\theta}}) = m^{-1} \sum_{i=1}^{m} |\breve{\theta}_i - \theta_i| \\ \text{average squared deviation:ASD}(\breve{\boldsymbol{\theta}}) = m^{-1} \sum_{i=1}^{m} (\breve{\theta}_i - \theta_i)^2 \\ \text{average absolute relative deviation:AARD}(\breve{\boldsymbol{\theta}}) = m^{-1} \sum_{i=1}^{m} |(\breve{\theta}_i - \theta_i)/\theta_i| \\ \text{average squared relative deviation:ASRD}(\breve{\boldsymbol{\theta}}) = m^{-1} \sum_{i=1}^{m} \{(\breve{\theta}_i - \theta_i)/\theta_i\}^2 \end{cases} \tag{2.10}$$

We also compare the above estimation methods based on interval estimates. Suppose for a method M a 95% prediction interval for θ_i is $(l_{i,M}, u_{i,M})$. Based on the ground truth θ_i, we also compute empirical average length (AL), empirical coverage probability (CP) and empirical interval score (IS) for an interval estimator. While the first two quantities are well known, the third quantity is inspired by the interval score in eqn. (43) of Gneiting and Raftery, 2007. We compute this score using

$$IS_M = \frac{1}{m} \sum_{i=1}^{m} [u_{i,M} - l_{i,M}] + \frac{\sum_{i=1}^{m} (l_{i,M} - \theta_i) I(\theta_i < l_{i,M})}{\max[1, \sum_{i=1}^{m} I(\theta_i < l_{i,M})]} + \frac{\sum_{i=1}^{m} (\theta_i - u_{i,M}) I(u_{i,M} < \theta_i)}{\max[1, \sum_{i=1}^{m} I(u_{i,M} < \theta_i)]}. \tag{2.11}$$

We report the four deviation measures for point estimates and three interval estimation measures in Table 2.2 for all our seven estimates. The row “Direct" provides these values for the sample-based (direct) estimates from the CPS.

The direct estimates are poor compared to the model-based estimates in terms of all seven measures. It may be partly due to differences between the wording of income questions in the CPS and the Census. Several experienced Census Bureau researchers think the regression function used in the model-based estimate is effective in explaining the true median values. Since all six model-based estimators are based on this effective mean function, they perform substantively better than the direct estimates for all seven measures. Also, since the sampling variances D_i of the direct estimators are usually larger than the estimated model error variance A, the EBLUP estimates of θ_i

Table 2.2: Comparison of various estimators for the median income data.

Estimate	AAD	ASD	AARD	ASRD	AL	CP	IS
Direct	2928.8	13.81×10^6	0.0735	0.0084	11424.6	0.9020	13541.9
ML	1394.9	3.19×10^6	0.0348	0.0019	7710.5	0.9608	8789.9
REML	1454.9	3.45×10^6	0.0363	0.0021	7716.3	0.9608	8866.0
FH	1464.4	3.49×10^6	0.0365	0.0021	7762.1	0.9608	8901.9
PR	1438.1	3.38×10^6	0.0359	0.0020	7824.6	0.9608	8905.9
OBP	1652.4	4.15×10^6	0.0420	0.0026	8058.3	0.9412	8718.9
PBE	1610.4	3.98×10^6	0.0409	0.0025	7165.6	0.9608	8177.7

Columns $2-5$ are defined in eqn. (2.10) and the last column in eqn. (2.11). Columns AL and CP give average length and average coverage probability, respectively.

tend to assign less weight to the direct estimates compared to the OBP and the PBE estimates, and thus the former predictors do better than the latter group in terms of the four deviation measures. In terms of these measures the PBE method is better than the OBP method; these measures for the PBE estimates are less than those for the OBP estimates by 2.5%, 4.1%, 2.6% and 3.8%, respectively. We also consider the respective predictive/credible intervals for θ_is obtained from the four EBLUPs, OBP, and PBE methods. The last three columns of Table 2.2 present the average length of the 51 estimated intervals (AL), the coverage rate (CP) and the interval score (IS) for each considered method. Here, in terms of the three evaluation measures for interval estimation, the PBE method performs uniformly better than the remaining methods. Finally, despite having the highest average length, the intervals associated with the direct estimates are severely anti-conservative (approximately 90% CP) and have the worst performance. In contrast, the model-based intervals have nearly nominal coverage rates for all methods.

Figure 2.6 displays 95% prediction/credible intervals. It includes only the ML EBLUP as it is the best of the EBLUPs in Table 2.2. As demonstrated in this figure, the EBLUP, OBP, and PBE intervals cover the true (state) means in the majority of states, with exceptions for CA and DE (all three methods) and AK (OBP and PBE).

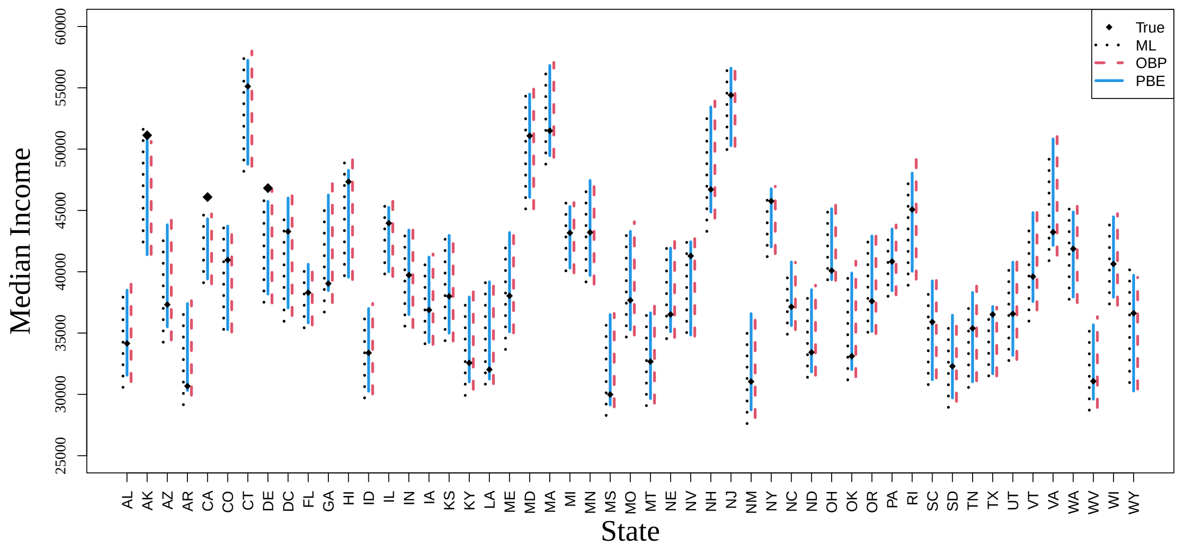


Figure 2.6: 95% prediction/credible intervals for θ_i based on ML EBLUP (black dotted lines), OBP (red dashed lines) and PBE (blue solid lines). Black diamonds are true values, three shown larger which are outside the OBP and PBE intervals.

On an average, the PBE intervals are about 11% shorter than the OBP intervals. From Table 2.2 and Figure 2.6 we conclude that the PBE method, which is inherently a Bayesian method, is competitive with and often superior to the other model-based methods based on these various frequentist evaluation measures.

2.5 Simulation Studies

In our simulations, as in the applications, the prior on A takes $r = 1$. In this section, we present two simulation studies similar to those in Jiang et al., 2011, examining our proposed pseudo-Bayes method as well as the existing methods (i.e., EBLUP and OBP). In the third study we mimic the median incomes estimation problem.

2.5.1 Example 1: A Simple Fay-Herriot Model with no Covariate

Jiang et al., 2011 showed that for a simple Fay-Herriot model with no covariate the gain for OBP over EBLUP can be substantial if the underlying model is misspecified. So we compare the PBE to the EBLUP and OBP under the same setup. Suppose that the true model is

$$y_i = \begin{cases} \mu_1 + v_i + e_i, & 1 \leqslant i \leqslant n \\ \mu_2 + v_i + e_i, & n+1 \leqslant i \leqslant m, \end{cases} \tag{2.12}$$

where $\mu_1 \neq \mu_2$, $m = 2n$, and the sampling variances are $D_i = \sigma_1^2$, $1 \leqslant i \leqslant n$, and $D_i = \sigma_2^2$, $n+1 \leqslant i \leqslant m$. Now fit a misspecified Fay-Herriot model (2.1) with $\mathbf{x}_i^\top \boldsymbol{\beta} = \boldsymbol{\beta}$, which assumes $\mu_1 = \mu_2$. We set $\mu_1 = 0$, $\sigma_1^2 = 4$, $\sigma_2^2 = 1$, while $\mu_2 = 1$ or 5 to produce a mild or a severe case of misspecification. We take $A = 0.2$ for the variance of the v_is and $m = 50$ or 100 small areas. For both scenarios, $K = 500$ simulations are performed.

For each simulation, a data set is generated under the true model (2.12), but a misspecified model with a common intercept is fit using the EBLUP, OBP, and PBE. Let $\boldsymbol{\theta}_{(k)} = (\theta_{i(k)})_{1 \leqslant i \leqslant m}$, $k = 1, \ldots, K$, denote the true $\boldsymbol{\theta} = (\theta_i)_{1 \leqslant i \leqslant m}$ from the kth run, where $\theta_i = \mu_1 + v_i$, $1 \leqslant i \leqslant n$, and $\theta_i = \mu_2 + v_i$, $n+1 \leqslant i \leqslant m$. Also, let $\breve{\boldsymbol{\theta}}_{(k)} = (\breve{\theta}_{i(k)})_{1 \leqslant i \leqslant m}$ denote an estimator of $\boldsymbol{\theta}_{(k)}$. The empirical MSPE for this estimator is given by

$$\text{MSPE}^* = \frac{1}{K} \sum_{k=1}^{K} |\breve{\boldsymbol{\theta}}_{(k)} - \boldsymbol{\theta}_{(k)}|^2 = \sum_{i=1}^{m} \frac{1}{K} \sum_{k=1}^{K} \{\breve{\theta}_{i(k)} - \theta_{i(k)}\}^2 = \sum_{i=1}^{m} \text{MSPE}_i^*.$$

Table 2.3 gives the empirical MSPEs. The numbers in the parentheses are the percentage increase in MSPE for the EBLUP and OBP over the PBE; negative percentages indicate decrease. All four EBLUPs perform similarly. Among the four, the FH EBLUP is slightly better when $\mu_2 = 1$, and the PR EBLUP is slightly better when $\mu_2 = 5$. The OBP and PBE methods have similar MSPEs. The table provides some evidence of approximate equivalence of the OBP and the PBE methods in terms of empirical MSPE. As in Jiang et al., 2011, OBP is substantially better than the EBLUPs, especially for gross misspecification. The same is true for PBE as well, with the percentage increase in MSPE for the EBLUPs ranging between 16% and 43%.

Table 2.3: Empirical MSPEs with percentage increases over PBE (for no-covariate misspecified Fay-Herriot model)

μ_2	m	ML	REML	FH	PR	OBP	PBE
1	50	25.72	25.59	25.56	26.45	22.85	22.11
		(16.2%)	(15.7%)	(15.5%)	(19.5%)	(3.3%)	
1	100	48.83	48.60	48.08	49.51	41.34	40.78
		(19.6%)	(19.0%)	(17.7%)	(21.2%)	(1.2%)	
5	50	99.09	98.27	96.62	96.13	68.76	69.21
		(43.1%)	(41.9%)	(39.5%)	(38.8%)	(−0.7%)	
5	100	188.66	187.95	185.29	184.44	132.40	132.84
		(41.9%)	(41.3%)	(39.3%)	(38.7%)	(−0.4%)	

Columns $3-6$ are for ML, REML, FH and PR EBLUP methods, and columns $7-8$ are for OBP and PBE methods.

We now compare PBE, OBP, and EBLUPs in terms of area-specific MSPEs. Although the OBP is defined by minimizing the overall (observed) MSPE and the PBE is a pseudo-Bayesian alternative to it, there is no guarantee that their area-specific MSPEs are minimal as well. But area-specific MSPEs are important in small area estimation. Moreover, such comparisons provide further insight into the results presented in Table 2.3. Figure 2.7 presents boxplots of empirical area-specific MSPEs. Notice that the OBP and PBE distributions are highly similar, with both being substantively less variable than any of their considered EBLUP counterparts.

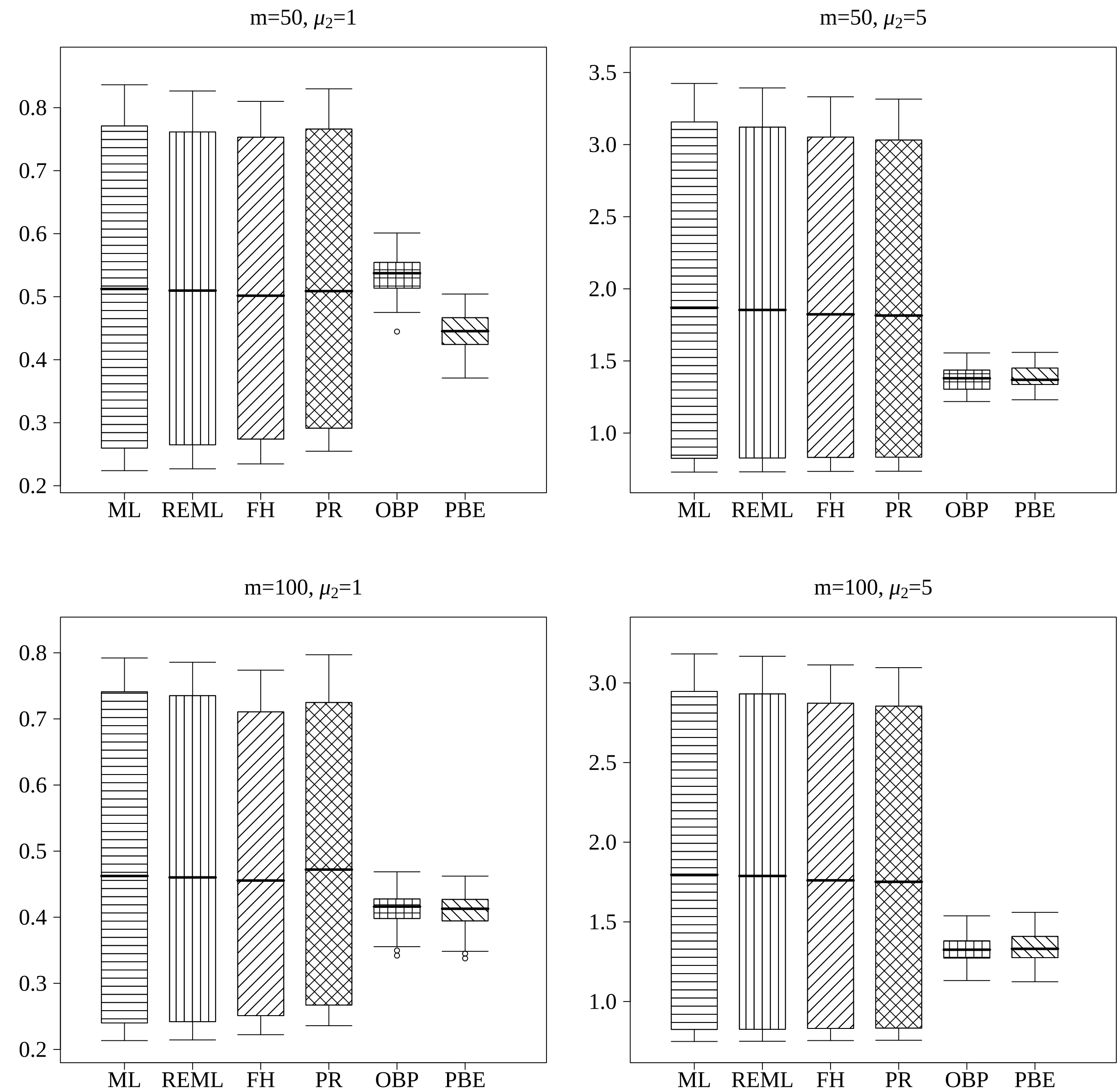


Figure 2.7: Boxplots of area-specific empirical MSPEs for six considered methods for misspecified intercept-only Fay-Herriot model for 50 and 100 areas, and with a minor misspecification: $\mu_2 = 1$ or a major misspecification: $\mu_2 = 5$.

The PBE and OBP are similar to each other, and different from the four EBLUPs. The boxplots show that in three out of four scenarios the OBP, and in all four scenarios the PBE, have median empirical MSPEs smaller than for the EBLUPs. Also, in all four scenarios both the OBP and the PBE have empirical MSPEs with smaller variability than the four EBLUPs.

Table 2.4: Comparison of methods based on 95% prediction/credible intervals (for no-covariate misspecified Fay-Herriot model)

μ_2	m		ML	REML	FH	PR	OBP	PBE
1	50	CP	0.839	0.830	0.854	0.915	0.907	0.909
		AL	2.304	2.268	2.434	3.055	2.701	2.508
		IS	2.670	2.668	2.835	3.319	3.390	2.891
1	100	CP	0.820	0.816	0.854	0.886	0.909	0.911
		AL	2.070	2.058	2.261	2.624	2.514	2.432
		IS	2.472	2.478	2.649	2.932	2.899	2.765
5	50	CP	0.939	0.940	0.943	0.945	0.943	0.919
		AL	4.906	4.908	4.956	4.983	4.942	4.517
		IS	5.465	5.476	5.535	5.573	5.384	5.055
5	100	CP	0.945	0.945	0.948	0.949	0.938	0.923
		AL	4.877	4.878	4.925	4.947	4.765	4.524
		IS	5.502	5.501	5.563	5.595	5.249	5.092

Columns $4-7$ are for ML, REML, FH and PR EBLUPs, and the last two are for OBP and PBE. For each combination of μ_2 and m, CP is 95% coverage probability, AL represents average length, and IS represents interval score defined in eqn. (2.13).

Table 2.4 compares the considered methods in terms of interval estimation by aggregating over all areas. For the kth simulation, denote the $1-\alpha$ prediction interval of θ_i by $(l_{i,M,k}, u_{i,M,k})$ for a method M. Computation of the empirical coverage probability corresponding to θ_i is obvious. With empirical average length $AL_{i,M}$ as $AL_{i,M} = \frac{1}{K}\sum_{k=1}^{K}(u_{i,M,k} - l_{i,M,k})$, we compute empirical mean interval score for predicting θ_i as

$$IS_{i,M} = AL_{i,M} + \frac{\sum_{k=1}^{K}(l_{i,M,k} - \theta_{i,k})I(\theta_{i,k} < l_{i,M,k})}{\max[1, \sum_{k=1}^{K} I(\theta_{i,k} < l_{i,M,k})]} + \frac{\sum_{k=1}^{K}(\theta_{i,k} - u_{i,M,k})I(u_{i,M,k} < \theta_{i,k})}{\max[1, \sum_{k=1}^{K} I(u_{i,M,k} < \theta_{i,k})]}. \tag{2.13}$$

In Table 2.4 we provide averages over small areas of CP, AL and IS measures for six different methods for nominally 95% intervals. For a method M, we report $\frac{1}{m}\sum_{i=1}^{m} CP_{i,M}$, $\frac{1}{m}\sum_{i=1}^{m} AL_{i,M}$ and $\frac{1}{m}\sum_{i=1}^{m} IS_{i,M}$ for CP, AL and IS, respectively. Notice that the coverage is nominal or nearly nominal for the four EBLUPs and OBP methods with more severe model misspecification ($\mu_2 = 5$), whereas the corresponding PBE intervals demonstrate undercoverage (perhaps due to their shorter average length). For severe misspecification, both the average length and the interval score for PBE are less than their counterparts for all competitors.

Table 2.4 shows that for minor misspecification ($\mu_2 = 1$), the PBE is better than the PR EBLUP and the OBP. Here, the ML, REML and the FH EBLUPs have very low coverage, and frequentist methods do not have better coverage than the PBE. On the other hand, the interval score for PBE is greater than those of the ME, REML and FH EBLUPS for both large and small sample sizes. We notice that the mean empirical coverage probabilities of the PBE method are stable at about 91% for all four combinations of misspecification and sample size.

For minor misspecification, the empirical mean lengths of the frequentist intervals vary substantially over m. The range of variation is between 8% and 17%, but it is only 3% for PBE. The reason for this variation in frequentist method is that for smaller sample sizes, more variation in the frequentist estimates of A contributes significantly to the estimated mean squared error. To clarify it, we present the following. If $\hat{\theta}_i$ denotes an EBLUP of θ_i, from eqns. (5)-(6) of Datta et al., 2005 we get that

$$MSE(\hat{\theta}_i) = g_{1i}(A) + g_{2i}(A) + g_{3i}(A) + o(m^{-1}),$$

where $g_{1i}(A) = A\gamma_i$, $g_{2i}(A) = \gamma_i^2 V(\hat{\mu})$ and $g_{3i}(A) = (D_i + A)^{-1}\gamma_i^2 V(\hat{A})$. In this approximation, g_{1i} is the dominating term and the other two terms are lower-order terms. The usual lower-order and less dominating g_{3i} term sometimes contributes substantially more than the "usual dominating order $O(1)$ term g_{1i}" to the MSE estimate. In our simulations g_{3i} is several hundred times larger than g_{1i} for PR EBLUPs in at least 25% of the small areas, at least two times larger for ML EBLUPs in about 25% of the areas, at least 1.6 times larger for REML EBLUPs for about 12% areas, and g_{3i} is larger than g_{1i} for FH EBLUPs for about 10% areas. Thus, while the ML, REML or the FH intervals are shorter than the PBE intervals, their coverage probabilities are as low as 0.85.

2.5.2 Example 2: A Simulation Study Imitating the Hospital Data

The second simulation study imitates the hospital data considered by Jiang et al., 2011. We mentioned earlier that there is a potential outlier in the hospital data, which is the point at the upper right corner of the scatter plot in Figure 1 of main manuscript. To incorporate this outlying observation in their analysis, Jiang et al., 2011 considered a quadratic outlying Fay-Herriot model by adding an extra intercept term to the model for the outlier. However, we do not consider this to be a good practice since this model prevents the outlier from playing any role in model-fitting. This amounts to fitting separate models for the two groups; it forces the direct estimator for the outlier to be its fitted value, and fits the quadratic regression model based on the remaining 22 data. Accordingly, we exclude the covariate for the outlier observation and conduct a simulation based on the remaining 22 observations to generate data using the model

$$y_i = \theta_i + e_i,\ \theta_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + v_i,\ i = 1, \ldots, m, \tag{2.14}$$

where $v_i \overset{\text{iid}}{\sim} \text{N}(0, A)$ independent of $e_i \overset{\text{ind}}{\sim} \text{N}(0, D_i)$ to generate data, and $D_i's$ are the same as in the real hospital data without outlier (see Table 1 of Section 4.1 of main manuscript). Following these authors we used $\beta_0 = -1.1, \beta_1 = 20, \beta_2 = -50$ and $A = 0.0016$.

For the actual model to be fitted to the data, we consider two scenarios, namely, a misspecified case (linear in x_i) by eliminating the quadratic term, and a correctly specified model. We consider moderate misspecification (Mis) by taking $\beta_2 = -50$ to generate θ_i in (2.14). We note that the

MSPEs of any of these predictors (the EBLUPs, OBP or the PBE) when we fit the correctly specified model do not depend on the true values of the regression parameters.

Table 2.5 presents the total empirical MSPEs (multiplied by 100), with percentage increases over the PBE estimates are reported in parantheses below. The MSPEs of both the OBP and the PBE predictors are larger than those of the EBLUPs in both the misspecified and correctly specified cases. This is not quite consistent with the results reported by Jiang et al., 2011, where the OBP outperformed the EBLUPs in the misspecified case. As for the PBE, its MSPE is about 1.5% larger than that of the OBP in the misspecified case, and the PBE and the OBP have about the same MSPE for the correct model.

Table 2.5: Empirical MSPEs of the six predictors (multiplied by 100) with percentage increases over PBE reported in parentheses (second simulation study). The correct model is quadratic. Misspecified model is linear.

Model	**ML**	**REML**	**FH**	**PR**	**OBP**	**PBE**
MIS	3.134	3.128	3.131	3.133	3.198	3.245
	(−3.42%)	(−3.61%)	(−3.51%)	(−3.45%)	(−1.45%)	
Correct	2.138	2.074	2.077	2.099	2.216	2.207
	(−3.13%)	(−6.03%)	(−5.89%)	(−4.89%)	(0.41%)	

Table 2.6: Empirical coverage probabilities of 95% prediction/credible intervals for θ_i's with average lengths of the intervals (second simulation study). The correct model is quadratic. Misspecified model is linear.

Model		**ML**	**REML**	**FH**	**PR**	**OBP**	**PBE**
MIS	CP	0.951	0.951	0.951	0.951	0.948	0.932
	AL	0.147	0.147	0.147	0.147	0.147	0.139
	IS	0.171	0.172	0.171	0.171	0.171	0.164
Correct	CP	0.939	0.939	0.940	0.947	0.943	0.887
	AL	0.121	0.119	0.119	0.122	0.126	0.104
	IS	0.148	0.145	0.145	0.147	0.152	0.132

Next, we compute the empirical coverage probabilities of 95% prediction/credible intervals for θ_i's by averaging over the areas. The coverage probabilities are reported in Table 2.6. Our simulations suggest for the correct model that the PBE credible intervals are substantially narrow (at least 15% narrower than the counterparts of the five frequentist predictors), resulting in average frequentist coverage probabilities over the areas which is about 7% lower than the nominal value. Even though we created all our PBE credible intervals based on lower and upper 2.5% quantiles, as previously discussed in Subsection 4.1 in the main manuscript that the length of the credible intervals are proportional to posterior SDs. The frequentist undercoverage of the PBE credible intervals is caused by negative bias of the posterior variance being used as an estimator of the frequentist MSE of the PBE predictor. Indeed our simulations show that under the correct model the empirical total MSPE of the PBE predictors is substantially (about 35%) larger than the empirical average of the posterior variances associated with the PBE predictors. Under the misspecified situation, this difference is less pronounced. Also, although the PBE intervals are subject to undercoverage, the extent of this undercoverage is not severe. All six methods have good average coverage probabilities in the misspecified case, with the PBE intervals exhibiting slightly more undercoverage and smaller average interval lengths than the other methods, but retaining the lowest empirical average interval score. Overall, the PBE method is a viable method which always yields positive measure of uncertainty and is to implement to obtain both point and interval estimates.

We also compare the four EBLUPs, the OBP and the PBE in terms of their biases (which are not reported in any of thables here). While all of them are nearly unbiased under the correct model, all of them display some bias under the misspecified model. We did not report individual area level biases here but the mean (over small areas) of absolute deviations (of the estimators from the values they are estimating) under the misspecified model for the EBLUP(ML), OBP and the PBE are 0.011, 0.012 and 0.014, respectively. Corresponding three means of absolute relative deviations (in percentages) are 2.9%, 3.7% and 4.2%, respectively.

2.5.3 Example 3: A simulation study imitating the median incomes data

Our final simulation study uses the setup of the median incomes estimation application. We use the parameter values $\boldsymbol{\beta} = (12000, -0.3, 0.8)$ and $A = 6 \times 10^6$, obtained by rounding the estimates in the application in Section 4.1 of main manuscript to generate data from the correct model that involves both the covariates x_1 and x_2, introduced earlier in the same section. We use the given D_i

values for sampling variances in our simulations. We introduce misspecification by excluding x_2 from the regression model.

Table 2.7: Empirical MSPEs (divided by 10^6) with percentage increases over PBE (median incomes application). The correct model is multiple linear regression with x_1 and x_2. The misspecified model excludes x_2.

	ML	REML	FH	PR	OBP	PBE
Mis	312.09	311.10	311.20	311.81	321.90	323.39
Percent	-3.49	-3.80	-3.77	-3.58	-0.46	
Correct	181.82	180.88	181.32	185.75	195.91	195.29
Percent	-6.90	-7.38	-7.15	-4.88	0.32	

Table 2.7 shows that the OBP and the PBE methods perform similarly, and total empirical MSPEs of the EBLUPs are lower than those for the OBP and the PBE methods by between 5.0% and 7.5% for the correct model, and between 3.5% and 3.8% for the misspecified model.

As in the last example, we also studied biases of the OBP and the PBE point estimates. While the biases are nearly zero under the correct model, for the misspecified model there are moderate biases (two means of absolute relative deviations, in percentages, are 2.1% and 2.2%, respectively).

Table 2.8 reports the empirical coverage probabilities of 95% prediction/credible intervals for θ_i's, empirical lengths and interval scores, averaged over the areas. These results show as in the earlier application a similar pattern of undercoverage for the PBE when the model is correctly specified. All six methods have good average coverage probabilities in the misspecified case.

To gain a further insight into the performances of these methods for individual small areas we provide boxplots of interval scores in Figure 2.8. These scores are computed from equation 2.13. We included only four methods in the plot, two EBLUPs (the best performing ML and the worst performing PR) and the alternative OBP and the PBE methods. Without going into elaboration, we expect that readers will agree that the proposed PBE method is a viable alternative to the considered frequentist methods.

Table 2.8: Mean empirical coverage probabilities, average empirical mean lengths and average empirical interval scores of approximate 95% prediction intervals (median income).

		ML	REML	FH	PR	OBP	PBE
Mis	CP	0.95	0.95	0.95	0.95	0.95	0.94
	AL	9488	9485	9515	9554	9837	9221
	IS	10641	10693	10718	10761	11010	10536
Correct	CP	0.95	0.95	0.95	0.95	0.95	0.92
	AL	7537	7534	7537	7654	7828	6968
	IS	8932	8895	8923	9179	9221	8508

2.6 Conclusions

This paper proposes a pseudo-Bayesian alternative (PBE) to OBP. We use independent priors on the model parameters $\boldsymbol{\beta}$ and A. We find a nearly uniform prior for $\boldsymbol{\beta}$. That structure leads to a tractable but minimally proper prior for A. We implement our PBE method by simulating samples from the pseudo-posterior distribution. Based on samples from the posterior distribution we computed posterior means and variances, and constructed credible intervals of the small area means.

Using real data examples and simulations, we demonstrated that the PBE is very similar to or better than the OBP in terms of overall and area-specific MSPEs. One major advantage of the PBE over the OBP is that the PBE automatically has a nonnegative measure of uncertainty, whereas the second-order unbiased MSPE estimator for the OBP proposed by Y. Liu et al., 2022 is technically difficult and can occasionally take negative values.

The data examples and simulations show the pseudo-Bayesian method is competitive in terms of frequentist criteria. Its credible intervals are somewhat narrow and achieve smaller than nominal coverage under the correct model, but they are nearly correct for misspecified models. The PBE method has good Bayesian and frequentist properties.

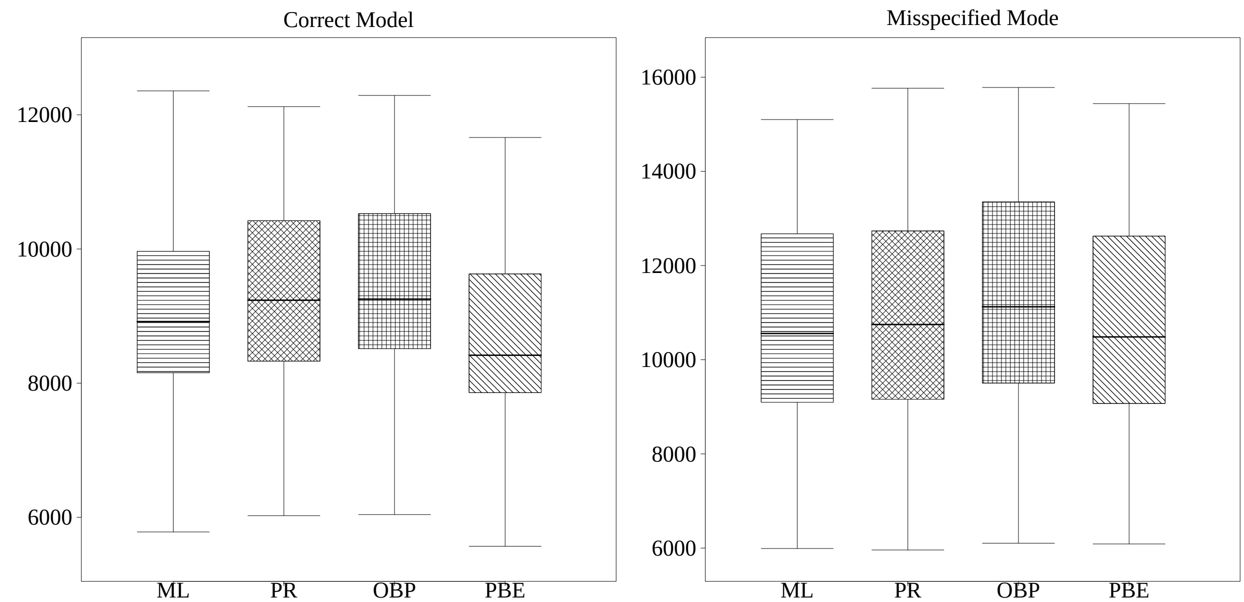
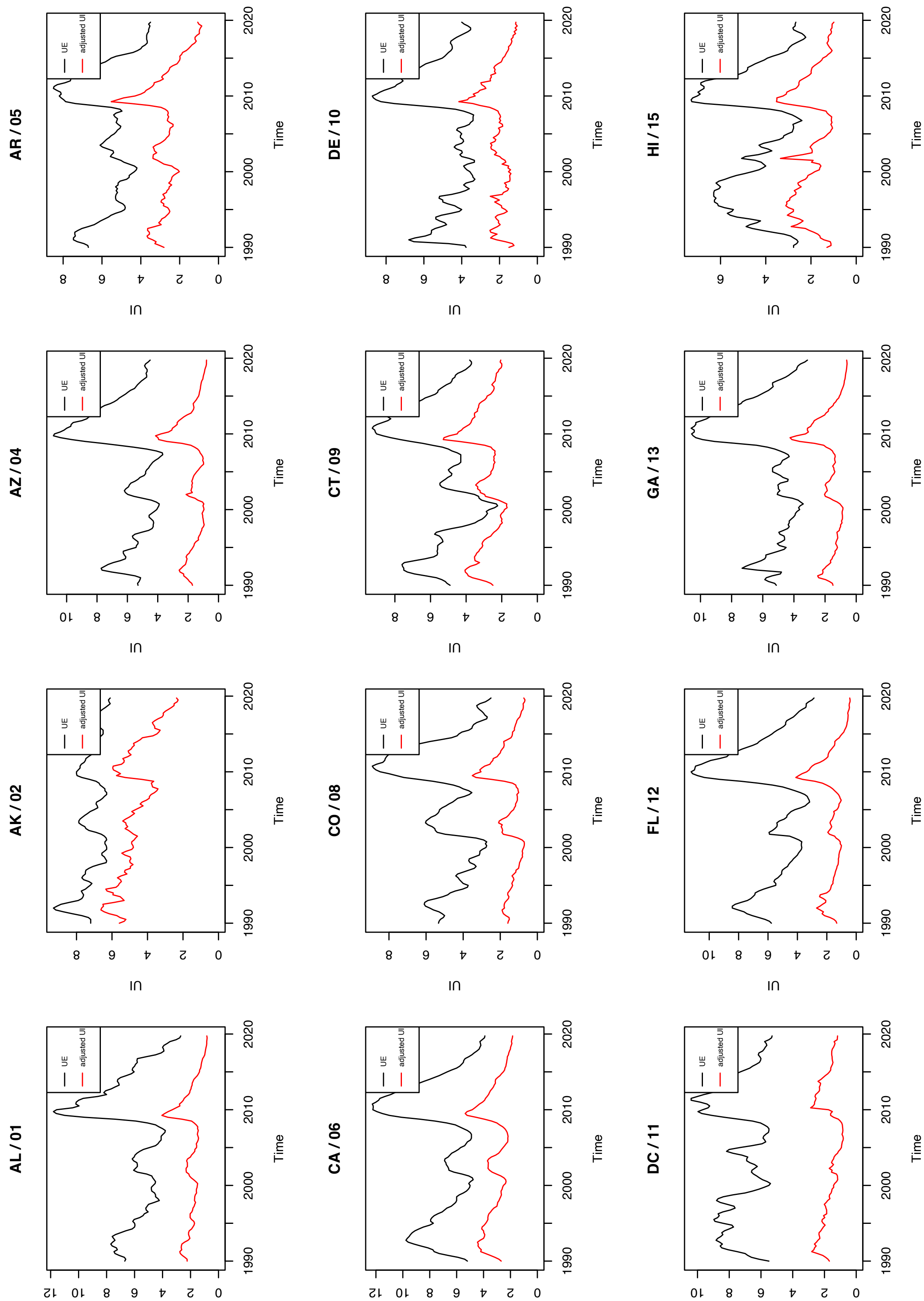
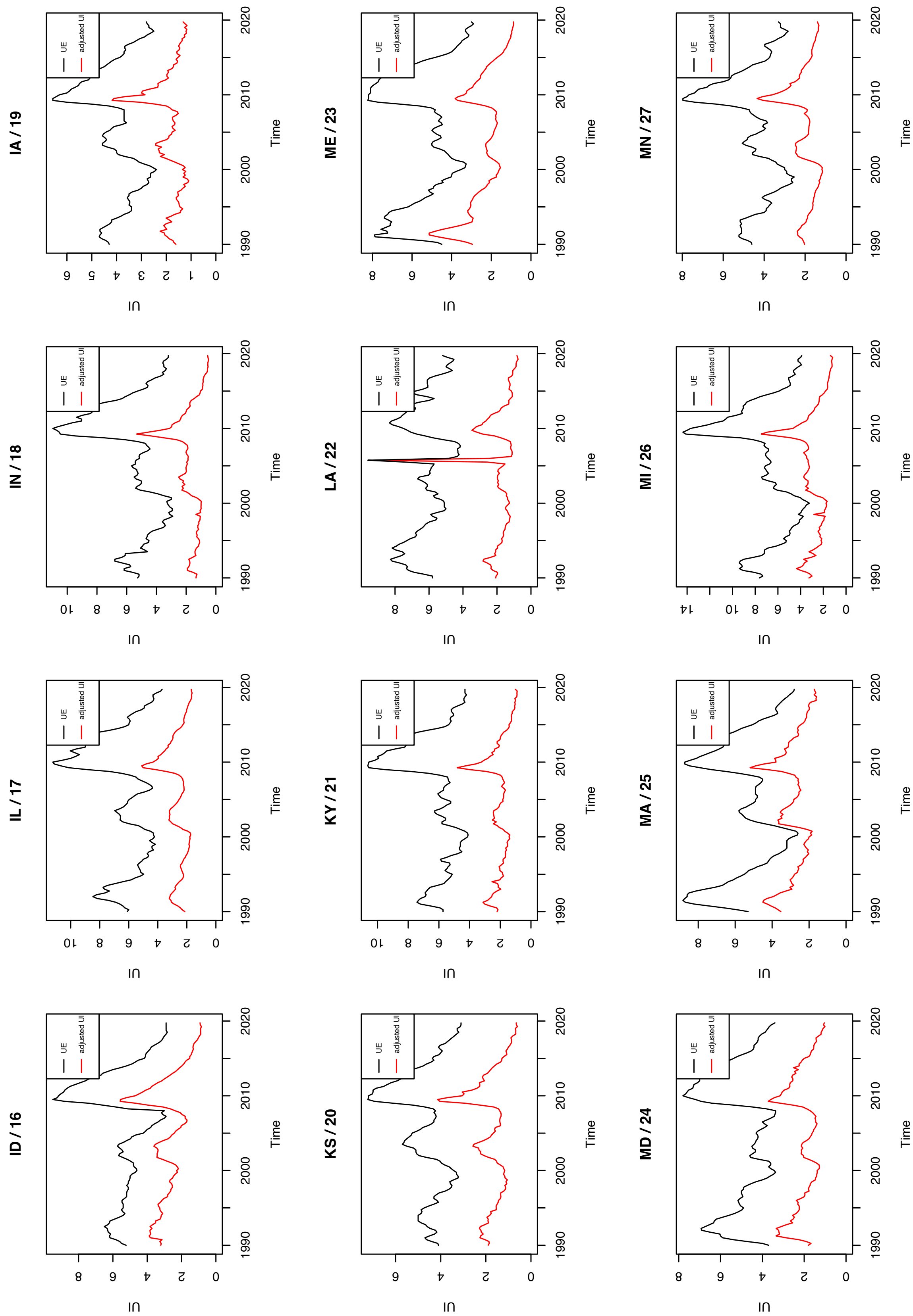
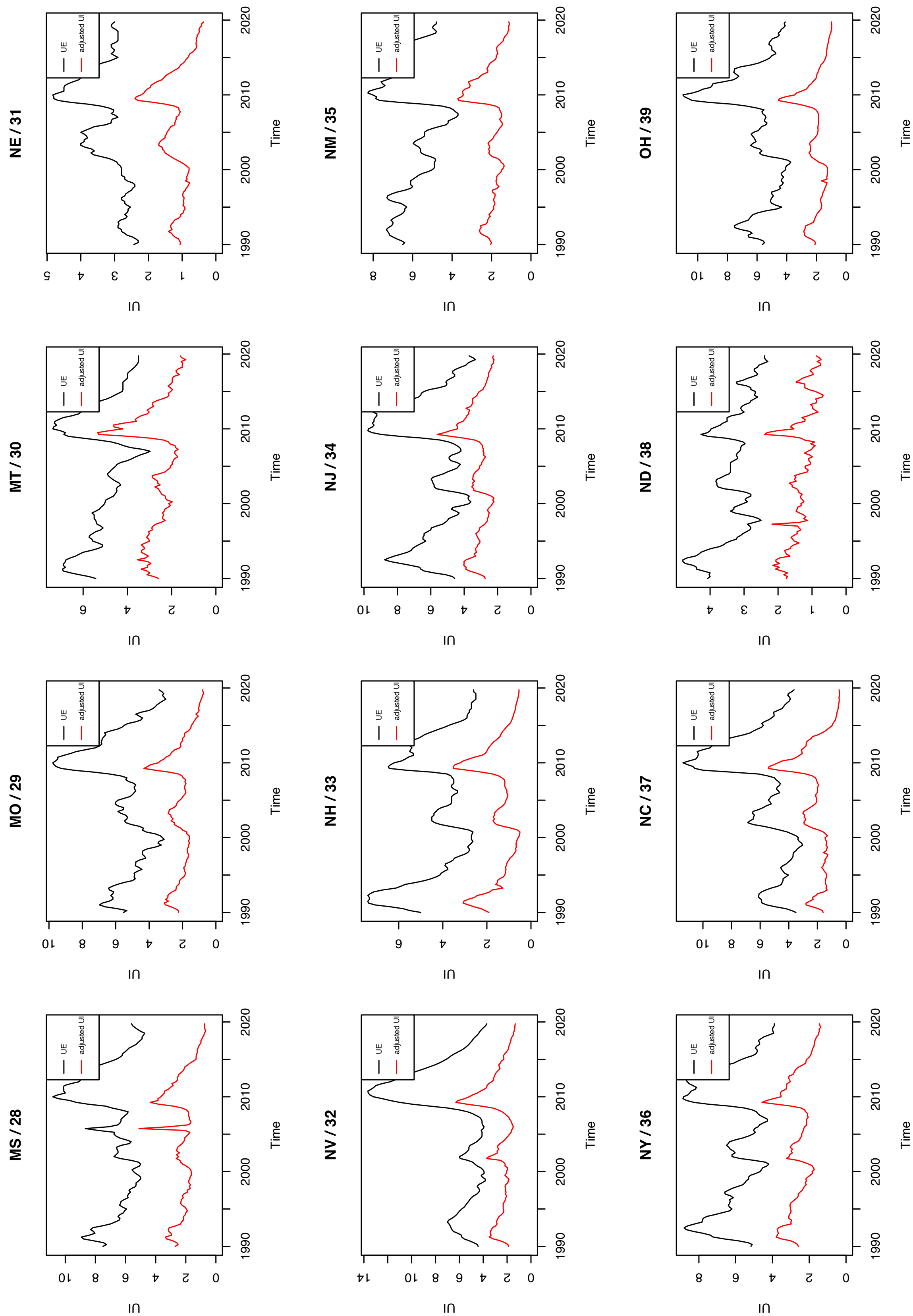


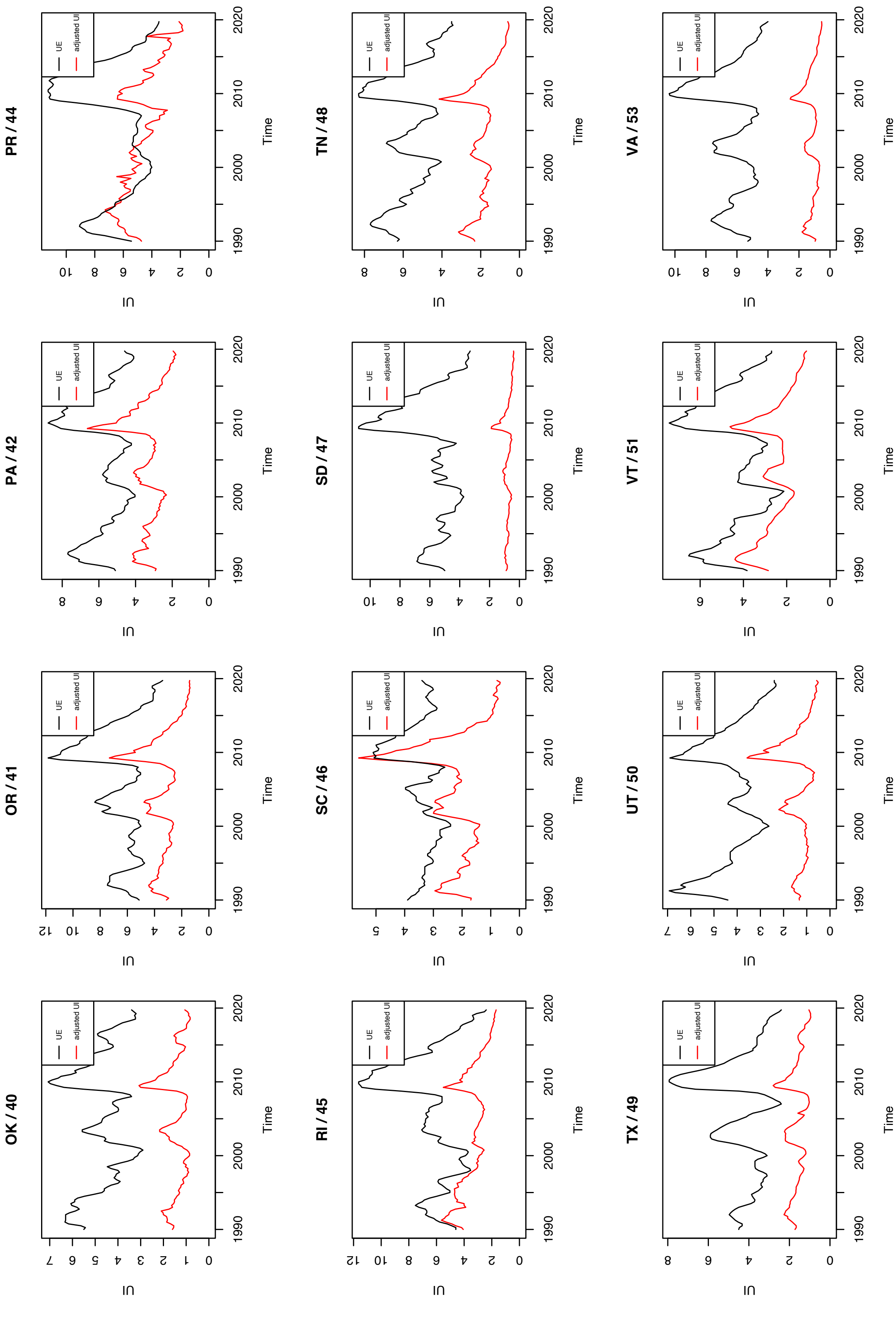
Figure 2.8: Boxplot for interval scores for the median incomes simulation. Left panel for the correct model, right panel for the misspecified model. Within each boxplot from left to right: ML, PR, OBP, and PBE.

Appendix A


AL / 01
AK / 02
AZ / 04
AR / 05
CA / 06
CO / 08
CT / 09
DE / 10
DC / 11
FL / 12
GA / 13
HI / 15
UE
adjusted UI
Time
UI


ID / 16
IL / 17
IN / 18
IA / 19
KS / 20
KY / 21
LA / 22
ME / 23
MD / 24
MA / 25
MI / 26
MN / 27
UE
adjusted UI
Time
UI


MS / 28
MO / 29
MT / 30
NE / 31
NV / 32
NH / 33
NJ / 34
NM / 35
NY / 36
NC / 37
ND / 38
OH / 39
UE
adjusted UI
UI
Time



OK / 40
OR / 41
PA / 42
PR / 44
RI / 45
SC / 46
SD / 47
TN / 48
TX / 49
UT / 50
VT / 51
VA / 53
UE
adjusted UI
UI
Time

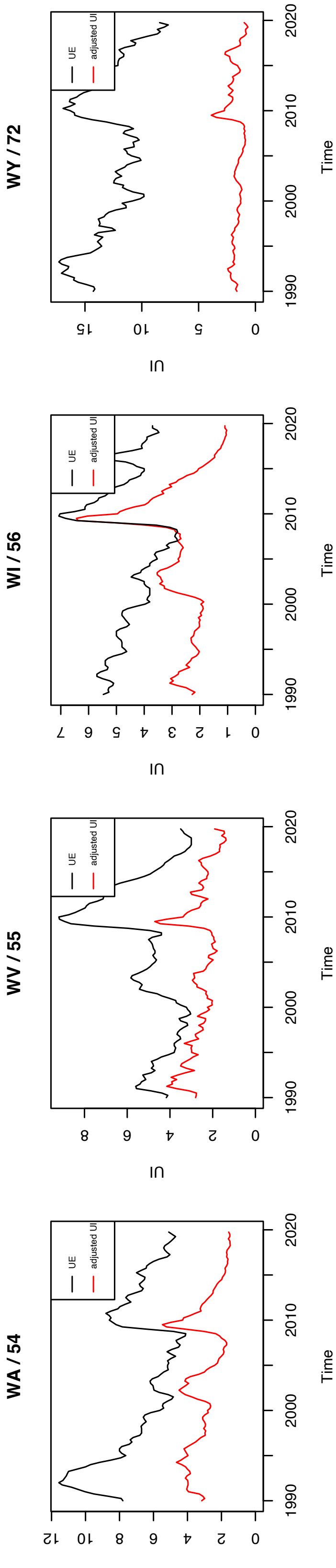

WA / 54
WV / 55
WI / 56
WY / 72
UE
adjusted UI
UI
Time

Table A.1: Bigtime method of Four steps ahead Forecasting Results for Group 1

state	step1	step2	step3	step4	MAPE
1:ME	0.413	4.514	5.064	2.029	3.005
2:ND	3.649	4.306	5.523	6.809	5.071
3:DE	7.468	8.443	11.837	11.986	9.934
4:MS	9.800	15.187	21.142	24.022	17.538
5:LA	16.933	18.040	17.425	17.109	17.377
6:ID	3.511	1.036	2.574	7.469	3.648
7:NM	2.008	2.988	7.081	8.646	5.181
8:AK	2.510	2.206	1.634	1.165	1.879
9:HI	1.022	10.919	19.448	23.337	13.681
10:WA	2.714	0.631	6.472	6.114	3.983
Sum	50.027	68.269	98.200	108.687	81.297

Table A.2: Bigtime method of Four steps ahead Forecasting Results for Group 2

state	step1	step2	step3	step4	MAPE
1:NJ	1.791	3.968	12.083	20.623	9.616
2:IN	2.496	3.501	3.695	7.485	4.294
3:MO	5.387	10.678	12.646	11.595	10.076
4:NE	6.414	4.495	5.124	6.217	5.563
5:NC	1.292	6.122	13.911	13.350	8.669
6:SC	0.855	2.379	4.815	6.336	3.596
7:KY	2.274	1.013	6.514	6.736	4.134
8:AR	1.693	0.787	4.146	6.726	3.338
9:CA	2.771	0.336	4.606	9.451	4.291
10:AL	1.740	0.718	4.539	9.750	4.187
11:OK	0.604	5.521	10.447	12.505	7.269
12:MT	3.905	5.519	6.974	9.651	6.512
13:WY	0.840	4.137	11.618	10.380	6.744
Sum	32.061	49.174	101.119	130.804	78.289

Table A.3: Bigtime method of Four steps ahead Forecasting Results for Group 3

state	step1	step2	step3	step4	MAPE
1:CT	1.492	0.770	7.503	10.847	5.153
2:IL	4.743	5.336	3.242	2.544	3.966
3:MA	2.176	4.016	10.387	15.761	8.085
4:NH	1.441	4.206	11.390	14.861	7.975
5:RI	1.686	4.175	2.203	1.601	2.416
6:VT	7.671	9.267	5.564	4.490	6.748
7:NY	2.109	1.210	9.563	17.107	7.497
8:PA	7.654	10.071	7.940	5.671	7.834
9:UT	1.156	3.744	10.293	15.499	7.673
10:MI	4.537	6.655	7.236	8.223	6.663
Sum	34.665	49.452	75.320	96.604	64.010

Table A.4: Bigtime method of Four steps ahead Forecasting Results for Group 4

state	step1	step2	step3	step4	MAPE
1:OH	9.150	9.444	10.095	10.687	9.844
2:KS	6.227	3.790	1.015	2.026	3.264
3:MN	3.379	1.464	2.861	10.951	4.664
4:VA	2.766	4.195	11.041	14.998	8.250
5:DC	7.045	13.737	17.796	15.771	13.587
6:TN	8.087	5.012	1.094	0.391	3.646
7:NV	3.203	2.698	9.564	14.905	7.592
8:WI	0.498	4.738	5.539	9.046	4.955
9:IA	8.302	9.497	11.158	9.992	9.737
10:SD	0.866	0.595	4.321	9.320	3.775
11:MD	0.403	2.451	3.684	3.544	2.520
12:WV	3.825	0.328	3.527	6.000	3.420
13:CO	4.708	1.009	1.257	2.648	2.405
Sum	58.459	58.958	82.952	110.279	77.659

Table A.5: Bigtime method of Four steps ahead Forecasting Results for Group 5

state	step1	step2	step3	step4	MAPE
1:FL	1.604	0.238	4.507	9.353	3.925
2:GA	1.925	7.444	9.376	12.279	7.756
3:TX	1.198	2.999	8.480	17.133	7.453
4:AZ	0.275	3.359	10.486	14.399	7.130
5:OR	1.165	6.027	10.171	16.547	8.477
Sum	6.168	20.067	43.019	69.711	34.741

Table A.6: Bigtime method of Four steps ahead Forecasting Results for Region 1

state	step1	step2	step3	step4	MAPE
1:CT	2.746	1.435	4.822	8.148	4.288
2:ME	0.101	2.148	2.455	0.471	1.293
3:MA	1.296	2.922	8.259	13.578	6.514
4:NH	1.036	2.926	9.541	12.997	6.625
5:RI	2.406	5.649	3.828	3.226	3.777
6:VT	9.457	12.970	9.912	8.873	10.303
7:NJ	0.626	1.663	8.693	17.027	7.002
8:NY	3.030	0.044	6.860	14.251	6.046
9:PA	5.928	8.860	7.515	5.239	6.885
Sum	26.626	38.616	61.885	83.809	52.733

Table A.7: Bigtime method of Four steps ahead Forecasting Results for Region 2

state	step1	step2	step3	step4	MAPE
1:IL	4.009	1.861	0.308	1.023	1.800
2:IN	4.882	5.855	6.036	9.724	6.624
3:MI	1.664	2.954	3.587	4.640	3.211
4:OH	6.787	7.108	7.796	8.424	7.529
5:WI	1.016	4.197	4.998	8.490	4.675
6:IA	7.474	8.683	10.362	9.189	8.927
7:KS	5.839	3.395	0.613	2.435	3.071
8:MN	1.677	0.249	4.621	12.819	4.842
9:MO	7.878	13.024	14.930	13.891	12.431
10:NE	9.606	7.771	8.374	9.408	8.790
11:ND	7.454	8.145	9.362	10.648	8.902
12:SD	1.066	0.392	4.514	9.503	3.869
Sum	59.354	63.634	75.500	100.196	74.671

Table A.8: Bigtime method of Four steps ahead Forecasting Results for Region 3

state	step1	step2	step3	step4	MAPE
1:DE	4.349	5.348	8.848	8.997	6.885
2:FL	4.170	2.345	1.851	6.613	3.745
3:GA	10.693	15.596	17.243	19.792	15.831
4:MD	1.574	0.441	1.660	1.534	1.302
5:NC	2.788	1.874	9.387	8.880	5.732
6:SC	5.812	7.241	9.548	10.995	8.399
7:VA	3.052	3.891	10.720	14.668	8.083
8:DC	9.811	16.300	20.234	18.261	16.151
9:WV	5.791	1.716	1.425	3.855	3.197
10:AL	6.274	3.924	0.268	4.711	3.794
11:KY	8.223	5.120	0.064	0.287	3.424
12:MS	13.600	18.770	24.485	27.249	21.026
13:TN	14.023	11.116	7.409	6.705	9.813
14:AR	4.023	1.583	1.717	4.253	2.894
15:LA	17.334	18.439	17.828	17.515	17.779
16:OK	7.562	2.978	1.643	3.644	3.957
17:TX	4.549	0.463	4.876	13.295	5.796
Sum	123.626	117.145	139.206	171.251	137.808

Table A.9: Bigtime method of Four steps ahead Forecasting Results for Region 4

state	step1	step2	step3	step4	MAPE
1:AZ	2.197	1.389	8.407	12.278	6.068
2:CO	4.416	0.705	1.568	2.964	2.413
3:ID	4.922	0.441	1.076	5.898	3.084
4:MT	0.596	2.141	3.547	6.149	3.108
5:NV	3.742	2.139	8.983	14.310	7.293
6:NM	2.687	2.284	6.360	7.925	4.814
7:UT	2.902	10.774	17.578	22.930	13.546
8:WY	1.674	3.262	10.682	9.455	6.268
9:AK	0.831	0.497	0.113	0.616	0.514
10:CA	2.999	0.569	4.363	9.199	4.282
11:HI	1.375	13.548	22.216	26.131	15.818
12:OR	1.226	3.548	7.623	13.876	6.568
13:WA	3.931	0.624	5.151	4.803	3.627
SUM	33.499	41.919	97.668	136.534	77.403

Bibliography

Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. *Selected papers of hirotugu akaike* (pp. 199–213). Springer.

Albert, J., & Chib, S. (1993). Bayesian inference of autoregressive time series with mean and variance subject to markov jumps," *Journal of Business and Economic Statistics*, *11*, 1–15.

Anderson, T. W. (1978). Repeated measurements on autoregressive processes. *Journal of the American Statistical Association*, *73*(362), 371–378.

Azzalini, A. (1981). Replicated observations of low order autoregressive time series. *Journal of Time Series Analysis*, *2*(2), 63–70.

Bissiri, P. G., Holmes, C. C., & Walker, S. G. (2016). A general framework for updating belief distributions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *78*(5), 1103–1130.

Bissiri, P. G., & Walker, S. G. (2019). On general bayesian inference using loss functions. *Statistics & Probability Letters*, *152*, 89–91.

BLS. (2015). How the government measures unemployment [Accessed: 2021-12-29].

Box, G. E., & Jenkins, G. M. (1976). Time series analysis. forecasting and control (rev. ed.)

Broemeling, L. D., & Shaarawy, S. (1986). A bayesian analysis of time series. *Bayesian Inference and Decision Techniques. Elsevier Science Publishers BV.*

Chen, C. W., & Lee, J. C. (1995). Bayesian inference of threshold autoregressive models. *Journal of time series analysis*, *16*(5), 483–492.

Chib, S., & Greenberg, E. (1994). Bayes inference in regression models with arma (p, q) errors. *Journal of Econometrics*, *64*(1-2), 183–206.

Datta, G. S., & Lahiri, P. (2000). A unified measure of uncertainty of estimated best linear unbiased predictors in small area estimation problems. *Statistica Sinica*, *10*, 613–627.

Datta, G. S., Lahiri, P., Maiti, T., & Lu, K. L. (1999). Hierarchical bayes estimation of unemployment rates for the states of the us. *Journal of the American Statistical Association*, *94*(448), 1074–1082.

Datta, G. S., Rao, J. N. K., & Smith, D. D. (2005). On measuring the variability of small area estimators under a basic area level model. *Biometrika*, *92*(1), 183–196.

Efron, B., & Morris, C. (1972). Empirical bayes on vector observations: An extension of stein's method. *Biometrika*, *59*(2), 335–347.

Fay, R. E. (1987). Application of multivariate regression to small domain estimation. *in R. Platek, J.N.K. Rao, C.-E. Sarndal, and M.P. Singh (Eds.), Small Area Statistics*, 91–102.

Fay, R. E., & Herriot, R. A. (1979). Estimates of income for small places: an application of James-Stein procedures to census data. *Journal of the American Statistical Association*, *74*(366a), 269–277.

Galbraith, R., & Galbraith, J. (1974). On the inverses of some patterned matrices arising in the theory of stationary time series. *Journal of applied probability*, *11*(1), 63–71.

Gelfand, A. E., & Smith, A. F. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American statistical association*, *85*(410), 398–409.

Gelman, A., Gilks, W. R., & Roberts, G. O. (1997). Weak convergence and optimal scaling of random walk metropolis algorithms. *The annals of applied probability*, *7*(1), 110–120.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical science*, *7*(4), 457–472.

Ghosh, M. (1992). Hierarchical and empirical bayes multivariate estimation. *in M. Ghosh and P.K. Pathak (Eds.), Current Issues in Statistical Inference: Essays in Honor of D. Basu, IMS Lecture Notes - Monograph Series*, *Volume 17*, Beachwood, OH: Institute of Mathematical Statistics.

Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American statistical Association*, *102*(477), 359–378.

Hastings, W. K. (1970). Monte carlo sampling methods using markov chains and their applications.

James, W., & Stein, C. (1961). Estimation with quadratic loss. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.*, *1*, 361–379.

Jiang, J., Nguyen, T., & Rao, J. S. (2011). Best predictive small area estimation. *Journal of the American Statistical Association*, *106*(494), 732–745.

Lindley, D. V., & Smith, A. F. (1972). Bayes estimates for the linear model. *Journal of the Royal Statistical Society: Series B (Methodological)*, *34*(1), 1–18.

Liu, X., Ma, H., & Jiang, J. (2022). That prasad-rao is robust: Estimation of mean squared prediction error of observed best predictor under potential model misspecification. *Statistica Sinica*, *32*, 2217–2240.

Liu, Y., Liu, X., Pan, Y., Jiang, J., & Xiao, P. (2022). An empirical comparison of various mspe estimators and associated prediction intervals for small area means. *Journal of Statistical Computation and Simulation*.

Marin, J.-M., Mengersen, K., & Robert, C. P. (2005). Bayesian modelling and inference on mixtures of distributions. *Handbook of statistics*, *25*, 459–507.

Marriott, J., Ravishanker, N., & Gelfand, A. E. (1993). *Bayesian analysis of arma processes: Complete sampling based inference under full likelihoods* (tech. rep.). STANFORD UNIV CA DEPT OF STATISTICS.

McCulloch, R., & Tsay, R. (1993). Bayesian analysis of threshold autoregressive processes with a random number of regimes. *Computing Science and Statistics*, 253–253.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *The journal of chemical physics*, *21*(6), 1087–1092.

Monahan, J. F. (1983). Fully bayesian analysis of arma time series models. *Journal of Econometrics*, *21*(3), 307–331.

Montgomery, A. L., Zarnowitz, V., Tsay, R. S., & Tiao, G. C. (1998). Forecasting the us unemployment rate. *Journal of the American Statistical Association*, *93*(442), 478–493.

Morris, C. N., & Christiansen, C. L. (1996). Hierarchical models for ranking and for identifying extremes, with applications. *Bayesian Statistics 5*, New York: Oxford University Press, 277–296.

Prasad, N. G. N., & Rao, J. N. K. (1990). The estimation of the mean squared error of small-area estimators. *Journal of the American Statistical Association*, *85*(409), 163–171.

Rao, J. N. K., & Molina, I. (2015). *Small area estimation*. Hoboken, NJ: John Wiley & Sons, Inc.

Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, 461–464.

Tanner, M. A., & Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American statistical Association*, *82*(398), 528–540.

Tierney, L. (1994). Markov chains for exploring posterior distributions. *the Annals of Statistics*, 1701–1728.

Tiller, R. B. (1992). Time series modeling of sample survey data from the us current population survey. *Journal of Official Statistics*, *8*(2), 149–166.

Wilms, I., Basu, S., Bien, J., & Matteson, D. S. (2021). Sparse identification and estimation of large-scale vector autoregressive moving averages. *Journal of the American Statistical Association*, 1–12.